

Analysing User Contribution Patterns of Drone Pictures to the dronestagram Photo Sharing Portal

Hartwig Hochmair and Dennis Zielstra

Drones, also known as unmanned aerial vehicles, are nowadays frequently used to supplement traditional airborne data collection methods such as aerial photography and satellite imagery. Dronestagram, launched in July 2013, is one of the first Web 2.0 projects that share geo-referenced drone pictures, providing a valuable source of VGI image data. This paper analyses spatial patterns of contributions to dronestagram world-wide and for two selected regions. Results show that the number of uploaded pictures is associated with the socio-economic development of a country and the presence of geographical features, and that pictures are clustered in sub-regions.

Key words: VGI, drones, pattern analysis, photo sharing

1. Introduction

Photo sharing Web sites are prominent examples for Web 2.0 platforms that facilitate the collection and distribution of Volunteered Geographic Information (VGI) (Goodchild 2007). Examples are Flickr (<http://www.flickr.com>) and Panoramio (<http://www.panoramio.com>). Both allow a Web user to upload geotagged images to an application server where the geographic position of an image is then visualized on a world map through a thumbnail image. While photo sharing services typically host imagery taken from the ground, recent Web 2.0 projects facilitate the sharing of geo-referenced drone pictures. A drone, which is also referred to as an unmanned aerial vehicle (UAV), is defined as a pilotless aircraft operated by remote control (Farlex 2014). Whereas historically UAVs were remotely piloted, autonomous control is increasingly being employed (Adams and Friedland 2011; Watts *et al.* 2012). Drones are used for numerous commercial and governmental applications. Examples are UAV based photogrammetry (Remondino *et al.* 2011) and LiDAR (Lin *et al.* 2011) for

mapping and 3D modelling, agriculture (Torres-Sanchez *et al.* 2013), and disaster research and management (Adams and Friedland 2011). A detailed overview of different UAV platforms and applications is provided by Watts *et al.* (2012). Drones are an important alternative to aerial images from piloted aircraft since they can obtain subdecimeter resolution imagery, be deployed quickly and repeatedly, and are less costly and safer than piloted aircraft (Laliberte and Rango 2009). As a VGI data source, drone imagery supplements existing photo sharing Web sites, such as Flickr and Panoramio, since drone pictures are taken from some distance above the ground, providing a more comprehensive view of a scene.

Public aviation authorities typically distinguish between UAVs (for commercial use), and model aircraft (for recreational use), with more legal restrictions on UAVs, e.g. requirements for certificates, than on model aircraft. The use of UAVs for commercial and professional use in the United States is mostly restricted to supporting public operations, such as military and border security tasks, which will change in 2015 through new regulations (Federal Aviation Administration 2013). The use of model aircraft in the US is permitted up to 400 feet above ground level and away from airports and air traffic (Federal Aviation Administration 1981). An overview of drone guidelines and laws for selected nations across the globe and associated estimated ratings of UAV friendliness is provided by Garcia (2013). On the one end of the spectrum are highly UAV friendly nations, such as Mexico, where no civil aviation authority regulations beset UAV users. On the other end are some Asian countries, such as China, Japan, and South Korea, where non-military use is limited to some branches of agriculture, such as rice fields of Japan. Member states of the European Union operate under the jurisdiction of the European Aviation Safety Agency (EASA) and are rated somewhere in-between in terms of UAV friendliness. UAVs of less than 150 kg are exempt from Agency

regulations and basic regulation is the responsibility of national member states, where in general more restrictions apply to heavier UAVs than for lighter ones or model aircraft (Bundesministerium für Verkehr 2012; UK Civil Aviation Authority 2012). For model aircraft, some operation restrictions for their usage, such as maximum flight heights or minimum distance away from airports, may apply. Some countries, such as Germany, require also a liability insurance even for users of model aircraft.

The dronestagram project (<http://www.dronestagr.am/>) is a photo sharing Web sites focussing on drone pictures. Drone models used for taking aerial images shared on that Web site indicate that these are light-weight models which commonly fall under the category model aircraft. The Web site has also been facilitating the sharing of drone videos since November 2013, which is, however, not further explored in our analysis. The project is based in Lyon, France, and the first drone picture was uploaded on 16 July, 2013. This paper describes the geographic coverage of uploaded pictures and further analyses the point pattern of image locations that evolved from user contributions within the first eight months between July 2013 and March 2014. This study will allow other users to better assess the suitability of dronestagram picture data for their projects. Somewhat related to dronestagram is the Balloon & Kite Mapping project of the Public Laboratory for Open Technology and Science (Public Lab 2013), which hosts an archive of imagery taken by balloons and kites and further provides guidance for building a balloon aerial photography system. Some of the crowd-sourced aerial images on that Web site are featured on Google Earth (Ingraham 2012). Drone imagery was recently used to trace features in OpenStreetMap (OSM) (Sudekum 2013).

The remainder of the paper is structured as follows: The next section reviews related work on VGI point data analysis. Section 3 describes the extraction and characteristics of used drone data, which is followed by a regression analysis of

dronestagram contribution patterns in section 4 and cluster analysis in section 5. A summary and outlook for future work are provided in section 6.

2. Related work

Elwood *et al.* (2012) distinguish between two methods of online georeferencing user generated content. The first method is online geotagging photographs, text, or other online media, such as tweets. The second method describes the sharing of activities along with current whereabouts, which is also referred to as geosocial networking. The analysis of contribution patterns in dronestagram is concerned with geotagged picture information, and therefore relates to the first method of georeferencing.

The position and content of VGI point data has been frequently used for the analysis of people's spatio-temporal travel patterns and perception of space. This includes the extraction of travel trajectories (Andrienko *et al.* 2009; Girardin *et al.* 2008; Zheng *et al.* 2012), events (Chen and Roy 2009), popular places (Schlieder and Matyas 2009), and vernacular regions (Hollenstein and Purves 2010) from geotagged images. Geotagged tweets were used to forecast the spread of diseases (Brennan *et al.* 2013) and people's activity patterns (Li and Shan 2013). Login information from Foursquare was used to identify movement patterns within cities (Noulas *et al.* 2012).

As opposed to this, only relatively few studies analyse user contribution patterns of VGI point data, in particular those of shared photos. No study could be identified in the literature that focuses on contributions to drone picture collections. Using California as a case study Li *et al.* (2013) found that well-educated people in the occupations of management, business, science, and arts are more likely to be involved in the generation of georeferenced tweets and Flickr photos than others. Antoniou *et al.* (2010) studied contribution patterns to photo sharing services in Great Britain, including Flickr, Panoramio, and Picasa Web (<http://picasa.google.com>), which revealed high clustering

in urban areas and around tourist attractions. The paper introduces an expectation surface which uses the chi-statistic. It is negative for the tiles where the number of geotagged photos is lower than expected (according to population data) and positive when it is greater than expected. Another paper analysed the change of contributor patterns to Flickr between different seasons in the area of South Tyrol, Italy (Sagl *et al.* 2012). Results showed that the main tourist season is summer, and that tourist destinations vary heavily between summer and winter seasons. Hecht and Gergle (2010) compared the localness of Flickr and Wikipedia contributions, which was found to be higher for Flickr. That is, approximately 53 percent of Flickr users contribute, on average, content that is 100 km or less from their specified home location, whereas this number drops to around 23 percent for the English Wikipedia. Hochmair (2010) found that the density of contributed Panoramio pictures is higher along the scenic route than along the corresponding fastest route for a given trip origin and destination.

3. Data

Extraction of drone picture data points

When uploading a new drone picture to dronestagram the user is asked to specify the picture location through country, county or state, city, and street name (but not a street number). The upload interface does not allow the user to input geographic coordinates or to position the image manually on a background map. Next dronestagram converts this information through address matching to geographic coordinates. If the user provides a street name that can be matched against the reference database the picture is placed at the midpoint of that street, otherwise at the city centre. Despite this geocoding, picture coordinates are not displayed or made accessible for download. Instead, the geocoded picture position is visualized on top of a zoomed background map. One workaround to access picture coordinates is to download the complete content of the

Web domain, where each picture is stored in a separate folder together with a Webpage. That Webpage is displayed when a user clicks on a thumbnail image within the various directory listings of the main page. Each Website displays, besides the drone picture, also user provided metadata, such as image title, camera and drone type, city and street, and posting statistics, including the number of views and comments from other users (Figure 1).

The geocoded coordinates of a drone picture with its latitude and longitude values can be retrieved from the HTML and JavaScript source code of the picture Web site. The downloaded domain content contained a total of 1837 picture folders, but only 1643 picture Web sites contained latitude and longitude information in their source code, meaning that 194 pictures were not successfully geocoded by the dronestagram server-sided application. A one by one check of the corresponding 194 picture Webpages revealed that 96 of them contained user provided city names and an additional 21 also the street name, which were for some reason not geocoded on the dronestagram Web site. Furthermore, we could identify the location of 48 additional images through the image title, which contained for example city or landmark names. This resulted in a total of 165 out of 194 picture Web sites containing geo-referencing information, whereas 29 contained none. Since the goal of this study is to analyse user contribution patterns based on a dataset that reflects picture uploads as completely as possible, we retrieved the geographic coordinates for these 165 pictures by typing provided location information (e.g. city, street, or landmark name) into the Google Earth Search field. We refer to this set of 165 pictures as manually geo-coded. Together with 1643 automatically geo-coded pictures this gives a total of 1808 (Figure 2). A one by one comparison between the automatically geo-coded position and the user provided location description of a picture revealed that in two cases pictures were significantly

misplaced, i.e., located in the wrong country. Geographic coordinates were then corrected for these two pictures. Some of the 29 image Web sites that had no spatial information showed commercial imagery, such as a company logo or sales pictures of cameras or drones.

Descriptive characteristics of picture contributions

Figure 3 shows the development of picture and contributor numbers between the project start on July 16, 2013 and March 10, 2014. Within these first 237 days 391 users contributed 1808 pictures with location information.

The number of pictures each user contributed varies heavily (Figure 4), demonstrating Participation Inequality (Nielsen 2006). This contribution pattern matches that of other community based projects, such as OSM (Neis and Zipf 2012) and Wikipedia (Javanmardi *et al.* 2009), where the majority of users makes few contributions, and only a small percentage of users is actively contributing. Among the users in dronestagram who contributed picture data, more than one third (147/391) uploaded only one picture. The highest contribution per user was 73 pictures for one user.

When uploading a picture to dronestagram the contributor is asked to assign the picture a pre-defined category, such as “Cityscape” or “Industrial”. The Website uses this information to facilitate searching for drone pictures by category. Most uploaded picture contributions were categorized as “Country” and “Urban” (Figure 5). The list of provided categories are modified from time to time. For example, the single category “Others & Crazy stuff”, which was available at the time of data download for this analysis, was recently split into the two categories “Others” and “Crazy Stuff”. Some of these changes are reported in the dronestagr.am blog, but the Web site does not provide definitions for the different categories and what criteria an image or video needs to

satisfy to be eligible for a specific category. It is therefore left to the contributor to decide under which category he or she uploads an image or video.

One can also determine for how many different countries of the world an individual user contributes pictures. Two country vector layers were used for this task and subsequent regression analyses. The first layer is the 1:10 mio scale Countries layer from the Natural Earth Web site (Natural Earth 2014), which hosts thematic layers from the United Nations (UN) and the US Central Intelligence Agency among others. The Countries layer distinguishes between metropolitan (homeland), and dependent portions of sovereign states. Examples are the United States of America (homeland) and Puerto Rico (dependent portion of the United States). Each of these areas (both homeland and dependent portions) is coded with a UN ISO alpha-3 code, totalling 177 areas. This layer was chosen because it provides the estimated population, a five-tiered categorical variable for income, and a seven-tiered categorical variable for economic development for each area. The second layer comes from the United Nations Environment Programme (UNEP) (UNEP 2014) and includes median age for world countries. This layer contains a total of 235 areas with UN ISO alpha-3 code, plus two US Pacific islands which were not assigned a UN ISO alpha-3 code. The UNEP layer maps dependent islands, such as Aruba, Dominica, or Martinique, more completely than the Natural Earth layer, which explains the difference in total number of features between both layers. For our analysis we used only those areas that were mapped in both layers as indicated by UN ISO alpha-3 code, which resulted in 171 areas. Further we removed drone points that were more than 10 km away from any of these 171 areas as mapped in the UNEP layer. The 10 km buffer was used to retain points for the analysis that were slightly outside the areas due to imprecision in point coordinates of drone images or due

to the generalized shape of the UNEP area polygons. This way 1775 out of the original geocoded 1808 drone pictures were retained for further analysis.

These 1775 pictures came from 386 users and were taken in 56 countries. Most pictures were contributed for France (482), followed by the US (256), the Netherlands (135), Germany (121), Switzerland (79), and Italy (75). Table 1 shows how many users contributed pictures from how many countries. When considering all 386 contributing users only 9% of them (33/386) contributed imagery from more than one country (Table 1, left). This increases to 14% (33/241) when considering users that contribute at least two pictures (Table 1, right). In summary these statistics indicate that most contributors are active in only one geographic region.

4. Analysis of spatial contribution patterns

This section introduces two negative binomial regression models for the prediction of uploaded pictures. The first model uses countries as spatial units and relates picture counts with socio-economic variables including median age, income, and population, in a global prediction model. The second model subdivides two selected regions, i.e., the contiguous United States and part of Europe, into $50 \times 50 \text{ km}^2$ grids and relates picture counts in grids with geographical features in these grids, including oceans and cities. Since the predicted variable, i.e., the number of uploaded dronestagram pictures, denotes count data, an Ordinary Least Square (OLS) regression model is not appropriate so that a negative binomial regression is used instead. As opposed to a Poisson regression, which can also be used with count data, the negative binomial regression can handle overdispersed count outcome variables. Thus the latter is a suitable extension of the Poisson regression which relaxes the assumption of equal mean and variance in observed count data.

For the negative binomial model the dependent variable Y (number of uploaded pictures) has a negative binomial distribution with count data y taking values 0, 1, 2, ... The expected value of Y, the mean μ , conditional on the p explanatory variables X_1, X_2, \dots, X_p , and p+1 parameters $\beta_0, \beta_1, \dots, \beta_p$, is an exponential function of independent variables and determined through

$$E(Y | X_1, X_2, \dots, X_p) = \mu = \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p) \quad (1)$$

or: $\ln(\mu) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$

Since the link function is the natural log, the interpretation of β coefficients is that each one-unit increase in X_i increases mean picture counts by a multiplication factor $\exp(\beta_i)$. A manual model building approach was applied due to known limitations of automated stepwise procedures in multiple regression, such as not being able to determine the variables that are most influential. In a first step a Pearson's r correlation coefficient between the count variable and explanatory variables under consideration were computed, where categories in the categorical variables (e.g., income group) were replaced by dummy variables. Variables significantly correlated with picture counts were then used as a starting point for building a regression model.

Socioeconomic factors – global analysis

The global analysis model relates picture counts per country with the country's socioeconomic characteristics. Since dronestagram does not provide residence information about its contributors, we use the assumption that the country which receives most picture uploads from a contributor is the contributor's home region. 382 out of 386 contributors were found to have exactly one country from which more pictures were contributed than from any other country. Picture counts from the home regions of these contributors were retained for further analysis. For four contributors a home region could not be identified based on this criterion. Picture counts from these

four contributors were therefore excluded. This left a total of 1601 points to be used for further analysis. For the negative binomial regression model the following explanatory variables were considered at the country level:

- median age (ratio, source: UNEP)
- income (ordinal, source: Natural Earth)
- economic development (ordinal, source: Natural Earth)
- population (ratio, source: Natural Earth)

The income variable has been regrouped into four ordered classes, which are 1. High income, 2. Upper middle income, 3. Lower middle income, and 4. Low income. The economic development variable consists of seven categories between 1. Developed region G7 and 7. Least developed region.

In a next step bivariate correlations were computed between the explanatory variables to identify potential problems with collinearity. Correlations were generally small. A few higher correlations between some income and economic development categories were found, such as between 4. Low income and 7. Least developed region ($r=0.749$, $p<0.001$), which were, however, not used together in the final model. Since collinearity is strictly a problem of correlations between explanatory variables that does not depend on the nature of the link function to the response, it can be diagnosed with ordinary least-square (OLS) procedures that provide collinearity diagnostics. Thus we tested each negative binomial regression model under consideration for collinearity with an OLS that used the same variables. For the final model (Table 2) this test procedure revealed a maximum Variance Inflation Factor (VIF) of 3.9, which is in the lower part of the recommended VIF threshold range of < 10 (DeMaris 2004).

Results in Table 2 show that median age, population, high income, and upper middle income are positively associated with numbers of uploaded pictures. This provided model gave the best model fit as measured by Akaike's Information Criterion (AIC). The results support a pattern found in other VGI contributor studies called the digital divide (Goodchild 2007; Heipke 2010), which describes a lack of affordable technology in many parts of the world to participate in VGI projects.

The last row in Table 2 shows a dispersion coefficient that was estimated by maximum likelihood. Since its confidence interval does not include zero, the negative binomial model provides a better model fit than the Poisson model to the analysed data.

Effect of geographical features – regional analysis

This section presents two regional models for the prediction of picture uploads. Two regions are used that exhibit relatively homogeneous socioeconomic conditions, but significant variation of geographical features within each region. Geographical features can be distinguished into natural geographical features, such as terrain types or bodies of water, and artificial geographical features, such as human settlements or engineered constructs, both of which were used in the regional models. The US region was mapped with the USA Contiguous Lambert Conformal Conic map projection (standard parallels at 33 and 45 degrees latitude), and Europe with the Europe Lambert Conformal Conic projection (standard parallels at 43 and 62 degrees latitude).

Each of the two regions was subdivided into 50x50km² grids, for which the dependent variable (number of picture uploads) and the independent variables were measured. This grid size was chosen since it approximates a person's home region and thus reduces the autocorrelation between picture counts. Next, grids were clipped to the outline of the United States (for region 1) and to landmasses within a pre-defined

rectangle covering part of Europe (region 2). Further, areas covered by hydrographic features were erased from the cells to obtain land-based elevation data only. The following explanatory variables were considered for the negative binomial regression models:

- adjacency to ocean (binary)
- adjacency to lake or wide river (binary)
- presence of city (binary)
- mean elevation (ratio)
- forest coverage (ratio)

Data for the first three variables were extracted from ESRI's Data and Maps for ArcGIS 2013 dataset, utilizing the worldwide "hydropolys" and "urban_area" polygon vector layer. The hydropolys layer contains water bodies representing rivers, lakes, seas, and oceans of the world, and the urban area layer contains urban areas with populations greater than 10,000. Both layers are designed for maps scaled 1:250,000 or larger. The hydropolys layer contains an attribute that allows separating oceans from inland features. Among inland features only those with an area larger than 5 mio m² and a shapefactor less than 1000 were considered lakes or segments of wide rivers. The shapefactor was computed as $(\text{feature perimeter})^2/\text{area}$ (P2A) (de Smith *et al.* 2013), which is a commonly used measure of shape since its value is not affected by the size of the feature. A higher P2A value indicates a more elongated feature.

Elevations for Europe south of 60 degrees latitude were obtained from USGS/NASA Shuttle Radar Topography Mission (SRTM) data with a 3 arc-second (90 m) resolution (Reuter *et al.* 2007). North of 60 degrees latitude the ASTER Global Digital Elevation Model V2 (Tachikawa *et al.* 2011) with a 1 arc-second (30 meter)

resolution was used and merged after re-sampling with the 3 arc-second SRTM elevation data. For the US the 100 m resolution elevation map layer from the National Atlas of the United States was used, which is derived from the National Elevation Dataset.

Forest coverage information for Europe was taken from the Corine land cover 2006 data provided by the European Environment Agency (Büttner *et al.* 2012). For the US, forest coverage information was based on land cover data from the USGS National Gap Analysis Program (GAP), which is based on multi-season satellite imagery (Landsat ETM+) from 1999-2001 in conjunction with digital elevation model derived datasets.

In a next step those $50 \times 50 \text{ km}^2$ grid cells from both analysed regions were extracted that were located within a 200km buffer around mapped drone pictures. The buffering was done to exclude areas without any data collection efforts that would not contribute to the regression model. While there is no hard criterion for the selection of the buffer size, visual inspection of the geographical features in both analysed regions suggested that such distance would cover a large variation in geographical features around drone image locations, making it suitable for the regression model. Figure 6 shows the grid cells located within a 200km buffer, which was used for the US and Europe regional models, together with geographical features. In addition, regression models for cells within 50km and 100km buffers around drone image locations were also estimated.

Bivariate correlations between all considered predictor variables were small for both regions (Pearson $|r| < 0.4$). Collinearity diagnostics for the final models using OLS showed a VIF smaller than 1.8 for both regions. To avoid model bias through clustered picture uploads of individual users within small regions, all picture uploads but one of

each individual user per grid cell were removed before picture counts. In other words, the picture count per cell corresponded to the number of different users contributing to a cell. This method reduced the effective overall counts of pictures taken in home regions from 230 to 104 for the US and from 830 to 341 for Europe. Using this method instead of all picture counts gave higher levels of significance in the regression models, but did not change the arithmetic sign of the regression coefficients. Count models need to consider the fact that counts can be made over different observation periods or different areas. This is done by including an exposure variable with the coefficient constrained to be one. Since the offset is the natural log of the exposure, the natural log of the grid area was used as the offset variable for the negative binomial regression models.

Table 3 shows the regression results for both regional models. Presence of ocean and city features in a grid cell are positively associated with picture counts, indicating that more pictures are uploaded along coast lines and in densely populated areas. Further both regions showed a higher density of drone pictures in their Western parts, e.g. California for the US and France for Europe, indicating areas of higher initial popularity of dronestagram in those parts. The negative coefficient of the interaction term for the US model indicates that canopy along coastal regions reduces the number of drone images taken, possibly due to flight obstruction. For Europe higher elevation was associated with fewer drone images, whereas lakes or wide rivers in higher elevations increase drone images. Some hydrographic features at higher altitude with numerous drone pictures include, for example, Lake Geneva (372m) and Lac d'Annecy (445m). The estimation of the dispersion coefficients suggest that the negative binomial model provides a better fit than the Poisson model for both regions. The models presented in Table 3 had the best model fit among several tested models based on Akaike's Information Criterion (AIC).

Estimating the same models for grid cells located within the 50km and 100km buffers around mapped image locations led to similar results as the models with the 200km buffer. That is, positive and negative arithmetic signs of coefficients were retained, with some changes in the magnitude of coefficients. The p-values were also similar to those obtained from the 200km buffer models, rendering all variables that were identified as significant predictors in the 200km buffer models significant in connection with smaller buffer sizes as well.

5. Cluster Analysis

The goal of this analysis is to identify if and where dronestagram picture uploads form high activity contribution clusters when controlled for other, more established VGI photo sources, and population, respectively. Whereas a variety of photo sharing Websites exist (<http://l-lists.com/en/lists/ndr9ye.html>), only a few use geo-coded images and provide free access to their images. Flickr and Panoramio are two of the most prominent photo sharing Websites, both of which allow the image download through API's. Compared to Flickr, Panoramio has the advantage that it features only outdoor pictures, similar to dronestagram. Further it provides also a better positional accuracy than Flickr (Zielstra and Hochmair 2013). Therefore Panoramio picture locations were used as a control dataset in the cluster analysis. Since the number of Panoramio images is large for the two test regions (over 2 mio points for the contiguous US and over 14 mio points for the analysed Europe area), a random sample of 20,000 Panoramio image locations were extracted as control points for both test regions for the cluster analysis.

The cluster analysis was conducted with the SaTScan 9.3 software (Kulldorff 2014) which applies a spatial scan statistics (Kulldorff 1997). The spatial scan statistics is a cluster detection test which detects the location of clusters and evaluates their

statistical significance while adjusting for multiple testing. It is based on the likelihood ratio associated with the number of events inside and outside a circular scanning window. The numerator of the ratio is associated with the hypothesis that the case rates inside and outside the scanning circle are different, whereas the denominator ratio is associated with the hypothesis that the two case rates are equal. Likelihood ratios are computed for circular scanning windows of various sizes, which move along a grid over space. The maximum observed ratio is then compared to ratios that are simulated by assuming the null hypothesis to be true. For cluster detection based on case and control point data, SaTScan provides various models, including a Bernoulli probability model with 0/1 event data, or a Discrete Poisson model for region based case and control counts. In this study both models are applied for the US, and the Bernoulli model for Europe. With the Bernoulli model, the individual locations of dronestagram pictures denote cases, and the sample of Panoramio image locations denotes control points. The discrete Poisson model requires case and population counts for a set of data locations such as counties. In this study dronestagram image locations were aggregated by county representing cases, and controlled for by population per county which was obtained from 2010 US Census data. The SaTScan output provides, among others, a cluster shapefile and a cluster information file. The latter reports the log likelihood ratio associated with a cluster, the cluster radius, the p value, and the numbers of observed and expected cases and their ratio.

For the US, the Bernoulli model identified 11 significant clusters ($p < 0.05$) as shown in Figure 7 and Table 4. Cluster #1 is the largest in terms of radius and number of observations, covering Southern California. This is followed by cluster #5 which covers the border region between Colorado and New Mexico. In addition to the SaTScan output we added the number of different users contributing to each cluster

(right-most column in Table 4). All but the two previously mentioned clusters are based on only one or two contributors, and have a small cluster radius. These cluster characteristics indicate that clusters stem from local mapping activities that do not represent general areas of higher image contribution activity. These smaller clusters are either spatially separated from other clusters, like the one in Miami (#2), or found within larger clusters, such as cluster #4, which is located inside cluster #5. Thus the most prominent cluster is the one covering parts of California, indicating that this region is a leader in applying drone technology for imagery purposes. Although this region already features a high density of Panoramio photos through various National Parks and other tourist attractions, it is even more so a hot spot of drone picture contributions.

Use of the discrete Poisson model, which controls for population per county, results in five significant clusters ($p < 0.05$). Figure 8 highlights counties that are part of a cluster, and Table 5 shows characteristics of these clusters. This cluster approach merges the two major clusters from the previous Bernoulli model to the West into one large cluster covering the southwestern states of the US and provides therefore a more general and less cluttered picture of clustered regions compared to the previous model. This cluster contains now drone image contributions from 24 users. The remaining smaller clusters mostly overlap with those from the Bernoulli model, e.g. Miami (cluster #3) or Tampa (#5), indicating that control for Panoramio images and population result in similar cluster results.

Since for Europe coherent population data at the county level was not readily available and since cluster results from the US did not reveal major differences between use of Panoramio images and population as a control, the cluster analysis for Europe was only conducted with the Bernoulli model, using Panoramio image locations as control. The spatial scan statistic detected 9 significant clusters as shown in Figure 9

and Table 6. Only two clusters have contributions of 3 or more users (#1 and #8). The largest cluster is centered around France, most likely because the dronestagram project was founded in Lyon and then promoted among local contributors. Another possible explanation is that France is among the first countries permitting unmanned drones in the civilian airspace (Masi 2013). The country has already authorized more than 220 operators, and there are 14 companies certified to design drones. Other smaller clusters in Table 6 are primarily the result of local efforts of individual image contributors.

6. Discussion and outlook

This paper started with an analysis of the development of photo contributions to the dronestagram photo sharing platform over time. A growth plot showed that new pictures are continuously uploaded and that the user community is steadily growing. Contribution analysis revealed also Participation Inequality among data contributors. It was found that 55% of participating users contribute only one or two images, and that only 11% of users contribute 10 or more pictures. Analysis showed also that 92% of users contributed pictures in only one country. It can be expected that special promotions, such as the 2014 Dronestagram Photo Contest, which was conducted in collaboration with National Geographic, will increase the awareness of this Web site and attract new users.

This study analysed further three aspects of picture contribution counts, which involves the role of socio-economic variables and geographical features in picture contribution frequency as well as spatial clustering under consideration of Panoramio image locations and population data as a control. The first analysis, which was conducted on worldwide data, revealed a clear relationship between the income category of a country and the number of uploaded drone images among other factors.

This result clearly supports the concept of the digital divide (Goodchild 2007; Heipke 2010), indicating that opportunities to contribute to VGI vary between different countries based on their socio-economic development.

The regional analysis within the contiguous US and parts of Europe showed that the number of contributed drone pictures is positively associated with coastal regions and populated areas, and that also elevation, forest, and lakes or wide rivers have some effect on picture contributions.

The cluster analysis for parts of Europe identified the largest cluster around France. One of the potential explanations is that France is the project home country. This effect is not uncommon in VGI projects. For example, previous studies on OSM data completeness found that one of the cities with the highest contributions of pedestrian segments in the US was San Francisco, which is the city where the US OSM project was launched (Zielstra and Hochmair 2011, 2012). The future development of dronestagram will reveal whether the location of the project home region stays an influential factor for data contributions. For the US the largest contribution cluster was identified in the Southwest, which is known to be one of the thriving regions with respect to IT and software development, and home to many start-up companies. We assume that the increased use of drones for imaging reflects the affinity of this region for technological innovation.

Given the initial success of the dronestagram project and the increased interest of the general public in drone based aerial mapping makes it likely that similar other Web 2.0 applications will be launched in the near future. As far as the US goes, the Federal Aviation Administration (FAA) recognized the potential use of drones for a broad range of commercial activities, which led to a roadmap towards the integration of civil UAVs in the National Airspace System by September 2015 (Federal Aviation

Administration 2013). These changes in regulations will most likely further boost the availability of community based drone pictures in the future, although residents start to raise concerns over the surveillance capabilities of UAVs. One example is a case in Seattle, Oregon, where the police department decided to terminate its drones program and agreed to return the purchased equipment to the manufacturer because of citizen concerns about their privacy (The Associated Press 2013). The emergence of a movement against the use of surveillance drones by law enforcement can also be observed in other states where state legislations require law enforcement to get a probable cause warrant before using a drone in an investigation (Bohm 2013).

Future work will include the analysis of picture content and contribution purpose based on user provided picture metadata, such as tags, to get a deeper understanding of the contribution patterns. These proposed analyses can in the future also be expanded to the analysis of drone video contributions made to dronestagram.

References

- Adams, S. M., & Friedland, C. J. (2011) A Survey of Unmanned Aerial Vehicle (UAV) Usage for Imagery Collection in Disaster Research and Management, *Ninth International Workshop on Remote Sensing for Disaster Response*.
- Andrienko, G., Andrienko, N., Bak, P., Kisilevich, S., & Keim, D. (2009) Analysis of community-contributed space- and time-referenced data (example of flickr and panoramio photos), *IEEE Symposium on Visual Analytics Science and Technology*.
- Antoniou, V., Morley, J., & Haklay, M. (2010) Web 2.0 geotagged photos: Assessing the spatial dimensions of the phenomenon. *Geomatica*, vol. 64, pp. 99-110.
- Bohm, A. (2013) *Status of Domestic Drone Legislation in the States* [online]. Available from: <https://www.aclu.org/blog/technology-and-liberty/status-domestic-drone-legislation-states> [Accessed 7/21/2014].
- Brennan, S., Sadilek, A., & Kautz, H. (2013) Towards Understanding Global Spread of Disease from Everyday Interpersonal Interactions, *23rd International Joint Conference on Artificial Intelligence (IJCAI 2013)*.
- Bundesministerium für Verkehr (2012) *Luftverkehrs-Ordnung (LuftVO)* [online]. Available from: <http://www.gesetze-im-internet.de/bundesrecht/luftvo/gesamt.pdf> [Accessed 7/21/2014].
- Büttner, G., Kosztra, B., Maucha, G., & Pataki, R. (2012) *Implementation and achievements of CLC2006* [online]. Available from: <http://www.eea.europa.eu/data-and-maps/data/clc-2006-vector-data-version-2/> [Accessed 7/21/2014].

- Chen, L., & Roy, A. (2009) Event Detection from Flickr Data through Wavelet-based Spatial Analysis, *18th ACM conference on Information and Knowledge Management*.
- de Smith, M. J., Goodchild, M. F., & Longley, P. A. (2013) *Geospatial Analysis (4th ed.)*. Matador, Leicester.
- DeMaris, A. (2004) *Regression with Social Data: Modeling Continuous and Limited Response Variables*. Wiley & Sons, Hoboken, NJ.
- Elwood, S., Goodchild, M. F., & Sui, D. Z. (2012) Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice. *Annals of the Association of American Geographers*, vol. 102, pp. 571-590.
- Farlex (2014) *TheFreeDictionary* [online]. Available from: <http://www.thefreedictionary.com/drone> [Accessed 7/21/2014].
- Federal Aviation Administration (1981) *Advisory Circular 91-57: Model Aircraft Operating Standards* [online]. Available from: [http://rgl.faa.gov/Regulatory_and_Guidance_Library/rgAdvisoryCircular.nsf/0/1acfc3f689769a56862569e70077c9cc/\\$FILE/ATTBJMAC/ac91-57.pdf](http://rgl.faa.gov/Regulatory_and_Guidance_Library/rgAdvisoryCircular.nsf/0/1acfc3f689769a56862569e70077c9cc/$FILE/ATTBJMAC/ac91-57.pdf) [Accessed 7/21/2014].
- Federal Aviation Administration (2013) *Integration of Civil Unmanned Aircraft Systems (UAS) in the National Airspace System (NAS) Roadmap* [online]. Available from: http://www.faa.gov/about/initiatives/uas/media/UAS_Roadmap_2013.pdf [Accessed 7/21/2014].
- Garcia, Z. (2013) *What Flies When it Comes to Drone Laws Across the Globe* [online]. Available from: <http://www.missouridronejournalism.com/2013/04/what-flies-when-it-comes-to-drone-laws-across-the-globe/> [Accessed 7/21/2014].
- Girardin, F., Blat, J., Calabrese, F., Fiore, F. D., & Ratti, C. (2008) Digital Footprinting: Uncovering Tourists with User-Generated Content. *Pervasive Computing*, vol. 7, pp. 36-43.
- Goodchild, M. F. (2007) Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0 (Editorial). *International Journal of Spatial Data Infrastructures Research (IJS DIR)*, vol. 2, pp. 24-32.
- Hecht, B., & Gergle, D. (2010) On the “Localness” of User-Generated Content *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, ACM, New York, NY, pp. 229-232.
- Heipke, C. (2010) Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, pp. 550-557.
- Hochmair, H. H. (2010) Spatial Association of Geotagged Photos with Scenic Locations. In: Car A., Griesebner, G., & Strobl, J. (eds.) *Proceedings of the Geoinformatics Forum Salzburg*, Wichmann, Heidelberg, pp. 91-100.
- Hollenstein, L., & Purves, R. S. (2010) Exploring place through user-generated content: Using Flickr to describe city cores. *Journal of Spatial Information Science*, vol. 1, pp. 21-48.
- Ingraham, N. (2012) *Google Earth now includes publicly-sourced aerial images from balloons and kites* [online]. Available from: <http://www.theverge.com/2012/4/18/2957154/google-earth-balloon-kite-sourced-imagery> [Accessed 7/21/2014].
- Javanmardi, S., Ganjisaffar, Y., Lopes, C., & Baldi, P. (2009) User Contribution and Trust in Wikipedia, *5th International ICST Conference on Collaborative Computing: Networking, Applications, Worksharing*.
- Kulldorff, M. (1997) A spatial scan statistic. *Communications in Statistics - Theory and*

- Methods*, vol. 26, pp. 1481-1496.
- Kulldorff, M. (2014) *SaTScan User Guide for version 9.3* [online]. Available from: <http://www.satscan.org/> [Accessed 7/21/2014].
- Laliberte, A. S., & Rango, A. (2009) Texture and Scale in Object-Based Analysis of Subdecimeter Resolution Unmanned Aerial Vehicle (UAV) Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, pp. 761-770.
- Li, L., Goodchild, M. F., & Xu, B. (2013) Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science*, vol. 40, pp. 61-77.
- Li, Y., & Shan, J. (2013) Understanding the Spatio-Temporal Pattern of Tweets. *Photogrammetric Engineering & Remote Sensing*, vol. September 2013, pp. 769-773.
- Lin, Y., Hyypä, J., & Jaakkola, A. (2011) Mini-UAV-Borne LIDAR for Fine-Scale Mapping. *IEEE Geoscience and Remote Sensing Letters*, vol. 8, pp. 426 - 430.
- Masi, A. (2013) *In France, Drones Are All the Rage* [online]. Available from: <http://www.vocativ.com/world/france-world/france-drones-rage/> [Accessed 7/21/2014].
- Natural Earth (2014) *Natural Earth* [online]. Available from: <http://www.naturalearthdata.com/> [Accessed 7/21/2014].
- Neis, P., & Zipf, A. (2012) Analyzing the Contributor Activity of a Volunteered Geographic Information Project - The Case of OpenStreetMap. *ISPRS International Journal of Geo-Information* vol. 1, pp. 46-165
- Nielsen, J. (2006) *Participation Inequality: Encouraging More Users to Contribute* [online]. Available from: <http://www.nngroup.com/articles/participation-inequality/> [Accessed 7/21/2014].
- Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., & Mascolo, C. (2012) A Tale of Many Cities: Universal Patterns in Human Urban Mobility. *PLoS ONE*, vol. 7, pp. e37027.
- Public Lab (2013) *The Public Laboratory for Open Technology and Science: Balloon & Kite Mapping* [online]. Available from: <http://publiclab.org/wiki/balloon-mapping> [Accessed 7/21/2014].
- Remondino, F., Barazzetti, L., Nex, F., Scaioni, M., & Sarazzi, D. (2011) UAV photogrammetry for mapping and 3D modeling - Current status and future perspectives. In: Eisenbeiss H., Kunz, M., & Ingensand, H. (eds.) *ISPRS Archives - International Conference on Unmanned Aerial Vehicle in Geomatics (UAV-g)* (38-1/C22), ISPRS, Zurich, Switzerland, pp. 25-31.
- Reuter, H. I., Nelson, A., & Jarvis, A. (2007) An evaluation of void-filling interpolation methods for SRTM data. *International Journal of Geographic Information Science*, vol. 21, pp. 983–1008.
- Sagl, G., Resch, B., Hawelka, B., & Beinath, E. (2012) From Social Sensor Data to Collective Human Behaviour Patterns – Analysing and Visualising Spatio-Temporal Dynamics in Urban Environments. In: Jekel T., Car, A., Strobl, J., & Griesebner, G. (eds.) *GI_Forum 2012: Geovisualization, Society and Learning*, Wichmann, Berlin, pp. 54-63.
- Schlieder, C., & Matyas, C. (2009) Photographing a City: An Analysis of Place Concepts Based on Spatial Choices. *Spatial Cognition and Computation*, vol. 9, pp. 212-228.
- Sudekum, B. (2013) *Drone Imagery for OpenStreetMap* [online]. Available from: www.mapbox.com/blog/drone-imagery-openstreetmap/ [Accessed 7/21/2014].
- Tachikawa, T., Kaku, M., & Iwasaki, A. (2011) *ASTER GDEM Version 2 Validation*

- Report* [online]. Available from:
https://lpdaacaster.cr.usgs.gov/GDEM/Appendix_A_ERSDAC_GDEM2_validation_report.pdf [Accessed 5/6/2014].
- The Associated Press (2013) *Seattle mayor ends police drone efforts* [online]. Available from: <http://www.usatoday.com/story/news/nation/2013/02/07/seattle-police-drone-efforts/1900785/> [Accessed 7/21/2014].
- Torres-Sanchez, J., Lopez-Granados, F., De Castro, A. I., & Pena-Barragan, J. M. (2013) Configuration and Specifications of an Unmanned Aerial Vehicle (UAV) for Early Site Specific Weed Management. *PLoS ONE*, vol. 7, pp. e37027.
- UK Civil Aviation Authority (2012) *Unmanned Aircraft System Operations in UK Airspace – Guidance* [online]. Available from: www.caa.co.uk/docs/33/cap722.pdf [Accessed 5/6/2014].
- UNEP (2014) *The UNEP Environmental Data Explorer, as compiled from World Population Prospects, the 2012 Revision (WPP2012), United Nations Population Division. United Nations Environment Programme* [online]. Available from: <http://geodata.grid.unep.ch> [Accessed 7/21/2014].
- Watts, A. C., Ambrosia, V. G., & Hinkley, E. A. (2012) Unmanned Aircraft Systems in Remote Sensing and Scientific Research: Classification and Considerations of Use. *Remote Sensing* vol. 4, pp. 1671-1692.
- Zheng, Y.-T., Zha, Z.-J., & Chua, T.-S. (2012) Mining Travel Patterns from Geotagged Photos. *ACM Transactions on Intelligent Systems and Technology*, vol. 3.
- Zielstra, D., & Hochmair, H. H. (2011) A Comparative Study of Pedestrian Accessibility to Transit Stations Using Free and Proprietary Network Data. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2217, pp. 145-152.
- Zielstra, D., & Hochmair, H. H. (2012) Using Free and Proprietary Data to Compare Shortest-Path Lengths for Effective Pedestrian Routing in Street Networks. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2299, pp. 41-47.
- Zielstra, D., & Hochmair, H. H. (2013) Positional accuracy analysis of Flickr and Panoramio images for selected world regions. *Journal of Spatial Science*, vol. 58, pp. 251-273.

Table 1. Number of different countries with upload activity per user

All users		Users with >1 picture	
Nr. of users	Nr. of countries	Nr. of users	Nr. of countries
353	1	208	1
24	2	24	2
6	3	6	3
1	4	1	4
1	5	1	5
1	9	1	9

Table 2. Negative Binomial Regression Estimates for the global model

Parameter	Coeff.	95% Wald Confidence Limits		p
Intercept	-19.569	-24.190	-14.949	.000
Median age	.147	.061	.232	.001
High income [†]	3.967	1.297	6.638	.004
Upper middle income [†]	3.812	1.441	6.184	.002
Lower middle income [†]	.601	-1.864	3.067	.633
ln of population	.773	.545	1.001	.000
(Dispersion)	3.062	2.104	4.456	

[†]...Low income is the base category

Table 3. Negative Binomial Regression Estimates for two regional models

	Parameter	Coeff.	95% Wald Confidence Limits		p
Contiguous US	Intercept	-11.745	-12.100	-11.390	.000
	Ocean [†]	1.754	1.223	2.285	.000
	City [‡]	2.015	1.556	2.474	.000
	Easting (1000 km)	-.266	-.417	-.114	.001
	Forest x Ocean	-.819	-1.372	-.267	.005
	(Dispersion)	.858	.284	2.592	
Europe	Intercept	-9.927	-10.220	-9.634	.000
	Ocean [†]	.888	1.037	1.636	.000
	City [‡]	1.336	1.037	1.636	.000
	Easting (1000 km)	-.871	-1.077	-.665	.000
	Elevation (km)	-.931	-1.563	-.299	.004
	Lake x Elevation (km)	.781	.124	1.439	.020
	(Dispersion)	3.486	2.676	4.542	

[†]...Absence of ocean is the base category, [‡]...Absence of city is the base category
 Model: Offset = ln(area)

Table 4. Cluster information for the contiguous US based on the Bernoulli model

ID	Latitude	Longitude	r [km]	LL ratio	p	Obs. #	Exp. #	O./E.	Users
1	35.36	-120.85	370.10	67.40	0.000	90	23.58	3.82	14
2	25.90	-80.12	5.01	56.45	0.000	15	0.23	65.97	2
3	37.68	-77.89	0.00	40.47	0.000	9	0.10	87.96	1
4	36.95	-107.07	45.79	27.35	0.000	8	0.15	54.13	2
5	36.30	-106.06	167.98	27.31	0.000	18	1.74	10.35	4
6	27.78	-82.63	2.22	22.46	0.000	6	0.09	65.97	1
7	35.16	-106.62	7.32	21.24	0.000	6	0.10	58.64	1
8	37.24	-112.96	0.00	17.94	0.000	4	0.05	87.96	1
9	47.36	-114.22	14.37	14.15	0.005	4	0.07	58.64	1
10	44.88	-91.92	0.00	13.45	0.029	3	0.03	87.96	1
11	33.55	-111.91	7.55	12.44	0.043	4	0.09	43.98	1

Table 5. Cluster information for the contiguous US based on the discrete Poisson Model

ID	Latitude	Longitude	r [km]	Counties	LL ratio	p	Obs. #	Exp. #	O./E.	Users
1	33.03	-116.77	1154.77	186	101.96	0.000	135	38.77	3.48	24
2	37.72	-77.92	0.00	1	48.19	0.000	9	0.02	563.02	1
3	25.61	-80.50	0.00	1	23.17	0.000	17	1.84	9.25	3
4	44.95	-91.90	41.23	2	14.72	0.001	4	0.04	105.88	2
5	27.90	-82.74	0.00	1	12.58	0.010	8	0.67	11.86	1

Table 6. Cluster information for parts of Europe based on the Bernoulli model

ID	Latitude	Longitude	r [km]	LL ratio	p	Obs. #	Exp. #	O./E.	Users
1	48.19	0.65	594.17	546.97	0.000	669	199.46	3.35	135
2	45.76	21.25	4.82	81.92	0.000	31	1.70	18.20	1
3	53.57	7.90	16.96	59.40	0.000	25	1.56	16.05	1
4	43.55	11.59	12.21	26.18	0.000	11	0.68	16.15	1
5	53.10	12.90	11.82	24.21	0.000	8	0.39	20.55	2
6	60.41	15.38	25.49	21.18	0.000	7	0.34	20.55	1
7	47.91	11.84	4.51	18.15	0.000	6	0.29	20.55	1
8	55.30	12.37	27.22	15.22	0.004	7	0.49	14.39	3
9	41.94	12.78	21.31	15.17	0.005	10	1.12	8.93	2

Figures:

Las Vegas, Nevada, USA

LIKE! SHARE

Added by romesoh on 14 September 2013 [Follow romesoh](#)



Share on Facebook Share on Twitter

Mandalay Bay Pool Area

A Google Maps interface showing the location of the uploaded drone picture. The map is centered on Las Vegas, Nevada, with a red pin indicating the location of the Mandalay Bay Pool Area. The map shows major roads like I-15 and I-215, and various neighborhoods including Summerlin, Spring Valley, and Paradise. Below the map, there is a search bar with the text "Las Vegas Boulevard South" and a list of metadata for the location.

Showing 1 to 1 of 1 entries

- Camera : GoPro Hero 3 Black Edition
- Drone : DJI Phantom
- City : Las Vegas
- Street : Las Vegas Blvd
- County or State : Nevada
- Country : USA

POST STATS

972	0
VIEWS	COMMENTS
2	
LIKES	

RECENT POSTS

- Bedford House Cemetery, Ieper, Belgium
12 mins ago
15 0 0
- Bedford House Cemetery, Ieper, Belgium
14 mins ago
14 0 0
- Vevey
4 hours ago

Figure 1. Website for an uploaded drone picture

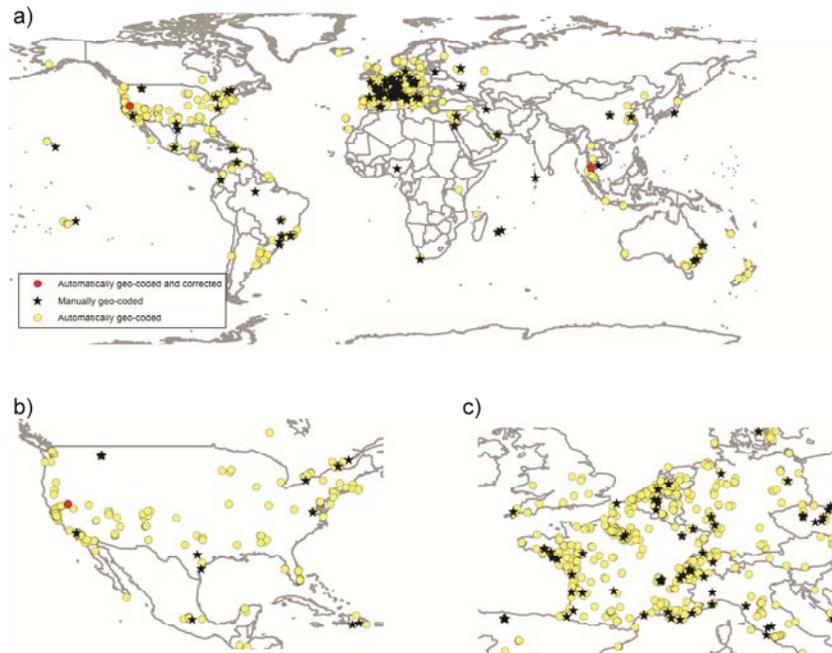


Figure 2. Location of mapped drone pictures worldwide (a), in the United States (b), and in central Europe (c)

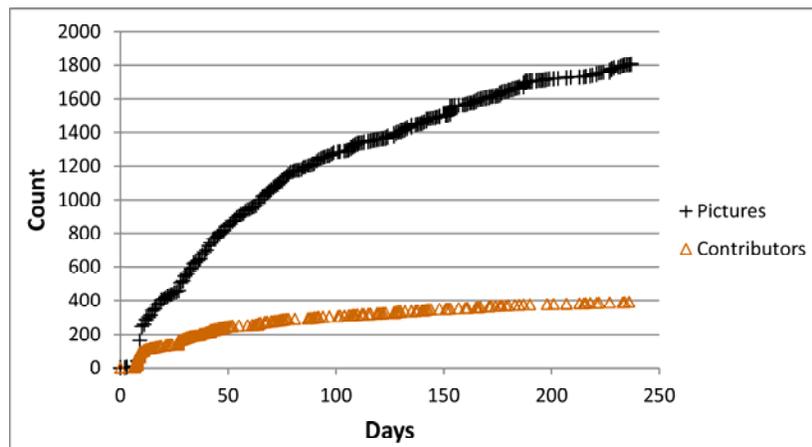


Figure 3. Number of contributors and uploaded pictures since the beginning of the project

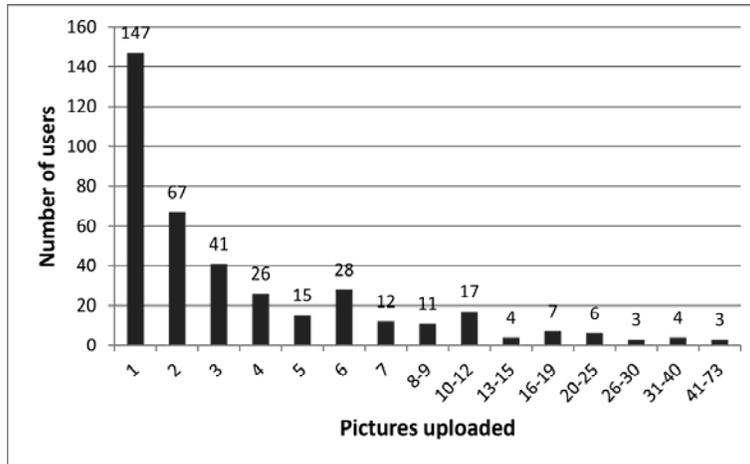


Figure 4. Picture uploads per user

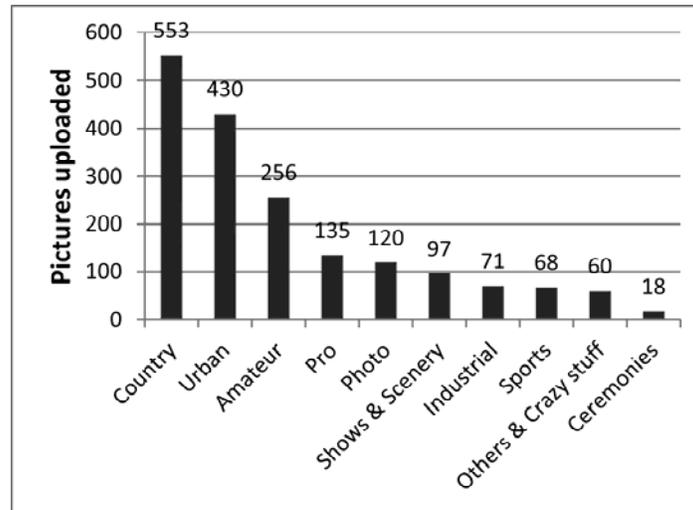


Figure 5. Number of contributions to each category

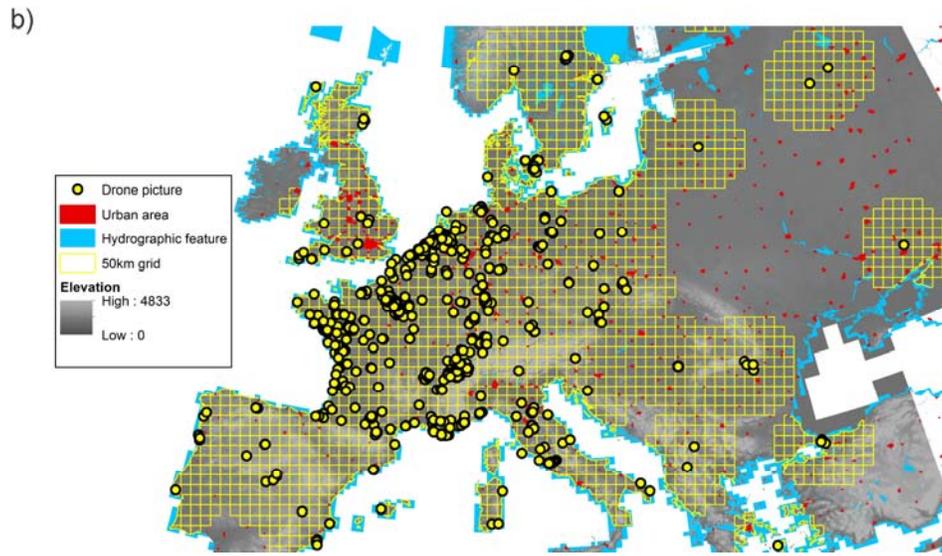
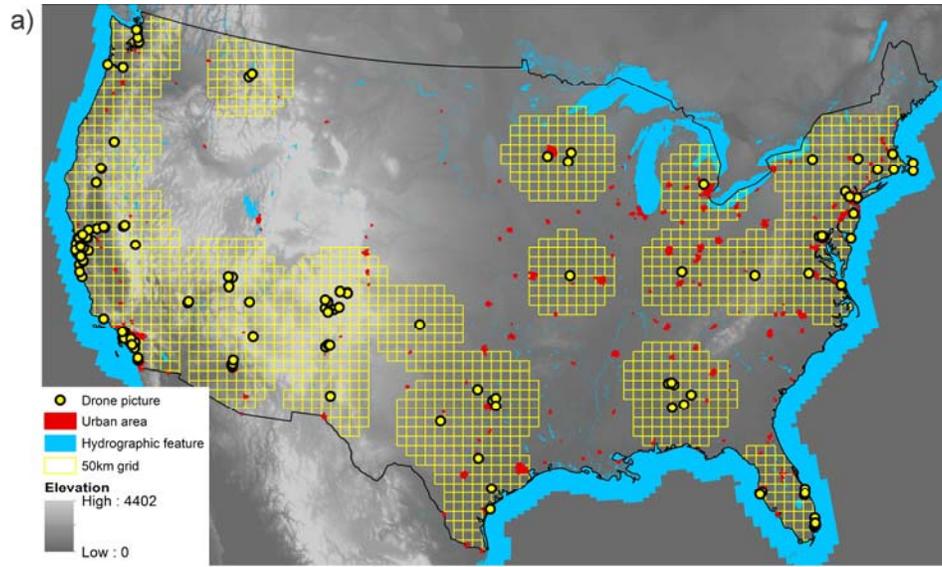


Figure 6. Grid cells and geographical features used in the United States (a) and Europe (b) regional models (elevation in m)

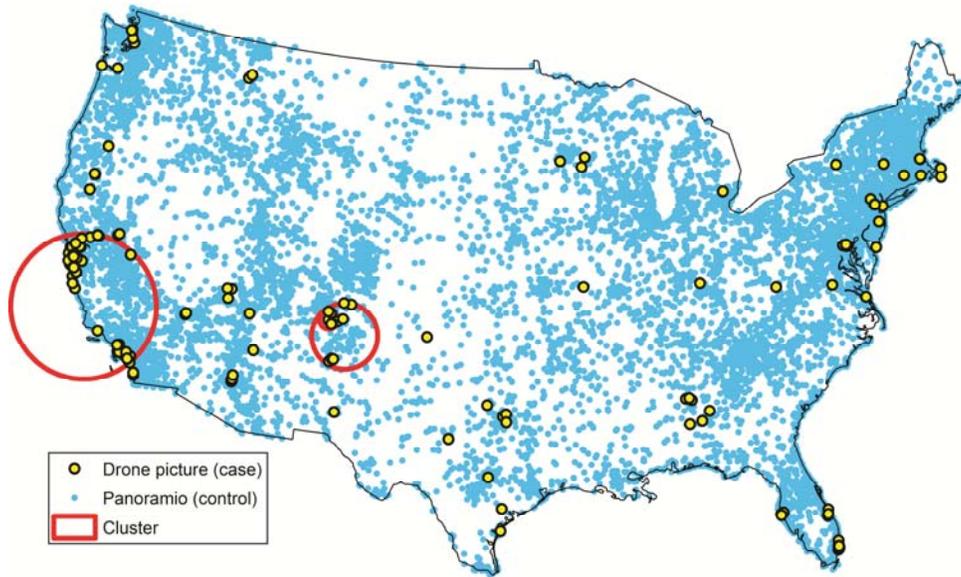


Figure 7. Clusters for the contiguous US based on the Bernoulli model

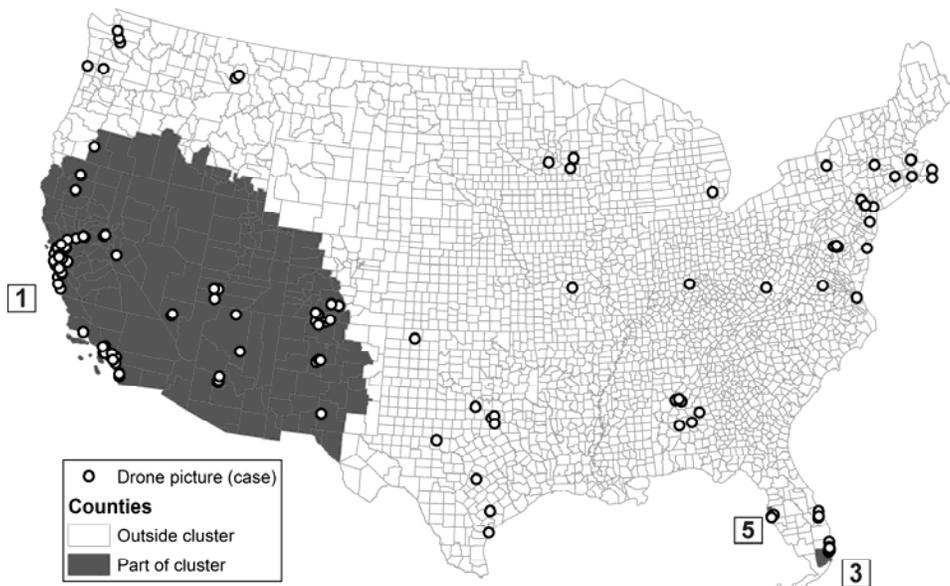


Figure 8. Clusters for the contiguous US based on the discrete Poisson model

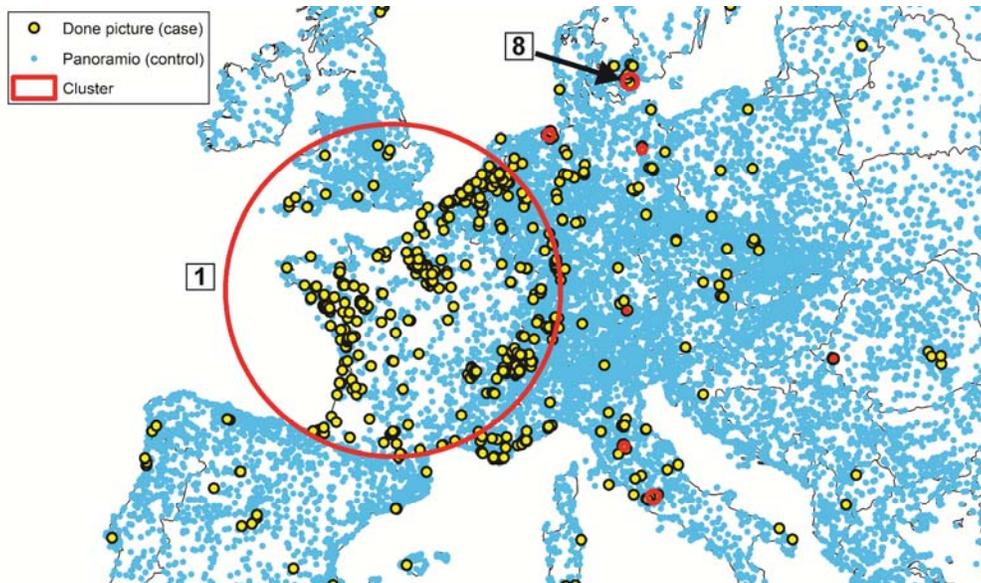


Figure 9. Clusters of contributed drone pictures for part of Europe