Blue Waters Professor Allocation Annual Report: February 1, 2018 – January 31, 2019

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): Shashank Bansal, Michelle Butler, Matias Carrasco Kind, Brandon Chen, Landon Clipp, Soumi Dutta, Dongwei Fu, Yizhao Gao, Yan Liu, Yulan Hong, Jonathan Kim, Jesse Loveridge, Arka Mitra, Don Petravick, David Raila, Puja Roy, Sean Stevens, John Towns, Javier Villegas Bravo, Shaowen Wang, Yizhe Zhan, Guangyu Zhao, *University of Illinois at Urbana-Champaign*; Mike Garay, *Jet Propulsion Laboratory*; Alexandra Jones, *Princeton University*; Ralph Kahn, James Limbacher, and Scott Sinno, *NASA Goddard Space Flight Center*; Lusheng Liang, *SSAI and NASA Langley Research Center*; Hyohyung Lee and Kent Yang, *HDF Group*; and Souichiro Hioki, Yi Wang, 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 Blue Waters, satellite observations and 3D radiative transfer has had an incredible year of success: (1) dissemination of research in 3 peer-reviewed publications and 16 conferences/meetings/seminars; (2) extensive progress on the Terra Data Fusion project with continued support from NASA and Illinois; (3) continued Terra data analysis for meteorological studies in SE Asia as part of NASA's Cloud and Aerosol Monsoonal Processes Philippines Experiment (CAMP2Ex) field campaign; (4) continued enhancements to our understanding of global microphysical properties of water and ice clouds; and (5) continued development of training data from Terra to support ML (CNN) cloud detection algorithm development for NASA's MAIA mission. Our initial request of 210K NH was on target, but was underutilized on Blue Waters because I/O limitations between Nearline and Scratch forced a reduced scope in our research.

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 70% 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 physical, chemical, and biological 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, in part, to solve these problems.

The TDF project has transferred years 2000-2015 of the Terra L1B data (~ 1.1 PB) from the three NASA data centers to Blue Waters, and the project has successfully generated 2.4 PB of Terra Basic Fusion files (currently residing on Nearline). The Terra Basic Fusion files provides the necessary stepping-stone for developing higher-level products and provides the framework for other flavors of fusion. We have compiled a draft version of the Terra Basic Fusion Algorithm Theoretical Basis and Data Specification document temporally posted for download at:

https://1drv.ms/w/s!Agotov0_Ayi7jBBH80y0gJG36_q7

Benchmarking 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. Data transfer between Nearline and Scratch has been by far the largest bottleneck, so much so that we needed to greatly reduce the scope of our research. More recently, issues with long waits in the queue hasn't helped.

The data format for the Terra Basic Fusion product is hdf5 with a file structure constructed to comply with Climate and Forecast(CF) conventions, which follows the netCDF-4 data model enabling NetCDF4 tools as well as HDF5 tools to access and explore file contents. All of the granule-level metadata for the entire mission complying perfectly with the NASA CMR format have been generated and stored on the BW nearline system and will be ingested into the NASA Earthdata system for a broad user community to search the Terra Basic Fusion product.

We have been working on three scientific use cases that help sharpen and push the TDF project forward: (1) monitoring climate change from Terra, (2) retrieving liquid water cloud drop size distribution, (3) retrieving cirrus cloud ice crystal shapes, and (4) studies of Arctic albedos. The first in a series of planned studies to monitor climate change from Terra was reported in earlier BW reports and in Zhao *et al*. (2016). The latest addition includes an examination of three of the Terra instruments (MISR, MODIS and CERES) to examine and explain region trends in the observed Terra instrument radiances, which also required a detailed examination of radiometric stability for these instruments.

We discovered a serious issue with MODIS calibration that appears in the data after 2014. We reported these results at the AGU Fall Meeting in DC (Zhan et al. 2018).

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. Our approach builds upon the approach we showed in Liang et al. (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 M.S. graduate student Dongwei Fu, which he has deposited in November. He has painted a new view of the global distribution of cloud drop sizes that are now in line with spot measurements had from field campaigns. He has presented these results at several venues (see presentation list below) and he is currently writing up the results for a peer-reviewed publication.

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. Results from his group were presented at several meeting and conference venues (see below), showing latitudinal dependences in ice crystal roughness not never observed before. This identifies that the need for improved treatment of ice cloud optical properties both within our climate models, as well as in many standard remote sensing algorithms. Our first paper on these results has been published in Wang et al. (2018).

On the forth use case, we have been examined the uncertainties in Arctic albedos under cloudy conditions derived from MISR and CERES instruments. The work is supported under JPL contract 147871 and the NSF Division of Polar Programs under contract 16-03544. Bringing the two instruments together allowed us to quantify and decouple errors caused by narrow-to-broadband conversions from errors caused by errors in the angular distribution models used in the radiance-to-irradiance conversion, showing differences in the errors caused by the angular distribution models to be less than 4% when the solar zenith angle was < 70°. However, for solar zenith angles > 80° errors reached 13.5%, indicating a strong need to improve upon our knowledge of the angular anisotropic scattering by cloud fields at very high solar zenith angles. The results are now published in Zhan et al. (2018).

Blue Waters, the Terra data, and other satellite datasets are also being used in analysis supporting the meteorological studies in SE Asia as part of the Cloud and Aerosol Monsoonal Processes Philippines Experiment (CAMP2Ex) field campaign, with NASA assets being deployed from the Philippines starting in August 2019 (delayed from the original deployment of July 2018 owing to mechanical problems with one of the aircrafts). This research is being carried out under NASA contract 80NSSC18K0144. Currently, postdoc Yulan Hong and student Soumi Dutta have been examining MISR and MODIS data from Terra to ascertain cloud characteristics for the study region, with a focus on cloud fraction and phase. Results were presented at the AGU Fall 2018 meeting in DC (see presentations below).

We are also using BW and the Terra data to support the development and testing of the cloud detection algorithm for NASA's upcoming Multi-Angle Imager for Aerosol (MAIA) mission (lunches in 2021) under JPL contract 1586704. I'm a Co-I on this mission, whose measurements will be combined with population health records to better understand the connections between aerosol pollutants and health problems such as adverse birth outcomes, cardiovascular and respiratory diseases, and premature deaths. Cloud detection issues will be a major source of error for this mission, as well as other research listed above using Terra data. Postdoc Yizhe Zhan is prototyping a new deep learning approach using TensorFlow on BW, with assistant from Aaron Saxton in NCSA, and a new graduate student, Javier Villegas Bravo, has been assigned to this project. Javier has been using the Terra data as proxy data for MAIA in developing the training set needed for the MAIA cloud detection algorithm.

In addition to the satellite data processing and analysis work described above, we also continued to work on advancing our 3-D radiative transfer models and research involving these new models on BW. Our first 3D radiative transfer model developed and benchmarked on BW has been published in (Jones and Di Girolamo 2018). With my collaborative grant with colleagues at Technion – Israel Institute of

Technology – under the Binational Science Foundation (BSF grant 2016325) we are using the 3D radiative transfer model in solving new tomographic approaches in retrieving cloud properties from multi-angle measurements, not only for use on Terra, but also for an upcoming mission selected last month by ESA for launch in 2023 called CloudCT (PI Prof. Yoav Scheschner, Technion).

Finally, we have been working on data transfer issues between BW and our collaborators at NASA's Goddard Space Flight Center, which does not support Globus. The Blue Waters network team has worked with NASA to resolve data transfer issues at NASA's end.

Publications and Presentations Associated this this Work

Peer-Reviewed Journals and Dissertations

Jones, A.L., and L. Di Girolamo, 2018: Design and verification of a new monochromatic thermal emission component of the I3RC Community Monte Carlo Model. *J. Atmos. Sci.,* **75**, 885-906*,* doi: 10.1175/JAS-D-17-0251.1

Zhan, Y., L. Di Girolamo, R. Davies, and C. Moroney, 2018: Instantaneous top-of-atmosphere albedo comparison between CERES and MISR over the Arctic. *Remote Sens.* **10(12)**, 1882, doi:10.3390/rs10121882.

Wang, Y., S. Hioki, P. Yang, M. King, L. Di Girolamo, D. Fu, and B.A. Baum, 2018: Inference of an optimal ice particle model through latitudinal analysis of MISR and MODIS data. *Remote Sens.* **10(12)**, 1981, doi:10.3390/rs10121981.

Fu, D., 2018: *Examination of the behavior of MODIS-retrieved cloud droplet effective radius through MISR-MODIS data fusion*. M.S. dissertation, University of Illinois, 65 pp.

Conference, meeting, and seminar presentations

Di Girolamo, L., et al., 2018: The Terra data fusion project: an update. *MISR Data Users Symposium*, February, Pasadena, CA.

Fu, D., L. Di Girolamo, L. Liang, and G. Zhao, 2018: The observed behavior of the bias in MODIS-retrieved cloud drop effective radius through MISR-MODIS data fusion. *MISR Data Users Symposium*, February, Pasadena, CA.

Dutta, S., S. Dey, and L. Di Girolamo, 2018: Observing the differences in cloud cover between MISR RCCM resolution-corrected and uncorrected versions. *MISR Data Users Symposium*, February, Pasadena, CA.

Yang, M., L. Clipp, H.-K. Lee, G. Zhao, and L. Di Girolamo, 2018: Generation of 16-years large-scale Terra Level 1B data fusion products. *NASA Earth Data Systems Working Group annual meeting*, April, Annapolis, MD.

Lee, H.-K., M. Yang, Y.L. Yo, G. Zhao, and L. Di Girolamo, 2018: Terra fusion metadata generation. NASA Earth Data Systems Working Group annual meeting, April, Annapolis, MD.

Di Girolamo, L., et al., 2018: The Terra data fusion project: an update. *Asia Oceania Geosciences Society*, June 3-8, Honolulu, HI.

Fu, D., L. Di **Girolamo**, L. Liang, and G. Zhao, 2018: Estimating the regional bias of MODIS-retrieved cloud droplet effective radius through MISR-MODIS data fusion. *AMS 15th Conf. Atmos. Rad*., 9 – 13 July, Vancouver, BC.

Jones, A.L., and L. Di Girolamo, 2018: Benchmark quality open source models for spectral or broadband 3D radiative transfer in cloudy atmospheres. *AMS 15th Conf. Atmos. Rad*., 9 – 13 July, Vancouver, BC.

Di Girolamo, L., Y. Zhan, G. Zhao, and Y. Hong, 2018: Cloud masking for MAIA. *MAIA Science Team Meeting*, Aug 7 – 8, Pasadena, CA.

Di Girolamo, L., Y. Zhan, and G. Zhao, 2018: MAIA Cloud Mask (MCM): Testing and Validation. *MAIA Science Team Meeting*, Aug 7 – 8, Pasadena, CA.

Di Girolamo, L., 2018: Perspectives on Terra data fusion for improved Earth science products and knowledge. *NASA Goddard Space Flight Center*, Oct 12, Greenbelt, MD. ***(invited)***

Fu, D., L. Di **Girolamo**, L. Liang, and G. Zhao, 2018: Estimating the regional bias of MODIS-retrieved cloud droplet effective radius through MISR-MODIS data fusion. *MODIS and VIIRS Science Team Meeting*, Oct 15 – 18, Silver Spring, MD.

Yang, M., L. Clipp, H.-K. Lee, G. Zhao, and L. Di Girolamo, 2018: Generation of 16-years large-scale Terra Level 1B data fusion products. *MODIS and VIIRS Science Team Meeting*, Oct 15 – 18, Silver Spring, MD.

Di Girolamo, L., 2018: Current data and computing challenges in addressing NASA's Earth science questions: Perspectives from the ACCESS to Terra Fusion Project. *NASA Science Mission Directorate Workshop on Maximizing the Scientific Return on NASA Data*, Oct. 30 – 31, Washington DC. ***(invited)***

Zhan, Y., L. Di Girolamo, and G. Zhao, 2018: Earth's spectral and textural trends as observed from Terra. *American Geophysical Union Fall Meeting*, Dec. 9 – 14, Washington, DC.

Hong, Y., L. Di Girolamo, and J.S. Reid, 2018: An investigation of aerosol, warm and mixed phase clouds overlapped by cirrus over the Southeast Asia Region. *American Geophysical Union Fall Meeting*, Dec. 9 – 14, Washington, DC.

Dutta, S., L. Di Girolamo, S. Dey, and Y. Zhan, 2018: Assessment of newly built resolution-corrected MISR cloud fraction product. *American Geophysical Union Fall Meeting*, Dec. 9 – 14, Washington, DC.

Di Girolamo, L., et al., 2018: The Terra data fusion project. *American Geophysical Union Fall Meeting*, Dec. 9 – 14, Washington, DC.

Plan for Next Year:

Our work for the upcoming year extends much of the work described above. The Terra Data Fusion (TDF) project is fully supported under NASA Grant Number NNX16AM07A, with additional R&D support under other NASA, JPL, BSF and NSF sponsored projects under grant numbers listed above. 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, and the NCSA director was gracious to commit an additional 2 PB. We are currently working with NCSA and NASA for the long-term archiving and access to the valuable data created under the TDF project. For the upcoming year, we anticipate 117,000 node hours on Blue Waters for the processing and analysis of the Terra data over all these projects that total 18 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.

We will be carrying out numerous 3D radiative transfer simulations to address specific problems in cloud tomography that associated with the CloudCT mission, as well as additional experiments designed to quantify the bias in satellite products due to 3D effects. Based on our experience with these models over the past few years and our planned experiments, we anticipate approximately 70,000 node hours to bring this work to completion.

We therefore request 187,000 node hours for next year. We expect the usage break down by quarter to be the following:

Q1: 25% Q2: 35% Q3: 25% Q4: 25%

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, the compressed Basic Fusion files will take up an additional 2.4 PB. We continue to retain a copy of the original Terra data on nearline, which amounts to 1.1 PB. The various mission-scale products we derive and ancillary products needed to derive them is about 0.5 PB. Therefore, with proper data management, we anticipate that our current allocation of 4 PB of Nearline storage will be sufficient for this year.