

TITLE PAGE

Classification:

PHYSICAL SCIENCES: Earth, Atmospheric, and Planetary Sciences

SOCIAL SCIENCES: Anthropology

Title:

Automated detection of archaeological mounds using machine learning classification of multi-sensor and multi-temporal satellite data

Author Line:

Hector A. Orengo^{1*}, Francesc C. Conesa^{2*}, Arnau Garcia-Molsosa², Agustín Lobo³, Adam S. Green², Marco Madella^{4,5,6} and Cameron A. Petrie^{2,7}

Author affiliation:

¹ Catalan Institute of Classical Archaeology, Plaça d'en Rovellat, s/n · 43003 Tarragona, Spain

² McDonald Institute for Archaeological Research, University of Cambridge. Downing Street, CB2 3ER, Cambridge, UK

³ Institute of Earth Sciences Jaume Almera, Spanish National Research Council, Carrer de Lluís Solé i Sabarís, 08028 Barcelona, Spain

⁴ CaSEs, Department of Humanities, Universitat Pompeu Fabra, c/Trias Fargas 25-27, 08005, Barcelona, Spain

⁵ ICREA, Passeig Lluís Companys 23 08010 Barcelona, Spain

⁶ School of Geography, Archaeology and Environmental Studies, The University of the Witwatersrand, Johannesburg, South Africa

⁷ Department of Archaeology, University of Cambridge, Downing Street, CB2 3DZ, Cambridge, UK

ORCID: H.A. Orengo [0000-0002-9385-2370], F.C. Conesa [0000-0002-4026-7266], A. Garcia-Molsosa [0000-0001-5416-2986], A. Lobo [0000-0002-6689-2908], A.S. Green [0000-0002-3324-5165], M. Madella [0000-0002-9324-1545], and C.A. Petrie [0000-0002-2926-7230]

*Corresponding author: horengo@icac.cat, phone +34 977249133

*Co-corresponding author: fconesa@icac.cat, phone +34 977249133

Abstract

This paper presents an innovative multi-sensor, multi-temporal machine-learning approach using remote sensing Big Data for the detection of archaeological mounds in Cholistan (Pakistan). The Cholistan Desert presents one of the largest concentrations of Indus Civilisation sites (from c. 3300 to 1500 BC). Cholistan has figured prominently in theories about changes in water availability, the rise and decline of the Indus Civilisation and the transformation of fertile monsoonal alluvial plains into an extremely arid margin. This paper implements the first archaeological application of a multi-sensor, multi-temporal machine learning approach for the remote detection of archaeological mounds. A classifier algorithm that employs a large-scale collection of Synthetic Aperture Radar and multispectral images has been implemented in Google Earth Engine, resulting in an accurate probability map for mound-like signatures across an area that covers c. 36,000 km². The results show that the area presents many more archaeological mounds than previously recorded, extending south and east into the desert, which has major implications for understanding the archaeological significance of the region. The detection of new small (<5 ha) to large mounds (>30 ha) suggests that there were continuous shifts in settlement location. These shifts are likely to reflect responses to a dynamic and changing hydrological network and the influence of the progressive northward advance of the desert in a long-term process that culminated in the abandonment of much of the settled area during the Late Harappan period.

Keywords

Multitemporal and multisensor satellite big data, machine learning, virtual constellations, archaeology, Indus Civilisation, South Asia.

Significance statement

This is the first paper illustrating the potential of machine learning-based classification of multi-sensor, multi-temporal satellite data for the remote detection and mapping of archaeological mounded settlements in arid environments. Our research integrates for the first time multitemporal Synthetic Aperture Radar and multispectral bands to produce a highly accurate probability field of mound signatures. The results largely expand the known concentration of Indus settlements in the Cholistan Desert in Pakistan (c. 3300-1500 BC), with the detection of hundreds of new sites deeper in the desert than previously suspected including several large-sized urban centres. These distribution patterns have major implications regarding the influence of climate change and desertification in the collapse of the largest of the Old-World Bronze Age civilisations.

Introduction

Artificial mounds are a characteristic feature of permanent and semi-permanent settlement locations in past cultural landscapes, particularly on sedimentary plains, but also in arid and semi-arid regions. These mounds can be readily visible due to their prominence and shape, and the fact that they are comprised of accumulated debris such as mud bricks and pottery sherds, which creates specific soil with distinct colour and surface texture. These characteristics make them detectable using different methods, and their number and distribution have seen them play an important role in addressing questions about the formation of early urbanism, states and economic systems.

The use of Remote Sensing (RS) to detect and map archaeological mounds has been attempted in many parts of the world (1–4). Much research has focused on arid and semi-arid areas in the Levant and the Near East, where the geomorphological and sedimentary properties of mounds make them highly visible in digital elevation models and aerial and satellite imagery (5). Mounds can also leave specific multispectral soil signatures in highly anthropised landscapes with levelled or irrigated fields (6). When available, the use of declassified historical photographs such as CORONA imagery has been critical to the detection of mounds (7–9). Georeferenced historical map series have also been used solely or in combination with contemporary declassified data (10–14). In recent years, RS-based archaeological research has gradually incorporated machine learning techniques and algorithms that facilitate the automated detection of sites and features. Most of those applications have focused on the detection of small-scale features using high-resolution datasets such as lidar (15) or WorldView imagery (16–18). In the Near East, Menze and Ur (19) applied a Random Forest classifier over a multi-temporal collection of ASTER imagery to identify probable anthrosols. Some other attempts have used multi-temporal data to monitor archaeological sites and human impact such as urban sprawl and looting (20–22). The detection of anthropic signatures, such as those that characterise mounded sites, across a very large area, remains seldom attempted, presumably due to the large computational resources, programming expertise and the large amount of satellite data required.

This paper presents a novel machine learning approach for the detection of mounded sites across one such very large area. It employs multi-temporal data in a way similar to the approach used by Menze and Ur (19), but rather than detecting anthrosols, which include modern towns, settlements and other areas of human use or occupation, this study outlines a method for restricting the algorithm detection to archaeological mounds. It does this by employing a new multi-sensor and multi-temporal approach that combines Synthetic Aperture Radar data and multispectral satellite imagery. The study area is the Cholistan Desert in Pakistan, which is a large arid area that has long been considered a core region of South Asia's Indus Civilisation (c. 3300–1500 BC). The results of the analysis are evaluated, compared to previous field survey data, and discussed in relation to the prevailing interpretations of the trajectories of settlement in the region, including the development and decline of the large urban centres of the Indus Civilisation.

Research background

The Cholistan Desert and the Indus Civilisation

The Cholistan Desert is the western extension of the Great Indian or Thar Desert, and stretches from the southern edge of the alluvial plains of Punjab to the north of Sindh province in Pakistan (Figure 1a–b). The area is usually described as a marginal arid region that is highly sensitive to the annual variation of the Indian Summer Monsoon (23), the intensity of which has significantly affected its population and ecological diversity throughout the Holocene (24). Today its landscape is characterised by fossilised sand dunes with shrub vegetation, small patches of trees around artificial water tanks, called *tobas*, and saline mud flats called *dahars* (25, Figure 1c). The region has played a significant historical role in transcontinental mobility between Central

and South Asia, as attested by the caravan routes that crossed the area and the numerous forts that protected them (26). The area has also been home to nomadic pastoralists who have moved with their livestock near *tobas* (23, 27). In recent decades, however, major investment in irrigation schemes in the western plains has changed the traditional subsistence strategies and movement patterns of local populations (28–30). Cholistan has figured prominently in discussions about the Indus Civilisation, and Possehl (31, 32) argued that it was the most important area of settlement concentration during the Mature Harappan period (also Harappa phase; c.2500-1900 BC), which is when South Asia's first cities flourished. There has been considerable discussion about the region's hydrological network and its relation to the former Ghaggar-Hakra River, the date at which it ceased to flow perennially is much debated (33–38). Despite its perceived archaeological importance, at present only survey data is available for the region.

The first archaeological explorations in Cholistan can be traced back to the early European officials operating in the area in the 1830-40s, when sporadic mounded sites were reported on traveller's notes (39, 40). Oldham (41) was among the first explorers to report scattered mounded sites in the desert in the late 1890s, and these observations were reiterated in the 1940s and 50s by Stein (42) and Field (43). The most extensive work in the region was led by Mughal, who executed an extensive survey between 1974 and 1977 (26). Additional areas of Cholistan were surveyed in the 1980s as part of an attempt to systematically survey the whole of Punjab (44). To date, Mughal's publications constitute the largest and most detailed archaeological legacy data for this area, and the reported sites have been integrated into the corpus of Indus Civilisation sites (45, 46). Through field walking and random surface collection, Mughal's team noted 414 locations associated to different chronological periods ranging from the early Indus phases to the early Islamic. Petrie and Lynam (47) reviewed Mughal's legacy data and incorporated sites reported by Stein in the 1940s, coming to a total of 462 archaeological site locations in the Cholistan region.

Some of the largest mounds discovered during the Mughal-led surveys have been the focus of recurrent visits by different teams (48, 49), particularly Ganweriwala, which traditionally has been considered one of the five major Indus cities together with Harappa, Mohenjo Daro, Rakhigarhi and Dholavira (46, Figure 1a, d), though its size and significance has been the focus of some discussion (48–51). Since Mughal's surveys no major large-scale field surveys have been conducted in the region, despite the substantial and continued interest that these sites have aroused (52–57).

Mughal's surveys: site distribution and settlement patterns

Mughal found evidence for settlement distributions that varied across what he characterised as the Hakra, Early Harappan, Mature Harappan, and Late Harappan periods (26). The earliest, Hakra period, sites appear to be clustered to the south of Cholistan. There was a reduction of settlement in the south and displacement to the north in the Early Harappan period, a reduction of settlement in the north and more extensive settlement in the south in the Mature Harappan period, and abandonment of the south and a return to the north in the Later Harappan period. This alternating pattern is illustrated in **Figure 2**. Mughal argued that these shifts in settlement concentrations were related to the movement of river channels and water availability (26:23-26, 57: Figs 3-6, 58). While the Hakra, Early Harappan and Mature Harappan period sites were distributed over the whole of the area, the Late Harappan period sites appear to have been restricted to more northern locations. This displacement of settlement was not reversed in the subsequent Painted Grey Ware period, suggesting that during or at least after the Mature Harappan period, the population made a major shift north.

Old limitations, novel approaches

The Stein, Field and Mughal data are inevitably constrained by the technological limitations of their time. The toponymy, location and extension of mounded sites and scattered materials were recorded using manual and non-systematic approaches, which makes it difficult to identify existing mound site locations and to quantify

their extension, exact distribution and pattern (59, 60). A second problem is the lack of quantitative parameters that can help evaluating their significance. These data were usually published in the form of list or gazetteers (e.g. 26, 46), lacking a quantified description of surface materials. Moreover, Mughal's data (26: Table 11) include both mounded locations and other types of short-term occupation, such as industrial sites or campsites, and therefore many sites do not correspond to long-term mounded settlements. In addition, reported coordinates are often inaccurate and duplicated.

The lack of more systematic field surveys in recent decades is largely due to the remote location of most of these mound sites, the harsh desert conditions of the area and its proximity to the India-Pakistan border. These limitations, along with the unique physiographic characteristics and archaeological significance of the region, make the Cholistan Desert an ideal scenario for testing new RS approaches.

Materials and methods

RS-based archaeological research in marginal or remote areas has often been limited by poor satellite coverage and limited temporal and spatial resolution. These limitations are changing thanks to: 1) the availability of time series of global, medium-resolution satellite imagery from Earth Observation missions, such as the Landsat and Copernicus programmes; and 2) the implementation of multi-petabyte image catalogues and geospatial datasets in cloud computing environments that allow for planetary-scale analysis. For the analysis presented in this paper, the Copernicus Sentinel open access satellite series, in particular Sentinel 1 (both ascending and descending sensors) and 2, have been employed as they offer a higher resolution and greater number of bands than available in other non-commercial satellite imagery. The original spatial resolution of the sensor bands employed, 10 and 20 m/px, is adequate for the detection of mound signatures in the area. These present a minimum diameter of around 100 m and, therefore, could incorporate enough number of pixels in sentinel imagery to be clearly identified. The Sentinel collections have been accessed and analysed using Google Earth Engine (GEE) in order to investigate and automatically identify surface sediments potentially related to archaeological mounds. The methodological workflow is illustrated in **Figure 3**. This novel combination of purposely built multi-source multitemporal data and methods based on Big Data analysis has allowed the examination of a very large-scale study area of c. 36,000 km².

Google Earth Engine (GEE) cloud computing geospatial platform

GEE is a fast-growing web-based geospatial platform seeing application within several academic disciplines, and in recent years it has boosted the emergence of RS-based automated applications at the continental and planetary scale of analysis (61). The archaeological application of GEE has been greatly extended only recently (12, 62–67). GEE is particularly suitable for implementing large-scale multi-temporal data analysis as it provides access to a 20-petabyte catalogue of satellite imagery and geospatial datasets, which includes the Sentinel series and most other publicly accessible satellite data acquired since the 1970s. GEE parallelises and executes code developed in JavaScript or Python using Google's cloud computing infrastructure, permitting work with intensive computational processes at unprecedented scales. GEE is also very useful for developing machine learning processes as it allows the computation of partial machine learning-based classifications within a few seconds (or minutes if large training sets and many bands are employed) using screen map area and resolution. This is an important advantage to traditional machine learning processes as it allows the evaluation of the results of new training data without having to compute a full-resolution classification of an entire area, and reduces the number of necessary iterations to achieve satisfactory results. GEE also incorporates high-resolution imagery (equal to that of Google Earth) that allows the evaluation of the results of the classification and the selection of new training data without needing to export them to GIS software. Furthermore, GEE provides vector drawing tools that simplify training data selection.

Multi-sensor Sentinel series

Sentinel 1 is a dual polarization C-band SAR with several scanning modes. The analysis presented here selected interferometric wide (IW) swath mode, which is the mainland operational mode with a ground resolution of 10 m/px and produces data in single (HH or VV) or double (HH + HV or VV + VH) polarization in both ascending and descending modes. Each scene available at GEE had been pre-processed using Sentinel-1 Toolbox to: 1) remove low intensity noise and invalid data on scene edges; 2) remove thermal noise; 3) radiometric calibration; and 4) terrain correction using SRTM 30 (spatial resolution of ~30 m, 1 arcsecond at equator, absolute horizontal accuracy ≤ 20 m, absolute vertical accuracy ≤ 16 m, and relative vertical accuracy ≤ 10 m). Sentinel 1 provides data starting from October 2014.

Sentinel 2 multispectral imagery incorporates 13 bands from which only the visible/near infrared bands (VNIR B2-B8A) and the short-wave infrared bands (SWIR B11-B12) were employed. Bands B1, B9 and B10 (60 m/px each) correspond to aerosols, water vapour and cirrus respectively and they are not employed in this study except for the use of the cirrus-derived cloud mask applied. Visible (B2-B4) and NIR (B8) bands provide a ground resolution of 10 m/px, while red-edge (B5-B7 and B8A) and SWIR (B11-B12) bands present a 20 m/px spatial resolution. Specifically, for this research Sentinel 2 Level 1C products representing Top of Atmosphere (TOA) reflectance were preferred due to the larger span of its mission (starting from June 2015) and excellent resolution.

Machine learning-based approaches to the detection of archaeological sites have previously employed a single type of multispectral imagery source. For the first time, this research has combined large multi-temporal series of multispectral (Sentinel 2) and SAR (Sentinel 1) satellite data for the detection of archaeological mounds. Its use responds to SAR capacity to reflect soil roughness, texture and dielectric property (68), and other ground physical conditions such as compactness. The characteristic compact soil of archaeological mounds that have been formed by the decay of clay-based mud-bricks was assumed to provide a stark contrast with surrounding desert soil (see 68–70). Another advantage of the use of SAR is that it has a certain amount of soil penetration in very dry sandy loose soils, which makes it particularly adequate for this specific area. Initial tests confirmed that known mound sites provided a characteristic SAR signature that differentiates them from the surrounding terrain (Figure 4a-d).

A notable drawback of single SAR images, which has been reduced here with the use of multitemporal series, is the presence of noise (speckle), an artefact of microwave scattering. Furthermore, SAR alone is not able to provide enough information to isolate archaeological mounds from other types of similar clayish soils (such as modern desert seasonal settlements and dried up water *tobas*) that produce analogous responses. Equally, optical multispectral imagery is not single-handedly able to isolate mound spectral signatures in some other areas such those presenting natural accumulation of clays that produces similar spectral responses (Figure 4e-f). The different nature of these sensors was an important factor for ensuring that those elements that would produce similar values to those of mounds in one source were discriminated in the other. Giving the complementarity of SAR and multispectral imaging, their combined use was conceived as a way of providing discriminant values for mounded archaeological sites in the area.

Multi-temporal aggregates

The use of multiple images makes it possible to consider short- and long-term environmental variability and different visibility conditions, thereby reducing the impact that incidental circumstances such as the presence of clouds have when using a single image. The only previous instance of the incorporation of multi-temporal images for the machine learning-based detection of archaeological mounds has been by Menze and Ur (19). They used multiple ASTER satellite images of the same area acquired at different moments over a period of several years. Here, we have superseded that approach by employing a multi-temporal fusion that averages

1500 SAR images taken from 2014-20 in the case of Sentinel 1, and 3112 multispectral images acquired from 2015 to 2020 in the case of Sentinel 2.

An algorithm was developed (see **SI Appendix**) to combine all available Sentinel 1 images for the study area and create a composite image integrating polarisation and look angles to increase the amount of information available about the objects of interest. Median values were employed to integrate the different images available to ensure that we obtained a stable image and that radar speckle was eliminated. Medians were preferred to mean values to minimize the effect of eventual outliers.

While Sentinel 1 SAR is unaffected by clouds, Sentinel 2 TOA values may be affected by cloud cover. The S2 TOA collection reports coded information on quality concerns for each pixel in the QA60 bitmask. We integrated all multi-temporal images of the Sentinel 2 collection into a single multispectral image by averaging the pixel values per each band but ignoring those observations flagged as cloudy (opaque or cirrus) in the QA60 bitmask.

Sentinel 1 and 2 data aggregates made it possible to produce a 14-band multi-temporal and multi-sensor image composite. The image integrates 4 SAR bands (a single VV and a dual HH-HV band in both ascending and descending modes), and 10 optical multispectral bands (B2-B8A, B11 and B12). The high-quality optical and radar bands are not affected by particular environmental or visibility conditions and therefore reflect average reflectance values for the study area while significantly reducing the computational costs of the process. While incorporating seasonal information might have produced improved results (see for example 64, 72), we preferred to test the most straightforward approach of using aggregate averages, thus keeping computational cost relatively low. This is a particularly important point as the algorithm employs two sensors and a relatively high spatial resolution given the very large size of the study area.

The creation of multi-temporal and multi-sensor (active and passive) aggregates for the detection of archaeological sites constitutes an important new development. Previous research has emphasized the current need for ‘the harmonization and synergistic use of different sensors [...] to maximize the impact of earth observation sensors and enhance their benefit to the scientific community’ (73). In this regard, this research constitutes one of the first large-scale applications of the concept of virtual constellations, a ‘set of space and ground segment capabilities that operate in a coordinated manner to meet a combined and common set of Earth Observation requirements’, put forward by the Committee on Earth Observation Satellites (CEOS) (74, see also 75) and falls in line with current agendas for the advancement of archaeological remote sensing (76, 77). Virtual constellations aim to combine sensors with similar attributes to increase the efficiency of remote sensing processes. Here, we have gone beyond the initial definition to include sensors with very different principles (active and passive), but with the combined potential to produce superior results given the nature of our object of interest. The use of a machine learning algorithm provided a practical way to employ the multiplicity and disparity of data present in the image composite bands.

Machine learning algorithm

The steps for classification of mound-like signatures included gathering training data, training the classifier model, classifying the image composite and then validate the classifier with an independent validation set. We employ a selection of 25 mound-sites identified in Mughal’s survey on desert areas (26) as our training (n=5) and validation (n=20) sets. Despite the quality problems in Mughal’s data, we selected those sites that could be clearly identified and accurately located high-resolution imagery available in GEE. These corresponded to large and well-preserved sites. Polygons were drawn in GEE to define the areas of the selected mounds from which the values of the pixels in the image composite could be extracted for the training of the algorithm. The definition of spectral signatures and the evaluation of training data is described in detail within the **SI Appendix (Figures S1 and S2)**.

A Random Forest (RF) classifier was selected for the GEE machine learning implementation. A RF classifier builds a number of decision trees on bootstrapped training samples, but each time it considers a split in a tree.

For each split, a new random subset of predictors are considered when splitting nodes (78). The average of the resulting trees helps to avoid overfitting, and hence is less variable and more reliable than other decision tree-based classifiers (17, 79). Moreover, RF classifiers can handle a small number of training samples and it is possible to get the number of votes (i.e. the probability density) for each class. These are two advantages that are particularly useful for RS-based archaeology with limited land use/cover information.

In our GEE algorithm, the RF was composed of 128 trees, considered an adequate number to obtain optimal results without increasing computational cost unnecessarily (80). The RF was set in probability mode so that the results could be evaluated, filtered and refined to improve the algorithm's detection capabilities. The machine learning process underwent three iterations. The original iteration of the algorithm produced satisfactory results in that it was possible to clearly identify the 20 well-known mounds used as test data, and many more potential mounds through their higher RF probability values. Nonetheless, two more iterations were necessary to tune pixels with higher percentages that did not correspond to mound-sites, thus ensuring a good balance between low presence of non-mound pixels with high RF probability values and a high mound detection rate. The output of the RF classifier is a probability field in raster format, in which each value records the probability of a given pixel being a "mound".

In order to produce a map of archaeological mounds, a >0.55 RF probability threshold for mound values was selected after close inspection of the training data on the high-resolution imagery, which produced a raster map of clusters ("mounds") on a background of "no mound". A higher threshold resulted in the better delineation of big and clear mounds, but many small clusters of pixels corresponding to partially covered or small mounds were lost. We considered the >0.55 threshold a good compromise between a high mound detection capacity and a minimal inclusion of false positive pixels (mainly scattered isolated pixels). The algorithm's validation and quality assessment methods are outlined in the **SI Appendix (Figures S3 and S4)**.

Integration of complementary data, area estimates and GIS database

The resulting clusters of high RF probability pixels representing mounds were vectorised to reconstruct the areas of the mounds currently covered by sand dunes or desert shrubs. Photo-interpretation used high-resolution satellite imagery provided by several map services (including Google Earth and Bing Maps) and a limited collection of available WorldView 2-3 imagery. The combination of these sources provided enough temporal and environmental variability to evaluate and delineate the possible extent of the mounds identified by the algorithm. A final mound geodatabase was prepared in a GIS environment and compared with previous coordinates from legacy data (see 26, 44).

Data Availability Statement

All satellite data used in this study are freely available under the open data policy adopted by the Copernicus programme of the European Space Agency. The code developed for this research, which has been used to produce the results discussed in the paper is provided within the **SI Appendix**. This code has been written for Google Earth Engine (GEE)'s implementation of JavaScript. GEE provides free access to the satellite data and the processing necessary to conduct the analysis upon registration. The code is ready for the direct execution of the analysis discussed in this paper (including the gathering and treatment of Sentinel data) and it only requires to be pasted in GEE Code Editor and pressing the 'run' button. It includes instructions on how to use and modify it to be applied to other research needs or areas. The code is also available in an online repository https://github.com/horengo/Orengo_et_al_2020_PNAS where future updates will be implemented.

Results

RF probability field

Thresholding the RF probability field at >0.55 resulted in a map of 337 clusters that we propose as archaeological mound soil surfaces (**Figure 5a-e**). This set includes the 25 mounds selected from Mughal's surveys (26) used as training ($n: 5$) and validation ($n: 20$) data, which were all successfully identified by the algorithm. The newly proposed mounds are similar to the previously known mounds used as training data (**Figure 6a**). The capacity of the algorithm to detect mound-like signatures is probably related to: a) the high contrast in Sentinel 1 single and dual polarisation bands (81); b) the ability of the SAR C-band to penetrate loose dry sand (82-84); and c) the specific reflectance in the Sentinel 2 red edge, NIR and SWIR bands of anthropic sediments in mounded-sites (85, 86).

Due to vegetation and sand cover the RF probability only produced a few well-defined rounded shapes. Most of the newly proposed mounds present fragmentary rounded shapes, elongated strips or shapeless groupings of pixels (**Figure 6b**). It is entirely possible that beside these 337 detected mounds there are also remains of archaeological mounds partially covered by sand or shrub vegetation, with low RF probability values. It is worth stressing that the algorithm helped to identify small clusters of mound-like pixels even when visible high-resolution images or SAR backscatter alone do not show any significant change in surface land cover.

When possible, the new features were matched with legacy data from previous archaeological sites recorded by Mughal's team (26, 44). Out of a total of 337 clusters of high-probability pixels identified as archaeological mounds by the algorithm, only 71 (including the 25 employed to train and test the algorithm) could be linked with reasonable certainty to sites previously recorded. However, it is possible that some other mounds in the database were indeed recorded by Mughal, but their imprecise coordinates make it difficult to establish any secure spatial correlation.

Discussion

Detection and distribution of mound-like signatures

Despite the intensity of the regional surveys conducted by Mughal's teams in the 1970 and 1980s, the automated detection of mounds in Cholistan has significantly increased the number of mound-like settlements over a much larger area of Cholistan than previously thought. In particular, the distribution extends towards the southern part of the region and the inner Thar desert towards the east, which was an area that was virtually inaccessible for previous studies (**Figure 7**).

As expected, the distribution of the detected mounds by the algorithm was concentrated in the desert areas. Multispectral contrast and SAR texture differences among classes are accentuated in the desert, especially in the *dahar* flat surfaces, whereas the environmental variability in the agricultural lands towards the western edge of the desert is highly affected by several factors (e.g. mechanised agriculture, irrigation canals, roads and railways and irrigation-supported agricultural villages). As a result, the visibility of most existing archaeological mounds in this area has been strongly influenced by present-day taphonomy.

There is also a clear decrease in site density towards the middle of the overall site distribution (**Figure 7**). This can be attributed to an extended stretch of deeper dunes that potentially hides a similar site density to those documented west and north of the dune stretch. In a similar manner, the area where new mounds have been identified is delimited to the south and southeast by the presence of deep dunes. The distribution of documented sites in relation to the presence of dunes (**Figure 7**) strongly suggests that many more sites could be lying below dunes in the deeper desert, and therefore the depositional dynamics of aeolian sediments in the Cholistan desert may have played a crucial role in the settlement history of the region. It is also interesting to

note that the areas with a higher concentration of sites are situated in open mud flats (*dahars*) that are scattered through the region, though many of these mounds are still partially covered by sand dunes. Some *dahars* are used today to extract fine silty sediments or to excavate *tobas* (25, 27). These activities have accumulated silty soils on the *dahar* surface, leaving soil signatures and texture that are very similar to those of mounds in single Sentinel 1 and 2 scenes and high-resolution imagery. In all cases, the RF probability field was able to clearly discriminate between *dahars*, *tobas* and archaeological mounded areas.

Only 71 of the mounds located by Mughal match the new photo-interpreted RF probability mounds, and usually these are among the most well-preserved mounds in the landscape. A large number of nomadic sites and pottery distributions not related to long-term settlement were also reported by Mughal, but they were only distinguished on site by surface scatters of artefacts (27), and the resolution of our sources is not enough to locate them. An important aspect of the RF probability approach is that the new locations represent areas with a specific mound-like sediment signature, which suggests a relatively stable occupation, potentially for some time, and the continued use of construction materials such as mud brick. These characteristics are closely associated to sedentary dwellings in similar Indus contexts (51).

For each new location, the extension, visibility and preservation of surface archaeological sediments can be now further explored in relation to its immediate surroundings. Previous studies tended to separate a unique multi-phase location into several distinct locations based on the distribution of surface material culture, which are largely biased by the terrain view that was used (no aerial images were available), and the partial burying and occlusion of sites. For example, the site of Bokhariyanwala was described as having occupation during the Mature and Late Harappan periods on two adjacent mounds. Our analysis shows that the ridge of a fossilised dune crosses the area, suggesting that these sites formed a single large mound (Figure 8a). Similarly, the site of Changalawa was reported to have two distinct Mature Harappan period mounds and a third Late Harappan period mound. However, the RF-probability field only returns a single mound in this area. In this case, the distribution of materials may have been disturbed by the presence of a 1920s irrigation canal crossing the site, and different parts of the one site may have been occupied in each period (Figure 8b). However, until archaeological excavation can be carried out at these sites, the possibility that some of these sites may be comprised of discontinuous occupations on multiple overlapping mounds cannot be dismissed. These examples reflect the high heterogeneity in terms of site visibility and preservation across the study area and provide a cautionary tale on using the previous data to study site density, distribution and size.

Villages and towns: mound size estimates

An important result of the automated site detection has been the delineation of the estimated area size for the 337 mounds located in the area (Figure 9a-b). Mound area estimates in the Near East (87) have raised the question of whether these represent the total habitation size of a settlement, or wall debris and open-air spaces outside the settlement. For the results presented here, area estimates should be used with caution and be considered only as a tentative measure of the total mound area estimated through visual inspection of the RF probability field and high-resolution satellite imagery. A large proportion of the newly detected mounds are less than 5 ha in size ($n = 246$, 72.99%). Estimated site areas suggest a general pattern of small rural Indus settlements distributed between medium and large-size sites that are possibly urban in nature (26, 88), which is similar to other Indus core areas in north-western India such as Gujarat (89) and Haryana (90). However, the area for most of the known sites that we have identified is slightly different from those previously reported (26: table 13). For example, the Hakra period site of Lathwala was reported to be 26.3 ha, but the closest visible mound is around 5 ha, although it should be noted that the site is now clearly divided by a large sand ridge. The largest Early Harappan period site, Gamanwala, was reported as being 27.3 ha, but the closest mound to this location is a much smaller mound of c. 6.5 ha. Kudwala, the largest Late Harappan period mound at 38.1 ha could not be detected remotely, suggesting either that its original location was not correct or that it has since been obliterated by the expansion of agricultural land. Significantly, the RF probability field detected six

mounds that can undoubtedly be considered as large Indus settlements, potentially towns of more than 20 ha. The distribution of these mounds is highlighted in **Figure 9a-b**. The Mature Harappan period settlement of Ganweriwala, initially estimated to be 80 ha in size (26), has long been considered one of the major centres of the Indus Civilisation. However, recent reassessment has shown that it was much smaller, between 20 and 40 ha (49-51). Our data suggests that the mounded area is c. 33 ha, comparable to a large town (90), though it may well have functioned as an urban centre (51). It is possible, however, that Ganweriwala was not the only large town dating to the Mature Harappan period in the region. Two more sites identified by Mughal as dating to this period, Sanukewala and Rajbai, have a similar visible mound extension (c. 32 ha) and they could have been even larger in the past. The northwest section of Sanukewala is largely affected by an abandoned canal and Rajbai is partially covered by sand dunes. The Late Harappan period site of Siddhuwala, also have significant proportions (26.5 ha). The RF-probability field has also brought to light two previously unknown large sites, named Khundowala (29 ha) and Mulhiawala (20.5 ha) after the nearest toponym shown on historical maps. Mulhiawala is in the south-eastern margin of the desert, an area that was virtually empty of Indus sites before the automated detection of mounds. The site of Khundowala is unique in lying in the north-east margin of the desert, and it is completely isolated from other visible mounds. The site is partially covered by a fossilised sand dune, and indeed the RF probability only showed a small portion of the total mound that can be appreciated today in high-resolution satellite imagery. The identification of this site is highly relevant, as it documents the presence of a large mound in the northern parts of the historical Hakra River basin. Despite previous evidence of the presence of more Indus Civilisation sites in this area (26, 44), only a few small mounds have been detected by the algorithm there. The implications of the northern expansion of the desert, as evidenced by the large sand dunes that partially cover these sites, is further discussed below.

Long-term landscape dynamics

The new distribution of mounds in the study area indicates that the Thar Desert has expanded considerably since the Indus period. No current occupation except for seasonal pastoral camps is known in the southern and western sectors of the site distribution area. Medieval and late Medieval period settlements such as the Cholistan forts known in the area form an arc delimiting the north-western limit of the Thar desert at a more northerly and westerly location (26). These patterns might suggest that the expansion of the desert is a long-term process that has been progressing over several millennia, contributing to the abandonment of settlements at the desert margins.

At present, the lack of palaeoenvironmental evidence such as sedimentary records in the core area of the Cholistan Desert can only be partially complemented by data from the adjacent Thar Desert, especially in west and north Rajasthan. Durcan et al. (91) have suggested that the Holocene geomorphological dynamics were highly dynamic and the distinct phases of fluvial and aeolian deposition were not spatially or temporally instantaneous, but rather a synchronic alternation. Some studies have, however, highlighted a drier climatic condition on the northern margins of the Thar Desert sometime after $\sim 4.4 \pm 0.1$ ka BP (24, 59, 92–94), a phase that is consistent in other north-western regions such as mainland Gujarat (95, 96). Nonetheless, the cyclic fashion of the Quaternary aeolian activities in the Thar Desert started much earlier (c. 150 ka BP), and Singhvi and Kar (97) advocate for a phase of major aeolian activities during the Holocene Climatic Optimum (c. 5-3.5 ka BP), with century-scale rates of spatial shifts in dune forming in west Rajasthan up to 2 ka BP. Existing sedimentary records for the western margins of the Thar Desert, therefore, correlate well with a potential enhanced aeolian activity c. 4 ka BP, followed by declining rainfall, which, in turn, would have stabilised the now subdued sand dunes. Another burst of aeolian activity started c. 2 ka while the latest aeolian active phase was in historical times, when the rates of dune mobility in the desert increased mainly due to human pressure (98, 99).

The new data on mound distribution can also be compared to the documentation of multiple palaeoriver channels and seasonal streams in the study area. Just as with the mounds, these are visible only in *dahar*

surfaces and therefore it is difficult to offer a continuous picture of their shape and length. Overall, the hydrological network in the area has been related to the course of the former Ghaggar-Hakra River. Some authors have argued that the Ghaggar-Hakra system ceased to flow through this region before the Holocene (59, 100–103), thus suggesting that the alluvial plains of north-western India and eastern Pakistan were characterised by a continuum of fluvial environments interrelated through a seasonal precipitation gradient and local aeolian dynamics (36, 104). The river basin is well-documented in historical nineteenth century maps and narratives (Figure 9), showing recurring seasonal flooding during episodes of extreme rainfall such as in 1804, 1805 and 1871 (105). However, previous RS-based attempts to identify the former channel network across Cholistan have suggested several potential interconnected waterflows (106–109), but little is known about their chronology. Moreover, the detected relict network of fluvial environments across the region has several characteristics that differ from a major Ghaggar-Hakra course, specifically: 1) these palaeochannels form multiple courses with roughly similar orientation; 2) their morphology is relatively straight, at least by comparison with those documented to the north of the study area, where they tend to be very sinuous given the low slope of the alluvial plain (see 64, 65, 110); and 3) their orientation is coincident with that of the fossilized dunes and not with that of the alluvial plains just north of the desert edge. In this regard, it is important to note that while rivers flow following the aspect of the terrain, dunes are mostly influenced by wind direction in plain areas. Therefore, the flow of many of the rivers detected must have been influenced by the presence of earlier relict sand dunes. The many sub-parallel, ephemeral river traces in the area might be related to a process of river migration towards the north-eastern region as dunes expanded from the south. Consequently, it is possible to propose that water was still available, at least seasonally, in streams and flooded fertile *dahars* until recent historical times, as attested by the nomadic migration traditions in the area (27, 111). The monsoonal-fed hydrological network changed dramatically from 1897 onwards with the construction of the Ottu Barrage on the northern Indian course of the Ghaggar River, which increased desertification in the southern desert edges and led to the development of large irrigation schemes through the desert lands early in the 1930s (112).

Indus settlement trends

Although at the moment it is not possible to provide chronological information for the newly detected mounds, they can be combined with those previously identified and dated by Mughal (26) to attempt to understand changes in settlement distribution over time. The surveys of Cholistan documented all of the mounds that were encountered, including those of the Medieval period. Considering that the vast majority of the previously reported sites were settlements attributed to the periods of the Indus Civilisation, there is a high chance that most of the newly detected ones were also occupied during this same period. This assumption can be contrasted with similar Indus contexts. For example, in the plains of Haryana in north-west India, Green et al. (14) have highlighted that mound features visible on historical maps tend to be protohistoric or Bronze Age settlements when surveyed or validated on the ground, whereas Early Historic and Medieval settlements appear to be more frequently associated with modern settlement locations. In Cholistan, specifically, several mounds are partially covered by fossilised sand dunes and ridges indicating that these mounds largely pre-date the north-western expansion of the Thar Desert, and that mounds located deeper in the desert may therefore be particularly early in date.

Mughal's analysis highlighted the importance of the Cholistan settlement data for understanding the development of Indus urbanism, and Madella (113) has suggested that Cholistan was potentially a zone of intensive and extensive cultivation. Taken at face value, Mughal's settlement distribution data for Cholistan suggests that this was an intensively occupied area. However, Petrie and Lynam's (47) reanalysis of the data suggests that this settlement system may have been marked by displacement and considerable instability, which indicates that it was more unusual than is typically assumed. This interpretation has ramifications for the interpretation of agricultural practices and the sustainability of the region, indicating a high degree of flexibility and mobility. This insight is important, because there is currently no direct archaeobotanical evidence available

from sites in this region. This agricultural flexibility must have been essential to Cholistan, which was strategically located in a central point of the area occupied by Indus populations (51, 114), and was potentially a necessary transit area for movement between different regions. Cholistan may well have formed a communication node between surrounding Indus areas. The apparent loss of the more southern areas of Cholistan for settlement may have been an important factor in the breakdown of Indus interaction networks and the increase in more local scale interactions in the Late Harappan period (115). The decline of settlement in Cholistan might thus have created a “dead zone” of interaction, increasing the cost of communication and exchange beyond the point that de-urbanizing cities could maintain.

Conclusion

The new dataset provides the first collection of Sentinel 1 and Sentinel 2 spectral signatures for mound-like archaeological features in drylands, and the resulting new mound locations can be now addressed in terms of RF probability values. We present the first combination of multi-temporal, multi-polarisation and multi-angle SAR bands and multi-temporal optical bands (including visible, red-edge, NIR and SWIR) analysed using a machine learning algorithm in a cloud computing platform for the detection and analysis of archaeological mounds, which has the potential to transform archaeological site detection. The machine learning algorithm that has been employed was able to detect all previously known mounds in the study area for which we could gather accurate locations and large numbers of new ones well beyond the expectations laid out by previous research. The method provides results that are noticeably superior to the use of single source remote sensing approaches. RS-based applications in arid and semi-arid areas elsewhere can benefit from the integration of globally-available Sentinel data in GEE’s accessible, flexible and reproducible environment to perform and evaluate machine-learning workflows.

The new distribution of archaeological sites in the Cholistan Desert, in combination with legacy archaeological data, suggest that most of these mounds are proto-historic settlements, that they extended across a larger area than previously recognised, and that they include several previously unknown mounds that, considering their large size, can be classified as urban on their own right. Archaeological data in combination with landscape analysis, which includes the mapping of factors affecting site visibility, suggest that these were only a relatively small part of the mounds present in the area, many of which might lie below large dunes in the core study area and deeper into the desert.

The archaeological sites in Cholistan were occupied at points along a span of five millennia. The significant number of Indus Civilisation Mature Harappan period and urban-scale mounds that were documented and the shift in the concentrations of settlement to the north suggests that it is important to consider the impact of the advancement of the desert on settlement displacement of Indus populations. Given the centrality of this area, the displacement of occupation in Cholistan potentially played an important role in the generalised processes of de-urbanisation, decrease in settlement size and regionalisation that characterise the Late Harappan period across the Indus region.

While undoubtedly Cholistan was a significant zone of settlement for the populations of the Indus Civilisation, the nature of the settlement dynamics in this region and their relationship to water availability are in need of both ongoing re-evaluation and ground-truthing in the field. Indus populations had clearly adapted their behaviour to survive in this apparently unstable environment, and it appears to have remained an important area for settlement and a component in interactive networks for an extended period. Future research should perhaps also consider the role of pastoralism and the pastoral economy in this region, and its possible links to population mobility. The results of the study presented here provide critical new resources for the processes of re-thinking the dynamics of settlement distribution and the archaeological significance of the region.

Acknowledgements

This research arises from a collaboration between the ERC-CoG TwoRains project (HAO, CP, AG) and the MSCA-IF Actions MarginScapes (FC) and WaMStrIn (AGM). TwoRains is funded by the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement No. 648609) and the DST/UKIERI, the British Academy and the McDonald Institute for Archaeological Research. MarginScapes and WaMStrIn are funded under the European Commission Marie Skłodowska-Curie grant agreements No. 794711 and 746446, respectively. HAO is a Ramón y Cajal Fellow (Spanish Ministry of Science, Innovation and Universities), and ASG is a member of the GCRF TIGR2ESS project (BB/P027970/1). The authors are indebted to Gethin Rees (GIS Research Curator at the British Library) and the research staff at a) the British Library and b) the Map Room and Imaging Services at the University Library, University of Cambridge, for providing access to the Survey of India map collections. WorldView imagery used in this research was generously granted to FC by the DigitalGlobe Foundation. We would like to thank the anonymous reviewers for their helpful comments.

Supporting information

The Supporting Information (SI Appendix) of this article includes:

Data validation: detailed information and figures to support the quality assessments for the machine learning algorithm and the validation and treatment of the results.

JavaScript algorithm for GEE described and employed in the paper.

Author contributions

HAO, FC and CP conceived the study; HAO designed, wrote and implemented the GEE algorithm; AL performed data validation in R, with the support of FC; HAO and FC generated the mound database and performed spatial analyses; CP supervised the research; HAO, FC and CP wrote the paper with contributions from all co-authors.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. A. Beck, G. Philip, M. Abdulkarim, D. Donoghue, Evaluation of Corona and Ikonos high resolution satellite imagery for archaeological prospection in western Syria. *Antiquity* **81**, 161–175 (2007).
2. S. Parcak, Satellite Remote Sensing Methods for Monitoring Archaeological Tells in the Middle East. *J. F. Archaeol.* **32(1)**, 65–81 (2007).
3. A. Agapiou, D. G. Hadjimitsis, D. Alexakis, A. Sarris, Observatory validation of Neolithic tells ("Magoules") in the Thessalian plain, central Greece, using hyperspectral spectroradiometric data. *J. Archaeol. Sci.* **39(5)**, 1499–1512 (2012).
4. H. A. Orengo, A. Krahtopoulou, A. Garcia-Molsosa, K. Palaiochoritis, A. Stamati, Photogrammetric re-discovery of the hidden long-term landscapes of western Thessaly, central Greece. *J. Archaeol. Sci.* **64**, 100–109 (2015).
5. B. H. Menze, J. A. Ur, A. G. Sherratt, Detection of ancient settlement mounds: Archaeological survey based on the SRTM terrain model. *Photogramm. Eng. Remote Sens.* **72(3)**, 321–327 (2006).
6. M. Bini *et al.*, Identification of leveled archeological mounds (Höyük) in the alluvial plain of the Ceyhan River (Southern Turkey) by satellite remote-sensing analyses. *Remote Sens.* **10(2)**, 241 (2018).

7. G. Philip, D. N. M. Donoghue, A. Beck, N. Galiatsatos, CORONA satellite photography: an archaeological application from the Middle East. *Antiquity* **76**, 109–118 (2002).
8. T. J. Wilkinson, *Archaeological landscapes of the Near East* (The university of Arizona Press, Tucson, 2003).
9. J. Casana, J. Cothren, Stereo analysis, DEM extraction and orthorectification of CORONA satellite imagery: archaeological applications from the Near East. *Antiquity* **82**, 732–749 (2008).
10. B. Rondelli, S. Stride, J. J. García-Granero Soviet military maps and archaeological survey in the Samarkand region. *J. Cult. Herit.* **14**(3), 270–276 (2013).
11. F. C. Conesa *et al.*, CORONA photographs in monsoonal semi-arid environments: addressing archaeological surveys and historic landscape dynamics over North Gujarat, India. *Archaeol. Prospect.* **22**, 75–90 (2015).
12. A. Garcia, H. A. Orengo, F. C. Conesa, A. S. Green, C. A. Petrie, Remote Sensing and Historical Morphodynamics of Alluvial Plains. The 1909 Indus Flood and the City of Dera Gazhi Khan (Province of Punjab, Pakistan). *Geosciences* **9**(1), 21 (2019).
13. C. A. Petrie *et al.*, Mapping Archaeology While Mapping an Empire: Using Historical Maps to Reconstruct Ancient Settlement Landscapes in Modern India and Pakistan. *Geosciences* **9**(1), 11 (2019).
14. A. S. Green *et al.*, Re-discovering ancient landscapes: archaeological survey of mound features from historical maps in northwest India and implications for investigating the large-scale distribution of cultural heritage sites in South Asia. *Remote Sens.* **11**(18), 2089 (2019).
15. A. Guyot, L. Hubert-Moy, T. Lorho, Detecting Neolithic burial mounds from LiDAR-derived elevation data using a multi-scale approach and machine learning techniques. *Remote Sens.* **10**(2), 225 (2018).
16. W. P. Megarry, G. Cooney, D. C. Comer, C. E. Priebe, 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 Sens.* **8**(6), 1–18 (2016).
17. S. Biagetti *et al.*, High and medium resolution satellite imagery to evaluate late holocene human-environment interactions in arid lands: A case study from the Central Sahara. *Remote Sens.* **9**(4), 1–21 (2017).
18. O. L. Thabeng, S. Merlo, E. Adam, High-resolution remote sensing and advanced classification techniques for the prospection of archaeological sites' markers: The case of dung deposits in the Shashi-Limpopo Confluence area (southern Africa). *J. Archaeol. Sci.* **102**, 48–60 (2019).
19. B.H. Menze, J. A. Ur, Mapping patterns of long-term settlement in Northern Mesopotamia at a large scale. *Proc. Natl. Acad. Sci. U.S.A.* **109**(14), 778–787 (2012).
20. D. Tapete, F. Cigna, D. N. M. Donoghue, 'Looting marks' in space-borne SAR imagery: Measuring rates of archaeological looting in Apamea (Syria) with TerraSAR-X Staring Spotlight. *Remote Sens. Environ.* **178**, 42–58 (2016).
21. A. Agapiou, V. Lysandrou, D. Hadjimitsis, Optical Remote Sensing Potentials for Looting Detection. *Geosciences* **7**(4), 98 (2017).
22. F. Cigna, D. Tapete, Tracking human-induced landscape disturbance at the Nasca lines UNESCO World Heritage Site in Peru with COSMO-SkyMed InSAR. *Remote Sens.* **10**(4), 572 (2018).
23. F. Ahmad, Agro-Pastoral Systems in Cholistan. *Pakistan Geogr. Rev.* **60**(2), 65–69 (2005).
24. Y. Dixit *et al.*, Intensified summer monsoon and the urbanization of Indus Civilization in northwest

India. *Sci. Rep.* **8**(1), 1–8 (2018).

25. G. Akbar, T. Naseeb, A. Mohammad, Cholistan Desert, Pakistan. *Rangelands* **18**(4), 124–128 (1996).
26. M. R. Mughal, *Ancient Cholistan: Archaeology and architecture* (Ferozsons Ltd., Lahore, 1997).
27. M. R. Mughal, "The Harappan nomads of Cholistan" in *Living Traditions. Studies in the Ethnoarchaeology of South Asia*, B. Allchin, Ed. (Oxford & IBH Publishing, New Delhi, 1994), pp. 53–68.
28. S. Malik, F. Rahman, A. Ali, Sustainability of Subsistence Livelihoods of Agro-Pastoralists in Changing Socioeconomic Environment of Cholistan Desert-Pakistan. *Pakistan J. Commer. Soc. Sci.* **11**(3), 1100–1133 (2017).
29. M. Soharwardi, K. Ali, M. Arshad, Migration of Cholistani People from Desert areas toward Irrigated areas: Causes and Consequence (A Case Study of Cholistan, Pakistan). *Int. J. Soc. Sci. Educ.* **1**(3), 262–271 (2011).
30. S. Ashraf *et al.*, Prospective of dairy farming in Cholistan. *SciInt (Lahore)* **25**(2), 345–346 (2013).
31. G. L. Possehl, The Transformation of the Indus Civilization. *J. World Prehistory* **11**(4), 425–472 (1997).
32. E. Cork, *Rethinking the Indus myths A comparative re-evaluation of the Indus Civilisation as an alternative paradigm in the organisation and structure of early complex societies*. (Archaeopress BAR International Series, Oxford, 2006).
33. B. B. Lal, *The Sarasvati Flows On: The Continuity of Indian Culture* (Aryan Books, New Delhi, 2002).
34. H. S. Saini *et al.*, Reconstruction of buried channel-floodplain systems of the northwestern Haryana Plains and their relation to the "Vedic" Saraswati. *Curr. Sci.* **11**(11), 1634–1643 (2009).
35. M. Danino, *On the Trail of the Lost Saraswati* (Penguin, New Delhi, 2010).
36. L. Giosan *et al.*, Fluvial landscapes of the Harappan civilization. *Proc. Natl. Acad. Sci. U.S.A.* **109**(26), 1–7 (2012).
37. P. D. Clift *et al.*, U–Pb zircon dating evidence for a pleistocene Sarasvati River and capture of the Yamuna River. *Geology* **40**, 211–214 (2012).
38. A. Singh *et al.*, Counter-intuitive influence of Himalayan river morphodynamics on Indus Civilisation urban settlements. *Nat. Commun.* **8**(1), 1617 (2017).
39. A. Conolly, *Journey to the North of India, overland from England, through Russia, Persia, and Affghaunistaun* (Richard Bentley, New Burlington Street, London, 1834).
40. C. Masson, *Narrative of various journeys in Balochistan, Afghanistan, the Panjab, Kalat, during a residence in those countries to which is added* (Richard Bentley, New Burlington Street, London, 1844).
41. C. F. Oldham, The Sarasvati and the lost river of the Indian desert. *J. R. Asiat. Soc.* **34**, 49–76 (1893).
42. A. Stein, Survey of ancient sites along the 'lost' Saraswati river. *A. Geogr. J.* **99**, 137–182 (1942).
43. H. Field, *An Anthropological Reconnaissance in West Pakistan, 1955* (The Peabody Museum, Massachusetts, 1955).
44. M. R. Mughal, F. Iqbal, M. Afzal, M. Hassan, Archaeological sites and monuments in Punjab: Preliminary results of explorations, 1992–96. *Pakistan Archaeology* **29**, 1-474 (1996).
45. J. P. Joshi, M. Bala, J. Ram, "The Indus Civilisation: a reconsideration on the basis of distribution maps" in *Frontiers of the Indus Civilization*, B. Lal, S. P. Gupta, Eds. (New Delhi, 1984), pp. 511–531.
46. G. L. Possehl, Revolution in the Urban Revolution: The Emergence of Indus Urbanization. *Annu. Rev.*

Anthropol. **19**, 261–282 (1999).

47. C. A. Petrie, F. Lynam, Revisiting settlement contemporaneity and exploring stability and instability: case studies from the Indus Civilisation. *J. F. Archaeol.* **45(1)**, 1-15 (2018).
48. S. Gulzar, "Harappan Cities: Landscape and Surface Survey of Ganweriwala" in *South Asian Archaeology and Art*, L. Greaves, A. Hardy, Eds. (Dev Publishers, New Delhi, 2018).
49. F. Masih, "Ganweriwala – A New Perspective" in *Walking with the Unicorn. Social Organization and Material Culture in Ancient South Asia*, D. Frenez, G. Jamison, R. Law, M. Vidale, R. Meadow, Eds. (Archaeopress Publishing Ltd, Oxford, 2018), pp. 377–383.
50. J. M. Kenoyer, South Asia. The Indus Civilization. *Encycl. Archaeol.* 715–733 (2008).
51. C. A. Petrie, "South Asia" in *The Oxford Handbook of Cities in World History*, P. Clark, Ed. (Oxford University Press, Oxford, 2013), pp. 83–104.
52. K. A. R. Kennedy, *Ancient Cities of the Indus Valley Civilization* (Oxford University Press, Islamabad, 2005).
53. D. P. Agrawal, *The Indus civilization: an interdisciplinary perspective* (Aryan Books, New Delhi, 2007).
54. G. L. Possehl, *The Indus Civilization: A Contemporary Perspective* (Altamira Press, Walnut Creek, 2002).
55. D. Chakrabarti, *India an Archaeological History* (Oxford University Press, Oxford, 2012).
56. D.K. Chakrabarti, *The archaeology of ancient Indian cities* (Oxford University Press, Oxford, 2013).
57. R. Wright, C. Hritz, "Satellite Remote Sensing Imagery: New Evidence for Site Distributions in the Upper Indus" in *South Asian Archaeology 2007. Prehistoric Periods*, D. Frenez, M. Tosi, Eds. (BAR International Series, Oxford, 2013), pp. 315–321.
58. M. R. Mughal, "The Protohistoric Settlement Patterns in the Cholistan Desert" in *South Asian Archaeology 1987*, M. Taddei, Ed. (Istituto Italiano per il Medio ed Estremo Oriente, Rome, 1990).
59. C. A. Petrie *et al.*, Adaptation to Variable Environments, Resilience to Climate Change: Investigating *Land, Water and Settlement* in Indus Northwest India. *Curr. Anthropol.* **58(1)**, 1–30 (2017).
60. A. S. Green, C. A. Petrie, Landscapes of Urbanization and De-Urbanization: A Large-Scale Approach to Investigating the Indus Civilization's Settlement Distributions in Northwest India. *J. F. Archaeol.* **43(4)**, 284–299 (2018).
61. N. Gorelick *et al.*, Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **202**, 18–27 (2017).
62. A. Agapiou, Remote sensing heritage in a petabyte-scale: satellite data and heritage Earth Engine© applications. *Int. J. Digit. Earth* **10(1)**, 85–102 (2017).
63. B. Liss, M. D. Howland, T. E. Levy, Testing Google Earth Engine for the automatic identification and vectorization of archaeological features: A case study from Faynan, Jordan. *J. Archaeol. Sci. Reports* **15**, 299–304 (2017).
64. H. A. Orengo, C. A. Petrie, Large-scale, multi-temporal remote sensing of palaeo-river networks: A case study from Northwest India and its implications for the indus civilisation. *Remote Sens.* **9(7)**, 735 (2017).
65. H. A. Orengo, C. A. Petrie, Multi-scale relief model (MSRM): a new algorithm for the visualization of subtle topographic change of variable size in digital elevation models. *Earth Surf. Process. Landforms* **43(6)**, 1361–1369 (2018).
66. L. Rayne, D. Donoghue, A remote sensing approach for mapping the development of ancient water

management in the Near East. *Remote Sens.* **10(12)**, 2042 (2018).

67. H. A. Orenco, A. Garcia-Molsosa, A brave new world for archaeological survey: Automated machine learning-based potsherd detection using high-resolution drone imagery. *J. Archaeol. Sci.* **112**, 105013 (2019).
68. D. C. Comer, R. G. Blom, "Detection and Identification of Archaeological Sites and Features Using Synthetic Aperture Radar (SAR) Data Collected from Airborne Platforms" in *Remote Sensing in Archaeology*, J. Wiseman, F. El-Baz, Eds. (Springer New York, New York, NY, 2007), pp. 103–136.
69. F. Richason, C. Hritz, "The use of digitally enhanced Radarsat SAR imagery in the interpretation of archaeological sites in the Nippur, Iraq, area of the Lower Mesopotamian Plain" in *Proceedings of the International Conference on Remote Sensing in Archaeology from Spacecraft, Aircraft, on Land and in the Deep Sea* (Boston University Press, Boston, 1998).
70. F. Cigna, D. Tapete, R. Lasaponara, N. Masini, Amplitude Change Detection with ENVISAT ASAR to Image the Cultural Landscape of the Nasca Region, Peru. *Archaeol. Prospect.* **147**, 133–147 (2013).
71. T. Balz, G. Caspari, B. Fu, M. Liao, Discernibility of burial mounds in high-resolution X-Band SAR images for archaeological prospections in the altai mountains. *Remote Sens.* **8(10)**, 1–10 (2016).
72. F. C. Conesa, N. Devanthéry, A. L. Balbo, M. Madella, O. Montserrat, Use of satellite SAR for understanding long-term human occupation dynamics in the monsoonal semi-arid plains of North Gujarat, India. *Remote Sens.* **6**, 11420–11443 (2014).
73. A. Agapiou, D. D. Alexakis, D. G. Hadjimitsis, Potential of virtual earth observation constellations in archaeological research. *Sensors* **19(19)**, 4066 (2019).
74. Generic, CEOS—Committee on Earth Observation Satellites. Available at: <http://ceos.org/ourwork/virtual-constellations/> [Accessed March 19, 2020].
75. M. A. Wulder *et al.*, Virtual constellations for global terrestrial monitoring. *Remote Sens. Environ.* **170**, 62–76 (2015).
76. R. Opitz, J. Herrmann, Recent Trends and Long-standing Problems in Archaeological Remote Sensing. *J. Comput. Appl. Archaeol.* **1(1)**, 19–41 (2018).
77. L. Luo *et al.*, Airborne and spaceborne remote sensing for archaeological and cultural heritage applications: A review of the century (1907–2017). *Remote Sens. Environ.* **232**, 111280 (2019).
78. L. Breiman, Random Forests. *Mach. Learn.* **45(1)**, 5–32 (2001).
79. E. R. Crema, T. Kerig, S. Shennan, Culture, space, and metapopulation: a simulation-based study for evaluating signals of blending and branching. *J. Archaeol. Sci.* **43**, 289–298 (2014).
80. T. M. Oshiro, P. S. Perez, J. A. Baranauskas, "How Many Trees in a Random Forest?" in *Machine Learning and Data Mining in Pattern Recognition*, P. Perner, Ed. (Springer-Verlag, Berlin, 2012), pp. 154–168.
81. C. Stewart, R. Lasaponara, G. Schiavon, Multi-frequency, polarimetric SAR analysis for archaeological prospection. *Int. J. Appl. Earth Obs. Geoinf.* **28(1)**, 211–219 (2014).
82. F. El-Baz, M. Maingue, C. Robinson, Fluvio-aeolian dynamics in the north-eastern Sahara: the relationship between fluvial/aeolian systems and ground-water concentration. *J. Arid. Environ.* **44(2)**, 173–183 (2000).
83. D. W. Holcomb, I. Shingiray, "Imaging Radar in Archaeological Investigations: An Image Processing perspective" in *Remote Sensing in Archaeology*, J. Wiseman, F. El-Baz, Eds. (Springer, New York, 2007),

pp. 11–46.

84. H. Kumar, A. S. Rajawat, Potential of RISAT-1 SAR data in detecting palaeochannels in parts of the Thar Desert, India. *Curr. Sci.* **113**(10), 1899–1905 (2017).
85. A. Agapiou, D. D. Alexakis, A. Sarris, D. G. Hadjimitsis, Evaluating the potentials of Sentinel-2 for archaeological perspective. *Remote Sens.* **6**(3), 2176–2194 (2014).
86. D. Tapete, F. Cigna, Appraisal of opportunities and perspectives for the systematic condition assessment of heritage sites with copernicus Sentinel-2 high-resolution multispectral imagery. *Remote Sens.* **10**(4), 1–22 (2018).
87. D. Lawrence, G. Philip, H. Hunt, L. Snape-Kennedy, T. J. Wilkinson, Long Term Population, City Size and Climate Trends in the Fertile Crescent: A First Approximation. *PLoS One* **11**(3), e0152563 (2016).
88. G. L. Possehl, The Indus Civilization. *Man Environ.* **19**(1–2), 103–113 (1994).
89. M. Madella, *et al.*, "The North Gujarat Archaeological Project – NoGAP. A multi-proxy and multi-scale study of long-term socio-ecological dynamics" in Walking with the Unicorn. Social Organization and Material Culture in Ancient South Asia, D. Frenez, G. Jamison, R. Law, M. Vidale, R. Meadow, Eds. (Archaeopress Publishing Ltd, Oxford, 2018), pp. 342–358.
90. D. Parikh, C. A. Petrie, 'We are inheritors of a rural civilisation': rural complexity and the ceramic economy in the Indus Civilisation in northwest India. *World Archaeol.* **51**(2), 252–272 (2019).
91. J. A. Durcan *et al.*, Holocene landscape dynamics in the Ghaggar-Hakra palaeochannel region at the northern edge of the Thar Desert, northwest India. *Quat. Int.* **501**, 317–327 (2019).
92. R. Bryson, M. Swain, Holocene variations of monsoon rainfall in Rajasthan. *Quat. Res.* **16**(2), 135–145 (1981).
93. Y. Dixit, D. A. Hodell, C. A. Petrie, Abrupt weakening of the summer monsoon in northwest India 4100 yr ago. *Geology* **42**(4), 339–342 (2014).
94. S. Prasad, Y. Enzel, Holocene paleoclimates of India. *Quat. Res.* **66**(3), 442–453 (2006).
95. V. Prasad *et al.*, Mid–late Holocene monsoonal variations from mainland Gujarat, India: A multi-proxy study for evaluating climate culture relationship. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **397**, 38–51 (2014).
96. R. Raj *et al.*, Holocene climatic fluctuations in the Gujarat Alluvial Plains based on a multiproxy study of the Pariyaj Lake archive, western India. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **421**, 60–74 (2015).
97. A. K. Singhvi, A. Kar, The aeolian sedimentation record of the Thar desert. *Proc. Indian Acad. Sci. Earth Planet. Sci.* **113**(3), 371–401 (2004).
98. F. Ahmad, Historical and archaeological perspectives of soil degradation in Cholistan. *Soc. Naturez* **1**(1), 864–870 (2005).
99. A. Srivastava, D. S. G. Thomas, J. A. Durcan, Holocene Dune Activity in the Thar Desert, India. *Earth Surf. Process. Landforms* **44**(7), 1407–1418 (2019).
100. A. Chatterjee, J. S. Ray, Geochemistry of Harappan potteries from Kalibangan and sediments in the Ghaggar River: Clues for a dying river. *Geosci. Front.* **9**(4), 1203–1211 (2018).
101. A. Chatterjee, J. S. Ray, Sources and depositional pathways of mid-Holocene sediments in the Great Rann of Kachchh, India: Implications for fluvial scenario during the Harappan Culture. *Quat. Int.* **443**, 177–187 (2017).

102. A. K. Dave, M. A. Courty, K. E. Fitzsimmons, A. K. Singhvi, Revisiting the contemporaneity of a mighty river and the Harappans: Archaeological, stratigraphic and chronometric constraints. *Quat. Geochronol.* **49**, 230–235 (2019).
103. N. Khonde *et al.*, Tracing the Vedic Saraswati River in the Great Rann of Kachchh. *Sci. Rep.* **7(1)**, 1–6 (2017).
104. M. Jain, S. K. Tandon, Fluvial response to Late Quaternary climate changes, western India. *Quat. Sci. Rev.* **22(20)**, 2223–2235 (2003).
105. Punjab State Gazetteers, *Bahawalpur State with Maps (1904)* (The Civil and Military Gazette Press, Lahore, 1908).
106. B. Ghose, A. Kar, Z. Husain, The Lost Courses of the Saraswati River in the Great Indian Desert: New Evidence from Landsat Imagery. *Geogr. J.* **145(3)**, 446–451 (1979).
107. M. B. Rajani, A. S. Rajawat, Potential of satellite based sensors for studying distribution of archaeological sites along palaeo channels: Harappan sites a case study. *J. Archaeol. Sci.* **38(9)**, 2010–2016 (2011).
108. J. R. Sharma, A. K. Gupta, B. K. Bhadra, Course of Vedic River Saraswati as Deciphered from Latest Satellite Data. *Puratattva* **26**, 187–195 (2005).
109. Z. Islam, J. Iqbal, J. Aziz, Q. Qazi, Paleochannel delineation using Landsat 8 OLI and Envisat ASAR image fusion techniques in Cholistan desert, Pakistan. *J. Appl. Remote Sens.* **10(4)**, 046001 (2016).
110. S. Muntazir, N. Pant, H. S. Saini, S. A. I. Muftaba, P. Pande, Identification of palaeochannel configuration in the Saraswati River basin in parts of Haryana and Rajasthan, India, through digital remote sensing and GIS. *Episodes* **39(1)**, 29–38 (2015).
111. M. Arshad, A. Rao, G. Akbar, Masters of Disaster in Cholistan Desert, Pakistan: pattern of nomadic migration. *Desertif. Control Bull.* **35**, 33–38 (1999).
112. A. Hussain, Z. Akram, Agrarian economy of the princely state of Bahawalpur: a historical perspective (1866-1924). *J. Res. Soc. Pakistan* **54(2)**, 15–26 (2017).
113. M. Madella, "Of Crops and Food: A Social Perspective on Rice in the Indus Civilization" in *Ancient Plants and People: Contemporary Trends in Archaeobotany*, M. Madella, L. Lancelotti, M. Savard, Eds. (University of Arizona Press, Tucson, 2014), pp. 218–236.
114. J. M. Kenoyer, *Ancient Cities of the Indus Valley Civilization* (Oxford University Press, Carachi, 1998).
115. R. Wright, *The ancient Indus: urbanism, economy, and society* (Cambridge University Press, Cambridge, 2010).

Figure legends

Figure 1. Location map of the Cholistan Desert: a) distribution of Indus sites with the location of the five major Indus cities during the urban period (c. 2500-1900 BC); b) distribution of Indus sites in the Cholistan Desert, after Mughal (26, 47); c) inset showing characteristic fossilised sand dunes, mud flats (*dahars*) and water ponds (*tobas*); d) the well-known mound of Ganweriwala, partially covered by a sand dune.

Figure 2. Spatial distribution of Mughal's sites per period. Approximate distribution of previously known Indus sites in the area (see 26, recently revised by 47). The smooth regression lines represent the spatial trends in the

distribution of sites. Although many of these coordinates need re-visiting in the field, the chronological distribution of Mughal's sites evidence continuous shifts in settlement.

Figure 3. Schematic workflow used in this study. The code available within the SI Appendix follows the three main steps of this research performed in GEE: 1) development of a multi-sensor, multi-temporal image composite, 2) train and apply the RF classifier and 3) export the resulting probability raster field. In addition, data validation and statistics were performed using R software.

Figure 4. SAR and multispectral mound visibility: a) Google Earth basemap showing the location of a well-preserved mound (yellow circle) and three main land cover types in the desert edge: *dahars* or mud flat surfaces, stabilised sand dunes and spots of irrigated lands. Note the differences in mound and land cover visibility in the following band combinations from the multi-temporal image composite: b) dual Sentinel 1 band [VV,VH] in ascending mode ; c) single Sentinel 1 band [VV] in ascending mode; d) Sentinel 1 false composite in RGB; e) Sentinel 2 visible composite (B4-B3-B2); f) Sentinel 2 false colour composite (B8-B4-B3).

Figure 5. Results of the RF classifier: a) in red, extent of the study area, showing the distribution of new RF probability mounds; b) inset showing the RF probability at the desert edge, note the white dots scattered through the region indicating high-probability mounds; c) visible high-resolution imagery (Google Earth basemap) with virtual absence of mounds; d) the same area as c), showing high RF probability mound-like signatures in *dahar* surfaces; e) inset showing filtered pixels at >0.55 RF probability threshold, suggesting the presence of mounds partially covered by sand dunes.

Figure 6. RF probability outputs: a) example of detected mound-like surfaces in well-known mounds used as validation set; b) clusters of high-probability pixels in the area, photo-interpreted as new archaeological mounds. Note the distinct preservation of mounds due to partial coverage of sand dunes and desert shrubs.

Figure 7. Distribution of newly detected mounds in relation to regional landcover.

Figure 8. Visibility of RF mounds and legacy data: Google basemaps and RF probability fields showing a) the vectorised new mound of Bokhariyanwala, closely located to multiple legacy coordinates for the same site; b) the vectorised new mound of Chagalawa, also reported as multiple locations in legacy data probably due to the partial obliteration of the site by an irrigation canal.

Figure 9. Mound area estimates: a) map showing the mound size estimates. The location of Ganweriwala and the other 5 large mounds (>20 ha) located in the region is highlighted. Note the extension of the Ghaggar-Hakra River, as digitised from historical maps; b) histogram with the size density for the mounds. Mean size is indicated by the red dashed line.

















