We appreciate your time to read our manuscript and give us these comments. Below please find our reply.

Anonymous Referee #1

1 General Comments

This manuscript presents an evaluation of several aerosol-optical depth products derived from satellite sensor data with ground-based observations in a region of East Asia. The text is well-structured and mostly well-written. However, in my view, two major items remain to be addressed in a revision:

- Several choices in study design are not fully explained and require additional justification (see below).
- It appeared to me that not all numerical preconditions for correlation and regression analyses, both central to the presented study, were met in all situations. Also, statistical significance of regression was not tested for. Details below.

Response: According to your comments in the "individual issues/questions" section, we added some description and explanation of our data processing and analyzing method. We hope this information can help you evaluate our study.

2 Individual Issues/Questions

- 20710-24 (henceforth "10-24" etc.): Is the bias systematic?

Response: There is evidence that this positive bias includes both random error and systematic error due to improper characterization of surface reflectance, uncertainties in the assumed aerosol model, and cloud masking. The 3 km MODIS products sample fewer reflectance pixels to retrieve aerosol pixels relative to the 10 km products, introducing sporadic extreme values of AOD that are avoided more successfully by the 10 km products. Previous studies also indicated that this positive bias in urban areas resulted from improper characterization of bright urban surfaces, a known difficult situation for the Dark Target algorithm (Munchak et al., 2013;Remer et al., 2013). The VIIRS IP product is retrieved at the reflectance pixel level without aggregation, so it is expected to include more noise. Moreover, VIIRS IP is also affected by factors that impede the Dark Target algorithm; thus, this positive bias is due to both random error and algorithm issues. We added the following sentences to explain this on page 21, line 25- page 22, line 6: "There is evidence that this positive bias includes systematic errors due to improper characterization of surface reflectance, uncertainties in the assumed aerosol model, and cloud masking. The 3 km MODIS products sample fewer reflectance pixels to retrieve aerosol pixels relative to the 10 km products, introducing sporadic unrealistic high AOD retrievals that are avoided more successfully by the 10 km products (Munchak et al., 2013). Previous studies also reported that improper characterization of bright urban surfaces, a known difficult situation for

the Dark Target algorithm, led to positive bias in urban/suburban regions (Munchak et al., 2013; Remer et al., 2013). The VIIRS IP product is retrieved at the reflectance pixel level without aggregation, thus it is expected to include more noise.”

• 13-12: what does a value of -0.1 indicate? This should probably not be referred to as a "value".

Response: Since the MODIS Collection 5 algorithm, negative retrieval values have been allowed. In this study, both MODIS C6 3 km and GOCI aerosol products have negative retrievals. Though such a negative AOD value is not physically possible, it statistically represents small positive AOD values in the overall data distribution. In other words, removing these negative values would in fact truncate the lower tail of the AOD distribution. Most previous evaluation studies include these negative retrievals as valid values (Levy et al., 2013;Munchak et al., 2013;Remer et al., 2013); thus, we also included these negative values in our analyses.

• 14-7: All AERONET observations are point observations. In evaluating the accuracy of the satellite products, why would there be a need for spatially continuous ground-based observations? I would expect a multi-temporal evaluation using a wide range of AERONET stations to allow for a fairly representative assessment of product quality. Or do you expect distinct spatial patterns in the satellite products? As this point is the central motivation for this study as I understand it, I suggest that you elaborate your argument in this respect.

Response: Thank you for this suggestion. Since we aim to evaluate the performance of high-resolution satellite aerosol products, we need intensive ground observations as “ground truth”. For example, the spatial resolution of MODIS C6 3 km products is 3 km, thus in theory the variation in aerosol loading at two locations that are 3 km apart can be detected by these products. However, if the ground stations are 10 km away from each other, we cannot validate this 3 km product at their designed resolution; even if these products can detect the aerosol loading variability across stations separated by 10 km, they may not perform well at a 3-km resolution. The DRAGON-Asia Campaign provides intensive ground measurements and makes it possible to validate these high-resolution satellite aerosol products. To make the motivation of this study clear, we modified these sentences as follows on page 6, line 20–25: “Evaluation of satellite aerosol products’ ability to track small-scale aerosol spatial variability is limited due to lack of intensive ground observations of AOD: the permanent AERONET stations can be tens or even hundreds of kilometers apart, leading to insufficient information of the small-scale horizontal distribution of aerosol loading that is required for a precise evaluation at high resolution.”

• 14-8: A point observation at the ground does not have a ’spatial resolution’ at all. You may be referring to the distance between observations. Please clarify, and change the terminology here and elsewhere.

Response: We modified the sentences as follows on page 6, line 25–28: “In response to the lack of intensive ground AOD observations, AERONET conducted several campaigns, which deployed additional temporary sunphotometers in selected regions and provided valuable information of

small-scale AOD distribution."

• 15-14: What does "high quality" refer to in EDR/IP?

Response: There are several quality assurance steps in the retrieval process and both EDR and IP are assigned quality flags of "high", "degraded", or "low", indicating the confidence of retrievals. The "high" quality AOD is suggested for scientific research and applications by the VIIRS aerosol products team. We added the following sentence in the revised manuscript on page 5, line 9–10: "Detailed description of the quality assurance of VIIRS aerosol products is documented by Liu et al. (2014)." We also modified the sentence as follows on page 8, line 3-4: "Thus, only EDR and IP pixels from May 2012 to June 2013 with high quality (Quality Flag = "high") were processed."

• 16-3: "AERONET stations" do not "measure AOD". Please increase precision of statement.

Response: Thank you for this suggestion. We modified this statement through the manuscript, e.g. we changed "AOD measurements" to "AOD observations".

• 16-9: How can anyone "assure" the quality? Did you perchance mean "quality assessed"?

Response: The phrase "quality assured" is developed by the AERONET science team and is widely used in related articles. Quality-assured data (Level 2.0) have both pre- and post- deployment calibration, leading to uncertainty of about 0.01–0.02. We added the following sentence on page 8, line 23–26: "The Level 2.0 (quality assured) AOD data have both pre- and post-deployment calibration, leading to an uncertainty of about 0.01–0.02 while the Level 1.5 AOD data are cloud-screened but not quality-assured (Otter et al., 2002)."

• 16-27: Why did you reproject the data? This will certainly lead to sampling induced errors!

Response: The original satellite aerosol products are in a geographic coordinate system with latitude and longitude information; however, to build a 3-km/6-km fixed grid, we need to covert the latitude/longitude coordinate system to a projected coordinate system. To clarify the data processing method, we modified the sentence as follows on page 9, line 18: "All the data were converted to the JGD_2000_UTM_Zone_52N coordination system."

• 16-27: Please give details of the averaging/pixel combination method used in the reprojection process.

Response: In the reprojection process, we did not conduct any averaging or pixel combination. This process was basically conducted using ArcGIS to convert satellite and ground AOD data from a latitude/longitude geographic coordinate system to a projected coordinate system, thus allowing us to match satellite retrievals and ground stations based on the distance between them.

• 16-28: What do you mean by "data integration"?

Response: The data integration here means to spatially join satellite retrievals with ground stations based on their locations and to develop collections of coincident satellite–ground AOD pairs for following comparisons. We modified the sentences as follows in this section to clarify this data processing step: "All the data were converted to the JGD_2000_UTM_Zone_52N coordination system. For the matchup process, a 6-km grid and a 3-km grid covering the whole study domain were constructed, corresponding to the spatial resolution of each satellite product. Satellite aerosol data from different sensors were mapped and spatially joined to this 6-km grid (for VIIRS EDR and GOCI products) or 3 km grid (for VIIRS IP and MODIS C6 3 km products) to construct coincident satellite-ground AOD pairs."

• 16-28: Why would "data integration" be necessary? Why not leave all data at their original aspects and resolutions and compare them based on location alone?

Response: The data integration, or the matchup process was necessary because it provided satellite–ground AOD pairs for following comparisons. This process did not change the original aspects or resolutions of the data, and our comparisons were based on location and time.

• 17-12: "maximum sample size" - in what respect?

Response: Since the DRAGON-Asia Campaign and the Beijing sampling experiment were conducted in different time periods, if we used the overlapping periods of these two experiments, we would lose many ground observations, leading to an insufficient sample size. To include the maximum number of ground observations, we allowed the spatial comparisons in Beijing and the Japan − South Korea region to differ in time periods. To clarify this, we modified the sentence as follows on page 10, line 12–16: "Temporal comparisons and spatial comparisons differ in study periods (Table 2): the temporal comparison period was the longest overlap period covered by all five satellite products and the spatial comparison periods in Beijing and the Japan–South Korea region are different in order to include the maximum number of ground observations."

• 17-17: Why did you average 3x3 grid cell environments if your main aim was to assess the quality of high-spatial-resolution data?

Response: We developed two comparison methods in this study: the temporal comparison and the spatial comparison. For the spatial comparison, the intensive ground observations from the DRAGON-Asia Campaign and the Beijing Sampling Experiment provided sufficient satellite–ground AOD pairs to validate satellite aerosol products performance at their designed resolution. For the temporal comparison, since we aimed to validate the ability of satellite aerosol products to track the day-to-day variation of aerosol to improve coverage and benefit from a collection of AOD retrievals (Ichoku et al., 2002), we used average AOD from the 3x3 grid cell buffer. This average method is widely used in previous evaluation studies.

• 18-6: How do you choose a 4x4 pixel window? Do you use the coordinates of the point between the four central pixels for comparison with other data sets?

Response: We used the 3-km grid for comparisons with VIIRS IP data because we did not have intensive ground sampling data to create a 0.75-km grid. We did not choose a 4x4 pixel window, but the 3-km grid cell sampling buffer cover a roughly 4x4 pixel window. To make this clear, we modified sentences as follows on page 11, line 21–25: "For the temporal comparison, we averaged valid IP AOD retrievals falling in the 3-km grid cell centered at each ground AERONET station and the mean and median CV were 0.33 and 0.25, respectively, within the 3 km grid cell buffer. This sampling buffer roughly covered a 4 × 4 pixel group."

• 18-8: "due to the lack of..." - I don't understand this argument. What do you mean by fine-resolution ground-based observations here? What would you ideal ground-based comparison data look like?

Response: The ideal ground-based observation for validation of the VIIRS IP aerosol product should be distributed roughly 0.75 km apart, thus we can test if the VIIRS IP AOD can detect variations in AOD at its 0.75-km nominal resolution. We modified this sentence as follows on page 11, line 26–28: "In the spatial comparison of VIIRS IP, we also used the 3-km sampling buffer due to a lack of more intensive ground AOD observations."

• 18-25: If the distortion towards the fringes of the pass impedes study results, why not use a dynamic spatial averaging approach that takes pixel size into account and tries to keep averaging area approximately constant, regardless of location and satellite system?

Response: Thank you for this suggestion. The major effect of the distortion towards the edge of the scan is that we may lose some ground–satellite AOD pairs because a satellite pixel at the edge of the swath covers a larger area than the nadir pixel size. However, there is no evidence that such missing will bias the comparison results. Using a dynamic spatial averaging approach, like kriging or interpolation, may introduce new error. Moreover, previous evaluation studies rarely used dynamic spatial averaging approach to fill the missing data due to the stretch. To make our results comparable with previous studies, we decided to use the method similar to previous studies.

• 19-9: "grid cell centered on the ground stations" - how does this apply to the 4x4 pixel averaging described above?

Response: This did not apply to IP data because we did not create a 0.75-km grid. To explain this, we added the following sentence in the revised manuscript on page 13, line 4–6: "Since we did not create a 750 -m grid for the VIIRS IP product, VIIRS IP-ground AOD pairs were assigned either "High Quality" or "Low Quality"."

• 19-28: Your figure 5 suggests that the data were used 'as-is'. A correlation analysis assumes normally distributed data, so in the case of AOD a logarithmic transformation would be required.

Did you perform this? If not, what is the rationale?

Response: There are a few important issues that go against log-transformation in the context of this study. First, due to the existence of valid negative AOD values, log transformation cannot be applied to MODIS and GOCI products directly. One solution is to add a fixed small positive number—e.g. 0.05—to both satellite retrievals and AERONET values; however, doing so changed the reference range of EE (± 0.05 ± 0.15 AOD) and made the evaluation metrics incomparable across different satellite aerosol products. Moreover, with log-transformation, linear regression intercepts and slopes lack clear physical meanings. Second, the distributions of AOD values from different sensors, as shown in Figure 4, were not significantly skewed. Due to the existence of small positive AOD values, log-transformation actually introduced slight skews to the left. Third, previous evaluation studies rarely used log-transformation (Levy et al., 2013;Liu et al., 2014;Munchak et al., 2013). Since one of the objectives of this study is to compare the performance of these emerging finer-resolution products in urban regions to their global evaluation results, log-transformation made the evaluation metrics incomparable with previous studies. All things considered, we decided to use the original data in this analysis.

- 21-25: I suggest moving this sentence to the discussion/conclusions section Results/analysis section: You analyse regression slopes and intercepts. I see two potential problems:

1. Like correlation, regression analysis assumes normally distributed date. If no log transformation of the AOD data was performed this condition is probably not met, statistically invalidating the analysis.
2. In regression analysis, a p value is always computed, indicating the probability that the results were purely due to random variation. It is commonly accepted practice to set a significance level before the analysis (e.g. 90%, 95% etc. probability of the relationship NOT being random) and then to discard all relationships outside that frame (p value gt; 0.1, 0.05 etc.) as not statistically significant. A slope and intercept could be the result of random variation in your data set, or they could be statistically significant. Without ap value, no one can tell.

Response: Regarding your first comment that the non-normal distribution of AOD data violated the assumption of linear regression, as we explained in a previous response, there are a few important issues that go against log-transformation in the context of this study, including the existence of valid negative AOD values and inconsistency with previous evaluation studies. Thus, we decided to use the original data in this analysis.

Regarding your second comment, we added an indicator of significance level based on the p-values of the regression slopes and intercepts in Table 3 and Table 4 in the revised manuscript.

- 25-20: "cautious" - how?

Response: Researchers need to calibrate these high-resolution satellite aerosol products in their study regions before applying them. Researchers may need to develop specific methods to process these data: for example, filtering AOD retrievals based on land use information. We

modified the sentence as follows on page 19, line 8–12: "In general, these finer resolution aerosol products included larger bias relative to lower resolution products and researchers must be cautious when applying them, e.g. calibrate these high resolution satellite aerosol products in specified study regions and implement appropriate data filtering strategies."

- Tables 3 and 4: Why are no p values given?

Response: We added an indicator of significance level based on p-value for linear regression slopes and intercepts in Tables 3 and Table 4.

- Figure 5: Since AOD is not normally distributed, it should be shown on a log scale or another suitable transformation.

Response: As explained in our previous response, there are a few important issues that go against log-transformation in the context of this study, including the existence of valid negative AOD values and inconsistence with previous evaluation studies. Thus, we decided to use the original data in this analysis.

3 Technical Details

- 11-15: ground-based
- 12-1 and 12-16: different time formats. Please harmonize throughout manuscript in accordance with journal requirements.
- 13-8 replace "that were" by a comma
- 13-12 remove "range"
- 14-5: small-scale
- 14-7: remove "required"
- 15-3: The size/extent etc. of the study area...
- 15-21: Ground-based measurements (here and elsewhere)
- 15-25: were/are distributed
- 16-2: approximately 10km apart -> with an averate distance of about 10km between two stations (surely 10 km isn't the distance between Osaka and Seoul...)
- 16-2: which can be... check wording
- 16-6: in THE Japan-South Korea region
- 16-17: "that distributed" -> selected sites roughly 6km apart from each other along
- 17-14: cells -> cell
- 20-5: metrics -> metric
- 21-5: results ... suggest

- 21-6: among -> between
- 23-8: over THE Japan-...
- 23-10: DRAGON
- Tables 3 and 4: The "Spatial Comparison" part should be more clearly visually distinct from the "Temporal Comparison" part.
- Figure 3, line 3: observations -> observation
- Figure 3, line 4: retrievals -> retrieval
- Figure 3: red and green are hard to impossible to distinguish for a of humanity (including me :). I suggest using a different pair of colors (e.g. red and blue)
- Figure 5: In their current form, the individual figures seem too small.
- Figure 5: in dash line -> as a dashed line
- Figure 5: in gray solid -> as gray solid

Response: Thank you for these suggestions/corrections, we changed the words and modified the figures in the revised manuscript accordingly.

References:

Ichoku, C., Chu, D. A., Mattoo, S., Kaufman, Y. J., Remer, L. A., Tanré, D., Slutsker, I., and Holben, B. N.: A spatio - temporal approach for global validation and analysis of MODIS aerosol products, Geophysical Research Letters, 29, MOD1-1-MOD1-4, 2002.

Levy, R., Mattoo, S., Munchak, L., Remer, L., Sayer, A., Patadia, F., and Hsu, N.: The Collection 6 MODIS aerosol products over land and ocean, Atmospheric Measurement Techniques, 6, 2989-3034, 2013.

Liu, H., Remer, L. A., Huang, J., Huang, H. C., Kondragunta, S., Laszlo, I., Oo, M., and Jackson, J. M.: Preliminary evaluation of S - NPP VIIRS aerosol optical thickness, Journal of Geophysical Research: Atmospheres, 119, 3942-3962, 2014.

Munchak, L., Levy, R., Mattoo, S., Remer, L., Holben, B., Schafer, J., Hostetler, C., and Ferrare, R.: MODIS 3 km aerosol product: applications over land in an urban/suburban region, Atmospheric Measurement Techniques Discussions, 6, 1683-1716, 2013.

Remer, L., Mattoo, S., Levy, R., and Munchak, L.: MODIS 3 km aerosol product: algorithm and global perspective, Atmospheric Measurement Techniques Discussions, 6, 69-112, 2013.

Anonymous Referee #2

This work studies the spatial and temporal characteristics of satellite remote sensing of aerosol products against ground measurements of AERONET, the DRAGON-Asia campaign, and data from a mobile sunphotometer sampling campaign in Beijing. Five emerging satellite aerosol products from three different platforms (i.e. MODIS, VIIRS, GOCI) are evaluated over East Asia in 2012-2013.

In general, the manuscript is well written and organized in a clear and logical way. This manuscript is, as far as I know, the first to compare these five satellite AOD products in one study. Moreover, the VIIRS and GOCI products are rather new and have not yet explored in depth. As such, this study adds knowledge to the atmospheric research community and could be published after addressing the following comments:

Major Comments:

âA˘ c The authors use VIIRS products and comment in page 20712, lines 16-17 that "The VIIRS aerosol product reached validated maturity level in January 2013". In the NASA LAADS website it is written in relation to the use of VIIRS products that "All Suomi NPP VIIRS EDRs are currently beta quality (with known problems) and are not intended for scientific use". A clarification is therefore needed as the data sources for VIIRS and GOCI satellite products are missing.

Response: The VIIRS aerosol product science team published a global evaluation study, reporting that the VIIRS AOD at the provisional maturity level is validated. The provisional maturity level is defined as: "product quality may not be optimal" but it is "ready for operational evaluation". To make this clear, we cited this study and added the following sentence on page 4, line 25-28: "The VIIRS aerosol product reached provisional maturity level in January 2013, which means the "product quality may not be optimal" but it is "ready for operational evaluation" (Liu et al., 2014)."

The GOCI science team recently submitted an evaluation paper of the GOCI aerosol product and it has been published on Atmospheric Measurement Techniques Discussions (Choi et al., 2015). In addition, the GOCI science team published a study about monitoring transboundary particulate pollution using the GOCI aerosol product, indicating that this product can be used for quantitative studies (Park et al., 2014). We hope our evaluation study can contribute to the validation of the GOCI aerosol product. To make this clear, we added the following sentence on page 5, line 28-page 6, line 2: "A recently published evaluation study reported that from March to May 2012, the GOCI AOD had a linear relationship with AERONET AOD with a slope of 1.09 and an intercept of -0.04 (Choi et al., 2015)."

âA˘ c This work presents data from sources with very different temporal and spatial resolutions including a changing footprint (e.g. MODIS) compared to a fixed footprint (i.e. GOCI). It is not clear how these differences have been taken into account? How has data fusion to one grid been done?

Response: We compared the satellite data with ground observations using sampling buffers with

respect to satellite products' resolutions. Both satellite and ground observations were fused to a fixed 3-km/10-km grid based on their locations, processed with ArcGIS. In addition, for polar orbit sensors (VIIRS and MODIS) that provide one observation per day, we used the 1-h time window (± 30 min of satellite pass-over time) for comparisons; for the geostationary orbit sensor (GOCI) that provides multiple observations per day, we conducted comparisons during the 1-h window around 13:30 that overlaid with other sensors, as well as during each of its 8 hourly observation periods.

âA˘ c This In page 20717, line 2 the authors write that the data was "remapped". A detailed explanation in the text of the remapping methodology is missing. I find it an important stage of the work and a detailed explanation will able the reader to understand and reproduce the methodology in a future work. Furthermore, is the remapping a daily procedure? What is the possible bias due to the remapping procedure?

Response: The remapping process here means spatially joining the satellite data with the fixed grid in a projected coordination system. To avoid any confusion, we modified this sentence as follows on page 9, line 21–23: "Satellite aerosol data from different sensors were mapped and spatially joined to this 6-km grid (for VIIRS EDR and GOCI products) or 3-km grid (for VIIRS IP and MODIS C6 3 km products) with respect to their spatial resolution." This process was conducted at daily level. Due to the stretch of MODIS and VIIRS pixels toward the edge of the scan, joining the satellite pixels with the fixed grid may lead to some missing satellite − ground AOD pairs, but there is no evidence that this missing will introduce a significant systematic bias.

âA˘ c I suggest to put more emphasis in the conclusion (and abstract) and throughout the manuscript on the better performance of satellite aerosol products in tracking the day to-day variability than in tracking/representing the spatial variability at high resolution. For example, in the Conclusion the authors claim that small scale variability and point sources can be detected. Unless point source has the size of 3-6 km I do not see how this claim is supported by the results in this manuscript. Also, individual exposure is mentioned on line 10 of p. 20729 – individual exposure estimation in urban areas may be obtained if we assume uniform exposure for all the people that live in a 3-6 km grid cells. If this is what the authors mean this needs to be clarified. Otherwise, I suggest to reduce expectations rather than increase them based on the reported MS results.

Response: We modified these sentences as follows on page 4, line 12-16: "The variability of aerosol loading at local scales in urban areas with complex land surface and meteorological conditions are expected to be greater (Li et al., 2005). Accurately characterizing local-scale $PM_{2.5}$ heterogeneity is critical for assessing population PM exposure, detecting air pollution sources, and monitoring air quality." and on page 23, lines 27-29: "High-resolution satellite aerosol products provide valuable information for the spatial and temporal characterization of $PM_{2.5}$ at local scales."

âA˘ c Sections 3.2, 3.3 – it will be very valuable to show performance metrics for the different satellite aerosol products after they were calibrated against ground measurements. Namely, once these products are calibrated it is very interesting to know which in fact performs better. Clearly, the calibration should be based on a complete leave one-out cross validation process, such that the model parameters are “optimal” in the sense that they represent all the data but not overfitting the data. Model parameterization should be developed on a regional (spatial) scale and then applied locally on AOD measurements, such that the spatial variability is still evident.

Response: We conducted 10-fold cross-validation analyses for temporal comparisons of VIIRS and GOCI data in the Japan–South Korea region and the regression statistics are similar to the original regression statistics. Due to the small sample size, cross-validation was not conducted for MODIS products. Since we aimed to evaluate rather than calibrate these satellite aerosol products, we did not create a table showing the performance metrics for the calibrated AOD. We added the following sentences on page 19, line 24-28: “Ten-fold cross validation was conducted for the comparison of VIIRS and GOCI products to detect overfitting. The linear regression statistics of cross validation did not change significantly relative to the statistics of comparisons. The cross validation R^2 values of VIIRS EDR, VIIRS IP, GOCI at 13:00, and GOCI 8 observations data were 0.73, 0.51, 0.78, and 0.82, respectively.”

All of these satellite aerosol products have their own advantages and disadvantages and are suitable for different research objectives, thus it is hard to say which one performed the best. We added the following sentences on page 23, line 19-26: “These satellite aerosol products have their own advantages and disadvantages. For example, the GOCI aerosol product provides high accuracy AOD retrievals eight times per day, but it only covers East Asia; the VIIRS EDR product provides high accuracy AOD retrievals and global coverage once per day, but its 6 km resolution is relatively low; the MODIS C6 3 km products provide high resolution AOD retrievals with global coverage, but have positive bias in urban regions. Researchers need to apply these aerosol products according to specified research objectives and study design.”

Using GOCI 8 observations per day data, we applied the regionally developed linear regression parameters to individual station data in the Japan–South Korea region. The linear regressions with the satellite AOD as a dependent variable and the fitted AOD from a regional model as an independent variable have an R^2 greater than 0.75 at all sites except the AERONET site ‘Nara’ and ‘Osaka’, two stations located in Osaka. Limited by sample size, we cannot apply this method to other aerosol products. However, since the spatial distribution of satellite aerosol products from different sensors are similar in this region, we believe that parameters from regional datasets were also valid locally. We added the following sentences on page 19, line 28-page 20, line 7: “In addition, to detect the spatial variability of the satellite retrieval performance, we applied the regionally developed linear regression parameters of GOCI 8 observations data to individual AERONET station in the Japan–South Korea region. The linear regressions with the satellite AOD as the dependent variable and the fitted AOD from a regional model as the independent variable yielded R^2 larger than 0.75 at all sites except the AERONET sites ‘Nara’ and ‘Osaka’, two stations located in Osaka. This result indicated that parameters from the regional dataset were valid locally. Limited by sample size, we did not apply this method to other aerosol products.”

Minor Comments:

âA˘ c Figure S1 presents the spatial distribution of the stations with the different buffers. (a) The size of the ground station symbols is not proportional and I recommend to reduce the symbol size. (b) I recommend using a scale bar of 3-6-9 km, which is more relevant, instead of 5-10-20 km. (c) The different sample size boxes are not very clear: 3x3, 4x4, 6x6, 9x9? An additional table at the bottom of the figure with an explanation in the manuscript and next to each cell size can possibly make this clearer.

Response: We modified this figure according to your suggestion and added the following explanation in the captions: "The temporal comparison figure (left) shows the buffer of 3 x 3 grid cells for MODIS (pink), VIIRS EDR and GOCI products (blue), as well as the single grid cell buffer for VIIRS IP product (green); the spatial comparison figure (right) shows the single grid cell buffer for each sensor."

âA˘ c Table S2- How was the number of observations (N) from each data source taken into account? Show that the results are affected/not affected by this parameter (N).

Response: Since coverage and accuracy are two major metrics used to evaluate the performance of satellite aerosol products, the number of coincident satellite-ground AOD pairs in this table was aimed to reflect the coverage of each satellite aerosol product. Since the estimated slopes and intercepts were significant, the sample size was sufficient and the results were not affected by N.

âA˘ c The standard deviation within the 3x3 cells isn't reported. I think it is important to report it before averaging the cells in order to study/observe the distance between values within the 3x3 boxes is low.

Response: We calculated the coefficient of variation (CV), which is the standard deviation divided by the mean, of AOD retrievals in the temporal-comparison buffer from various sensors. To avoid effects from large within-buffer variation in aerosol loading, we removed satellite pixels with CV outside the range of ± 1.0. Doing so led to less than 10% missing data and the regression statistics remained almost the same. We reported CVs of AOD retrievals from each sensor in section 2.4 of the revised manuscript and we added the following sentences on page 10, line 16-27: "The coefficients of variation (CV), which is standard deviation divided by mean of AOD retrievals, from various sensors in temporal-comparison sampling buffers were calculated and reported below to assess the homogeneity of aerosol loading within buffers. The mean CV from various aerosol products ranged between 0.18 and 0.35, indicating that, as expected, certain heterogeneity in aerosol loading existed within the temporal-comparison buffer. This relatively small heterogeneity should not be a detriment to the temporal comparison, however; some extremely large CV values that were probably due to very small mean AOD values were observed. In order to avoid potentially large variations in aerosol loading within buffers, we removed satellite pixels with CVs outside the range of ± 1.0

(Liu et al., 2007) in temporal comparisons. Moreover, the existing heterogeneity of AOD loading encouraged us to conduct spatial comparisons implementing smaller sampling buffers."

âA˘ c P 20720, lines 2-4. "slope is the slop e of the linear regression with satellite retrievals as the dependent variable and ground AOD measurements as the independent variable;" it should be exactly the opposite. We want to predict ground PM by AOD so satellite AOD should be the independent variable and ground measurements (hear ground AOD) be the dependent variable. This way the satellite AOD will be consistently used as the independent variable.

Response: Since ground AOD is considered as "true value", we used ground AOD as the independent variable and the satellite retrievals as the dependent variable. In this study, we did not want to estimate the "true" AOD from satellite retrievals; in contrast, we wanted to validate satellite retrievals with ground truth and tested by how much the satellite retrievals deviated from the ground truth; thus, the satellite AOD was the dependent variable. In most previous evaluation studies, the satellite AOD was the dependent variable and the AERONET AOD was the independent variable. When predicting ground-level PM concentrations using satellite AOD, satellite AOD—together with other parameters—are independent variables, but the objectives and interpretations of these two kinds of studies are different.

âA˘ c Page 20720, lines 10-16. Consider moving these lines to the introduction and method sections.

Response: We moved these sentences to the introduction and method section.

âA˘ c p 20721 line 3. Figure 2b shows the site specific average AOD with the regional average AOD subtracted in these three cities – how was the background calculated?

Also, please explain what is the meaning of 0.01 increase in AOD as represented by different colors in Figure 2(b). Moreover, the manuscript (page 20721, line 20) refers to a difference of AOD of 0.4 between stations, a value not represented in the figure.

Response: The regional average (background) AOD was calculated as the average of AOD from all the ground stations located in this region. The background color, mainly green, denotes the elevation of this region with the same color scale as in Figure 1. To clarify this, we changed the color scale of AOD in Figure 2(b) and added the following sentence in the caption of Figure 2: "The background color shows the elevation with the same color scale as in Figure 1." The different colors in Figure 2(b) indicate the difference between AOD from each ground station and the regional average AOD. We added two more colors to this color scale to show that the difference in AOD between two nearby stations in Beijing is about 0.4.

âA˘ c P 20721 lines15-18. I assume that the higher variability in Beijing comes from the (a) poorer performance of the hand held device (e.g. instrument quality), (b) the use of daily average AOD values in DRAGON sites vs. momentarily measurements (in each site-day) in Beijing (e.g. measurement noise, un-representativeness of the measurments in Beijing), and (c) in Beijing the measurement may have been performed when the devices does not exactly face the sun due to operation errors. I suggest to discuss all these optional sources of errors.

Response: We added the following sentences in the revised manuscript on page 14, line 25-page 15, line 4: "Second, the handheld sunphotometer may introduce larger measurement errors than DRAGON stations, due to both instrument quality and operation errors. Previous evaluation indicates that handheld stability and inaccurate pointing to the Sun significantly affects the accuracy of measurements by Mocrotops II (Ichoku et al., 2002; Morys et al., 2001). Our comparison of Microtops II AOD with nearby AERONET data yielded a slope of ~0.95, a correlation coefficient of ~0.8, and an intercept of 0.16 (Supplemental Material, Text S1), indicating that the handheld sunphotometer AOD are usable."

âA˘ c Page 20722, lines 4-5. Compare the availability of different satellite-based data and AOD from AERONET at 13:00. Terra overpass is at 10:30 local time, it hasn't been mentioned throughout the manuscript if the Terra data was compared to AERONET data at 10:30. One can understand from the text that the Terra observations were compared to AERONET at 13:00. Yet, later in the manuscript, in the first paragraph in page 20724, the overpass time difference of Terra is mentioned. I recommend to either make this clearer or to consider excluding the Terra dataset from this study.

Response: We compared the availability of Terra data with AERONET from 10:00-11:00 am. We kept Terra in this study because its aerosol products are widely used and it provides additional information about aerosol distribution. To make this clear, we added the definition of the 1-h window used for comparisons on page 15, line 21 and page 17, line 14.

âA˘ c As written in page 20724, line 3, the Y-axis in Figure 4 is "relative frequency rather than the total number of retrievals". If the frequency is relative to the number of observations (N) than it (i.e. N) should be specified in the text and/or in the figure. Moreover, as the number of satellite observations has seasonal variation (e.g. due to clouds), I suggest to add the number of observations per satellite per month, possibly in a separate figure/table.

Response: The frequency is relative to the total number of matched AOD retrievals from the corresponding sensor. We modified the sentence on page 17, line 19-23: "This histogram is plotted with the frequency of AOD retrievals from each sensor relative to the total number of matched AOD retrievals from the corresponding sensor rather than the count of AOD retrievals because these aerosol products differ in sampling strategies, leading to different total numbers of coincident satellite–ground AOD pairs." We also added the following sentence to clarify this in the caption of Figure 4: "The x-axis shows AOD values and the y-axis shows the frequency of AOD observations from each sensor relative to the total number of matched AOD observations from the corresponding sensor."

This figure compared the distribution of AOD from each satellite dataset to AOD from AERONET, the ground truth. Thus, we can detect systematic bias. The variation in the number of observations (N) across satellite aerosol products due to differences in the aerosol products'

resolutions and masking strategies does not necessarily lead to different retrieval quality, so we did not specify N in this figure. Since we used AERONET AOD as ground truth and showed the distribution of AOD from matched satellite-ground AOD pairs, this figure indicated distribution of AOD retrievals in cloud-free conditions. The seasonal missing pattern of each AOD dataset due to cloud and weather conditions is out of the scope of this figure. We added the following sentence to clarify this on page 17, line 11-15: "It is notable that the seasonal missing pattern due to cloud cover and weather conditions may vary across these satellite aerosol products. However, since we did not have enough coincident satellite-ground AOD pairs to conduct seasonal evaluation, the seasonal missing patterns and seasonal performance of these satellite aerosol products were not analyzed in this study."

âA˘ c Page 20729 top. Clearly, the conclusion that the 6 km products provide more accurate data than the 3 km products results from the spatiotemporal averaging. This may be useful in some cases but is huge disadvantage in other cases, in particular for environmental health and exposure estimation, which is one of the applications declared by the authors as their interest.

Response: We understand that one major application of aerosol satellite remote sensing is exposure assessment and that's why we introduced quality flags for coincident satellite-ground AOD pairs. There is a trade-off between satellite retrieval coverage and accuracy, and we tried to increase the coverage without significantly decreasing accuracy. We understand that the 3 km products and products at even higher resolution will contribute to fine-scale exposure assessment; however, these products showed higher bias in this and previous evaluations. Researchers need to use these products with caution. We added the following sentence on page 23, line 12-16: "however, VIIRS IP and MODIS C6 3 km products provide additional information about fine-resolution aerosol spatial distribution and will benefit exposure assessments at local scales;"

âA˘ c Figure 6. The color scale should be the same for all figures for a clearer interpretation.

Response: We used the same color scale for all figures with different minimum and maximum values. The minimum value is 0 for VIIRS products and -0.05 for MODIS and GOCI products, and the maximum value is 2.0 for VIIRS products and >2.0 for MODIS and GOCI products. This difference is related to retrieval algorithms and we wanted to indicate this difference in figures, but the fact that these color scales differed in maximum and minimum values did not affect comparisons across these figures.

âA˘ c Table 3. The temporal comparison section and the spatial comparison section should be separated, e.g. by a line above the spatial comparison section.

Response: We modified table 3 and table 4 to make the temporal and spatial comparison sections more visually separated from each other.

âA˘ c Caption to Fig. 2a – what is "Loess curvy" ? Fig. 2b – what is the meaning of the green background color in non-measurement locations?

Response: The Loess curve is a smooth curve based on a non-parametric regression. We used this curve to show the trend of the correlation coefficient of AOD from two stations with increasing distance. The green background color shows the elevation with the same color scale as in Figure 1. To make this clear, we added the following sentence in the caption of Figure 2: "The background color shows the elevation with the same color scale as in Figure 1."

âA˘ c Fig. 5 is too small and its details cannot be seen. There is a need to improve the presentation of this fig.

Response: We modified the arrangement of Figure 5 and enlarged each of the figures.

References:

Choi, M., Kim, J., Lee, J., Kim, M., Je Park, Y., Jeong, U., Kim, W., Holben, B., Eck, T. F., Lim, J. H., and Song, C. K.: GOCI Yonsei Aerosol Retrieval (YAER) algorithm and validation during DRAGON-NE Asia 2012 campaign, Atmos. Meas. Tech. Discuss., 8, 9565-9609, 10.5194/amtd-8-9565-2015, 2015.

Ichoku, C., Chu, D. A., Mattoo, S., Kaufman, Y. J., Remer, L. A., Tanré, D., Slutsker, I., and Holben, B. N.: A spatio - temporal approach for global validation and analysis of MODIS aerosol products, Geophysical Research Letters, 29, MOD1-1-MOD1-4, 2002.

Levy, R., Mattoo, S., Munchak, L., Remer, L., Sayer, A., Patadia, F., and Hsu, N.: The Collection 6 MODIS aerosol products over land and ocean, Atmospheric Measurement Techniques, 6, 2989-3034, 2013.

Liu, H., Remer, L. A., Huang, J., Huang, H. C., Kondragunta, S., Laszlo, I., Oo, M., and Jackson, J. M.: Preliminary evaluation of S - NPP VIIRS aerosol optical thickness, Journal of Geophysical Research: Atmospheres, 119, 3942-3962, 2014.

Munchak, L., Levy, R., Mattoo, S., Remer, L., Holben, B., Schafer, J., Hostetler, C., and Ferrare, R.: MODIS 3 km aerosol product: applications over land in an urban/suburban region, Atmospheric Measurement Techniques Discussions, 6, 1683-1716, 2013.

Park, M., Song, C., Park, R., Lee, J., Kim, J., Lee, S., Woo, J.-H., Carmichael, G., Eck, T. F., and Holben, B. N.: New approach to monitor transboundary particulate pollution over Northeast Asia, Atmospheric Chemistry and Physics, 14, 659-674, 2014.

Remer, L., Mattoo, S., Levy, R., and Munchak, L.: MODIS 3 km aerosol product: algorithm and global perspective, Atmospheric Measurement Techniques Discussions, 6, 69-112, 2013.

Evaluation of VIIRS, GOCI, and MODIS Collection 6 AOD retrievals against ground sunphotometer observations over East Asia

Q. Xiao[1], H. Zhang[2], M. Choi[3], S. Li[1, 4], S. Kondragunta[5], J. Kim[3], B. Holben[6], R. C. Levy[6], Y. Liu[1]

[1] {Emory University, Rollins School of Public Health, Atlanta, GA, USA}

[2] {I.M. Systems Group,Inc., College Park, MD, USA}

[3] {Yonsei University, Seoul, South Korea}

[4] {State Key Laboratory of Remote Sensing Science, Beijing, China}

[5] {National Oceanic and Atmospheric Administration, Greenbelt, MD, USA}

[6] {NASA Goddard Space Flight Center, Greenbelt, MD, USA}

Correspondence to: Y. Liu (yang.liu@emory.edu)

Abstract

Persistent high aerosol loadings together with extremely high population densities have raised serious air quality and public health concerns in many urban centers in East Asia. However, ground-based air quality monitoring is relatively limited in this area. Recently, satellite-retrieved Aerosol Optical Depth (AOD) at high resolution has become a powerful tool to characterize aerosol patterns in space and time. Using ground AOD observations from the Aerosol Robotic Network (AERONET) and the Distributed Regional Aerosol Gridded Observation Networks (DRAGON)-Asia Campaign, as well as from handheld sunphotometers, we evaluated emerging aerosol products from the Visible Infrared Imaging Radiometer Suite (VIIRS) aboard the Suomi National Polar-orbiting Partnership (S-NPP), the Geostationary Ocean Color Imager (GOCI) aboard the Communication, Ocean, and Meteorology Satellite (COMS), and Terra and Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) (Collection 6) in East Asia in 2012 and 2013. In the case study in Beijing, when compared with AOD observations from handheld sunphotometers, 51% of VIIRS Environmental Data Record (EDR) AOD, 37% of GOCI AOD, 33% of VIIRS Intermediate Product (IP) AOD, 26% of Terra MODIS C6 3 km AOD, and 16% of Aqua MODIS C6 3 km AOD fell within the reference expected error (EE) envelop (±0.05±0.15AOD). Comparing against AERONET AOD over the the Japan-South Korea region, 64% of EDR, 37% of IP, 61% of GOCI, 39% of Terra MODIS and 56% of Aqua MODIS C6 3 km AOD fell within the EE. In general, satellite aerosol products performed better in tracking the day-to-day variability than tracking the spatial variability at high resolutions. The VIIRS EDR and GOCI products provided the most accurate AOD retrievals, while VIIRS IP and MODIS C6 3 km products had positive biases.

1. Introduction

Rapid economic growth and increasing fossil fuel usage have led to increasing air pollutant emission in East Asia. From 1980–2003, the emissions of black carbon, organic carbon, SO_2, and NOx increased by 28%, 30%, 119%, and 176%, respectively (Ohara et al., 2007). The continuous air quality degradation together with high population density have raised serious public health concerns in this region. Among commonly monitored air pollutants, particulate matter (PM), especially fine particulate matter ($PM_{2.5}$, airborne particles with an aerodynamic diameter less than or equal to 2.5 μm), is noted for its adverse health impacts, such as increased cardiovascular and respiratory morbidity and mortality (Holben et al., 1998; Li et al., 2005). The severe PM pollution in East Asia has attracted worldwide attention and ground PM monitoring networks have been developed in some East Asian countries like China, Japan and South Korea. For instance, in South Korea, PM_{10} together with other important air pollutants have been measured by a dense ground-based network, called 'Air Korea', by the Ministry of Environment (http://eng.airkorea.or.kr). However, ground-based monitoring networks have two main limitations: uneven distribution and limited coverage. For example, the majority of air quality monitoring stations in China are located in large cities and the monitoring network only covers about 360 out of the approximately 2,860 municipalities. These two limitations of ground PM measurements result in insufficient information to conduct studies about PM sources, distribution, and consequent health impacts in East Asia, which can negatively impact policymaking.

The extensive spatial coverage and growing time series of satellite retrievals allow researchers to better characterize aerosol patterns spatially and temporally. The most widely used satellite aerosol sensor, the Moderate Resolution Imaging Spectroradiometer (MODIS), has 36 spectral bands, acquiring data in wavelength from 0.41 μm to 15 μm and providing information about atmospheric aerosol properties (Anderson et al., 2003). Two identical MODIS instruments are aboard the National Aeronautics and Space Administration (NASA) Terra and Aqua satellites, which fly over the study area at around 10:30 and 13:30 LT, respectively. Several algorithms have been developed to retrieve aerosol optical depth (AOD) from MODIS data over land, such as the Dark-Target (Levy et al., 2013) algorithm and the Deep-Blue (Hsu et al., 2013) algorithm, providing AOD retrievals at 550 nm with global coverage. The widely used 10 km resolution MODIS aerosol products provides valuable information on aerosol distribution in space and time, and has been widely used to characterize aerosol dynamics and distribution, simulate climate change, and assess population PM exposure (Levy et al., 2013; Levy et al., 2010). However, the 10 km product cannot depict small-scale $PM_{2.5}$ heterogeneity. Though ~~A~~a previous study (Anderson et al., 2003) indicated that the aerosol loading is homogeneous at horizontal scales within 200 km~~. However~~, that study is conducted over the ocean, which provides a homogeneous surface, leading to reduced aerosol spatial variability. The variability of aerosol loading at local scales in urban areas with complex land surface and meteorological conditions are expected to be greater (Li et al., 2005). Accurately characterizing local-scale PM2.5 heterogeneity is critical for assessing population PM exposure, detecting air pollution sources, and monitoring air quality. To resolve small-scale aerosol features, satellite aerosol products with higher resolutions and acceptable accuracy are urgently needed.

In response to the requirement of aerosol retrievals with higher spatial resolution, several emerging satellite aerosol products have become available recently. The Visible Infrared Imaging Radiometer Suite (VIIRS), is a multi-disciplinary scanning radiometer with 22 spectral bands covering from 0.412–12.05 μm and is designed as a new generation of operational satellite sensors that are able to provide aerosol products with similar quality to MODIS (Jackson et al., 2013). VIIRS is on board the NASA-NOAA Suomi National Polar-orbiting Partnership (S-NPP) that launched in October 2011, and passes over the study area daily at approximately 13:30 LT. The VIIRS aerosol product reached provisional maturity level in January 2013, which means the “product quality may not be optimal” but it is “ready for operational evaluation” (Liu et al., 2014). The characteristics of the instrument and the aerosol retrieval algorithms are documented in detail elsewhere (Liu et al. (2014)) and briefly described here. VIIRS provides two AOD products: the Intermediate Product (IP) and the Environmental Data Record (EDR). The VIIRS aerosol retrieval is performed at pixel-level (~0.75 km) spatial resolution globally as the IP that employs information from Navy Aerosol Analysis and Prediction System (NAAPS) and Global Aerosol Climatology Project (GACP) to fill in missing observations (Vermote et al., 2014). The IP is then aggregated to 6-km spatial resolution as the EDR, a level 2 aerosol product, through quality checking and excluding information from the NAAPS and GACP models. Both VIIRS IP and EDR are assigned quality flags of “high”, “degraded”, or “low” and valid AOD values range between 0.0 and 2.0. Detailed description of the quality assurance of VIIRS aerosol products is documented by Liu et al. (2014). Previous global evaluation against AERONET AOD over all land use types indicates that 71% of EDR retrievals fell within the expected error (EE) envelope established by MODIS level 2 aerosol products over land (±0.05±0.15AOD), with a bias of -0.01 (Liu et al., 2014).

The Geostationary Ocean Color Imager (GOCI) is a geostationary Earth orbit sensor, providing hourly multi-spectral aerosol data eight times per day from 9:00 to 16:00 Korean LT. It covers a 2500 × 2500 km^2 sampling area, centered at [130E, 36N] in East Asia, at 500-m resolution with eight spectral channels at 412, 443, 490, 555, 660, 680, 745, and 865 nm, respectively (Park et al., 2014). GOCI is aboard South Korea’s Communication, Ocean, and Meteorology Satellite (COMS) that launched in June 2010. The retrieval algorithm of its aerosol product, Yonsei aerosol retrieval algorithm, was originally based on the NASA MODIS algorithm and provides level 2 AOD retrievals at 6-km spatial resolutions (Levy et al., 2007; Levy et al., 2010; Lee et al., 2010). The characteristics of the Yonsei retrieval algorithms and the aerosol product are documented in detail by Choi et al. (2015). The GOCI aerosol product allows AOD values ranging between -0.1 and 5.0. A previous study reported that during a two-month period (1 April to 31 May 2011), the GOCI AOD retrievals agreed well with AERONET AOD (~~r~~R^2 = 0.84) over East Asia (Park et al., 2014). A recently published evaluation study reported that from March to May 2012, the GOCI AOD had a linear relationship with AERONET AOD with a slope of 1.09 and an intercept of -0.04 (Choi et al., 2015).

To meet the need for finer resolution aerosol products, a 3 km aerosol product was introduced as part of the MODIS Collection 6 delivery. The 3 km aerosol product includes a quality flag ranging between 0 and 3 to indicate the quality of each retrieval and the valid AOD values range between -0.1 and 5.0. The retrieval algorithm of the 3 km product is documented in detail by Remer et al. (2013) and a global evaluation based on six months of Aqua data against ground sunphotometer AOD indicates that 63% of the retrievals fell into the EE with a bias of 0.03 over land (Remer et al., 2013). Munchak et al. (2013) reported that in the Baltimore–Washington, D.C. area, an urban/suburban region, 68% of the 3 km retrievals from June 20, 2011 to July 31, 2011 fell into the EE with a bias of 0.013.

The release of these fine-resolution satellite aerosol products has raised the question of whether these AOD retrievals can reflect the spatial pattern of aerosol loadings at their assigned resolutions. AERONET, a globally distributed federation of ground-based atmospheric aerosol observations, provides reliable “ground truth” of AOD that are widely used for the characterization of aerosol and validation of satellite retrievals (Morys et al., 2001; Holben et al., 1998). However, previous evaluation studies with AERONET data focused on the temporal accuracy (i.e., examined if the retrieved AOD can track the day-to-day variability of aerosol loadings). Evaluation of satellite aerosol products’ abilities to track small-scale aerosol spatial variability is limited due to a lack of intensive ground observations of AOD: the permanent AERONET stations can be tens or even hundreds of kilometers apart, leading to insufficient information on the small-scale horizontal distribution of aerosol loading that is required for a precise evaluation at high resolution. In response to the lack of intensive ground AOD observations, AERONET conducted several campaigns, which deployed additional temporary sunphotometers in selected regions and provided valuable information of small-scale AOD distribution. One of these campaigns, the Distributed Regional Aerosol Gridded Observation Network (DRAGON)-Asia Campaign in Japan and South

Korea, lasted from February 15, 2012 to May 31, 2012 and provided a rare opportunity to validate these emerging satellite aerosol products (Seo et al., 2014;Sano et al., 2012). Another issue with previous evaluation studies is that few of them focused specifically on urban areas with higher pollution levels, greater disease burdens, and more complex aerosol patterns. Our work contributes to the validation effort of these emerging satellite products by employing ground AOD observations at finer resolution, extending the study period to one year, and conducting a mobile sampling experiment in the urban core of Beijing.

In this work, we quantitatively evaluate whether the latest VIIRS, GOCI and MODIS aerosol products can provide reliable AOD retrievals and accurately characterize the spatial pattern of AOD over the urban areas in East Asia. Ground AOD from AERONET, DRAGON-Asia, and handheld sunphotometers were collected over a period of one and a half years. The rest of the paper is organized such that Section 2 describes data sources and evaluation methods used in this study, Section 3 presents the performance of various satellite AOD products in representing intra city as well as regional variability of aerosol loadings. Finally, we summarize our findings and described future study directions in section 4.

2. Data and Methods

2.1 Study Area

The extent of the study area is approximately 2500 × 1100 km^2, centered at [128.5E, 35.5N] in East Asia, covering eastern China, South Korea and Japan (Fig. 1). This domain is within the overlapping region of all satellite datasets and ground observations and covers large urban centers, suburban areas, and rural areas. We also conducted a mobile sampling study in Metro Beijing along three major roads (Fig. 1). The study period is from January 2012 to June 2013.

2.2 Remote Sensing Data

The satellite aerosol products used in this study were from VIIRS, GOCI, Aqua MODIS and Terra MODIS sensors (Table 1). VIIRS data before May 2012 are not available because the sensor was in an early checkout phase and lacked a validated cloud mask (Liu et al., 2014). Thus, only EDR and IP pixels from May 2012 to June 2013 with high quality (Quality Flag = "high") were processed. Similarly, GOCI aerosol retrievals from January 2012 to June 2013 were filtered by its assigned quality and only high quality (Quality Flag = 3) retrievals were included. The Aqua and Terra MODIS C6 3 km data from January 2012 to June 2013 were obtained from the Goddard Space Flight Center (http://ladsweb.nascom.nasa.gov/data). Only retrievals with high quality (Quality Flag = 3) were included in the analysis. The quality control criteria of these five satellite aerosol products are shown in Table 1.

2.3 Ground Observations

The characteristics of ground AOD datasets are shown in Table 2. There were 18 permanent AERONET stations in the study area during the study period, supplemented by 24 temporary stations during the DRAGON-Asia Campaign. The DRAGON stations were distributed nearly uniformly with approximately 10 km apart from each other in two urban centers: Osaka in Japan (7 stations) and Seoul in South Korea (11 stations). Other DRAGON stations, which can be tens to hundreds of kilometers apart, were located across Japan and South Korea. AERONET stations observe AOD at eight spectral bands between 340 nm and 1020 nm. To compare with satellite retrievals, AOD at 550 nm was calculated using a quadratic log-log fit from AERONET AOD at wavelengths 440 nm and 675 nm. Near-real time level 2.0 AERONET/DRAGON data in the Japan-South Korea region and level 1.5 AERONET data in Beijing were downloaded from the Goddard Space Flight Center (http://aeronet.gsfc.nasa.gov/). The Level 2.0 (quality assured) AOD data have both pre- and post-deployment calibration, leading to an uncertainty of about 0.01–0.02 while the Level 1.5 AOD data are cloud-screened but not quality-assured (Otter et al., 2002). However, our preliminary results indicate that the level 1.5 daily average AOD values agreed well with the level 2.0 data, with a slope of 1.0 and zero intercept. Thus, we used the level 1.5 data in the case study in Beijing because level 2.0 data are not available for some AERONET stations.

To analyze the intra-city aerosol variability, we conducted ground measurements of AOD by a handheld sunphotometer (model 540 Microtops II, Solar Light Company, Inc.) at the Metro Beijing area in 2012 and 2013. Microtops II provide accurate AOD retrievals and is widely used for ground AOD observations (Morys et al., 2001; Tiwari and Singh, 2013; Otter et al., 2002). Previous calibration reported that the root-mean square differences in AOD from Microtops and corresponding AERONET stations were about ± 0.02 at 340 nm (Ichoku et al., 2002). In this study, ground observations were conducted on every cloud-free day at preselected sites that were roughly 6 km apart from each other along the 3rd and the 5th Ring Roads and the Chang'an Avenue of Beijing. This sampling took place between 9:30 and 14:00 LT, and 5–10 repeated measurements were made at each site. To control the quality of the ground data, we used the median value of the repeated observations as ground truth to eliminate the impact of extreme values and only included AOD with the ratio of standard deviation over median AOD less than 2.0. Our comparison of Microtops AOD retrievals with nearby AERONET data yielded a slope of ~0.95 and a correlation coefficient of ~0.8 (Supplemental Material, Text S1).

2.4 Data Integration and Analytical Methods

All the data were converted to the JGD_2000_UTM_Zone_52N coordination system. For matchup process, a 6-km grid and a 3 km grid covering the whole study domain were constructed, corresponding to the spatial resolution of each satellite product. Satellite aerosol data from different sensors were mapped and spatially joined to this 6-km grid (for VIIRS EDR and GOCI products) or 3 km grid (for VIIRS IP and MODIS C6 3 km products) to construct coincident satellite-ground AOD pairs.

To assess the intra-city spatial variations of aerosol loadings, we analyzed ground AOD observations over Beijing, Osaka, and Seoul from handheld sunphotometer and DRAGON-Asia stations in 2012. First, the great circle distance between each of two ground observation sites which are less than 20 km apart were calculated. Then we stratified the site-to-site distances by increments of 750 m, the resolution of VIIRS IP aerosol product, and calculated the station-to-station correlation coefficients of daily average AOD within each distance stratum. The observations from DRAGON sites in Osaka and Seoul and from handheld sunphotometers in Beijing were processed separately due to differences in instrumentation. Only handheld sunphotometer AOD observations in Beijing from February 15, 2012 to May 31, 2012 were included to ensure that the study period at these three locations is the same.

To validate the performance of high-resolution satellite aerosol products, two types of comparisons were conducted: the temporal comparison, which compared satellite AOD retrievals within 3 × 3 grid cells sampling buffers against ground AOD from AERONET stations during one year from July 2012 to June 2013; and the spatial comparison, which compared satellite AOD retrievals within single grid cell sampling buffers against intensive ground AOD from DRAGON stations or the handheld sunphotometer. Temporal comparisons and spatial comparisons differ in study periods (Table 2): the temporal comparison period was the longest overlap period covered by all five satellite products and the spatial comparison periods in Beijing and the Japan–South Korea region are different in order to include the maximum number of ground observations. The coefficients of variation (CV), which is standard deviation divided by mean of AOD retrievals, from various sensors in temporal-comparison sampling buffers were calculated and reported below to assess the homogeneity of aerosol loading within buffers. The mean CV from various aerosol products ranged between 0.18 and 0.35, indicating that, as expected, certain heterogeneity in aerosol loading existed within the temporal-comparison buffer. This relatively small heterogeneity should not be a detriment to the temporal comparison, however; some extremely large CV values that were probably due to very small mean AOD values were observed. In order to avoid potentially large variations in aerosol loading within buffers, we removed satellite pixels with CVs outside the range of ± 1.0 (Liu et al., 2007) in temporal comparisons. Moreover, the existing heterogeneity of AOD loading encouraged us to conduct spatial comparisons implementing smaller sampling buffers.

For the temporal comparison of VIIRS EDR data, we averaged valid AOD retrievals in each 3 × 3 grid cells sampling buffer (18 × 18 km^2) centered at each ground AERONET station. The mean and median CV were 0.25 and 0.21, respectively. The average AOD values were then compared with the mean AERONET AOD within a 1-h time window (± 30 min around the satellite overpass time). We employed this smaller spatial averaging window than the widely used 27.5 km-radius-circle buffer suggested by the Multi-sensor Aerosol Products Sampling System (MAPSS) (Seo et al., 2014) in order to examine the performance of these finer resolution products at the scale of their expected application conditions. We used the typical 1-h time window because a previous analysis indicated that changing the time window matters little to validation results (Remer et al., 2013) and the 1-h time window yields a larger database for the validation. For the spatial comparison of VIIRS EDR data, we used single 6-km pixels covering each ground observation location, i.e. DRAGON station or handheld sunphotometer measurement location, and compared the AOD retrieval values with the mean AOD from the corresponding DRAGON station within the 1-h time window or the median AOD from the handheld sunphotometer at the corresponding location. The temporal and spatial comparisons of GOCI data followed the same protocol as described above. Although GOCI provides eight hourly AOD retrievals per day, we only used retrievals at 1:00 pm LT in the comparison in order to make the validation results comparable among these satellite products. The mean and median CV of GOCI retrievals within the 3 × 3 grid cells sampling buffer were 0.35 and 0.15, respectively.

For the comparisons of VIIRS IP data, we used the 3 km grid because we did not have enough ground sampling data to create a 750-m grid. For the temporal comparison, we averaged valid IP AOD retrievals falling in the 3 km grid cell centered at each ground AERONET station and the mean and median CV were 0.33 and 0.25, respectively, within the 3 km grid cell buffer. This sampling buffer roughly covered a 4 × 4 pixel group. The average AOD values were compared against average AOD from the corresponding AERONET station within the 1-h time window. In the spatial comparison of VIIRS IP, we also used the 3 km sampling buffer due to a lack of more intensive ground AOD observations. Thus, the VIIRS IP data is oversampled in the spatial comparison. For the temporal comparison of Aqua and Terra MODIS C6 3 km data, we employed the 3 km grid and averaged valid AOD retrievals in each 3 × 3 grid cells centered at each ground AERONET station to compare with the mean AOD within the 1-h time window. The mean CV of Aqua and Terra MODIS within the 3 × 3 grid cells sampling buffer were 0.18 and 0.13, respectively. For the spatial comparison of MODIS C6 3 km data, we used the individual 3 km pixel AOD value falling on each ground observation location to compare with average AOD from the corresponding DRAGON station within the 1-h time window or the median AOD from the handheld sunphotometer at the corresponding location.

In summary, coincident satellite–ground AOD pairs were defined as average satellite AOD retrievals within the specific sampling buffer matched with average ground AOD observations of the corresponding site within 1-h time windows with respect to satellite pass over time. for VIIRS EDR and GOCI products, the temporal and spatial comparison buffer was 18 × 18 km^2 and 6 × 6 km^2, respectively. For the VIIRS IP product, the temporal and spatial comparison employed the same 3 × 3 km^2 buffer. For MODIS C6 3 km product, the temporal and spatial comparison buffer was 9 × 9 km^2 and 3 × 3 km^2, respectively. The examples of buffers used in the temporal and spatial comparisons for each satellite product are shown in Supplemental Material (Fig. S1). It is notable that both MODIS and VIIRS pixels were stretched toward the edge of the scan. For example, the 3 × 3 km^2 MODIS pixels become approximately 6 × 12 km^2 toward the edge. Thus, the spatial joining and our construction of coincident satellite-ground AOD pairs mayslightly decrease the coverage for MODIS and VIIRS products and may potentially affect the spatial comparison results.

In epidemiological studies, in order to improve the coverage of satellite aerosol data to provide exposure assessment, spatial aggregation is widely used. In our analysis, we constructed quality flags for each satellite–ground AOD collection to obtain better coverage without losing accuracy. For the temporal validation, coincident satellite–ground AOD pairs with at least 20% coverage of both satellite data and ground data (Levy et al., 2013) (e.g., having two or more satellite pixels within the sampling buffer and at least two AERONET/DRAGON AOD within the 1-h time window) were marked as “High Quality”; coincident satellite-ground AOD pairs with less than 20% satellite pixels falling in the sampling buffer but one or more pixels located within the grid cell centered on the ground stations were marked as “Medium Quality”; allother coincident satellite-ground AOD pairs were marked as “Low Quality”. Since we did not create a 750--m grid for the VIIRS IP product, VIIRS IP-ground AOD pairs were assigned either “High Quality” or “Low Quality”. In the spatial validation, because the best scenario satellite-ground AOD collection is to have one or more satellite pixels within the one-grid cell sampling buffer and two or more AERONET/DRAGON AOD during the one hour time window, we only assigned two quality levels: “High Quality” for coincident satellite-ground AOD pairs in the best scenario, and “Low Quality” for all others. Only coincident satellite-ground AOD pairs with high and medium quality were included in our validations. We also conducted a comparison, shown as Table S2, including all the satellite–ground AOD pairs—regardless of their quality—to examine the influence of sampling bias. In addition, we conducted sensitivity analyses on VIIRS IP AOD retrievals including both high- and degraded-quality retrievals (Supplemental Material, Table S1) and for the GOCI product at hourly scale (Supplemental Material, Table S4) with respect to its eight hourly observations per day. In the hourly comparison, we constructed hourly average AERONET AOD as the ground true value and employed the same 3 × 3 grid cells temporal comparison sampling buffer.

2.5 Evaluation Metrics

Several statistical metrics were used to describe the performance of satellite aerosol products in this study: coverage (%) describes the availability of site–day (or site–hour for GOCI data) satellite retrievals when the ground AERONET AOD were available in the temporal comparison. We include all available matched satellite retrievals when calculating the coverage regardless of the quality flag of the coincident satellite-ground AOD pairs; Pearson correlation coefficient describes the correlation between satellite retrievals and ground AOD; bias describes the average difference between satellite retrievals and ground AOD; slope is the slope of the linear regression with satellite retrievals as the dependent variable and ground AOD as the independent variable; and we calculated the percent of retrievals falling within the expected error (EE) range. For the consistency of the last~~is~~ metric among different aerosol products, we employed the same EE, ±(0.05+0.15AOD), that is established by MODIS C5 aerosol products over land in this study.

3 Results and Discussion

3.1 Spatial Variations of Aerosol Loadings

Figure 2 (a) shows the correlation coefficient of daily AOD by binned distance and Fig. 2 (b) shows the site-specific average AOD with the regional average AOD subtracted in these three cities. Figure 2 (a) indicates that the DRAGON AOD were highly correlated within a 20-km spatial range with a correlation coefficient larger than 0.9. However, results from handheld sunphotometer observations in Beijing suggest that the spatial correlation coefficients declined slowly as the distance between two measurement locations increased up to 12 km. The correlation coefficient increased slightly when the distance among two measurement locations are beyond 12 km. This can be explained by the clustered distribution of ground measurement locations in Beijing: these long location-to-location distances only occur when the two locations are located along the Chang'an Avenue and, since vehicle exhaust is one of the major sources of aerosol in Beijing, these AOD are highly correlated. The different aerosol spatial variability trends in Beijing and in the DRAGON domain can be attributed to the following reason: first, the DRAGON-Asia campaign provides real-time observation but our ground AOD observations in Beijing provide one observation at each site per day, so that the average daily AOD from DRAGON stations may have smoothed away some of the spatial heterogeneity. Second, the handheld sunphotometer may introduce larger measurement errors than DRAGON stations, due to both instrument quality and operation errors. Previous evaluation indicates that handheld stability and inaccurate pointing to the Sun significantly affects the accuracy of measurements by Mocrotops II (Ichoku et al., 2002; Morys et al., 2001). Our comparison of Microtops II AOD with nearby AERONET data yielded a slope of ~0.95, a correlation coefficient of ~0.8, and an intercept of 0.16 (Supplemental Material, Text S1), indicating that the handheld sunphotometer AOD are usable.

Even though the aerosol loadings are highly related spatially, the AOD value may differ among nearby stations (Fig. 2 (b)). In Beijing, the difference in average AOD between two neighboring sites that are ~6 km apart can be as high as 0.4, about 49% of the regional mean AOD value. The observations from DRAGON stations show smaller differences in average AOD relative to those in Beijing, but the difference between two neighboring sites can still be greater than 0.1 in Seoul—23% of the regional mean AOD value. These results indicate that spatial contrast in aerosol loading exists at local scale and finer resolution satellite aerosol products are needed to better characterize individual and population exposure of particulate pollution.

3.2 The Beijing Sampling Experiment

The GOCI aerosol product provided the highest coverage in the temporal comparison over Beijing with 73% available retrievals relative to AERONET AOD within the 1-h time window (± 30 min around the satellite overpass time), followed by the VIIRS IP (42%), VIIRS EDR (41%), MODIS Terra C6 3 km product (40%), and MODIS Aqua C6 3 km product (38%) (Supplemental Material, Table S1). Table 3 shows the statistical metrics from the temporal and spatial comparisons over Beijing. In the temporal comparison, the GOCI product provided the most accurate AOD retrievals, which slightly overestimated AOD by 0.02 on average. Other aerosol products significantly overestimated AOD with theaverage bias in the temporal comparison for VIIRS EDR, VIIRS IP, Aqua and Terra MODIS C6 3 km products equal to 0.11, 0.25, 0.21, and 0.29, respectively. Though GOCI AOD retrievals agreed well with ground AOD in the temporal comparison, with 55% of GOCI AOD retrievals at 13:00 falling within the EE, only 37% of GOCI AOD retrievals fell within the EE in the spatial comparison. The comparison including all eight hourly GOCI observations represented reduced coverage (59%), a smaller average bias (-0.006), and a larger proportion of retrievals fell within EE (59%). Thus, the GOCI product resolved the temporal and spatial variability of aerosol loadings at its designed temporal and spatial resolutions, but it tracked the small-scale spatial variability less well than the temporal variability in Beijing.

VIIRS EDR product performed well in Beijing in both the temporal and spatial comparisons, with 52% and 51% of retrievals falling within the EE in the temporal and spatial comparison, respectively. Although VIIRS IP had a relatively large positive bias (0.25) in the temporal comparison, it provided acceptable coverage with 33% retrievals falling within the EE in the spatial comparison, resolving valuable information of small-scale aerosol variability in urban areas. The MODIS C6 3 km product had the largest high bias and lowest %EE in this spatial comparison, with 16% and 26% of retrievals falling within the EE for Aqua and Terra MODIS, respectively. A previous validation study of the 3 km MODIS AOD data also reported similar retrieval errors in urban areas (Remer et al., 2013). It is notable that the ~~r~~R^2 values of the MODIS C6 3 km products is the highest in the spatial comparisons (0.68 for Aqua and 0.85 for Terra) and the linear regression statistics indicates that the low percent of retrievals falling within EE is mainly due to a relatively constant positive offset: the intercepts for Aqua and Terra are 0.22 and 0.30, respectively. One possible explanation of the positive bias of MODIS and VIIRS products is that our study domain is highly urbanized with bright surfaces, therefore is challenging for the Dark Target algorithm.

3.3 The Temporal Evaluation of AOD over the Japan-South Korea region

We first looked at the AOD retrievals distribution on one clear day, 7 May 2012, during the DRAGON period (Fig. 3). Figure 3 indicates that the sampling strategies and cloud masks differ in these five satellite aerosol products, resulting in different patterns of missing data. GOCI provided the best coverage with almost no missing data over this region. VIIRS products and MODIS products showed similar missing data in the center of the map but were less consistent at its edges; while VIIRS products showed more missing data in the lower right corner, MODIS products showed more missing in the upper right corner. VIIRS and MODIS pixels are stretched toward the edge of the scan.

VIIRS and MODIS products tended to overestimate AOD values in the urban area (Seoul), but GOCI provided accurate AOD estimates in this region. Though these 3 km products showed similar spatial distribution patterns to the 6-km products, the 3 km products demonstrated greater heterogeneity, which is valuable to analyze local aerosol sources and estimate personal air pollution exposure.

Similar to the comparisons in Beijing, the GOCI aerosol products provided the highest coverage in the temporal comparison over the Japan–South Korea region, with 74% retrievals relative to AERONET observations within the 1-h time window (±30 min around the satellite overpass time), followed by VIIRS EDR (63%), VIIRS IP (50%), Terra MODIS C6 3 km (26%), and Aqua MODIS C6 3 km (24%) (Supplemental Material, Table S1). It is notable that the seasonal missing pattern due to cloud cover and weather conditions may vary across these satellite aerosol products. However, since we did not have enough coincident satellite-ground AOD pairs to conduct seasonal evaluation, the seasonal missing patterns and seasonal performance of these satellite aerosol products were not analyzed in this study. The distributions of the coincident satellite-AERONET AOD pairs with high or medium quality are shown in Fig. 4. The distribution of the Terra MODIS C6 product is not shown here because it passes the study region in the morning, leading to potential differences in AOD distribution relatives to other sensors that pass the study region in the afternoon. This histogram is plotted with frequency of AOD retrievals from each sensor relative to the total number of matched AOD retrievals from the corresponding sensor rather than the count of AOD retrievals because these aerosol products differ in sampling strategies, leading to different total number of coincident satellite-ground AOD pairs. VIIRS EDR, VIIRS IP, and GOCI products showed a similar mode of distribution to AERONET AOD, with the peak probability around 0.2. The distribution of Aqua MODIS C6 3 km AOD had the peak around 0.3, indicating that the Aqua MODIS C6 3 km product tended to overestimate AOD in general. A previous study also reported that the MODIS C6 3 km product had a decreased proportion of low AOD values and an increased proportion of high AOD values (Remer et al., 2013) relative to the 10 km product over land, leading to a higher global average AOD. The VIIRS IP product also tended to overestimate AOD, with higher percentage of retrievals occurring at high AOD values. The distribution of GOCI data provided the best fit with AERONET data, with a correlation coefficient of 0.95, followed by VIIRS EDR (R^2 = 0.93), VIIRS IP (R^2 = 0.77), and MODIS Aqua C6 3 km product (R^2 = 0.76). The difference in the distributions of these satellite aerosol products can be partly explained by different retrieval assumptions including aerosol models, different surface reflectance and different global sampling strategies. Moreover, these satellite aerosol products differ in the valid AOD retrieval ranges, leading to differences in the distribution of extremely high and low AOD values.

The temporal comparisons over the Japan–South Korea region showed more retrievals falling within the EE and smaller biases relative to comparisons in Beijing. Figure 5 shows the frequency scatter plots showing the results of temporal comparisons over the Japan–South Korea region and the corresponding box plots showing the difference between satellite AOD retrievals and ground observations. GOCI retrievals at 13:00 LT were highly correlated with the ground AOD with an R^2 of 0.80. The linear regression of GOCI retrievals and ground AOD fell close to the 1:1 line with a small offset (0.04), and 61% of GOCI retrievals at 13:00 LT fell in the EE. Comparison including eight GOCI hourly retrievals showed a higher R^2 of 0.82 with a smaller average bias (0.02), with 66% of retrievals falling within the EE (Table 4, GOCI all obs.). The box plot indicates that GOCI retrievals overestimated AOD at high AOD values (AOD > 0.6) (Fig. 5). Thus, the GOCI product tracked the daily variability of aerosol loadings well and it provided additional information to study short-term aerosol trends. Similarly, 64% of VIIRS EDR retrievals fell into the EE with a slightly higher bias (0.05) and a slightly lower R^2 of 0.73 (Table 4). This positive bias is consistent with a previous global validation study, which reports a 0.01 bias of VIIRS EDR in East Asia (Liu et al., 2014). Though the VIIRS EDR product tended to overestimate AOD at low (AOD < 0.3) and high AOD values (AOD > 1.0), it agreed well with the AERONET observations when AOD ranged between 0.3 and 1.0 (Fig. 6).

The VIIRS IP had a linear regression slope close to 1 (1.03) against AERONET observations, but it had a consistent positive bias of 0.15 on average. Only 37% of VIIRS IP retrievals fell within the EE. The scatter plot indicates that the IP retrievals varied substantially, especially when the AOD values were low. MODIS C6 3 km products had a high positive bias of 0.08 for Aqua and 0.16 for Terra. Consistent with what was reported by a previous global evaluation study, we observed that the MODIS C6 3 km products tended to overestimate AOD and the bias increased with AOD values (Remer et al., 2013). 56% of the Aqua MODIS C6 3 km retrievals and 39% of the Terra MODIS C6 3 km retrievals fell within the EE. In general, these finer resolution aerosol products included larger bias relative to lower resolution products and researchers must be cautious when applying them by, for example,calibrating these high resolution satellite aerosol products in specified study regions and implementing appropriate data filtering strategies.

Since the GOCI product provides eight hourly observations per day, to examine the temporal variability in the accuracy of GOCI aerosol retrievals, we compared the GOCI AOD retrievals with AERONET AOD stratified by hour (Supplemental Material, Table S4). In general, the GOCI product provided high quality retrievals consistently throughout the day except that it tended to slightly overestimate AOD in the morning and underestimate AOD in the afternoon. Such temporal variability in accuracy was also reported by a previous evaluation study of the Geostationary Operational Environmental Satellite (GOES) aerosol product (Morys et al., 2001). The daily variability in the quality of GOCI retrievals may be due to changes in scattering angle, clouds and the associated Bidirectional Reflectance Distribution Function (BRDF) effects.

Ten-fold cross validation was conducted for the comparison of VIIRS and GOCI products to detect overfitting. The linear regression statistics of cross validation did not change significantly relative to the statistics of comparisons. The cross validation R^2 values of VIIRS EDR, VIIRS IP, GOCI at 13:00, and GOCI 8 observations data were 0.73, 0.51, 0.78, and 0.82, respectively. In addition, to detect the spatial variability of the satellite retrieval performance, we applied the regionally developed linear regression parameters of GOCI 8 observations data to individual AERONET station in the Japan–South Korea region. The linear regressions with the satellite AOD as the dependent variable and the fitted AOD from a regional model as the independent variable yielded R2 larger than 0.75 at all sites except the AERONET sites ‘Nara’ and ‘Osaka’, two stations located in Osaka. This result indicated that parameters from the regional dataset were valid locally. Limited by sample size, we did not apply this method to other aerosol products.

3.4 The Spatial Evaluation of AOD over the Japan-South Korea region

The mean daily AOD from different sensors and AERONET stations during the one-year period from July 2012 to June 2013 are shown in Fig. 6. These five aerosol products provided similar distributions of average AOD during the one-year period, with the highest values occurring in northeastern China and the Yangtze River delta, and the lowest values occurring in southern China and Japan. Several high-AOD-value spots appeared along the west coast of South Korea and surrounded the Seto Inland Sea, likely due to emissions from urban centers in these regions. These five maps differ in missing patterns due to their different masking approaches. The VIIRS algorithms did not retrieve AOD over inland lakes (e.g. the Taihu Lake); the GOCI product retrieved AOD over inland water; while MODIS products provided some AOD retrievals over inland lakes, with some missing data. The GOCI product did not provide high-quality retrievals at some locations in central Japan due to snow coverage in this mountain region. To maintain a consistentevaluative data filtering strategy, the inland water AOD retrievals and ground observations were removed from the validation. The VIIRS EDR product showed lower AOD values in northeastern China and South Korea relative to AOD retrievals from other sensors. The VIIRS IP product also showed lower AOD values in northeastern China, but provided higher AOD retrievals in northern Japan. This can be explained by the system bias reported in a previous study that VIIRS retrievals tend to underestimate AOD when NDVI value is low and overestimate AOD over vegetated surfaces (Liu et al., 2014). The VIIRS IP product had higher AOD values relative to the EDR product, especially over the Korean Peninsula and northern Japan.

This may be due to IP's ability to track small-scale variability which were smoothed in the EDR retrievals, or may result from the positive bias of IP observed in the temporal comparison. Because VIIRS aerosol products restrict valid AOD values to between 0.0 and 2.0, they may underestimate AOD values when the aerosol loadings are extremely high, like in northeastern China, though we lacked ground AOD data in this region to test this hypothesis. Aqua and Terra MODIS C6 3 km aerosol products showed similar spatial distribution in AOD retrievals, with higher AOD values in urban areas (e.g., over the Yangtze River Delta and North China Plain in China). GOCI presented some high AOD values in local regions such as western South Korea, around the Seto Inland Sea, and over northeastern China. However, it showed lower AOD values over the Yangtze River Delta in China. This result is consistent with the temporal comparison results shown in Fig. 5 that the GOCI product slightly overestimated AOD at high AOD values (AOD>0.6). Compared with ground AOD, all these five aerosol products overestimated AOD in Japan, where the average AOD values were relatively low. VIIRS EDR tended to slightly underestimate AOD over the Seoul region. The lack of ground AOD, especially in northeast China, makes it impossible to quantitively evaluate the spatial distribution of these aerosol products in China.

Results of the spatial comparison over DRAGON-Asia region are shown in Table 4. Satellite aerosol products performed better in tracking the day-to-day variability relative to tracking their spatial patterns. In the spatial comparison, all the satellite aerosol products showed lower ~~r~~R^2 and larger offset with less retrievals falling into the EE. GOCI product provided the highest accuracy, with a small positive bias of 0.03 and 48% of retrievals falling in the EE, followed by VIIRS EDR, with a positive offset of 0.16 and 41% of retrievals falling in the EE. In contrast, VIIRS IP and MODIS C6 3 km had large positive biases, and less than 30% of retrievals fell within the EE due to larger noise (related to the finer resolutions). There is evidence that this positive bias includes systematic errors due to improper characterization of surface reflectance, uncertainties in the assumed aerosol model, and cloud masking. The 3 km MODIS products sample fewer reflectance pixels to retrieve aerosol pixels relative to the 10 km products, introducing sporadic unrealistic high AOD retrievals that are avoided more successfully by the 10 km products (Munchak et al., 2013). Previous studies also reported that improper characterization of bright urban surfaces, a known difficult situation for the Dark Target algorithm, led to positive bias in urban/suburban regions (Munchak et al., 2013; Remer et al., 2013). The VIIRS IP product is retrieved at the reflectance pixel level without aggregation, thus it is expected to include more noise. Though these finer resolution aerosol products did not fully track the spatial trends of aerosol loading at their designed resolution, they provide additional information about aerosol spatial distribution and will benefit exposure assessments at local scales.

To examine possible sampling bias due to our data inclusion criteria, we performed temporal and spatial comparisons including all the coincident satellite-ground AOD pairs over the Japan–South Korea region (Supplemental Material, Table S2). There is no significant change in the evaluation metrics after including pairs with low quality. Thus, the validation results are robust and there is no evidence for sampling bias. We validated the VIIRS IP AOD retrievals with degraded quality over the Japan–South Korea region and observed lower correlation coefficients, higher biases, and less retrievals falling within the EE in both the temporal and spatial comparisons (Supplemental Material, Table S3). This result suggests to use only high-quality VIIRS IP retrievals. We also validated the GOCI AOD retrievals with different quality over the Japan–South Korea region. Including medium- and low-quality GOCI retrievals decreased the accuracy, but significantly increased the coverage (Supplemental Material, Table S5). By including the retrievals having quality flags equal to both 3 and 2, the coverage increased from 27% to 38% in the temporal comparison over the Japan–South Korea region, while the average bias increased by 0.01 and the percentage of retrievals falling within the EE decreased by 7%. Thus, including retrievals with medium quality might be acceptable, depending on study objectives. Due to the relatively small number of matched observations, analysis of the correlation between quality of satellite aerosol retrievals and satellite viewing angles were beyond the scope of this analysis. However, previous studies reported that towards the edge of the scan, VIIRS EDR tends to underestimate AOD over land (Liu et al., 2014).

4 Conclusion

In this work, the intra-city variability of aerosol loadings were examined with ground AOD from the DRAGON-Asia campaign and our mobile sampling campaign in Beijing. Five emerging high-resolution satellite aerosol products are evaluated by comparing them with ground AOD from AERONET, DRAGON, and handheld sunphotometers over East Asia in 2012 and 2013. We observed variability in both correlation coefficients and average AOD values among ground AOD observation sites in three urban centers in Asia. Evaluation results indicated a) that the 6-km resolution products—VIIRS EDR and GOCI—provided more accurate retrievals with higher coverage relative to the higher resolution products—VIIRS IP, Terra and Aqua MODIS C6 3 km products—in both temporal comparisons and spatial comparisons; however, VIIRS IP and MODIS C6 3 km products provide additional information about fine-resolution aerosol spatial distribution and will benefit exposure assessments at local scales; b) satellite aerosol products resolved the day-to-day aerosol loading variability better than the spatial aerosol loading variability; and c) satellite products performed less well in Beijing relative to the Japan-South Korea region, indicating that retrieval in urban areas is challenging. These satellite aerosol products have their own advantages and disadvantages. For example, the GOCI aerosol product provides high accuracy AOD retrievals eight times per day, but it only covers East Asia; the VIIRS EDR product provides high accuracy AOD retrievals and global coverage once per day, but its 6 km resolution is relatively low; the MODIS C6 3 km products provide high resolution AOD retrievals with global coverage, but have positive bias in urban regions. Researchers need to apply these aerosol products according to specified research objectives and study design. The performance of these aerosol products over Beijing and the Japan-South Korea region demonstrates that satellite aerosol products can track the small-scale variability of aerosol loadings. High-resolution satellite aerosol products provide valuable information for the spatial and temporal characterization of

$PM_{2.5}$ at local scales. Future studies with additional ground AOD observations at fine spatial and temporal scale will help us analyze air pollution patterns and further validate satellite products.

Acknowledgment

The work of Liu and Xiao was partially supported by the NASA Applied Sciences Program (grants NNX11AI53G and NNX14AG01G, PI: Liu). We would like to acknowledge the AERONET team, Prof. I. Sano and the DRAGON-Japan team, the Yonsi team and their collaborators in S. Korea, and CARSNET and CAS teams in and around Beijing for providing data support in this study. The AERONET project is supported NASA EOS project office, and by Hal B. Maring, Radiation Sciences Program, NASA Headquarters.

References

Anderson, T. L., Charlson, R. J., Winker, D. M., Ogren, J. A., and Holmén, K.: Mesoscale Variations of Tropospheric Aerosols*, Journal of the Atmospheric Sciences, 60, 119-136, 2003.

Choi, M., Kim, J., Lee, J., Kim, M., Je Park, Y., Jeong, U., Kim, W., Holben, B., Eck, T. F., Lim, J. H., and Song, C. K.: GOCI Yonsei Aerosol Retrieval (YAER) algorithm and validation during DRAGON-NE Asia 2012 campaign, Atmos. Meas. Tech. Discuss., 8, 9565-9609, 10.5194/amtd-8-9565-2015, 2015.

Holben, B., Eck, T., Slutsker, I., Tanre, D., Buis, J., Setzer, A., Vermote, E., Reagan, J., Kaufman, Y., and Nakajima, T.: AERONET—A federated instrument network and data archive for aerosol characterization, Remote sensing of environment, 66, 1-16, 1998.

Hsu, N. C., Jeong, M. J., Bettenhausen, C., Sayer, A. M., Hansell, R., Seftor, C. S., Huang, J., and Tsay, S. C.: Enhanced Deep Blue aerosol retrieval algorithm: The second generation, Journal of Geophysical Research: Atmospheres, 118, 9296-9315, 10.1002/jgrd.50712, 2013.

Ichoku, C., Levy, R., Kaufman, Y. J., Remer, L. A., Li, R. R., Martins, V. J., Holben, B. N., Abuhassan, N., Slutsker, I., and Eck, T. F.: Analysis of the performance characteristics of the five‐channel Microtops II Sun photometer for measuring aerosol optical thickness and precipitable water vapor, Journal of Geophysical Research: Atmospheres (1984–2012), 107, AAC 5-1-AAC 5-17, 2002.

Jackson, J. M., Liu, H., Laszlo, I., Kondragunta, S., Remer, L. A., Huang, J., and Huang, H. C.: Suomi‐NPP VIIRS aerosol algorithms and data products, Journal of Geophysical Research: Atmospheres, 118, 12,673-612,689, 2013.

Lee, J., Kim, J., Song, C. H., Ryu, J.-H., Ahn, Y.-H., and Song, C.: Algorithm for retrieval of aerosol optical properties over the ocean from the Geostationary Ocean Color Imager, Remote sensing of environment, 114, 1077-1088, 2010.

Levy, R., Mattoo, S., Munchak, L., Remer, L., Sayer, A., Patadia, F., and Hsu, N.: The Collection 6 MODIS aerosol products over land and ocean, Atmospheric Measurement Techniques, 6, 2989-3034, 2013.

Levy, R. C., Remer, L. A., and Dubovik, O.: Global aerosol optical properties and application to Moderate Resolution Imaging Spectroradiometer aerosol retrieval over land, Journal of Geophysical Research: Atmospheres (1984–2012), 112, 2007.

Levy, R. C., Remer, L. A., Kleidman, R. G., Mattoo, S., Ichoku, C., Kahn, R., and Eck, T.: Global evaluation of the Collection 5 MODIS dark-target aerosol products over land, Atmospheric Chemistry and Physics, 10, 10399-10420, 2010.

Li, C., Lau, A.-H., Mao, J., and Chu, D. A.: Retrieval, validation, and application of the 1-km aerosol optical depth from MODIS measurements over Hong Kong, Geoscience and Remote Sensing, IEEE Transactions on, 43, 2650-2658, 2005.

Liu, H., Remer, L. A., Huang, J., Huang, H. C., Kondragunta, S., Laszlo, I., Oo, M., and Jackson, J. M.: Preliminary evaluation of S‐NPP VIIRS aerosol optical thickness, Journal of Geophysical Research: Atmospheres, 119, 3942-3962, 2014.

Liu, Y., Franklin, M., Kahn, R., and Koutrakis, P.: Using aerosol optical thickness to predict ground-level PM 2.5 concentrations in the St. Louis area: a comparison between MISR and MODIS, Remote sensing of environment, 107, 33-44, 2007.

Morys, M., Mims, F. M., Hagerup, S., Anderson, S. E., Baker, A., Kia, J., and Walkup, T.: Design, calibration, and performance of MICROTOPS II handheld ozone monitor and Sun photometer, Journal of Geophysical Research: Atmospheres (1984–2012), 106, 14573-14582, 2001.

Munchak, L., Levy, R., Mattoo, S., Remer, L., Holben, B., Schafer, J., Hostetler, C., and Ferrare, R.: MODIS 3 km aerosol product: applications over land in an urban/suburban region, Atmospheric Measurement Techniques Discussions, 6, 1683-1716, 2013.

Ohara, T., Akimoto, H., Kurokawa, J.-i., Horii, N., Yamaji, K., Yan, X., and Hayasaka, T.: An Asian emission inventory of anthropogenic emission sources for the period 1980–2020, Atmospheric Chemistry and Physics, 7, 4419-4444, 2007.

Otter, L., Scholes, R., Dowty, P., Privette, J., Caylor, K., Ringrose, S., Mukelabai, M., Frost, P., Hanan, N., and Totolo, O.: The Southern African regional science initiative (SAFARI 2000): wet season campaigns, South African Journal of Science, 98, p. 131-137, 2002.

Park, M., Song, C., Park, R., Lee, J., Kim, J., Lee, S., Woo, J.-H., Carmichael, G., Eck, T. F., and Holben, B. N.: New approach to monitor transboundary particulate pollution over Northeast Asia, Atmospheric Chemistry and Physics, 14, 659-674, 2014.

Remer, L., Mattoo, S., Levy, R., and Munchak, L.: MODIS 3 km aerosol product: algorithm and global perspective, Atmospheric Measurement Techniques Discussions, 6, 69-112, 2013.

Sano, I., Mukai, S., Holben, B., Nakata, M., Yonemitsu, M., Sugimoto, N., Fujito, T., Hiraki, T., Iguchi, N., and Kozai, K.: DRAGON-West Japan campaign in 2012: regional aerosol measurements over Osaka, SPIE Asia-Pacific Remote Sensing, 2012, 85231M-85231M-85236.

Seo, S., Kim, J., Lee, H., Jeong, U., Kim, W., Holben, B., Kim, S., Song, C., and Lim, J.: Spatio-temporal variations in PM 10 concentrations over Seoul estimated using multiple empirical models together with AERONET and MODIS data collected during the DRAGON-Asia campaign, Atmospheric Chemistry and Physics Discussions, 14, 21709-21748, 2014.

Tiwari, S., and Singh, A.: Variability of aerosol parameters derived from ground and satellite measurements over Varanasi located in the Indo-Gangetic Basin, Aerosol Air Qual Res, 13, 627-638, 2013.

Vermote, E., Justice, C., and Csiszar, I.: Early evaluation of the VIIRS calibration, cloud mask and surface reflectance Earth data records, Remote sensing of environment, 148, 134-145, 2014.

Table 1. Characteristics and quality control criteria of satellite aerosol products.

Dataset	Including Criteria	Resolution	Coverage
VIIRS EDR	Quality Flag=High	6 km, daily	Global
VIIRS IP	Quality Flag=High	0.75 km, daily	Global
GOCI	Quality Flag=3	6 km, 8 hourly obs. per day	East Asia
Aqua MODIS C6 3 km	Quality Flag=3	3 km, daily	Global
Terra MODIS C6 3 km	Quality Flag=3	3 km, daily	Global

Table 2. Characteristics of ground AOD measurement datasets.

		Temporal Comparison	Spatial Comparison
Beijing	Data Set	AERONET	Microtops II
	Including Criteria	Level 1.5	Median/Std. Dev. <2
	Study Period	Jul. 2012 – Jun. 2013	Jan. 2012 – Jun. 2013
East Asia	Data Set	AERONET	DRAGON
	Including Criteria	Level 2.0	Level 2.0
	Study Period	Jul. 2012 – Jun. 2013	Feb. 15 – May 31, 2012

Table 3. Statistics of the temporal and spatial comparisons between satellite retrievals and ground AOD measurements at 550 nm in Beijing.

	N	R^2	Slope	Intercept	Bias	%EE
Temporal Comparison						
VIIRS EDR	90	0.70	0.96**	0.12**	0.11	52
VIIRS IP	133	0.63	1.00**	0.25**	0.25	32
GOCI	142	0.88	0.95**	0.05	0.02	55
GOCI all obs.	957	0.88	0.98**	0.008	-0.006	59
Aqua MODIS C6 3 km	119	0.81	1.05**	0.19**	0.21	44
Terra MODIS C6 3 km	133	0.80	0.99**	0.30**	0.29	25
Spatial Comparison						
VIIRS EDR	108	0.14	0.25**	0.34**	0.04	51
VIIRS IP	150	0.16	0.34**	0.45**	0.18	33
GOCI	~~208~~124	~~0.44~~ 0.51	~~0.75~~0.74**	~~0.07~~0.23**	-~~0.11~~0.00	~~31~~37
Aqua MODIS C6 3 km	77	0.68	1.19**	0.22**	0.31	16
Terra MODIS C6 3 km	73	0.85	1.00**	0.30**	0.30	26

** p-value < 0.01

Table 4. Statistics of the temporal and spatial comparisons between satellite retrievals and ground AOD measurements at 550 nm over Japan-South Korea region.

	N	R^2	Slope	Intercept	Bias	%EE
Temporal Comparison						
VIIRS EDR	~~601~~600	0.74 ~~0.73~~	0.96**	0.06**	0.05	64
VIIRS IP	~~437~~424	0.55	1.03**	0.14**	0.15	37
GOCI	~~343~~317	0.80	~~1.04~~1.02**	~~0.02~~0.04**	~~0.04~~0.05	~~62~~61
GOCI all obs.	~~2774~~2547	0.82	~~1.03~~1.02**	~~0.00~~0.01*	~~0.01~~0.02	66
Aqua MODIS C6 3 km	~~180~~179	0.71	1.00**	0.08**	0.08	56
Terra MODIS C6 3 km	197	0.70	1.06**	0.14**	0.16	39
Spatial Comparison						
VIIRS EDR	144	0.53	0.96**	0.18**	0.16	41
VIIRS IP	229	0.60	1.11**	0.21**	0.26	26
GOCI	196	0.79	1.19**	-0.09**	0.03	48
Aqua MODIS C6 3 km	108	0.81	1.26**	0.07*	0.19	28
Terra MODIS C6 3 km	132	0.73	1.00**	0.23**	0.23	27

* p-value < 0.05

** p-value < 0.01

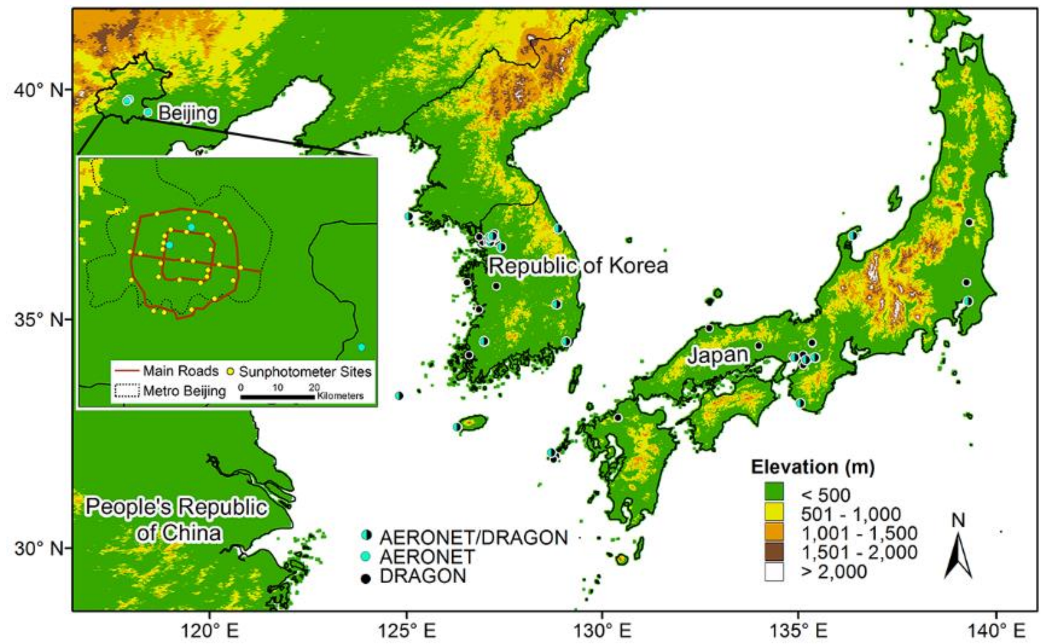

Figure 1. Study area showing all the ground AOD measurement sites.

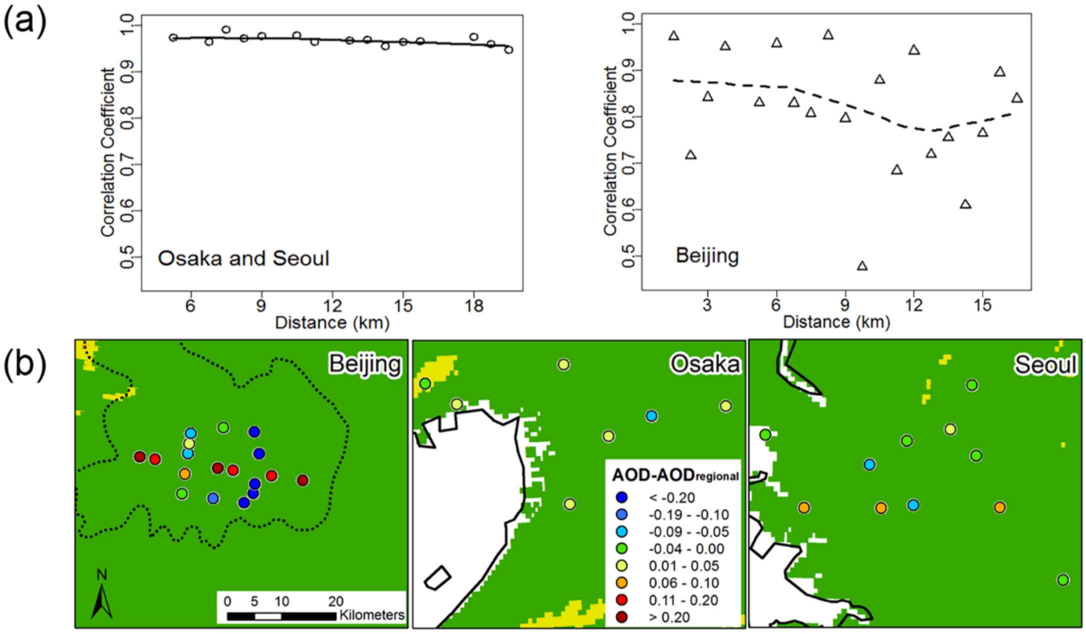

Figure 2. (a) The station to station correlation coefficients of daily mean AOD stratified by distance over (left) DRAGON-Asia region (right) Beijing region. The line is the Loess curvy. (b) The spatial distribution of average AOD in these three cities. The background color shows the elevation with the same color scale as in Figure 1.

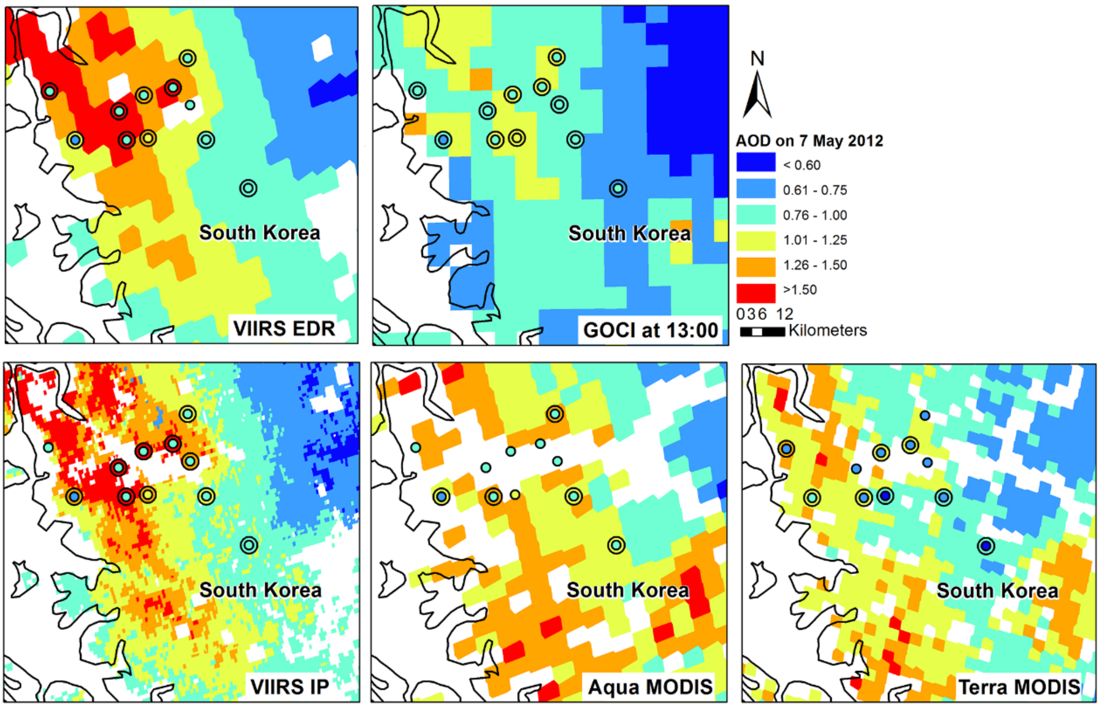

Figure 3. The AOD retrievals at 550 nm from different satellite aerosol products at their designed resolution on 7 May 2012. Coincident Satellite-DRAGON AOD pairs are shown in double circles: the inner circle is the average DRAGON observation within ±30 min of satellite overpass and the outer circle is the satellite retrieval that the DRAGON stations falls in.

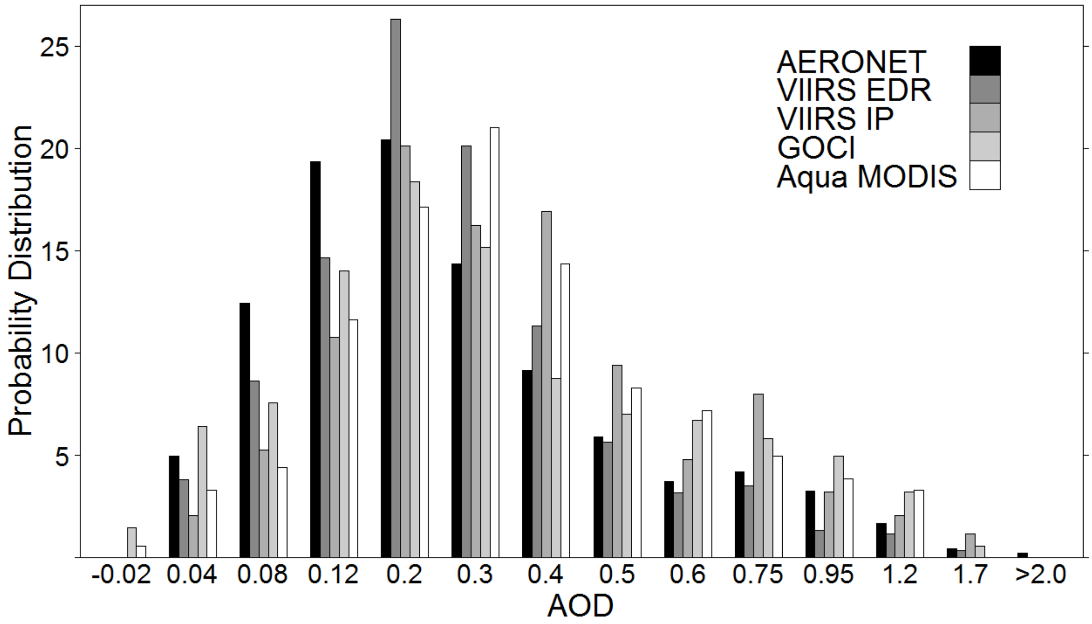

Figure 4. Histogram for the matched satellite AOD retrievals and AERONET measurements. The x-axis shows AOD values and the y-axis shows the frequency of AOD observations from each sensor relative to the total number of matched AOD observations from the corresponding sensor.

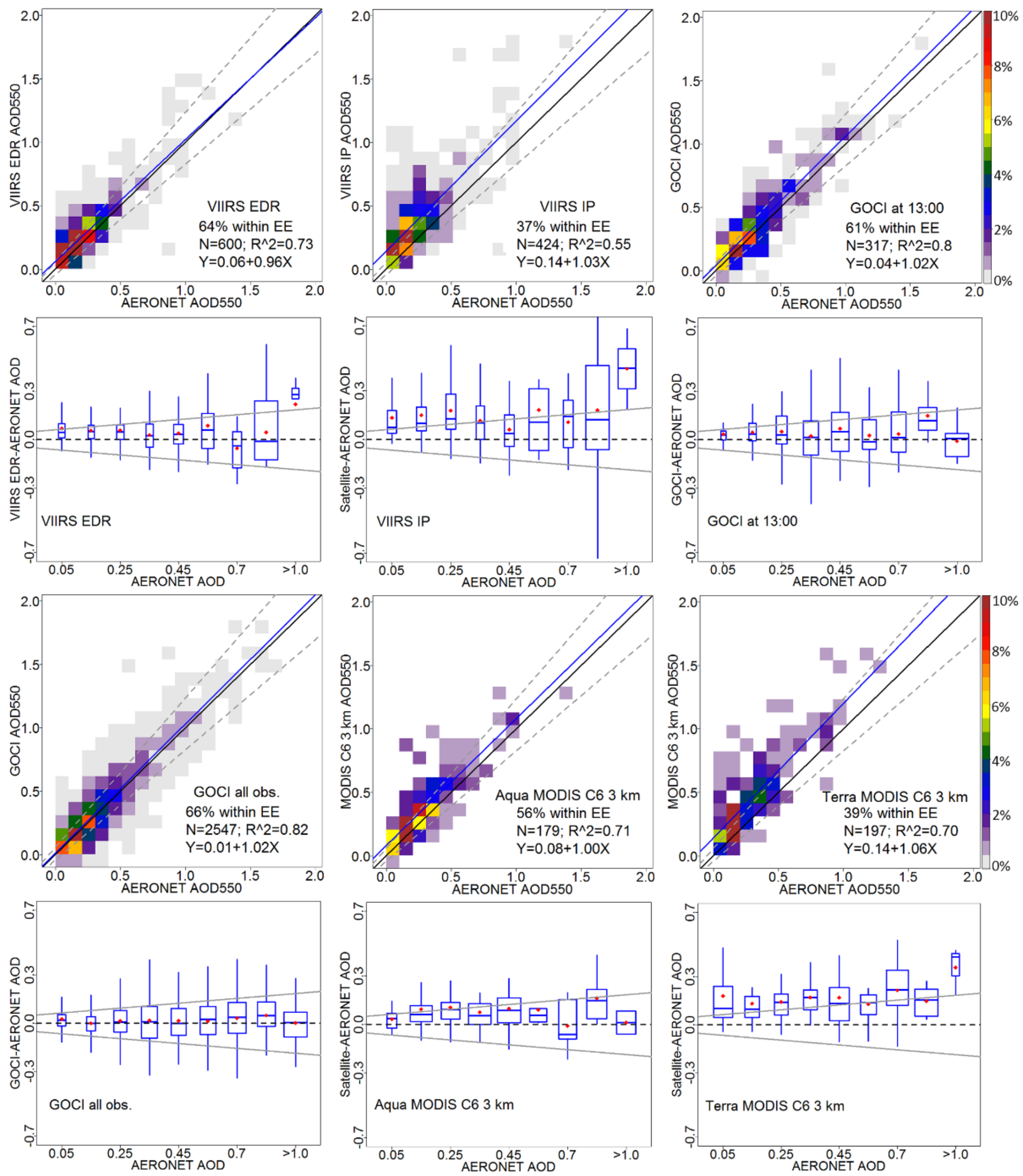

Figure 5. Upper - frequency scatter plots of satellite AOD retrievals against AERONET AOD measurements at 550 nm over the Japan-South Korea region. The linear regression is shown as solid blue line, the boundary lines of the expected error are shown in the dash lines, and the one-one line is shown as solid black lines for reference. Lower - box plots of AOD errors (satellite – AERONET) versus AERONET AOD over the Japan-South Korea region. The one-one line (zero error) is shown as a dash line and the boundary lines of the expected error are shown as gray solid lines. For each box-whisker, its properties and representing statistics include: width is σ of the satellite AOD; height is the interquartile range of AOD error; whisker is the 2 σ of the AOD error; middle line is the median of the AOD error; and

red dot is the mean of the AOD error.

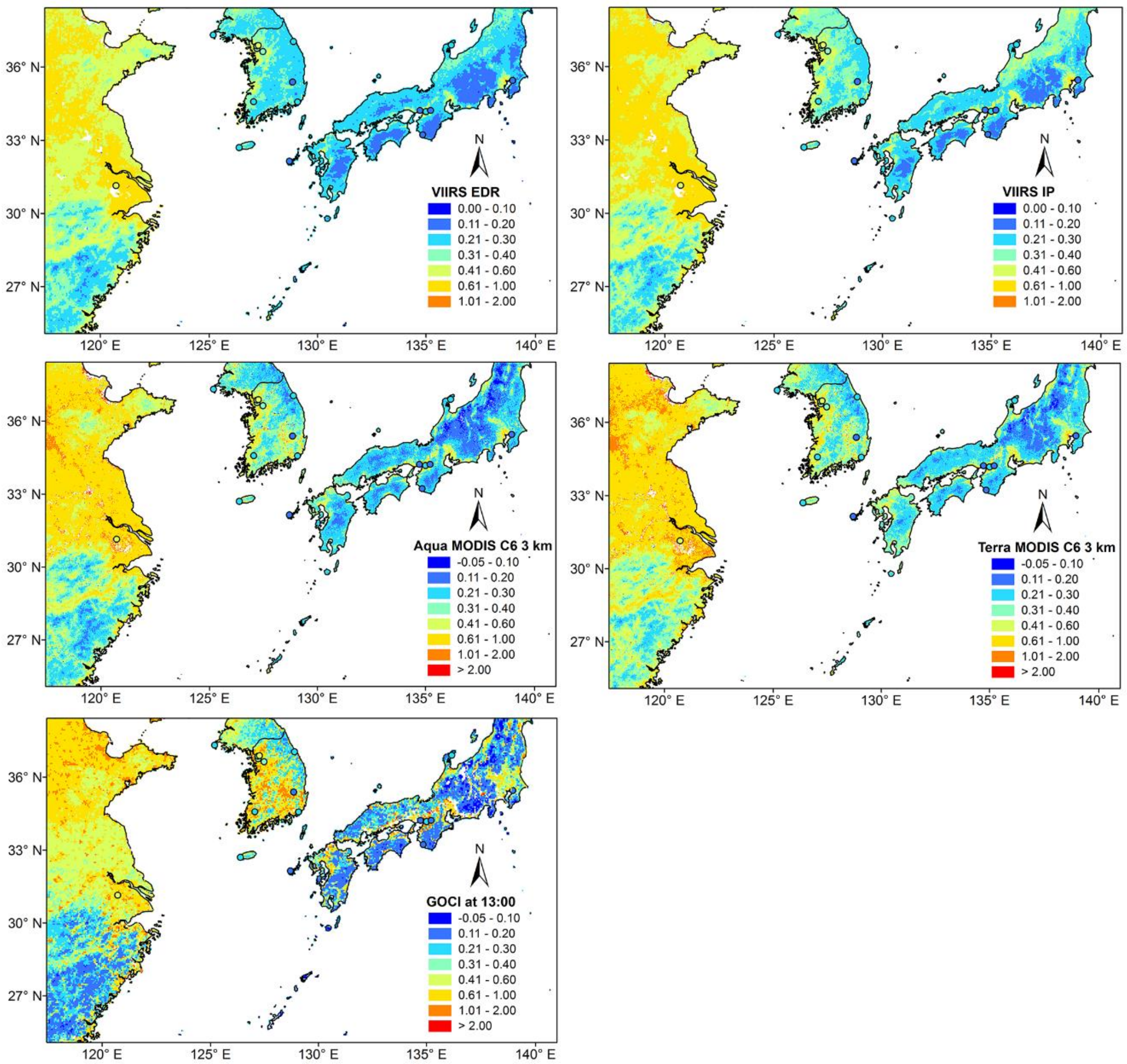

Figure 6. The distributions of the twelve months average AOD values from July 2012 to June 2013 from VIIRS EDR, VIIRS IP, Aqua MODIS C6 3 km, Terra MODIS C6 3 km, and GOCI datasets.

We appreciate your time to read our manuscript and give us these comments. Below please find our reply.

Main review points
=================

The authors have addressed many of my original concerns as well as points raised by reviewer 2. However, the following major items remain to be resolved:

1. The motivation of the study is still unclear:

- Why would the spacing of ground-based measurements make a difference for satellite product evaluation, in which each individual retrieval is compared to only one ground-based measurement at a time?

The retrieval errors of satellite aerosol products are affected by various factors such as meteorological conditions, terrain, emissions source properties, and changing aerosol mixtures. In previous global validation studies, AOD retrievals and their errors are treated as spatially independent because the validation sites are far apart and the AOD retrieval resolutions are relatively low. In other words, these studies essentially evaluated whether a satellite product can accurately track AOD values in time. At a smaller scale, another important question arises: can a satellite product accurately represent the aerosol characteristics in space? This question can only be answered with the help of spatially dense ground measurements. Motivated by this question, our spatial analysis evaluates whether these AOD products can accurately reflect the fine-scale aerosol characteristics.

We modified the sentences as follows on page 7, line 12-19, "However, previous evaluation studies comparing these emerging satellite aerosol retrievals with AERONET data were mostly at the global scale. AOD retrievals and their errors are treated as spatially independent because the validation sites are far apart and the AOD retrieval resolutions are relatively low. Therefore, these studies evaluated how accurately a satellite product can track AOD values in time. With the help of spatially dense ground measurements, a regional-scale evaluation can evaluate satellite aerosol products' abilities to accurately reflect the fine-scale aerosol characteristics in space."

- While particulate matter, including PM2.5, will certainly contribute to atmospheric aerosol loading, there is no way to directly infer PM from AOD. Also, smaller particles in particular tend to travel far, reducing the utility of high-resolution aerosol retrievals for PM2.5 in particular. I agree that monitoring air pollution is important, but I do not see how particulate matter distribution would *directly* motivate this study.

Response: We modified the introduction section as follows on page 3-5:" Aerosols play a critical role in atmospheric processes as well as global climate change. Rapid economic growth and increasing fossil fuel usage have significantly affected aerosol formation and transportation in East Asia. From 1980–2003, the emissions of black carbon, organic carbon, SO_2, and NO_x increased by 28%, 30%, 119%, and 176%, respectively (Ohara et al., 2007). Aerosols are also noted for its adverse health impacts, such as increased cardiovascular and respiratory morbidity and mortality (Lim et al., 2013). The continuous air quality degradation together with high population density have raised serious public health concerns in East Asia.

Satellite remote sensing data have been applied to characterize aerosol global distribution and temporal variation. Although the primary goal of satellite observations is to advance our understanding of the climate system, the comprehensive spatial coverage and growing time series of satellite retrievals benefit various applications, including monitoring ground level air pollution, especially particulate matter (PM). The traditional ground-based air quality monitoring networks are expensive to operate and have limited spatial coverage. For example, most PM monitoring stations in China are located in urban centers and the monitoring network only covers about 360 out of the more than 3,000 counties. Most developing countries, where PM levels are dangerously high, have little or no regular ground monitoring network. These limitations of ground measurements result in insufficient information to conduct studies about pollution sources, distribution, and consequent health impacts. Satellites provide continuous, high-coverage observations of aerosol loadings and various approaches have been developed to estimate ground-level PM concentrations from satellite retrievals (Ma et al., 2014;Xu et al., 2015). Estimates of ground-level PM concentrations from satellite observations have been used in epidemiological studies and benefited policy making (Strickland et al., 2015;Evans et al., 2013).

The most widely used satellite aerosol sensor, the Moderate Resolution Imaging Spectroradiometer (MODIS), has 36 spectral bands, acquiring data in wavelength from 0.41 μm to 15 μm and providing information about atmospheric aerosol properties (Anderson et al., 2003). Two identical MODIS instruments are aboard the National Aeronautics and Space Administration (NASA) Terra and Aqua satellites, which fly over the study area at around 10:30 and 13:30 LT, respectively. Several algorithms have been developed to retrieve aerosol optical depth (AOD) from MODIS data over land, such as the Dark-Target (Levy et al., 2013) algorithm and the Deep-Blue (Hsu et al., 2013) algorithm, providing AOD retrievals at 550 nm with global coverage. The widely used 10 km resolution MODIS aerosol products provides valuable information on aerosol distribution in space and time, and has been widely used to characterize aerosol dynamics and distribution, simulate climate change, and assess population PM exposure (Levy et al., 2013;Levy et al., 2010). However, the 10 km product cannot depict small-scale PM heterogeneity. Though a previous study (Anderson et al., 2003) indicated that the aerosol loading is homogeneous at horizontal scales within 200 km, that study is conducted over the ocean, which provides a homogeneous surface, leading to reduced aerosol spatial variability. The variability of aerosol loading at local scales in urban areas with complex land surface and meteorological conditions are expected to be greater (Li et al., 2005). Accurately characterizing local-scale aerosol heterogeneity is critical for assessing population PM exposure, detecting small smoke plums, and analyzing aerosol-cloud process. To resolve small-scale aerosol features, satellite aerosol products with higher resolutions and accuracy are urgently needed."

2. If this is not a communication problem, reprojection and data loss thus incurred does not seem necessary and is very likely detrimental to the reliability of the evaluation.

Response: Reprojection is necessary because the original satellite data is in geographic coordinate systems, with longitude and latitude in decimal degrees. We need to acquire the Euclidean distances between satellite pixels and ground stations to construct the coincident satellite-ground AOD pairs, thus we need to project the data to a projected coordinate system that better preserve distance. There is no data loss in this process of converting decimal degrees to meters. To clarify this issue, we modified the sentence as follows on page 10, line 21-24, "Since satellite pixel coordinates are provided in a geographic coordinate system, to acquire the accurate Euclidean distance between satellite pixels and ground measurement locations, the coordinates of all the data were converted to the JGD_2000_UTM_Zone_52N coordination system."

3. I still have strong concerns regarding the statistical analysis of data sets not normally distributed. If this cannot be resolved, at least it needs to be explicitly discussed, and all results need to be carefully appraised in this light.

Response: We evaluated these satellite aerosol products based on evaluation metrics including coverage, Pearson correlation coefficient, bias, percent of retrievals falling within the previously established expected error range, and slope and intercept from linear regression, and these statistics, except the regression slope and intercept, are not limited to normally distributed data. Our linear regression analysis results are consistent with our findings based on other evaluation metrics. To clarify this, we modified the sentences as follows on page 14, line 29-page 15, line 14,

"Several metrics were used to evaluate the performance of satellite aerosol products in this study. Coverage (%) describes the availability of site–day (or site–hour for GOCI data) satellite retrievals when the ground AERONET AOD were available in the temporal comparison. We include all available matched satellite retrievals when calculating the coverage regardless of the quality flag of the coincident satellite-ground AOD pairs. Pearson correlation coefficient describes the correlation between satellite retrievals and ground AOD. Bias describes the average difference between satellite retrievals and ground AOD. We calculated the percent of retrievals falling within the expected error (EE) range. For the consistency of this metric among different aerosol products, we employed the same EE, ±(0.05+0.15AOD), which is established during the global validation of MODIS C5 aerosol product over land, in this study. In addition, linear regression with satellite retrievals as the dependent variable and ground AOD as the independent variable was employed. The slopes and intercepts from linear regressions were reported."

Furthermore, we conducted tests for normality of residuals from linear regressions for the temporal comparisons over Japan-South Korea region and the results are listed below (Table 1). In general, the residuals from original data were right skewed and the residuals from log-transformed data were left skewed. Although the log transformation made some datasets, e.g. VIIRS IP, Aqua MODIS C6 3 km, more symmetrical, it decreased the symmetry of other datasets, e.g. GOCI. We evaluated the performance of satellite retrievals by using log-transformed data, after adding 0.05 to MODIS and GOCI retrievals and corresponding AERONET retrievals (Table 2). To discuss these results, we added the following sentences on page 15, line 17-26,

"The residuals of the linear regressions were slightly skewed (Supplemental Material, Table S1), indicating that one assumption of linear regression, normality of the residual distribution, was not fully met. However, log-transformation did not necessarily make the residual distribution more normal (Supplemental Material, Table S1) and log-transformation led to loss of physical meaning of the evaluation metrics and made the evaluation metrics incomparable to previous studies. All things considered, we used the original data in this analysis. We conducted a sensitivity analysis using log-transformed data after adding 0.05 to GOCI, Aqua and Terra MODIS C6 3 km satellite retrievals as well as the corresponding AERONET retrievals over Japan-South Korea region."

and on page 24, line 20-22,

"Results from linear regressions with log-transformed data (Supplemental Material, Table S7) indicated that GOCI aerosol products provided the best estimate of ground measured AOD, followed by VIIRS EDR and MODIS Aqua C6 3 km products."

We also added two tables, Table S1 and Table S7.

Table 1. Statistics of tests of normality of residuals from linear regressions for temporal comparisons between satellite retrievals and ground AOD measurements at 550 nm over Japan-South Korea region

Satellite	Skewness	P-value of tests of normality		
		Kolmogorov-Smirnov	Cramer-von Mises	Anderson-Darling
Original				
VIIRS EDR	0.72	<0.01	<0.01	<0.01

VIIRS IP	1.31	<0.01	<0.01	<0.01
GOCI	0.81	<0.01	<0.01	<0.01
GOCI all obs.	0.63	<0.01	<0.01	<0.01
Aqua MODIS C6 3 km	1.85	<0.01	<0.01	<0.01
Terra MODIS C6 3 km	1.24	<0.01	<0.01	<0.01
Log-transformed				
VIIRS EDR	-0.43	<0.01	<0.01	<0.01
VIIRS IP	0.10	<0.01	<0.01	<0.01
GOCI	-2.40	<0.01	<0.01	<0.01
GOCI all obs.	-1.75	<0.01	<0.01	<0.01
MODIS_A	0.25	<0.01	<0.01	<0.01
MODIS_T	0.76	<0.01	<0.01	<0.01

Table 2. Statistics of the temporal comparisons between satellite retrievals and ground AOD measurements at 550 nm over Japan-South Korea region, with and without log-transformation.

		Original			Log-transformed		
	N	R^2	Slope	Intercept	R^2	Slope	Intercept
VIIRS EDR	600	0.74	0.96	0.06	0.60	0.77	-0.07
VIIRS IP	424	0.55	1.03	0.14	0.50	0.69	-0.01
GOCI	317	0.80	1.02	0.04	0.66	0.91	-0.01
GOCI all obs.	2547	0.82	1.02	0.01	0.66	1.00	-0.01
Aqua MODIS C6 3 km	179	0.71	1.00	0.08	0.68	0.90	0.03
Terra MODIS C6 3 km	197	0.70	1.06	0.14	0.60	0.74	0.03

Detailed comments
=================

- 4-8, 4-14 and 23-29: I do not see from the manuscript how PM2.5 characterization at the ground would be helped by having a high-resolution AOD product available. AOD is a very different parameter and represents a vertical integral rather than information pertaining to ground level only. There may of course be links between particulate matter (of whatever size fraction) and AOD, but this link is complex and not self-evident.

Response: The original design objective of most satellite aerosol observations is to support the research of global climate processes. However, the advantages of satellite observations, including comprehensive spatial coverage and historical data records, make satellite AOD retrievals a powerful tool to monitor particle air pollution especially in regions with little or no ground monitoring. Developing innovative models to estimate ground level PM concentrations from satellite AOD retrievals is becoming a rapidly evolving new research area. To make this clear, we modified these sentences in the introduction section as follows on page 3, line 15-page 4, line 2,

"Satellite remote sensing data have been applied to characterize aerosol global distribution and temporal variation. Although the primary goal of satellite observations is to advance our understanding of the climate system, the comprehensive spatial coverage and growing time series of satellite retrievals benefit various applications, including monitoring ground level air pollution, especially particulate matter (PM). The traditional ground-based air quality monitoring networks are expensive to operate and have limited spatial coverage. For example, most PM monitoring stations in China are located in urban centers and the monitoring network only covers about 360 out of the more than 3,000 counties. Most

developing countries, where PM levels are dangerously high, have little or no regular ground monitoring network. These limitations of ground measurements result in insufficient information to conduct studies about pollution sources, distribution, and consequent health impacts. Satellites provide continuous, high-coverage observations of aerosol loadings and various approaches have been developed to estimate ground-level PM concentrations from satellite retrievals (Ma et al., 2014;Xu et al., 2015). Estimates of ground-level PM concentrations from satellite observations have been used in epidemiological studies and benefited policy making (Strickland et al., 2015;Evans et al., 2013)."

- 6-22: In your response to one of the comments in the previous review round you argue that you require spatially concentrated ground-based observations, and proceed to calling this "intensive" observations.
I have two queries: 1) I don't think "intensive" is any clearer than the previously used term "high resolution". I would suggest "spatially concentrated", "closely spaced" or something similar, or possibly a description instead of an adjective. 2) I am not convinced that the evaluation of a 3km product requires closely spaced observations. Just like a 10km, the performance of the algorithm can be compared to ground-based observations. For the quality of the individual retrieval, any potential spatial heterogeneity hardly seems to be relevant. Or is there a particular reason to suspect that the adjoining 3km pixel would have reduced quality? Since this continues to be the central motivation of the study, I would suggest some further explanations.

Response: Thank you for this suggestion. We changed the word "intensive" to "spatially concentrated" in the text. We need spatially concentrated ground measurements to evaluate how satellite AOD can accurately reflect the gradients of aerosol loading in space. The retrieval errors of satellite aerosol products are affected by various factors and are spatially dependent in most cases. Previous global validation studies treated AOD retrievals and their errors as spatially independent and evaluated accuracy of satellite product in time because the validation sites are far apart and the satellite AOD resolutions are relatively low. To evaluate these high-resolution aerosol products, we would like to test the accuracy of satellite product in space. Thus we need spatially dense ground measurements. We modified the sentences as follows on page 7, line 12-19,

"However, previous evaluation studies comparing these emerging satellite aerosol retrievals with AERONET data were mostly at the global scale. AOD retrievals and their errors are treated as spatially independent because the validation sites are far apart and the AOD retrieval resolutions are relatively low. Therefore, these studies evaluated how accurately a satellite product can track AOD values in time. With the help of spatially dense ground measurements, a regional-scale evaluation can evaluate satellite aerosol products' abilities to accurately reflect the fine-scale aerosol characteristics in space."

- 9-18 (was: 16-27 in previous round). Yes, I understand that reprojection is needed to create a grid.

1. But why do you need a NEW grid to compare gridded satellite data to non-gridded ground-based data? The quality of the satellite data set will be reduced.

Response: These satellite data are provided in unprojected geographic coordinates, making distance calculation inaccurate in our regional study domain. We need the distance between satellite pixels and AERONET stations to construct coincident satellite-ground AOD pairs for comparisons, thus we must project the data. The grid was used to constructed comparison buffers. To ensure the consistence of the comparison buffers, we created the grid. We modified the sentence as follows on page 10, line 21-24,

"Since satellite pixel coordinates are provided in a geographic coordinate system, to acquire the accurate Euclidean distance between satellite pixels and ground measurement locations, the coordinates of all the data were converted to the JGD_2000_UTM_Zone_52N

coordination system."

2. In your second response you claim that reprojection does not include averaging of any kind. Every reprojection includes changes to the original data, I would think. Can you please elaborate?

Reviewer 2 also pointed out a possible lack of clarity regarding the 'remapping' strategy employed. While you have addressed this point in principle, I am still not sure I fully understand the procedure from the manuscript.

Response: We did not change the statistics or distribution of AOD. The reprojection only allowed us to calculate the distance between satellite pixel centroids and AERONET stations more accurately. To clarify this, we modified the sentence as follows on page 10, line 21-24, "Since satellite pixel coordinates are provided in a geographic coordinate system, to acquire the accurate Euclidean distance between satellite pixels and ground measurement locations, the coordinates of all the data were converted to the JGD_2000_UTM_Zone_52N coordination system."

- (was 19-28): I agree that there is little physical meaning in slopes and intercepts in log-transformed data. In the same way, however, if the condition required for the numerical method (i.e. normal distribution) is not fulfilled the physical meaning of your results will be equally limited. It does not matter in the least whether other studies in the past made the same mistake, as you argue. In your response you imply that the actual distribution of the values is closer to a normal distribution than the log-transformed distribution (although this cannot be seen from Figure 4, where values on the horizontal axis are unevenly spaced). Can you support this with a test for normal distribution? Are there other transformations that would get us closer to a normal distribution? In any case, I think this problem needs to be explicitly addressed in the paper. Also, and accordingly, all discussion based on the results of the numerical analysis need to be treated and interpreted with great care on this basis.

Response: We tested for normality of the residuals from linear regressions and conducted comparisons using log-transformed data, after adding 0.05 to MODIS and GOCI retrievals and corresponding AERONET retrievals, using dataset for temporal comparisons over the Japan-South Korea region. Results of these analyses were shown in Supplemental Material, Table S1 and Table S2. We added the following sentences on page 15, line 17-26,

"The residuals of the linear regressions were slightly skewed (Supplemental Material, Table S1), indicating that one assumption of linear regression, normality of the residual distribution, was not fully met. However, log-transformation did not necessarily make the residual distribution more normal (Supplemental Material, Table S1) and log-transformation led to loss of physical meaning of evaluation metrics as well as made the evaluation metrics incomparable to previous studies. All things considered, we used the original data in this analysis. We conducted a sensitivity analysis using log-transformed data after adding 0.05 to GOCI, Aqua and Terra MODIS C6 3 km satellite retrievals as well as corresponding AERONET retrievals over Japan-South Korea region."

- Figure 5: Are the regression lines shown here statistically significant at a certain level? If so, please state this in the figure caption. Otherwise I suggest removing the regression lines.
By the way: The first and third rows clearly show how the AOD distribution is dominated by small values – there does not seem to be a normal distribution, so linear regression will not yield reliable results here.

Response: There regression lines were statistically significant at the alpha level of 0.01. We added the following words in the figure caption, "all the linear relationships are statistically significant at the alpha level of 0.01".

- Figure 6: Reviewer 2 noted that color scale limits vary among figures, impairing comparability. I agree with reviewer 2 and think the sclaes should use identical limits.

Response: We used the same color scale for all figures with different minimum and maximum values, corresponding to different satellite products. The minimum value is 0 for VIIRS products and -0.05 for MODIS and GOCI products, and the maximum value is 2.0 for VIIRS products and >2.0 for MODIS and GOCI products. This difference is related to retrieval algorithms and we wanted to indicate this difference in figures, but the fact that these color scales differed in the maximum and minimum values did not affect comparisons across these figures since except the first and last color, the range of each color across figures was the same.

Technical comments
==================

- 6-26: ...which temporarily deployed additional sun photometers in...

Response: We modified this sentence as follows on page 7, line 26-29, "In response to the lack of spatially concentrated ground AOD observations, AERONET conducted several campaigns, which temporarily deployed additional sunphotometers in selected regions and provided valuable information on small-scale AOD distribution."

- 8-18: You changed the wording from "measure" to "observe" - still, the station itself does not measure or observe anything. The measurement occurs AT the station. Please increase precision in wording here.

Response: We changed the description of ground AOD measurements to "the sunphotometer at each AERONET station measures AOD at eight spectral bands between 340 nm and 1020 nm."

- 29-2: marking both slope AND intercept values with significance flags seems redundant

Response: We removed the significance flags of slope and added the following footnote to the table, "All the slopes are statistically significant at the alpha level of 0.01."

- Figure 5: labels on horizontal axes are inconsistent – this may lead to confusion on whether there is a difference between AOD and AOD550. Suggestion: decide on one and use consistently

Response: We changed the labels on horizontal and vertical axes to "AOD".

Evaluation of VIIRS, GOCI, and MODIS Collection 6 AOD retrievals against ground sunphotometer observations over East Asia

Q. Xiao[1], H. Zhang[2], M. Choi[3], S. Li[1, 4], S. Kondragunta[5], J. Kim[3], B. Holben[6], R. C. Levy[6], Y. Liu[1]

[1] {Emory University, Rollins School of Public Health, Atlanta, GA, USA}

[2] {I.M. Systems Group,Inc., College Park, MD, USA}

[3] {Yonsei University, Seoul, South Korea}

[4] {State Key Laboratory of Remote Sensing Science, Beijing, China}

[5] {National Oceanic and Atmospheric Administration, Greenbelt, MD, USA}

[6] {NASA Goddard Space Flight Center, Greenbelt, MD, USA}

Correspondence to: Y. Liu (yang.liu@emory.edu)

Abstract

Persistent high aerosol loadings together with extremely high population densities have raised serious air quality and public health concerns in many urban centers in East Asia. However, ground-based air quality monitoring is relatively limited in this area. Recently, satellite-retrieved Aerosol Optical Depth (AOD) at high resolution has become a powerful tool to characterize aerosol patterns in space and time. Using ground AOD observations from the Aerosol Robotic Network (AERONET) and the Distributed Regional Aerosol Gridded Observation Networks (DRAGON)-Asia Campaign, as well as from handheld sunphotometers, we evaluated emerging aerosol products from the Visible Infrared Imaging Radiometer Suite (VIIRS) aboard the Suomi National Polar-orbiting Partnership (S-NPP), the Geostationary Ocean Color Imager (GOCI) aboard the Communication, Ocean, and Meteorology Satellite (COMS), and Terra and Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) (Collection 6) in East Asia in 2012 and 2013. In the case study in Beijing, when compared with AOD observations from handheld sunphotometers, 51% of VIIRS Environmental Data Record (EDR) AOD, 37% of GOCI AOD, 33% of VIIRS Intermediate Product (IP) AOD, 26% of Terra MODIS C6 3 km AOD, and 16% of Aqua MODIS C6 3 km AOD fell within the reference expected error (EE) envelop (±0.05±0.15AOD). Comparing against AERONET AOD over the the Japan-South Korea region, 64% of EDR, 37% of IP, 61% of GOCI, 39% of Terra MODIS and 56% of Aqua MODIS C6 3 km AOD fell within the EE. In general, satellite aerosol products performed better in tracking the day-to-day variability than tracking the spatial variability at high resolutions. The VIIRS EDR and GOCI products provided the most accurate AOD retrievals, while VIIRS IP and MODIS C6 3 km products had positive biases.

1. Introduction

Aerosols play a critical role in atmospheric processes as well as global climate change. Rapid economic growth and increasing fossil fuel usage have significantly affected aerosol formation and transportation led to increasing air pollutant emission in East Asia. From 1980–2003, the emissions of black carbon, organic carbon, SO_2, and NOx increased by 28%, 30%, 119%, and 176%, respectively (Ohara et al., 2007). The continuous air quality degradation together with high population density have raised serious public health concerns in this region. Among commonly monitored air pollutants, particulate matter (PM), especially fine particulate matter ($PM_{2.5}$, airborne particles with an aerodynamic diameter less than or equal to 2.5 μm), Aerosolsis are also noted for its adverse health impacts, such as increased cardiovascular and respiratory morbidity and mortality (Lim et al., 2013). The continuous air quality degradation together with high population density have raised serious public health concerns in East Asia.

Satellite remote sensing data have been applied to characterize aerosol global distribution and temporal variation. Although the primary goal of satellite observations is to advance our understanding of the climate system, the comprehensive spatial coverage and growing time series of satellite retrievals benefit various applications, including monitoring ground level air pollution, especially particulate matter (PM). The traditional ground-based air quality monitoring networks are expensive to operate and have limited spatial coverage. For example, most PM monitoring stations in China are located in urban centers and the monitoring network only covers about 360 out of the the more than 3,000 counties. Most developing countries, where PM levels are dangerously high, have little or no regular ground monitoring network. These limitations of ground measurements result in insufficient information to conduct studies about pollution sources, distribution, and consequent health impacts. Satellites provide continuous, high-coverage observations of aerosol loadings and various approaches have been developed to estimate ground-level PM concentrations from satellite retrievals (Ma et al., 2014;Xu et al., 2015). Estimates of ground-level PM concentrations from satellite observations have been used in epidemiological studies and benefited policy making (Strickland et al., 2015;Evans et al., 2013).

(Holben et al., 1998; Li et al., 2005). The continuous air quality degradation together with high population density have raised serious public health concerns in this region. The severe PM pollution in East Asia has attracted worldwide attention and ground PM monitoring networks have been developed in some East Asian countries like China, Japan and South Korea. For instance, in South Korea, PM_{10} together with other important air pollutants have been measured by a dense ground-based network, called 'Air Korea', by the Ministry of Environment (http://eng.airkorea.or.kr). However, ground-based monitoring networks have two main limitations: uneven distribution and limited coverage. For example, the majority of air quality monitoring stations in China are located in large cities and the monitoring network only covers about 360 out of the approximately 2,860 municipalities. These two limitations of ground PM measurements result in insufficient information to conduct studies about PM sources, distribution, and consequent health impacts in East Asia, which can negatively impact policymaking.

The extensive spatial coverage and growing time series of satellite retrievals allow researchers to better characterize aerosol patterns spatially and temporally. The most widely used satellite aerosol sensor, the Moderate Resolution Imaging Spectroradiometer (MODIS), has 36 spectral bands, acquiring data in wavelength from 0.41 μm to 15 μm and providing information about atmospheric aerosol properties (Anderson et al., 2003). Two identical MODIS instruments are aboard the National Aeronautics and Space Administration (NASA) Terra and Aqua satellites, which fly over the study area at around 10:30 and 13:30 LT, respectively. Several algorithms have been developed to retrieve aerosol optical depth (AOD) from MODIS data over land, such as the Dark-Target (Levy et al., 2013) algorithm and the Deep-Blue (Hsu et al., 2013) algorithm, providing AOD retrievals at 550 nm with global coverage. The widely used 10 km resolution MODIS aerosol products provides valuable information on aerosol distribution in space and time, and has been widely used to characterize aerosol dynamics and distribution, simulate climate change, and assess population PM exposure (Levy et al., 2013;Levy et al., 2010). However, the 10 km product cannot depict small-scale PM~~$_{2.5}$~~ heterogeneity. Though a previous study (Anderson et al., 2003) indicated that the aerosol loading is homogeneous at horizontal scales within 200 km, that study is conducted over the ocean, which provides a homogeneous surface, leading to reduced aerosol spatial variability. The variability of aerosol loading at local scales in urban areas with complex land surface and meteorological conditions are expected to be greater (Li et al., 2005). Accurately characterizing local-scale ~~PM2.5~~aerosol heterogeneity is critical for assessing population PM exposure, detecting small smoke plums ~~air pollution sources~~, and analyzing aerosol-cloud process ~~monitoring air quality~~. To resolve small-scale aerosol features, satellite aerosol products with higher resolutions and ~~acceptable~~ accuracy are urgently needed.

In response to the requirement of aerosol retrievals with higher spatial resolution, several emerging satellite aerosol products have become available recently. The Visible Infrared Imaging Radiometer Suite (VIIRS), is a multi-disciplinary scanning radiometer with 22 spectral bands covering from 0.412–12.05 μm and is designed as a new generation of operational satellite sensors that are able to provide aerosol products with similar quality to MODIS (Jackson et al., 2013). VIIRS is on board the NASA-NOAA Suomi National Polar-orbiting Partnership (S-NPP) that launched in October 2011, and passes over the study area daily at approximately 13:30 LT. The VIIRS aerosol product reached provisional maturity level in January 2013, which means the "product quality may not be optimal" but it is "ready for operational evaluation" (Liu et al., 2014). The characteristics of the instrument and the aerosol retrieval algorithms are documented in detail elsewhere (Liu et al. (2014)) and briefly described here. VIIRS provides two AOD products: the Intermediate Product (IP) and the Environmental Data Record (EDR). The VIIRS aerosol retrieval is performed at pixel-level (~0.75 km) spatial resolution globally as the IP that employs information from Navy Aerosol Analysis and Prediction System (NAAPS) and Global Aerosol Climatology Project (GACP) to fill in missing observations (Vermote et al., 2014). The IP is then aggregated to 6-km spatial resolution as the EDR, a level 2 aerosol product, through quality checking and excluding information from the NAAPS and GACP models. Both VIIRS IP and EDR are assigned quality flags of “high”, “degraded”, or “low” and valid AOD values range between 0.0 and 2.0. Detailed description of the quality assurance of VIIRS aerosol products is documented by Liu et al. (2014). Previous global evaluation against AERONET AOD over all land use types indicates that 71% of EDR retrievals fell within the expected error (EE) envelope established by MODIS level 2 aerosol products over land (±0.05±0.15AOD), with a bias of -0.01 (Liu et al., 2014).

The Geostationary Ocean Color Imager (GOCI) is a geostationary Earth orbit sensor, providing hourly multi-spectral aerosol data eight times per day from 9:00 to 16:00 Korean LT. It covers a 2500 × 2500 km^2 sampling area, centered at [130E, 36N] in East Asia, at 500-m resolution with eight spectral channels at 412, 443, 490, 555, 660, 680, 745, and 865 nm, respectively (Park et al., 2014). GOCI is aboard South Korea’s Communication, Ocean, and Meteorology Satellite (COMS) that launched in June 2010. The retrieval algorithm of its aerosol product, Yonsei aerosol retrieval algorithm, was originally based on the NASA MODIS algorithm and provides level 2 AOD retrievals at 6-km spatial resolutions (Levy et al., 2007;Levy et al., 2010;Lee et al., 2010). The characteristics of the Yonsei retrieval algorithms and the aerosol product are documented in detail by Choi et al. (2015). The GOCI aerosol product allows AOD values ranging between -0.1 and 5.0. A previous study reported that during a two-month period (1 April to 31 May 2011), the GOCI AOD retrievals agreed well with AERONET AOD (R^2 = 0.84) over East Asia (Park et al., 2014). A recently published evaluation study reported that from March to May 2012, the GOCI AOD had a linear relationship with AERONET AOD with a slope of 1.09 and an intercept of -0.04 (Choi et al., 2015).

To meet the need for finer resolution aerosol products, a 3 km aerosol product was introduced as part of the MODIS Collection 6 delivery. The 3 km aerosol product includes a quality flag ranging between 0 and 3 to indicate the quality of each retrieval and the valid AOD values range between -0.1 and 5.0. The retrieval algorithm of the 3 km product is documented in detail by Remer et al. (2013) and a global evaluation based on six months of Aqua data against ground sunphotometer AOD indicates that 63% of the retrievals fell into the EE with a bias of 0.03 over land (Remer et al., 2013). Munchak et al. (2013) reported that in the Baltimore–Washington, D.C. area, an urban/suburban region, 68% of the 3 km retrievals from June 20, 2011 to July 31, 2011 fell into the EE with a bias of 0.013.

The release of these fine-resolution satellite aerosol products has raised the question of whether these AOD retrievals can reflect the spatial pattern of aerosol loadings at their assigned resolutions. AERONET, a globally distributed federation of ground-based atmospheric aerosol observations, provides reliable "ground truth" of AOD that are widely used for the characterization of aerosol and validation of satellite retrievals (Morys et al., 2001;Holben et al., 1998). However, previous evaluation studies comparing these emerging satellite aerosol retrievals with AERONET data were mostly at the global scale. AOD retrievals and their errors are treated as spatially independent because the validation sites are far apart and the AOD retrieval resolutions are relatively low. Therefore, these studies evaluated how accurately a satellite product can track AOD values in time. With the help of spatially dense ground measurements, a regional-scale evaluation can evaluate satellite aerosol products' abilities to accurately reflect the fine-scale aerosol characteristics in space. ~~However, previous evaluation studies with AERONET data focused on the temporal accuracy (i.e., examined if the retrieved AOD can track the day-to-day variability of aerosol loadings). Evaluation of satellite aerosol products' abilities to track small-scale aerosol spatial variability is limited due to a lack of intensive ground observations of AOD: the permanent AERONET stations can be tens or even hundreds of kilometers apart, leading to insufficient information on the small-scale horizontal distribution of aerosol loading that is required for a precise evaluation at high resolution.~~ In response to the lack of ~~intensive~~spatially concentrated ground AOD observations, AERONET conducted several campaigns, which temporarily deployed additional ~~temporary~~ sunphotometers in selected regions and provided valuable information on~~of~~ small-scale AOD distribution. One of these campaigns, the Distributed Regional Aerosol Gridded Observation Network (DRAGON)-Asia Campaign in Japan and South Korea, lasted from February 15, 2012 to May 31, 2012 and provided a rare opportunity to validate these emerging satellite aerosol products in East Asia (Seo et al., 2014;Sano et al., 2012). Another issue with previous evaluation studies is that few of them focused specifically on urban areas with higher pollution levels, greater disease burdens, and more complex aerosol patterns. Our work contributes to the validation effort of these emerging satellite products by employing ground AOD observations at finer resolution, extending the study period to one year, and conducting a mobile sampling experiment in the urban core of Beijing.

In this work, we quantitatively evaluate whether the latest VIIRS, GOCI and MODIS aerosol products can provide reliable AOD retrievals and accurately characterize the spatial pattern of AOD over the urban areas in East Asia. Ground AOD from AERONET, DRAGON-Asia, and handheld sunphotometers were collected over a period of one and a half years. The rest of the paper is organized such that Section 2 describes data sources and evaluation methods used in this study, Section 3 presents the performance of various satellite AOD products in representing intra city as well as regional variability of aerosol loadings. Finally, we summarize our findings and described future study directions in section 4.

2. Data and Methods

2.1 Study Area

The extent of the study area is approximately 2500 × 1100 km^2, centered at [128.5E, 35.5N] in East Asia, covering eastern China, South Korea and Japan (Fig. 1). This domain is within the overlapping region of all satellite datasets and ground observations and covers large urban centers, suburban areas, and rural areas. We also conducted a mobile sampling study in Metro Beijing along three major roads (Fig. 1). The study period is from January 2012 to June 2013.

2.2 Remote Sensing Data

The satellite aerosol products used in this study were from VIIRS, GOCI, Aqua MODIS and Terra MODIS sensors (Table 1). VIIRS data before May 2012 are not available because the sensor was in an early checkout phase and lacked a validated cloud mask (Liu et al., 2014). Thus, only EDR and IP pixels from May 2012 to June 2013 with high quality (Quality Flag = “high”) were processed. Similarly, GOCI aerosol retrievals from January 2012 to June 2013 were filtered by its assigned quality and only high quality (Quality Flag = 3) retrievals were included. The Aqua and Terra MODIS C6 3 km data from January 2012 to June 2013 were obtained from the Goddard Space Flight Center (http://ladsweb.nascom.nasa.gov/data). Only retrievals with high quality (Quality Flag = 3) were included in the analysis. The quality control criteria of these five satellite aerosol products are shown in Table 1.

2.3 Ground Observations

The characteristics of ground AOD datasets are shown in Table 2. There were 18 permanent AERONET stations in the study area during the study period, supplemented by 24 temporary stations during the DRAGON-Asia Campaign. The DRAGON stations were distributed nearly uniformly with approximately 10 km apart from each other in two urban centers: Osaka in Japan (7 stations) and Seoul in South Korea (11 stations). Other DRAGON stations, which can be tens to hundreds of kilometers apart, were located across Japan and South Korea. The sunphotometer at each AERONET stations ~~observe~~measures AOD at eight spectral bands between 340 nm and 1020 nm. To compare with satellite retrievals, AOD at 550 nm was calculated using a quadratic log-log fit from AERONET AOD at wavelengths 440 nm and 675 nm. Near-real time level 2.0 AERONET/DRAGON data in the Japan-South Korea region and level 1.5 AERONET data in Beijing were downloaded from the Goddard Space Flight Center (http://aeronet.gsfc.nasa.gov/). The Level 2.0 (quality assured) AOD data have both pre- and post-deployment calibration, leading to an uncertainty of about 0.01–0.02 while the Level 1.5 AOD data are cloud-screened but not quality-assured (Otter et al., 2002). However, our preliminary results indicate that the level 1.5 daily average AOD values agreed well with the level 2.0 data, with a slope of 1.0 and zero intercept. Thus, we used the level 1.5 data in the case study in Beijing because level 2.0 data are not available for some AERONET stations.

To analyze the intra-city aerosol variability, we conducted ground measurements of AOD by a handheld sunphotometer (model 540 Microtops II, Solar Light Company, Inc.) at the Metro Beijing area in 2012 and 2013. Microtops II provide accurate AOD retrievals and is widely used for ground AOD observations (Morys et al., 2001;Tiwari and Singh, 2013;Otter et al., 2002). Previous calibration reported that the root-mean square differences in AOD from Microtops and corresponding AERONET stations were about ± 0.02 at 340 nm (Ichoku et al., 2002). In this study, ground observations were conducted on every cloud-free day at preselected sites that were roughly 6 km apart from each other along the 3rd and the 5th Ring Roads and the Chang'an Avenue of Beijing. This sampling took place between 9:30 and 14:00 LT, and 5–10 repeated measurements were made at each site. To control the quality of the ground data, we used the median value of the repeated observations as ground truth to eliminate the impact of extreme values and only included AOD with the ratio of standard deviation over median AOD less than 2.0. Our comparison of Microtops AOD retrievals with nearby AERONET data yielded a slope of ~0.95 and a correlation coefficient of ~0.8 (Supplemental Material, Text S1).

2.4 Data Integration and Analytical Methods

Since satellite pixel coordinates are provided in a geographic coordinate system, to acquire the accurate Euclidean distance between satellite pixels and ground measurement locations, the coordinates of all the data were converted to the JGD_2000_UTM_Zone_52N coordination system.~~All the data were converted to the JGD_2000_UTM_Zone_52N coordination system.~~ For matchup process, a 6-km grid and a 3 km grid covering the whole study domain were constructed, corresponding to the spatial resolution of each satellite product. Satellite aerosol data from different sensors were mapped and spatially joined to this 6-km grid (for VIIRS EDR and GOCI products) or 3 km grid (for VIIRS IP and MODIS C6 3 km products) to construct coincident satellite-ground AOD pairs.

To assess the intra-city spatial variations of aerosol loadings, we analyzed ground AOD observations over Beijing, Osaka, and Seoul from handheld sunphotometer and DRAGON-Asia stations in 2012. First, the great circle distance between each of two ground observation sites which are less than 20 km apart were calculated. Then we stratified the site-to-site distances by increments of 750 m, the resolution of VIIRS IP aerosol product, and calculated the station-to-station correlation coefficients of daily average AOD within each distance stratum. The observations from DRAGON sites in Osaka and Seoul and from handheld sunphotometers in Beijing were processed separately due to differences in instrumentation. Only handheld sunphotometer AOD observations in Beijing from February 15, 2012 to May 31, 2012 were included to ensure that the study period at these three locations is the same.

To validate the performance of high-resolution satellite aerosol products, two types of comparisons were conducted: the temporal comparison, which compared satellite AOD retrievals within 3 × 3 grid cells sampling buffers against ground AOD from AERONET stations during one year from July 2012 to June 2013; and the spatial comparison, which compared satellite AOD retrievals within single grid cell sampling buffers against ~~intensive~~spatially concentrated ground AOD from DRAGON stations or the handheld sunphotometer. Temporal comparisons and spatial comparisons differ in study periods (Table 2): the temporal comparison period was the longest overlap period covered by all five satellite products and the spatial comparison periods in Beijing and the Japan–South Korea region are different in order to include the maximum number of ground observations. The coefficients of variation (CV), which is standard deviation divided by mean of AOD retrievals, from various sensors in temporal-comparison sampling buffers were calculated and reported below to assess the homogeneity of aerosol loading within buffers. The mean CV from various aerosol products ranged between 0.18 and 0.35, indicating that, as expected, certain heterogeneity in aerosol loading existed within the temporal-comparison buffer. This relatively small heterogeneity should not be a detriment to the temporal comparison, however; some extremely large CV values that were probably due to very small mean AOD values were observed. In order to avoid potentially large variations in aerosol loading within buffers, we removed satellite pixels with CVs outside the range of ± 1.0 (Liu et al., 2007) in temporal comparisons. Moreover, the existing heterogeneity of AOD loading encouraged us to conduct spatial comparisons implementing smaller sampling buffers.

For the temporal comparison of VIIRS EDR data, we averaged valid AOD retrievals in each 3 × 3 grid cells sampling buffer (18 × 18 km^2) centered at each ground AERONET station. The mean and median CV were 0.25 and 0.21, respectively. The average AOD values were then compared with the mean AERONET AOD within a 1-h time window (± 30 min around the satellite overpass time). We employed this smaller spatial averaging window than the widely used 27.5 km-radius-circle buffer suggested by the Multi-sensor Aerosol Products Sampling System (MAPSS) (Seo et al., 2014) in order to examine the performance of these finer resolution products at the scale of their expected application conditions. We used the typical 1-h time window because a previous analysis indicated that changing the time window matters little to validation results (Remer et al., 2013) and the 1-h time window yields a larger database for the validation. For the spatial comparison of VIIRS EDR data, we used single 6-km pixels covering each ground observation location, i.e. DRAGON station or handheld sunphotometer measurement location, and compared the AOD retrieval values with the mean AOD from the corresponding DRAGON station within the 1-h time window or the median AOD from the handheld sunphotometer at the corresponding location. The temporal and spatial comparisons of GOCI data followed the same protocol as described above. Although GOCI provides eight hourly AOD retrievals per day, we only used retrievals at 1:00 pm LT in the comparison in order to make the validation results comparable among these satellite products. The mean and median CV of GOCI retrievals within the 3 × 3 grid cells sampling buffer were 0.35 and 0.15, respectively.

For the comparisons of VIIRS IP data, we used the 3 km grid because we did not have enough ground sampling data to create a 750-m grid. For the temporal comparison, we averaged valid IP AOD retrievals falling in the 3 km grid cell centered at each ground AERONET station and the mean and median CV were 0.33 and 0.25, respectively, within the 3 km grid cell buffer. This sampling buffer roughly covered a 4 × 4 pixel group. The average AOD values were compared against average AOD from the corresponding AERONET station within the 1-h time window. In the spatial comparison of VIIRS IP, we also used the 3 km sampling buffer due to a lack of more ~~intensive~~spatially concentrated ground AOD observations. Thus, the VIIRS IP data is oversampled in the spatial comparison. For the temporal comparison of Aqua and Terra MODIS C6 3 km data, we employed the 3 km grid and averaged valid AOD retrievals in each 3 × 3 grid cells centered at each ground AERONET station to compare with the mean AOD within the 1-h time window. The mean CV of Aqua and Terra MODIS within the 3 × 3 grid cells sampling buffer were 0.18 and 0.13, respectively. For the spatial comparison of MODIS C6 3 km data, we used the individual 3 km pixel AOD value falling on each ground observation location to compare with average AOD from the corresponding DRAGON station within the 1-h time window or the median AOD from the handheld sunphotometer at the corresponding location.

In summary, coincident satellite–ground AOD pairs were defined as average satellite AOD retrievals within the specific sampling buffer matched with average ground AOD observations of the corresponding site within 1-h time windows with respect to satellite pass over time. for VIIRS EDR and GOCI products, the temporal and spatial comparison buffer was 18 × 18 km^2 and 6 × 6 km^2, respectively. For the VIIRS IP product, the temporal and spatial comparison employed the same 3 × 3 km^2 buffer. For MODIS C6 3 km product, the temporal and spatial comparison buffer was 9 × 9 km^2 and 3 × 3 km^2, respectively. The examples of buffers used in the temporal and spatial comparisons for each satellite product are shown in Supplemental Material (Fig. S1). It is notable that both MODIS and VIIRS pixels were stretched toward the edge of the scan. For example, the 3 × 3 km^2 MODIS pixels become approximately 6 × 12 km^2 toward the edge. Thus, the spatial joining and our construction of coincident satellite-ground AOD pairs mayslightly decrease the coverage for MODIS and VIIRS products and may potentially affect the spatial comparison results.

In epidemiological studies, in order to improve the coverage of satellite aerosol data to provide exposure assessment, spatial aggregation is widely used. In our analysis, we constructed quality flags for each satellite–ground AOD collection to obtain better coverage without losing accuracy. For the temporal validation, coincident satellite–ground AOD pairs with at least 20% coverage of both satellite data and ground data (Levy et al., 2013) (e.g., having two or more satellite pixels within the sampling buffer and at least two AERONET/DRAGON AOD within the 1-h time window) were marked as “High Quality”; coincident satellite-ground AOD pairs with less than 20% satellite pixels falling in the sampling buffer but one or more pixels located within the grid cell centered on the ground stations were marked as “Medium Quality”; allother coincident satellite-ground AOD pairs were marked as “Low Quality”. Since we did not create a 750-m grid for the VIIRS IP product, VIIRS IP-ground AOD pairs were assigned either “High Quality” or “Low Quality”. In the spatial validation, because the best scenario satellite-ground AOD collection is to have one or more satellite pixels within the one-grid cell sampling buffer and two or more AERONET/DRAGON AOD during the one hour time window, we only assigned two quality levels: “High Quality” for coincident satellite-ground AOD pairs in the best scenario, and “Low Quality” for all others. Only coincident satellite-ground AOD pairs with high and medium quality were included in our validations. We also conducted a comparison, shown as Table S3~~2~~, including all the satellite–ground AOD pairs—regardless of their quality—to examine the influence of sampling bias. In addition, we conducted sensitivity analyses on VIIRS IP AOD retrievals including both high- and degraded-quality retrievals (Supplemental Material, Table S2~~1~~) and for the GOCI product at hourly scale (Supplemental Material, Table S5~~4~~) with respect to its eight hourly observations per day. In the hourly comparison, we constructed hourly average AERONET AOD as the ground true value and employed the same 3 × 3 grid cells temporal comparison sampling buffer.

2.5 Evaluation Metrics

Several ~~statistical~~ metrics were used to ~~describe~~ evaluate the performance of satellite aerosol products in this study.~~;~~ C~~c~~overage (%) describes the availability of site–day (or site–hour for GOCI data) satellite retrievals when the ground AERONET AOD were available in the temporal comparison. We include all available matched satellite retrievals when calculating the coverage regardless of the quality flag of the coincident satellite-ground AOD pairs.~~;~~ Pearson correlation coefficient describes the correlation between satellite retrievals and ground AOD.~~;~~ bias describes the average difference between satellite retrievals and ground AOD.~~;~~ We calculated the percent of retrievals falling within the expected error (EE) range. For the consistency of the last metric among different aerosol products, we employed the same EE, ±(0.05+0.15AOD), which~~that~~ is established during the global validation of MODIS C5 aerosol product over land~~by MODIS C5 aerosol products over land~~, in this study. In addition, linear regression with satellite retrievals as the dependent variable and ground AOD as the independent variable was employed. The slopes and intercepts from linear regressions were reported. ~~slope is the slope of the linear regression with satellite retrievals as the dependent variable and ground AOD as the independent variable; and we calculated the percent of retrievals falling within the expected error (EE) range. For the consistency of the last metric among different aerosol products, we employed the same EE, ±(0.05+0.15AOD), that is established by MODIS C5 aerosol products over land in this study.~~ The residuals of the linear regressions were slightly skewed (Supplemental Material, Table S1), indicating that one assumption of linear regression, normality of the residual distribution, was not fully met. However, log-transformation did not necessarily make the residual distribution more normal (Supplemental Material, Table S1) and log-transformation led to loss of physical meaning of evaluation metrics as well as made the evaluation metrics incomparable to previous studies. All things considered, we used the original data in this analysis.– We conducted a sensitivity analysis using log-transformed data after adding 0.05 to GOCI, Aqua and Terra MODIS C6 3 km satellite retrievals as well as corresponding AERONET retrievals over Japan-South Korea region.

3 Results and Discussion

3.1 Spatial Variations of Aerosol Loadings

Figure 2 (a) shows the correlation coefficient of daily AOD by binned distance and Fig. 2 (b) shows the site-specific average AOD with the regional average AOD subtracted in these three cities. Figure 2 (a) indicates that the DRAGON AOD were highly correlated within a 20-km spatial range with a correlation coefficient larger than 0.9. However, results from handheld sunphotometer observations in Beijing suggest that the spatial correlation coefficients declined slowly as the distance between two measurement locations increased up to 12 km. The correlation coefficient increased slightly when the distance among two measurement locations are beyond 12 km. This can be explained by the clustered distribution of ground measurement locations in Beijing: these long location-to-location distances only occur when the two locations are located along the Chang'an Avenue and, since vehicle exhaust is one of the major sources of aerosol in Beijing, these AOD are highly correlated. The different aerosol spatial variability trends in Beijing and in the DRAGON domain can be attributed to the following reason: first, the DRAGON-Asia campaign provides real-time observation but our ground AOD observations in Beijing provide one observation at each site per day, so that the average daily AOD from DRAGON stations may have smoothed away some of the spatial heterogeneity. Second, the handheld sunphotometer may introduce larger measurement errors than DRAGON stations, due to both instrument quality and operation errors. Previous evaluation indicates that handheld stability and inaccurate pointing to the Sun significantly affects the accuracy of measurements by Mocrotops II (Ichoku et al., 2002;Morys et al., 2001). Our comparison of Microtops II AOD with nearby AERONET data yielded a slope of ~0.95, a correlation coefficient of ~0.8, and an intercept of 0.16 (Supplemental Material, Text S1), indicating that the handheld sunphotometer AOD are usable.

Even though the aerosol loadings are highly related spatially, the AOD value may differ among nearby stations (Fig. 2 (b)). In Beijing, the difference in average AOD between two neighboring sites that are ~6 km apart can be as high as 0.4, about 49% of the regional mean AOD value. The observations from DRAGON stations show smaller differences in average AOD relative to those in Beijing, but the difference between two neighboring sites can still be greater than 0.1 in Seoul—23% of the regional mean AOD value. These results indicate that spatial contrast in aerosol loading exists at local scale and finer resolution satellite aerosol products are needed to better characterize individual and population exposure of particulate pollution.

3.2 The Beijing Sampling Experiment

The GOCI aerosol product provided the highest coverage in the temporal comparison over Beijing with 73% available retrievals relative to AERONET AOD within the 1-h time window (± 30 min around the satellite overpass time), followed by the VIIRS IP (42%), VIIRS EDR (41%), MODIS Terra C6 3 km product (40%), and MODIS Aqua C6 3 km product (38%) (Supplemental Material, Table S2~~1~~). Table 3 shows the statistical metrics from the temporal and spatial comparisons over Beijing. In the temporal comparison, the GOCI product provided the most accurate AOD retrievals, which slightly overestimated AOD by 0.02 on average. Other aerosol products significantly overestimated AOD with theaverage bias in the temporal comparison for VIIRS EDR, VIIRS IP, Aqua and Terra MODIS C6 3 km products equal to 0.11, 0.25, 0.21, and 0.29, respectively. Though GOCI AOD retrievals agreed well with ground AOD in the temporal comparison, with 55% of GOCI AOD retrievals at 13:00 falling within the EE, only 37% of GOCI AOD retrievals fell within the EE in the spatial comparison. The comparison including all eight hourly GOCI observations represented reduced coverage (59%), a smaller average bias (-0.006), and a larger proportion of retrievals fell within EE (59%). Thus, the GOCI product resolved the temporal and spatial variability of aerosol loadings at its designed temporal and spatial resolutions, but it tracked the small-scale spatial variability less well than the temporal variability in Beijing.

VIIRS EDR product performed well in Beijing in both the temporal and spatial comparisons, with 52% and 51% of retrievals falling within the EE in the temporal and spatial comparison, respectively. Although VIIRS IP had a relatively large positive bias

(0.25) in the temporal comparison, it provided acceptable coverage with 33% retrievals falling within the EE in the spatial comparison, resolving valuable information of small-scale aerosol variability in urban areas. The MODIS C6 3 km product had the largest high bias and lowest %EE in this spatial comparison, with 16% and 26% of retrievals falling within the EE for Aqua and Terra MODIS, respectively. A previous validation study of the 3 km MODIS AOD data also reported similar retrieval errors in urban areas (Remer et al., 2013). It is notable that the R^2 values of the MODIS C6 3 km products is the highest in the spatial comparisons (0.68 for Aqua and 0.85 for Terra) and the linear regression statistics indicates that the low percent of retrievals falling within EE is mainly due to a relatively constant positive offset: the intercepts for Aqua and Terra are 0.22 and 0.30, respectively. One possible explanation of the positive bias of MODIS and VIIRS products is that our study domain is highly urbanized with bright surfaces, therefore is challenging for the Dark Target algorithm.

3.3 The Temporal Evaluation of AOD over the Japan-South Korea region

We first looked at the AOD retrievals distribution on one clear day, 7 May 2012, during the DRAGON period (Fig. 3). Figure 3 indicates that the sampling strategies and cloud masks differ in these five satellite aerosol products, resulting in different patterns of missing data. GOCI provided the best coverage with almost no missing data over this region. VIIRS products and MODIS products showed similar missing data in the center of the map but were less consistent at its edges; while VIIRS products showed more missing data in the lower right corner, MODIS products showed more missing in the upper right corner. VIIRS and MODIS pixels are stretched toward the edge of the scan. VIIRS and MODIS products tended to overestimate AOD values in the urban area (Seoul), but GOCI provided accurate AOD estimates in this region. Though these 3 km products showed similar spatial distribution patterns to the 6-km products, the 3 km products demonstrated greater heterogeneity, which is valuable to analyze local aerosol sources and estimate personal air pollution exposure.

Similar to the comparisons in Beijing, the GOCI aerosol products provided the highest coverage in the temporal comparison over the Japan–South Korea region, with 74%

retrievals relative to AERONET observations within the 1-h time window (±30 min around the satellite overpass time), followed by VIIRS EDR (63%), VIIRS IP (50%), Terra MODIS C6 3 km (26%), and Aqua MODIS C6 3 km (24%) (Supplemental Material, Table S2~~1~~). It is notable that the seasonal missing pattern due to cloud cover and weather conditions may vary across these satellite aerosol products. However, since we did not have enough coincident satellite-ground AOD pairs to conduct seasonal evaluation, the seasonal missing patterns and seasonal performance of these satellite aerosol products were not analyzed in this study. The distributions of the coincident satellite-AERONET AOD pairs with high or medium quality are shown in Fig. 4. The distribution of the Terra MODIS C6 product is not shown here because it passes the study region in the morning, leading to potential differences in AOD distribution relatives to other sensors that pass the study region in the afternoon. This histogram is plotted with frequency of AOD retrievals from each sensor relative to the total number of matched AOD retrievals from the corresponding sensor rather than the count of AOD retrievals because these aerosol products differ in sampling strategies, leading to different total number of coincident satellite-ground AOD pairs. VIIRS EDR, VIIRS IP, and GOCI products showed a similar mode of distribution to AERONET AOD, with the peak probability around 0.2. The distribution of Aqua MODIS C6 3 km AOD had the peak around 0.3, indicating that the Aqua MODIS C6 3 km product tended to overestimate AOD in general. A previous study also reported that the MODIS C6 3 km product had a decreased proportion of low AOD values and an increased proportion of high AOD values (Remer et al., 2013) relative to the 10 km product over land, leading to a higher global average AOD. The VIIRS IP product also tended to overestimate AOD, with higher percentage of retrievals occurring at high AOD values. The distribution of GOCI data provided the best fit with AERONET data, with a correlation coefficient of 0.95, followed by VIIRS EDR (R^2 = 0.93), VIIRS IP (R^2 = 0.77), and MODIS Aqua C6 3 km product (R^2 = 0.76). The difference in the distributions of these satellite aerosol products can be partly explained by different retrieval assumptions including aerosol models, different surface reflectance and different global sampling strategies. Moreover, these satellite aerosol products differ in the valid AOD retrieval ranges, leading to differences in the distribution of extremely high and low AOD values.

The temporal comparisons over the Japan–South Korea region showed more retrievals falling within the EE and smaller biases relative to comparisons in Beijing. Figure 5 shows the frequency scatter plots showing the results of temporal comparisons over the Japan–South Korea region and the corresponding box plots showing the difference between satellite AOD retrievals and ground observations. GOCI retrievals at 13:00 LT were highly correlated with the ground AOD with an R^2 of 0.80. The linear regression of GOCI retrievals and ground AOD fell close to the 1:1 line with a small offset (0.04), and 61% of GOCI retrievals at 13:00 LT fell in the EE. Comparison including eight GOCI hourly retrievals showed a higher R^2 of 0.82 with a smaller average bias (0.02), with 66% of retrievals falling within the EE (Table 4, GOCI all obs.). The box plot indicates that GOCI retrievals overestimated AOD at high AOD values (AOD > 0.6) (Fig. 5). Thus, the GOCI product tracked the daily variability of aerosol loadings well and it provided additional information to study short-term aerosol trends. Similarly, 64% of VIIRS EDR retrievals fell into the EE with a slightly higher bias (0.05) and a slightly lower R^2 of 0.73 (Table 4). This positive bias is consistent with a previous global validation study, which reports a 0.01 bias of VIIRS EDR in East Asia (Liu et al., 2014). Though the VIIRS EDR product tended to overestimate AOD at low (AOD < 0.3) and high AOD values (AOD > 1.0), it agreed well with the AERONET observations when AOD ranged between 0.3 and 1.0 (Fig. 6).

The VIIRS IP had a linear regression slope close to 1 (1.03) against AERONET observations, but it had a consistent positive bias of 0.15 on average. Only 37% of VIIRS IP retrievals fell within the EE. The scatter plot indicates that the IP retrievals varied substantially, especially when the AOD values were low. MODIS C6 3 km products had a high positive bias of 0.08 for Aqua and 0.16 for Terra. Consistent with what was reported by a previous global evaluation study, we observed that the MODIS C6 3 km products tended to overestimate AOD and the bias increased with AOD values (Remer et al., 2013). 56% of the Aqua MODIS C6 3 km retrievals and 39% of the Terra MODIS C6 3 km retrievals fell within the EE. In general, these finer resolution aerosol products included larger bias relative to lower resolution products and researchers must be cautious when applying them by, for example, calibrating these high resolution satellite aerosol products in specified study regions and implementing appropriate data filtering strategies.

Since the GOCI product provides eight hourly observations per day, to examine the temporal variability in the accuracy of GOCI aerosol retrievals, we compared the GOCI AOD retrievals with AERONET AOD stratified by hour (Supplemental Material, Table S54). In general, the GOCI product provided high quality retrievals consistently throughout the day except that it tended to slightly overestimate AOD in the morning and underestimate AOD in the afternoon. Such temporal variability in accuracy was also reported by a previous evaluation study of the Geostationary Operational Environmental Satellite (GOES) aerosol product (Morys et al., 2001). The daily variability in the quality of GOCI retrievals may be due to changes in scattering angle, clouds and the associated Bidirectional Reflectance Distribution Function (BRDF) effects.

Ten-fold cross validation was conducted for the comparison of VIIRS and GOCI products to detect overfitting. The linear regression statistics of cross validation did not change significantly relative to the statistics of comparisons. The cross validation R^2 values of VIIRS EDR, VIIRS IP, GOCI at 13:00, and GOCI 8 observations data were 0.73, 0.51, 0.78, and 0.82, respectively. In addition, to detect the spatial variability of the satellite retrieval performance, we applied the regionally developed linear regression parameters of GOCI 8 observations data to individual AERONET station in the Japan–South Korea region. The linear regressions with the satellite AOD as the dependent variable and the fitted AOD from a regional model as the independent variable yielded R2 larger than 0.75 at all sites except the AERONET sites ‘Nara’ and ‘Osaka’, two stations located in Osaka. This result indicated that parameters from the regional dataset were valid locally. Limited by sample size, we did not apply this method to other aerosol products.

3.4 The Spatial Evaluation of AOD over the Japan-South Korea region

The mean daily AOD from different sensors and AERONET stations during the one-year period from July 2012 to June 2013 are shown in Fig. 6. These five aerosol products provided similar distributions of average AOD during the one-year period, with the highest values occurring in northeastern China and the Yangtze River delta, and the lowest values occurring in southern China and Japan. Several high-AOD-value spots appeared along the west coast of South Korea and surrounded the Seto Inland Sea, likely due to emissions from urban centers in these regions. These five maps differ in missing patterns due to their different masking approaches. The VIIRS algorithms did not retrieve AOD over inland lakes (e.g. the Taihu Lake); the GOCI product retrieved AOD over inland water; while MODIS products provided some AOD retrievals over inland lakes, with some missing data. The GOCI product did not provide high-quality retrievals at some locations in central Japan due to snow coverage in this mountain region. To maintain a consistentevaluative data filtering strategy, the inland water AOD retrievals and ground observations were removed from the validation. The VIIRS EDR product showed lower AOD values in northeastern China and South Korea relative to AOD retrievals from other sensors. The VIIRS IP product also showed lower AOD values in northeastern China, but provided higher AOD retrievals in northern Japan. This can be explained by the system bias reported in a previous study that VIIRS retrievals tend to underestimate AOD when NDVI value is low and overestimate AOD over vegetated surfaces (Liu et al., 2014). The VIIRS IP product had higher AOD values relative to the EDR product, especially over the Korean Peninsula and northern Japan. This may be due to IP's ability to track small-scale variability which were smoothed in the EDR retrievals, or may result from the positive bias of IP observed in the temporal comparison. Because VIIRS aerosol products restrict valid AOD values to between 0.0 and 2.0, they may underestimate AOD values when the aerosol loadings are extremely high, like in northeastern China, though we lacked ground AOD data in this region to test this hypothesis. Aqua and Terra MODIS C6 3 km aerosol products showed similar spatial distribution in AOD retrievals, with higher AOD values in urban areas (e.g., over the Yangtze River Delta and North China Plain in China). GOCI presented some high AOD values in local regions such as western South Korea, around the Seto Inland Sea, and over northeastern China. However, it showed lower AOD values over the Yangtze River Delta in China. This result is consistent with the temporal comparison results shown in Fig. 5 that the GOCI product slightly overestimated AOD at high AOD values (AOD>0.6). Compared with ground AOD, all these five aerosol products overestimated AOD in Japan, where the average AOD values were relatively low. VIIRS EDR tended to slightly underestimate AOD over the Seoul region. The lack of ground AOD, especially in northeast China, makes it impossible to quantitively evaluate the spatial distribution of these aerosol products in China.

Results of the spatial comparison over DRAGON-Asia region are shown in Table 4. Satellite aerosol products performed better in tracking the day-to-day variability relative to tracking their spatial patterns. In the spatial comparison, all the satellite aerosol products showed lower R^2 and larger offset with less retrievals falling into the EE. GOCI product provided the highest accuracy, with a small positive bias of 0.03 and 48% of retrievals falling in the EE, followed by VIIRS EDR, with a positive offset of 0.16 and 41% of retrievals falling in the EE. In contrast, VIIRS IP and MODIS C6 3 km had large positive biases, and less than 30% of retrievals fell within the EE due to larger noise (related to the finer resolutions). There is evidence that this positive bias includes systematic errors due to improper characterization of surface reflectance, uncertainties in the assumed aerosol model, and cloud masking. The 3 km MODIS products sample fewer reflectance pixels to retrieve aerosol pixels relative to the 10 km products, introducing sporadic unrealistic high AOD retrievals that are avoided more successfully by the 10 km products (Munchak et al., 2013). Previous studies also reported that improper characterization of bright urban surfaces, a known difficult situation for the Dark Target algorithm, led to positive bias in urban/suburban regions (Munchak et al., 2013;Remer et al., 2013). The VIIRS IP product is retrieved at the reflectance pixel level without aggregation, thus it is expected to include more noise. Though these finer resolution aerosol products did not fully track the spatial trends of aerosol loading at their designed resolution, they provide additional information about aerosol spatial distribution and will benefit exposure assessments at local scales.

To examine possible sampling bias due to our data inclusion criteria, we performed temporal and spatial comparisons including all the coincident satellite-ground AOD pairs over the Japan–South Korea region (Supplemental Material, Table S32). There is no significant change in the evaluation metrics after including pairs with low quality. Thus, the validation results are robust and there is no evidence for sampling bias. We validated the VIIRS IP AOD retrievals with degraded quality over the Japan–South Korea region and observed lower correlation coefficients, higher biases, and less retrievals falling within the EE in both the temporal and spatial comparisons (Supplemental Material, Table S43). This result suggests to use only high-quality VIIRS IP retrievals. We also validated the GOCI AOD retrievals with different quality over the Japan–South Korea region. Including medium- and low-quality GOCI retrievals decreased the accuracy, but significantly increased the coverage (Supplemental Material, Table S65). By including the retrievals having quality flags equal to both 3 and 2, the coverage increased from 27% to 38% in the temporal comparison over the Japan–South Korea region, while the average bias increased by 0.01 and the percentage of retrievals falling within the EE decreased by 7%. Thus, including retrievals with medium quality might be acceptable, depending on study objectives. Results from linear regressions with log-transformed data (Supplemental Material, Table S7) indicated that GOCI aerosol products provided the best estimate of ground measured AOD, followed by VIIRS EDR and MODIS Aqua C6 3 km products. Due to the relatively small number of matched observations, analysis of the correlation between quality of satellite aerosol retrievals and satellite viewing angles were beyond the scope of this analysis. However, previous studies reported that towards the edge of the scan, VIIRS EDR tends to underestimate AOD over land (Liu et al., 2014).

4 Conclusion

In this work, the intra-city variability of aerosol loadings were examined with ground

AOD from the DRAGON-Asia campaign and our mobile sampling campaign in Beijing. Five emerging high-resolution satellite aerosol products are evaluated by comparing them with ground AOD from AERONET, DRAGON, and handheld sunphotometers over East Asia in 2012 and 2013. We observed variability in both correlation coefficients and average AOD values among ground AOD observation sites in three urban centers in Asia. Evaluation results indicated a) that the 6-km resolution products—VIIRS EDR and GOCI—provided more accurate retrievals with higher coverage relative to the higher resolution products—VIIRS IP, Terra and Aqua MODIS C6 3 km products—in both temporal comparisons and spatial comparisons; however, VIIRS IP and MODIS C6 3 km products provide additional information about fine-resolution aerosol spatial distribution and will benefit exposure assessments at local scales; b) satellite aerosol products resolved the day-to-day aerosol loading variability better than the spatial aerosol loading variability; and c) satellite products performed less well in Beijing relative to the Japan-South Korea region, indicating that retrieval in urban areas is challenging. These satellite aerosol products have their own advantages and disadvantages. For example, the GOCI aerosol product provides high accuracy AOD retrievals eight times per day, but it only covers East Asia; the VIIRS EDR product provides high accuracy AOD retrievals and global coverage once per day, but its 6 km resolution is relatively low; the MODIS C6 3 km products provide high resolution AOD retrievals with global coverage, but have positive bias in urban regions. Researchers need to apply these aerosol products according to specified research objectives and study design. The performance of these aerosol products over Beijing and the Japan-South Korea region demonstrates that satellite aerosol products can track the small-scale variability of aerosol loadings. High-resolution satellite aerosol products provide valuable information for the spatial and temporal characterization of $PM_{2.5}$ at local scales. Future studies with additional ground AOD observations at fine spatial and temporal scale will help us analyze air pollution patterns and further validate satellite products.

Acknowledgment

The work of Liu and Xiao was partially supported by the NASA Applied Sciences Program (grants NNX11AI53G and NNX14AG01G, PI: Liu). We would like to acknowledge the AERONET team, Prof. I. Sano and the DRAGON-Japan team, the Yonsi team and their collaborators in S. Korea, and CARSNET and CAS teams in and around Beijing for providing data support in this study. The AERONET project is supported NASA EOS project office, and by Hal B. Maring, Radiation Sciences Program, NASA Headquarters.

References

Anderson, T. L., Charlson, R. J., Winker, D. M., Ogren, J. A., and Holmén, K.: Mesoscale Variations of Tropospheric Aerosols*, Journal of the Atmospheric Sciences, 60, 119-136, 2003.

Choi, M., Kim, J., Lee, J., Kim, M., Je Park, Y., Jeong, U., Kim, W., Holben, B., Eck, T. F., Lim, J. H., and Song, C. K.: GOCI Yonsei Aerosol Retrieval (YAER) algorithm and validation during DRAGON-NE Asia 2012 campaign, Atmos. Meas. Tech. Discuss., 8, 9565-9609, 10.5194/amtd-8-9565-2015, 2015.

Evans, J., van Donkelaar, A., Martin, R. V., Burnett, R., Rainham, D. G., Birkett, N. J., and Krewski, D.: Estimates of global mortality attributable to particulate air pollution using satellite imagery, Environmental research, 120, 33-42, 2013.

Holben, B., Eck, T., Slutsker, I., Tanre, D., Buis, J., Setzer, A., Vermote, E., Reagan, J., Kaufman, Y., and Nakajima, T.: AERONET—A federated instrument network and data archive for aerosol characterization, Remote sensing of environment, 66, 1-16, 1998.

Hsu, N. C., Jeong, M. J., Bettenhausen, C., Sayer, A. M., Hansell, R., Seftor, C. S., Huang, J., and Tsay, S. C.: Enhanced Deep Blue aerosol retrieval algorithm: The second generation, Journal of Geophysical Research: Atmospheres, 118, 9296-9315, 10.1002/jgrd.50712, 2013.

Ichoku, C., Levy, R., Kaufman, Y. J., Remer, L. A., Li, R. R., Martins, V. J., Holben, B. N., Abuhassan, N., Slutsker, I., and Eck, T. F.: Analysis of the performance characteristics of the five‐channel Microtops II Sun photometer for measuring aerosol optical thickness and precipitable water vapor, Journal of Geophysical Research: Atmospheres (1984–2012), 107, AAC 5-1-AAC 5-17, 2002.

Jackson, J. M., Liu, H., Laszlo, I., Kondragunta, S., Remer, L. A., Huang, J., and Huang, H. C.: Suomi‐NPP VIIRS aerosol algorithms and data products, Journal of Geophysical Research: Atmospheres, 118, 12,673-612,689, 2013.

Lee, J., Kim, J., Song, C. H., Ryu, J.-H., Ahn, Y.-H., and Song, C.: Algorithm for retrieval of aerosol optical properties over the ocean from the Geostationary Ocean Color Imager, Remote sensing of environment, 114, 1077-1088, 2010.

Levy, R., Mattoo, S., Munchak, L., Remer, L., Sayer, A., Patadia, F., and Hsu, N.: The Collection 6 MODIS aerosol products over land and ocean, Atmospheric Measurement Techniques, 6, 2989-3034, 2013.

Levy, R. C., Remer, L. A., and Dubovik, O.: Global aerosol optical properties and application to Moderate Resolution Imaging Spectroradiometer aerosol retrieval over land, Journal of Geophysical Research: Atmospheres (1984–2012), 112, 2007.

Levy, R. C., Remer, L. A., Kleidman, R. G., Mattoo, S., Ichoku, C., Kahn, R., and Eck, T.: Global evaluation of the Collection 5 MODIS dark-target aerosol products over land, Atmospheric Chemistry and Physics, 10, 10399-10420, 2010.

Lim, S. S., Vos, T., Flaxman, A. D., Danaei, G., Shibuya, K., Adair-Rohani, H., AlMazroa, M. A., Amann, M., Anderson, H. R., and Andrews, K. G.: A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010, The lancet, 380, 2224-2260, 2013.

Liu, H., Remer, L. A., Huang, J., Huang, H. C., Kondragunta, S., Laszlo, I., Oo, M., and Jackson, J. M.: Preliminary evaluation of S－NPP VIIRS aerosol optical thickness, Journal of Geophysical Research: Atmospheres, 119, 3942-3962, 2014.

Liu, Y., Franklin, M., Kahn, R., and Koutrakis, P.: Using aerosol optical thickness to predict ground-level PM 2.5 concentrations in the St. Louis area: a comparison between MISR and MODIS, Remote sensing of environment, 107, 33-44, 2007.

Ma, Z., Hu, X., Huang, L., Bi, J., and Liu, Y.: Estimating ground-level PM2. 5 in China using satellite remote sensing, Environmental science & technology, 48, 7436-7444, 2014.

Morys, M., Mims, F. M., Hagerup, S., Anderson, S. E., Baker, A., Kia, J., and Walkup, T.: Design, calibration, and performance of MICROTOPS II handheld ozone monitor and Sun photometer, Journal of Geophysical Research: Atmospheres (1984–2012), 106, 14573-14582, 2001.

Munchak, L., Levy, R., Mattoo, S., Remer, L., Holben, B., Schafer, J., Hostetler, C., and Ferrare, R.: MODIS 3 km aerosol product: applications over land in an urban/suburban region, Atmospheric Measurement Techniques Discussions, 6, 1683-1716, 2013.

Ohara, T., Akimoto, H., Kurokawa, J.-i., Horii, N., Yamaji, K., Yan, X., and Hayasaka, T.: An Asian emission inventory of anthropogenic emission sources for the period 1980–2020, Atmospheric Chemistry and Physics, 7, 4419-4444, 2007.

Otter, L., Scholes, R., Dowty, P., Privette, J., Caylor, K., Ringrose, S., Mukelabai, M., Frost, P., Hanan, N., and Totolo, O.: The Southern African regional science initiative (SAFARI 2000): wet season campaigns, South African Journal of Science, 98, p. 131-137, 2002.

Park, M., Song, C., Park, R., Lee, J., Kim, J., Lee, S., Woo, J.-H., Carmichael, G., Eck, T. F., and Holben, B. N.: New approach to monitor transboundary particulate pollution over Northeast Asia, Atmospheric Chemistry and Physics, 14, 659-674, 2014.

Remer, L., Mattoo, S., Levy, R., and Munchak, L.: MODIS 3 km aerosol product: algorithm and global perspective, Atmospheric Measurement Techniques Discussions, 6, 69-112, 2013.

Sano, I., Mukai, S., Holben, B., Nakata, M., Yonemitsu, M., Sugimoto, N., Fujito, T., Hiraki, T., Iguchi, N., and Kozai, K.: DRAGON-West Japan campaign in 2012: regional aerosol measurements over Osaka, SPIE Asia-Pacific Remote Sensing, 2012, 85231M-85231M-85236.

Seo, S., Kim, J., Lee, H., Jeong, U., Kim, W., Holben, B., Kim, S., Song, C., and Lim, J.: Spatio-temporal variations in PM 10 concentrations over Seoul estimated using multiple empirical models together with AERONET and MODIS data collected during the DRAGON-Asia campaign, Atmospheric Chemistry and Physics Discussions, 14, 21709-21748, 2014.

Strickland, M., Hao, H., Hu, X., Chang, H., Darrow, L., and Liu, Y.: Pediatric Emergency Visits and Short-Term Changes in PM2. 5 Concentrations in the US State of Georgia, Environmental health perspectives, 2015.

Tiwari, S., and Singh, A.: Variability of aerosol parameters derived from ground and satellite measurements over Varanasi located in the Indo-Gangetic Basin, Aerosol Air Qual Res, 13, 627-638, 2013.

Vermote, E., Justice, C., and Csiszar, I.: Early evaluation of the VIIRS calibration, cloud mask and surface reflectance Earth data records, Remote sensing of environment, 148, 134-145, 2014.

Xu, J.-W., Martin, R., van Donkelaar, A., Kim, J., Choi, M., Zhang, Q., Geng, G., Liu, Y., Ma, Z., and Huang, L.: Estimating ground-level PM 2.5 in eastern China using aerosol optical depth determined from the GOCI satellite instrument, Atmospheric Chemistry and Physics, 15, 13133-13144, 2015.

Table 1. Characteristics and quality control criteria of satellite aerosol products.

Dataset	Including Criteria	Resolution	Coverage
VIIRS EDR	Quality Flag=High	6 km, daily	Global
VIIRS IP	Quality Flag=High	0.75 km, daily	Global
GOCI	Quality Flag=3	6 km, 8 hourly obs. per day	East Asia
Aqua MODIS C6 3 km	Quality Flag=3	3 km, daily	Global
Terra MODIS C6 3 km	Quality Flag=3	3 km, daily	Global

Table 2. Characteristics of ground AOD measurement datasets.

		Temporal Comparison	Spatial Comparison
Beijing	Data Set	AERONET	Microtops II
	Including Criteria	Level 1.5	Median/Std. Dev. <2
	Study Period	Jul. 2012 – Jun. 2013	Jan. 2012 – Jun. 2013
East Asia	Data Set	AERONET	DRAGON
	Including Criteria	Level 2.0	Level 2.0
	Study Period	Jul. 2012 – Jun. 2013	Feb. 15 – May 31, 2012

Table 3. Statistics of the temporal and spatial comparisons between satellite retrievals and ground AOD measurements at 550 nm in Beijing.

	N	R^2	Slope	Intercept	Bias	%EE
Temporal Comparison						
VIIRS EDR	90	0.70	0.96**	0.12**	0.11	52
VIIRS IP	133	0.63	1.00**	0.25**	0.25	32
GOCI	142	0.88	0.95**	0.05	0.02	55
GOCI all obs.	957	0.88	0.98**	0.008	-0.006	59
Aqua MODIS C6 3 km	119	0.81	1.05**	0.19**	0.21	44
Terra MODIS C6 3 km	133	0.80	0.99**	0.30**	0.29	25
Spatial Comparison						
VIIRS EDR	108	0.14	0.25**	0.34**	0.04	51
VIIRS IP	150	0.16	0.34**	0.45**	0.18	33
GOCI	124	0.51	0.74**	0.23**	0.00	37
Aqua MODIS C6 3 km	77	0.68	1.19**	0.22**	0.31	16
Terra MODIS C6 3 km	73	0.85	1.00**	0.30**	0.30	26

** p-value < 0.01

All the slopes are statistically significant with p-value<0.01.

Table 4. Statistics of the temporal and spatial comparisons between satellite retrievals and ground AOD measurements at 550 nm over Japan-South Korea region.

	N	R^2	Slope	Intercept	Bias	%EE
Temporal Comparison						
VIIRS EDR	600	0.74	0.96**	0.06**	0.05	64
VIIRS IP	424	0.55	1.03**	0.14**	0.15	37
GOCI	317	0.80	1.02**	0.04**	0.05	61
GOCI all obs.	2547	0.82	1.02**	0.01*	0.02	66
Aqua MODIS C6 3 km	179	0.71	1.00**	0.08**	0.08	56
Terra MODIS C6 3 km	197	0.70	1.06**	0.14**	0.16	39
Spatial Comparison						
VIIRS EDR	144	0.53	0.96**	0.18**	0.16	41
VIIRS IP	229	0.60	1.11**	0.21**	0.26	26
GOCI	196	0.79	1.19**	-0.09**	0.03	48
Aqua MODIS C6 3 km	108	0.81	1.26**	0.07*	0.19	28
Terra MODIS C6 3 km	132	0.73	1.00**	0.23**	0.23	27

* p-value < 0.05

** p-value < 0.01

All the slopes are statistically significant with p-value<0.01.

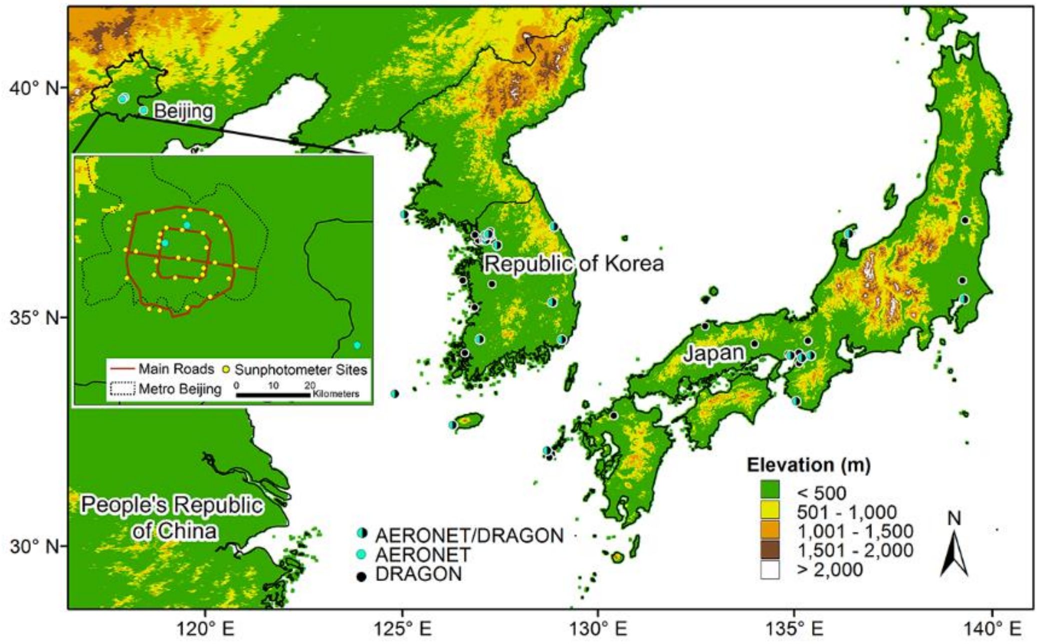

Figure 1. Study area showing all the ground AOD measurement sites.

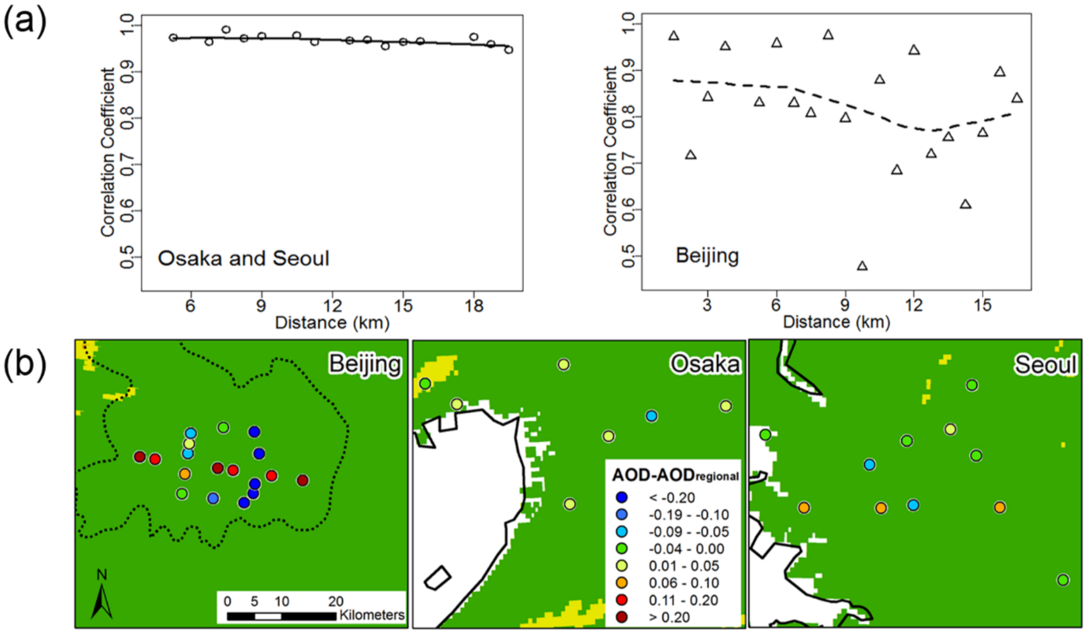

Figure 2. (a) The station to station correlation coefficients of daily mean AOD stratified by distance over (left) DRAGON-Asia region (right) Beijing region. The line is the Loess curvy. (b) The spatial distribution of average AOD in these three cities. The background color shows the elevation with the same color scale as in Figure 1.

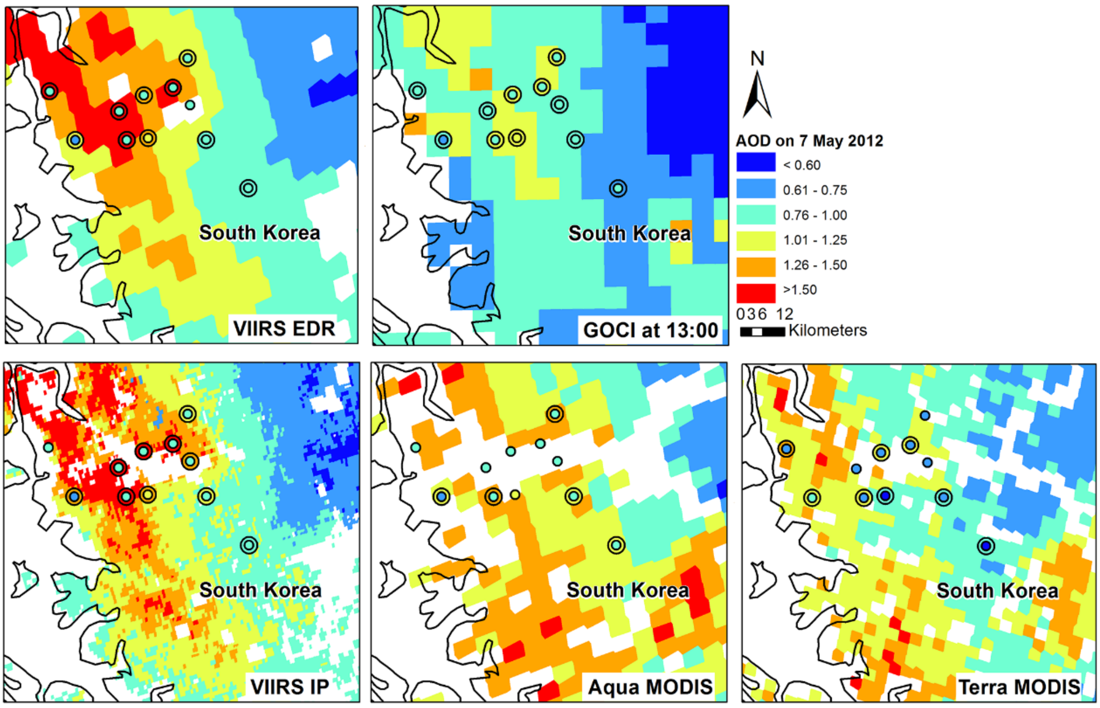

Figure 3. The AOD retrievals at 550 nm from different satellite aerosol products at their designed resolution on 7 May 2012. Coincident Satellite-DRAGON AOD pairs are shown in double circles: the inner circle is the average DRAGON observation within ±30 min of satellite overpass and the outer circle is the satellite retrieval that the DRAGON stations falls in.

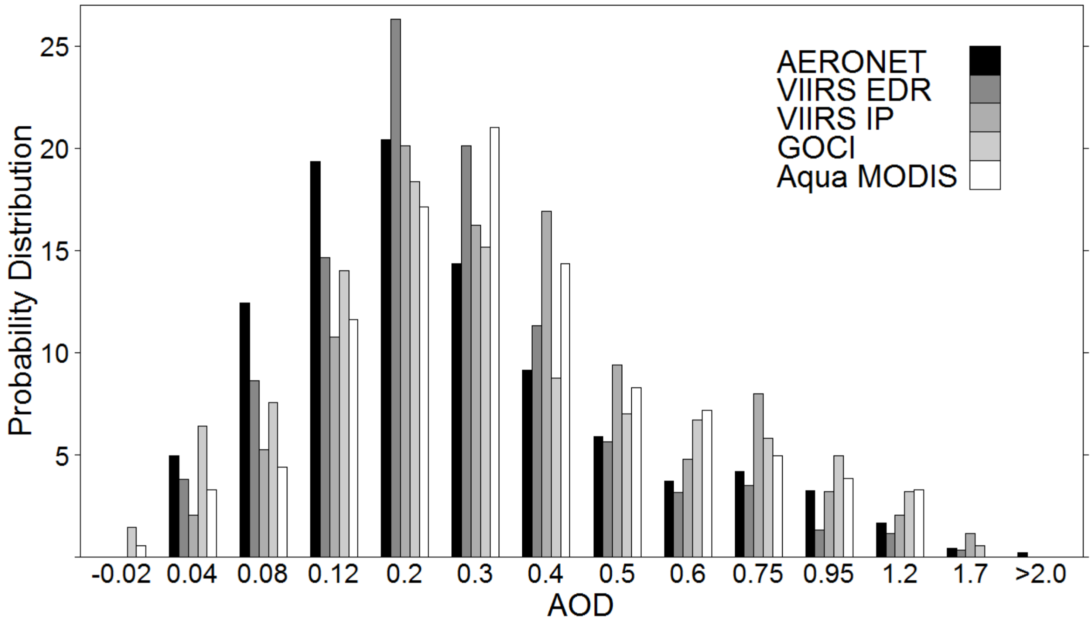

Figure 4. Histogram for the matched satellite AOD retrievals and AERONET measurements. The x-axis shows AOD values and the y-axis shows the frequency of AOD observations from each sensor relative to the total number of matched AOD observations from the corresponding sensor.

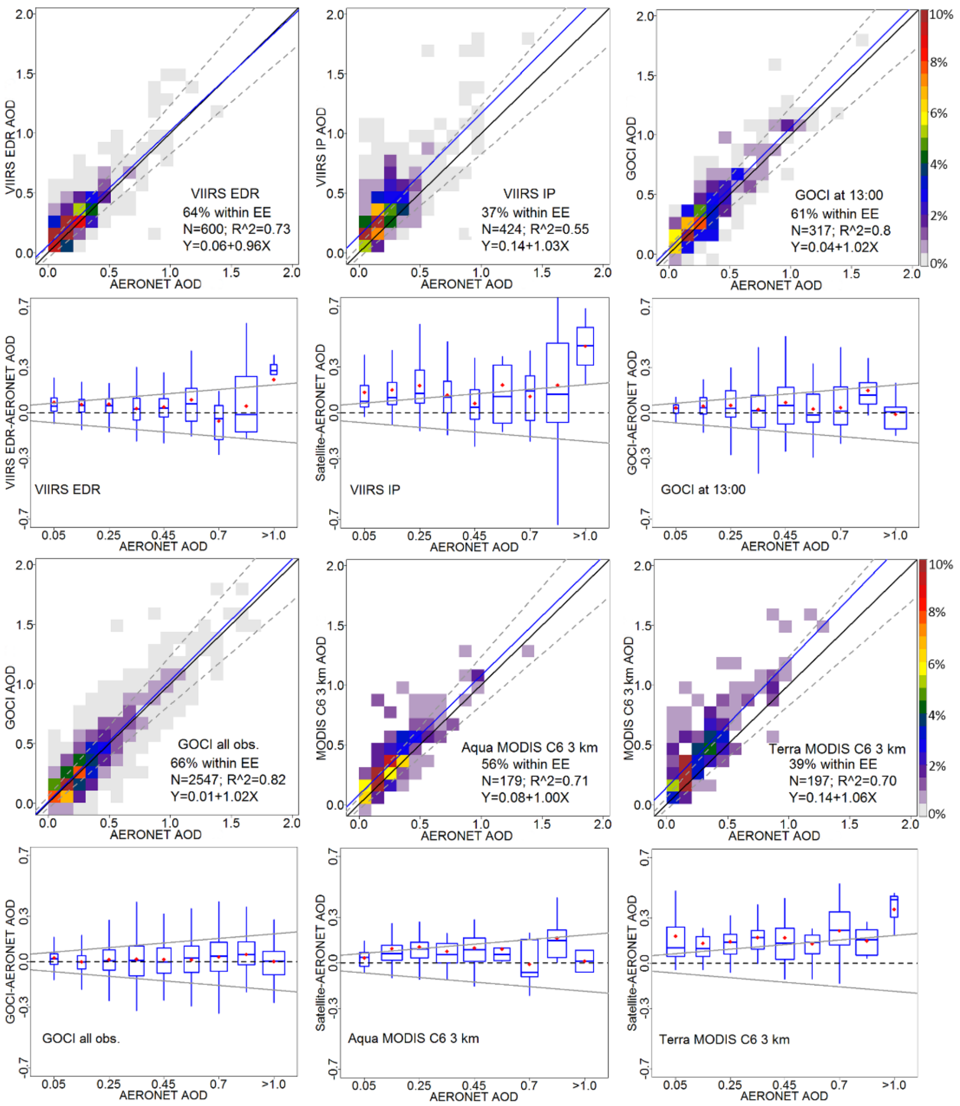

Figure 5. Upper - frequency scatter plots of satellite AOD retrievals against AERONET AOD measurements at 550 nm over the Japan-South Korea region. The linear regression is shown as solid blue line and all the linear relationships are statistically significant at the alpha level of 0.01.~~,~~ T~~t~~he boundary lines of the expected error are shown in the dash lines, and the one-one line is shown as solid black lines for reference. Lower - box plots of AOD errors (satellite – AERONET) versus AERONET AOD over the Japan-South Korea region. The one-one line (zero error) is shown as a dash line and the boundary lines of the expected error are shown as gray solid lines. For each box-whisker, its properties and representing statistics include: width is σ of the satellite AOD; height is the interquartile range of AOD error; whisker is the 2 σ of the AOD error; middle line is the median of the AOD error; and red dot is the mean of the AOD error.

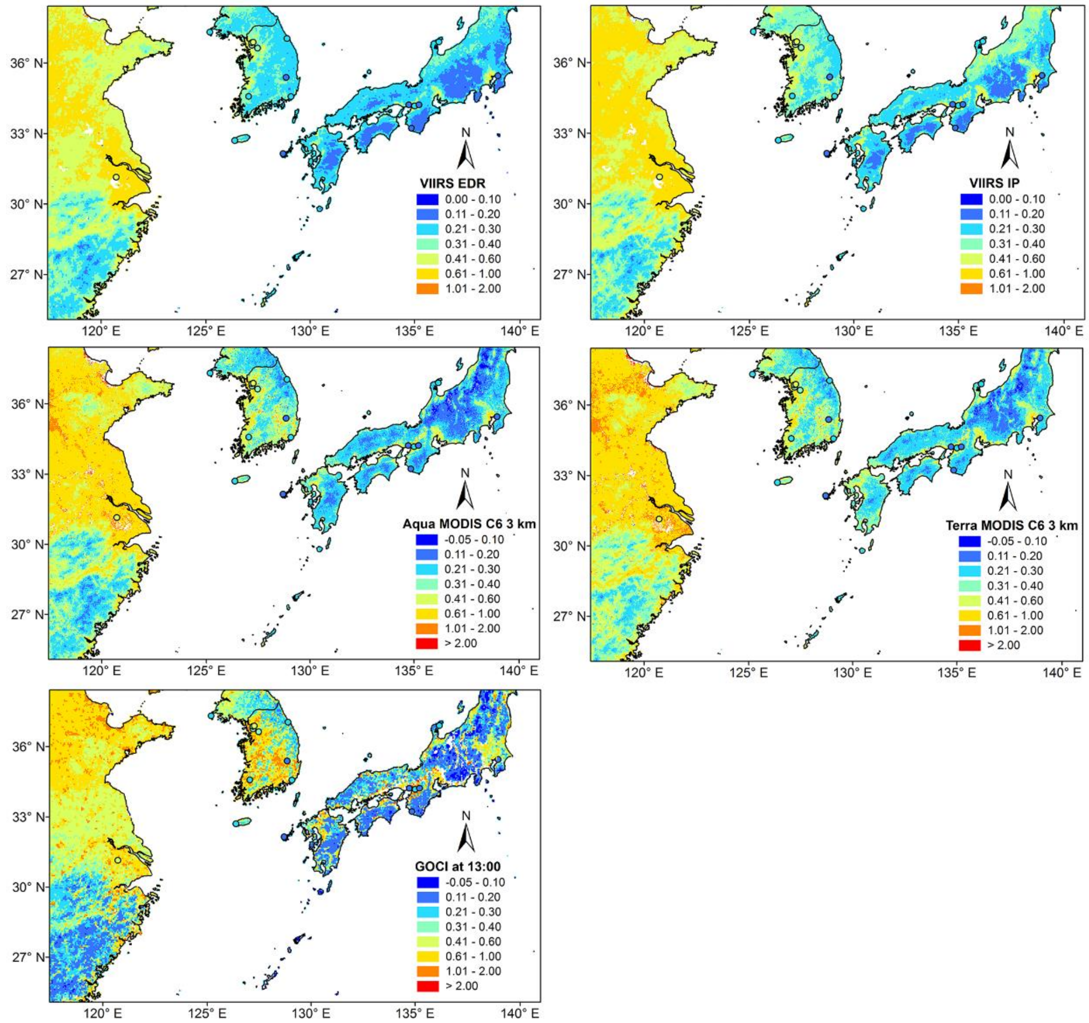

Figure 6. The distributions of the twelve months average AOD values from July 2012 to June 2013 from VIIRS EDR, VIIRS IP, Aqua MODIS C6 3 km, Terra MODIS C6 3 km, and GOCI datasets.