Using Online Media to Create A Real-time and Historic Database of Global Flood Events

Jens de Bruijn (1), Hans de Moel (1), Brenden Jongman (1,2), Marleen de Ruiter (1), Jurjen Wagemaker (3), and Jeroen Aerts (1)

(1) Institute for Environmental Studies, VU University, Amsterdam, the Netherlands, (2) World Bank Group, Washington D.C., USA, (3) FloodTags, The Hague, the Netherlands

Major disaster databases, such as Munich Re's NatCatSERVICE, provide accurate information about disasters around the world. While these databases provide valuable information for some applications, not all flood events are detected and compiling the database requires manual labour. Therefore, in this study, we analyze over 100 million tweets that mention flood-related keywords in 12 major languages to automatically detect flood events on a global scale. For this, we first filter the tweets using neural networks and keyword filters, and extract mentions of locations from the tweets. Based on these mentions, these flood-related tweets are assigned to countries or their first order administrative subdivisions. Then, we detect bursts within these geographical locations and label such burst as flood events. This enabled us to find over 10.000 flood events over a period of 3 years. Validation of these results shows that over 80% of the events are correctly detected and that ∼70% of the events in Nat-CatSERVICE are detected at the level of first order administrative subdivisions. The resulting events are freely accessible on www.globalfloodmonitor.org. This method can be applied to many other different hazard types, such as earthquakes, droughts and forest fires.