Spatial Aspects of Multi-Sensor Data Fusion: Aerosol Optical Thickness

Gregory Leptoukh
NASA Goddard Space Flight Center
Greenbelt, Maryland 20771, USA
Gregory.Leptoukh@nasa.gov

V. Zubko, A. Gopalan
RSIS
NASA Goddard Space Flight Center
Greenbelt, Maryland 20771, USA

Abstract—**The Goddard Earth Sciences Data and Information Services Center (GES DISC) investigated the applicability and limitations of combining multi-sensor data through data fusion, to increase the usefulness of the multitude of NASA remote sensing data sets, and as part of a larger effort to integrate this capability in the GES-DISC Interactive Online Visualization and Analysis Infrastructure (Giovanni). This initial study focused on merging daily mean Aerosol Optical Thickness (AOT), as measured by the Moderate Resolution Imaging Spectroradiometer (MODIS) onboard the Terra and Aqua satellites, to increase spatial coverage and produce complete fields to facilitate comparison with models and station data. The fusion algorithm used the maximum likelihood technique to merge the pixel values where available. The algorithm was applied to two regional AOT subsets (with mostly regular and irregular gaps, respectively) and a set of AOT fields that differed only in the size and location of artificially created gaps. The Cumulative Semivariogram (CSV) was found to be sensitive to the spatial distribution of gap areas and, thus, useful for assessing the sensitivity of the fused data to spatial gaps.**

Keywords-component; Remote sensing, Data tools, Data fusion, Value-added services, Aerosol optical thickness, MODIS, MISR, Terra, Aqua, Giovanni

I. INTRODUCTION

With the multitude of satellite data sets available from numerous missions and sensors, many of which are complementary to each other, there is an increasingly critical need to combine them through data fusion (DF) to derive the optimal benefits from the data. Often, information provided by an individual sensor might be incomplete, inconsistent, inadequate, and/or imprecise. Fusing of multi-sensor data, e.g., Aerosol Optical Thickness (AOT), can potentially create a more consistent, reliable, and complete picture of the space-time evolution of the underlying geophysical process (e.g., dust storms). Missing data from one sensor could be filled in with available co-located data from another sensor. For a given area, valid data from different sensors can be optimally combined (with error estimates) to produce a better estimate of some geophysical parameter. Although the Earth Observing System (EOS) program [1] significantly improved the interoperability of data from different sensors, two gridded products, for example, of the same parameter but from two different missions may still not be completely compatible with each other. Complications arise from the different spatial and temporal resolutions of the sensors, as well as the different sensor geometries.

The work described in this paper is part of the larger effort to enable DF in Giovanni. Our DF objective here is to increase the spatial coverage: filling orbital and other gaps through DF and incorporating the simplest DF algorithm of the Terra and Aqua AOTs. We provide a quick overview of Giovanni capabilities with emphasis on our plans for the added DF capability, using gridded daily mean AOT data from the Moderate Resolution Imaging Spectroradiometer (MODIS) onboard Terra and Aqua NASA satellites and MISR onboard Terra satellite. We concentrate on the following spatial aspects of differences in images to be fused: (a) degree of spatial overlap in Terra MODIS and MISR measurements over oceans; (b) benefits of fusing data from three sensors to spatial coverage increase; (c) effects of differences in spatial coverage in images on their properties expressed by various statistics; (d) effects of horizontal shifts (different measurement times) on these properties.

II. GIOVANNI

The NASA Earth Observing System (EOS) multi-satellite data archives are indispensable for studying regional or global atmospheric phenomena. Until recently, using this data required being able to locate and retrieve the relevant data coupled with a detailed understanding of the data's complicated internal structure. Consequently, this data was largely unusable to the public at large as gaining the knowledge required to carry out the data reduction is a time-consuming task which must be undertaken well in advance. Even for experienced users analysis of multi-sensor data sets that are typically in different formats, structures, and resolutions is a daunting task.

The NASA Goddard Earth Sciences Data and Information Services Center (GES DISC) has recognized this complexity and has taken a major step towards developing a user friendly Web interface that allows users to perform interactive analysis online without downloading any data, or needing to understand complicated data structures. The Goddard Interactive Online Visualization and Analysis Infrastructure or "Giovanni" (http://giovanni.gsfc.nasa.gov) addresses these objectives [2]. Giovanni has successfully demonstrated its utility as an interactive, online, analysis tool for data users to facilitate a wide spectrum of users in research, education, and the curious internet surfer.

One of the expressed interests of users worldwide has been to combine and fuse data from multiple sensors using Giovanni. GES DISC as a significant data archive location is uniquely positioned to address this need using data fusion (DF) techniques. DF is the intelligent merging or integration of data from multiple sources to extract more or better information than would be possible from the individual sources. With the vast quantity of satellite data sets available from numerous missions and sensors, many of which are complementary to each other, there is an increasingly critical need to combine these data to derive the most benefits from the data. Often, information provided by an individual sensor may be incomplete, inconsistent, inadequate, and/or imprecise. Fusing of multi-sensor data, e.g., Aerosol Optical Thickness (AOT) can potentially create more consistent, reliable, and complete picture of the space-time evolution of the underlying geophysical process. Missing data from one sensor could be intelligently "filled in" with available co-located data from another sensor to produce a better estimate of geophysical parameters.

III. Difference in Spatial Coverage between MODIS and MISR

Giovanni can be used to rapidly and efficiently create and visualize daily global 1°x1° maps of AOT (at 0.55 micron) using Terra and Aqua MODIS, and MISR Level-3 data products. Actually, we used Giovanni during the course of this study to identify interesting cases for data fusion. The typical large gaps, especially near the equatorial regions in the AOT daily mean field for both Terra MODIS and Aqua MODIS shown in Fig. 1 result from a combination of factors, including gaps between swaths from different orbits, and problems in AOT retrievals due to sun glint (over water), cloud cover, or very bright surfaces like deserts [3-5]. It is interesting to note that over most of the oceans (± 60° latitude) Terra MODIS and Terra MISR never measure over the same location – MODIS can't measure over sun glint while the narrow MISR swath measures within the "MODIS" sun glint. Thus it is advantageous to combine those data.

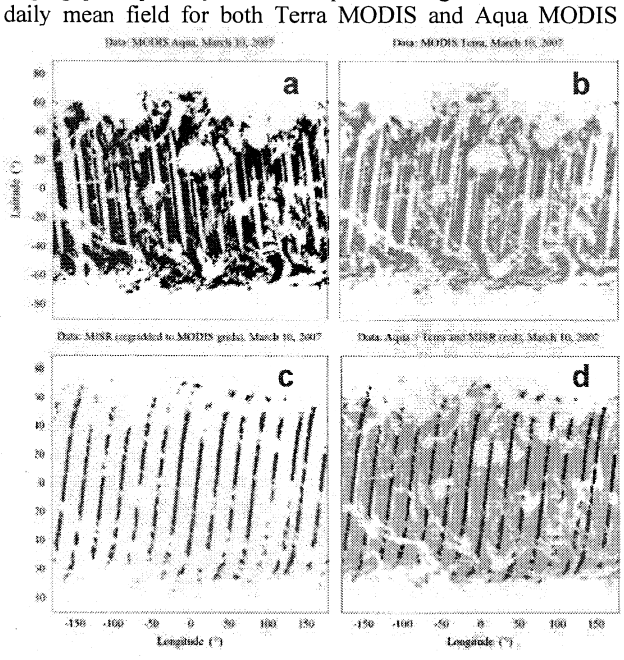

Figure 1. Sample global daily AOT coverage by indivividual sensors and the merged result: (a) Terra MODIS, (b) Aqua MODIS, (c) MISR, (d) merged.

IV. Data Merging

Our approach to data fusion is (1) to merge the data and then (2) interpolate to fill the gaps. This sequence is optimal in the sense that it preserves original data information most (least distortion) [6]. In this paper we mostly concentrate on the spatial aspects of the merging part.

For merging the data sets, we used weighted averaging, which is a family of methods based on arithmetic combinations of input values, such as linear combinations, weighted multiplication or ratios, and maximum likelihood estimate (MLE). The MLE emphasizes the use of different sources of data using statistics such as mean, standard deviation, and number of counts. For isotropic uncertainty, the MLE can provide a good approximation of the actual estimate of a feature from multiple observations. The MLE requires minimal a priori information, and it is easy to incorporate user-supplied weights for the data sources. For a set of N independent observations (F_k) of the same parameter, the MLE estimate is:

$$F_{\mathrm{MLE}} = \sum_{k=1}^{N} \frac{F_k}{\sigma_k^2} \Big/ \sum_{k=1}^{N} \frac{1}{\sigma_k^2}, \qquad (1)$$

where σ_k is the variance of the Gaussian noise affecting the observations. The σ_k is computed for each cell selected for

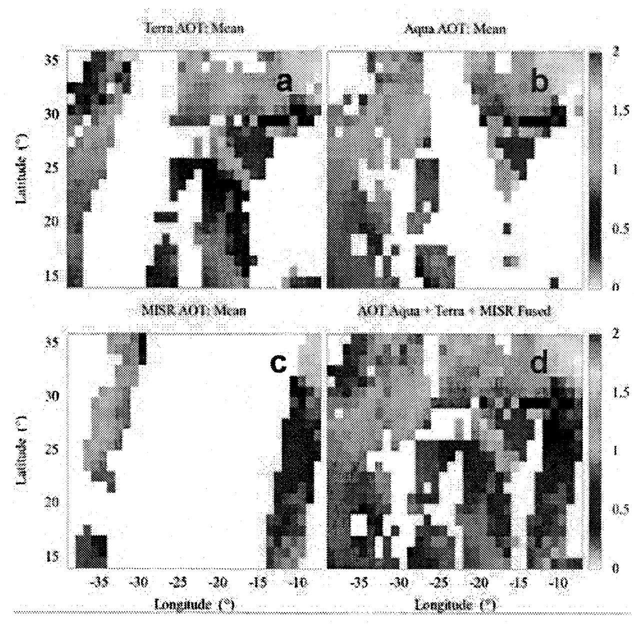

Figure 2. Single sensor AOT measurements and the merged result over a high dust event area: (a) Terra MODIS, (b) Aqua MODIS, (c) MISR, (d) merged.

data fusion and the expected estimate of F is calculated using (1). To illustrate the work of the merging algorithm, Fig.2 presents a subset of the original data sets for AOT MODIS Terra and Aqua and MISR for March 10, 2007 (a,b,c) and the result of their merging (d). Fig 2d demonstrates the large reduction in the fraction of pixels with no data.

V. Spatial Gap Study

In this section, we report on a study of a "single gap" experiment, the results of which suggested that the Cumulative Semivariogram (CSV) [7] of an area is sensitive to the spatial distribution and fraction of gap areas. We chose a 20°x30° region of the AOT field (from the March 2006 monthly Terra AOT) containing variations in gradient from low to high. The original AOT region as seen in Fig.3a did not contain any gaps. To study the sensitivity of the CSV to gaps, we generated a collection of data sets with a single gap each of size 10°x10° (Fig. 3c, 3d) in the original AOT data set. We then attempted to reconstruct the original input field by applying our data fusion approach. The respective CSVs are summarized in Fig. 3b. They indicate that CSV values were indeed sensitive to the gradients that might exist in the field but which might not have been explicitly reproduced because of measurement limitations. Generally, the CSV deviation is controlled by two factors. First, The CSV is a quantity normalized by the number of pairs of points, and when we remove an area from the analysis (create a gap), the number of pairs decreases, which will tend to increase the CSV. Second, the contribution of the low gradient areas to the CSV is certainly smaller than that of the high gradient areas. Thus, when we remove the low gradient area, the first factor prevails over the other, and the CSV increases. In the case when we remove the high gradient area, the second factor overwhelms the first, and the CSV decreases.

VI. Spatial Autocorrelations

Another question we were trying to answer was the following: What is the influence of spatial shifts, or in different words, how strongly images change when measurements by different sensors are separated in time by a couple of hours, and the studied event has moved due to winds. We assume here that winds are uniform across the selected area, and wind vectors point to the same direction at all altitudes.

Fig. 4 presents spatial autocorrelation properties for two distinct sample areas of MODIS Terra AOT (Collection 5) measured March 2006: one with high AOT gradient and the other with low AOT gradient. We calculated the Pearson correlation coefficient r and RMS error between the original dataset and itself spatially shifted along the lines North-South (N↔S) and East-West (E↔W). The plots demonstrate different behavior of the autocorrelation properties for various areas. For the high gradient area, the autocorrelation along N↔S disappears beyond 10° whereas the autocorrelation

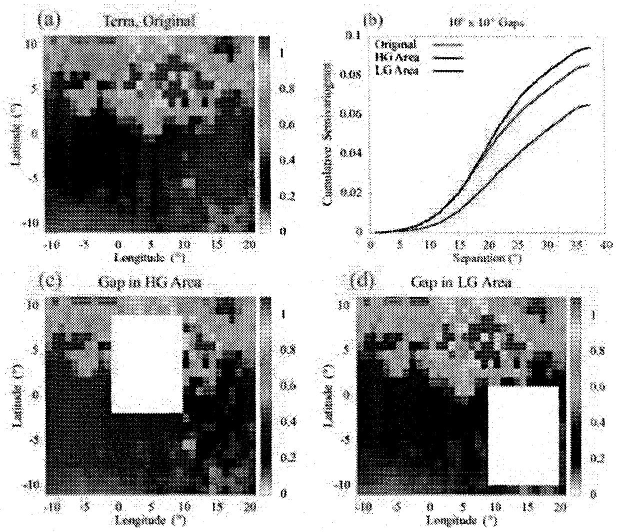

Figure 3. Cumulative Semivariogram sensitivity to simulated gaps in MODIS Terra AOT data: (a) original data; (b) Cumulative Semivariograms; (c, d) simulated 10°x10° gaps in high and low gradient areas.

along E↔W remains relatively high even at 25°. This is consistent with the dynamical properties of this area at this time of year (March), when streams of aerosol dust usually move to the West from Sahara. For another area corresponding to the Pacific Ocean, the autocorrelation properties along N↔S and E↔W seem quite close up to about 20° and demonstrate a relative uniformity in AOT.

VII. Conclusions

It was shown on the example of the AOT that combining data obtained from various sources can result in significant reduction in the fraction of pixels with no data, thus increasing the spatial coverage.

The results of our "spatial gap" and "spatial shift" experiments provided evidence that Cumulative Semivariogram and autocorrelation coefficient can serve as good indicators of the spatial variability in the data, reflecting both omnidirectional and anisotropic behavior.

We demonstrated that prior to fusing data from different sensors, one needs to assess differences in these images, and try to separate differences caused by differences in spatial coverage or due to horizontal shifts, from the "real" differences coming from differences in sensor capabilities, algorithms, or calibration. The former set of these spatial differences was addressed in the current paper, where we demonstrated significant differences appearing for pairs of absolutely identical images, after imposing obscurations in different areas or being shifting one image horizontally. These differences can be qualitatively seen in the resulting images, and quantitatively assessed by computing variograms or spatial autocorrelations, and also other not-spatial statistics. However, we still don't know how to separate these spatial effects from the other "physical" effects.

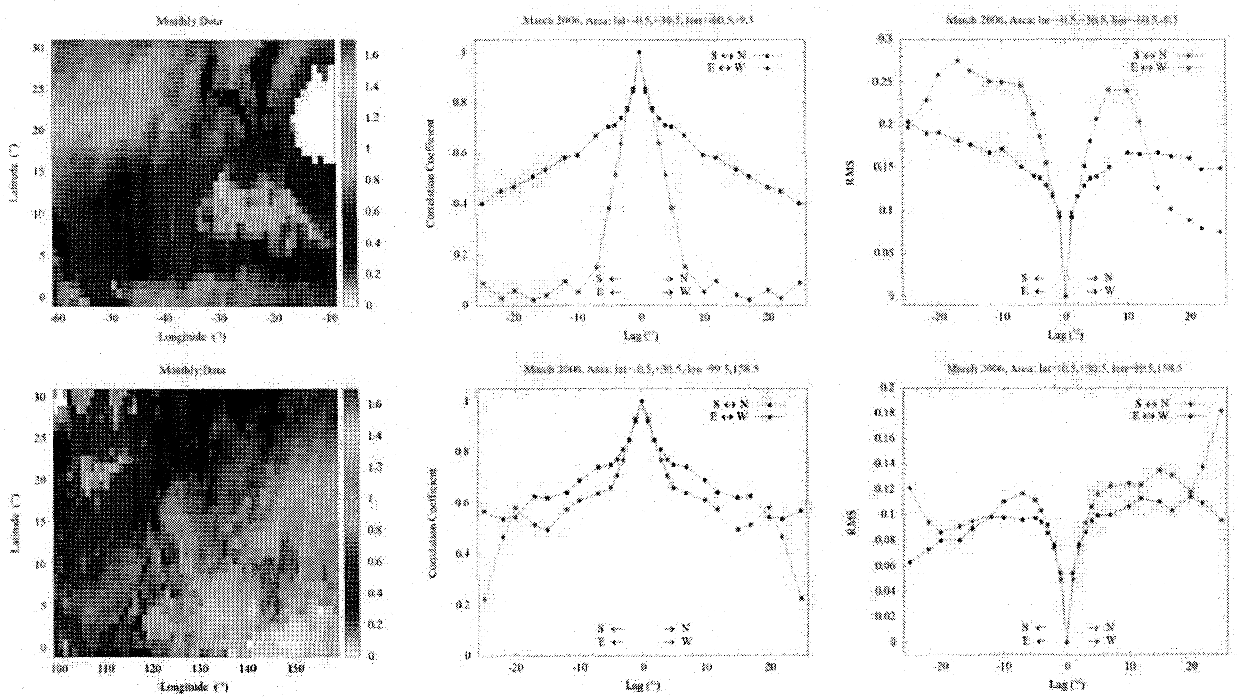

Figure 4. The correlation coefficient and RMS for the scatter plots for MODIS Terra AOT Monthly data for March 2006 for two areas of high (top) and low (bottom) AOT as a function of the spatial shift in various directions

The current study is a preparation for integration of the aerosol multi-sensor data intercomparison and fusion into Giovanni.

REFERENCES

[1] C.L Parkinson, A. Ward, and M.D. King, editors,. "Earth Science Reference Handbook, a Guide to NASA's Earth Science Program and Earth Observing Satellite Missions", National Aeronautics and Space Administration, Washington, D.C., 2006

[2] J. Acker, and G. Leptoukh, "Online analysis enhances use of NASA earth science data," EOS, Transactions, American Geophysical Union, vol. 88, pp 14-17, 2007.

[3] R. Kahn, M. Garay, D. Nelson, K. Yau, M. Bull, B. Gaitley, and J. Martonchik, "Satellite-derived aerosol optical thickness over dark water from MISR and MODIS: comparisons with AERONET and implications for climatological studies", Journal of Geophysical Research, 2006, In review.

[4] L.A. Remer, Y.J. Kaufman, D. Tanre, S. Mattoo, D.A. Chu, J.V.,Martins, R.-R. Li, C. Ichoku, R.C. Levy, R.G. Kleidman, T.F. Eck, E. Vermote, and B.N. Holben, "The MODIS aerosol algorithm, products and validation," Journal of the Atmospheric. Sciences, vol. 62, pp 947-973, 2005.

[5] N.C. Hsu, S-C.Tsay, M.D. King, and J.R. Herman, "Aerosol properties over bright-reflecting source regions," IEEE Transactions on Geoscience and Remote Sensing, vol. 42, #3, pp 557-569, 2004

[6] G. Leptoukh, V. Zubko, M. Nirala, A. Gopalan, and W. Teng. "Multi-sensor data fusion of NASA satellite measurements: Aerosol optical thickness", Photogrammetric Engineering & Remote Sensing, submitted, 2007

[7] Z. Şen. "Cumulative semivariogram models of regionalized variables," Mathematical Geology, vol. 21(8), pp. 1712-1724, 1989