Influence of Averaging Method on the Evaluation of a Coastal Ocean Color Event on the U.S Northeast Coast

James G. Acker[1]
Stephanie Schollaert Uz[2]
Suhung Shen[3]
Gregory G. Leptoukh[4]

[1]Goddard Earth Sciences Data and Information Services Center / Wyle IS LLC
[2]Earth System Science Interdisciplinary Center, University of Maryland
[3]Goddard Earth Sciences Data and Information Services Center / George Mason University
[4]Goddard Earth Sciences Data and Information Services Center / NASA

Presented at the American Geophysical Union 2010 Ocean Sciences Meeting, Portland, Oregon, February 22-26

ABSTRACT

Application of appropriate spatial averaging techniques is crucial to correct evaluation of ocean color radiometric data, due to the common log-normal or mixed log-normal distribution of these data. Averaging method is particularly crucial for data acquired in coastal regions. The effect of averaging method was markedly demonstrated for a precipitation-driven event on the U.S. Northeast coast in October-November 2005, which resulted in export of high concentrations of riverine colored dissolved organic matter (CDOM) to New York and New Jersey coastal waters over a period of several days. Use of the arithmetic mean averaging method created an inaccurate representation of the magnitude of this event in SeaWiFS global mapped chl *a* data, causing it to be visualized as a very large chl *a* anomaly. The apparent chl *a* anomaly was enhanced by the known incomplete discrimination of CDOM and phytoplankton chlorophyll in SeaWiFS data; other data sources enable an improved characterization. Analysis using the geometric mean averaging method did not indicate this event to be statistically anomalous. Our results predicate the necessity of providing the geometric mean averaging method for ocean color radiometric data in the Goddard Earth Sciences DISC Interactive Online Visualization ANd aNalysis Infrastructure (Giovanni).

Introduction

The widespread availability of mapped remote-sensing data products which have been binned and averaged spatially and temporally (*Level 3 data*) makes this type of data useful for many different varieties of Earth Science research. It is recognized, however, that the methodology of binning and averaging can introduce biasing in the presentation and interpretation of these data. The output result of the averaging method is directly related to the statistical distribution of the data.

While many types of Earth Science data are normally distributed, ocean color radiometric data, particularly the chlorophyll *a* (chl *a*) data product, usually have a log-normal or mixed log-normal distribution. This characteristic of the data is most commonly observed in coastal regions which exhibit wide ranges in chl *a* values.

In this presentation, we examine how the arithmetic mean averaging method, as well as the radiometric characteristics of the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) data products – overestimated the magnitude of an apparent chl *a* anomaly caused by a unique precipitation-driven event on the U.S. East Coast in November 2005.

October-November 2005 Precipitation Event

In October 2005, anomalously high precipitation occurred in the U.S. Northeast. The elevated precipitation events were unusual for autumn, and several Northeast states set all-time records for October precipitation. (Figure 1)

To better characterize this event, Tropical Rainfall Measuring Mission (TRMM) data in Giovanni was utilized to show the monthly precipitation pattern and the total precipitation amount (Figure 2). We also selected stream gauge stations from the United States Geological Survey "Real-Time Water Data for the Nation" Web site (http://waterdata.usgs.gov/nwis/rt) to examine the streamflow response to the October precipitation (Figure 3). An ocean monitoring site off of Sandy Hook, New Jersey (NOAA) also provided conductivity data which demonstrated the effect of the increased flow of the Hudson River on New York and New Jersey coastal waters (Figure 4).

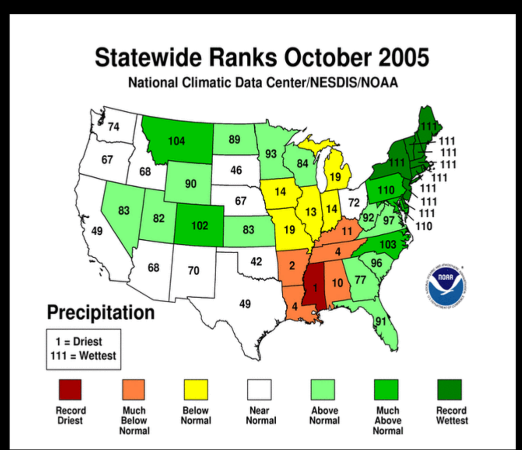


Figure 1: National Climatic Data Center ranking of precipitation amounts for October 2005.

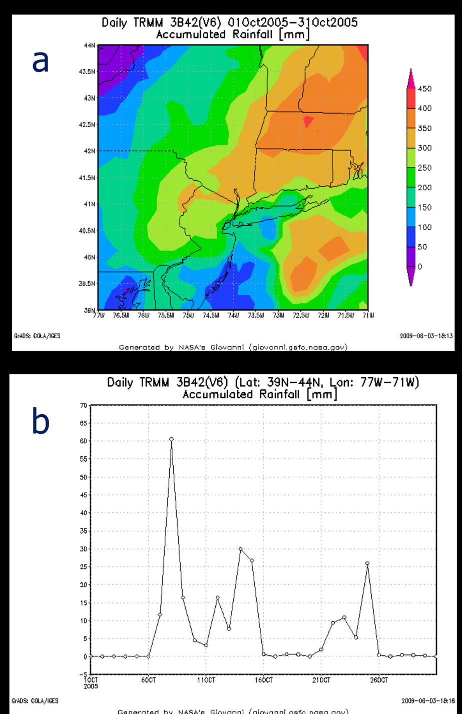


Figure 2: (a) TRMM daily accumulated rainfall (mm) for October 2005 over southern New England, eastern New York and Pennsylvania, and New Jersey. (b) Time-series of TRMM daily accumulated rainfall for this region in October 2005.

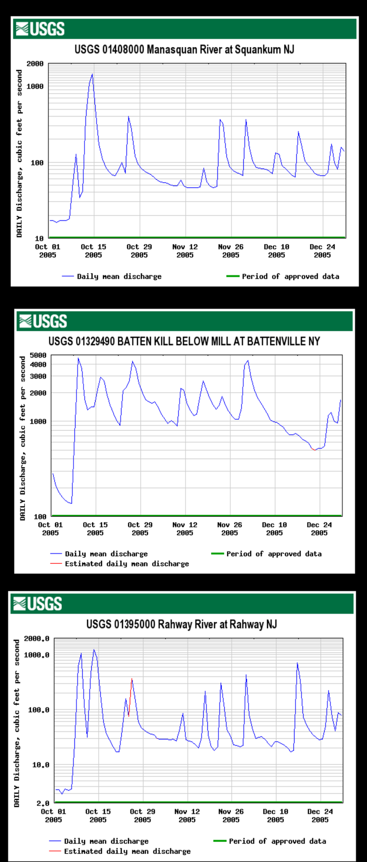


Figure 3: USGS streamflow (cfs) for selected gauge stations in New York and New Jersey, October-December 2005.

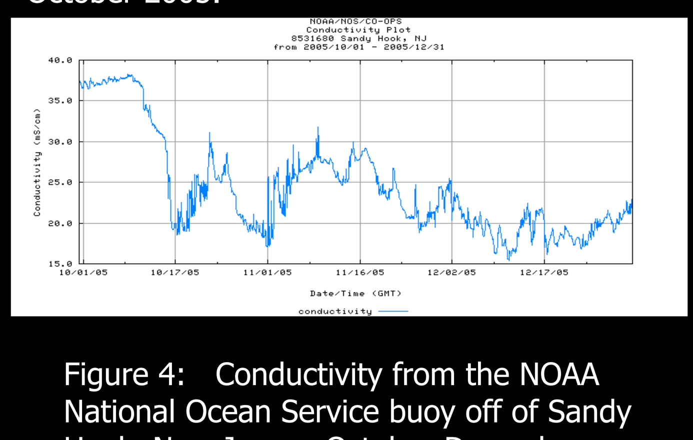
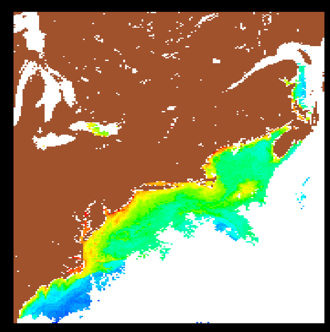


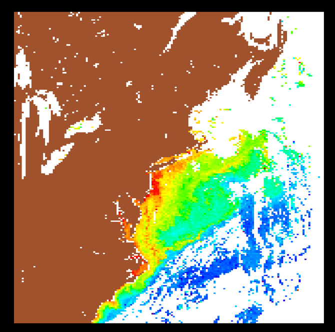
Figure 4: Conductivity from the NOAA National Ocean Service buoy off of Sandy Hook, New Jersey, October-December 2005, showing the effect of increased flow in the Hudson River caused by the elevated rainfall in October 2005.

EFFECTS OF THE EVENT: Coastal Observations by SeaWiFS and MODIS

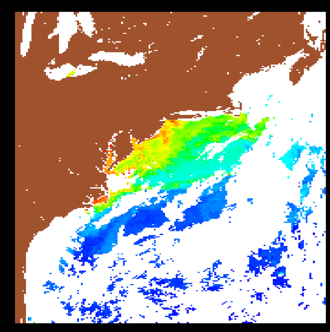
Data from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) and the Moderate Resolution Imaging Spectro-radiometer (MODIS) on the Aqua satellite was examined to determine the approximate length of time that northeastern coastal waters were influenced by the elevated rainfall event n October 2005. This examination indicated that the period of time with the strongest influence occurred from October 30 – November 6, 2005. Imagery prior to October 30 was sparse, likely due to cloud cover associated with the elevated rainfall. An image acquired on November 20 indicated that there may have been a waning influence through that date.



MODIS-Aqua chl *a* browse image, October 30, 2005.



SeaWiFS chl *a* browse image, November 3, 2005.



MODIS-Aqua chl *a* browse image, November 6, 2005.

The SeaWiFS chl *a* anomaly plot generated for the study region (Figure 5) indicated a large chl *a* anomaly in November 2005. The time-series of SeaWiFS chl *a* anomalies (Figure 6) showed that this anomaly was unprecedented over the length of the mission. The time-series values were calculated by arithmetic mean.

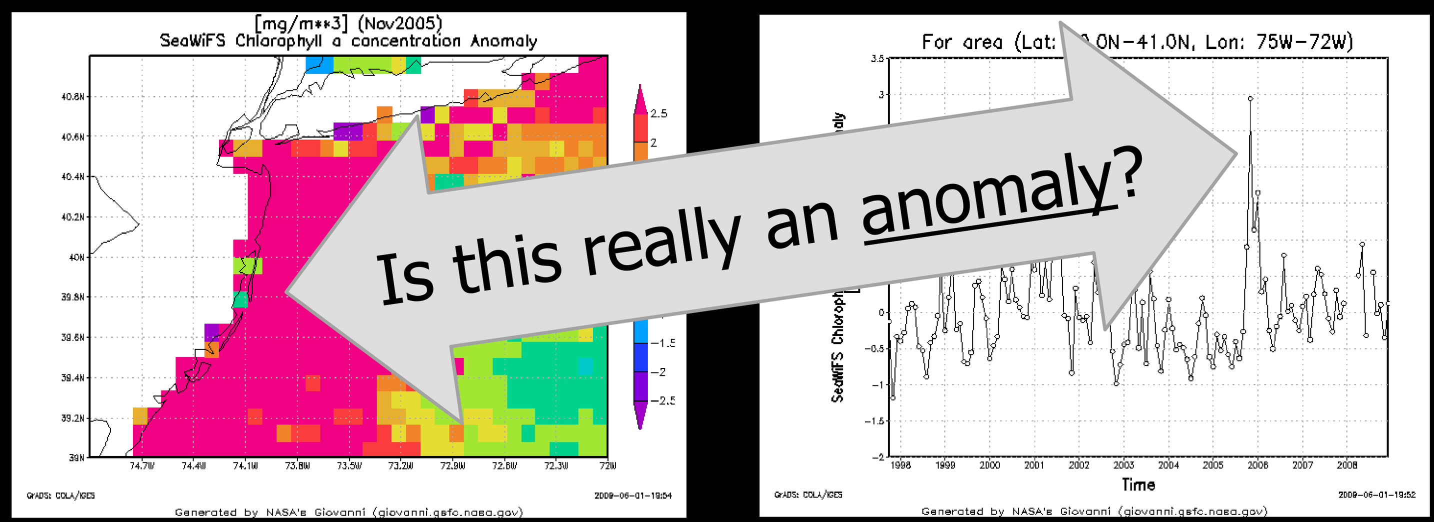


Figure 5: SeaWiFS chl *a* anomaly plot generated by Giovanni for the study region, November 2005, showing the high positive chl *a* anomaly along the coast of New Jersey and western Long Island. The anomaly palette range was extended to values of -2.5 (purple) to +2.5 (red) for this plot.

Figure 6: SeaWiFS chl *a* anomaly time-series generated by Giovanni for the study region, September 1997 through December 2008. The extremely high chl *a* anomaly value for the study region in November 2005 is clearly visible.

Calculation of mean values and climatological anomalies

Giovanni provides an anomaly analysis capability utilizing climatologies generated by the data provider – in this case, the ten-year SeaWiFS chl a climatology is provided by the Ocean Biology Processing Group (OBPG) at Goddard Space Flight Center (GSFC). The anomaly is calculated for the designated time period, and can be averaged over several months for periods up to one year. The anomaly is calculated by

$$X_a(i) = X(i) - X_m(i)$$

where:

- *i* denotes month (1, 2, ... , 12)
- $X_a(i)$ represents the anomaly data for month *i*
- $X(i)$ is the original data for month *i*
- $X_m(i)$ is the monthly climatology data for month *i*

Mean calculation methods

Given a selected area with *n* data points X_1, X_2, ... X_n, X_i denotes the Level 3 mapped data points within a selected area.

The arithmetic mean is calculated by:

$$X_{avg} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

The geometric mean is calculated by $X_{geo} = e^{m_i}$

where $m_i = \frac{1}{n}\sum_{i=1}^{n}\ln(X_i)$

Analyzing the influence of averaging method

Utilizing the Garver-Siegel-Maritorena (GSM) chl *a* and a_{cdm} data products, we sought to determine if the elevated coastal data values in November 2005 were anomalous. This analysis was performed independent of Giovanni. Geometric mean chl *a* was calculated by taking the median value of each scene for each time step to create a time series; the seasonal mean was calculated by averaging each month for the whole time series to create a 12-month climatology; subtracting the climatology from the time series yielded the chl *a* anomaly. Bio-optical data such as chl *a* have a lognormal distribution at a variety of spatial and temporal scales (Campbell, 1995) and therefore must be averaged geometrically, by taking the average of the lognormal distribution of the data or by taking the median value.

The time series and anomaly time series were calculated for GSM chl *a* and GSM a_{cdm}, and are shown in Figure 7.

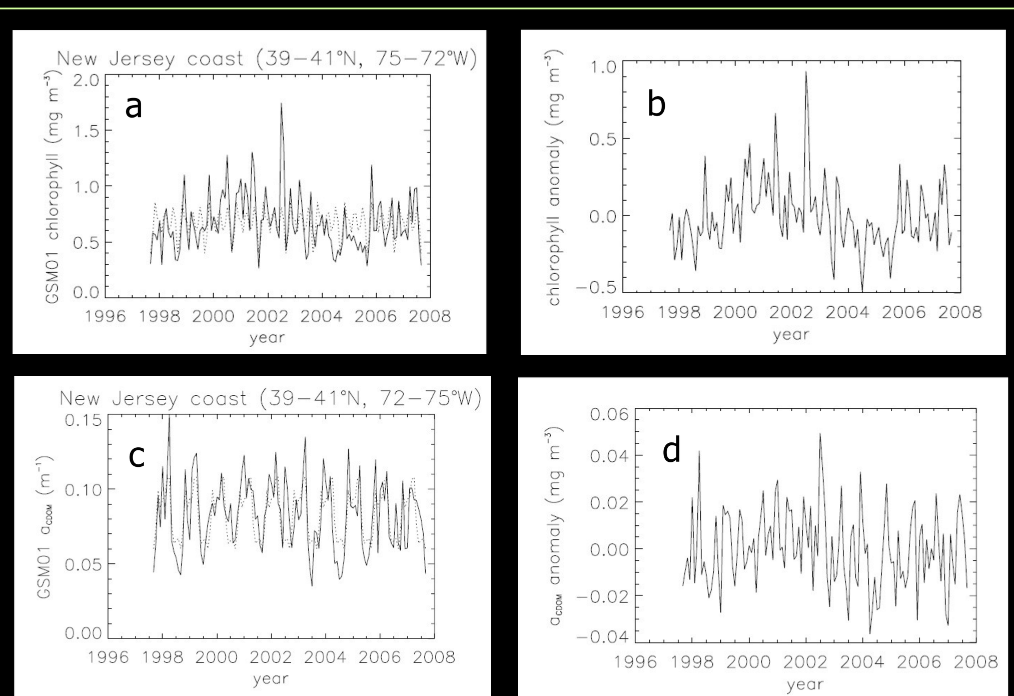


Figure 9. GSM chl *a* and a_{cdm} time series, with accompanying anomaly time series. (a) GSM chl *a*. (b) GSM chl *a* anomaly. (c) GSM a_{cdm}. (d) GSM a_{cdm} anomaly. Dotted line in (a) and (c) is the climatology that was subtracted to create the anomaly time series.

Comparison of chl *a* data products in Giovanni

Giovanni currently provides three different chl *a* data products: MODIS–Aqua chl *a*, SeaWiFS chl *a*, and Garver-Siegel Maritorena (GSM) merged (MODIS+SeaWiFS) chl *a*. The GSM merged chl *a* data product is less sensitive to interference from CDOM than either the MODIS-Aqua or SeaWiFS chl *a* products. (Note that the most recent SeaWiFS reprocessing has improved the discrimination of CDOM and chl *a*; a comparison will be shown subsequently.) The three data products are shown side-by-side in Figure 8 for the study area in November 2005. The SeaWiFS data set is SeaWiFS.R5.2.

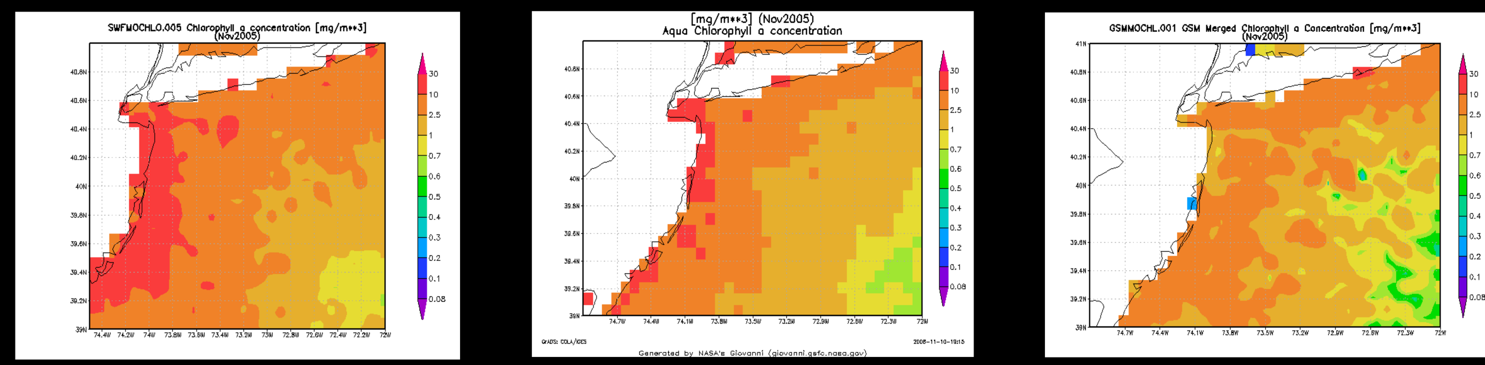


Figure 8. Comparison of SeaWiFS chl *a* (left), MODIS-Aqua chl *a* (center) and GSM chl *a* (right).

It was initially thought that the high SeaWiFS chl *a* values might be due to a late-autumn phytoplankton bloom, caused by the delivery of nutrients to the coastal zone by the increased river flows. It is clear from this comparison, however, that SeaWiFS chl *a* is higher than MODIS-Aqua chl *a*, and significantly higher than GSM chl *a*. Due to the known incomplete discrimination of chl *a* and CDOM in SeaWiFS data, the GSM a_{cdm} data product (absorption coefficient due to detrital and dissolved organic matter) was examined. This data product, shown in Figure 9, indicated a significant contribution to the absorption signal from dissolved detrital and organic matter.

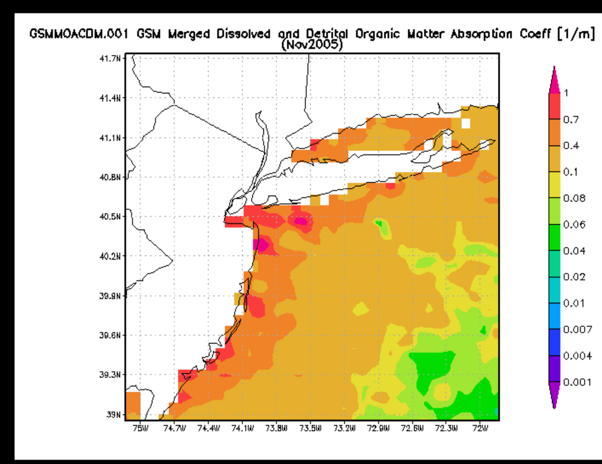


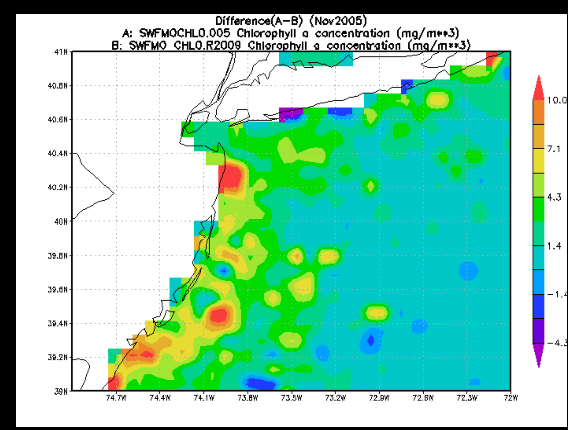
Figure 9. GSM a_{cdm} data product for the study area, November 2005.

Comparison of SeaWiFS.R5.2 and SeaWiFS.R2009 data products

Following the SeaWiFS data reprocessing in December 2009, the reprocessed SeaWiFS data products (R2009) were added to *Ocean Color Radiometry* in Giovanni. The new climatology has not yet been added, so anomalies cannot be calculated for the R2009 data products. It is possible, however to directly compare the R2009 SeaWiFS data to the previous version (R5.2) of the data products.

According to the OBPG, the refinements to the data substantially improve agreement in chlorophyll retrievals relative to ground truth measurements in turbid and highly productive regions. Figure 10 shows a difference plot between the R5.2 and the R2009 chl *a* data for the study area in November 2005, calculated as (R5.2 – R2009). Thus, positive values indicate where the R5.2 chl *a* exceeds R2009 chl *a*. The predominance of positive values indicates that R2009 chl *a* was reduced, further supporting that the high chl *a* values in the SeaWiFS R5.2 data product off the U.S. northeast coast in November 2005 were primarily due to increased export of riverine CDOM and detritus by the elevated river flow.

Figure 10. Difference plot of SeaWiFS. R5.2 chl a and SeaWiFS.R2009 chl *a* for November 2005, showing that R5.2 chl *a* is significantly greater than R2009 chl *a* near the coast. This plot demonstrates that the reprocessing reduced the influence of CDOM on chl *a* retrievals in turbid waters.



Discussion

The GSM chl *a* and a_{cdm} time series do not show significantly elevated values of either data product, nor a noteworthy anomaly for either data product, in November 2005. We attribute this distinct difference from the SeaWiFS data in Giovanni to two factors; one, the incomplete discrimination of phytoplankton chlorophyll and CDOM in SeaWiFS data; and two, more importantly, the difference in results from the use of the arithmetic mean and the geometric mean averaging method. The lognormal distribution of chl *a* data, combined with the sensitivity of the arithmetic mean to outliers, caused the SeaWiFS data anomaly calculation in Giovanni to produce an inaccurately high monthly anomaly for November 2005. This anomaly was likely enhanced by the limited amount of data available for the month, causing the elevated SeaWiFS chl *a* values occurring early in the month, due to the export of riverine CDOM, to dominate the monthly statistics.

Because the current arithmetic mean averaging method in Giovanni was inherited from GrADS, and because Giovanni serves many different types of remote-sensing data, the need for the geometric mean calculation for ocean color data has not been a high priority. Our results indicate that in order to provide the most accurate presentation of chl *a* data in Giovanni, it is necessary to add the geometric mean calculation for ocean color data. This improvement will allow more representative results and improved comparisons with new data products that will be added to Giovanni by upcoming projects.

Reference: Campbell, J. W. (1995), The lognormal distribution as a model for bio-optical variability in the sea. Journal of Geophysical Research, 100(C7), 13,237–13,254.