

How Do Eyewitness Media Reports

Reflect Socio-Economic Effects of Natural Hazards?

In this study we make use of the YFCC100M dataset in order to verify whether statistically robust relationships exist between the volumes of uploaded content during different natural hazards and estimated human and economic losses in the affected countries. Our findings demonstrate that Flickr reflect impacts of events with the highest frequency of occurrence (such as floods or storms) and/or with the recurring spatial structure (such as landslides or earthquakes).

Background

Natural hazard analytics using online activity traces and eyewitness accounts as additional or alternative data sources has been already applied to individual hazards, for example, flood impact prediction and semantic event forecasting). However, we argue that what is currently lacking from this emerging domain of analytics is records of how various hazards are represented in user generated content (UGC) sourced from social media platforms and whether it is possible to use UGC from the same platform to estimate impacts of different kinds of hazards. To answer this and related questions, we mined the entire content of the YFCC100m dataset in order to extract georeferenced entries for the main hazard keywords ('landslide', 'earthquake', 'flood', 'storm' and 'volcano/eruption') and these were benchmarked against the official, globally normalized estimates of socio-economic impacts contained in the EM-DAT dataset.

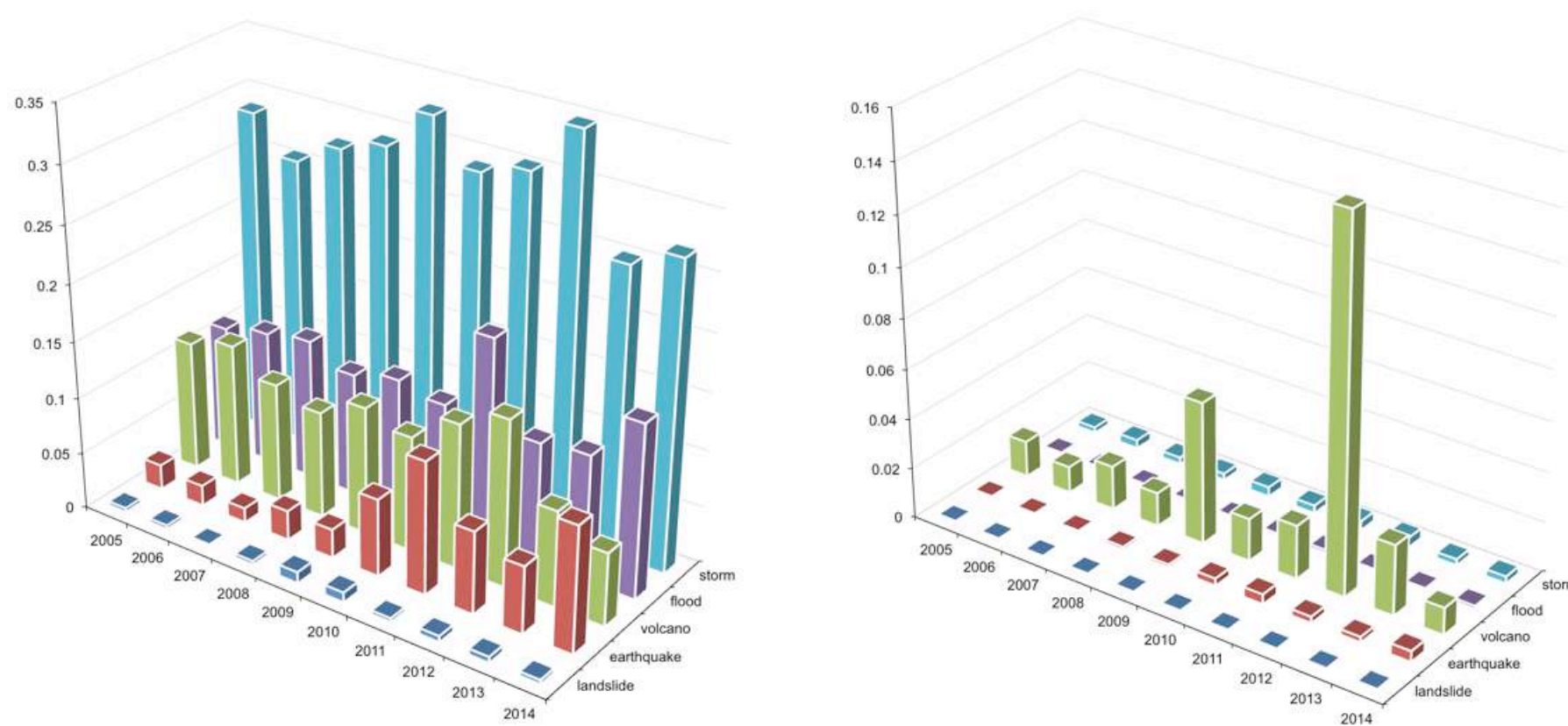


Figure 1: Fractions of hazard-related image and video material uploaded to the Yahoo! Flickr platform worldwide (2005–2014): normalized to total posted (left) and normalized to total posted and by number of events (right)

Results

| | landslide | earthquake | volcano | flood | storm |
|---|-----------|------------|---------|--------|--------|
| Worldwide, each year 2005-2014 | | | | | |
| Total affected | 0.508* | -0.262 | -0.034 | 0.204 | -0.318 |
| Total damage | 0.656** | 0.522* | -0.207 | 0.564* | 0.171 |
| Across countries, entire 2005-2014 period (volume) | | | | | |
| Total affected | -0.007 | 0.064 | 0.040 | -0.004 | -0.018 |
| Total damage | 0.001 | 0.064 | 0.023 | -0.004 | -0.005 |
| Across countries, entire 2005-2014 period (density) | | | | | |
| Total affected | 0.227 | 0.123 | 0.052 | 0.245 | 0.285 |
| Total damage | 0.217 | 0.121 | 0.080 | 0.399* | 0.499* |
| Statistical significance: (**): $p < 0.05$, (*): $p < 0.1$, ():not significant. | | | | | |

The results presented in the table above illustrate relationships between all three indicators derived from social media activity on Yahoo! Flickr and the actual impact metrics extracted from EM-DAT dataset (v.2016). We can observe straightaway that social media activity indicators do not exhibit any consistent behaviour across all five hazards within the scope of this study. However, we do observe a slightly stronger relation between social media activity and infrastructure damage costs across all three scenarios in our FII framework.

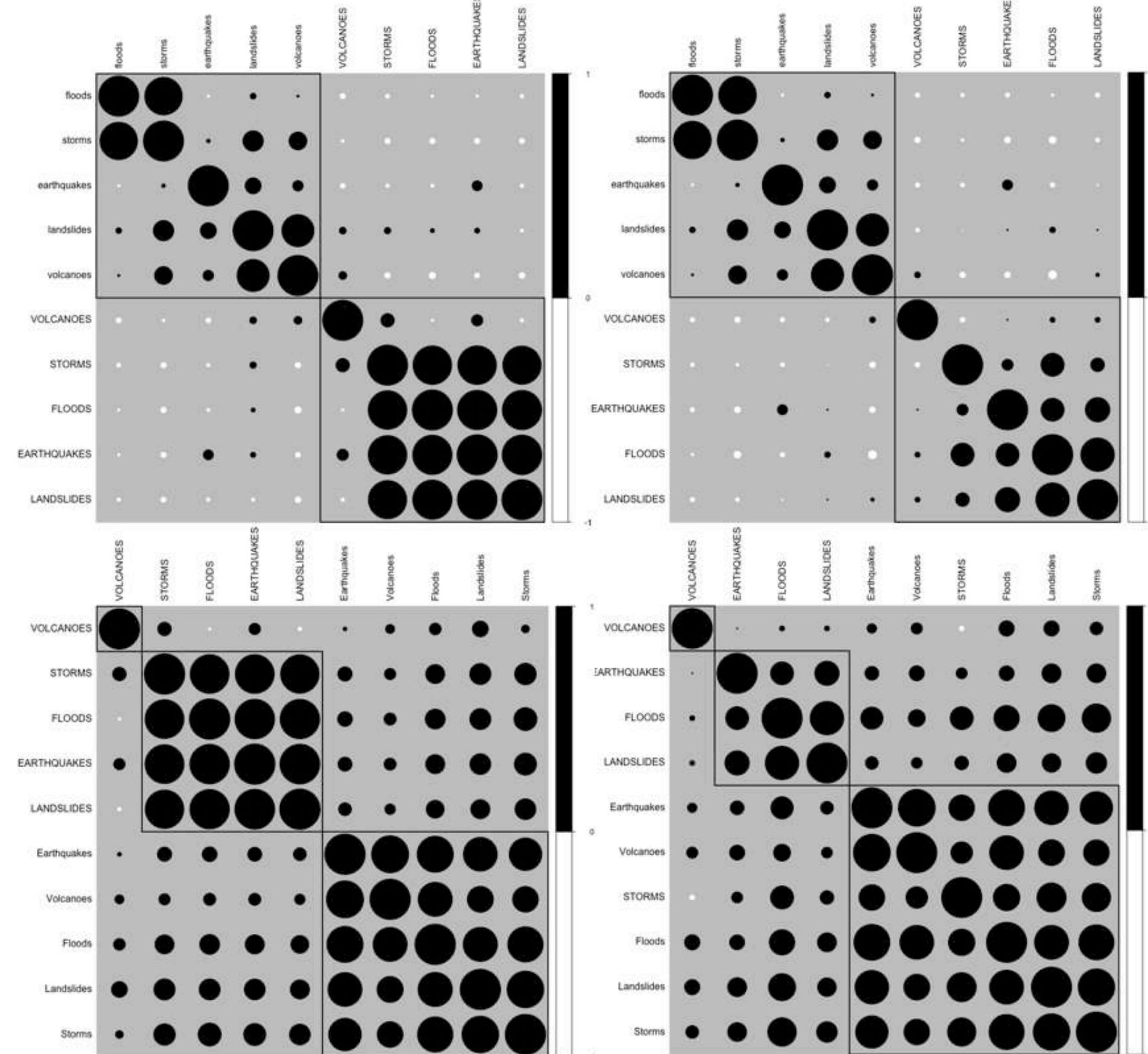


Figure 2: Expanded version of the FII matrices, calculated for the entire 2005–2014 period across all countries, between official EM-DAT metrics of socio-economic impact due to natural hazards (in capitals) and the volume of social media activity (top row) or its spatial density (bottom row), where left and right hand columns correspond to human losses and infrastructure damage respectively

"Global warming is partly to blame for the growing problem of flooding, but so too is thoughtless development."

The Economist
September 2017

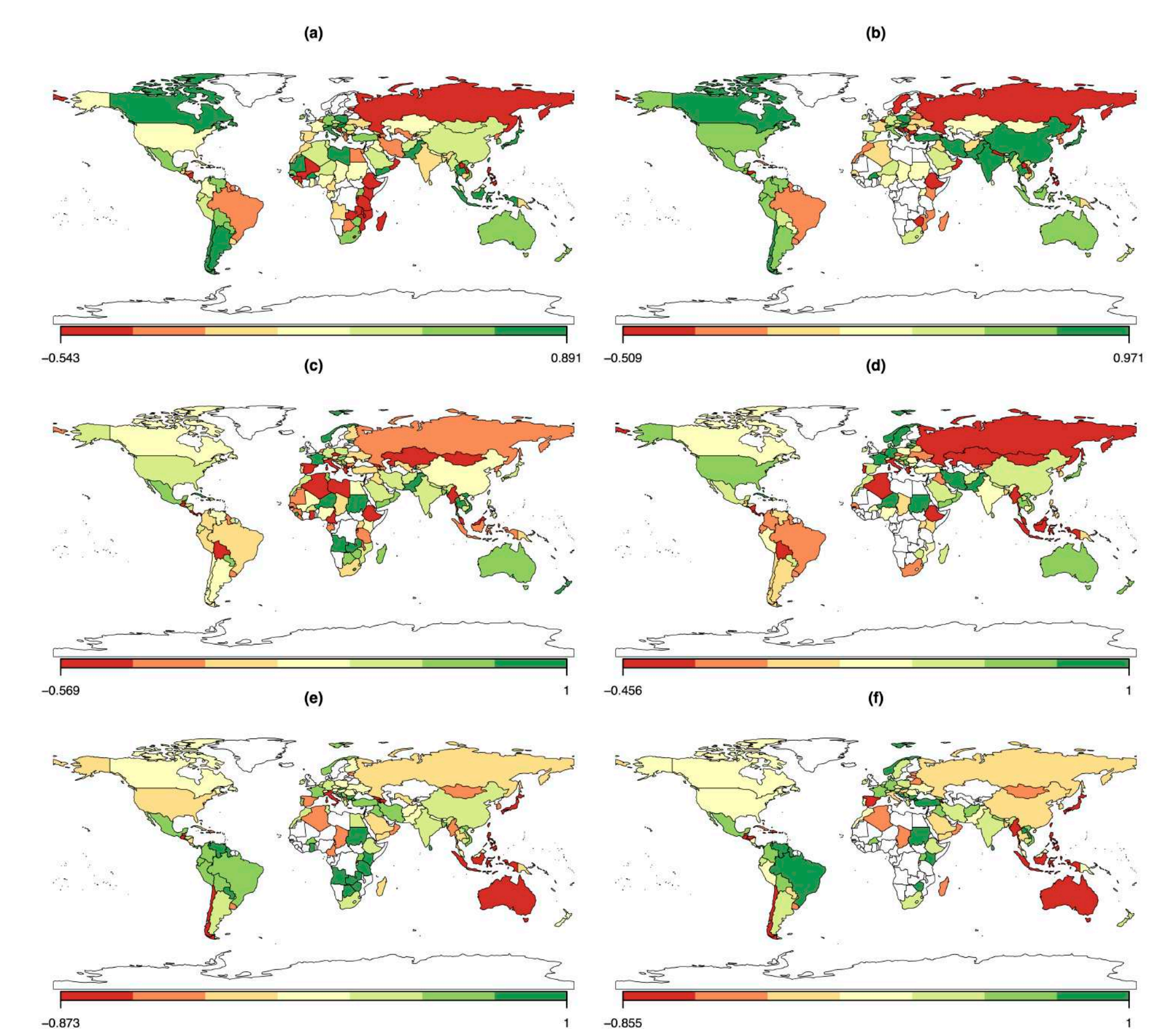


Figure 3: Statistical relationships between topical online image and video material uploaded to the Yahoo! Flickr platform (2005–2014) and the following socio-economic indicators: (a) number of people affected (including Dead, Injured and Homeless) by any type of emergency taken place each year from 2005 to 2014, (b) economic losses (e.g., damage to infrastructure) due to any type of emergency taken place each year from 2005 to 2014, (c) number of people affected (including Dead, Injured and Homeless) by different types of hazards, occurring during the entire period 2005–2014, (d) economic losses (e.g., damage to infrastructure) due to different types of hazards, occurring during the entire period 2005–2014, (e) number of people affected (including Dead, Injured and Homeless) by different types of hazards, occurring during the entire period 2005–2014 (accounting for spatial density of social media posts), (f) economic losses (e.g., damage to infrastructure) due to different types of hazards, occurring during the entire period 2005–2014 (accounting for spatial density of social media posts)

Nataliya Tkachenko
N.Tkachenko@warwick.ac.uk / @FloodSmartCity