WeatherCitizen: A Software Suite for the Collection, Assimilation, and Distribution of Traditional and Crowd Sourced Environmental Observations

Marc Shapiro (presenter), Jerry Bieszczad, David Callender
Creare (http://www.creare.com/)

Tuesday, 8 January 2019: 8:30 AM-10:00 AM

Crowdsourcing measurements and observations of environmental conditions using mobile devices promises rich, widespread, valuable data. Collection, assimilation, and distribution of such real-time observations from a broadly distributed and spatially dense network of mobile devices would lead to: improved awareness of current environmental conditions; data for assimilation into numerical weather forecasts; and increased access to data for researchers and forecasters. However, collection and management of these observations often require overcoming technical hurtles and result in low quality data with high analytical overhead. Assembling traditional and crowd-sourced environmental observations should be modular and intuitive, resulting in an approachable, transparent data products that can be audited, analyzed, and presented by researchers, forecasters, and citizens.

This presentation will discuss development and deployment of a modular full stack software suite (*WeatherCitizen*) for collecting, assimilating, serving and harmonizing traditional and crowd-sourced environmental observations. The goal of the system is to reduce the barrier for data producers to collect relevant, quality controlled environmental observations and for data consumers to create accurate, translatable insights in near real time. The core modules of the stack include: (1) a cross platform mobile application for collecting manual and automated crowd-sourced observations, (2) a geospatially aware cloud database for aggregating, filtering, and curating crowd-sourced observations alongside other traditional data sources, (3) a production ready application programming interface (API) for querying and serving data to researchers and third party software, (4) a web map interface for visualizing collected data, and (5) a set of open source tools for downloading, harmonizing, and analyzing collected data using MATLAB, Python, and commercial GIS programs.

We will present the initial deployment of this system and describe several approaches for estimating indirect environmental conditions based on collected traditional and crowd-sourced data. In particular, we will describe the initial deployment of WeatherCitizen in a marine environment to improve situational awareness and forecasts for mariners. In this context, we will discuss our workflow for collecting data and assess different data reduction strategies, including confidence filtering, machine learning, and physics-based modeling. We will discuss the issues with collecting crowd sourced data and mitigation strategies to increase the quality and density of observations. Finally, we discuss our preliminary experience utilizing this information, including data assimilation for improved marine weather forecasting, and the roadmap of WeatherCitizen into the future.