AURORASAURUS:

Citizen Science, Early Warning Systems and Space Weather

Andrea Tapia, Nicolas Lalone

Penn State University

atapia@ist.psu.edu and nick.lalone@gmail.com

Elizabeth MacDonald

NASA Goddard

eliz.macdonald@gmail.com

Nathan Case

NASA Goddard, New Mexico Consortium

nathan.a.case@nasa.gov

Michelle Hall

Science Education Solutions

hall@scieds.com

Matt Heavner

New Mexico Consortium, White House OSTP,

Los Alamos National Laboratory

Matthew_J_Heavner@ostp.eop.gov

Abstract

We have created Aurorasaurus, a website, a mobile application and a scientific tool that allows a community of users to better predict sightings of the aurora borealis. We focus on the aurora borealis as a rare and unpredictable event (as a proxy for a natural disaster), as it is in the middle latitudes, highly populated areas in North American and Europe. In the northern half of the continental United States the aurora may be visible once or twice per year. In the Southern half, perhaps only once every 20 years, especially during a solar maximum. We feel that the similarities between natural disaster occurrence and auroral occurrence can offer a chance for researchers to test elements of an Early Warning System.

The years around 2014 are the latest solar maximum recurring on an eleven-year solar cycle. Now is the time when aurora is more likely over populated areas, and this is the first solar maximum with social media, an unprecedented opportunity to engage the public, the scientific community, and the media.

Space weather scientists have only coarse means to predict where the aurora will occur. Forecasts derived from state-of-the- solar wind models and satellite-based observations of the Sun estimate the arrival of the coronal mass ejection material at Earth to a typical accuracy of ±8 hours. There is no way to accurately predict the strength of an incoming storm. Space weather observations are simply not plentiful enough to accurately forecast the 93 million miles of interplanetary plasma between the Sun and the Earth. However, a network of observers on the ground can provide ground truth, which could enable nowcasts. We anticipate an order of magnitude increase in accurate knowledge of where and when the aurora are visible by augmenting existing observations and models with real-time ground truth (thus going from 100s to 10s of km, and from hour to minute time-scales).

Using these new data we anticipate three major impacts on geospace science: 1) Significantly more accurate, timely, and relevant methods of auroral predictions. 2) New cost-effective sources of human-derived data for observing the aurora, including volunteered geographic information, images and videos, which may aid scientists in understanding their space-based data. 3) The ability to study auroras at mid-latitudes during storms. A storm has never been imaged at mid-latitudes and the last global imaging satellites with significantly lower resolution ended their missions in 2006. Space scientists are still trying to connect many magnetospheric regions to their visible manifestations in the ionosphere in this data-starved, vast, and relatively young field. Other citizen science activities may yield discoveries on the shape of the dynamic auroral oval

versus time and activity due to the current lack of data on these phenomena.

On October 24th, 2011 aurora was visible as far south as Alabama. Due to the excellent timing and widespread visibility in clear skies over heavily populated areas, this storm was covered by more than 500 traditional and online news outlets from large international newspapers to blogs. In real-time, tweets exceeded one per second during the event. Additionally, thousands of Twitter messages (*tweets*) containing text and images documented the visible aurora in real-time (MacDonald, 2012). During the testing of our prototype, we found that the *Kp index*, a key measure of auroral activity, correlates strongly with the number of aurora-related tweets. This suggests that volunteered Twitter data is a valid indicator of real-world events.

AURORASAURUS

Aurorasuarus combines limited space weather science data, participant sightings and the analysis of social media data into a better prediction engine. This then serves to alert interested parties as to nearby sightings. We find that citizens are curious about auroral activity, and this motivates them to seek information about when and where these events occur. We also find that citizens are willing to share information about auroral sightings with others. The vision of Aurorasaurus leverages this to gather real-time data about the aurora in two ways, by direct entry into an online form, and by continuous scanning of Twitter for tweets about sightings. Combined with data from Earth-based and satellite observatories, this allows Aurorasaurus to offer near real-time predictions of auroral activity in both text and map form, likely with much greater accuracy and timeliness than the current state of the art.

Aurorasaurus has been live on the web for one year and has shown promising results. At its core, Aurorasaurus offers an estimate of the aurora's location and character while simultaneously seeking participatory data for improving this nowcast. User observations of the aurora — collected from Twitter as well as an entry form offered on the website itself — are displayed on a map as pins. Users can click on these pins to see more detail about an observation; sometimes this detail includes photos or videos. Aurorasaurus also displays a current and 1-hour coarse estimate of the auroral oval's location based on satellite observations of the near Earth solar wind environment. We integrate user observations and these space-based estimations as part of the now-casting algorithm.

The Twitter data stream requires a significant level of filtering to select aurora-related tweets from the daily 300-500 million tweets shared in total. Our current classification algorithm, trained on a set of tweets manually classified by us, divides a tweet stream initially filtered by key-word into the following classes: positive sighting (i.e., someone saw the aurora), negative sighting (someone looked for the aurora but failed to see it), desire to see the aurora, not aurora related, and unclassifiable. Aurorasaurus brings users into the tweet filtering loop by allowing them to give feedback on the classification algorithm's judgments, using up- and down- votes similar to those found on the Reddit social news site; this will both improve the algorithm's accuracy and make it more nimble in the face of changing language. Second, we integrate improved location inference techniques. Geo-location is accomplished either by using the (typically derived from GPS or other automated location services) geotag provided by Twitter directly or a gazetteer-based approach applied to toponyms (place names) in the tweet text or user profile. This yields approximately 400 tweets on a typical day and 1,000 or more during a non-routine event.

We augment this web implementation with a mobile application, in order to make Aurorasaurus more fully available in the field, where its services are most useful. If the user makes a positive aurora sighting the user may then use the mobile application to submit a positive sighting of the aurora directly. Users can take a photo or video of the even and tag it with location, color, and height in the sky and aurora type. The user enters this data on a simple mobile form. This entry contributes to improving the prediction algorithms of aurora activity in general and contributes to the predicted aurora oval made by Aurorasaurus in the specific at that time and location, allowing other users to be more confident in their seeking a sighting.

Through this we are also building a *citizen science community*. That is, while many people have heard of the aurora and seen photographs, fewer have seen the curious and remarkable phenomenon personally, and fewer still can describe its causes in terms. Our vision is to engage citizen scientists in every phase of the research, from scientific observations contributed as tweets, structured "lab notebook" logs, images, and video, to verification and cleaning of data, to analysis; the results will inform scientific research in space science, machine learning, and other fields. We will offer a suite of citizen activities that advance current understanding of space weather. Finally, we offer learning modules to support the specific citizen science projects that can be undertaken by a wide variety of audiences for both deep and meaningful learning as well as science.

We expect that our tools will serve as a proof-of-concept for crowdsourcing other forms of alert and early warning system. Such a system could be employed during other kinds of disasters such as earthquakes, tornados, tsunamis, and flooding. Current attempts to integrate data from the crowd or social media have not yet moved beyond the stage of contributing the results of social media monitoring to existing monitoring systems. As an event is triggered it

appears on a map for all users as a form or continual, geographic notification. In addition, we provide targeted, local, custom notifications (push text alerts) to each user. We close that loop when a user receives the notification and acts upon it. We believe that translating heterogeneous motoring data into nowcasts and alerts that can be acted upon in near real-time, is the most important element in an early warning systems. Our early results from our beta phase point to potential successes in combining both algorithmic and participant filtering of social media data into faster more accurate nowcasts.

Another potential contribution of the Aurorasaurus project to early warning systems science and design is the ability to track action taken once notification is given. A problem in the design-science of early warning systems is the inability to know if an alert or notification was heard, heeded and acted upon by the receiver. Through Aurorasaurus we can follow the path of a notification through a user into action. A user only receives notification if they are registered and there is reported aurora activity in the area. If they receive it, then contribute to the site a post in the form of either a positive or negative sighting, we can know that the notification was received and acted upon.