



Alternative Earth Science Datasets For Identifying Patterns and Events

Kaylin Bugbee (1), Robert Griffin (1), Brian Freitag (1), Jeffrey Miller (1), Rahul Ramachandran (2), and Jia Zhang (3)

(1) University of Alabama in Huntsville, Huntsville, United States (kaylin.m.bugbee@nasa.gov), (2) NASA Marshall Space Flight Center, (3) Carnegie Mellon University

The availability of Earth observation data has grown significantly since the first satellites were launched in the 1960s and is expected to grow exponentially in the future with the launch of more powerful and accurate sensors. In light of this data deluge, new scientific and technological methods are needed in order to better understand and analyze Earth observation data.

Alternative data sources offer a solution to more accurately and quickly identify interesting scientific events within these large volume Earth observation datasets. Alternative data sources are data which are extracted or generated from non-traditional sources and can include social media data, point of sale transactions and product reviews. While alternative data was originally used in the investment world, there are other non-traditional, domain specific data sources that can be similarly leveraged within the Earth sciences. These data sources include, but are not limited to, the information found in numerous unstructured text documents such as flight reports for airborne field campaigns, agricultural reports and weather forecast discussions. Data extracted from these documents can be used to generate datasets that can be analyzed for spatial, temporal and climatological patterns in order to more effectively identify interesting events or trends. Events and patterns identified in these alternative data sources assist the geosciences community in more efficiently identifying interesting events or use cases and can also help decision makers better understand reporting of anticipated hazards and disasters. These data can in turn be leveraged to build an event database that will help the remote sensing community more effectively discover and use Earth observation data for research and for labelling data for machine learning applications.

This presentation will describe the creation of an alternative Earth science dataset from the U.S. National Weather Service's Area Forecast Discussion (AFD) documents. The AFD documents describe the most significant weather issues facing each of the National Weather Service's 122 weather forecast offices. The AFDs are written approximately every six hours, making these documents an optimal choice for identifying interesting atmospheric events or use cases. We will also describe the results of an exploratory use case that leverages the AFD dataset to investigate methods for identifying scientifically relevant events and will also provide a motivating example for the use of these data.