Statistical Considerations of Data Processing in Giovanni Online Tool

[1] Shen, Suhung (sshen@disc.gsfc.nasa.gov)

[2] G. Leptoukh (leptoukh@disc.gsfc.nasa.gov)

[3] J. Acker (acker@disc.gsfc.nasa.gov)

[2] S. Berrick (sberrick@disc.gsfc.nasa.gov)

[1] George Mason University, GSFC, Code 610.2, Greenbelt, MD 20771
[2] NASA/GSFC, Code 610.2, Greenbelt, MD 20771
[3] SSAI, GSFC, Code 610.2, Greenbelt, MD 20771

The GES DISC Interactive Online Visualization and Analysis Infrastructure (Giovanni) is a web-based interface for the rapid visualization and analysis of gridded data from a number of remote sensing instruments. The GES DISC currently employs several Giovanni instances to analyze various products, such as Ocean-Giovanni for ocean products from SeaWiFS and MODIS-Aqua; TOMS & OMI Giovanni for atmospheric chemical trace gases from TOMS and OMI, and MOVAS for aerosols from MODIS, etc. (http://giovanni.gsfc.nasa.gov)

Foremost among the Giovanni statistical functions is data averaging. Two aspects of this function are addressed here. The first deals with the accuracy of averaging gridded mapped products vs. averaging from the ungridded Level 2 data. Some mapped products contain mean values only; others contain additional statistics, such as number of pixels (NP) for each grid, standard deviation, etc. Since NP varies spatially and temporally, averaging with or without weighting by NP will be different. In this paper, we address differences of various weighting algorithms for some datasets utilized in Giovanni.

The second aspect is related to different averaging methods affecting data quality and interpretation for data with non-normal distribution. The present study demonstrates results of different spatial averaging methods using gridded SeaWiFS Level 3 mapped monthly chlorophyll *a* data. Spatial averages were calculated using three different

methods: arithmetic mean (AVG), geometric mean (GEO), and maximum likelihood estimator (MLE). Biogeochemical data, such as chlorophyll *a,* are usually considered to have a log-normal distribution. The study determined that differences between methods tend to increase with increasing size of a selected coastal area, with no significant differences in most open oceans. The GEO method consistently produces values lower than AVG and MLE. The AVG method produces values larger than MLE in some cases, but smaller in other cases. Further studies indicated that significant differences between AVG and MLE methods occurred in coastal areas where data have large spatial variations and a log-bimodal distribution instead of log-normal distribution.