Peer-reviewed version available at *Geosciences* **2017**, *7*, 127; doi:10.3390/geosciences7040127

Article

Semi-automatic detection of indigenous settlement features on Hispaniola through remote sensing data

Till F. Sonnemann[1*], Douglas C. Comer[2], Jesse L. Patsolic[3], William P. Megarry[4], Eduardo Herrera Malatesta[5], Corinne L. Hofman[6]

[1] ERC-Synergy project Nexus1492 (NEXUS1492.eu). Institute of Archaeology, Heritage Sciences and Art History (IADK), Otto-Friedrich-Universität Bamberg, Am Kranen 12, 96047 Bamberg, Germany; till.sonnemann@uni-bamberg.de

[2] Cultural Site Research and Management (CSRM), 2113 St. Paul Street, 21218 Baltimore, MD, USA; dcomer@culturalsite.com

[3] Center for Imaging Science, Johns Hopkins University, Baltimore, MD 21218, USA; jpatso11@jhu.edu.

[4] School of Natural and Built Environment, Queen's University Belfast, Belfast, Northern Ireland, UK. BT7 1NN; W.Megarry@qub.ac.uk

[5] ERC-Synergy project Nexus1492. Faculty of Archaeology, Universiteit Leiden, Einsteinweg, 2, 2333 CC Leiden, The Netherlands; c.l.hofman@arch.leidenuniv.nl

[6] ERC-Synergy project Nexus1492. Faculty of Archaeology, Universiteit Leiden, Einsteinweg, 2, 2333 CC Leiden, The Netherlands; e.n.herrera.malatesta@arch.leidenuniv.nl

* Correspondence: till.sonnemann@uni-bamberg.de, Tel: +49 951 863-3930

Abstract: Satellite imagery has had limited application in the analysis of pre-colonial settlement archaeology in the Caribbean; visible evidence of wooden structures perishes quickly in tropical climates. Only slight topographic modifications remain, typically associated with middens. Nonetheless, surface scatters, as well as the soil characteristics they produce, can serve as quantifiable indicators of an archaeological site, which can be detected by analysis of remote sensing imagery. A variety of data sets were investigated, with the intention to combine multispectral bands to feed a direct detection algorithm, providing a semi-automatic process to cross-correlate the datasets. Sampling was done using locations of known sites, as well as areas with no archaeological evidence. The pre-processed very diverse remote sensing data sets have gone through a process of image registration. The algorithm was applied in the northwestern Dominican Republic on areas that included different types of environments, chosen for having sufficient imagery coverage, and a representative number of known locations of indigenous sites. The resulting maps present quantifiable statistical results of locations with similar pixel value combinations as the identified sites, indicating higher probability of archaeological evidence. The results show the variable potential of this method in diverse environments.

Keywords: Remote sensing; direct detection; GIS mapping; Caribbean Archaeology; landscape archaeology

1. Introduction

The fascination with feature identification and mapping of geometric archaeological alignments by means of remote sensing is as old as the first appearance of aerial photos [1–3]. Throughout the last centuries, it has advanced significantly, leading to new archaeological discoveries using imagery from satellites and drones [4, 5]. The human eye remains an adept feature extractor and can distinguish linear or circular structures and earthworks easily from the natural soil [6]. More recently, however, automatic approaches in pattern recognition have also become common, often based on computer algorithms adopted from other disciplines [7–10], and tested for archaeological purposes to detect color [11, 12], changes in topography [13, 14] or different reflection patterns [15].

A different challenge is the identification of non-geometric archaeological features with more amorphous shape and structure. Without any clear geometry, they pose a special problem, as the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2017 doi:10.20944/preprints201710.0070.v1

Peer-reviewed version available at *Geosciences* **2017**, *7*, 127; doi:10.3390/geosciences7040127

most prominent parameter for successful recognition is missing. This is the case for indigenous settlements in the Caribbean, which have been identified through assemblages of shells, ceramics bone remains, and stone tools; but not by traces of extant or sub-surface structural remains [16, 17]. The irregular pattern of pre-colonial settlement vestiges has made their detection challenging for remote sensing [18], and previous work has been dominated by traditional archaeological survey methods: the identification of surface material based on the knowledge of local scouts or landowners, and defining an approximate delineation of areas based on the surface finds on site [19, 20]. The trial approach presented here is an example for a novel statistical, systematic, and therefore more objective method.

Developed at *Cultural Site Research and Management (CSRM)* [21, 22], the *Direct Detection Model (DDM)* identifies the probability of sites by comparing single pixel values. This approach presupposes that anthropogenic activities at archaeological sites, often over long periods of time, have impacted these parts of the landscape in ways that if they persist are statistically measureable in remote sensing data. The DDM has therefore two sets of input data. The first set has two parts. The first are the locations of known archaeological sites. In the trial area of the northwestern Dominican Republic and Haiti, the archaeological 'sites' were identified over several years by different archaeologists, mostly with the help of local guides. Each site visited was named, a number of archaeological samples taken, and georeferenced by taking one or more GPS points at the site using a handheld device. The second part of the first is represented by areas with presumably no sites; these are equally important for the study. A second data set comprises a variety of remote sensing imagery. The subtle variation between already discovered areas of human activity, the sites, and areas of no human activity (non-sites) within each remote sensing band, can be used to detect difference. The difference is more likely to be detected when many different bands of available satellite or aerial data sets are combined.

The area of interest, northern Hispaniola, presents a highly diverse environment. Along the coast runs the 200 kilometers long Cordillera Septentrional, a several hundred meter high mountain range, partly covered by temperate to tropical forest, separating the coast from the fertile floodplain of the *Valle de Cibao*. Large parts of the hills and the plains north of the cordillera have been cleared for pasture. The northern coast is protected by coral reefs and mangrove forests. The region has been settled through waves of immigrations, archaeologically divided into earliest lithic age period since 4000 BC [23], the archaic period from 2500 BC the later ceramic ages distinguishable by *ostionoid*, *meillacoid* and *chicoid* ceramics [24–26, 23]. Shortly after the arrival of Columbus, and the foundation of the first Spanish town in the Americas at La Isabella in 1493 [27], evidence for Amerindian activity declines rapidly [28] from the archaeological record [29, 30]. We can therefore postulate that most sites marked in the map are either from prehistoric or very early colonial times. Variations in topography, land use and vegetation have created a landscape that changes over few kilometers, which also affected the indigenous settlement strategy [31]. Accumulations of shells indicate Amerindian use of marine resources [32], while other sites, often on prominent location overseeing the landscape, have been identified as settlements due to their particular topographic attributes consisting of mounds and flattened areas that served as base for house construction [33, 34, 30].

3. Materials and Methods

Based on the availability of remote sensing datasets, and samples of already identified archaeological sites, three areas of 5 x 5 kilometers in different environments were initially identified for trials. All existing archaeological site datasets were merged into a single point shape file, and then split for each of the trial regions. Polygons were created for the identified areas, whenever the site dimension had been measured. Only the two areas in Puerto Plata, DR (1) and Montecristi, DR (2) were ultimately trialed (Figure 1). The third area in Meillac (Dep. Nord-Est), Haiti, (3) was excluded following the second round of trials. An additional 1.5 x 5 kilometer area (4) that had been focus of a systematic total area survey [35] was initially thought to be well suited for comparing remote sensing and ground interpretation. Unfortunately it had to be discarded as there were not enough known sites in the area.

Peer-reviewed version available at Geosciences 2017, 7, 127; doi:10.3390/geosciences7040127

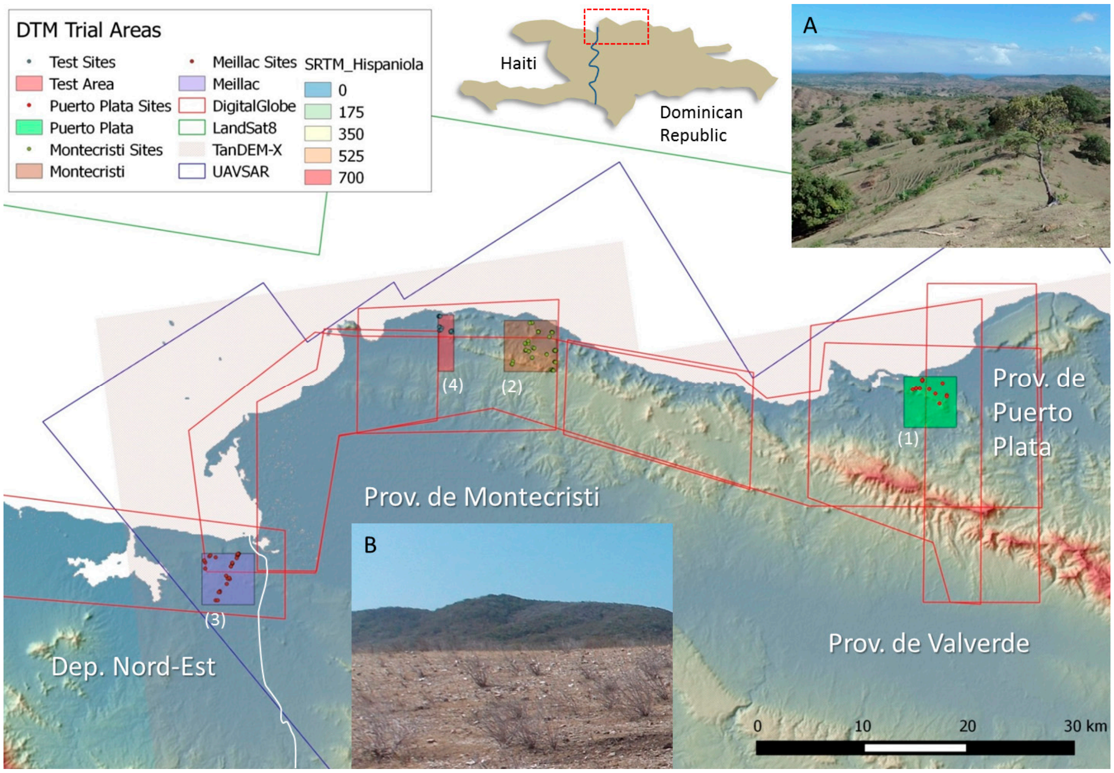

Figure 1. Initially selected trial areas in northern Hispaniola and the available remote sensing data sets superimposed on a modified NASA SRTM background. The small images display the (A) landscape in the Puerto Plata and (B) the view from an archaeological site in the Montecristi region (right).

Table 1. Availability of initially acquired data sets for sample sites Puerto Plata, DR (1) Montecristi, DR (2) Meillac, Haiti (3) and the Test site in the Montecristi (4). The regions were picked for very light or non-existing cloud cover within the images. In bold are the data sets included in the survey, while data sets in *italic* were rejected later for various reasons. MS = multispectral; PC = panchromatic.

	Dataset	Source	Bands	Resolution [m]	(1)	(2)	(3)	(4)
A	SRTM	USGS	1	30	x	x	x	X
B	LandSat-8	NASA/USGS	7 (MS) 1 (PC)	30 (MS) 15 (PC)	x	x	x	x
C	ASTER	NASA/METI/AIST	9	15	x	x	x	x
D	**UAVSAR**	**NASA/ JPL**	**6 (9)**	**5.7**	**x**	**x**	**x**	**x**
E	**WorldView-2**	**Digital Globe Foundation**	**8**	**1.85-2.07 (MS)**	**x**	**x**	**x**	**x**
F	Aerial	CNIGS (Govt. of Haiti)	3	.7			x	
G	TanDEM-X	DLR. e. V.	1	3	x	x	x	x

The passive remote sensing data sets *Landsat-8*, *Worldview-2*, *ASTER*, and active sensors *UAVSAR* as well as *TanDEM-X* stripmap (see Table 1 and Figure 2) were chosen based on resolution, availability, accessibility and practicality; they were either freely available, or acquired through generous data grants. Aerial imagery of northern Haiti was provided free of charge by Haiti's *Centre National de l'Information Géo-Spatiale*. Because of insufficient spatial resolution, work with Landsat-8, ASTER, and TanDEM-X was discontinued after consideration, leaving UAVSAR and Worldview-2 for further steps. The latter, multispectral data set, made available by the DigitalGlobe Foundation, covers the regions of interest in two-meter-resolution with one panchromatic and eight multispectral bands (see Table 2). The data set, with bands in the visible and near-visible range, was atmospherically corrected to reflectance values [36]. This standardized imagery removing artefacts caused by atmospheric interference. While often neglected, atmospheric correction is important and can significantly impact subsequent processing techniques like indices [37].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2017 doi:10.20944/preprints201710.0070.v1

Peer-reviewed version available at *Geosciences* **2017**, *7*, 127; doi:10.3390/geosciences7040127

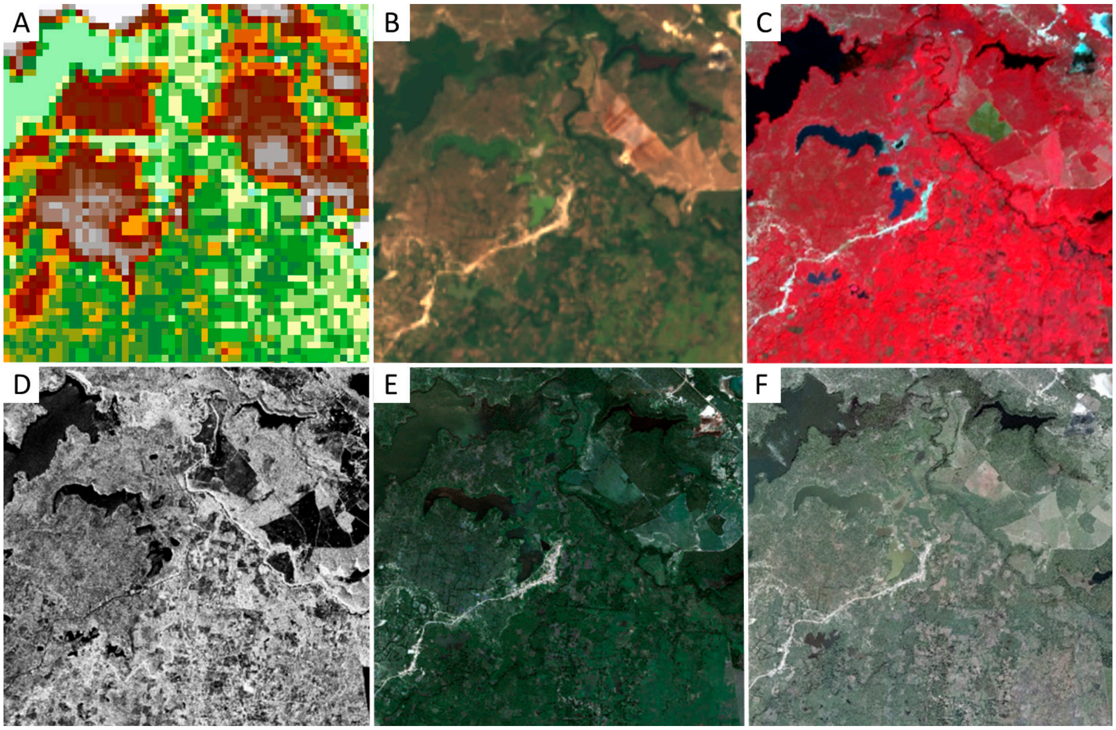

Figure 2. Overview of the initially trialed remote sensing data sets for the region Nordest Haiti. A) SRTM, B) LandSat, C) ASTER, D) UAVSAR, E) Worldview-2, F) Aerial.

Table 2. Band distribution and wavelength of Worldview-2 satellite**.**

Band	0	1	2	3	4	5	6	7	8
Color	Pan	Coastal	Blue	Green	Yellow	Red	Red Edge	NIR1	NIR2
λ in nm		400-450	450-510	510-580	585-625	630-690	705-745	770-895	860-1040

From the original Worldview-2 data, the transformations *NDVI*, *PCA* and *Tasseled Cap* were applied with the purpose to create additional bands that may improve the site identification regarding their environmental discrimination. Of these, the NDVI (Normalized Difference Vegetation Index) [38] is a unidimensional spectral index, adjusting the band information based on the principle that healthy vegetation absorbs most of the VIS light and reflects most of the NIR light. Unhealthy or sparse vegetation reflects more VIS light and less NIR light. The formula applied used the bands red and red edge:

NDVI: Float ("Red Edge"-"Red") / ("Red Edge"+"Red")

Principal Component Analysis (PCA) was applied with the intention to reduce the data dimensionality of correlated bands [39]. The method rotates the original space of features into a space where the transformed features are pairwise orthogonal. This creates an n-dimensional space of eigenvectors, where n is the number of input dimensions (features), with the goal to orthogonalize the data set. The first principal component accounts for the maximum proportion of variance from the original dataset, the following, being orthogonal to the first one, for the next principal components, creating eventually a new coordinate system of orthogonal axes. A subset of the components is usually chosen for subsequent analysis. The method used to select these components varies by application. The first three components were included in the algorithm while the latter components were discarded as redundant. For a more detailed explanation, see [40].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2017 doi:10.20944/preprints201710.0070.v1

Peer-reviewed version available at Geosciences 2017, 7, 127; doi:10.3390/geosciences7040127

Tasseled Cap Transformation or *K-T transform*, as originally developed by [41] for LandSAT imagery, was applied on each Worldview-2 data set using bands one to eight in accordance with [42]. Tasseled cap applies predefined correction coefficients to each band and will produce eight new bands. This spectral index conversion intends to highlight changes in vegetation and soil, where the pixel values are being transferred into a new orthogonal axial system; of these the first three new bands are the most important, representing *Brightness* (red), *Wetness* or yellowness of vegetation (blue) and *Greenness* (green). Based on the reflectance values given by [42] for each Worldview-2 component, the formulas go as such:

Brightness: Float (0.060436*"Coastal"+0.012147*"Blue"+0.125846*"Green"+0.313039*"Yellow" +0.412175*"Red"+0.482758*"Red Edge"-0.160654*"NIR1"+0.67351*"NIR2")

Greenness: Float (-0.140191*"Coastal"-0.206224*"Blue"-0.215854*"Green"-0.314441* "Yellow"-0.410892*"Red"+0.095786*"Red Edge" +0.600549*"NIR1"+0.503672*"NIR2")

Wetness & Shadow: Float (-0.270951*"Coastal" -0.315708*"Blue" -0.317263*"Green" -0.242544*"Yellow" -0.256463*"Red"-0.096550*"Red Edge"-0.742535*"NIR1"+0.202430*"NIR2")

In addition to passive remote sensing data, NASA had captured UAVSAR (Uninhabited Aerial Vehicle Synthetic Aperture Radar) polarimetric L-band data of ~5.7 meter, over the large fault zones of Hispaniola. Publicly available, this data set also covers the areas of interest. Accessed through the JPL/ASF website, the data was extracted to single band TIFF-files using [43] (downloaded product: PolSAR- polarimetric SAR – MLC). Different materials reflect radar waves with different intensities and polarizations. Among the feature differentiated are smoothness, homogeneity, and correlation, as well as soil moisture, and vegetation discrimination revealed by variation in density and structure. This is highlighted by the three color channels of a synthesized *Pauli Decomposition* image (see Table 3). From the original seven bands, Pauli decomposition bands were produced through [44], to represent all the polarimetric information in a single SAR image.

Table 3. UAVSAR bands as extracted to create Pauli decomposition bands.

Band	HHHH	HHHV	HHVV	HVHV	HVVV	VVVV		HV	HH-VV	HH+VV
Real	+	+	+	+	+	+	**Name**	Pauli3	Pauli2	Pauli1
Imaginary	-	+	+	-	+	-	**Code**	Red	Green	Blue

Atmospherically corrected 15m ASTER (*Advanced Spaceborne Thermal Emission and Reflection*) data was acquired through NASA/METI/AIST/Japan Space Systems and U.S./Japan ASTER Science Team, the low resolution however made the imagery of limited use. Additionally, several TanDEM-X data sets were acquired through a DLR e.V. research grant, but the uncorrected data was not utilized for the DDM. All remote sensing data went through a series of image registration protocols to render them standard in pixel size, resolution and angle that allowed exact correlation between pixels of different data sets. To achieve this goal, all datasets were initially converted to the same georeferenced system: WGS 84, UTM 19N for the Dominican Republic (20N respectively for Haiti) and a 5000 x 5000 meter area was resampled using a 2m grid. This created a 2500 x 2500 grid of points which were then used to sample data from each dataset (Figure 3). These values were then interpolated into a stack of registered raster dataset. After this transformation, the pixels and their attributes of each band were exactly overlapping, diminishing the possibility of corner and border uncertainties. In addition, the prepared Worldview-2 data set served as base for land cover classification (Table 4), to better distinguish the variety of surface coverage.

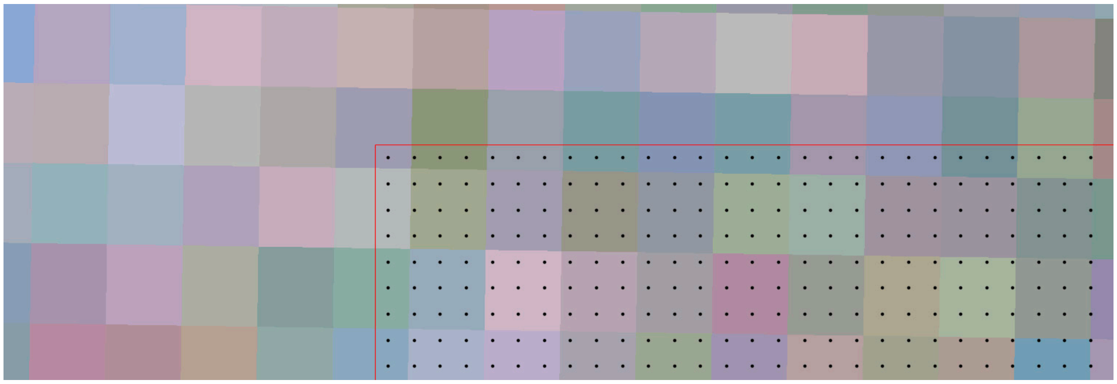

Figure 3. The image displays the necessity for point distribution and rearrangement of pixels on UAVSAR Pauli decompensated data.

Table 4. The three-band RGB combination of Worldview 2 was used to create land cover classification for each site.

	(1) Puerto Plata	(2) Montecristi	(3) Haiti	(4) Test site
1	Water	Water	Water	Water
2	Flat Surfaces	Mangrove	Mangrove	Bare Soil
3	Mangrove	Bare Earth	Structures & Roads	Forest
4	Forest	Built	Forest	Shrub
5	Eroded Land	Forest	Shrubs	
6	Pasture	Shrub	Pasture	
7	Structure	Clouds	Dump Site	

3. Results

3.1 Posterior Probability Approach

Initial tests on a modified version of the algorithm, using a posterior probability modeling [45–48] to define difference between potential areas of sites and non-sites were conducted with focus on a trial area in Puerto Plata, DR. This approach had been successfully applied elsewhere and involved using known sites and alleged non-sites to build a binary classifier where each cell was assigned a posterior probability of being an archaeological site. Datasets included Woldview-2 imagery and band difference ratios similar to the NDVI, which were then reduced using PCA. Results were positive with two caveats. Firstly, the algorithm was based on a binary classification might be more effective when identifying homogenous site-types like lithic scatters. Secondly, the algorithm performed better with a larger and very accurate sample of known sites and checked known non-sites of a surveyed area. In this project, the heterogeneous nature of the sites coupled with a small number of known sites must be regarded as an impediment to this approach. The sites in the original dataset were represented by single artefact find spots around a central point which represented the site proper. Polygons were digitized around all points to delineate sites. There were 12 sites in the area with an average area of 4338 m². Non-Sites were generated according to the following rules:

- In surveyed areas
- > 100 meters from known sites
- Buffers with a radius of 37 meters (with an area of 4300 m²) were generated around each site.

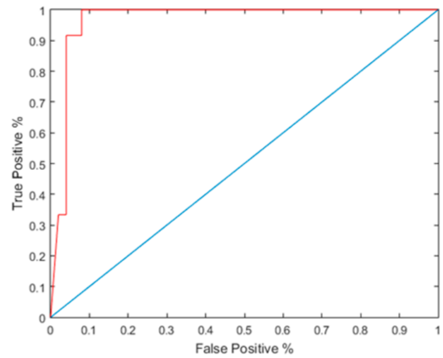

Figure 4. Receiver Operating Characteristics (ROC curve) of the posterior probability approach from Puerto Plata.

The approach may have been hindered by the dominance of sites in the mangroves, and on hilltops; the algorithm favored these areas for probable site locations. The trial results were plotted on a ROC curve (Figure 4), which demonstrates that sites were much more likely to be found in the high or higher probability areas of the posterior probability maps of Puerto Plata (Figure 5), and much less likely in the low probability areas.

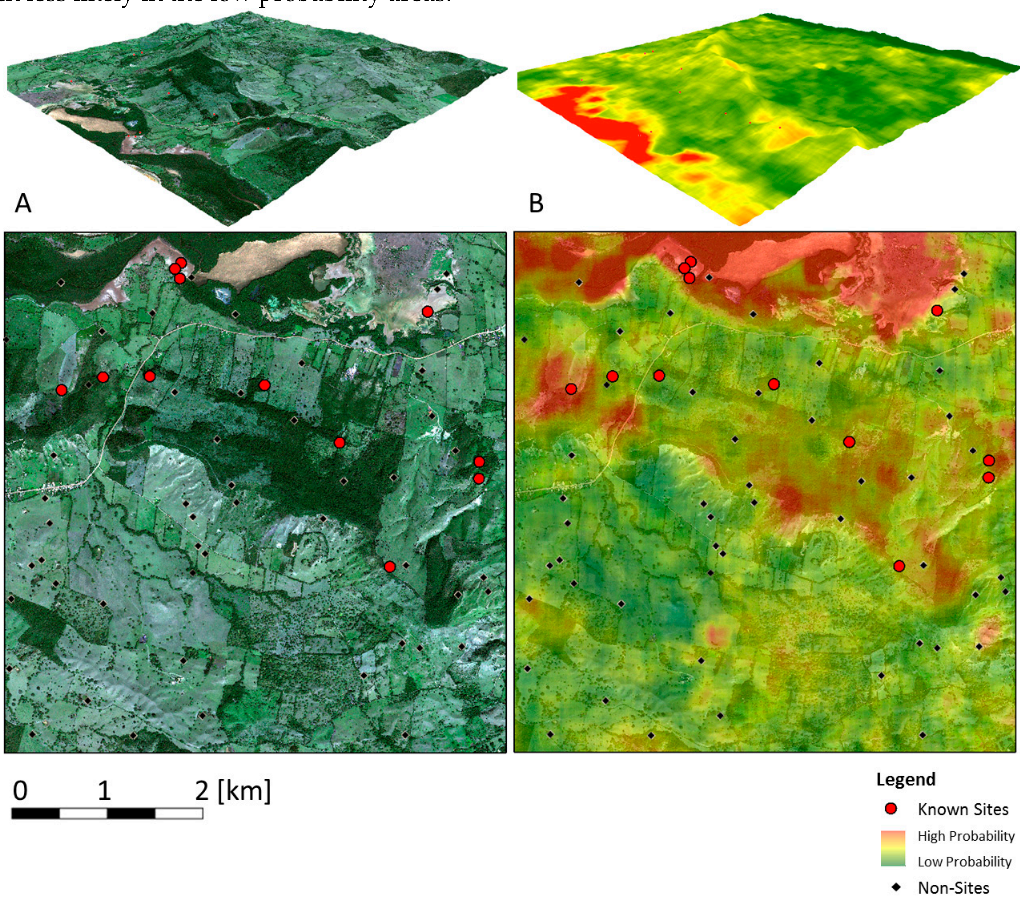

Figure 5. Posterior probability results from Puerto Plata in topographic and top down view. A) the original RGB data, B) overlaid by the resulting posterior probability map.

3.2 Frequentist protocols

A frequentist [22] approach was applied for the Montecristi area (Figure 6). This had been programmed in the statistical software *R* [49]. This 5 x 5 kilometers area is located in a hilly part of the coastal region of Montecristi, where new sites had recently been identified [31, 50, 35], of which 16 sites were chosen.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2017 doi:10.20944/preprints201710.0070.v1

Peer-reviewed version available at Geosciences 2017, 7, 127; doi:10.3390/geosciences7040127

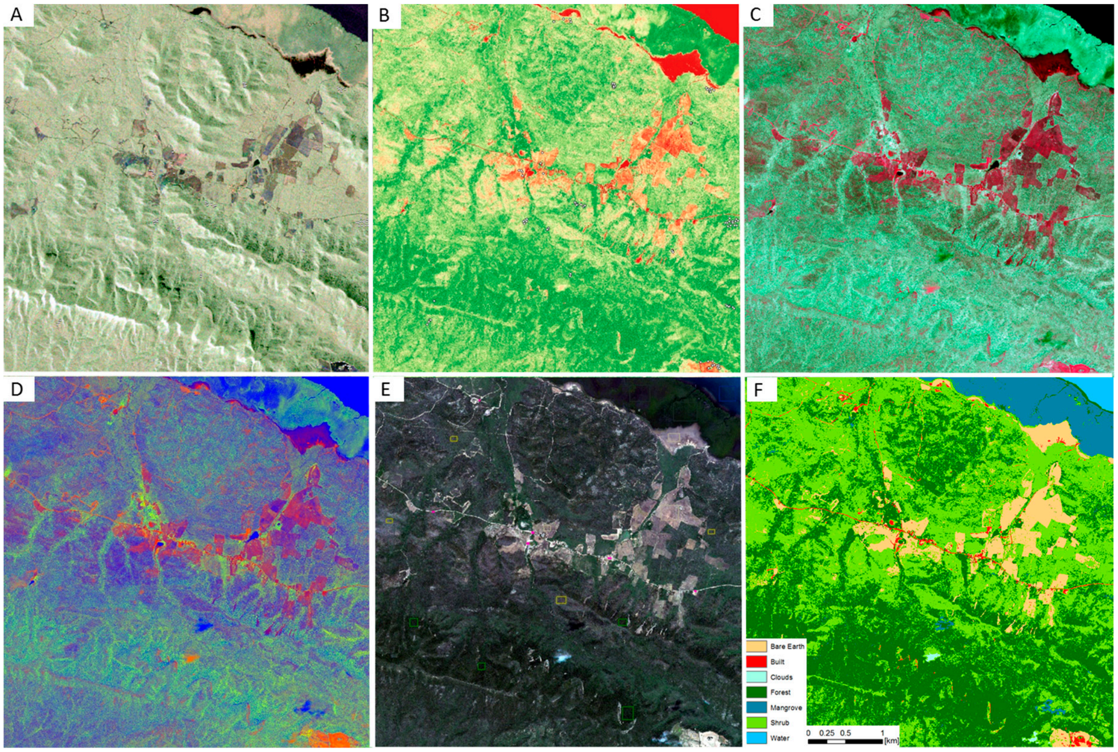

Figure 6. Various combined remote sensing data sets of the Montecristi trial area. A) UAVSAR Pauli Decomposition. B) Worldview-2 NDVI. C) Worldview PCA. D) Worldview Tasseled Cap. E) Worldview RGB with rectangles defined for F) land cover classification. A white cloud can be seen in the lower center of (E).

A window of 81 x 81 pixels (160x160meters) was set around the single pixels picked as the center of the known sites (KS) and randomly selected non-sites (NS) creating a base of information of 6561 pixels, across each band, for each point of interest. The same number of known sites and non-sites was considered. Histograms were generated for each separate band across sites and non-sites. The histograms were binned in 100 equally spaced separations (see Figure 7). A student t-test/Wilcoxon rank sum test was conducted to see if there is a significant dissimilarity between sites and non-sites.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2017 doi:10.20944/preprints201710.0070.v1

Peer-reviewed version available at Geosciences 2017, 7, 127; doi:10.3390/geosciences7040127

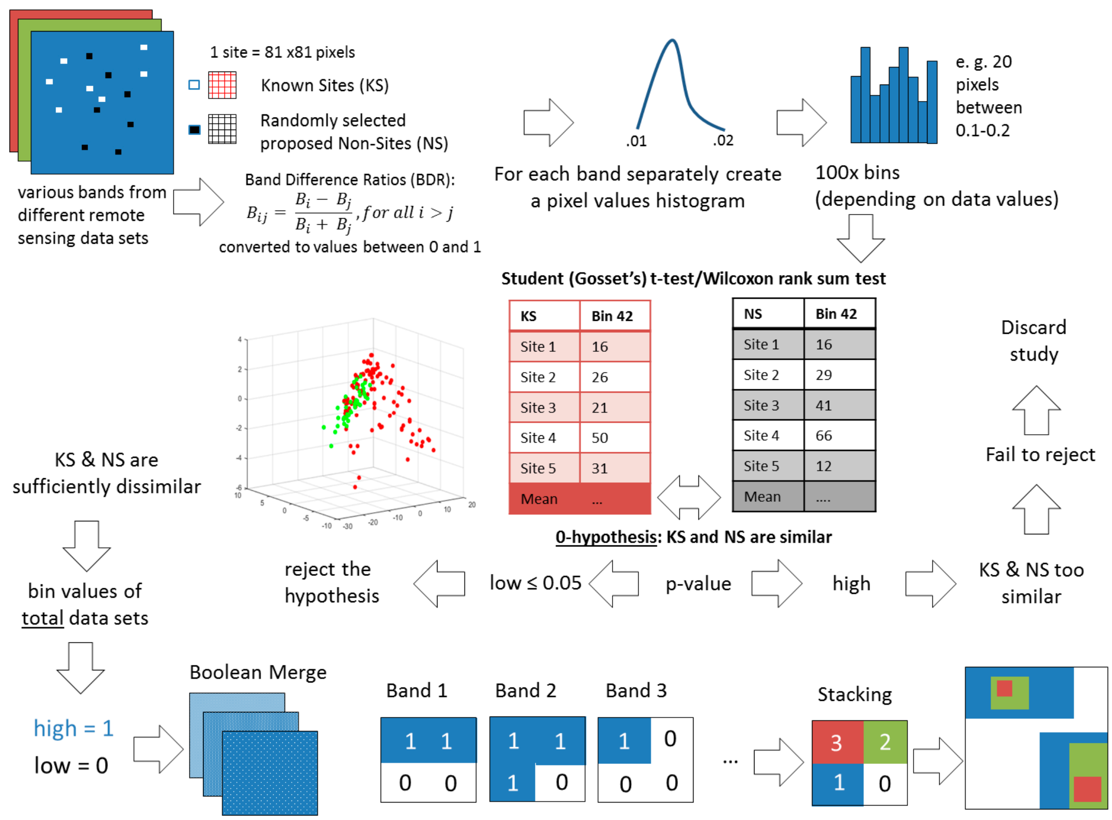

Figure 7. Sketch of the frequentist protocol algorithm. Student's (Gosset's) T-test/Wilcoxon rank sum test is applied to determine, if the distributions of site and non-site pixels in individual bins are statistically significantly different, with 0- hypothesis being they are from the same distribution.

3.3. Dominance of bands

A variety of statistical trial calculations were applied. A band-difference ratio (BDR) was generated among every band included in the algorithm data set, to reduce the dominance of particular as well as essentially redundant data sets. These ratios were indices similar to the NDVI and normalized datasets. Only bands with the lowest positive response rate, a low p-value (cause for rejecting the 0-hypothesis that KS and NS were similar) were further considered for the tests. The highly diverse environment, as made visible in the land cover maps, would, one might expect, influence the success of the approach in comparison with other areas where land cover was more homogenous [51].

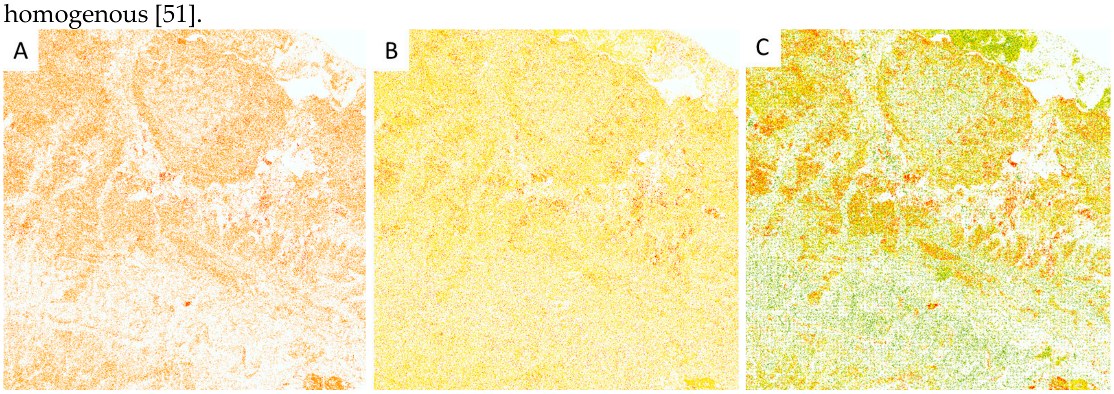

Figure 8. Results from the frequentist tests in the Montecristi area, using a total of 21 bands. A) Sturges rule: highest value after Boolean merge: 2, B) based on Scott: highest value after Boolean merge: 4 C) Sturges rule using BDR: highest combination after Boolean merge: 7.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2017 doi:10.20944/preprints201710.0070.v1

Peer-reviewed version available at Geosciences 2017, 7, 127; doi:10.3390/geosciences7040127

The frequentist protocol from [21, 51] was implemented in R with different binning strategies [52], and [53] using Student's (Gosset's) T-Test or the Wilcoxon rank sum test (for explanations of these tests, see [54]). The bulk of the work in R was done as exploratory data analysis with mixing and matching binning strategies and hypothesis testing. The results vary strongly on different numbers and combinations, based on the variety band different ratios, and statistical tests (Figure 8).

4. Discussion

The final result that incorporated the land cover information shows a definite response to the diverse landscape represented in the image (Figure 9). Several aspects are notable: As anticipated, without sites mapped in the mangrove area, No. (1), this area remains completely void of site activity. The forested areas also appear relatively unresponsive. Since large parts of the survey area are covered by areas defined as dense forest, the random distribution algorithm put more non-sites into forested areas which may likely have had an effect on the non-sites statistics. Most significant high response areas are found in locations with little vegetation where the dimension of these areas can be better defined. This expected best response rate is confirmed by the bright red colored areas surrounding these sites, showing that in these locations the algorithm shows its best strength. However, a significant number of sites had been marked in areas covered by forest and shrub, which, in our tests at least, do not respond well to the DDM search protocols using only the data sets available. It can be expected that the reflection value of sites in bare earth areas should differ significantly from non-sites here than in forested sites, as the scatter of archaeological material is better displayed on the surface, particularly in ploughed areas, while canopy vegetation does not appear particularly affected by it.

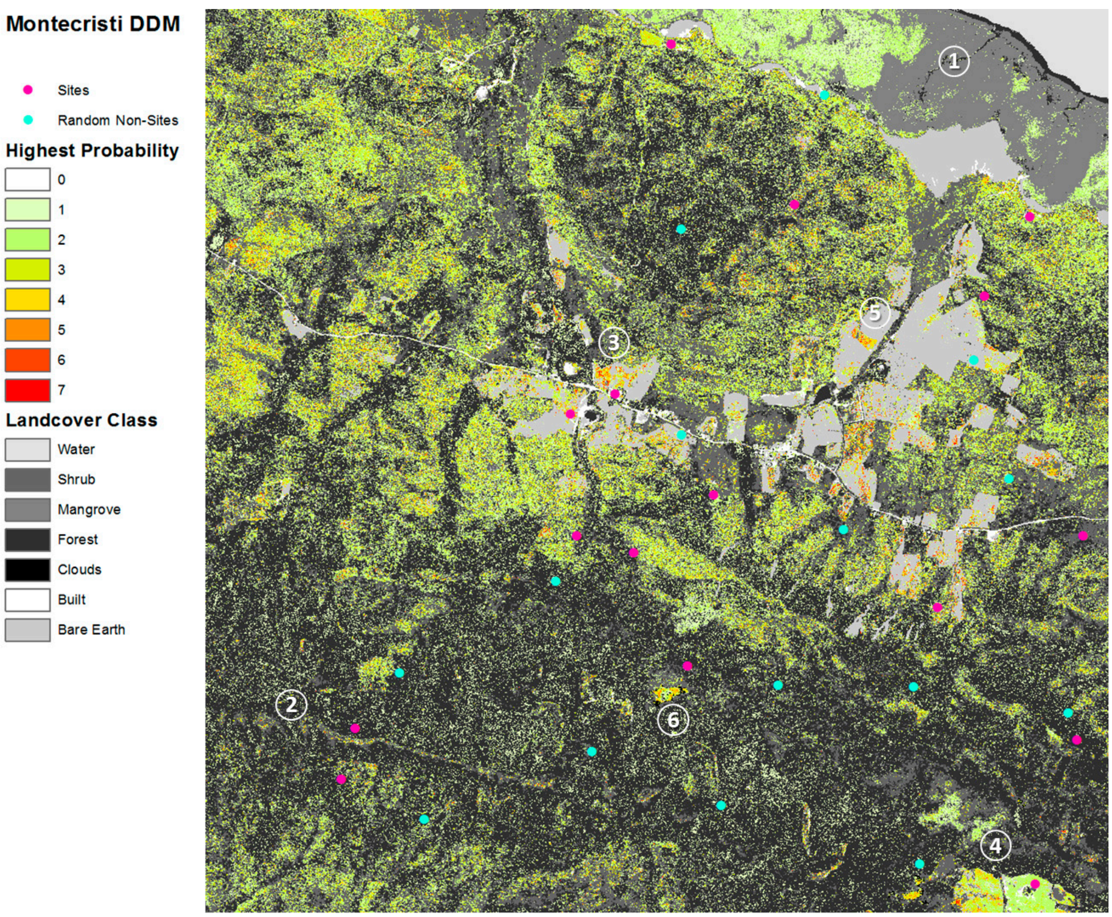

Figure 9. DDM frequentist results based on 8 UAVSAR bands, 3 Pauli Decomposition, 8 WV-2, NDVI, KTT and best BDR outcome.

Peer-reviewed version available at Geosciences 2017, 7, 127; doi:10.3390/geosciences7040127

From an archaeological interpretative view, the higher values do not necessarily represent an ancient pattern of settlement selection, but a combination of features that seem to be the trend in this particular area. It remains uncertain if the DDM corresponds to areas that follow attributes based on previous [20] and current research that served predictive models [35], a pattern that expresses tendencies, such as proximity to the sea, or other sea features such as mangrove forest, proximity to brooks, proximity to flat lands (usually less forested), and elevation less than 100m. Modern settlements have been built near areas that combined the aforementioned features, as these also allow the development of crops. The high valued pixels of the DDM show zones in which these features have been combined, and could be a reason to have also highlighted certain areas of current habitation. For the northern part the model created seems to correlate this indigenous activity pattern, possibly though related to other factors. While the topography was not taken in consideration due to its low resolution, interpreted visually, these areas respond to areas of low probability. In the center and south of the area the two known sites at location (3) in Fig. 9 are from two extensive sites on grassland, surrounding a former school yard. Here the results appear to delineate the area of the assemblage of material. Location (4) in the southwest corner seems to pick up a small site near the large site of El Manantial (MC-44, [18] only separated by a small gorge. The intensity showing (5) was identified as a modern dump site, (6) represents the above mentioned small cloud.

5. Conclusion

Automatic detection models for archaeology, particularly the idea of predictive modeling, have been under heavy scrutiny since their appearance in archaeological research [55–57], with predominant questioning as to whether the time and effort invested served the outcome. Leaving decisions not completely to the machine but guiding the solution finding the improving and more advanced and fast algorithms semi-automatically, shows great potential for breakthroughs in the detection to amorphous archaeological features in the future.

Regarding the applied frequentist algorithm, it was shown in previous studies at non-forested locations that the applied algorithms was particularly useful in otherwise uniform environments to identify archaeological [21] or geological features [47]. The anticipated significant differentiation between sites and non-sites on northern Hispaniola was overshadowed by the immense environmental variation in the surveyed region. Many strong factors weigh in that made it particularly difficult for the algorithm to distinguish archaeological sites from areas with little archaeological potential.

Improvements could be made, by using instead of a single point with a square of 81x81 pixels, an average of the pixel values inside an actually determined area extent of a site, as it could have been used for the small trial area, where sites had been identified by systematic survey. This would have provided a more precise fingerprint in comparison to the non-sites. Also, picking non-sites randomly from different environments may have enhanced the probability that with very bad luck an actual not yet identified site would have been selected. A point of critique could also be the use of only two completely different datasets; another might be that these data sets were used to produce synthetic bands. Also, the vegetation types or patterns in forested covered areas produce a diversity that could only be differentiated with additional data. A highly distinguishable feature of some identified sites, the topography, as identified through drone photogrammetry [34] could be an important factor to significantly improve the study, but for this the access to high resolution regional LiDAR data would be crucial.

To conclude, the study has to be seen as a trial to test and improve possibilities to semi-automatically identify areas with non-structural archaeological potential in diverse environments: this leaves great potential for future tasks to evaluate regions for unknown and potentially threatened heritage and archaeology automatically.

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n°319209. We

Peer-reviewed version available at *Geosciences* **2017**, *7*, 127; doi:10.3390/geosciences7040127

would like to thank Jorge Ulloa Hung who provided the archaeological sites test data set for the Puerto Plata region. The Digital Globe Foundation provided a generous data grant of Worldview-2 imagery for the region. We thank NGA/NASA/USGS for making their SRTM DEM data available to us. The ASTER L1A data product is courtesy of the online Data Pool at the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota (https://lpdaac.usgs.gov/data_access).

Author Contributions

T.S. managed and oversaw all the sub elements of the study, undertook the image and data preparation in the GIS. D.C. conceived, designed the direct detection and posterior probability approach and coordinated its implementation. J.P. implemented and performed the analysis of the direct detection protocols in R. W.M. implemented the data into the posterior probability approach. E.H. contributed the archaeological site data and assisted in the archaeological interpretation. As a principal investigator of the ERC-Nexus1942, C.H. provided the funding for study, coordinated the field work in the Dominican Republic, and contributed on discussing the archaeological validity of the study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Batut, A. *La Photographie aérienne par cerf-volant;* Gauthier-Villars: Paris, 1890.
2. Capper, J.E. Photographs of Stonehenge as seen from a War Balloon. *Archaeologia: or Miscellaneous Tracts Relating to Antiquity* **1907**, *60*, 571.
3. Millhauser, J.K.; Morehart, C.T. The Ambivalence of Maps: A Historical Perspective on Sensing and Representing Space in Mesoamerica. In *Digital Methods and Remote Sensing in Archaeology*; Forte, M., Campana, S., Eds.; Springer Books: New York, 2016; pp 247–268.
4. Wiseman, J.; El-Baz, F. *Remote sensing in archaeology;* Springer: New York NY, 2007.
5. Comer, D.C.; Harrower, M.J. *Mapping Archaeological Landscapes from Space: In Observance of the 40th Anniversary of the World Heritage Convention;* Springer Press: New York, 2013.
6. Carrol, D.M.; Evans, R.; Bendelow, V.C. *Air photo-interpretation for soil mapping;* Soil Survey: Harpenden, 1977.
7. Trier, O.D.; Larsen, S.O.; Solberg, R. Automatic detection of circular structures in high-resolution satellite images of agricultural land. *Archaeological Prospection* **2009**, *16*, 1–15, doi:10.1002/arp.339.
8. Di Iorio, A.; Bridgwood, I.; Schultz Rasmussen, M.; Kamp Sorensen, M.; Carlucci, R.; Bernardini, F.; Osman, A. Automatic detection of archaeological sites using a hybrid process of Remote Sensing, Gis techniques and a shape detection algorithm. *Proceedings of the 3rd International Conference on Information and Communication Technologies: From Theory to Applications*, 2008.
9. Schuetter, J.; Goel, P.; McCorriston, J.; Park, J.; Senn, M.; Harrower, M. Autodetection of ancient Arabian tombs in high-resolution satellite imagery. *International Journal of Remote Sensing* **2013**, *34*, 6611–6635.
10. Zingman, I.; Saupe, D.; Lambers, K. Automated search for livestock enclosures of rectangular shape in remotely sensed imagery. *Proc. of SPIE - Image and Signal Processing for Remote Sensing*, 2013.
11. Traviglia, A. MIVIS Hyperspectral Sensors for the Detection and GIS Supported Interpretation of Subsoil Archaeological Sites. In *Digital Discovery. Exploring New Frontiers in Human Heritage*, CAA2006. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 34th Conference; Clark, J.T., Hagemeister, E.M., Eds.; Archaeolingua: Budapest, 2006; CD-ROM 302-314.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 October 2017 doi:10.20944/preprints201710.0070.v1

Peer-reviewed version available at *Geosciences* **2017**, *7*, 127; doi:10.3390/geosciences7040127

12. Doneus, M.; Verhoeven, G.; Atzberger, C.; Wess, M.; Ruš, M. New ways to extract archaeological information from hyperspectral pixels. *Journal of Archaeological Science* **2014**, *52*, 84–96, doi:10.1016/j.jas.2014.08.023.
13. Menze, B.H.; Ur, J.A.; Sherratt, A.G. Detection of ancient settlement mounds: archaeological survey based on the SRTM terrain model. *Photogrammetric Engineering and Remote Sensing* **2006**, *72*, 321–327.
14. Opitz, R.S.; Cowley, D. *Interpreting archaeological topography. Airborne laser scanning, 3D data and ground observation; [3D data, visualisation and observation]*, Oxford [u.a.] : Oxbow Books, 2013.
15. Stewart, C.; Lasaponara, R.; Schiavon, G. ALOS PALSAR Analysis of the Archaeological Site of Pelusium. *Archaeol. Prospect.* **2013**, *20*, 109–116, doi:10.1002/arp.1447.
16. Rouse, I. Pattern and Process in West Indian Archaeology. *World Archaeology* **1977**, *9*, 1–11.
17. *The Oxford Handbook of Caribbean Archaeology;* Keegan, W.F.; Hofman, C.L.; Ramos, R.R., Eds.; Oxford University Press, 2013.
18. Sonnemann, T.F.; Herrera Malatesta, E.; Hofman, C.L. Applying UAS photogrammetry to analyse spatial patterns of Amerindian settlement sites in the northern DR. In *Digital Methods and Remote Sensing in Archaeology*; Forte, M., Campana, S., Eds.; Springer Books: New York, 2016; pp 71–87.
19. Ortega, E.; Denis, P.; Olsen Bogaert, H. Nuevos yacimientos arqueológicos en Arroyo Caña. *Boletín del Museo del Hombre Dominicano* **1990**, *23*, 29–40.
20. Ulloa Hung, J. *Arqueología en la Línea Noroeste de La Española Paisajes, Cerámicas e Interacciones;* Instituto Tecnológico de Santo Domingo: Santo Domingo, Dominican Republic, 2014.
21. Comer, D.C. *Merging Aerial Synthetic Aperture Radar (SAR) and Satellite Multispectral Data to Inventory Archaeological Sites*. *Report on file*, 2007. https://www.ncptt.nps.gov/download/28370/ (accessed on 9 June 2017).
22. Comer, D.C.; Blom, R.G. Detection and identification of archaeological sites and features using synthetic aperture radar (SAR) data collected from airborne platforms. In *Remote sensing in archaeology*; Wiseman, J., El-Baz, F., Eds.; Springer: New York NY, 2007; pp 103–136.
23. Keegan, W.F.; Hofman, C.L. *The Carribean before Columbus;* Oxford University Press: New York, 2017.
24. Veloz Maggiolo, M.; Ortega, E.; Caba, Á. *Los Modos de Vida Meillacoides y sus Posibles Orígenes;* Editora Taller: Santo Domingo, Dominican Republic, 1981.
25. Sinelli, P.T. Meillacoid and the Origins of Classic Taíno Society. In *The Oxford Handbook of Caribbean Archaeology*; Keegan, W.F., Hofman, C.L., Ramos, R.R., Eds.; Oxford University Press, 2013.
26. Ting, C.; Neyt, B.; Ulloa Hung, J.; Hofman, C.; Degryse, P. The production of pre-Colonial ceramics in northwestern Hispaniola: A technological study of Meillacoid and Chicoid ceramics from La Luperona and El Flaco, Dominican Republic. *Journal of Archaeological Science: Reports* **2016**, *6*, 376–385, doi:10.1016/j.jasrep.2016.02.031.
27. Deagan, K.A.; Cruxent, J.M. *Archaeology at La Isabela. America's first European town;* Yale University Press: New Haven, 2002.
28. Casas, B. de las. *Historia de Las Indias TII;* Imprenta de Miguel Ginesta: Madrid, Spain, 1875.
29. Rouse, I. *The Tainos. ;. Rise and Decline of the People Who Greeted Columbus;* Yale University Press, 1993.
30. Hofman, C.L.; Ulloa Hung, J.; Herrera Malatesta, E.; Jean, J.S.; Hoogland, M.L. Indigenous Caribbean perspectives.: Archaeologies and legacies of the first colonized 416 region in the New World. *Antiquity* (in press).

Peer-reviewed version available at *Geosciences* **2017**, *7*, 127; doi:10.3390/geosciences7040127

31. Ulloa Hung, J.; Herrera Malatesta, E. Investigaciones Arqueológicas en el Norte de La Española, Entre Viejos Esquemas y Nuevos Datos. *Boletín del Museo del Hombre Dominicano* **2015**, *46*, 75–107.
32. Rouse, I. Areas and Periods of Culture in the Greater Antilles. *Southwestern Journal of Anthropology* **1951**, *7*, 248–264.
33. Hofman, C.L.; Hoogland, M.L. Investigaciones arqueológicas en los sitios El Flaco (Loma de Guayacanes) y La Luperona (Unijica).: Informe pre-liminar. *Boletín del Museo del Hombre Dominicano* **2015**, *46*, 61–74.
34. Sonnemann, T.F.; Ulloa Hung, J.; Hofman, C.L. Mapping indigenous settlement topography in the Caribbean using drones. *Remote Sensing* **2016**, *8*, *doi:*10.3390/rs8100791.
35. Herrera Malatesta, E. Una isla, dos mundos: Sobre la transformación del paisaje indígena de Bohío a La Española. PhD dissertation; Leiden University, Leiden, The Netherlands, in press.
36. Bernstein, L.S. Quick atmospheric correction code: Algorithm description and recent upgrades. *Opt. Eng* **2012**, *51*, 111719, doi:10.1117/1.OE.51.11.111719.
37. Hadjimitsis, D.G.; Papadavid, G.; Agapiou, A.; Themistocleous, K.; Hadjimitsis, M.G.; Retalis, A.; Michaelides, S.; Chrysoulakis, N.; Toulios, L.; Clayton, C.R.I. Atmospheric correction for satellite remotely sensed data intended for agricultural applications: Impact on vegetation indices. *Nat. Hazards Earth Syst. Sci.* **2010**, *10*, 89–95, doi:10.5194/nhess-10-89-2010.
38. Rouse, J.W.; Haas, R.H.; Scheel, J.A.; Deering, D.W. Monitoring Vegetation Systems in the Great Plains with ERTS. *Proceedings, 3rd Earth Resource Technology Satellite (ERTS) Symposium* **1974**, 48–62.
39. Eastman, J.R.; Filk, M. Long sequence time series evaluation using standardized principal components. *Photogrammetric Engineering and Remote Sensing* **1993**, *59*, 991–996.
40. Faraway, J.J. *Linear Models with R, Second Edition*, 2nd ed.; CRC Press: Hoboken, 2015.
41. Kauth, R.J.; Thomas, G.S. The tasseled Cap: A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by LANDSAT. *LARS Symposia*; Purdue University: West Lafayette, Indiana, 1976.
42. Yarbrough, L.D.; Navulur, K.; Ravi, R. Tassled Cap Presentation of the Kauth–Thomas transform for WorldView-2 reflectance data. *Remote Sensing Letters* **2014**, *5*, 131–138.
43. Alaska Satellite Facility. *MapReady;* NASA, 2014.
44. IETR (Institute of Electronics and Telecommunications of Rennes - UMR CNRS 6164. *PolSARpro;* ESA, 2014.
45. Chen, L.; Priebe, C.E.; Sussmann, D.L.; Comer, D.C.; Megarry, W.P.; Tilton, J.C. Enhanced Archaeological Predictive Modelling in Space Archaeology. http://arxiv.org/abs/1301.2738 (accessed on 15 January 2016).
46. Chen, L.; Comer, D.C.; Priebe, C.E.; Sussmann, D.; Tilton, J.C. Refinement of a method for identifying probable archaeological sites from remotely sensed data. In *Mapping Archaeological Landscapes from Space: In Observance of the 40th Anniversary of the World Heritage Convention*; Comer, D.C., Harrower, M.J., Eds.; Springer Press: New York, 2013; pp 251–258.
47. Megarry, W.P.; Cooney, G.; Comer, D.C.; Priebe, C.E. Posterior Probability Modeling and Image Classification for Archaeological Site Prospection: Building a Survey Efficacy Model for Identifying Neolithic Felsite Workshops in the Shetland Islands. *Remote Sensing* **2016**, *8*.
48. Comer, D.C. *Institutionalizing Protocols for Wide-Area Inventory of Archaeological Sites by the Analysis of Aerial and Satellite Imagery*. Project Number 11-158: United States of America Department of Defense, Legacy Program, Washington, DC. Manuscript on File, 2014.
49. R Core Team. *R;* R Foundation for Statistical Computing: Vienna, Austria, 2015.

50. Herrera Malatesta, E. Understanding ancient patterns: Predictive Modeling for field research in Northern Dominican Republic. In *Proceedings of the 26th Congress of the IACA.* International Association of Caribbean Archaeologists, Sint Maarten; Velasquez, C.B., Haviser, J.B., Eds., 2017; pp 88–97.
51. Tilton, J.C.; Comer, D.C. Identifying Probable Archaeological Sites on Santa Catalina Island, California Using SAR and Ikonos Data. In *Mapping Archaeological Landscapes from Space: In Observance of the 40th Anniversary of the World Heritage Convention*; Comer, D.C., Harrower, M.J., Eds.; Springer Press: New York, 2013.
52. Sturges, H.A. The choice of a class interval. *Journal of the American Statistical Association* **1926**, *21*, 65–66.
53. Scott, D.W. On optimal and data-based histograms. *Biometrika* **1979**, *66*, 605–610, doi:10.1093/biomet/66.3.605.
54. Rice, J.A. *Mathematical statistics and data analysis*, 3. ed., internat. ed.; Thomson/Brooks/Cole: Belmont, Calif., 2007.
55. Kvamme, K.L. Development and testing of quantitative models. In *Quantifying the Present and Predicting the Past: Theory, Methods, and Applications of Archaeological Predictive Modeling;*; Judge, W.J., Sebastian, L., Eds.; US Department of Interior, Bureau of Land Management Service Center: Denver, CO, USA, 1988; pp 325–428.
56. *Archaeological prediction and risk management*. *Alternatives to current practice;* Kamermans, H.; van Leusen, M.; Verhagen, P., Eds.; Leiden Univ. Press: Leiden, 2009.
57. Verhagen, P.; Whitley, T.G. Integrating Archaeological Theory and Predictive Modeling: A Live Report from the Scene. *J Archaeol Method Theory* **2012**, *19*, 49–100, doi:10.1007/s10816-011-9102-7.