The Efficacy and Analytical Importance of Manual Feature Extraction Using Lidar Datasets

Seth Quintus, Stephanie S. Day, and Nathan J. Smith

There is little doubt that the acquisition and availability of airborne lidar (Light Detection and Ranging) datasets have revolutionized archaeology (Chase et al. 2012), and the utility of lidar has increased as commercial datasets and derivative products become publically available. For some, lidar-derived products provide another avenue of data visualization that aids in understanding the spatial relationships among archaeological remains (Bennett et al. 2012; Challis et al. 2011; Pingel et al. 2015; Prufer and Thompson 2016; Schindling and Gibbes 2014; Stark et al. 2015; Štular et al. 2012). For others, these datasets have allowed the identification and analysis of archaeological remains at regional scales in heavily

ABSTRACT

The availability of lidar datasets has led to several advances in archaeology, notably in the process of site prospection. Some remote sensing practitioners have aimed to create automated feature extraction (AFE) techniques that increase the efficiency and efficacy of identification and analysis. While these advances have been successful, many archaeological professionals who might have an interest in lidar-derived products do not have the technical experience to modify or create AFE techniques for particular regions or environments. Additionally, some features are not appropriate for AFE. Instead, the most widely used technique is still likely to be visually based manual feature identification. Using authors of different experience levels, we seek to evaluate the use of manual techniques for feature identification and subsequent analysis by implementing a publicly available lidar-derived digital elevation model (DEM). We demonstrate that manual feature extraction (MFE) can be accurate when more than one researcher is involved in a sort of "checks and balances" process. We also show that the use of confidence ratings can be an important part of this process if those ratings have some systematic and clearly defined underpinning. Finally, we argue, using a case study from American Samoa, that manually identified features can be analytically important as part of larger landscape studies.

La disponibilidad de conjuntos de datos lídar ha permitido varios avances en arqueología, notablemente en el proceso de prospección de sitios. Algunos profesionales de teledetección han apuntado a crear técnicas de extracción de características automatizadas (AFE por sus siglas en inglés) que aumentan la eficiencia y eficacia de la identificación y análisis. Aun cuando estos avances han sido exitosos, muchos arqueólogos interesados en el conjunto de datos lídar no tienen la experiencia técnica para modificar o crear técnicas AFE para su uso en regiones o ambientes particulares. Adicionalmente, algunos rasgos podrían no ser apropiados para el uso de AFE. Por lo tanto, es probable que la técnica mayormente usada continúe siendo la identificación manual de características por medio visual. Usando tres autores con diferentes niveles de experiencia, buscamos evaluar el uso de técnicas manuales para la identificación de rasgos y análisis subsecuentes usando un modelo de elevación digital de acceso público derivado de datos lídar. Demostramos que la extracción manual de características (MFE por sus siglas en inglés) puede ser precisa cuando más de un investigador participa en una especie de sistema de controles y balances. Demostramos que el uso de índices de confianza puede ser una parte importante de este proceso si las clasificaciones tienen bases claramente definidas y sistemáticas. Finalmente, usando el estudio de un caso de Samoa Estadounidense, argumentamos que la identificación manual de características puede ser analíticamente importante como parte de estudios de paisaje más amplios.

Advances in Archaeological Practice 5(4), 2017, pp. 351–364

DOI:10.1017/aap.2017.13

vegetated environments that were previously unexplored (Chase et al. 2014; Yaeger et al. 2016). Certainly, these two objectives are not mutually exclusive, and the use of lidar-derived imagery often encapsulates both goals. In fact, the technology has been of such influence that new finds routinely are made at some of the best-known archaeological landscapes in the world (Bewley et al. 2005; Evans and Fletcher 2015).

These global developments in remote sensing have also been realized in the Pacific Islands. In Hawai'i, Ladefoged et al. (2011) and McCoy et al. (2011) used lidar datasets to bridge previous survey data in order to understand regional patterning in archaeological data and document the full extent of prehistoric dryland and wetland agricultural systems. Building on this earlier research, Quintus et al. (2015) applied a semi-automated feature extraction (AFE[1]) technique to the heavily forested environment of Olosega Island, American Samoa, to document and analyze the distribution of residential terraces in a highly topographic environment. More recently, Freeland and colleagues (2016) used two AFE techniques in addition to more limited manual feature extraction (MFE[2]) to map the distribution of mound features in the low-relief but vegetated landscape of Tongatapu Island, Kingdom of Tonga.

The basic utility of lidar imagery for archaeological research has been demonstrated, but questions persist regarding how far these applications can be pushed to examine different feature types and the analytical skill needed to identify archaeological features in different environments. Whereas previous researchers were often responsible for the processing of lidar point clouds (see Devereux et al. 2005)—the raw datasets produced by any lidar investigation—the publication of processed data, often in the form of bare-earth DEMs (Digital Elevation Models; also known as DTMs [Digital Terrain Models]), and open-source data make the information more accessible to the archaeological community. Many practitioners of archaeology do not have the requisite knowledge of computer programming or access to specialized software that more easily enables the processing of raw lidar datasets. Because of this, and even though strides are being made to develop accurate and efficient AFE techniques (e.g., Freeland et al. 2016; Schneider et al. 2015), the more common method to identify features remains visual interpretation using publically available lidar-derived imagery.

MFE has a potentially important role to play for these reasons, especially if the product can be shown to be analytically useful. Although previous research has shown the effectiveness of MFE for the simple identification of a wide range of archaeological features (Chase 2016; Johnson and Ouimet 2014, 2016; Štular et al. 2012), very little research has focused on the analytical potential of these data. Here, we analyze a publically available DEM derived from a lidar dataset of Ta'u Island, American Samoa, to examine the validity and utility of MFE. We compare the visual interpretations of three individuals (the authors) to the results of a field-based mapping program. We seek to address two questions: (1) is MFE an effective means of documenting anthropogenic landscape modifications; and (2) does assigning a confidence rating to manually extracted features provide analytically important information?

STUDY AREA

The island of Ta'u is located in the West Polynesian archipelago of Samoa, which is divided into two modern political units: the Independent State of Samoa in the west and the US territory of American Samoa in the east. American Samoa is itself separated into two groups. The islands of Tutuila and Aunu'u form one group in the western half of the territory, and the islands of Ofu, Olosega, and Ta'u collectively form the Manu'a Group to the east (Figure 1). The uninhabited Rose Atoll and Swains Island are also part of American Samoa.

The islands of the Manu'a Group are small even by Polynesian standards. Ta'u is the largest at 36 km^2, with the smaller islands of Ofu (7.3 km^2) and Olosega (5 km^2) located some 10 km to the northwest. These islands are volcanic in origin and feature high topographic relief. Slopes range from near flat along the coastline to between 5 and 40° within the habitable areas of the interior. Their youthful age, less than 300,000 years (McDougall 2010), has precluded the development of dissected valleys, but some streams have formed, which run following heavy precipitation events. Given the amount of precipitation that the islands receive (more than 3,000 mm annually), vegetation cover is dense, especially in the interiors of these islands, where canopies can grow as high as 25 m (Whistler 1992:14). Much of the vegetation is the result of human manipulation, and forest in the interior portions of all islands in Manu'a includes a high proportion of secondary and economic species (Liu et al. 2011; Whistler 1992).

While archaeological research has been conducted on the coast of Ta'u (e.g., Addison 2008; Cleghorn and Shapiro 2000; Hunt and Kirch 1988), knowledge of the interior uplands is limited (but see Clark 1990; Herdrich et al. 1996). The documentation of the full range and distribution of features in these areas is impeded by several factors. The most important of these is the density of vegetation that limits the size of survey programs, but the lack of easy accessibility to some areas also hampers efforts. Fortunately, most of these archaeological features are landscape modifications visible in imagery derived from a lidar dataset (Quintus et al. 2015).

This article concerns the northeast corner of the interior uplands of Ta'u on the Luatele formation (Figure 2). The area is situated above the modern village of Fitiuta, downslope and seaward of a late Pleistocene volcanic crater (Luatele). Archaeological remains are found in an elevation range between 100 and 375 m with slope values from 5 to 25°. The area includes dense vegetation comprised of secondary (e.g., *Hibiscus tiliaceus*) and economic species (e.g., *Cocos nucifera*) and is bounded on the northwest and southeast by stream banks.

Residential/agricultural terraces and stone/earth walls or embankments are the most common archaeological features in the project area. The walls, embankments, or low linear mounds range in width from approximately 1 to 3 m and in length between approximately 20 and 450 m. In the project area, these

https://doi.org/10.1017/aap.2017.13 Published online by Cambridge University Press

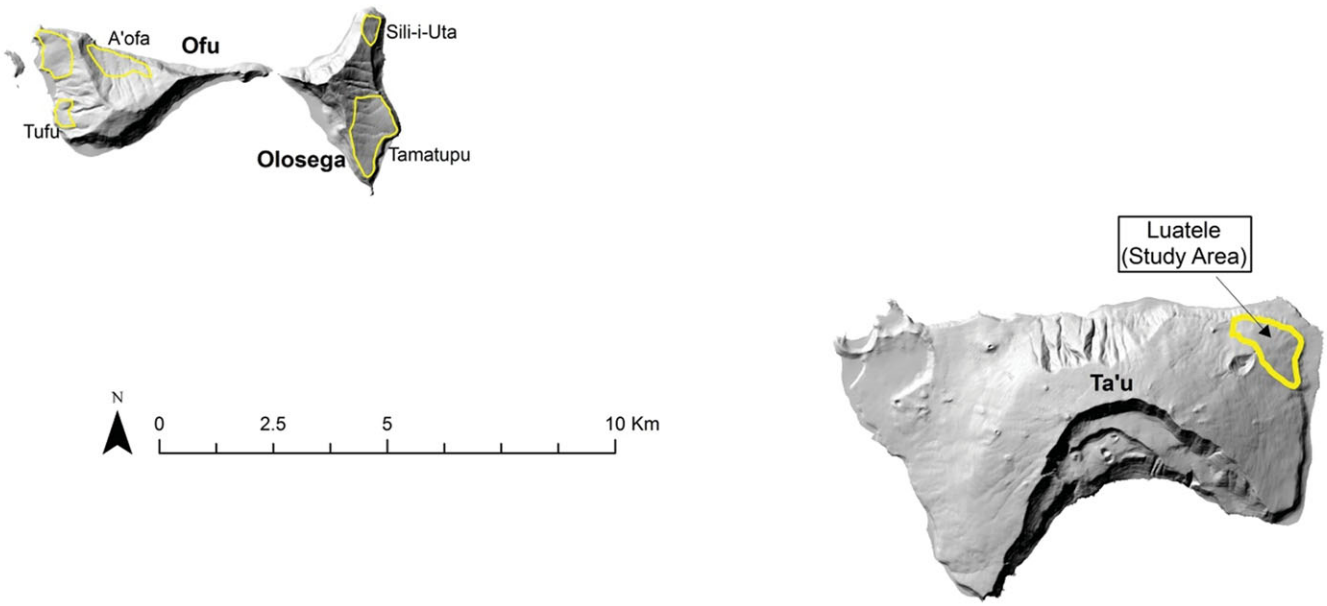

FIGURE 1. Manu'a Islands with the area of interest outlined.

features run both parallel and perpendicular to the slope. Their function is a matter of some speculation but, based on regional analogy (Ladefoged et al. 2003; McCoy and Hartshorn 2007), these might have served to mark boundaries between social units or agricultural fields, reduce erosion, or act as surfaces for cultivation. Terraces in Samoa vary by size and morphology, but, in general, are defined as flat features with 2–3 free-standing sides. While those on Ofu and Olosega are constructed of earth with minimal stone facing, those on Ta'u, at least in the project area, exhibit a relatively higher proportion of stone. These features likely served a variety of functions, ranging from domestic sleeping and eating quarters to foundations for resting and/or working during slope cultivation. The discrimination of function has been difficult, but previous researchers have posited that the presence of secondary features (e.g., coral or waterworn basalt paving, curbing stones) might be used as an indicator of domestic function (Quintus 2015; Quintus and Clark 2012). The identification and analysis of terraces is the subject of this study.

METHODS

Aerial lidar data were collected for American Samoa in June and July of 2012 by the National Oceanic and Atmospheric Administration (NOAA) Coastal Services Center in collaboration with the American Samoa Government Department of Commerce and the US National Park Service (Raber 2012). Data were collected by Photo Science, Inc., using an Optech lidar system at a height of 1,219 m. Line spacing occurred at 395 m with 50% overlap. Average point spacing was 1.43 pts/m^2 with a maximum point spacing of .838 m and a root mean square error of 0.074. Point clouds generated from data collection were processed and classified, also by Photo Science, Inc., in TerraScan and TerraModeler. It should be noted that classification is an imperfect process. Areas of thick vegetation, like those within the project area, can reduce the density of ground returns, and low vegetation might be erroneously classified as bare-earth returns. It is from these bare-earth or ground-classified returns that a DEM (or DTM) was created resulting in a raster with 1 m pixel resolution and better than 15 cm vertical accuracy. This research made use of this open-source DEM (DTM) that is publically available through NOAA with associated metadata.

The three authors have variable experience in the archaeology of the region and the use of GIS for the analysis of spatial datasets. One has extensive experience investigating the interior uplands of the Manu'a Group and has modest experience using lidar datasets for the identification and analysis of archaeological features (R1). Another has extensive experience working with spatial datasets, with a particular expertise in aerial and terrestrial laser scanning systems, and modest knowledge of the archaeology of the interior uplands of Manu'a (R2). The final author (R3) is a student with limited experience in the archaeology of the interior uplands and analysis of lidar datasets, but extensive knowledge of GIS.

Pedestrian survey data used for ground-truthing purposes were collected prior to digital survey during July of 2015 by a crew of four individuals spaced 10 m apart. Terrace features were recorded from three transects that spanned the length of the field and settlement system in the seaward:inland direction (Figure 2). Each feature was outlined as a polygon using iGIS on an Apple iPad tablet with < 5 m accuracy (error range provided by software). Maximum length and width measurements taken in the field with a 30 m tape were compared with these outlines. Photographs were taken of each feature, and other relevant information was recorded (e.g., presence/absence of retaining wall or secondary features). After fieldwork was complete, feature polygons were manually smoothed in ArcGIS to limit the effects of GPS error. Feature area was measured in two ways, by

https://doi.org/10.1017/aap.2017.13 Published online by Cambridge University Press

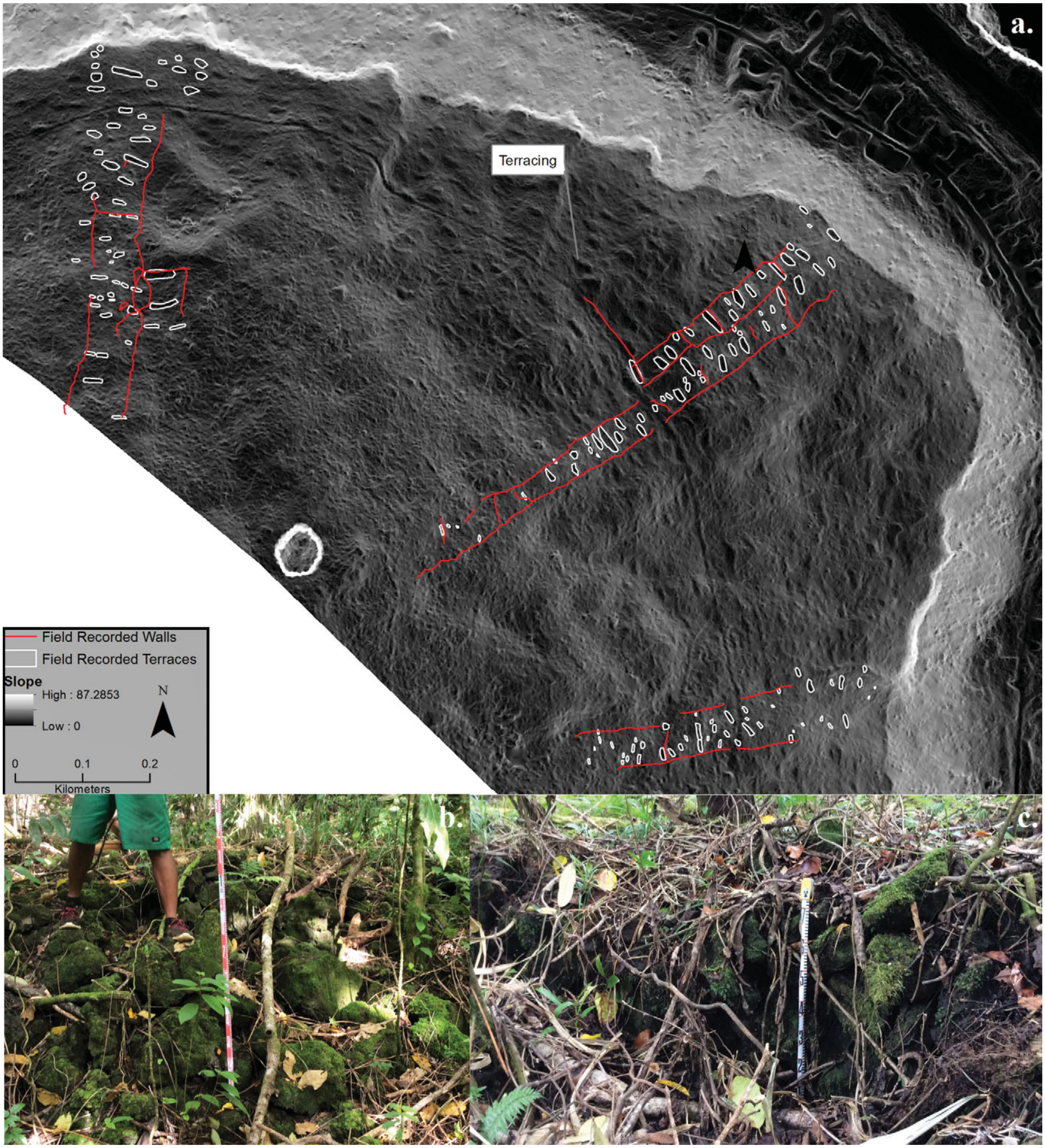

FIGURE 2. (a) results of the initial 2015 pedestrian survey in the Luatele site; (b) stone retaining wall of terrace in foreground, terrace in the background; (c) stone wall running seaward-inland.

multiplying length by width and by using the geometry tool in ArcGIS after the feature polygons had been modified.

Roughly a year after pedestrian survey, each author was provided with the approximate location of the settlement zone and a copy of the lidar-derived DEM. All derivative products (i.e., slope maps, hillshades) were created using the spatial analysis toolkit in ESRI ArcGIS 10.3. The hillshades (grayscale images of surface topography that use the sun's relative position for shading), created by each author, used variable azimuths and altitudes

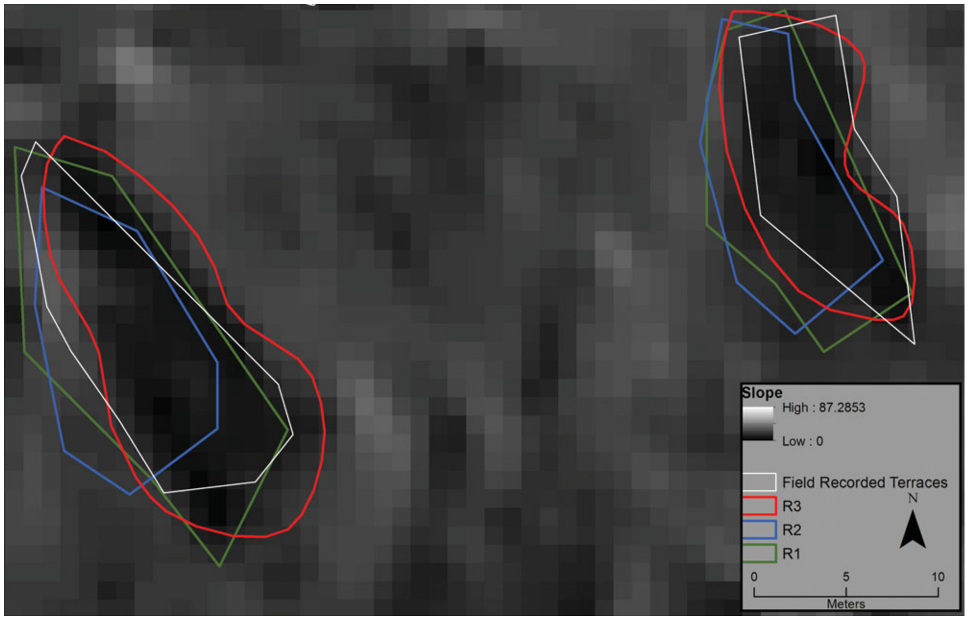

FIGURE 3. An example of positive identification between field and digitally identified features overlying slope map.

to model the sun's relative position. Slope maps with a Z-factor of 1 using stretched (percent clip) display in grayscale and in color were generally most useful. The authors also found it useful to classify the slope map in ArcGIS using a two-class technique (one class below 10–15° slope and the other above this). This method of visualization is similar to that successfully employed previously in Oceania (McCoy et al. 2011; Quintus et al. 2015). R1 and R2 had knowledge of features in the project area, as these authors were part of the crew that mapped the field features used in this research for ground-truthing purposes, but digital feature identification was independent of these previously recorded field data (blind identification). Neither R1 nor R2 knew the exact location or number of terrace features prior to the completion of digital identification, though both did have knowledge of feature morphology.

All digital identifications were completed independently of the other authors within a set time period (approximately two weeks). Each author assigned a three-scale (high [3], medium [2], and low [1]) confidence rating to each digital feature identified. Those rated high were features that the researcher thought had a >75 % chance of being a terrace based on morphology and location; those in the medium category were rated to have a 50–75% chance of being a terrace; and those in the low category were thought to have a < 50% chance of being a terrace but still had some morphological attributes (e.g., contiguous areas of low slope) that suggested they could be anthropogenic features. Finally, a cumulative confidence rating was assigned to each digitally identified feature, adding the confidence ratings of each author together if more than one author identified the feature. This cumulative score ranged from 1 to 9.

The field and digital datasets were integrated into a single GIS and compared. If digitally identified features of two or more authors overlapped, they were noted as a positive identification. Positive identifications were further categorized based on whether that identification was one-to-one (Figure 3) or multiple-to-one (Figure 4) (overlap features; multiple features identified by one author fitting within one feature identified by another). These same principles were applied to the comparison between digitally identified features and those mapped in the field. From this analysis, we focused on and calculated the rate of true positives (TP), false positives (FP), and false negatives (FN) (Table 1). This was done so that the results of identification could be quantitatively assessed using the F_1 measure (Freeland et al. 2016). This is a measure of the accuracy of a binary classification that relies on the calculation of the harmonic mean of sensitivity (R) and the precision (P) (Table 1).

TABLE 1. Definitions of Terms Used in the Text.

Term	Definition
True Positive (TS)	Field-identified terrace that was also identified digitally
False Positive (FS)	Digitally identified terraces that were not documented in the field
False Negative (FN)	Terraces recorded in the field that were not identified digitally
Precision (P)	Portion of positive identifications that are true positives; calculated as TP/TP+FP
Sensitivity (R)	True positive rate; calculated as TP/TP+FN

$$F_1 = 2PR/(P + R)$$

This measure, as opposed to the use of confusion matrices, does not take into account true negative results; these were not

https://doi.org/10.1017/aap.2017.13 Published online by Cambridge University Press

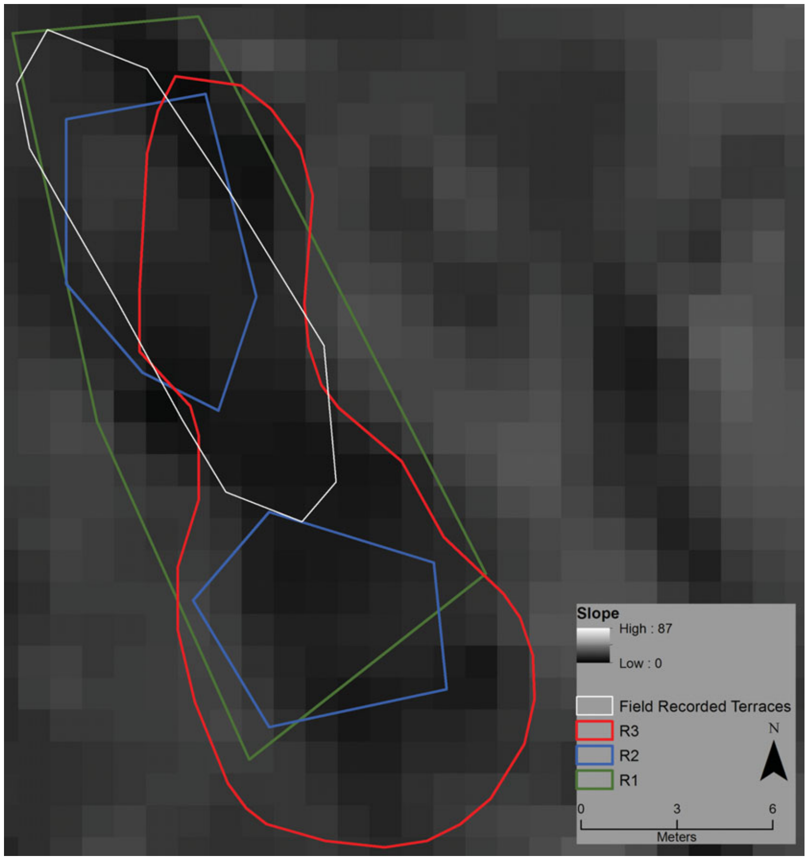

FIGURE 4. An example of a multiple-one (overlap) positive identification overlying slope map. The smaller features were given separate feature numbers and referred to as overlap features.

considered since the project was not able to quantify this in any meaningful way (area of each transect that did not include terracing would not be a realistic measure of true negative). The closer the score is to 1, the better the classification.

RESULTS AND ANALYSIS

Results of Surveyed Tracts and Comparison with Field Data

In total, 161 terraces were identified in the three field transects using pedestrian survey; these have a mean area of 110 m^2. Of those terraces identified, 55 were identified in Transect 1 (T1), 59 in T2, and 47 in T3. Terraces in each transect exhibited slightly different characteristics, notably in terms of size and variance, which is probably the result of functional differences and social processes (e.g., residential status) (Table 2). Even though these terrace samples are significantly different ($F = 13.48$; $p < 0.01$[3]), all three transects include a similar number of terraces that measure less than 200 m^2 (T1 = 46; T2 = 49; T3 = 47). T2 included more terraces measuring over 200 m^2 than any other transect, with no terrace over 200 m^2 identified in T3.

Each author digitally identified a variable number of features within the boundaries of these three transects. In total, R1 identified 145 (average area = 140 m^2) features, R2 identified 103 (average area = 65 m^2), and R3 identified 81 (average area = 144 m^2). In all three cases, TP outnumbered FP by a substantial number (Table 3). However, precision and sensitivity ranged considerably among the authors, and the number of FN recorded by each was over 45.

Results were more successful when author identifications were considered together. Because many of the digital features were recorded by more than one author, the total number of individual terraces identified was 181. Seventy-seven percent of the features identified in the field were identified by at least one of the authors. The vast majority of FP were recorded only by one, instead of more than one, author (70%), while TP were often recorded by two or three authors collectively (Table 4). Of the features identified by all three authors ($n = 53$), 94% were TP. Given the rate of TP within this group, it might be that three other features identified by all three authors are TP but were not recorded during field survey, given low visibility in this densely vegetated environment. Not surprisingly, the cumulative confidence ratings were related to the rate of TP and FP, with TP representing 90% of features with cumulative confidence ratings of 3 or higher (87 of 97) (Table 5). The results of feature identification, however,

TABLE 2. Descriptive Statistics of Features in Each Transect.

Transect	Field Count	Mean Size (m^2)	Max (Min) Size (m^2)	Variance	Lidar Count	Mean Size (m^2)	Max (Min) Size (m^2)	Variance
1	55	122	390 (20)	6599	59	103	373 (17)	**4,957**
2	59	138	428 (24)	9428	69	132	404 (10)	**6,651**
3	47	61	162 (10)	1552	53	71	202 (14)	**2,590**
	161				181			

[a]After combining the datasets of the three authors.

TABLE 3. F_1 Score Assessment of Feature Extraction.

	F_1-Score	FN	FP	TP	Field Identified	Author Total	Precision	Sensitivity
All	0.72	37	60	121	161	181	0.67	0.77
R1	0.71	48	37	108	161	145	0.74	0.69
R2	0.55	88	31	72	161	103	0.70	0.45
R3	0.62	82	9	72	161	81	0.89	0.47
>75 m^2	0.73	8	60	91	94	151	0.60	0.92

TABLE 4. The Differences in True and False Positive Rates When One or Multiple Authors Identify a Feature.

Number Identified	True Positive	False Positive	Total
1	32	42	**74**
2	39	15	**54**
3	50	3	**53**
Total	121	60	181

TABLE 5. The Relationship between Confidence Rating and True and False Positive Rates.

Rating	True Positive	False Positive	Total
9	7	0	**7**
8	14	0	**14**
7	9	0	**9**
6	12	2	**14**
5	12	1	**13**
4	11	2	**13**
3	22	5	**27**
2	16	15	**31**
1	18	35	**53**
Total	121	60	181

were not even across the three transects (Table 6). Instead, the ratio of TP to FP was highest in T2, which also had a low number of FN ($n = 5$).

The average size of terraces that were FN is significantly smaller than those that were positively identified ($U = 989$; $z = 5.24$; $p < 0.01$[4]). In other words, larger terraces were more accurately identified than smaller ones. Not surprisingly, but even more telling, the average size of digital features (average of three authors' feature outlines) is correlated with cumulative confidence rating ($n = 181$; $r_s = 0.47$; $p < 0.01$[5]). This signifies that the authors themselves were more confident in the identification of larger features. The impact of average size is confirmed by the significant increase in the sensitivity of identification when field features that measure less than 75 m^2 are excluded from analysis (Table 3).

The methods used to manually identify features relied heavily on the contrast between the slope of terraces and that of the surrounding area (similar to McCoy et al. 2011; Quintus et al. 2015). Because of this, it was predicted that surrounding slope might have been a limiting factor in identification; terraces that exhibit less contrast with the surrounding slope would be more difficult to identify. This was assessed by calculating the average slope within a 10 m buffer surrounding field-observed terraces, with the overlaps between buffers of different terraces dissolved. The results suggest that there are no differences in the slope of areas surrounding terraces that were positively identified (mean = 15.71) by the authors and those that were not (mean = 15.43; $U = 2280$; $z = -0.05$; $p = 0.96$). Therefore, slope does not appear to have been a factor in feature identification in this study.

For remotely sensed features to be analytically important, they must be reasonable representations of on-the-ground features. A correlation between the field-acquired and remotely sensed data on terrace size discussed here is generally supported by

https://doi.org/10.1017/aap.2017.13 Published online by Cambridge University Press

TABLE 6. The Relationship between Transect Number and the Results of Feature Identification.

Transect	True Positive	False Positive	False Negative
1	39	20	16
2	49	20	5
3	33	20	16
Total	121	60	37

TABLE 7. Attributes of Identified Features Also Identified by Another Author.

ALL	Count	Mean Size (m^2)	Identified[a]	Not Identified	Proportion (%)
R1	701	99	500	201	71.30
R2	752	61	507	245	67.40
R3	484	111	399	85	82.40
High confidence					
R1	337	127	280	57	83.10
R2	89	120	88	1	98.80
R3	76	186	76	0	100
Medium confidence					
R1	223	79	145	78	65.00
R2	188	82	172	16	91.50
R3	207	96	185	22	89.40
Low confidence					
R1	141	62	75	66	53.20
R2	475	42	247	228	52.00
R3	201	97	138	63	68.70

[a]Feature was identified by at least one other researcher.

the similarity of sample means[6] ($n = 103$; Field = 123, Digital = 124; $U = 5035$; $z = -0.63$; $p = 0.53$), but this might mask significant individual feature-level variation. To assess the relationship between field and digital features in our dataset at a finer scale, we compared terrace area measured in the field to that measured digitally using features with a one-to-one positive identification ($n = 103$). Simply, we calculated a proportion statistic (field-measured area/average digitally measured area), subtracted that statistic by 1 to examine distance from a perfect correlation, and then made all numbers positive. The resulting average distance away from a perfect correlation was 0.47. In other words, the field-measured area was roughly 47% larger or smaller than that measured digitally. While this seems large, the mean is affected by a set of outliers with proportions over 1.5, and the median measures 0.28. Moreover, there is a strong negative correlation between confidence rating and variation, with those terraces possessing higher confidence ratings exhibiting less variation from the field measurements ($n = 103$; $r_s = -0.40$; $p < 0.01$). These differences might also reflect the subjective placement of vertices by the authors (see Figures 3 and 4). Still, it appears that there is a general correlation between the field and digital dataset, and that confidence ratings can be used to predict variation between digital and field data.

Results of Non-Surveyed Tracts and Intergroup Comparison

Similar methods of identification were used to digitally survey previously unexplored regions of the Luatele site, with all authors surveying the same area. R1 identified 701 features as terraces, R2 identified 752, and R3 identified 484. On average, these features measured 99 m^2, 61 m^2, and 111 m^2 respectively, and these differences in terrace size among samples were significant ($F = 55.15$; $p < 0.01$). As was the case with feature identification in field-surveyed areas, each author assigned a confidence rating to their identifications. The details of these identifications are provided in Table 7.

The majority of terraces documented by each author were documented by at least one other author, with the rate of repeat identification ranging from 67.4% (R2) to 82.4% (R3). As would be expected, the repeat identification rate correlates with the confidence rating assigned by each author. The rate of repeat identification ranged from 83.1% (R1) to 100% (R3) for high confidence features, from 65% (R1) to 91.5% (R2) for medium confidence features, and from 52% (R2) to 68.7% (R3) for low confidence features (Table 7).

The combination of the three datasets resulted in the identification of 1,111 individual features, 1,009 of which either were identified by a single author or were one-to-one positive identifications identified by multiple authors. One hundred and two were overlap features. While these were not considered cases of one-to-one identifications, they were considered positive repeat identifications in reference to the individual author datasets. A total of 477 features were identified by more than one researcher (mean = 95 m^2), while the remaining 532 were recorded by only one researcher (mean = 53 m^2). Those recorded by one researcher were significantly smaller than those recorded by multiple researchers ($U = 194005$; $z = -14.5$; $p < 0.01$). As was the case in regards to field data, there is a strong positive correlation between confidence rating and average surface area ($n = 1009$; $r_s = 0.63$; $p < 0.01$). Those terraces possessing a combined confidence rating of three or more ($n = 483$; mean = 100) are larger than those with combined confidence ratings of two or less ($n = 526$; mean = 48; $U = 207071.5$; $z = -17.31$; $p < 0.01$).

An analytical dataset was created for further analysis and comparison that included those features that were identified by more than one researcher or were assigned a cumulative confidence rating of three or more. These attributes were chosen based on the ground-truthing results. Namely, a high percentage of features with cumulative confidence ratings of three or higher were true positives (90%), which was also the case for features identified by two or more researchers (83%). Unfortunately, it is likely that this dataset includes a small proportion of FP and excludes a small number of FN. Two separate analytical datasets are considered: one that includes overlap features (cumulative analytical dataset; $n = 637$) and one that does not (analytical dataset; $n = 535$).

The proportion of terraces of different size classes is comparable between the field and analytical datasets (Figure 5), hinting that the digital sample is a valid representation of the terrace

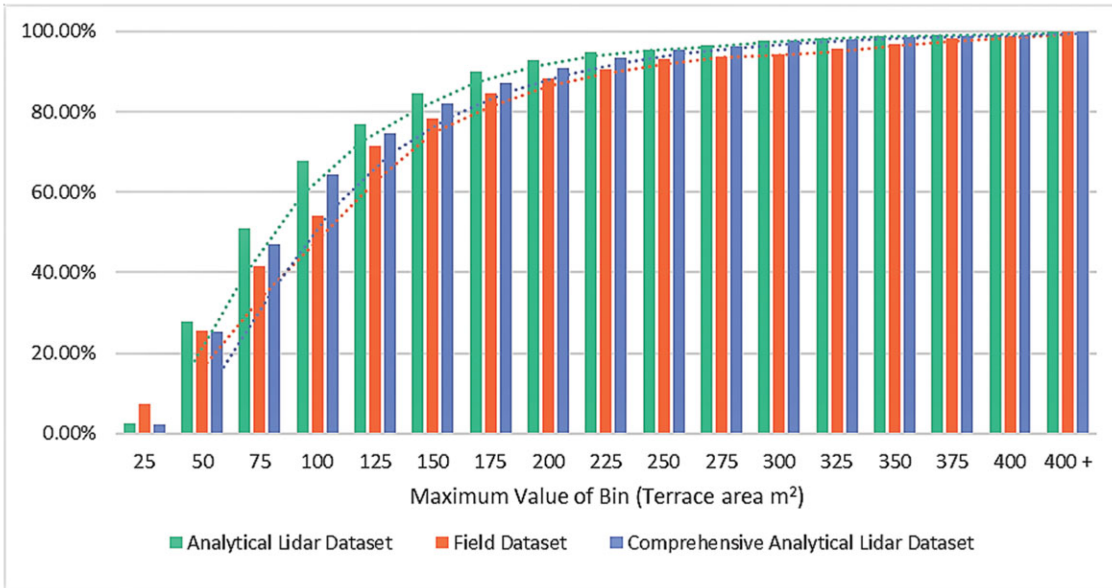

FIGURE 5. Comparison of terrace sizes measured in lidar and the field. Note the discrepancy under 100 m^2.

population. Furthermore, the mean of the analytical sample ($n = 535$; mean = 96) is similar to that of the field sample ($n = 161$; mean = 110), though differences between the samples are significant ($U = 47410$, $z = -1.94$, $p = 0.05$). The differences between the cumulative analytical dataset ($n = 637$; mean = 101 m^2) and field dataset are not significant ($U = 54066$, $z = -1.07$, $p = 0.28$).

Digitally measured features are smaller on average relative to those measured in the field. There could be several reasons for this slight discrepancy, including that the digital datasets include an underestimation of terrace size or that the difference reflects real variation in features found throughout the project area. The latter is more likely, as the digital datasets include more terraces from higher elevations relative to the field dataset. Smaller terrace sizes have been demonstrated to correlate with higher elevations elsewhere in Manu'a (Quintus 2015; Quintus and Clark 2016).

DISCUSSION

Archaeologists are an adventurous lot. We explore remote locations for evidence of past human activity. However adventurous we are, some environments exhibit more difficult conditions of site-prospecting than others, and it is not always possible to achieve 100% coverage, especially across large spatial scales. Tropical rainforests, including those throughout American Samoa, have been especially problematic for traditional field-based prospecting techniques, given the dense vegetation cover that masks even large structures. It is in these environments that the use of airborne lidar has been particularly helpful (Chase et al. 2012; Freeland et al. 2016; Ladefoged et al. 2011; McCoy et al. 2011; Opitz et al. 2015; Quintus et al. 2015). As an attempt to build on this recent research, this article has sought to evaluate the utility of MFE using preprocessed and publically available DEMs. Notably, we have demonstrated the effectiveness of commercial grade airborne lidar datasets for use in identifying and analyzing archaeological landscapes in forested environments (see also Howey et al. 2016). Moreover, we have documented that digital surveys using lidar-derived DEMs, simple MFE techniques, and limited training can provide positive results (F_1 score of 0.72). However, this exercise was not perfect, errors were common, and limitations were demonstrated.

It should also be noted that feature and measurement ambiguity is not unique to digital survey. As all practicing archaeologists know, field survey suffers from ambiguity as well (e.g., measurements, feature interpretations, etc.). For example, both field and digital measurements suffer from surrounding "noise." In the field, this noise is created by the dense tropical vegetation present in the project area. As mentioned above, it is conceivable that we failed to field-record three features later digitally identified as terraces by all three authors as a result of vegetation. Noise in digital settings is created by difficulties processing all vegetation returns out of an image, especially when vegetation is near ground surface, as well as by the error range built into datasets and images used in the analysis (resulting from the creation of the DEM or by rounding surfaces with steep sides). At least some of the FP identified by the authors might have been caused by areas of contiguous, thick, and low-lying vegetation classified as bare earth (see similar situation in Reese-Taylor et al. 2016). The elevation difference between the actual surface and this vegetation might have caused these areas to appear anthropogenic. Because of this, digital and field measurements allow for a check of each other.

Based on ground-truthing results, MFE is more accurate when the identifications of multiple researchers are combined. Our

https://doi.org/10.1017/aap.2017.13 Published online by Cambridge University Press

TABLE 8. Comparison among the Documented Settlement Zones of Manu'a.

	Features Identified	Mean Terrace Area (m^2)	Terrace Size Range (m^2)	Settlement Zone Area (ha)
Luatele (field)	161	110 ± 84	10–418	~115
Luatele (lidar and field)	696	99 ± 78	10–658	~115
Luatele (comprehensive)	798	103 ± 79	10–658	~115
Tufu	58	174 ± 133	18–636	~18
A'ofa	50	194 ± 129	35–650	~49
Sili-i-uta	104	184 ± 103	34–701	~20
Tamatupu	188	289 ± 231	27–2,035	~120

quantitative results confirm that the analysis of a single user includes lower levels of precision and sensitivity, no matter their knowledge of feature morphology. It is essential, based on these results, that researchers applying MFE techniques use and compare the identification interpretations of multiple individuals.

Our results also document some factors that might impact variation in identification. For the terrace feature class on Ta'u, the most pronounced limitation is feature size, with the identification of terraces smaller than 75 m^2 being ambiguous. These small terraces constituted the majority of both FP (77%) and FN (78%) even though they represented only 45% of the collective digital and 42% of the collective field datasets. When terraces of 75 m^2 or less were excluded from comparison, the sensitivity of feature identification improved dramatically, though precision dropped because there was a higher ratio of FP to TP. While it was originally thought to be an important factor, the slope values of the surrounding landscape do not appear to have had an effect on the results of feature identification *in this case*. However, it might be that the contrast between slope and terrace becomes more pronounced with larger features; this lack of *visible* contrast in small features might be a reason why they were difficult to identify. The density of vegetation in some of these areas might also be a key factor in feature identification and the quality of the lidar dataset, though this cannot be assessed here since we used a publically available DEM. These limitations are in part reconciled by the use of confidence ratings. These ratings are often thought of as subjective evaluations, but we contend and have shown that they provide a means to further refine the results of feature identification when they are assigned in a systematic way and have consistent definitions. Researchers must weigh increased sensitivity against decreased precision.

Even with the limitations discussed above, our results indicate that datasets built through MFE can be analytically important. In our datasets, most features with higher confidence ratings were better approximations of the field features than those of lower confidence ratings; those features identified by a single researcher as low confidence were particularly problematic. The blind identification of features by three separate individuals, along with confidence ratings, allowed us to examine the archaeological landscape beyond what has been previously surveyed in the Luatele site. The utility of this is demonstrated by comparison with the broader archaeological landscape of Samoa.

Comparison with Previous Landscape Studies in Manu'a and Regional Implications

The availability of airborne lidar datasets in Oceania has been increasing rapidly. Focusing on Samoa, this ability to survey large tracts of land under tropical vegetation has already proven groundbreaking; airborne lidar is becoming an essential tool for cultural resource management. Several extensive, but previously under-researched, areas in Manu'a have been documented through visual investigation of lidar-derived imagery, including Luatele on Ta'u and Sili-i-uta on Olosega. As modern development continues to intrude on these interior landscapes, airborne lidar will serve as a valuable tool to ensure that cultural resources are identified and sufficiently protected.

With the addition of this analysis, at least partial investigations of interior archaeological landscapes have now been completed on all three islands of the Manu'a group (Quintus 2011, 2012, 2015; Quintus and Clark 2012, 2016; Quintus et al. 2015). These areas, which include the vast majority of habitable land on each island, were virtually unknown just 10 years ago. What has become apparent is the near continuous distribution of archaeological remains across the landscape. Given the size of the islands in Manu'a, the sheer density of terracing speaks to the size of the prehistoric population and the carrying capacity of these environments.

As elsewhere in the world, the examination of variation within and among settlement zones has led to advances in our understanding of a range of social practices in Manu'a (Quintus 2012, 2015; Quintus and Clark 2012, 2016; Quintus et al. 2016). The localized nature of archaeological investigation prior to the availability of lidar in the archipelago largely prohibited this type of comparative analysis. Of particular importance in the analysis of variation, average terrace size differs significantly between documented settlement zones, which are perhaps comparable to the Samoan *nu'u*, or, roughly, village, on the three islands. Whereas there is broad comparability in average terrace size between one settlement zone on Olosega (Sili-i-uta) and the two on Ofu (A'ofa and Tufu), one settlement zone exhibits larger terraces (Tamatupu on Olosega) and one has smaller terraces (Luatele on Ta'u) (Table 8). The range of terrace sizes is consistent between all but Tamatupu (Figure 6), though a higher proportion of small terraces appear to be present in Luatele relative to other settlement zones (Figure 7). On almost all accounts, Tamatupu is an outlier in the Manu'a group.

https://doi.org/10.1017/aap.2017.13 Published online by Cambridge University Press

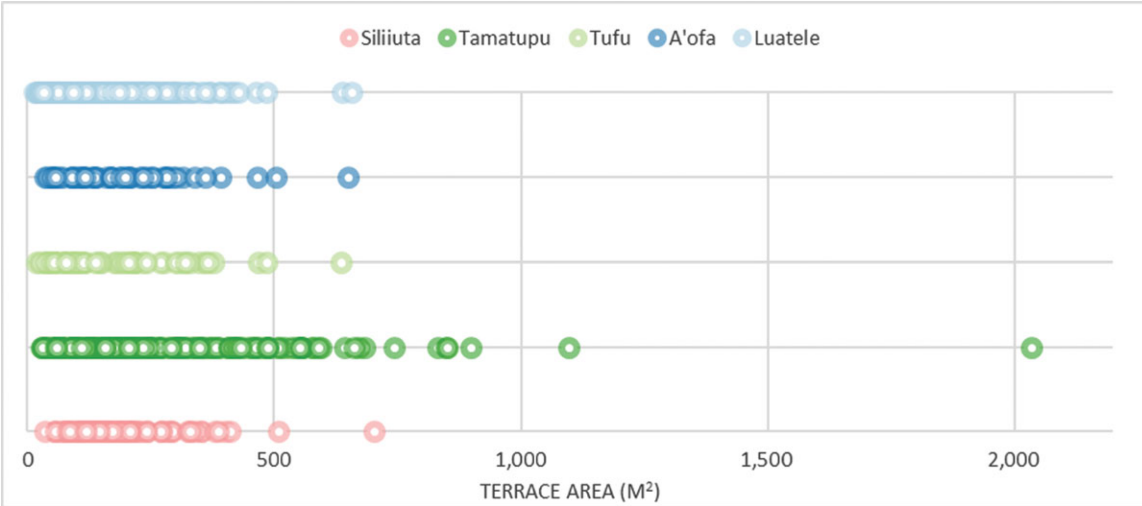

FIGURE 6. Comparison between documented settlement zones in Manu'a. The Luatele sample is the comprehensive dataset (field and all digital features).

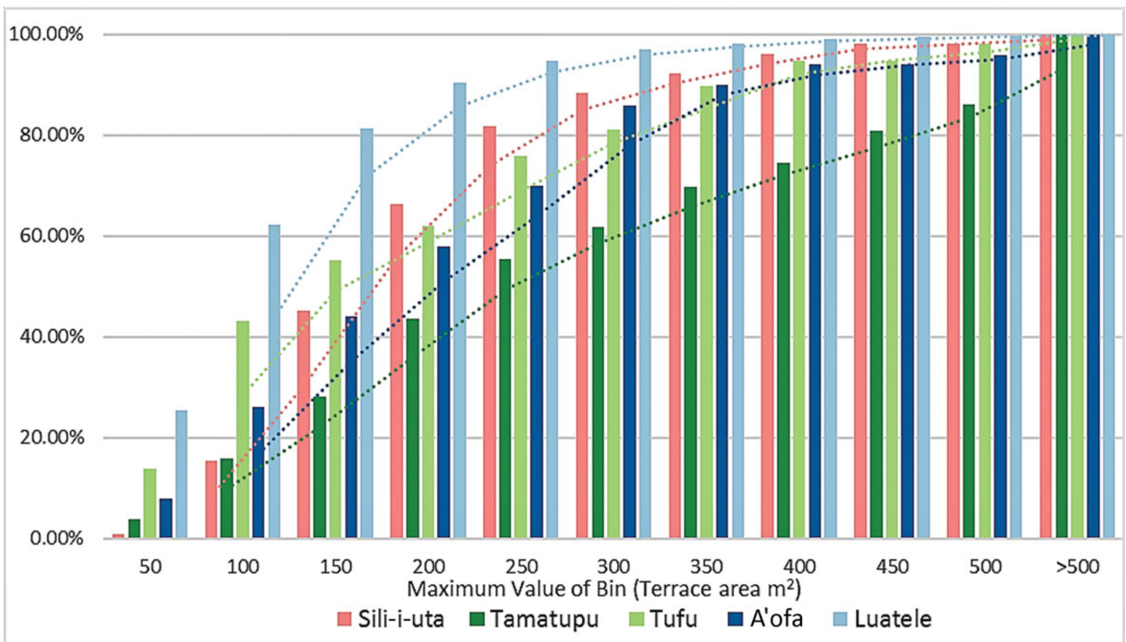

FIGURE 7. Comparative histogram of the five settlement zones documented in Manu'a.

The difference between Tamatupu and the other settlement zones has been hypothesized to relate to political considerations (Quintus et al. 2015). This is based not only on the proportion of large terraces that might reflect status but also on the number of star mounds, local examples of monumental architecture (Herdrich and Clark 1993), associated with the settlement zone. In contrast, the smaller average terrace size in Luatele might relate to differences in activities undertaken in at least portions of the zone. Luatele is unique in Manu'a in that small linear mounds of earth and stone are situated running both perpendicular and parallel to the slope. It might be that the smaller terraces relate to activities associated with these linear mounds, namely food production. Such an interpretation could be assessed with additional pedestrian survey to gather data on feature function. Alternatively, it might be that the smaller average terrace size in Luatele is a product of error associated with digital feature measurement. This is unlikely to explain the entire variation apparent between Luatele and other settlement zones, though, especially given the general comparability between field-based measurements and digital-based measurements discussed above.

The discovery and preliminary documentation of interior upland settlement zones in Manu'a hints of significant unknown archaeological landscapes on the larger islands

https://doi.org/10.1017/aap.2017.13 Published online by Cambridge University Press

of the archipelago. This has already proven true for Tutuila (Cochrane personal communication), but airborne lidar datasets are just becoming available for the Independent State of Samoa. Vast tracts of archaeological features have already been documented in some areas of ʻUpolu and Savaiʻi (Holmer 1980; Wallin and Martinsson-Wallin 2007), but the vast majority of land on each island remains unexplored. If the density of archaeological remains on adjacent Manono is any indication (Sand et al. 2012), we can expect important finds to be made if/when airborne lidar data is analyzed. We are hopeful that continued demonstration of the utility of commercial quality airborne lidar datasets serves as a proof-of-concept to justify the acquisition of comparable datasets by small island nations.

CONCLUSION

With the results of this research in mind, we make the following recommendations about the use of lidar datasets in archaeological research, especially in regards to MFE:

1. Airborne lidar should be paired with pedestrian survey data and the process of feature interpretation should be iterative (following Freeland et al. 2016; Opitz et al. 2015; Quintus et al. 2015; Reese-Taylor et al. 2016).
2. Projects employing MFE should include multiple individuals who undertake feature identification independently of other researchers.
3. Systematic and well-defined confidence ratings should be assigned for all identified features as a quality-control mechanism.
4. The relationship between field and digital data should be assessed quantitatively.
5. The analysis of datasets derived from digital survey that makes use of airborne lidar should provide a set of working hypotheses from which expectations and predictions can be formulated, evaluated, and refined as additional pedestrian survey data are acquired.

We agree with the arguments of previous researchers that the most effective use of lidar datasets is in conjunction with traditional field survey (after Optiz et al. 2015), and that remote sensing does not and should not replace localized pedestrian survey (Freeland et al. 2016). The data collected by different techniques do not necessarily overlap, as they are situated at different analytical scales. Pedestrian survey remains an important tool for documenting the range of sites, both standing and dispersed, within a region. In the case of Samoa, this is important, as secondary features (e.g., foundations, pavings, artifacts) have important implications for feature function analysis, as discussed above. In other regions of the world, lack of pedestrian survey might mask variation in construction materials, artifact densities, and other important data frequently used by archaeologists. However important lidar data might be, it is essential to recognize these limitations. Field and digital surveys are complementary tools, and it is the use of these strategies as complementary tools *by all archaeologists* that presents an opportunity for the addition of lidar to prove revolutionary in archaeology.

Acknowledgments

We thank the people of Taʻu for allowing us to conduct research on their island. Permits were not required for this project. We thank Nolita Motu for her assistance during fieldwork and Jamie Viteri for translating the abstract. We are grateful for the critical comments provided by Donald Schwert, Miram Stark, and four anonymous reviewers on an earlier draft of this manuscript.

Data Availability Statement

All data files will be made publically available online (https://www.ndsu.edu/researchgroups/samoanarchaeology/).

REFERENCES

Addison, David J.
2008 *Report on Archaeological Test Excavations at the New Taʻu Dispensary Site*. Samoa Environmental and Archaeological Consultants. Submitted to the Department of Public Works, American Samoa Government, Pago Pago, American Samoa.

Bennett, Rebecca, Kate Welham, Ross A. Hill, and Andrew Ford
2012 A Comparison of Visualization Techniques for Models Created from Airborne Laser Scanned Data. *Archaeological Prospection* 19: 41–48.

Bewley, R.H., S.P. Crutchley, and C.A. Shell
2005 New Light on an Ancient Landscape: Lidar Survey in the Stonehenge World Heritage Site. *Antiquity* 79:636–647.

Challis, Keith, Paolo Forlin, and Mark Kincey
2011 A Generic Toolkit for the Visualization of Archaeological Features on Airborne LiDAR Elevation Data. *Archaeological Prospection* 18:279–289.

Chase, Adrian S.Z.
2016 Beyond Elite Control: Residential Reservoirs at Caracol, Belize. *WIREs Water* 3(6):885–897.

Chase, Arlen F., Diane Z. Chase, Christopher T. Fisher, Stephen J. Leisz, and John F. Weishampel
2012 Geospatial Revolution and Remote Sensing LiDAR in Mesoamerican Archaeology. *Proceedings of the National Academy of Science* 109:12916–12921.

Chase, Arlen F., Diane Z. Chase, Jaime J. Awe, John F. Weishampel, Gyles Iannone, Holley Moyes, Jason Yaeger, Kathryn Brown, Ramesh L. Shrestha, William E. Carter, and Jaun Fernandez Diaz
2014 Ancient Maya Regional Settlement and Inter-Site Analysis: The 2013 West-Central Belize LiDAR survey. *Remote Sensing* 6:8671–8695.

Clark, Jeffrey T.
1990 *The Taʻu Road Archaeological Project*. North Dakota State University. Submitted to the American Samoa Historic Preservation Office, Pago Pago, American Samoa.

Cleghorn, Paul L., and William A. Shapiro
2000 *Archaeological Data Recovery Report for the Proposed Taʻu Road Reconstruction, at Faga and Fitiuta, Taʻu Island, Manuʻa, American Samoa*. Pacific Legacy. Submitted to the American Samoa Historic Preservation Office, Pago Pago, American Samoa.

Devereux, B.J., G.S. Amable, P. Crow, and A.D. Cliff
2005 The Potential of Airborne Lidar for Detection of Archaeological Features under Woodland Canopies. *Antiquity* 79:648–660.

Evans, Damien, and Roland Fletcher
2015 The Landscape of Angkor Wat Redefined. *Antiquity* 89:1402–1419.

Freeland, Travis, Brandon Heung, David V. Burley, Geoffrey Clark, and Anders Knudby
2016 Automated Feature Extraction for Prospection and Analysis of Monumental Earthworks from Aerial LiDAR in the Kingdom of Tonga. *Journal of Archaeological Science* 69:64–74.

https://doi.org/10.1017/aap.2017.13 Published online by Cambridge University Press

Herdrich, David J., and Jeffrey T. Clark
1993 Samoan Tia 'Ave and Social Structure: Methodological and Theoretical Considerations. In *The Evolution and Organisation of Prehistoric Society in Polynesia*, edited by Michael W. Graves and Roger C. Green, pp. 52–63. New Zealand Archaeological Association Monograph Series, Auckland.

Herdrich, David J., James R. Moore, Nick Kilzner, and Joseph Kennedy
1996 *A Cultural Resource Evaluation (Phase I and II) for a Portion of Road 1b, Phase I of the Ta'u Road Reconstruction Located on Ta'u Island, Manu'a, American Samoa.* Archaeological Consultants of the Pacific. Submitted to the American Samoa Historic Preservation Office, Pago Pago, American Samoa.

Holmer, Richard N.
1980 Mt. Olo Settlement Pattern Interpretation. In *Archaeological Excavations in Western Samoa*, edited by J.D. Jennings and R.N. Holmer, pp. 93–103. Pacific Anthropological Records, No. 32, Honolulu.

Howey, Meghan C., Franklin B. Sullivan, Jason Tallant, Robert Vande Kopple, and Michal W. Palace
2016 Detecting Precontact Microtopographic Features in a Forested Landscape with Lidar: A Case Study from the Upper Great Lakes Region, AD 1000–1600. *PLos One* https://doi.org/10.1371/journal.pone.0162062

Hunt, Terry L., and Patrick V. Kirch
1988 An Archaeological Survey of the Manu'a Islands, American Samoa. *Journal of the Polynesian Society* 97:153–183.

Johnson, Katharine M., and William B. Ouimet
2014 Rediscovering the Lost Archaeological Landscape of Southern New England Using Airborne Light Detection and Ranging (LiDAR). *Journal of Archaeological Science* 43:9–20.
2016 Physical Properties and Spatial Controls of Stone Walls in the Northeastern USA: Implications for Anthropocene Studies of 17th Century to Early 20th Century Agriculture. *Anthropocene* 15:22–36.

Ladefoged, Thegn N., Michael W. Graves, and Mark D. McCoy
2003 Archaeological Evidence for Agricultural Development in Kohala, Island of Hawai'i. *Journal of Archaeological Science* 30:923–940.

Ladefoged, Thegn N., Mark D. McCoy, Gregory P. Asner, Patrick V. Kirch, Cedric O. Puleston, Oliver A. Chadwick, and Peter M. Vitousek
2011 Agricultural Potential and Actualized Development in Hawai'i: An Airborne LIDAR Survey of the Leeward Kohala Field System (Hawai'i Island). *Journal of Archaeological Science* 38:3605–3619.

Liu, Zhanfeng, Neil E. Gurr, Mark A. Schmaedick, W. Arthur Whistler, and Lisa Fischer
2011 *Vegetation Mapping of American Samoa.* General Technical Report (R5-TP-033). Submitted to the U.S. Department of Agriculture, Forest Service, Pacific Southwest Region, Vallejo, California.

McCoy, Mark D., and Anthony S. Hartshorn
2007 Wind Erosion and Intensive Prehistoric Agriculture: A Case Study from the Kalaupapa Field System, Moloka'i Island, Hawai'i. *Geoarchaeology* 22:511–532.

McCoy, Mark D., Gregory P. Asner, and Michael W. Graves
2011 Airborne Lidar Survey of Irrigated Agricultural Landscapes: An Application of the Slope Contrast Method. *Journal of Archaeological Science* 38:2141–2154.

McDougall, Ian
2010 Age of Volcanism and Its Migration in the Samoa Islands. *Geological Magazine* 147:705–717.

Opitz, Rachel S., Krysta Ryzewski, John F. Cherry, and Brenna Moloney
2015 Using Airborne LiDAR Survey to Explore Historic-Era Archaeological Landscapes of Montserrat in the Eastern Caribbean. *Journal of Field Archaeology* 40:523–541.

Pingel, Thomas, J., Keith Clarke, and Anabel Ford
2015 Bonemapping: A Lidar Processing and Visualization Technique in Support of Archaeology under the Canopy. *Cartography and Geographic Information Science* 42:18–26.

Prufer, Keith, and Amy E. Thompson
2016 Lidar-Based Analyses of Anthropogenic Landscape Alterations as a Component of the Built Environment. *Advances in Archaeological Practice* 4:393–409.

Quintus, Seth J.
2011 Land Use and the Human-Environment Interaction on Olosega Island, Manu'a, American Samoa. Master's thesis, Department of Sociology and Anthropology, North Dakota State University, Fargo.
2012 Terrestrial Food Production and Land Use in Prehistoric Samoa: An Example from Olosega Island, Manu'a, American Samoa. *Archaeology in Oceania* 47:133–140.
2015 Dynamics of Agricultural Development in Prehistoric Samoa: The Case of Ofu Island. PhD thesis, Department of Anthropology, University of Auckland, Auckland, New Zealand.

Quintus, Seth J., and Jeffrey T. Clark
2012 Between Chaos and Control: Spatial Perception of Domestic, Political, and Ritual Organization in Prehistoric Samoa. *Journal of the Polynesian Society* 121:275–302.
2016 Space and Structure in Polynesia: Instantiated Spatial Logic in American Sāmoa. *World Archaeology*, DOI: 10.1080/00438243.2016.1195576.

Quintus, Seth, Jeffrey T. Clark, Stephanie S. Day, and Donald P. Schwert
2015 Investigating Regional Patterning in Archaeological Remains by Pairing Extensive Survey with a Lidar dataset: The Case of the Manu'a Group, American Samoa. *Journal of Archaeological Science: Reports* 2:677–687.

Quintus, Seth, Melinda S. Allen, and Thegn N. Ladefoged
2016 In Surplus and in Scarcity: Agricultural Development, Risk Management, and Political Economy, Ofu Island, American Samoa. *American Antiquity* 81:273–293.

Raber, Steven
2012 *Post-flight Aerial Acquisition and Calibration Report.* Contract No. EA133C-11-CQ-0009. Report completed by Photo Science, Inc. Colorado Springs, Colorado. https://coast.noaa.gov/htdata/lidar1_z/geoid12a/data/2490/supplemental/PhotoScience-AmSam_PostFlightAcquisitionReport_FINAL.pdf, accessed April 28, 2017.

Reese-Taylor, Kathryn, Armando Anaya Hernández, F. C. Atasta Flores Esquivel, Kelly Monteleone, Alejandro Uriarte, Christopher Carr, Helga Geovannini Acuña, Juan Carlos Fernandez-Diaz, Meaghan Peuramaki-Brown, and Nicholas Dunning
2016 Boots on the Ground at Yaxnohcah: Ground-Truthing Lidar in a Complex Tropical Landscape. *Advances in Archaeological Practice* 4:314–338.

Sand, Christophe, Jacques Bole, and David Baret
2012 *Archaeology of Manono Island (Samoa): First Field Program Report.* Rapport de mission d' archeology de la Nouvelle-Calédonie et du Pacifique, Nouméa.

Schindling, James, and Cerian Gibbes
2014 LiDAR as a Tool for Archaeological Research: A Case Study. *Archaeological and Anthropological Sciences* 6:411–423.

Schneider, Anna, Melanie Takla, Alexander Nicolay, Alexandra Raab, and Thomas Raab
2015 A Template-Matching Approach Combining Morphometric Variables for Automated Mapping of Charcoal Kiln Sites. *Archaeological Prospection* 22:45–62.

Stark, Miriam T., Damien Evans, Chay Rachna, Heng Pihpal, and Alison Carter
2015 Residential Patterning at Angkor Wat. *Antiquity* 89:1439–1455.

Štular, Benjamin, Žiga Kokalj, Kristof Oštir, and Laure Nuninger
2012 Visualization of Lidar-Derived Relief Models for Detection of Archaeological Features. *Journal of Archaeological Science* 39:3354–3360.

Wallin, Paul, and Helene Martinsson-Wallin
2007 Settlement Patterns-Social and Ritual Space in Prehistoric Samoa. *Archaeology in Oceania* 42 Supplement:83–89.

https://doi.org/10.1017/aap.2017.13 Published online by Cambridge University Press

Whistler, W. Arthur
1992 *Botanical Inventory of the Proposed Ta'u Unit of the National Park of American Samoa*. Technical Report 83. Submitted to the National Park Service, Honolulu.
Yaeger, Jason, M. Kathryn Brown, and Bernadette Cap
2016 Locating and Dating Sites Using Lidar Survey in a Mosaic Landscape in Western Belize. *Advances in Archaeological Practice* 4:339–356.

NOTES

1. Extraction techniques that make use of machine learning algorithms or similar techniques.
2. Extraction or identification of features based on visual interpretation.
3. ANOVA.
4. Mann-Whitney Test.
5. Spearman's rank correlation coefficient.
6. Only true positives with one-to-one identification.

AUTHORS INFORMATION

Seth Quintus ■ Department of Anthropology, University of Hawai'i at Mānoa, Honolulu, HI

Stephanie S. Day ■ Department of Geosciences, North Dakota State University, Fargo, ND

Nathan J. Smith ■ Department of Sociology and Anthropology, North Dakota State University, Fargo, ND

https://doi.org/10.1017/aap.2017.13 Published online by Cambridge University Press