

Major goals of the project:

Machine learning algorithms will be used to analyze the simulation data in situ and perform anomaly detection to determine whether local dynamics are exhibiting unusual behavior. Data saves would be triggered if behavior in a local region deviated from previously recorded dynamics. Therefore, at the beginning of the simulation, data from the full field would be saved, since no dynamics would have been previously recorded. As the simulation evolved, data would be recorded as new dynamics were exhibited. It is noteworthy that these dynamics might not be anomalous from a physical standpoint: e.g., an extreme weather event in a climate simulation. These events are anomalous in the sense that they differ from previously recorded dynamics. While anomaly detection algorithms have been extensively developed for other applications, our research would focus on the following two algorithmic innovations:

- 1) Scalability for in situ implementation on HPC platforms and
- 2) Generalization across multiple scientific domains without problem-specific hand-tuning.

Accomplishments under these goals:

In an extensive collaboration with Sandia National Laboratory, an initial algorithm for detecting events of interest based on feature importance metrics has been tested. For this work, the features are state variables calculated at each grid point. On each processor, we used Kernel Density Estimation to estimate the probability density function of the state variables at a given time step. We then used these density estimates to train an ensemble of decision trees. The feature importance metrics of this ensemble were compared to those at previous time steps and on other processors to detect events of interest (i.e., a tropical cyclone). The communication requirements for this technique scales with the number of features, and in most scientific computing applications relevant to climate science, the number of feature is in the tens requiring minimal communication. Furthermore, it can be deployed naturally in an online fashion.

In particular, the Community Atmosphere Model version 5 (CAM5), routinely used for climate change projections, was configured to simulate a tropical cyclone over a short simulation period (i.e., 7.5 day) with data recorded at 12-hour intervals. The model grid was uniform in longitude and latitude, and used hybrid-pressure coordinates in the vertical, with 512, 257, and 30 points in each of those directions, respectively. The simulation tracked the development of an idealized tropical cyclone in one region of the global domain. The idealized test initialization was develop previously by PI Kevin Reed, which is described in Reed and Jablonowski (2011), is spun-up from an initially weak vortex in a favorable environment for tropical cyclone development. Seven standard (though not necessary optimal from a detection standpoint) state variables were saved during the simulation, including the vorticity, the temperature, the zonal velocity, the meridional velocity, the specific humidity, the cloud ice concentration and the cloud liquid water concentration. In this idealized test the event detection metric for spatial

locality is able to successfully detect the presence of the tropical cyclone of interest. However, the temporal event detection metric does not detect any event of interest at any time step, likely due to the gradual strengthening of the tropical cyclone in the idealized simulation.

Additional idealized frameworks were used to implement algorithms into the CAM5 modeling framework, including the use of radiative convective equilibrium as in Chavas et al. (2017).

References:

Reed, K. A. and C. Jablonowski (2011a), **An analytic vortex initialization technique for idealized tropical cyclone studies in AGCMs**, *Mon. Wea. Rev.*, 139, 689-710, doi:10.1175/2010MWR3488.1.

Chavas, D. R., K. A. Reed and J. A. Knaff (2017), **Physical understanding of the tropical cyclone wind-pressure relationship**, *Nature Communications*, 8, 1360, doi: 10.1038/s41467-017-01546-9.

Opportunities for training and professional development provided:

This project has provided partial support for Xiaoning Wu and John Landy, graduate student researchers in the School of Marine and Atmospheric Sciences at Stony Brook University during the academic year. The project also provided partial summer support for Elizabeth Najman a first-year graduate student researcher in the School of Marine and Atmospheric Sciences at Stony Brook University. Therefore, the project is helping to train and develop a next generation researchers in the practical needs of high performance computing and data science, and able to communicate across disciplinary and professional boundaries. Students have attended the Community Earth System Model Tutorial to get hands on work with the CAM5 framework.

Dissemination of results to communities of interest:

Initial results of this project, as described above, have been disseminated to the community through a submission of a conference paper, and subsequent presentation, at 2017 IEEE 7th Symposium on Large Data Analysis and Visualization (LDAV). More detail of the presentation is provided here:

Ling, J., W. P. Kegelmeyer, K. Aditya, H. Kolla, K. A. Reed, T. M. Shedd and W. L. Davis (2017), **Using Feature Importance Metrics to Detect Events of Interest in Scientific Computing Applications**, *2017 IEEE 7th Symposium on Large Data Analysis and Visualization (LDAV)*, October 2, 2017.

Related to this work, PI Reed also presented at two conference related to tropical cyclones:

Reed, K. R., C. M. Zarzycki, M. F. Wehner, J. J. A. Huff and J. T. Bacmeister, **Investigating potential changes in tropical cyclone overland precipitation in the U.S. due to future climate change** (Apr. 16, 2018), Oral presentation at the 33rd AMS Conference on Hurricanes and Tropical Meteorology, Ponte Vedra Beach, FL, USA, April 16-20, 2018.

Reed, K. R., X. Wu, M. F. Wehner and J. T. Bacmeister, **Exploring the Use of High-Resolution CAM5 for Basin-Scale Projections of Tropical Cyclone Activity** (Jun. 5, 2018), Oral presentation at the 15th Annual Meeting of the Asia Oceania Geosciences Society, Honolulu, HI, USA, June 3-8, 2018.

Reed, K. A., A. R. Herrington, B. Medeiros and D. R. Chavas, **Reduced Complexity Frameworks for Exploring Resolution Dependence in CAM** (Jun. 18, 2018), Poster presentation at the 23rd Annual CESM Workshop, Boulder, CO, USA, June 18-20, 2018.

Davis, W. L., K. A. Reed, T. Shead, H. Kolla, G. Popoola, P. Kegelmeyer and A. Konduri, **The Potential of Integrated Machine Learning Algorithms for Tropical Cyclone Detection in Advanced Climate Modeling**, Poster presentation at the AGU Fall Meeting 2019, San Francisco, USA, December 9-13, 2019.

In addition, PI Reed attended the ASCR PI meeting in August 2018 to present results of the project to the PI community.

Finally, two additional scientific publications resulted, in part, from this work:

Chavas, D. R., and K. A. Reed (2019), **Dynamical aquaplanet experiments with uniform thermal forcing: system dynamics and implications for tropical cyclone genesis and size**, *J. Atmos. Sci.*, 76, 2257-2274, doi: [10.1175/JAS-D-19-0001.1](https://doi.org/10.1175/JAS-D-19-0001.1).

Reed, K. A., J. T. Bacmeister, J. J. A. Huff, X. Wu, S. C. Bates and N. A. Rosenbloom (2019), **Exploring the impact of dust on North Atlantic hurricanes in a high-resolution climate model**, *Geophys. Res. Lett.*, 46, 1105–1112, doi: [10.1029/2018GL080642](https://doi.org/10.1029/2018GL080642).