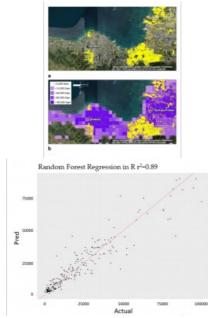## *PREDICTING OPENSTREETMAP COMPLETENESS IN RISK AREAS*

Getting ready for disaster events with remote sensing

Random Forest Regression in R r²=0.89

In a project in Haiti, several remotely sensed indicators were assessed to predict the completeness of Open Street Map for use in disaster relief and to guide future mapping efforts.

Timely information related to the distribution of vulnerable populations and critical infrastructure is key for effective disaster relief, especially in view of the increased frequency and severity of natural disasters. OpenStreetMap shows great potential to support humanitarian mapping tasks and has provided vital information in many past major disasters. Publicly available remotely sensed measurements can be utilized to identify areas that have not yet been fully mapped and help guide and prioritize future mapping efforts in preparation for future disasters.

On the afternoon of 12 January 2010, the island of Haiti suffered its most devastating natural disaster ever: a magnitude-7.0 earthquake. An estimated 250,000 people lost their lives, 300,000 people were injured and close to 1.5 million people were forcibly displaced. Six years later, in October 2016, Hurricane Matthew struck southwestern Haiti, leaving a further 900 people dead and 28,000 homes damaged. In fact, Haiti has been hit by ten hurricanes and other tropical storms since 1998, with every landfall causing widespread loss of life and flooding. More than 96% of its population is exposed to various types of natural hazards, particularly hurricanes, coastal and riverine floods, and earthquakes.

Devastating consequences of natural disasters

Haiti is not alone. Over the last few decades, many countries have been challenged by the devastating consequences of natural disasters which pose a significant threat to human health and safety, property, critical infrastructure and homeland security. Every year, natural disasters impact close to 160 million people worldwide, causing destruction of the physical, biological and social environments, impacting food security and causing global losses that amount to over US$100 billion. The Overseas Development Institute (ODI) estimates that, by 2030, up to 325 million extremely poor people will live in the 49 most hazard-prone countries, the majority of them in South Asia and sub-Saharan Africa.

When a disaster strikes, first responders are the first to react. They support search, rescue and evacuation efforts, aiming to meet the immediate and pressing needs of vulnerable communities for shelter, food, water and medical treatment. Timely information about the location, availability and functionality of critical infrastructure such as hospitals, shelters, water and sanitation facilities, roads and public transportation, as well as about where people live and the location of potentially isolated communities, is critical for effective disaster relief aimed at saving lives. Accurate, complete and accessible geospatial information is the foundation of a successful disaster relief effort.

Created in 2004 by Steve Coast, OpenStreetMap (OSM) is a collaborative project aiming to â€œcreate a free editable map of the worldâ€. As of today, there are more than 5.5 million OSM users and one million contributors who generate more than three million changes every day, as well as specialized groups such as Humanitarian OpenStreetMap Team (HOT-OSM) that conduct activities aimed at enriching OSM data to support emergency relief operations.

Volunteered geographic information

Until recently, governmental agencies and the commercial sector were the two primary sources of geospatial data for disaster management purposes. In the past decade, however, citizens have been increasingly recognized as a valuable source of such information too. Volunteered geographic information (VGI), or geographic information collected by individuals often on a voluntary basis, has been playing a growing role in the support of humanitarian relief. This trend was even recognized by *Time* magazine in 2006, when 'You' were chosen as the magazine's Person of the Year.

When the 2010 earthquake struck Haiti, accurate and timely information on the distribution of critical infrastructure was exactly what the disaster response community was looking for. But existing data, including from Google Maps, did not offer comprehensive coverage of the necessary features. In the days following the earthquake, many organizations released high-resolution satellite imagery under open licensing schemes, catalysing a worldwide effort to map Haiti and support the recovery operations. Over 450 OpenStreetMap volunteers from around 29 countries used the imagery to digitize roads, buildings and other features, creating the most detailed map of Haiti in existence in just a few weeks.

Created in 2004 by Steve Coast, OpenStreetMap (OSM) is a collaborative project aiming to "create a free editable map of the world". As of today, there are more than 5.5 million OSM users and one million contributors who generate more than three million changes every day, as well as specialized groups such as Humanitarian OpenStreetMap Team (HOT-OSM) that conduct activities aimed at enriching OSM data to support emergency relief operations.

OSM coverage

According to a [recent study](#), the world's user-generated OSM road map is more than 80% complete. However, OSM's coverage of features varies significantly between – and often within – countries. For example, its coverage of remote areas is often lower than coverage of highly populated urban areas, and the coverage of developed countries tends to be lower than that of developing countries. A complete and accurate OSM map, especially in countries vulnerable to disasters events, is essential for the support of timely and effective search and rescue operations during and immediately after disasters.

In 2019, the World Bank's Global Facility for Disaster Reduction and Recovery (GFDRR), the [GFDRR Innovation Lab](#), together with a research team at New Light Technologies Inc., evaluated the potential use of satellite data and other types of remotely sensed derived products as predictors of OSM coverage. The objective of the project was to leverage different types of remote sensing measurements, together with machine learning approaches, in order to identify geographical areas where OSM coverage of building footprints is incomplete. A robust methodology to identify areas that are not yet fully mapped by OSM and that are at risk of environmental and natural hazards would allow prioritization of mapping efforts and preparation for future disasters.

The ratio between the actual and the predicted area of OSM building footprint in a grid cell (shown in a & c). Grid cells exhibiting the lowest ratio (shown in b & d) represent locations where the actual area of OSM building footprints is much lower than the predicted area.

Haiti was chosen as a case study for this project because, despite some densely mapped zones of Port-au-Prince, large portions of the country remain unmapped. This is potentially the result of the episodic engagement of community mapping volunteers and the definition of mapping 'tasks' on a neighbourhood scale through OSM editing tools. To predict the completeness of OSM building footprints in Haiti, the research team first created a tessellated grid of cells, each cell $0.25km^2$ in size, spanning the entire country. The completeness of OSM building footprints in close to 1,600 cells was visually assessed against high-resolution satellite imagery as a base layer. Half of these cells, which were relatively fully covered (i.e. more than 75% of the buildings in a cell were estimated to be mapped), were used as a training set in order to evaluate the predictive power of different remotely sensed indicators.
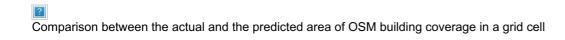
Several remotely sensed indicators were assessed as potential predictors of OSM completeness, including intensity of light emitted at night (based on VIIRS measurements), spectral indices (NDVI, NDBI, SAVI, UI) derived from Sentinel-2 satellites, surface texture (based on Sentinel-1 SAR measurements), elevation and slope. Additional remote sensing classification schemes representing the distribution of forest and built-up land cover were also assessed as predictors.

Predicting building footprint variation

The analysis showed that the combinatorial effect of nine of the variables together explains up to 82% of the variation in OSM building footprints in Haiti. This is much higher than the predictive power of each variable independently. The team found that Random Forest regression predicts up to 89% of the variation in OSM building footprints in a cell, suggesting the potential utilization of remotely sensed measurement indicators of OSM building footprints. The trained Random Forest regression model was then used to predict the coverage of OSM building footprints over the entire country. This allowed the team to identify cells that are predicted to be covered but that are actually not mapped yet – areas that would require extensive mapping efforts to ensure complete coverage to, for example, support disaster management in the future.

Evaluation of this method in a case study of two additional small island states (Dominica and St. Lucia) produced similar results, with a prediction of up to 94% of the variation in OSM building footprints in those states.

To summarize, with the increase in the frequency and severity of disaster events – especially in developing countries – it is essential to ensure the availability, accessibility and accuracy of geographical information to support disaster management operations. Although VGI platforms, specifically OpenStreetMap (OSM), show great potential to support humanitarian mapping tasks, significant areas – including countries vulnerable to natural disasters – remain unmapped. This study showed that the increasing availability of free and open-source remotely sensed data can be utilized to identify locations where built-up features still need to be mapped. Overlaying the model's predictions with hazard-risk data could allow not only the disaster management community, but also the OSM community to prepare for future disaster events and ensure that essential geospatial information is available to support the response and recovery efforts during and following major disasters.

Comparison between the actual and the predicted area of OSM building coverage in a grid cell