Enabling Automated Graph-based Search for the Identification and Characterization of Mesoscale Convective Complexes in Satellite Datasets through Integration with the Apache Open Climate Workbench

Lewis J. McGibbney Ph.D[1], Kim D. Whitehall [1], Chris A. Mattmann [1], Cameron E. Goodale [1], Michael J. Joyce[1], Paul. M. Ramirez[1], Paul Zimdars[1]

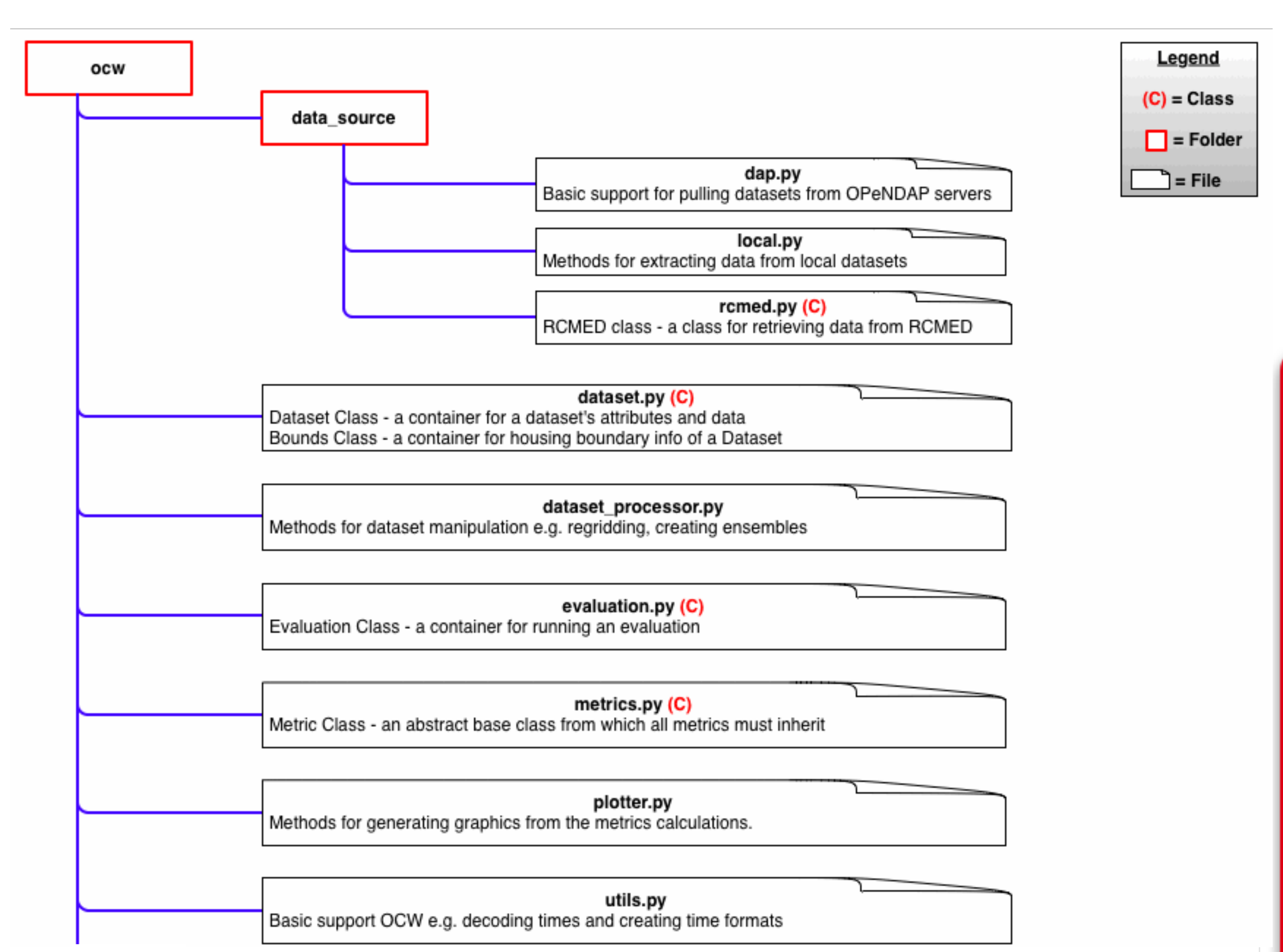
[1] NASA Jet Propulsion Laboratory/California Institute of Technology, Pasadena, CA, United States

Abstract

Modeling climate and Earth system processes on regional scales is essential for projecting the impacts of climate change on society and our natural resources. This research provides an enabling automated graph-based search tool suited to the identification and characterization of Mesoscale Convective Complexes (MCCs) in satellite datasets through integration with the Apache Open Climate Workbench (OCW).

Introduction

Apache OCW [0] (recently open sourced under the Apache Software v2.0 License by NASA JPL) is an an effort to develop software that performs climate model evaluation using model outputs from a variety of different sources (the Earth System Grid Federation, the Coordinated Regional Downscaling Experiment, the U.S. National Climate Assessment and the North American Regional Climate Change Assessment Program) and temporal/spatial scales with remote sensing data from NASA, NOAA and other agencies. The toolkit includes capabilities for rebinning, metrics computation and visualization.



OCW User Case

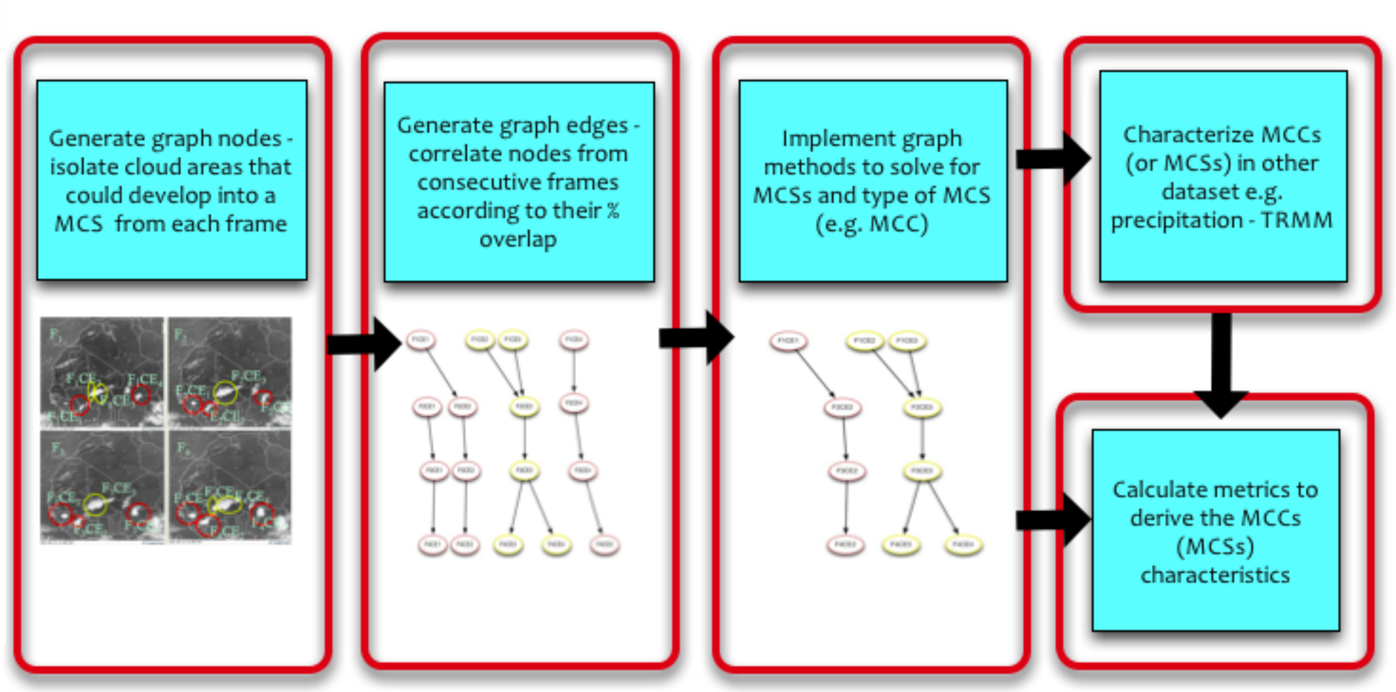
Why MCCs?

MCCs are convectively-driven weather systems with a average duration of 10 to 24 hours that contribute large amounts to daily and monthly rainfall totals within the areas they are located. They are identified (mostly through manually methods) from hourly infrared satellite datasets. More than 400 MCCs occur annually over various locations on the globe. The convective nature of MCCs raises questions regarding their impacts on local and global heat and water budgets, their seasonal variability and frequency in current and future climates, amongst others. However, in spite of the formal observation criteria of these features in 1980, these questions have remained comprehensively unanswered because of the untimely and subjective methods for identifying and characterizing MCCs due to limitations data-handling limitations.

The main outcome of this work therefore documents how the graph-based search algorithm, referred to as the Grab 'em, Tag 'em, Graph 'em" (GTG) method (Whitehall et al. [3] was implemented on top of the OCW stack with the ultimate goal of improving fully automated end-to-end identification and characterization of MCCs in high resolution observational datasets.

The GTG

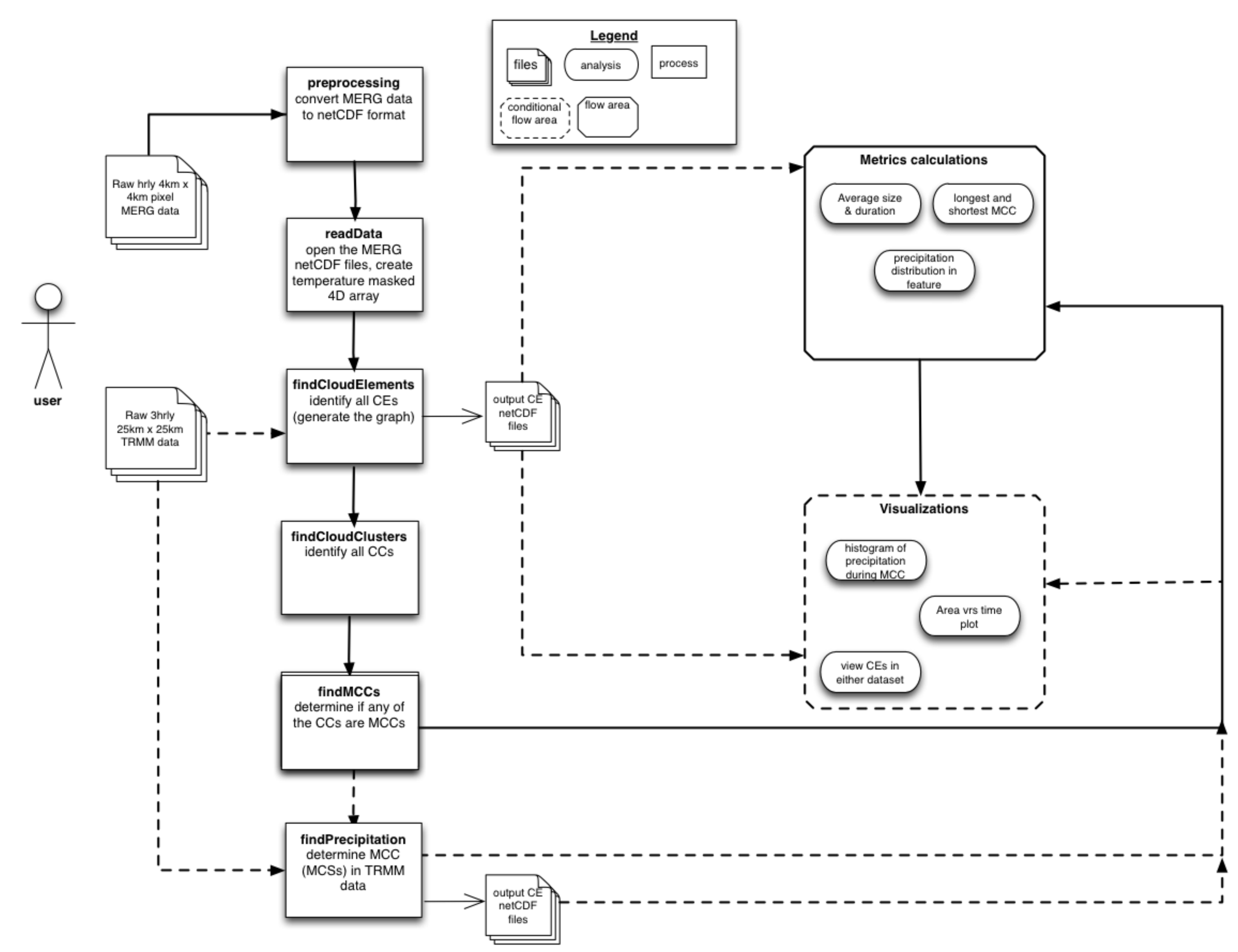
The GTG identifies and characterizes in a three-step process that involves the analysis of spatio-temporal characteristics within infrared satellite datasets and gridded precipitation datasets.



The implementation is modular and utilizes the OCW stack philosophy of data manipulation and

GTG Implementation

The implementation is modular and utilizes the OCW stack philosophy of data manipulation.



Data within the GTG implementation follow the dimension order of OCW datasets, although to date the data is not loaded into OCW datasets. As such functionality pertaining to dataset processing, including opening files, extracting data and regridding datasets was leveraged from the OCW dataset processor. Visualization and metric capabilities in the GTG were mirrored off the OCW stack.

Lessons learnt

Leveraging OCW allowed for:

1. fast implementation of the GTG, which out having to worry about unrelated tasks concerning data manipulation e.g. regridding.
2. Quick metric calculations and visualization of raw outputs.
3. Development of a basis for similar implementations for different scientific use cases.

Future work

1. Fully integrate the OCW stack into the MCC search e.g. using OCW Datasets.
2. Push the functionality of using sub-daily data in evaluations back to the OCW stack
3. Push graph visualizations to the OCW stack.

Conclusion

Apache OCW as an open source project was demonstrated from inception and we display how it was again utilized to advance understanding and knowledge within the climate domain. The project was born out of refactored code donated by NASA JPL from the Earth science community's Regional Climate Model Evaluation System (RCMES) [1], a joint project between the Joint Institute for Regional Earth System Science and Engineering (JIFRESSE), and a scientific collaboration between the University of California at Los Angeles (UCLA) and NASA JPL. The Apache OCW project was then integrated back into the donor code with the aim of more efficiently powering that project. Notwithstanding, the object-oriented approach to creating a core set of libraries Apache OCW has scaled the usability of the project beyond climate model evaluation as displayed in the MCC use case detailed herewith.

Resources

[0] Apache OCW – http://climate.apache.org

[1] Regional Climate Modeling and Evaluation System – http://rcmes.jpl.nasa.gov

[2] Whitehall, K.*et al.* Exploring a graph theory based algorithm for automated identification and characterization of large mesoscale convective systems in satellite datasets. *Earth Science Informatics*, 1-13

Acknowledgements

The authors would like to sincerely acknowledge support from the Regional Climate Modeling and Evaluation System (RCMES) project at NASA JPL. We would like to thank Dr. Kim D. Whitehall for having the vision to contribute her Ph.D work to the open source community. Additionally we would like to acknowledge the Apache OCW community who develop the OCW software for the benefit of the climate scientists, developers and users alike.

Contacts

Please direct any correspondence or feedback to Dr. Kim D. Whitehall – kim.d.whitehall@jpl.nasa.gov