

Chapter 6

Mapping Human Impact Using Crowdsourcing

Linda See, Steffen Fritz, Christoph Perger, Christian Schill, Franziska Albrecht, Ian McCallum, Dmitry Schepaschenko, Marijn Van der Velde, Florian Kraxner, Ujjal Deka Baruah, Anup Saikia, Kuleswar Singh, Sergio de-Miguel, Rubul Hazarika, Ankita Sarkar, Abel Alan Marcarini, Mrinal Baruah, Dhruvajyoti Sahariah, Trishna Changkakati, and Michael Obersteiner

Abstract This chapter outlines how crowdsourcing and Google Earth have been used to create the first global crowdsourced map of human impact. Human impact in this context refers to the degree to which the landscape has been modified by humans as visible from satellite imagery on Google Earth. As human impact is measured on a continuum, it could be used to indicate the wildest areas on the Earth. This bottom-up approach to mapping using the crowd is in contrast to more traditional GIS-based wilderness mapping methods, which integrate proximity-based layers of remoteness and indicators of biophysical naturalness in a top-down manner. Data on human impact were collected via a number of different data collection campaigns using Geo-Wiki, a tool for visualization, crowdsourcing and validation of global land cover. An overview of the crowdsourced data is provided, along with the resulting map of human impact and a visual comparison with the map of human

L. See (✉) • S. Fritz • C. Perger • C. Schill • F. Albrecht • I. McCallum • D. Schepaschenko
M. Van der Velde • F. Kraxner • M. Obersteiner
International Institute of Applied Systems Analysis (IIASA), Laxenburg, Austria
e-mail: see@iiasa.ac.at

U.D. Baruah • A. Saikia • K. Singh • A. Sarkar • M. Baruah • D. Sahariah
T. Changkakati
Gauhati University, Guwahati, Assam, India

R. Hazarika
B. Borooah College, Guwahati, Assam, India

S. de-Miguel
Faculty of Science and Forestry, University of Eastern Finland, P.O. Box 111, 80101,
Joensuu, Finland

Departament de Producció Vegetal i Ciència Forestal, Universitat de Lleida-Agrotecnio
Center (UdL-Agrotecnio), Av. Rovira Roure, 191, E-25198 Lleida, Spain

A.A. Marcarini
AgroParis Tech, Nancy, France

footprint (Sanderson EW, Jaiteh M, Levy MA, et al. *BioSci* 52:891–904. doi:[10.1641/0006-3568\(2002\)052\[0891:THFATL\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2002)052[0891:THFATL]2.0.CO;2), 2002).

Keywords Wilderness • Human impact • Human footprint • Geo-Wiki • Crowdsourcing • Land cover

6.1 Introduction

The large and ever-increasing influence of humans on the Earth's ecosystems is acknowledged as an important driver of environmental change, particularly through continued deforestation, large-scale land acquisition and the expansion of agriculture. Areas of true wilderness are diminishing as humans encroach upon more of the Earth's surface, which has motivated the need to map the spatial distribution of wilderness or wild land areas globally, regionally and at the national level. As there is no agreed-upon definition for what constitutes wilderness (Applet et al. 2000), many of the attempts to map wilderness have combined multiple input layers that reflect two main concepts: (i) remoteness from human influence and (ii) naturalness, which were set out originally by Lesslie et al. (1993) as part of a comprehensive inventory of wilderness undertaken by the Australian Heritage Commission. In particular, the approach considered remoteness from settlements; remoteness from access, i.e. road and rail networks; naturalness of the landscape in terms of the degree to which it is free from buildings and other permanent structures such as electricity pylons; and biophysical naturalness, which is the degree to which an area is free from modern technological society, e.g. through disturbance of the vegetation (See Chap. 2). Other wilderness mapping inventories have used similar approaches, e.g. mapping the wilderness of the Arctic Barents region (Henry and Husby 1995). Modifications have also been made for area-specific studies, e.g. the addition of factors that characterize ruggedness or the physically challenging nature of the terrain in Scotland (Carver et al. 2012). In mapping the wilderness of the United States, population density was added as a way of characterizing solitude or remoteness from permanent inhabitants, the number of dams was considered to be a reflection of uncontrolled processes, and pollution and night time lights were used to develop indicators of naturalness (Applet et al. 2000). At the global level, an evaluation of wilderness was undertaken by the Sierra Club and the World Bank (McCloskey and Spalding 1989) while more recent efforts have attempted to reflect anthropogenic influences even more heavily, e.g. in the mapping of the human footprint by Sanderson et al. (2002) and by Ellis and Ramankutty (2008) in their classification of anthropogenic biomes, which is based on input layers of population density, land cover and land use.

This chapter presents an entirely different approach to mapping wilderness using the concept of human impact. With the help of Geo-Wiki, which is a visualization, crowdsourcing and validation tool for improving global land cover (Fritz et al. 2012), volunteers were asked to identify the degree of human impact (on a scale

from 0 to 100 %) which is visible from Google Earth imagery; the concept is explained in more detail in section 5.2. Through different Geo-Wiki crowdsourcing competitions, more than 150,000 samples of human impact were collected globally at more than 100,000 unique locations. An overview of this dataset is provided, which was then used to create a map of human impact using simple interpolation; this approach draws upon a simplified remoteness concept, i.e. distance to visible human influence. The implications and limitations of such an approach are discussed along with plans for future research.

6.2 The Concept of Human Impact

Theobald (2004) developed a human modification framework that characterizes landscapes based on two criteria, i.e. how natural versus artificial the landscape is and how free versus controlled the natural processes are that act upon a particular landscape. Wilderness is at the two extremes of these criteria, i.e. natural landscapes with little evidence of human influence such as settlements or roads, which are characterized by natural processes that are devoid of control, e.g. wild fires that are allowed to occur freely or a low density of dams within a watershed. At the opposite end of the two spectra are urban areas that are highly artificial where natural processes are heavily controlled, e.g. presence of flood defences, urban gardening and landscaping. In between these two extremes are different kinds of landscapes that have been modified by humans to differing degrees, e.g. croplands, rangelands, ex-urban areas, etc.

The framework of Theobald (2004) has been used to guide the concept of human impact as used in this research. Human impact in our context refers specifically to evidence of human modification of the landscape that can be seen from Google Earth imagery, captured as a value between 0 and 100 %. Table 6.1 provides an overview of the gradient of human impact where 0 % indicates no evidence of human activity and would be the wildest landscapes visible, 100 % would be urban

Table 6.1 Overview of human impact

Human impact	Description
0 %	No evidence of any human activity visible
1–50 %	Some visible evidence of human activities such as tracks/roads; evidence of managed forests; some evidence of deforestation; some scattered human dwellings, some scattered agricultural fields; some evidence of grazing
51–80 %	Increasing density of agriculture from subsistence on the lower end to intensive, commercial agriculture with large field sizes on the upper end
81–99 %	Urban areas with decreasing amounts of green space and increasing density of housing
100 %	A built up urban area with no green space, typically the business district of a city

Source: See et al. (2013)

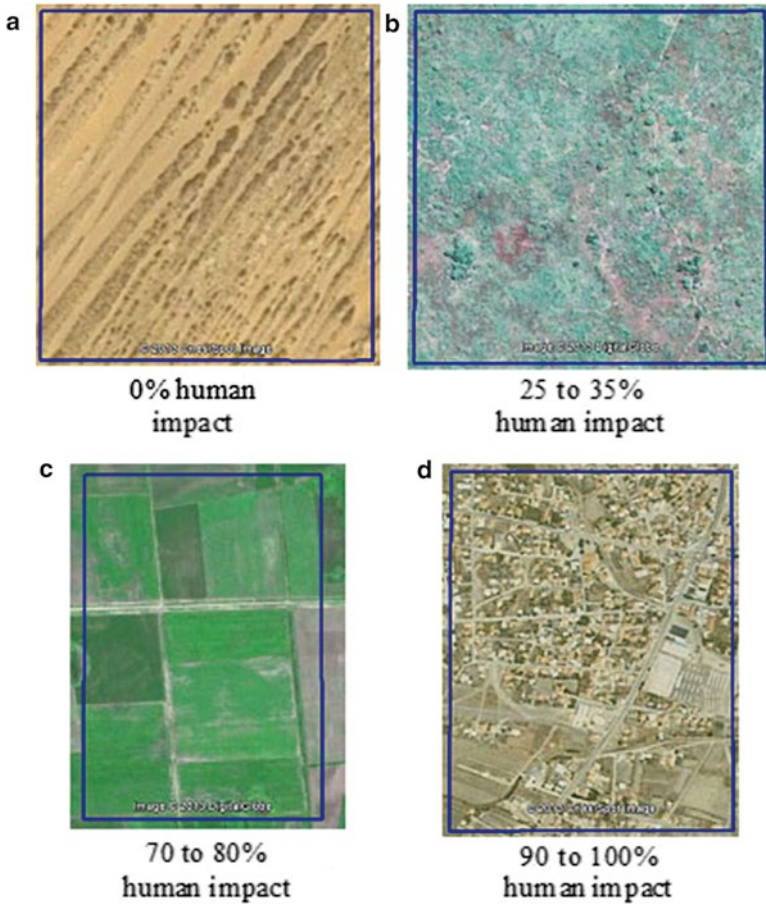


Fig. 6.1 Examples of different degrees of human impact as shown on Google Earth: **(a)** No evidence of human impact; **(b)** some evidence of human impact in an area of largely tree cover; **(c)** an intensively cultivated area with human impact between 70 and 80 %; and **(d)** an urban landscape between 90 and 100 % human impact

areas with no visible green space, and other types of modification would be located along this gradient. Managed forests, or forests with evidence of tracks and some deforestation, indicate increasing human impact followed by grazing and rangelands. Croplands are located in the upper half of the human impact scale depending on field size and intensity followed by urban areas, which have varying types of land use. Thus, suburban areas with green spaces have a lower human impact than fully built-up areas. These types of landscapes are characterized similarly by Theobald (2004) within his framework.

Figure 6.1 provides examples of different landscapes that are visible from Google Earth and their corresponding degree of human impact from an area of wilderness in a desert landscape to a highly built up urban area.

6.3 The Human Impact Dataset

This section describes how the data were collected using Geo-Wiki and then examines how human impact varies across different land cover types and by competition. A comparison of the crowdsourced data with a set of control values, i.e. a set of samples where experts agreed upon the value of the human impact in order to assess the quality of the crowd, is also presented.

6.3.1 *Collection of the Data Via Geo-Wiki*

The data on human impact (as well as land cover) were collected during four different Geo-Wiki campaigns where each one had a different theme or research question that drove the data collection competition and hence the geographical sampling of pixels on Google Earth. The first competition was driven by the need to validate a map of land availability for biofuel production (Perger et al. 2012; Fritz et al. 2013) while the second one was focused on sampling values from areas where global land cover maps currently disagree (Fritz et al. 2011). The third competition was aimed at specifically collecting human impact and land cover at points on the Earth that correspond to the same locations as those used to validate a new 30 m global land cover map¹ (FROM-GLC) produced by Tsinghua University (Gong et al. 2013) while the final competition was run at these same locations with the purpose of building up a robust crowdsourced dataset for the validation of land cover products more generally. Figure 6.2 shows the Geo-Wiki data collection interface from the third competition where users were asked to indicate the three main land cover types and the human impact within a 1 km pixel shown by the dark outline.

6.3.2 *An Overview of the Crowdsourced Data*

Over the four competitions, 151,942 validation points were collected (Table 6.2). In some competitions, the participants were provided with the same set of control points in order to monitor their performance. Thus there are a small number of locations for which many validations are available. In other competitions we wanted the same location to be validated at least twice. For these reasons the total number of unique locations for which there is information on human impact and land cover type is 103,509.

There is a highly skewed distribution in terms of the number of validation points contributed by the different users as shown in Fig. 6.3, i.e. of the 1500 registered Geo-Wiki users, only a small percentage contributed the vast majority of validations.

¹The land cover map can be downloaded from: <http://data.ess.tsinghua.edu.cn/>



Fig. 6.2 A screenshot from the third Geo-Wiki campaign to collect data on human impact in order to characterize wilderness extent. This particular competition included a tutorial at the start to help train participants

Table 6.2 Number of data points collected by each competition

Competition	Number of contributions	Number of unique locations
1	53,278	33,815
2	30,359	8571
3	32,861	27,814
4	35,444	33,309
Total	151,942	103,509

This is partly due to the nature of the prizes awarded, i.e. in most competitions the prize was co-authorship for the top 10 validators based on a combination of quantity and quality. Thus the number of volunteers per competition was on the order of 50 while in competition four, nine students were paid to collect the data.

In competition 1, the volunteers were asked to identify the dominant land cover type from one of 10 simple classes: (1) Tree cover; (2) Shrub cover; (3) Herbaceous vegetation/Grassland; (4) Cultivated and managed; (5) Mosaic of cultivated and managed/natural vegetation; (6) Flooded/wetland; (7) Urban; (8) Snow and ice; (9) Barren; and (10) Open Water, and to then assess the human impact across the entire 1 km² pixel. Feedback from competition 1 indicated that volunteers found it hard to determine a dominant land cover type in numerous cases and that specifying the percentage of different land cover types across the pixel would be easier. As a result the Geo-Wiki interface was changed and in competitions 2 to 4, volunteers entered up to 3 land cover types and the percentage of each one. Then for each of these different land cover types, a value for human impact was entered separately.

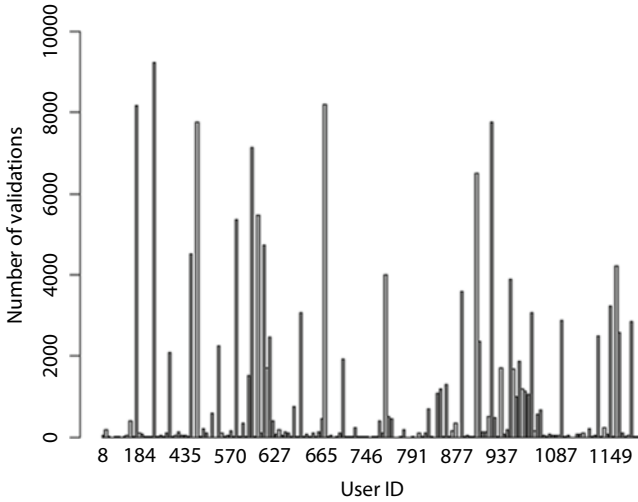


Fig. 6.3 Number of contributions by participant across all competitions

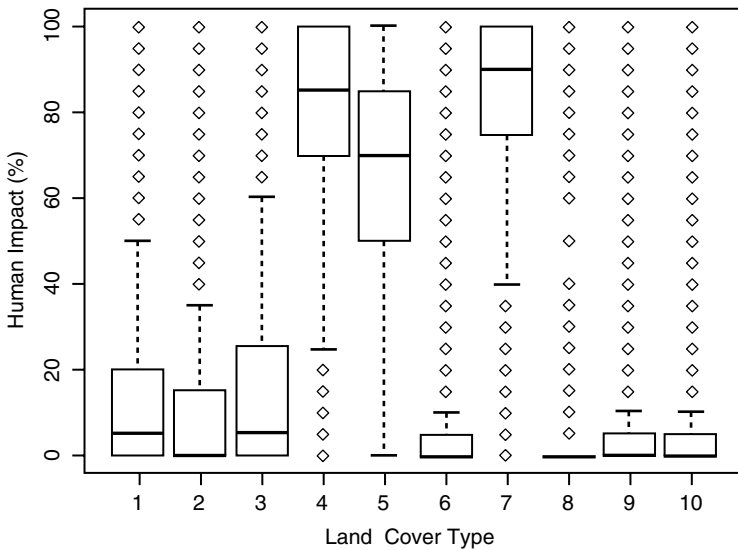


Fig. 6.4 Distribution of human impact across different land cover types: (1) Tree cover; (2) Shrub cover; (3) Herbaceous vegetation/Grassland; (4) Cultivated and managed; (5) Mosaic of cultivated and managed/natural vegetation; (6) Flooded/wetland; (7) Urban; (8) Snow and ice; (9) Barren; and (10) Open Water

The overall human impact for each validation point was then calculated as a weighted average of the individual human impact values based on the percentage of each land cover type occupying the pixel.

Figure 6.4 shows the human impact across each of the ten land cover types where the results are as expected, i.e. land cover types (4) Cultivated and managed, (5)

Mosaic of cultivated and managed/natural vegetation and (7) Urban are all on the higher end of human impact while the rest, which are generally more natural land cover types, have much lower values of human impact.

Manual checking of some of the outliers revealed examples of what were often very complex landscapes, particularly in competition 1 where participants were required to choose the dominant land cover type. The result was higher values of human impact than might have been expected for certain land cover types but which reflected human impact visible in these complex, mixed pixels.

The values of human impact at the locations of the crowdsourced data points were then compared to the values of the human footprint extracted from the map of Sanderson et al. (2002) at each point location. The result was a correlation coefficient of 0.56, which indicates that there is clearly some correspondence between the two approaches. In order to examine these differences, a crowdsourced map of human impact was created as outlined in the next section.

6.4 A Crowdsourced Map of Human Impact

A simple inverse distance weighted interpolation method was used to create the crowdsourced map of human impact. This interpolation method is based on Tobler's first law of geography, i.e. things that are close together are more related to one another than things further away (Tobler 1970). For each grid point to be interpolated, the algorithm identifies all the other points within a certain neighbourhood and calculates a weighted vector, w , based on a simple inverse power function:

$$w(d) = \frac{1}{d^x} \quad (6.1)$$

where d is the distance and x governs the rate of distance decay. Each interpolated point is then calculated as a weighted average of its neighbours. In this study the default values in ArcGIS were used, i.e. a power of 2 and a neighbourhood of 12 points. Although different settings and interpolation methods could be employed, the point was to demonstrate how a simple interpolation method can effectively be used to create a crowdsourced map of human impact, which is shown below in Fig. 6.5. No attempt was made at this point to experiment with different interpolation algorithms or the default settings in ArcGIS but this will be undertaken in further research using this dataset.

From this Fig. 6.5 one can see areas of high human impact in the agricultural belts of Canada, the USA and Brazil as well as the big cities on the western coast of the USA and in Mexico. Likewise there is high human impact across most of Europe, India, eastern China, and along the coastal fringes of Australia and North Africa, which reflects areas of high population density. However, looking across Africa as a continent, there is much less human impact evident. Where there are pockets of higher human impact, they clearly reflect locations of agricultural areas

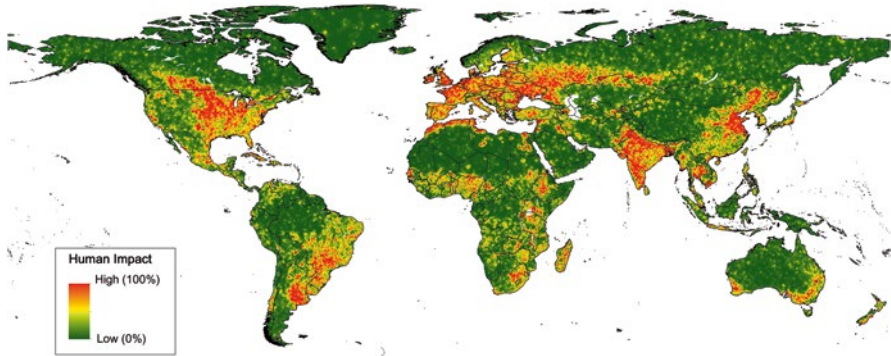


Fig. 6.5 The degree of human impact interpolated from pixels that were interpreted by the crowd using Google Earth

and urban centres. Interestingly, Madagascar shows considerable evidence of human impact throughout the island, which is in line with the relatively small amounts of rainforest left. Large areas of lower human impact coincide with deserts, large areas of tropical rainforest and the temperate forests of the northern latitudes as well as the tundra. Although this would be expected, these areas are also where the lowest resolution satellite imagery is available on Google Earth, i.e. the base Terrametrics imagery at a resolution of 15 m. Thus, evidence of human impact can be difficult to see clearly on some of these images. These areas of low human impact also show the artificial effects of the interpolation where the sampled areas are shown as small pockets across an area that looks otherwise to be devoid of human impact. With a much greater sample size, the visible effect of these artefacts would be minimized and a smoother transition would be produced. Despite these limitations, the crowd-sourced map appears to reflect an overall picture of human impact that conforms with areas of human habitation and activity.

These spatial patterns can be compared with the map of human footprint (Sanderson et al. 2002), which is also on a scale of 0 to 100 % and is shown below in Fig. 6.6.

The Sanderson map was developed using a top-down approach whereby global datasets were combined into a single indicator of human impact. Four categories of input data were used: gridded population density, land transformation (based on existing maps of land cover, built up areas and settlements), accessibility (based on access to roads, rivers and coastlines) and presence of electricity infrastructure (determined through remotely sensed data on night time lights). These layers were combined to produce a human influence index, which is based on the concept of ‘remoteness’ from existing anthropogenic influences. As outlined previously, this is a commonly used approach to determine areas of wilderness. Sanderson et al. (2002) then normalized the human influence index to account for the presence of different biomes and produced a spatial distribution of the human footprint on a scale from 0 to 100.

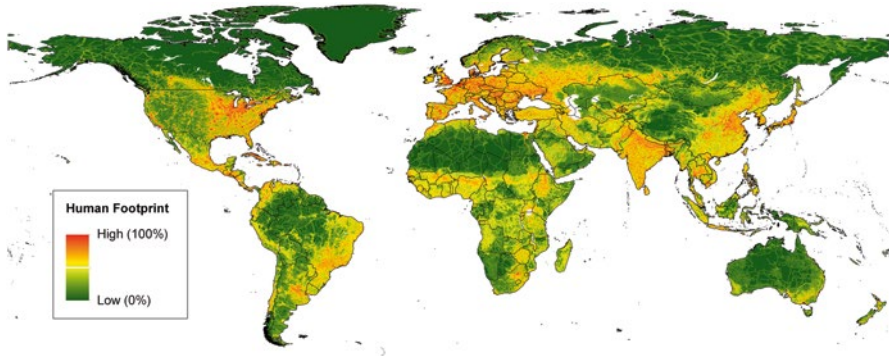


Fig. 6.6 Map of human footprint produced by Sanderson et al. (2002)

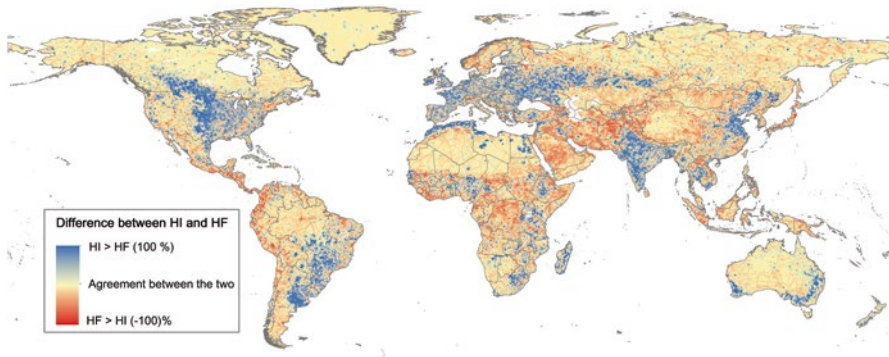


Fig. 6.7 Difference between crowdsourced map of human impact and the map of human footprint from Sanderson et al. (2002)

The maps in Figs. 6.5 and 6.6 have the same scales so it is immediately clear that the amount of human impact in the Sanderson et al. map (2002) is much lower than the crowdsourced map and that the areas with highest human impact are less widespread. However, there is a clearly an overall agreement between the two maps with many of the agricultural and urban areas corresponding to the areas with highest human impact. In contrast, because of the approach used by Sanderson et al. (2002), the map of human footprint shows road patterns and national/sub-national borders as a result of some of the input datasets used, unlike the much smoother pattern shown by the crowdsourced map. These border effects are a result of the input datasets that they used.

In order to compare these two different approaches in a more quantitative way, the maps were subtracted from one another to create a difference image as shown in Fig. 6.7. The blue shading shows areas where the crowdsourced map shows higher human impact compared to the human footprint, yellow denotes areas of agreement while red areas show where the map of human footprint indicates areas of greater human impact than the crowdsourced map.

Figure 6.7 shows that there are large differences primarily in areas of agriculture where there are higher values of human impact using the crowdsourced approach compared to the human footprint. Since the human footprint is based partly on gridded population density, there may be many areas where density is low in rural areas but Google Earth shows a different picture. The presence of road networks and borders also show up as visible differences between the two images. Moreover, areas of Africa, China and the Middle East all have a higher human impact according to Sanderson et al. (2002) compared to the crowdsourced map. Google images suggest landscapes that are less influenced by humans than that which results from a more top-down methodology. Some of these differences may arise because of the fact that the two approaches may actually be measuring slightly different concepts, e.g. assessing human impact via Google Earth considers the direct impact of man-made features on the Earth's surface but it does not directly take remoteness from anthropogenic features into account, as with more traditional wilderness mapping.

Each approach to mapping human impact clearly has advantages and disadvantages. In the case of the crowdsourced approach, the data can be collected very easily but this approach relies on only a sample of data, where some areas may need a much denser representation to accurately reflect human impact. Scaling up this approach to produce a truly representative map of human impact may require far more data than have currently been collected. Moreover, the data have been collected at a resolution of 1 square km. As this resolution contains many heterogeneous pixels of mixed land cover types, determination of human impact is complicated. Increasing the resolution of the sample may improve the ease with which human impact can be identified although this will result in a trade-off in terms of how much data can be collected.

There are also issues which arise regarding the quality of crowdsourced data. Although control points of known human impact have been used to determine overall quality, this applies only to a very small number of points. More systematic methods of bias correction and more interactive and ongoing crowd training need to be incorporated into future crowdsourcing campaigns. The temporal element of Google Earth images is another issue, where images are available from different time periods. However, we have started to collect this information so that we can use it to filter out data from old images or provide some indication of certainty based on the currency of the data.

In contrast the map of human footprint was created using global datasets that have comprehensive spatial coverage. Yet Sanderson et al. (2002) clearly acknowledge that there are potential problems with these datasets. Land cover data have been shown to have high spatial uncertainties and accuracies that even today are still only between 65 and 75 % (Fritz et al. 2011; Gong et al. 2013). There are also issues with all of the other datasets used, as pointed out by Sanderson et al. (2002), e.g. potential incompleteness of the road data, and gridded population data have problems with accuracy and representation, particularly in rural areas. The input datasets were combined without considering weighting as there is no guidance to indicate whether one factor is more important than another, and a sensitivity analysis was not carried out. Finally, how can you really validate the results coming out of such a

top-down approach? Despite these limitations, Sanderson et al. (2002) are quick to point out that the map of human footprint is much too inexact for direct conservation purposes, and that what the map serves to do is illustrate the global picture of our current human footprint.

6.5 Conclusions

This chapter has outlined how data collected through the Geo-Wiki crowdsourcing tool has been used to create a map of human impact, which could be used as a spatial indicator of wilderness in a similar way to that undertaken by Sanderson et al. (2002) in identifying the 10 % wildest areas on Earth. As more crowdsourced data are collected in future Geo-Wiki campaigns and a denser network of points becomes available, we can continue to improve the crowdsourced map of human impact in the future.

This bottom-up approach represents a very different way of creating a map of human impact, which is normally derived through the top-down combination of various input datasets that reflect remoteness and/or biophysical naturalness such as the map of human footprint created by Sanderson et al. (2002). However, both bottom-up and top-down approaches have advantages and disadvantages. One way forward may be combining both approaches. For example, simple interpolation was used to illustrate how the map could be produced but there are better interpolation methods available that can use additional input data layers, e.g. road networks, population density, etc. to help guide the spatial allocation of human impact. Another approach would be to combine the global input data layers used by Sanderson et al. (2002) – although updated with the most current products available – in combination with the crowdsourced map of human impact in a type of multi-criteria hybrid approach. These types of approaches will be investigated as part of future work.

Acknowledgements This research was supported by the Austrian Research Funding Agency (FFG) via the LandSpotting (No. 828332) and FarmSupport (No. 833421) projects.

References

- Applet, G., Thomson, J., & Wilbert, M. (2000). Indicators of wilderness: Using attributes of the land to assess the context of wilderness. *USDA forest service proceedings*, USDA, pp. 89–98.
- Carver, S., Comber, A., McMorran, R., & Nutter, S. (2012). A GIS model for mapping spatial patterns and distribution of wild land in Scotland. *Landscape and Urban Planning*, *104*, 395–409. doi:10.1016/j.landurbplan.2011.11.016.
- Ellis, E. C., & Ramankutty, N. (2008). Putting people in the map: Anthropogenic biomes of the world. *Frontiers in Ecology and the Environment*, *6*, 439–447. doi:10.1890/070062.
- Fritz, S., See, L., McCallum, I., et al. (2011). Highlighting continued uncertainty in global land cover maps for the user community. *Environmental Research Letters*, *6*, 044005. doi:10.1088/1748-9326/6/4/044005.

- Fritz, S., McCallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D., van der Velde, M., Kraxner, F., & Obersteiner, M. (2012). Geo-Wiki: An online platform for land cover validation and the improvement of global land cover. *Environmental Modelling and Software*, *31*, 110–123.
- Fritz, S., See, L., van der Velde, M., et al. (2013). Downgrading recent estimates of land available for biofuel production. *Environmental Science and Technology*, *47*, 1688–1694. doi:[10.1021/es303141h](https://doi.org/10.1021/es303141h).
- Gong, P., Wang, J., Yu, L., et al. (2013). Finer resolution observation and monitoring of global land cover: First mapping results with landsat TM and ETM+ data. *International Journal of Remote Sensing*, *34*, 2607–2654. doi:[10.1080/01431161.2012.748992](https://doi.org/10.1080/01431161.2012.748992).
- Henry, D., & Husby, E. (1995). *Wilderness quality mapping in the Euro-Arctic Barents region: A potential management tool*. Northern wilderness areas: Ecology, sustainability (pp. 319–339). Rovaniemi: University of Lapland Arctic Centre Publications, University of Lapland, Arctic Centre.
- Lesslie, R., Taylor, D., & Maslen, M. (1993). *National wilderness inventory. Handbook of principles, procedures and usage*. Canberra: Australian Heritage Commissions.
- McCloskey, J. M., & Spalding, H. (1989). A reconnaissance level inventory of the amount of wilderness remaining in the world. *Ambio*, *18*, 221–227.
- Perger, C., Fritz, S., See, L., et al. (2012). A campaign to collect volunteered geographic information on land cover and human impact. In T. Jekel, A. Car, J. Strobl, & G. Griesebner (Eds.), *GI_forum 2012: Geovisualisation, society and learning* (pp. 83–91). Berlin/Offenbach: Herbert Wichmann Verlag.
- Sanderson, E. W., Jaiteh, M., Levy, M. A., et al. (2002). The human footprint and the last of the wild. *BioScience*, *52*, 891–904. doi:[10.1641/0006-3568\(2002\)052\[0891:THFATL\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2002)052[0891:THFATL]2.0.CO;2).
- See, L., Comber, A., Salk, C., et al. (2013). Comparing the quality of crowdsourced data contributed by expert and non-experts. *PLoS ONE*, *8*, e69958. doi:[10.1371/journal.pone.0069958](https://doi.org/10.1371/journal.pone.0069958).
- Theobald, D. M. (2004). Placing exurban land-use change in a human modification framework. *Frontiers in Ecology and the Environment*, *2*, 139–144. doi:[10.1890/1540-9295\(2004\)002\[0139:PELCIA\]2.0.CO;2](https://doi.org/10.1890/1540-9295(2004)002[0139:PELCIA]2.0.CO;2).
- Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*, 234–240.