Mapping Land Use Dynamics Using the Collective Power of the Crowd

Christoph Aubrecht, Joachim Ungar, Dilek Ozceylan Aubrecht, Sérgio Freire, and Klaus Steinnocher

Concept and Context

Traditional land use and land cover (LULC) mapping has long relied strongly on input from Earth Observation (EO) data sources at various resolutions and scale levels. With high performance and cloud computing on the rise, rapid processing of large volumes of very high resolution (VHR) satellite imagery—big EO data—is becoming less problematic. Consequently, scientific challenges in that topical domain move on to the next level.

"Remote Sensing Science 2.0" has been coined as having a primary focus on the advanced consideration of temporal scale, i.e. change and dynamics, in addition to the traditional spatial aspects (Herold 2011). The recent emergence of EO *satellite* constellations (some already operational, some in the planning stage) set up in temporally shifted coplanar orbits (e.g. ESA's Sentinels, Airbus' Pléiades, Planet's RapidEye, UrtheCast's OptiSAR) as well as smallsat and nanosat swarms (e.g. Planet's Doves, BlackSky's Pathfinders, Terra Bella's SkySats) comes in line with

C. Aubrecht (✉)
ESA/ESRIN & World Bank, Washington, DC, USA
e-mail: caubrecht@worldbank.org; christoph.aubrecht@esa.int

J. Ungar
EOX IT Services GmbH, Vienna, Austria

D. Ozceylan Aubrecht
Management Information Systems Department, Sakarya University, Sakarya, Turkey

S. Freire
Disaster Risk Management Unit, Joint Research Centre (JRC), European Commission, Ispra, Italy

K. Steinnocher
Energy Department, AIT Austrian Institute of Technology, Vienna, Austria

increased awareness and efforts to tackle the space-time resolution dichotomy of traditional space-based Earth Observation and related analytics. Highly frequent revisit times and acquisition intervals associated with such constellations guarantee the availability of data at unprecedented spatio-temporal resolution, opening up a wealth of new possibilities for dynamic mapping of environmental changes (Aubrecht et al. 2017a). Complementing space-based data, unmanned aerial vehicles (UAVs), also commonly referred to as drones, provide similar rapid on-demand high-quality data acquisition capability in the *airborne* domain. Benefits include below-cloud operation, thus decreased weather-dependency in that regard, as well as low cost and VHR imagery (up to 5 cm). With fast-pace technological developments in recent years drone data has left the exclusiveness of military applications and can now rather easily be created and made accessible by and to the general public (Tiwari and Dixit 2015). Drone mapping is particularly addressing the on-demand application market, with continuous long-term change monitoring certainly remaining in the satellite EO domain.

Building upon this new applied scientific and more widespread appreciation of spatio-temporal analytics, traditional "static" methods of LULC mapping seem to be in urgent need of a conceptual revisit. Both *timeliness* (i.e. long processing and release cycles, irregular updating) and inherent *temporal granularity* (i.e. land use being commonly represented as simplified static non-dynamic attributive data) are clearly restricted in traditional approaches. Technological advancements and innovations in remotely sensed EO and data acquisition can provide improved input for highly frequent analysis of changing physical features of the Earth (i.e. referring to *land cover*), thus well-addressing the timeliness factor. Understanding *land use* and its dynamic variations, however, requires additional contextual information. This is particularly relevant when analyzing urban areas that feature highly variable patterns of human activities where remote sensing is restricted to providing identification of built-up and basic physical characteristics with limited thematic context and human usage variation.

Major innovation in LULC assessment therefore now needs to happen through information integration rather than via incremental improvements at technical level. Volunteered Geographic Information (VGI), i.e. data created and shared by anybody anywhere at any time, has the potential to cover the full dynamic range of human activities spatio-temporally linked to their respective location patterns. This paper takes the discussion on advancing LULC mapping one step further from the traditional schematic remote sensing-based approach by introducing VGI as ancillary data source for integrative dynamic analytics. The presented concepts illustrate the way forward with regard to advanced spatio-temporal decision support for a variety of applications stretching across the traditional LULC domain including improved population dynamics modeling, fine-scale human-adjusted "smart" urban development, and real-time disaster risk reduction.

The Collective Power of the Crowd

LULC mapping is one of the oldest and best established fields in the wider scope of EO and remote sensing research. It draws upon several decades of scientific studies and developments, with the "modern" era of land cover mapping starting in the early 1970s, in line with the rise of civil space-based remote sensing (Loveland 2012). The basic understanding of what is considered adequate output of a LULC model has not changed much ever since however. It is still mainly referring to characterizing use and/or land cover classes at one defined point in time, i.e. a static snapshot and abstract simplification of reality.

In particular in urban settings, changes of physical features commonly occur at a faster pace than the usual map update cycles. Advanced remote sensing and processing technology now has the power to catch up with this speed and provide near-real time coverage of the physical environment. Even more dynamic, however, are *human activity patterns* that form the backbone for a progressive understanding of space and place, closer to reality, which is essential in making customized decisions in the increasingly relevant smart city context. The emerging research domain of integrating big data, urban informatics, and associated geo-analytics (Thakuriah et al. 2017) aims at addressing these issues and fostering innovation through joint use of human and technical sensors (in situ and remote) to reveal previously un-identifiable information in near real time.

VGI has made a substantial impact on how data is understood, addressed, and analyzed in the geospatial domain since it was first conceptually highlighted about a decade ago (Goodchild et al. 2017). It is no longer an exotic prospect but has become a widely recognized additional stream of geospatial information opening up new research fields and application areas. While the scientific use of VGI and crowd sourcing is certainly still increasing, it is far from being mainstream in most traditional geographic application domains. Focus is mostly put on individual case studies and often VGI is seen as an add-on to potentially enhance established processes. Conceptual innovation and disruptive thinking is still rare and limited.

Under the umbrella of the EU FRESHER project we evaluated how *volunteered geo-dynamic information* (labeled VGDI) can contribute to exposure assessment for health risk analyses. The dynamic component was thereby particularly highlighted as innovative input factor for spatio-temporal population modeling as well as time-dependent urban zone characterization (Aubrecht and Steinnocher 2016). The Greater Lisbon area (Portugal) served as one of the test study areas. We used data from the location-based social network platform Foursquare/Swarm to illustrate the collective power of the crowd inherent in VGDI. In this article, we now focus on underlying aspects of data access and integration and conclude with an outlook to where innovation is driving future LULC mapping concepts.

VGDI Access and Integration: The Foursquare Use Case

There are various ways of categorizing VGI. One approach is to assess whether the data is implicitly or explicitly collective (Aubrecht et al. 2017b). Implicitly collective in that regard would signify that contributing users provide certain pieces of information for personal purposes (such as connecting with friends) but without the main intention of creating a "bigger picture". In contrast to the motivation of its individual users, the underlying data infrastructure and curation practice are precisely designed for that purpose, i.e. for understanding interrelations and deriving insight from collective patterns. Implicitly collective information is therefore usually considered the most valuable asset of service providers like Google, Facebook, or Foursquare. For this reason, public access to such databases in full detail is commonly technically limited as well as legally restricted. For effective integration in LULC mapping the collective input is required, which implies the need to identify efficient ways to access and compile this type of data.

Foursquare maintains an API for app developers which allows gathering information about specific *venues* (location of facilities where activities take place) and—if the user grants access—user interaction. For our case study of analyzing human activity dynamics and use patterns (1) a georeferenced list of all venues in the area of interest and (2) temporal information about the number of users checked in at each venue are required. The Foursquare API does not support bulk data download for a defined geographic area. It does, however, enable caching some data in restricted form for certain applications. Using a latitude/longitude coordinate feed for the API call, venues surrounding these coordinate pairs are returned together with categorical metadata and time-specific current user counts (i.e. users checked in at that venue at the time of the request). The number of venues returned in that process is limited to a maximum of 30. Furthermore, also the number of requests per developer are restricted to 5000 per hour in order to preclude large scale scraping.

To comprehensively extract all venue data for the area of interest numerous requests need to be performed in parallel, each returning the maximum of 30 entries. When implementing evenly distributed quadratic request point grids, dense activity areas are insufficiently covered (i.e. more venues exist than can be extracted) while output in more remote areas shows largely redundant information (i.e. neighbouring request points returning identical venues). With the maximum number of requests per hour restricted to 5000 (preventing simple request grid densification) the spatial distribution of request points needs to be optimized, thus minimizing return duplicates while maximizing spatial reach. We developed a novel extraction method[1] using hexagonal request grids to optimize spatial coverage. Implementing a system similar to the R-tree method (Guttman 1984), grid density is adapted in iterative steps to account for the spatially uneven distribution of venues. Additional request points are thereby placed in dense activity areas by creating smaller hexagons

---

[1]Code available on GitHub https://github.com/ungarj/fsq_crawler (02/16/2017)

around every relevant point from the prior iteration (using half distance and a 30°
rotation). If all extracted venues (up to 30) for a certain request point were already
covered by prior requests, hexagon densification around that point is stopped for the
next iteration. For further optimization, low profile venues with less than five total
check-ins are omitted for insignificance throughout the process.

For the Lisbon study area we extracted approximately 40,000 venues from
about 4000 request points using the described method. During testing, however,
the effective number of successful requests per hour proved to actually be lower,
likely due to performance restrictions imposed by the API. We therefore conducted
the data harvesting in two parallel processes, each using its own developer account.
Outputs of the two processes were then combined for the final compilation.

To integrate the extracted human activity information into a dynamic land use
model we created spatially interpolated thematic time-dependent activity surfaces
(see Fig. 1 for aggregated day and night activity in a generalized "events" category
including restaurants, bars, and nightlife spots). These activity surfaces can then
subsequently be integrated with EO based land cover maps, thus refining human use
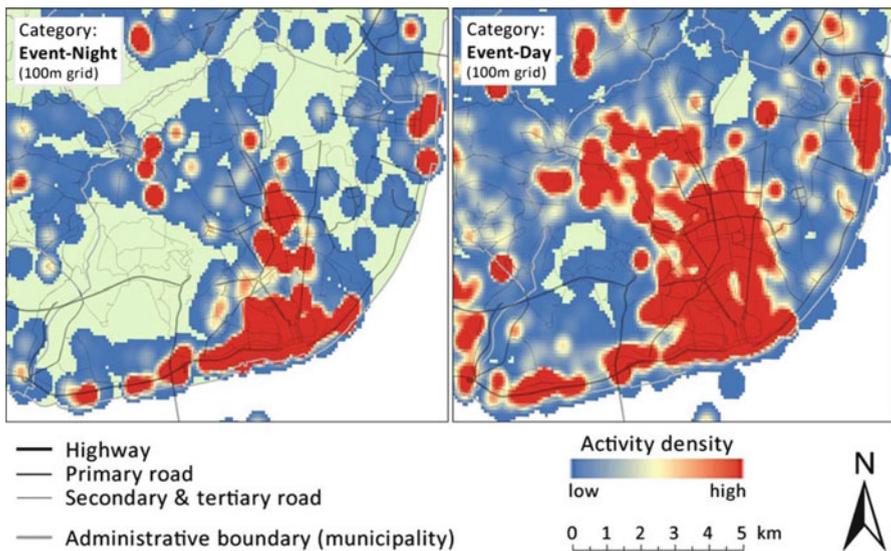categories beyond the visible physical parameters.

Fig. 1 Time-dependent activity surfaces for Lisbon (Portugal), spatially interpolated from 2014
Foursquare *venue* and user *check-in* data

What Is Next and Where Are We Headed?

Integrating VGI-based dynamic human activity data into land use mapping constitutes a major refinement over traditional static approaches (Jiang et al. 2015). While this chapter focuses on the human activity component, near real-time land use identification would in parallel require dynamic monitoring of the physical environment. Much reduced revisit and data acquisition times of novel satellite constellations and continuous data flow allow quick access while cloud processing environments such as the Amazon Web Services, the Google Cloud, or Microsoft Azure enable on-the-fly identification of rapid changes of the environment.

A major constraint of location-based social network (LBSN) data and VGI in general is their varying representativeness and partly low and uneven penetration rate across different sectors of society and regions. In Foursquare, while the completeness of the venue database is not considered to be affected, this is particularly relevant for venue-specific temporal considerations. Spatio-temporal integration of per se unrelated social media feeds like Facebook, Twitter and Instagram has potential to increase the reach in terms of social profiling and reduce bias effects. Service providers, however, are very hesitant in allowing such operations. Foursquare, for example, explicitly prohibits "*combin[ing] or aggregate[ing] Foursquare location information with location data from other sources*" (Foursquare 2017).

A glimpse at the future in terms of contextual real-time location tracking is provided by the recently activated live service of Google Maps "Popular Times". Google uses aggregated and anonymized data from users who have opted in to Google Location History to determine temporal usage profiles of specific locations. The new service feature integrates live visit data thus indicating relative current popularity compared to the usual trend (Aspinall 2016).

Combining live human activity patterns not impaired by LBSN penetration constraints with real-time mapping of land cover dynamics would imply the transition from static snapshots via time-dependent models to observation of current and ongoing activity flows and eventually predictive analytics of future patterns.

References

Aspinall J (2016) Know before you go, with Google. Google Products Blog. https://blog.google/products/search/know-you-go-google/. Accessed 29 Jan 2017

Aubrecht C, Meier P, Taubenböck H (2017a) Speeding up the clock in remote sensing – identifying the 'black spots' in exposure dynamics by capitalizing on the full spectrum of joint high spatial and temporal resolution. Nat Hazards 86(1): 177 – 182

Aubrecht C, Özceylan Aubrecht D, Ungar J, Freire S, Steinnocher K (2017b) VGDI - advancing the concept: volunteered geo-dynamic information and its benefits for population dynamics modeling. Trans GIS 21(2):253–276

Aubrecht C, Steinnocher K (2016) Volunteered geo-dynamic information for health-risk exposure assessment – a FRESHER case study. GI_Forum J Geogr Inf Sci 2:157–163

Foursquare (2017) Foursquare for Developers. Venues Service – Venues Database Usage Rules. https://developer.foursquare.com/overview/venues.html. Accessed 29 Jan 2017

Goodchild MF, Aubrecht C, Bhaduri B (2017) New questions and a changing focus in advanced VGI research. Trans GIS 21(2):189–190

Guttman A (1984) R-trees: a dynamic index structure for spatial searching. Proceedings of the 1984 ACM SIGMOD international conference on management of data, pp 47–57

Herold M (2011) Remote Sensing Science 2.0. Inaugural lecture, Professor of Geo-information Science, Wageningen University, 09/08/2011

Jiang S, Alves A, Rodrigues F, Ferreira J Jr, Pereira FC (2015) Mining point-of-interest data from social networks for urban land use classification and disaggregation. Comput Environ Urban Syst 53:36–46

Loveland T (2012) History of land-cover mapping. In: Giri CP (ed) Remote sensing of land use and land cover, principles and applications. CRC Press, Boca Raton, pp 13–22

Thakuriah PV, Tilahun N, Zellner M (2017) Seeing cities through big data - research, methods and applications in urban informatics. Springer, Cham, Switzerland. 559 pp

Tiwari A, Dixit A (2015) Unmanned aerial vehicle and geospatial technology pushing the limits of development. Am J Eng Res 4(1):16–21