Blue Waters Professor Allocation Annual Report: February 1, 2016 – January 31, 2017

Title: Satellite remote sensing and 3D radiative transfer modeling for improved weather and climate predictions

PI: Larry Di Girolamo, Department of Atmospheric Sciences, University of Illinois at Urbana-Champaign

Collaborator(s): Michelle Butler, Landon Clipp, Dongwei Fu, Yizhao Gao, Yan Liu, David Raila, Sean Stevens, John Towns, Kandace Turner, Shaowen Wang, Guangyu Zhao, University of Illinois at Urbana-Champaign; Mike Garay, Jet Propulsion Laboratory; Alexandra Jones, Princeton University; Ralph Kahn, NASA Goddard Space Flight Center; Lusheng Liang, SSAI and NASA Langley Research Center; Kent Yang, HDF Group; and Ping Yang, Texas A&M University

Corresponding Author: Larry Di Girolamo, gdi@illinois.edu

Executive Summary:

Our multi-pronged approach for tackling key problems in weather and climate research using satellite observations and 3D radiative transfer has had an incredible year of success: (1) The launch of the Terra Data Fusion project with the support of NASA, which aims to tackle challenges for data fusion of Terra's five instruments, (2) improvements to NASA's infrastructure for the transfer of a PB of Terra data to Illinois, (3) research results highlighted on the front cover of IEEE TGRS, (4) enhanced understanding of global microphysical properties of water and ice clouds, and (5) progress towards finalizing two manuscripts for publication that have emerged from former Blue Water Graduate Fellow Alexandra Jones' Ph.D. dissertation, which developed and benchmarked the first spectrally integrating, atmospheric 3D Monte Carlo radiative transfer model. Blue Waters allocation was underutilized owing to Jones' commitment to her position at Princeton, continued use of ROGER, the delay in the NASA contract, and the slower than anticipated data transfer rates from NASA.

Description of Research Activities and Results:

Research in weather and climate has massive societal benefits, and indeed has been one of the leading drivers for advancing supercomputing infrastructures. One of least understood and most important aspect of the weather and climate system are Earth's clouds. Clouds cover about 68% of our planet. They are one of the most interconnected components of the Earth System, playing a key role in the Earth's hydrological cycle, regulating the incident solar radiation field more than any other atmospheric variable, and acting as the most important greenhouse constituent in our atmosphere. As such, they modulate a wide range of processes on Earth. The Intergovernmental Panel on Climate Change (IPCC) affirms that the role of clouds remains the leading source of uncertainty in anthropogenic climate change predictions. In addition, the role of cloud microphysics and cloud-radiation interactions in the timing and intensity of weather events remains an active area of research.

To make headway in reducing uncertainty in weather and climate predictions, the World Meteorological Organization and the IPCC defined a list of Essential Climate Variables (ECVs) requiring global satellite observations (http://www.wmo.int/pages/prog/gcos). It has been established that ~2/3 of the ECVs derived from satellite do not meet accuracy requirements, therefore calling for improvements in the algorithms and technologies used by satellites. For improving algorithms, one of the key recommendations from the NRC 2007 Decadal Survey on Earth Science and Applications from Space (NRC 2007) is clear: *"… experts should… focus on providing comprehensive data sets that combine measurements from multiple sensors."* This, in part, targets NASA's flagship of the Earth Observing System called Terra. Terra was launched in 1999 and continues to collect data for Earth sciences using five instruments: the Moderate-resolution Imaging Spectroradiometer (MODIS), the Multi-angle Imaging SpectroRadiometer (MISR), the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), the Clouds and Earth's Radiant Energy System (CERES), and the Measurements of Pollution in the Troposphere (MOPITT). Terra data is amongst the most popular NASA datasets, serving not only the scientific community, but also governmental, commercial, and educational communities.

While the need for data fusion and the ability for scientists to perform large-scale analytics with long records have never been greater, the challenge is particularly acute for Terra, given its growing data volume (>1 petabyte), the storage of different instrument data at different NASA centers, the different data file formats and projection, and inadequate cyberinfrastructure. We recently initiated the Terra Data Fusion (TDF) Project, supported under NASA Grant Number NNX16AM07A, to tackle two long-standing problems: 1) How do we efficiently generate and deliver Terra data fusion products; 2) How do we facilitate the use of Terra data fusion products by the community in generating new products and knowledge through national computing facilities, and disseminate these new products and knowledge through national data sharing services? Blue Waters provides the computational resources needed to solve these problems.

The TDF project requires the transfer of the Terra L1B data (~ 1 PB) from the three NASA data centers to Blue Waters. This has been challenging owing to the inadequate infrastructure that NASA has for filling and delivering such large data orders. Our work with NASA over the past year has resulted in a 2 to 5 x increase in NASA's ability to deliver Terra instrument data to the community. Still, this is only averaging ~150 to 200 MB/s, so the transfer is still slow. But at this point, we have retrieved about ¾ of the Level 1B data and anticipate the transfer to be complete in about 2 months. The slow transfer rates have postponed some of our use of budgeted BW node hours.

Code development for fusion of Terra instrument data is progressing on schedule. Our so called "basic fusion" product is near completion, and testing on Blue Waters has shown that mission-scale processing requires only ~ 32,000 node hours, which is exciting as it points to our ability to derive other mission-scale products with low amounts of node hours. However, storage cost is high for the basic fusion files at 1 PB, compressed.

We have been working on three scientific use cases that help sharpen and push the project forward: (1) monitoring climate change from Terra, (2) retrieving liquid water cloud drop size distribution, and (3) retrieving cirrus cloud ice crystal shapes. The first in a series of planned studies to monitor climate change from Terra was reported in last year's BW report, where we noted that the first paper of our finding was in review. The paper has since made it to print (Zhao *et al*. 2016), and we were especially thrilled that the journal selected to highlight our results on the front cover of the issue! See Figure 1. We broadly advertised the results of this paper at several meetings and conferences (see below).

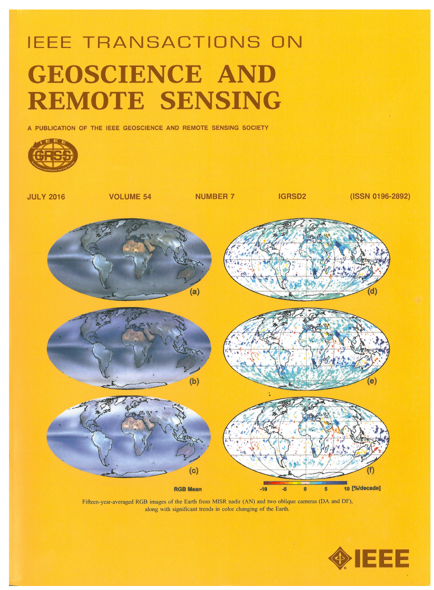

Figure 1. Front cover of the July 2016 issue of IEEE Transactions on Geoscience and Remote Sensing, showcasing results from "Regional changes in Earth's color and texture observed from space over a 15-year period" by Zhao et al. (2016).

Results of the second use case, where we fused the Terra MISR and MODIS data to determine biases in cloud drop size datasets that are in widespread use, were also presented in last year's BW report and came out in journal print in 2015. In the past year, we extended that study under NASA Grant Number NNX14AJ27G, to examine the underlying causes of the biases. This has become the research thesis of graduate student Dongwei Fu. This required additional code development for large scale analysis on BW, with some initial success already made. For example, to our surprise, our analysis has shown that cloud spatial heterogeneity alone (thought to be the leading cause of the bias) only explained about 30 to 50% of the bias, raising concerns that our understanding of the issues that led to the bias are incomplete. We are currently testing additional hypotheses that could explain the remaining bias.

On The third use case, we are working closely with Prof. Ping Yang at Texas A&M on a specialized MISR and MODIS fusion dataset designed for retrieving ice cloud microphysical properties. This work is supported under NASA Grant Number NNX15AQ25G. The codes for this are now complete, and we have produced one day of test data to Prof. Yang and his students for analysis and feedback. Early results from his group were presented at several meeting and conference venues (see below), showing an altitude (hence temperature) dependence on ice crystal structure. We are gearing up to process one full year of MISR+MODIS data on BW during the next few weeks, which will likely be followed by a mission-scale processing at the end of this year. This will provide the first regional to global view of the distribution of cirrus ice crystal morphology.

The TDF project has many big data science issues that are common throughout the sciences and that are part of larger discussion underway by the National Data Services consortium. As such, the TDF project has been identified as a use case under development for the National Data Services (https://nationaldataservice.atlassian.net/wiki/pages/viewpage.action?pageId=4358159).

In addition to the satellite data processing and analysis work described above, we also continued to work on bringing the 3-D radiative transfer models and research to publication. Model development and research were carried out by former Blue Waters Graduate Fellow Alexandra Jones, the results of which were described in last years BW report. Dr. Jones has since moved to Princeton University as a postdoc, and we have been working on two manuscripts for publication. The first manuscript is near complete and should be submitted sometime within the next month. One of the models is the first spectrally integrating, Monte Carlo 3D radiative transfer model that includes both internal and external sources, and accounts for absorption, emission, and scattering by the atmosphere, clouds, and the surface. This required the development of large databases of atmospheric optical properties. For example, the optical properties for a particular gamma distribution of liquid water droplets at approximately 250 Hz spectral resolution was created. This database requires about 5TB of storage and is the most comprehensive database of its kind. An unanticipated benefit of developing this database is that it can be mined to provide updated parameterizations of optical properties at lower spectral resolution; the most popular of the sort haven't been updated in almost 30 years. The accuracy of our RT model has been verified with extensive comparison to analytical solutions and results from the world's most advanced 1D Line-by-Line RT model. Our model is now ready to act as the first 3D broadband standard of accuracy for comparison by other RT models that make simplifying assumptions. It will be released for public use and development as a community model upon publication (see Plan for Next Year). With Dr. Jones new position and duties at Princeton, progress in finalizing these manuscripts have been slower than anticipated, and with much lower usage of node hours than planned, but this is only because of the demand on Dr. Jones time in her new position rather than any issues with the models.

Publications and Presentations Associated this this Work

Zhao, G., L. Di Girolamo, D. J. Diner, C.J. Bruegge, K. Muller, and D.L. Wu (2016), Regional changes in Earth's color and texture as observed from space over a 15-year period. *IEEE Trans. Geosci. Remote*

Sens., **54**(7), 4240-4249, doi:10.1109/TGRS.2016.2538723.

Di Girolamo, L., G. Zhao, J. Towns, S. Wang, Y. Liu, and K. Yang, 2016: The Terra Data Fusion Project. *2016 Blue Waters Annual Report*, University of Illinois Press, Urbana, IL

Di Girolamo, L,, L. Liang, and W. Sun, 2016: Bias in MODIS cloud drop effective radius for oceanic water clouds as deduced from measured cloud optical thickness variability across scattering angles. *MISR Data Users Symposium*, February, Pasadena, CA.

Hioki, S., P. Yang, and L. Di Girolamo, 2016: Retrieval of surface roughness of ice crystals in cirrus clouds from MISR observations. *MISR Data Users Symposium*, February, Pasadena, CA.

Zhao, G., *et al*., 2016a: Regional changes in Earth's color and texture as observed from space over a 15-year period. *MISR Data Users Symposium*, February, Pasadena, CA.

Zhao, G., *et al*., 2016b: Regional changes in Earth's color and texture as observed from space over a 15-year period. *School of Earth, Society and Environment Research Review*, February, Urbana, IL.

Di Girolamo, L., M. Su, L. Liang, and W. Sun, 2016: Bias in MODIS cloud drop effective radius for oceanic water clouds as deduced from measured cloud optical thickness variability across scattering angles. *International Radiation Symposium, April 2016, Auckland, New Zealand.*

Di Girolamo, L., 2016: The Terra Data Fusion Project. *NCSA 2015-2016 Faculty Fellows Colloquium*, May, Urbana, IL.

Fu, D., L. Di Girolamo, L. Liang, and G. Zhao, 2016: The effect of cloud heterogeneity on microphysical properties through MISR-MODIS data fusion. *MISR Data Users Symposium*, December, Pasadena, CA.

Hioki, S., P. Yang, A. Bell, Y. Wang, and L. Di Girolamo, 2016 Retrieving degree of ice particle surface roughness in clouds from multi-angle measurements. *MISR Data Users Symposium*, December, Pasadena, CA.

Di Girolamo, L., *et al*., 2016: The Terra data fusion project. *MISR Data Users Symposium*, December, Pasadena, CA.

Di Girolamo, L., G. Zhao*, et al.*, 2016: Regional changes in Earth's color and texture as observed from space over a 15-year period. *MISR Data Users Symposium*, December, Pasadena, CA.

Bell, A., S. Hioki, Y. Wang, P. Yang, and L. Di Girolamo, 2016: Sensitivity analysis of observed reflectivity to ice particle surface roughness using MISR satellite observations. *American Geophysical Union 2016 Fall Meeting*, December 12-16, San Francisco, CA.

Zhao, G., *et al*., 2017: Regional changes in Earth's color and texture as observed from space over a 15-year period. *American Meteorological Society 97th Annual Meeting*, January 22-26, Seattle, WA.

Wang, Y., S. Hioki, A. Bell, P. Yang, and L. Di Girolamo, 2017: Assessing the effect of satellite viewing geometry on retrieved ice cloud particle surface roughness using MISR satellite observations. *American Meteorological Society 97th Annual Meeting*, January 22-26, Seattle, WA.

Plan for Next Year:

Our work for the upcoming year extends much of the work we did this past year. The Terra Data Fusion (TDF) project is fully supported under NASA Grant Number NNX16AM07A, with additional cloud product R&D supported under two other NASA-sponsored projects: NNX14AJ27G for the cloud drop effective radius of liquid water clouds and NNX15AQ25G for ice crystal roughness for cirrus clouds. In all cases, allocation on Blue Waters within Di Girolamo's current Blue Waters Professorship allocation was defined. In addition, the OVCR has committed 2 PB of storage on Nearline for the TDF project. We have also launched an educational initiative that processes the Terra data in a manner that is both informative and appealing to a broader public (like Figure 1) who are interested in Earth sciences. For the upcoming year, we anticipate 130,000 node hours on Blue Waters for the processing and analysis of the Terra data over all these projects that total 20 mission-scale processing activities. This is based on our experience in current processing of the Terra data, where we estimate 2000 to 32,000 node hours per activity, depending on activity.

The spectrally integrating, atmospheric Monte Carlo 3D radiative transfer model, and its sister monochromatic version, will be made available to the public. These models have been developed in an object-oriented style, meant to allow for further community development. Before public release later this year, we will work on developing the needed peer-reviewed publications and finalize user and developer documentation for these models. Some work to improve the memory utilization and fine tune the performance of the model at scale on Blue Waters will be carried out. Optimization of these codes on Blue Waters is essential since the intent is to make them available to the broader community, some of whom may wish to carry out their research with these codes on Blue Waters. The models will then be utilized to provide highly accurate standards of comparison for other radiative transfer models. This contribution will allow for inclusion of parameterizations of 3D effects in the simpler, less expensive, more commonly used radiative transfer models that are either coupled to climate and weather models, or used to invert satellite observations, for example. Additional experiments, highlighting the bias in satellite products due to 3D effects will also be conducted. Based on our experience with these models over the past couple of years and our planned experiments to finalize the two papers, we anticipate approximately 50,000 node hours to bring this work to completion.

We therefore request 180,000 node hours for next year. We expect the usage break down by quarter to be the following:

Q1: 20% Q2: 15% Q3: 35% Q4: 30%

The storage requirement for the 3D radiative transfer modeling work is not anticipated to be large. Tables of scattering properties will need to be retained for each unique atmospheric domain, however total storage for those tables and the corresponding output should not exceed 50 TB. The model requires only two input files and produces one output file, so there is no anticipated taxing of the file system expected due to large numbers of files. The radiative transfer model is comprised mainly of logical operations to determine the fate of the bundle of light, i.e. comparisons of random numbers to cumulative distribution functions and simple arithmetic calculations to tally the contribution of each bundle as it travels through the domain. Memory usage will depend on domain size.

The storage requirement for the Terra work will be large. At the moment, we have stored ~0.75 PB of Terra data on Nearline and on ROGER (some of the Terra data needed to be transferred to ROGER before moved to Nearline owing to how NASA has set things up), and another ~0.25 PB is currently being transferred and anticipated to be completed within the next 2 months. The compressed Basic Fusion files will take up an additional 1 PB. The Basic Fusion files will replace all original Terra data, so there will be no need to keep the original Terra files once the Basic Fusion files are verified and complete. The various mission-scale products that will be derived from the Basic Fusion files are also projected to be about 1 PB compressed. Therefore, with proper data management, we anticipate that our current allocation of 2 PB of Nearline storage will be sufficient for this year.