Mapathons versus automated feature extraction: a comparative analysis for strengthening immunization microplanning

Amalia Mendes (opf9@cdc.gov)
Agency for Toxic Substances and Disease Registry https://orcid.org/0000-0003-1020-8633
Tess Palmer
Agency for Toxic Substances and Disease Registry
Andrew Berens
Agency for Toxic Substances and Disease Registry
Julie Espey
Agency for Toxic Substances and Disease Registry
Rhiannan Price
Maxar Technologies
Apoorva Mallya
Gates Foundation: Bill & Melinda Gates Foundation
Sidney Brown
Gates Foundation: Bill & Melinda Gates Foundation
Maureen Martinez
Centers for Disease Control and Prevention
Noha Farag
Centers for Disease Control and Prevention
Brian Kaplan
Agency for Toxic Substances and Disease Registry

Research Article

Keywords: Feature extraction, mapathon, essential immunization, population estimates, microplanning, satellite imagery, building footprints

Posted Date: March 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-312199/v1

Abstract

Background

The barriers to global essential immunization (EI) coverage are increasingly due to social instability and logistical factors like the displacement of vulnerable populations, the difficulty of accessing these populations, and the lack of geographic information for hard-to-reach areas. Microplanning, a population-based, healthcare intervention planning method has begun to leverage geographic information system (GIS) technology and geospatial methods to improve the remote identification and mapping of vulnerable populations to ensure inclusion in outreach and immunization services, when feasible. We compare two methods of accomplishing a remote inventory of building locations to assess their accuracy and similarity to currently employed microplan line-lists in the study area.

Methods

The outputs of a crowd-sourced digitization effort, or mapathon, were compared to those of a machine-learning algorithm for digitization, referred to as automatic feature extraction (AFE).

The accuracy assessments were conducted: 1) an agreement analysis of the two methods assessed the occurrence of matches across the two outputs, where agreements were labeled as “befriended” and disagreements as “lonely”; 2) true and false positive percentages of each method were calculated in comparison to satellite imagery; and 3) counts of features generated from both the mapathon and AFE were statistically compared to the number of features listed in the microplan line-list for the study area.

Results

The mapathon and AFE outputs detected 92,713 and 53,150 features, respectively. A higher proportion (30%) of AFE features were befriended compared with befriended mapathon points (28%). The AFE had a higher true positive rate (90.5%) of identifying structures than the mapathon (84.5%). The difference in the average number of features identified per area between the microplan and mapathon points was larger (t = 3.56) than the microplan and AFE (t = -2.09) (alpha = 0.05).

Conclusions

Our findings indicate AFE outputs had higher agreement (i.e., befriended), slightly higher likelihood of correctly identifying a structure, and were more similar to the local microplan line-lists than the mapathon outputs. These findings suggest AFE may be more accurate for identifying structures in high-resolution satellite imagery than mapathons. However, they both had their advantages and the ideal method would utilize both methods in tandem.

Background

Of the 20 million children across the world with incomplete or no essential immunization (EI) for vaccine-preventable diseases, nearly half live in countries with conflicts and population displacement (e.g., Afghanistan, Central African Republic, Iraq, Mali, Nigeria, Pakistan, and Somalia) (1). Conflicts and regional instabilities generally lead to poor vaccination coverage and interrupted vaccine schedules (2) due to disruption of health systems and impeded access to care resulting in vaccine delivery inequities. Currently, the barriers to vaccine preventable disease control are increasingly less about pathogen biology and more about the identification of sub-populations missed by the Expanded Programme on Immunization and therefore left without equitable access to interventions like essential immunization and supplementary vaccination campaigns (3, 4). Immunization programs miss or underserve hard-to-reach sub-populations for various reasons including geographic inaccessibility, irregular population migration due to regional instabilities, and nomadic lifestyles. Public health must harness innovative and effective technologies to improve the remote identification of vulnerable, hard-to-reach sub-populations in order to conduct outreach when feasible and to deliver services to those populations during periods of accessibility.

Understanding the geographic distribution of target populations for health interventions is a critical component of microplanning — a population-based, detailed planning method aimed at delivering health-care interventions like childhood essential immunizations by addressing the implementation demands of a specific setting (5). Microplans critically inform decisions regarding appropriate delivery strategies (i.e., fixed-post, outreach, or mobile) and logistics needed to reach children targeted for the intervention (i.e., target populations) (6). Despite the utility of current microplans, arguments have been made for updated methods of microplanning that leverage Geographic Information Systems (GIS) and satellite imagery to generate high quality and up-to-date maps of target population distributions and maps of built features such as residential structures and settlements (7, 8). In their Reach Every District (RED) strategy for essential immunization, the World Health Organization (WHO) and the United Nations Children's Fund (UNICEF) recognized the need for these updated methods and outlined new GIS-enhanced microplanning tactics for improved location surveillance of some populations

In some situations, GIS-based microplanning incurs higher costs than traditional, non-GIS based microplanning; however, this does not necessarily imply cost ineffectiveness. A recent analysis cost-effectiveness conducted in two Nigerian states determined that increased cost for GIS-based microplanning was mostly due to purchasing additional vaccines for populations previously uncounted and unreached by traditional microplanning methods (7). Not only does GIS-based microplanning save resources when executed appropriately, it also protects the lives of field workers in settings where conflict could compromise their security by reducing the need for deployment to high-risk areas (6). When in-person access is safe and feasible, having field workers physically present in the region of interest allows for ground-truthing which is needed to validate maps generated remotely (i.e., generated using imagery and without physical access to the area of interest). Supplementing microplanning methods with the integration of GIS technologies could further support other public health interventions, such as spraying

insecticides for mosquito abatement and malaria prevention (9, 10) and the provision of maternal and child health care services (7).

To support the integration of GIS technology in public health planning, researchers take advantage of high- or very high-resolution (VHR) satellite imagery generated by satellites like GeoEye, QuickBird, RapidEye, and WorldView. Sub-meter resolution imagery from these satellites allows analysts to digitize features such as buildings, rooftops, roads, nomadic camps, and informal settlements. The size of a population can even be modeled from these footprints.

Large-scale feature digitization (e.g., digitization of individual structures across multiple districts or provinces) from imagery without automation methods is very time-consuming for a small group of analysts, especially when the features of interest are sparse in the imagery. Consequently, a method of participatory data acquisition has gained popularity – the “mapathon” – which is a time-limited, crowd-sourced effort by a group of trained participants with or without formal geospatial analysis backgrounds. Participants, used in this paper to describe both the group of contributors and validators, generate spatial data of features like residential structures or informal settlements within a specific area of interest by using GIS platforms, such as OpenStreetMap and ArcGIS Online. Generally there is no financial incentive for contributions made during a mapathon (11) and anyone with a computer and internet connection can contribute. Consequently, humanitarian efforts frequently rely on mapathons to identify mobile populations and undetected settlements (11, 12). Similarly, data generated from mapathons are useful for detecting and enumerating populations missed during immunization campaigns; thereby, optimizing immunization campaign microplans. Mapathons also provide data that are used to map health facility catchment areas when merged with other key information (12).

An alternative method to using mapathons is automated feature extraction (AFE), a type of model-based feature generation, which can be semi- (i.e., some human support) or fully automated (i.e., no human support). After an initial time investment to manually develop training data using selected examples of features of interest (e.g., man-made structures) and examples of features not of interest (e.g., large boulders), AFE does not require time- consuming and labor-intensive steps such as identifying structures and placing points or drawing polygons manually on a computer. AFE relies on computer algorithms and models to learn patterns, edges, and shapes of features (e.g., rooftops or settlement footprints) to digitize and categorize. Machine learning algorithms are designed to enhance performance by effectively teaching the computer how to extract the desired spatial data from imagery with both precision and accuracy. AFE has been leveraged for a myriad of purposes, such as mapping agricultural land use (13–16) and water boundaries (17, 18), estimating human and livestock populations (19, 20), road feature extraction (21, 22), building feature extraction (23–29), and to support disaster relief efforts (30, 31).

Like mapathons, AFE relies on high-resolution imagery for optimal performance, but image collection parameters can be refined to account for cloud cover, thick vegetation, and low spectral resolution. Additionally, using a time-series of images can improve the accuracy of feature detection by minimizing false-positives (14, 18) and are especially helpful when analyzing pre- and post-disaster impacts to roads

(30) and facilities (31). AFE could be particularly useful for essential immunization efforts because it generates spatial data from imagery rapidly and has the potential to be more accurate than mapathons.

There is currently no information on how results from participatory mapping compare to the results from AFE; if researchers determine AFE to be as accurate and precise as mapathons but faster at generating spatial data, increasing its use could save valuable resources and time for public health programs without compromising quality. Additionally, as geospatial professionals gain a deeper understanding of the strengths of each method, future projects can more optimally combine the two to complement and enhance their end-products.

Disparities in equitable access to health services will decrease when additional sub-populations are identified in microplans and serviced by EI campaigns and other public health interventions. Here, we seek to explore and compare the accuracy of two methods of feature generation – mapathons and AFE – to provide evidence for the suitability of each method in identifying hard-to-reach populations vulnerable to vaccine-preventable diseases in inaccessible areas and whether the two methods can work in a complementary or synergistic way.

Methods

Both feature generation events (i.e., mapathon and AFE) studied here used the same satellite imagery. The study area comprises two districts in the Central Asia region of the world that were inaccessible to EI at the time of the study. To protect the security of populations living in our study region, the specific geographic areas will not be disclosed. The varied terrain of the urban and rural study areas included rocky and forested mountainous regions, low plateau areas with desert terrain, and some fertile plains used for farmland. The climate in the study area is arid to semiarid with low rainfall in most areas of the region. Permanent and temporary housing structures were visible in the satellite imagery used for the study and included small mud free-standing structures, larger mud-brick and stone compounds surrounded by walls, modern free-standing structures and housing complexes in urban areas, small cliff dwellings, and temporary tents or yurts.

Mapathon

We conducted the mapathon for this project under the guidance of the WHO and the Geospatial Research, Analysis, and Services Program (GRASP) at the Centers for Disease Control and Prevention (CDC). The mapathon coordinators created a dedicated ArcGIS Online (ESRI, Redlands, CA) hub page for enrolling and training both novice and experienced mappers. This mapathon resource repository included registration information, tutorials on digitizing and application use, a real-time monitoring dashboard to analyze participant progress links to communication channels, and sections on frequently asked questions. Mapathon participants from the CDC and WHO were recruited via emails and hardcopy informational posters. Participants logged into an ArcGIS Online web application with basic data editing

functionality to view current, high-resolution (0.3–0.5 meters) satellite imagery downloaded from DigitalGlobe (DG) with the goal of identifying structures inside the two study districts.

The imagery for both districts covered a total area of 6,146 square kilometers. The entire area of interest was divided into 1km x 1km grid cells for the contributors who then digitized features of interest within one cell at a time. For each cell, contributors placed one spatially linked point on the center of any man-made structure that was larger than $9m^2$ in the image (Fig. 1). If multiple structures existed within a compound (i.e., several structures surrounded by a common wall), the contributors digitized any eligible structure within the compound rather than counting the compound grouping of structures as just one point. Digitized features could be structures used for any purpose. Additionally, contributors were instructed to place points on structures that seemed to be under construction, regardless of shape, while avoiding those that appeared to be destroyed. Structures partially within the grid were treated as within the grid and were digitized. When a contributor marked a cell as complete, all man-made structures larger than $9m^2$ and visible in the imagery should have been digitized as point feature class data. GIS experts served as validators within the mapathon coordination team, using a separate ArcGIS Online web application to validate any cells marked as complete by the contributors. Validators ensured that all features of interest in the underlying satellite image were correctly digitized by contributors and made edits as needed before finalizing each cell. Because the mapathon contributors and validators had little-to-no knowledge about the setting, they did not make any classifications regarding the current use of the buildings they digitized.

Automated Feature Extraction

The alternative method of acquiring spatial data for this project leveraged semi-automated feature extraction using the results from machine-learning deployments on millions of structures across various developing countries. These data supported Ecopia Tech's (Toronto, ON, Canada) proprietary machine-learning models in generating building footprints for structures of interest detected in the imagery.

To support the models in extracting building footprints, relevant imagery was broken into a grid of 265 x 256-pixel chips. Within each chip, a classifier ran through every pixel and assigned each one a probability score for containing a feature of interest, using a variety of textural feature data from neighboring pixels in its calculations. The classifier algorithm used is a proprietary algorithm developed by Ecopia which measures the shearing of pixels along with color gradients to determine the likelihood that a structure falls within a pixel. Shearing in straight and/or circular lines can be indicative of man-made materials. If sheering, texture and contrast scores exceeded Ecopia's internal threshold of 1 then the pixels were classified as likely containing or being a part of a structure. The classifier algorithm then digitized each structure's boundary, using the confidence scores previously generated for each pixel. Any chips that did not contain structures were removed from the algorithm's output. A team of former geospatial professionals and remote sensing enthusiasts who are expert annotators then reviewed the resulting data sets, manually corrected any errors, and provided any necessary updates to the classifier algorithm. Using a "CrowdRank" algorithm we were able to classify users who perform better when compared against other users completing the same task. Users who regularly fall below a pre-defined benchmark are

removed from the project in an iterative fashion to promote the highest accuracy possible. Informed by the updated and improved data, the classifier algorithm then iteratively reproduced the process to increase overall accuracy. During these iterations, the annotators continued to manually revise any incorrectly generated vector edges and updated the classifier algorithm accordingly. Furthermore, the annotators manually digitized obscured structures to reflect accurate footprints of structures of interest. Prior to this deployment, Ecopia Tech developed an AFE algorithm capable of classifying footprints and partnered with Maxar Technologies (Westminster, CO) to utilize their very high-resolution imagery mosaics and guidance on categorizing the building footprint outputs. To accurately categorize footprints as commercial, compound, or residential, expert imagery analysts from Ecopia manually identified examples of each from the imagery to use in training data for the machine-learning model.

Guided by discussions with local consultants, Ecopia defined a compound as typically including several structures along with a yard surrounded by a wall.
Non-walled, free-standing structures were then categorized as either commercial or residential depending on other contextual factors, such as the proximity and presence of latrines, farmlands, vehicles, and other indicators of human activity (Fig. 2). The AFE algorithm excluded round structures and structures smaller than $9m^2$. The outputs of the model were polygons drawn around the perimeter of all free-standing structures larger than $9m^2$ (whether categorized as commercial or residential) and all residential compounds (Fig. 1). Unlike the mapathon, compounds, regardless of the number of structures contained inside, were treated as one polygon feature.

An estimate of the population inhabiting the structures captured by each method was calculated employing the assumption that a compound includes three households, where each household has an average of 2 adults and 5 children. Therefore, it was estimated that each compound housed an average of 21 individuals.

Accuracy Assessment

The study areas were selected for two reasons: their geographic heterogeneity despite the low spectral diversity (e.g., deserts, arid mountainous, alluvial plains, etc.) and the inaccessibility of local ground-truth data due to continuous insurgencies.

We employed the following accuracy assessment techniques to determine how well each feature generation method – mapathon or AFE – captured the actual location of features of interest.

Assessment 1

We conducted an agreement analysis of the two feature classes to assess matches across the two outputs. To ensure a uniform comparison across both sets of features, we only considered the non-compound AFE features and the mapathon points that were not part of a compound. A simple ‘select by location’ query was used within ArcGIS Pro, whereby both feature types, point and polygon, were analyzed together.

To allow for small shifts in geographic location when comparing mapathon points and AFE polygons, features within 5m of each other were considered a match and labeled as “befriended” (Fig. 3). If a point from the mapathon did not fall within an AFE polygon or have an AFE polygon within 5 meters of it, we labeled that point as “lonely”. Similarly, if a polygon from AFE did not have a corresponding mapathon point within 5 meters of the polygon’s edge, we labeled that polygon as “lonely”. Five meter buffers were applied to points and polygons separately, rather than simultaneously, such that the consideration of a potential 5 meter shift in the location of the polygon or the point was analyzed first for one of the feature types and then the other.

Assessment 2

We conducted a subset analysis of the data from both feature generation methods. Two GIS experts independently analyzed the same set of 100 random, lonely points and 100 random, lonely polygons against the same high-resolution imagery. Points and polygons correctly corresponding to a structure in the imagery were classified as true positives (TP), while those remaining were classified as false positives (FP). Finally, we calculated the true positive percentage by averaging the number of TP and FP yielded by the two GIS experts.

Assessment 3

The third accuracy assessment involved statistically comparing the features generated from both the mapathon and AFE to a microplan, considered the gold-standard data, shared by the local-level team from one of the study districts.
The microplan was developed by the local health authorities and is created by listing out the known settlements in the areas targeted for vaccination and estimating the number of households that vaccination teams should expect to find in each settlement. The study district consisted of 44 operational sub-districts called clusters and one vaccination team was assigned to work in each cluster. The microplan included cluster names, number of households per cluster, number of vaccination teams, the population aged 0–59 months (i.e., target age for vaccination), and total population.

To account for differences in feature extraction techniques and parameters, we analyzed residential or compound polygons and mapathon points. Because mapathon points captured structures of any use while the AFE and microplan indicated household structures of residential use only, the count of mapathon points per cluster was recalculated to better approximate the number of households in the cluster. To count only one mapathon point per compound, we first removed mapathon points that fell inside of AFE polygons categorized as compounds from the analysis. We then multiplied the percentage of AFE compounds containing at least one mapathon point (83%) with the number of compounds in each of the 44 clusters and added that number to the mapathon points for each cluster, thus creating a more comparable dataset to the AFE polygons and microplan.

The null hypothesis for this assessment assumed no significant differences between the average number of features per cluster in the microplan in comparison with the average number of features per cluster

obtained from each feature generation method. To test this assumption, we conducted 2-sample t-tests: (1) comparing the average number of points per cluster from the mapathon to the microplan and (2) comparing the average number of polygons per cluster from AFE to the microplan. T-statistics indicated whether there were significant differences ($p < 0.05$) between each feature generation method and the microplan.

Assessment 4: Estimating Population

Population in the study area was estimated by applying the following assumptions to the mapathon and AFE data - A compound consists of 3 households and a household consists of 7 individuals. This assumption was based on advice from in-country colleagues. Therefore, population estimates were calculated based on the number of free-standing (7 individuals) residences and the number of compound residences (21 individuals).

Results

Descriptive Statistics of Mapathon and Validation

The mapathon took place in August 2018 over five days and recruited 107 participants. Seven organizers spent approximately 840 hours, or approximately 120 hours per person, preparing for and conducting the event. The contributors and validators captured a total of 92,713 individual structures across an area of 6,146 km^2 during the mapathon. The number of digitized features differed widely between participants, with a minimum point count of 1 and a maximum of 10,134. Participants spent a total of 98 hours digitizing, averaging 748 features per person.

Of the 92,713 structures digitized during the mapathon, the vast majority (n = 79,640, 85.9%) required no revision by a validator, a sizeable proportion (n = 12,608, 13.6%) were uniquely generated by the validators because contributors missed these structures entirely, and < 0.5% were edited by the validators or contributors themselves.

Descriptive Statistics of Automated Feature Extraction

The AFE process required a total of nine days to complete, costing $25,000. This cost included image mosaic preparation, training data development, model deployment and iterations, and quality checks for a total area of 6,146 km^2. The semi-automated method identified 53,150 individual structures and compounds. The combined use of Maxar satellite imagery processing and Ecopia algorithms enabled the generation of building footprints and consequent categorization of those footprints. Due to the difference in methodologies, it is expected that the AFE method would result in fewer features than the mapathon. The AFE classifier categorized 80.7% (n) of the building footprints as compound structures, 16.4% (n) as commercial structures, and 2.9% (n) as residential structures. The average area of all AFE polygons, representing the footprints of compounds, was 808.2m^2, while the average area for commercial and residential building footprints were 24.4m^2 and 65.3m^2, respectively.

Assessment 1: Comparing Mapathon and AFE

Based on the matches assessed across the two feature generation outputs (Fig. 3), 30% (n/N) of the non-compound AFE-identified structures intersected or were within 5 meters of a mapathon point, while 70% (n/N) were not. Comparatively, 28% (n/N) of the mapathon points that were not part of a compound fell inside of or were within 5m of an AFE polygon, while 72% (n/N) did not. A slightly higher proportion of AFE features were befriended (30%) than the proportion of mapathon points that were befriended (28%). 2% more identified polygons were corroborated by a mapathon point than identified mapathon point were corroborated by a polygon.

Assessment 2: Accuracy Assessment

The subset analysis demonstrated that the AFE method, including residential and commercial structures, had a higher true positive percent (90.5%) than the mapathon (84.5%) in identifying structures (see Appendix I).

Assessment 3: Comparing Average Number of Features per Cluster against the Microplan

When compared against the 25,141 total features included in the microplan, the mapathon identified an additional 20,804 features. The difference between the microplan and the AFE results was smaller (8,142). Figure 4 shows the variation of all features, resulting from the three different techniques in one district of the study area.

The average number of features per cluster in the microplan was 571.39 and the average number of mapathon points per cluster was statistically significantly higher (mean = 1044.20, $t = 3.56$, $p < 0.001$) as was the average number of AFE polygons per cluster (mean = 756.43, $t = 2.09$, $p = 0.04$). Both comparisons indicate that the microplans were missing structures in the clusters reviewed or that both the methods overestimated the number of structures in the microplan clusters. The p-value was significant (alpha = 0.05) for both t-tests, providing sufficient support to reject the null hypothesis, which assumed no significant difference between the average number of features obtained through both extraction methods and the microplan (Table 1).

Assessment 4: Estimating Population

The population in the study area estimated from mapathon results was 648,991 and the population based on AFE results was 911,302.

Table 1
Two-Sample t-Test results comparing features per cluster

t-Test: Two-Sample Assuming Unequal Variances					
	Mapathon Points	Microplan		AFE Features	Microplan
Mean	1044.20	571.39	Mean	756.43	571.39
Variance	717621.79	60285.87	Variance	284076.30	60285.87
Observations	44	44	Observations	44	44
Hypothesized Mean Difference	0		Hypothesized Mean Difference	0	
df	50		df	60	
t Stat	3.56		t Stat	2.09	
P(T < = t) two-tail	0.00083		P(T < = t) two-tail	0.04071	
t Critical two-tail	2.01		t Critical two-tail	2.00	

Table 2
Comparison of feature generation techniques

Mapathon	Indicator	Automated Feature Extraction
A. Generic Indicators		
60 days	**Time**	9 days
Cost of organizational ArcGIS Online licenses (creator license = $1000/yr, Editor license = $200/yr). Labor cost for coordinators (based on salary of coordinators). Participants were unpaid volunteers.	**Cost* to conduct project**	$25,000 for 6,146 km2
Specialized application development expertise required of coordinators	**Skill Level**	Specialized machine-learning expertise required
Smaller geographic regions	**Area (best suited for)**	Larger geographic areas
B. Performance Indicators		
92,713	**Number of Structures Identified**	53,150
+ 20,804	**Difference in structures identified compared to microplan**	+ 8,142
28%	**Non-Compound Match Rate (%Befriended)**	30%
30,904	**Number of Compounds Identified**	43,395
84.50%	**True Positive Percent**	90.50%
648,991	**Estimated Population****	911,302
*Cost does not take cost of imagery into account, as imagery did not have a stand-alone procurement fee for the specific event studied here		
**Population was estimated by applying the following assumptions to the mapathon and AFE data: A compound consists of 3 households and a household consists of 7 individuals.		

Discussion

The results obtained from both feature generation methods were compared to estimates from the current field-level source, a microplan line-list. Even though the study compared differing methodologies for feature generation, measures to ensure a uniform comparison were taken into consideration, including the exclusion of commercial structures and recalculation of mapathon points to better approximate the number of households in the cluster. Results of the t-tests indicated statistically significant differences for each technique in comparison with the microplan, with the total number of features per cluster larger than the microplan and in the case of the mapathon, almost twice the average. The AFE was found to be similar to the microplan when looking at absolute feature counts. While the accuracy of the field-level microplan itself it unknown, it is the best comparison dataset the authors had to compare mapathon and AFE results to. AFE outputs had a higher true positive percent (90.5%) than the mapathon (84.5%), meaning the AFE was slightly better at correctly identifying a structure in the satellite imagery as a structure. The two techniques could be optimized to more accurately detect structures, as both were subject to false positives and an unknown number of false negatives. Large boulders and trees were accidentally digitized manually as structures in the mapathon which could be avoided by using various indices that enhance spectral diversity and by conducting more training with the participants. Likewise, numerous structures amidst cliffs and hilly terrain were not captured by the AFE technique.

Population estimates in inaccessible regions are often difficult to ascertain due to dynamic population changes and the enumeration process being labor-intensive (12). This has important consequences for planning immunization campaigns and estimating vaccination coverage for EI. Acquiring precise population estimates translates into improved vaccine delivery programs once areas become accessible and more accurate evaluations on the coverage of the campaign (32). This analysis was able to produce rough population estimates for the study area based on each feature generation method, based on structure to population ratio assumptions supplied by country level partners. Our overarching purpose in comparing both feature generation methods was to determine which method more accurately identified structures in high-resolution satellite imagery and how the two methods might best complement one another. The most accurate population estimates are a result of optimum accuracy in structure identification.

As populations and population movements continue to fluctuate across large geographic areas, the availability of up-to-date information on the distribution of human settlements constantly changes (33). These are circumstances in which AFE that can be rerun and retrained quickly could make valuable contributions to data availability compared to mapathons that require time-consuming manual inspections for updates. This AFE process required two days for mosaic preparation, and seven days for training data curation, model deployment iterations, and quality checks. Algorithms like the ones used here are useful for expediting work while maintaining or enhancing quality; however, these algorithms are costly. The cost of the pilot AFE project was $25,000 across 6,146 km^2 of the study area and required highly specialized technical expertise (Table 2). As this technology becomes more commonly used and explored, it is expected that the cost could decrease over time, making it more accessible. In contrast,

participatory and collaborative mapping like mapathons require an extensive amount of manpower and time and are therefore most valuable when timeliness is less of a priority, the geographic scale of the study is limited, and current, high-resolution imagery is available (Table 2).

The use of mapathons for public health interventions has increased meaningfully in recent years (34). Mapathons have the ability to promote effective community engagement, creating a sustainable mechanism of generating geographic data that can be used by local immunizers during campaigns, ensuring the inclusion of all settlements. This collaborative style of mapping can recruit a range of expertise and be conducted mostly free of cost; however, the two main methodological challenges are the uncertainty of the quality of data generated by participants and the number of hours it takes to organize and conduct a mapathon.

Manual feature generation and model-driven feature generations are also useful methods to utilize in tandem to exploit the merits of each and develop a product superior to that which would be created by using only one method alone. For example, smaller scale mapathon efforts are an efficient means of training data creation for AFE. Additionally, AFE footprints can be added to an online application to assist mapathon validators in assessing incoming results during a mapathon event. Both AFE footprints and mapathon points, or a combination of the two, can be utilized as inputs for population estimation models whereby structures are a proxy for the population. The researchers suggest that parameters, such as terrain type, delivery deadline, budget, human resources, computing resources, imagery availability, and requirements of the data output should be considered to strike an appropriate balance in using these two methods together. Although assessing whether the mapathons together with AFE provide more accurate results was outside of the scope of this paper, future work might include an analysis of how these two methods could be used in tandem.

Limitations

While the strengths of this study included the use of current, equivalent satellite imagery across the two feature generation methods compared, multiple assessments to understand the ways in which the results of each method were similar and different, and a thorough logistical comparison on how to decide which method to employ (when you must choose only one), the study also had limitations. An important caveat for interpreting our findings is the lack of a true gold standard, in the form of ground-truthed data collection, which made it impossible to calculate the false-negative rate for each method. Furthermore, due to the insecurity of the area, building footprints and points generated through the study could not be validated in the field for potential inaccuracies. Instead, the findings were compared to a microplan line-list developed by the country's local teams and considered closest to ground truth. However, microplans are also limited because they capture known areas of settlements and may not reflect newly established or abandoned settlements. Another inherent challenge in this study is the use of different feature extraction techniques. The mapathon participants were instructed to place one point per unique rooftop included within compounds, while the AFE grouped numerous structures into one feature (as shown in Fig. 1) when the polygon feature represented a compound. This resulted in a considerable

underestimation of individual structures (39,563 fewer structures) with the AFE technique, which can be an issue if using individual structure counts to estimate population

Another important limitation is the lack of equivalent, constant oversight and the introduction of human bias by the mapathon participants in comparison with AFE. While mapathon participants were provided with training resources and had access to constant communication with GIS experts through an online chat application, mapathon coordinators were not able to monitor every point placed by novice contributors. Mapathon validators did confirm the accuracy of each digitized point and made any necessary edits following submission by the contributors, but human error could still be present in this validation process. The AFE method involved in this study also utilized human input as part of the validation process after the classification algorithm was run, but the possibility of human error is lower than that of the mapathon because much less human input was involved.

Conclusions

We presented results comparing two feature extraction methods with the objective of determining how accurately each method identified settlements in hard-to-reach areas for the purpose of improving EI efforts. The findings suggest that the AFE is more robust in detecting structures when compared to the mapathon; however, the results need to be validated in the field when feasible in order to calculate sensitivity and specificity compared to the gold standard of data collected on the ground (i.e., ground truthing). The geographic data obtained from this study will be used to improve existing microplans with the intent of increasing the EI coverage rate in our study area. Future comparison studies must consider a consistent methodological framework across both feature extraction techniques to improve the findings presented in this study. Although both feature generation techniques could be improved further, this study is a step towards strengthening the understanding of potential methods of mapping population distribution in inaccessible areas to support public health interventions.

Appendix

Appendix I:

Assessment 2: Accuracy Assessment Results

GIS Analyst	Mapathon Points		AFE Polygons	
	True Positive	False Positive	True Positive	False Positive
A	91	9	94	9
B	78	22	87	13
Total	169	31	181	22
Average Percent	84.5%	15.5%	90.5%	11%

Abbreviations

AFE
Automatic Feature Extraction
CDC
Centers for Disease Control and Prevention
DG
DigitalGlobe
EI
Essential Immunization
FP
False Positives
GIS
Geographic Information System
GRASP
Geospatial Research, Analysis, and Services Program
RED
Reach Every District
TP
True positives
UNICEF
United Nations Children's Fund
VHR
Very High-Resolution
WHO
World Health Organization

Declarations

Ethics approval

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets generated during and/or analyzed during the current study are not publicly available to protect the security of populations living in our study region but are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' contributions

TP participated in study design and coordination, drafted the manuscript, and contributed to the data collection, analysis, and paper conceptualization. AB contributed to the paper conceptualization. AM participated in coordination, drafted manuscript, and contributed to the data collection, analysis, and paper conceptualization. JE participated in coordination, drafted the manuscript, conducted analyses and contributed to the paper conceptualization. BK conceived and designed study and contributed to data collection and coordination. RP conceived of study, conducted analyses, and participated in study design. AM contributed to the paper conceptualization. SB contributed to the paper conceptualization. MM conceived of study and participated in coordination and study design. NH contributed to the paper conceptualization. All authors reviewed and provided feedback to the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank all the mapathon volunteers from World Health Organization and the Centers for Disease Control and Prevention.

Additional Information

Disclaimers

The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention, the Agency for Toxic Substances and Disease Registry, or other institutions to which the authors belong.

This research was supported in part by an appointment to the Research Participation Program at the Centers for Disease Control and Prevention administered by the Oak Ridge Institute for Science and Education, Training Programs in Epidemiology and Public Health Interventions Network (TEPHINET) and DRT Strategies through an interagency agreement between the U.S. Department of Energy and CDC.

References

1. WHO. National Immunization Coverage Scorecards Estimates for 20182018. Available from: https://www.who.int/docs/default-source/immunization/pertussis/gvap-national-immunization-coverage-scorecards-estimates-2018.pdf?sfvrsn=46a24831_2.
2. Sodha S, Dietz V. Strengthening routine immunization systems to improve global vaccination coverage. British medical bulletin. 2015;113.
3. WHO. Global Vaccine Action Plan 2011-20202013. Available from: https://www.who.int/immunization/global_vaccine_action_plan/GVAP_doc_2011_2020/en/.
4. Polio Endgame GPEI. Strategy 2019–2023: Eradication, integration, certification and containment2019. Available from: http://polioeradication.org/wp-content/uploads/2019/06/english-polio-endgame-strategy.pdf.
5. WHO. Global Polio Eradication Initiative. Best Practices in Microplanning for Polio Eradication2018. Available from: http://polioeradication.org/wp-content/uploads/2018/12/Best-practices-in-mircoplanning-for-polio-eradication.pdf.
6. WHO. Microplanning for immunization service delivery using the Reaching Every District (RED) strategy2009. Available from: https://www.who.int/immunization/sage/9_Final_RED_280909.pdf.
7. Ali D, Levin A, Abdulkarim M, Tijjani U, Ahmed B, Namalam F, et al. A cost-effectiveness analysis of traditional and geographic information system-supported microplanning approaches for routine immunization program management in northern Nigeria. 2020.
8. WHO. Global Routine Immunization Strategies and Practices (GRISP)-a companion document to the Global Vaccine Action Plan (GVAP)2016. Available from: https://apps.who.int/iris/bitstream/handle/10665/204500/9789241510103_eng.pdf.
9. Kamanga A, Renn S, Pollard D, Bridges DJ, Chirwa B, Pinchoff J, et al. Open-source satellite enumeration to map households: planning and targeting indoor residual spraying for malaria. Malaria Journal. 2015;14(1):345.
10. Kelly G, Seng CM, Donald W, Taleo G, Nausien J, Batarii W, et al. A spatial decision support system for guiding focal indoor residual spraying interventions in a malaria elimination zone. 2011.
11. See L, Mooney P, Foody G, Bastin L, Comber A, Estima J, et al. Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. ISPRS International Journal of Geo-Information. 2016;5(5):55.
12. de Albuquerque JY, Pitidis G, Ulbrich V. P. Towards a participatory methodology for community data generation to analyse urban health inequalities: a multi-country case study. 52nd Hawaii International Conference on System Sciences; Hawaii2019.
13. Brown M, McCarty J. Remote Sensing Data and Methods for Identifying Urban and Peri-Urban Smallholder Agriculture in Developing Countries and in the United States. 2017.
14. Debats SR, Luo D, Estes LD, Fuchs TJ, Caylor KK. A generalized computer vision approach to mapping crop fields in heterogeneous agricultural landscapes. Remote Sens Environ. 2016;179:210–21.

15. Ellis P, Griscom B, Walker W, Gonçalves F, Cormier T. Mapping selective logging impacts in Borneo with GPS and airborne lidar. Forest Ecology Management. 2016;365:184–96.
16. North HCP, Belliss D. S.E. Boundary Delineation of Agricultural Fields in Multitemporal Satellite Imagery. IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing. 2019;12(1):237–51.
17. Rishikeshan CR. H An automated mathematical morphology driven algorithm for water body extraction from remotely sensed images. ISPRS journal of photogrammetry remote sensing. 2018;146:11–21.
18. Zimba H, Kawawa B, Chabala A, Phiri W, Selsam P, Meinhardt M, et al. Assessment of trends in inundation extent in the Barotse Floodplain, upper Zambezi River Basin: A remote sensing-based approach. Journal of Hydrology: Regional Studies. 2018;15:149–70.
19. Kellenberger B, Marcos D, Tuia D. Detecting mammals in UAV images: Best practices to address a substantially imbalanced dataset with deep learning. Remote Sens Environ. 2018;216:139–53.
20. Wania A, Kemper T, Tiede D, Zeil P. Mapping recent built-up area changes in the city of Harare with high resolution satellite imagery. Appl Geogr. 2014;46:35–44.
21. Miao Z, Shi W, Gamba P, Li Z. An object-based method for road network extraction in VHR satellite images. IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing. 2015;8(10):4853–62.
22. Nunes DM. Medeiros NdG, Santos AdPd. Semi-automatic road network extraction from digital images using object-based classification and morphological operators. Boletim de Ciências Geodésicas. 2018;24(4):485–502.
23. Arun P, Katiyar S. An intelligent approach towards automatic shape modelling and object extraction from satellite images using cellular automata-based algorithms. GIScience remote sensing. 2013;50(3):337–48.
24. Dumitru CO, Cui S, Schwarz G, Datcu M. Information content of very-high-resolution SAR images: Semantics, geospatial context, and ontologies. IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing. 2014;8(4):1635–50.
25. Hung C-LJ, James LA, Hodgson ME. An automated algorithm for mapping building impervious areas from airborne LiDAR point-cloud data for flood hydrology. GIScience remote sensing. 2018;55(6):793–816.
26. Konstantinidis D, Stathaki T, Argyriou V, Grammalidis N. Building detection using enhanced HOG–LBP features and region refinement processes. IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing. 2016;10(3):888–905.
27. Manno-Kovács A, Ok AO. Building detection from monocular VHR images by integrated urban area knowledge. IEEE Geoscience Remote Sensing Letters. 2015;12(10):2140–4.
28. Sedaghat A, Ebadi H. Distinctive order based self-similarity descriptor for multi-sensor remote sensing image matching. ISPRS Journal of Photogrammetry Remote Sensing. 2015;108:62–71.

29. Yousefı B, Mirhassani SM, AhmadiFard A, Hosseini MM. Hierarchical segmentation of urban satellite imagery. International Journal of Applied Earth Observation Geoinformation. 2014;30:158–66.
30. Coulibaly I, Spiric N, Lepage R, St-Jacques M. Semiautomatic road extraction from VHR images based on multiscale and spectral angle in case of earthquake. IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing. 2017;11(1):238–48.
31. Dubois D, Lepage R. Fast and efficient evaluation of building damage from very high resolution optical satellite images. IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing. 2014;7(10):4167–76.
32. Kamadjeu R. Tracking the polio virus down the Congo River: a case study on the use of Google Earth™ in public health planning and mapping. Int J Health Geogr. 2009;8(1):4.
33. Esch T, Heldens W, Hirner A, Keil M, Marconcini M, Roth A, et al. Breaking new ground in mapping human settlements from space – The Global Urban Footprint. ISPRS Journal of Photogrammetry Remote Sensing. 2017;134:30–42.
34. Coetzeea SM, Solisc M, Rautenbacha P, Greena V, C. Towards understanding the impact of mapathons - reflecting on Youthmappers experiences. The International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences. 2018.
35. Declarations.

Figures

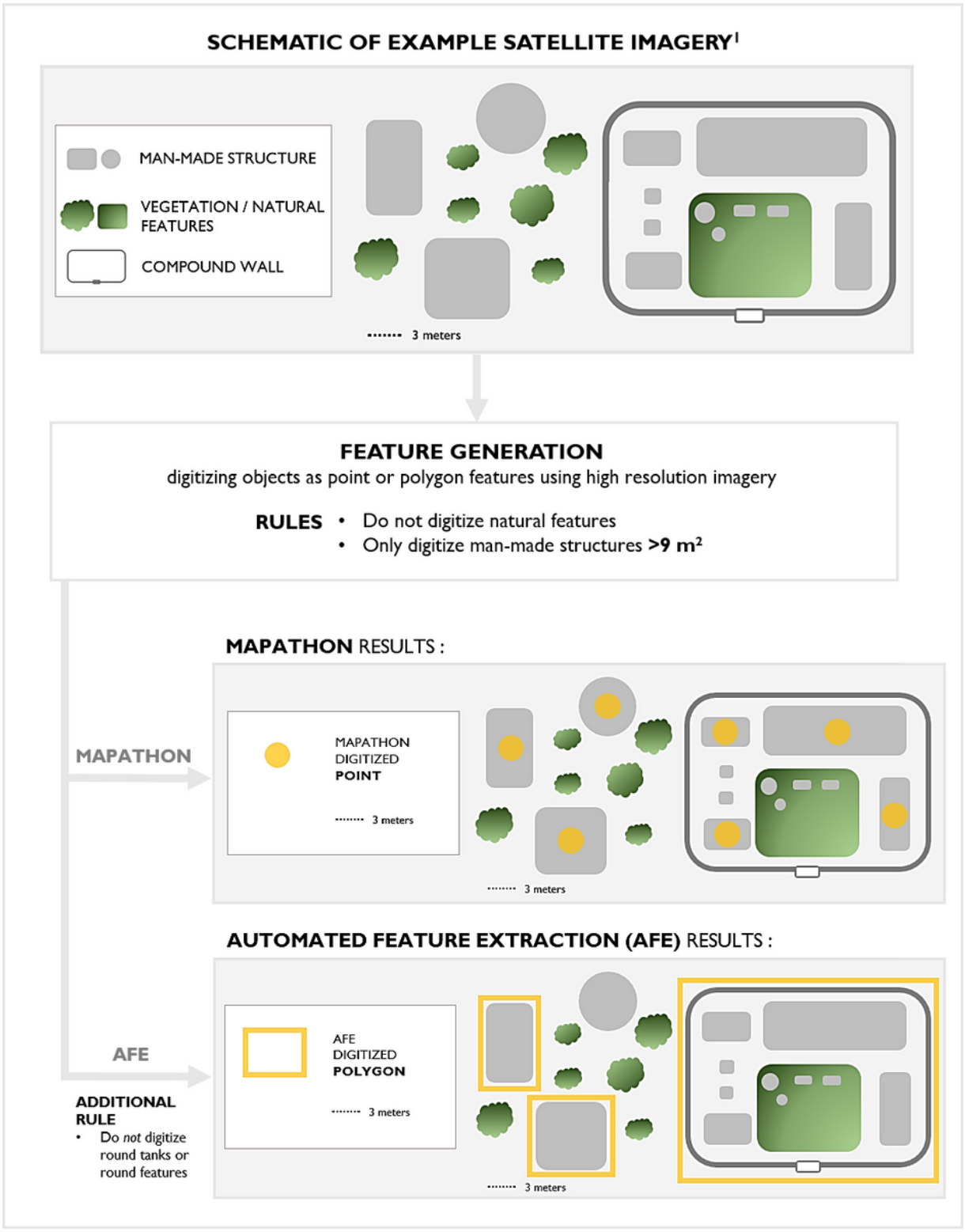

Figure 1

Feature generation using two methods: mapathon (point) and automated feature extraction algorithm (polygon)

Figure 2

Example of free-standing structures in the study area, categorized as commercial (left) and residential (right). (© 2020 Maxar Technologies)

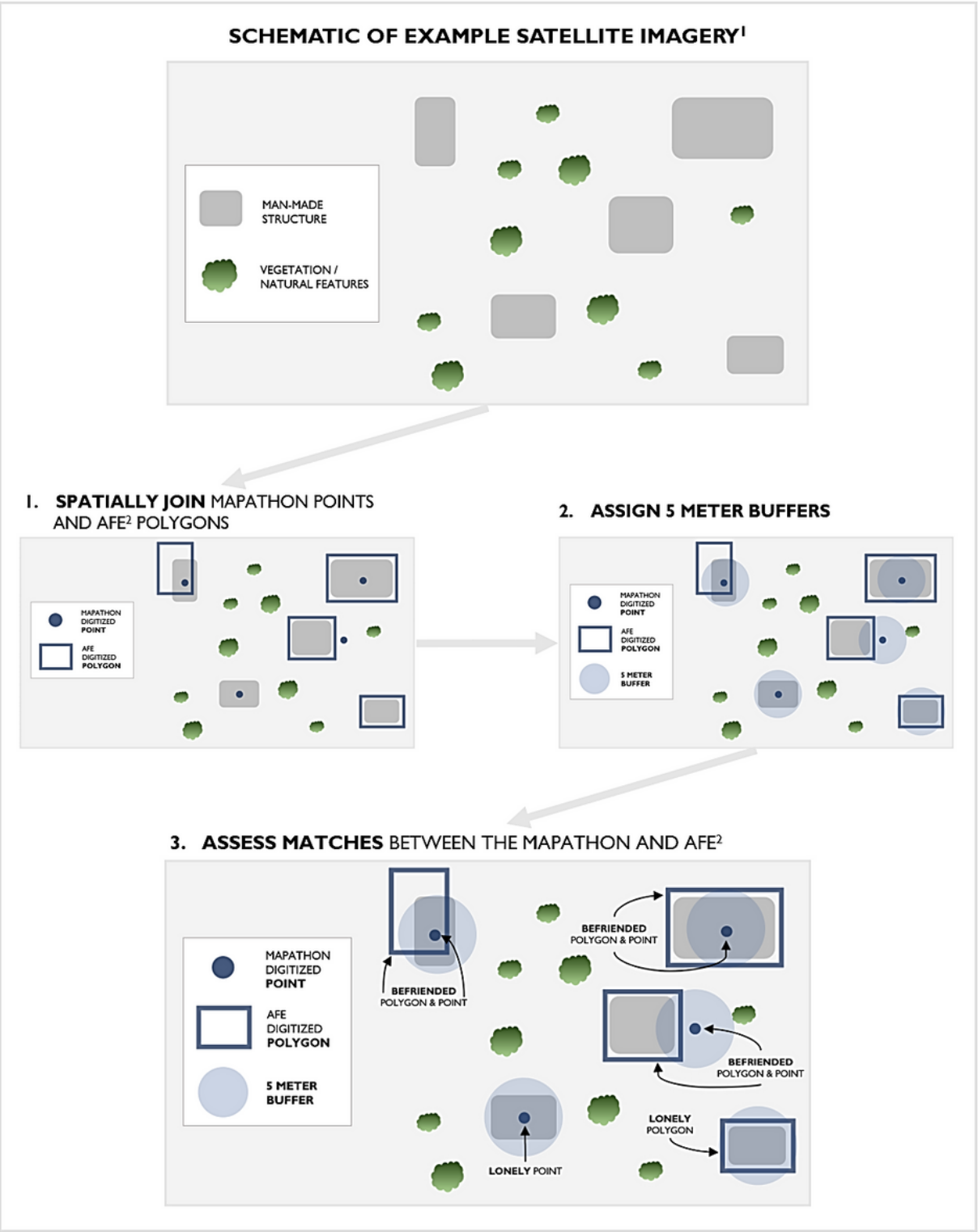

Figure 3

Comparing the results of two feature generation methods: match assessments categorized as “befriended” or “lonely” 1For illustrative purposes only. 2 AFE = automated feature extraction