Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Hexagon KH-9 Imagery using Sunspot: Monitoring At-Risk Heritage and Development Pressure in Haryana India (1915-2021)

Petrus J. Gerrits

p.jg65@cam.ac.uk

University of Cambridge

Carla Klehm

Center for Advanced Spatial Technologies

Jackson Cothren

Center for Advanced Spatial Technologies

Hayley Hames

Center for Advanced Spatial Technologies

Rebecca Roberts

University of Cambridge

Mou Sarmah

University of Cambridge

Jack Tomaney

University of Cambridge

Vaneshree Vidyarthi

University of Cambridge

Seth Warn

Center for Advanced Spatial Technologies

Cameron A. Petrie

University of Cambridge

Article

Keywords: Declassified imagery, Hexagon (KH-9), Sunspot, heritage monitoring, land-use change, historical maps, Survey of India, Haryana

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11406373/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Hexagon KH-9 Imagery using Sunspot: Monitoring At-Risk Heritage and Development Pressure in Haryana India (1915-2021)

Authors: Petrus J. Gerrits[1], Carla Klehm[2,3], Jackson Cothren[2,4], Hayley Hames[2], Rebecca Roberts[1], Mou Sarmah[1], Jack Tomaney[1], Vaneshree Vidyarthi[1], Seth Warn[2,4] and Cameron A. Petrie[1,5]

Affiliations:
[1]McDonald Institute for Archaeological Research, University of Cambridge, Cambridge, United Kingdom
[2]Center for Advanced Spatial Technologies, University of Arkansas, Fayetteville, AR, United States
[3]Department of Anthropology, University of Arkansas, Fayetteville, AR, United States
[4]Department of Geosciences, University of Arkansas, Fayetteville, AR, United States
[5]Department of Archaeology, University of Cambridge, Cambridge, United Kingdom

Corresponding author: Petrus J. Gerrits, pjg65@cam.ac.uk; Courtyard Building, Downing Street, Cambridge, CB2 3ER, The United Kingdom

Abstract:

Archaeological mounds in Haryana, India, are being lost to agricultural intensification and urban growth, yet the most destructive decades fall between the two records commonly used to monitor them: early twentieth-century Survey of India maps and recent satellite products. We conduct a regional-scale application of declassified Hexagon (KH-9) panoramic imagery to heritage monitoring, using the updated Sunspot platform to georeference six 1975 frames with 20-25 ground control points each. Land cover for 1914-15 (maps), 1975 (Hexagon) and 2021 (ESA WorldCover) was classified independently, harmonized and compared within distance zones around twelve sites. Ground not classed as cropland fell from 35% of the observed area in 1915 to 14% in 1975 and 4% in 2021, whereas built-up land within 1 km of sites grew later and unevenly, reaching 75% at the most affected site. Separating these stages differentiates site-level risk and demonstrates the value of archival imagery and historical maps for heritage monitoring.

Keywords: Declassified imagery, Hexagon (KH-9), Sunspot, heritage monitoring, land-use change, historical maps, Survey of India, Haryana

1. Introduction

Heritage sites across the globe are at risk. Monitoring, management, and even intervention are imperatives to prevent severe degradation or complete loss [1–4]. Such sites include, for example, some of the world's earliest urban centers, such as those associated with the Indus Civilization (c. 2600-1500 BC) in South Asia [5]. The state of Haryana, India, hosts important archaeological landscapes, including a large share of the known settlement record of the Indus Civilization, along with later Early Historic and Medieval urban centers [6–8]. Many of its archaeological sites survive as raised mounds that remain visible today. At the same time, the rapid urbanization and agricultural development, along with increasing environmental stressors from the effects of climate change, mean that sites in Haryana are at high risk of destruction within the next few decades. The same processes that historically made the region one of India's most productive agricultural zones, namely canal irrigation and rapid urban development, are gradually erasing them [9,10]. These are not sudden changes but part of a gradual process: land is first, if possible, brought under cultivation and only later built over, and the two damage a mound differently and at different times. Tracking them requires spatial records spanning a long period, which this paper assembles by pairing an early 20th-century colonial-era map series, Cold War spy satellite imagery, and a more recent satellite imagery product in order to understand the impacts of change over the past century in Haryana, and when and how these changes compound over time.

The Mapping Archaeological Heritage in South Asia (MAHSA, https://www.mahsa.arch.cam.ac.uk) project, based at the University of Cambridge and supported by Arcadia, documents endangered archaeological heritage throughout the Indus River Basin in India and Pakistan, the Ganges basin, and the coasts of Pakistan and northwestern India. MAHSA, along with other Arcadia projects including Mapping Africa's Endangered Archaeological Sites and Monuments (MAEASaM) and Endangered Archaeology in the Middle East and North Africa (EAMENA) are among the largest-scale heritage documentation and management efforts globally. MAHSA works with in-country partners and other collaborators to co-create and enrich database records for its open-access database, while also working to support heritage management by the relevant authorities. The database comprises information concerning the location, size, shape, periods of occupation, degree of preservation, and risk level of archaeological heritage; where available, it includes map images, photographs, features, finds, and bibliographic information for sites that have been studied [8]. All information is structured and recorded in such a way as to give due credit to those involved in its creation at every step, including researchers creating primary records from their archaeological research, organizations responsible for managing archaeological data, and those working with the records through activities such as auditing data or verifying the current condition of sites on the ground.

Legacy data (i.e., published information sources, including historical imagery) can help provide crucial time series datasets to identify the pace and impacts of development on heritage landscapes. Low-cost, high-resolution imagery for archaeology and heritage management is essential for

monitoring regions at a large scale [11], as are historical maps [7,8,12–14]. MAHSA and its predecessors, the TwoRains and Land, Water, Settlement projects, have demonstrated the utility of the Survey of India (SoI) one-inch-to-one-mile map series, which was created as part of large-scale mapping efforts in the late 19th and early 20th centuries to support British colonial control of territories, including the assessment of land revenue [7,8,15]. Through documenting elevated features in the landscape, SoI mapping teams incidentally conducted a systematic survey of thousands of mound features across South Asia, many of which are important archaeological sites, which were therefore recorded decades before their formal recognition as such [8]. The maps provide a baseline indication of thousands of archaeological heritage locations prior to the rapid expansion of agriculture during the 'Green Revolution' in the 1960s [12]. By using machine learning, these features can be automatically traced and extracted from thousands of map sheets at scale, removing the need for time-consuming manual extraction of map features, meaning that researchers can instead focus their energies on verification and cross-checking of these potential archaeological sites [16]. Digitized SoI map features are matched with modern ESA WorldCover satellite imagery and legacy data records to determine the current state of preservation and level of risk to archaeological heritage [8,17]. The problem for large-scale risk monitoring in Haryana, not least attempting to determine the scale and risk, however, is one of timing: the most destructive decades, roughly the 1950s to the 1990s, fall mostly between the two most commonly used records, namely the historical maps of the early 1900s and modern satellite imagery [18].

Declassified US spy satellite imagery collected as part of Cold War reconnaissance efforts during the 1960s through the 1980s provides near-global coverage in high resolution: up to 2-3m in the late 1960s and early 1970s and 0.6m by the late 1970s. This declassified imagery, available through the USGS's EarthExplorer archive, is available for low-to-no cost: $30 for the initial scanning of a film, if it has not been digitized, and available perpetually for free afterwards for any interested user. Declassified imagery is often referred to by code names associated with the various missions, and many publications refer specifically to these names. For the two high-resolution, panoramic camera missions mentioned earlier, these are Corona (KH-4, and more specifically KH-4A/B) and Hexagon (KH-9). Corona imagery has been extensively used by the archaeological and heritage community for prospection and monitoring (e.g., [19,20]), as well as for land use and land cover change (e.g., [21–25]). Recently, for example, Corona led to the discovery of nearly 400 Roman-era forts in the northern portion of the Fertile Crescent (modern-day Syria and Iraq) [26]. This eastern frontier of the Roman Empire becomes interpretable in an otherwise inaccessible region, one which has seen rapid development and destruction of heritage sites over the last several decades [27]. More than 10,000 archaeological sites have been discovered in the northern Fertile Crescent alone with Corona imagery [20]. Declassified imagery is not only valuable, but is often a unique resource for heritage monitoring, given its window into the past with resolution where even the layout of sites is visible.

Use of Corona imagery by the archaeological community became popularized after the first release of Sunspot, a free, web-based photogrammetry and georeferencing platform developed by the Center for Advanced Spatial Technologies (CAST) at the University of Arkansas (https://sunspot.cast.uark.edu) [20,28,29]. Sunspot helps correct the complicated distortions associated with declassified panoramic imagery, such as Corona, allowing users to use the full imagery strip, rather than a small, isolated area. Given the coverage of a single image (e.g., with Corona, 13.8 x 188 km), landscape-scale analyses become possible with georeferenced, orthorectified products that can be then brought into common geospatial software such as ArcGIS and QGIS. Alternative approaches also exist (e.g., [30–34]), though they may require more expertise (i.e., command line proficiency), which may be beyond many heritage practitioners; more ground control points, which may not be possible in rural areas; and/or, given limitations of the user and the study area, may only be used reliably for small, discrete areas of interest, such as monitoring of a single site at nadir (see [35]).

Hexagon mapping camera and panoramic camera imagery, collected between 1971 and 1984, has become available during the last dozen years (declassified over two events, 2002 and 2018). However, the ability for researchers to process and model this imagery has been extraordinarily challenging, which in turn has limited its adoption and scientific applications. This is due to the complex distortions associated with the panoramic cameras and issues with the lack of metadata. Internal and external camera parameters are difficult to discern, and the camera tilts and orientations vary between missions. This allowed for more tailored imagery acquisition, including wide swaths of the landscape in a single image; it also means a single, simple solution for georeferencing is hard to model and that off-nadir locations in raw images can be heavily distorted (see [36–39]). Commercial geospatial software currently does not handle panoramic cameras well, especially these, and each mission requires different corrections. There is growing awareness, and excitement, across the archaeological and cultural heritage community about Hexagon because of its greater coverage and clarity even over Corona (e.g., [40]). Yet, its complex nature means that use has been extremely limited, and primarily to only small geographic areas. However, Sunspot has recently been updated to accommodate both Hexagon panoramic (the aforementioned high resolution) cameras as well as Hexagon mapping camera imagery.

We present the first case study about how Hexagon panoramic imagery can enhance heritage monitoring at larger, regional scales. Just as with the SoI historical maps, Hexagon as a data source was originally made to control and document land, rather than to conserve it; SoI maps to assess it for revenue, and Hexagon to surveil it. Nevertheless, both datasets recorded the land use categories and the interval that makes a long-term mapping of land development pressure on heritage possible. We also introduce the updated version of Sunspot, which now handles both Corona and Hexagon imagery at scale. Sunspot models the full image with relatively few ground control points and provides accurate, georeferenced, and orthorectified map products suitable for modern GIS and analytical workflows. The Sunspot tool is appropriate for both novice and

experienced users of remote sensing imagery, reduces the time required for georeferencing dramatically, and allows for multi-user, multi-location projects to take place. These new capabilities also enable multi-user, multilocation teams through an online platform without the need to upload or process data-heavy imagery, which extends possibilities for researchers in lower-resourced technology environments, such as those based in the Global South, to coordinate and share [41]. We see this as an innovative, essential resource for large-scale heritage management that is, importantly, low cost for the end user.

Hexagon panoramic imagery processed with Sunspot is incorporated into the MAHSA workflow alongside SoI maps and ESA WorldCover land-cover data. Together, these sources enable a quantitative assessment of changing land use and land cover and associated risks with cultural heritage sites across Haryana over approximately a century. This paper focuses on twelve selected sites (Figure 1) to assess the value of this multi-temporal approach for more localized risk assessment. By integrating declassified satellite imagery with historical maps and contemporary land-cover products, MAHSA and Sunspot demonstrate how archival remote sensing can support the documentation and monitoring of archaeological heritage threatened by rapid development.

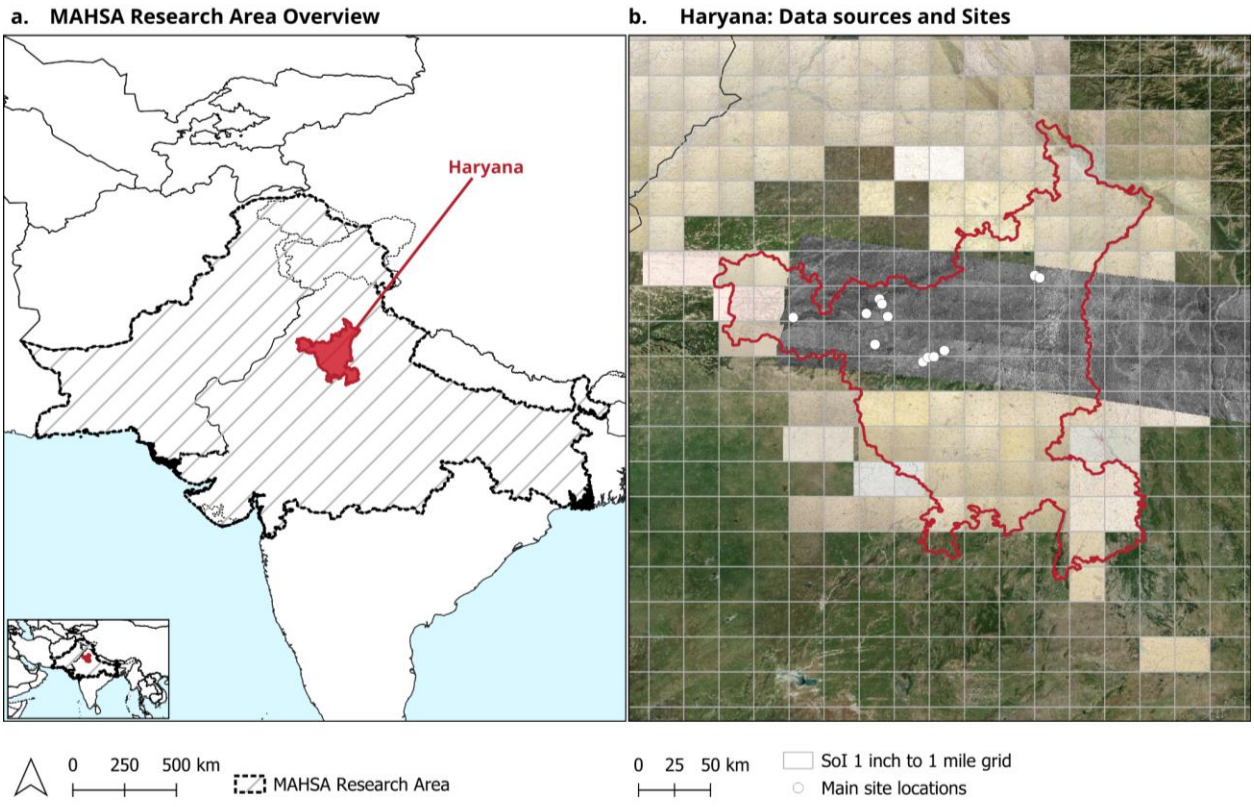


Figure 1. The study area and its data sources. ***a.*** *The MAHSA research area (hatched), with Haryana in red.* ***b.*** *Haryana (red outline) with the Survey of India one-inch-to-one-mile sheet index; the seventeen sheets used here were surveyed in 1914–15. The gray band (b) is the combined footprint of the six Hexagon Mission 1211 frames of December 1975; it does not fully reach the west of the state. White circles mark the twelve study sites analyzed in this research.*

2. Methods

The combined use of historical map series, declassified satellite imagery, and recent satellite products for quantitative, large-scale research involves a complex and time-consuming approach. Often, the scale, geometric distortion, color and radiometric qualities, spatial coverage, and methods of recording are just some of the things that can vary across these sources. In order to prepare each dataset for analysis while maintaining the characteristics and limitations of its original source data, our methodology uses discrete, source-specific processing pipelines.

This study's initial methodological concept is that each epoch is processed individually. SoI maps, KH-9 Hexagon images, and ESA WorldCover data are all generated, classified, and measured using their own geometry first, before being jointly evaluated. No epoch is registered with or derived from another. Instead, the outputs are harmonized just at the land-cover class level and combined using summary statistics that are resilient to residual positional uncertainty.

Our workflow has four stages (Figure 2). First, data preparation transforms each source into an analysis-ready raster and generates heritage site-centered analysis zones for each epoch (Section 2.2). Second, data processing methods tailored to each media (such as color- and morphology-based classification for historical maps, deep-learning classification for KH-9 imagery, and class remapping for WorldCover) are used to create a unified land-cover scheme (Section 2.3). Third, the analysis compares each epoch's land-cover classification to its corresponding site-anchored zones. This yields long-term land-cover fractions, estimates of development in the distance bands surrounding each site, and the distance-to-nearest-built metrics described in Section 3. Finally, the per-epoch results are ultimately combined to build long-term landscape change trajectories and nearby encroachment estimates at the site level.

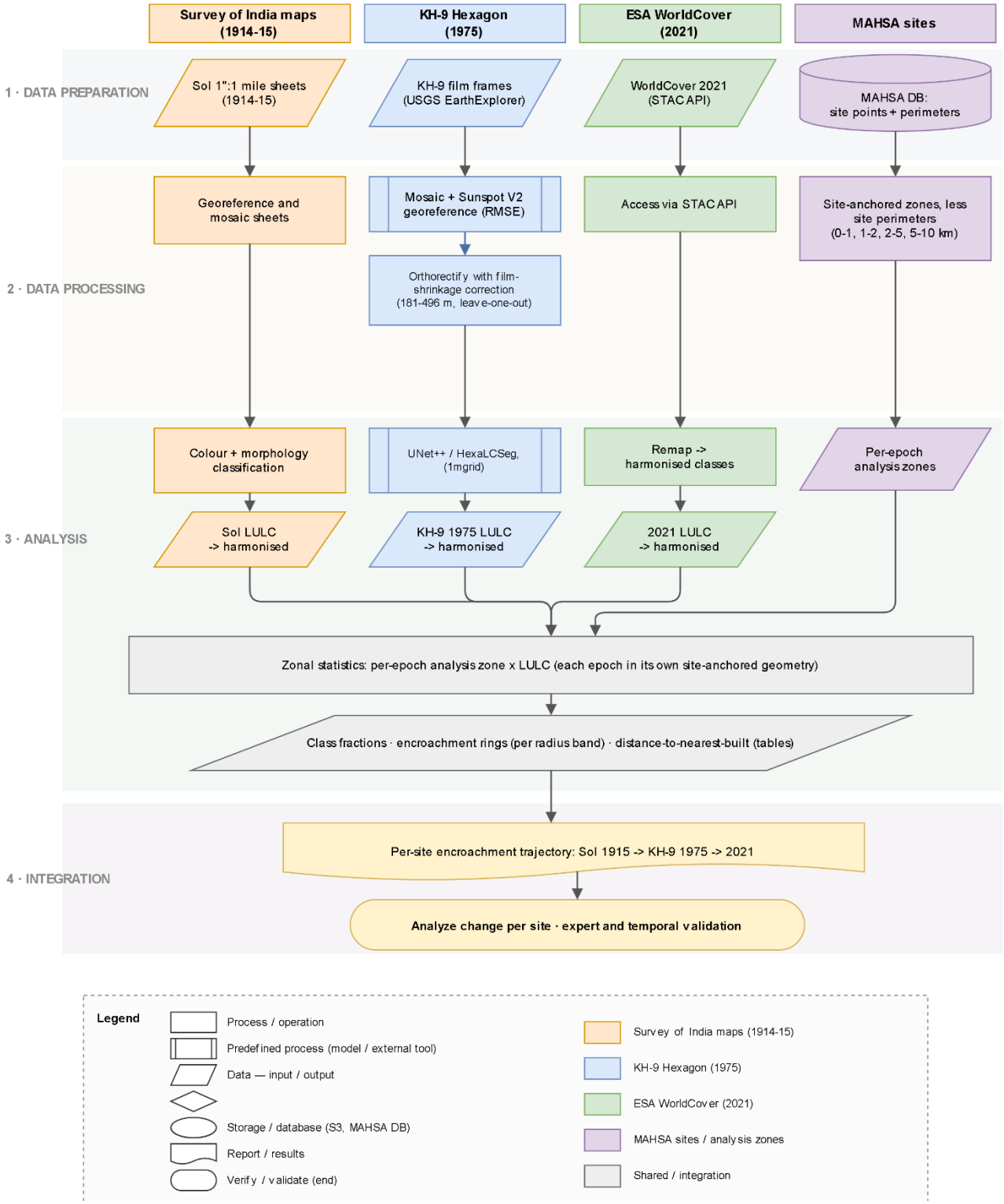
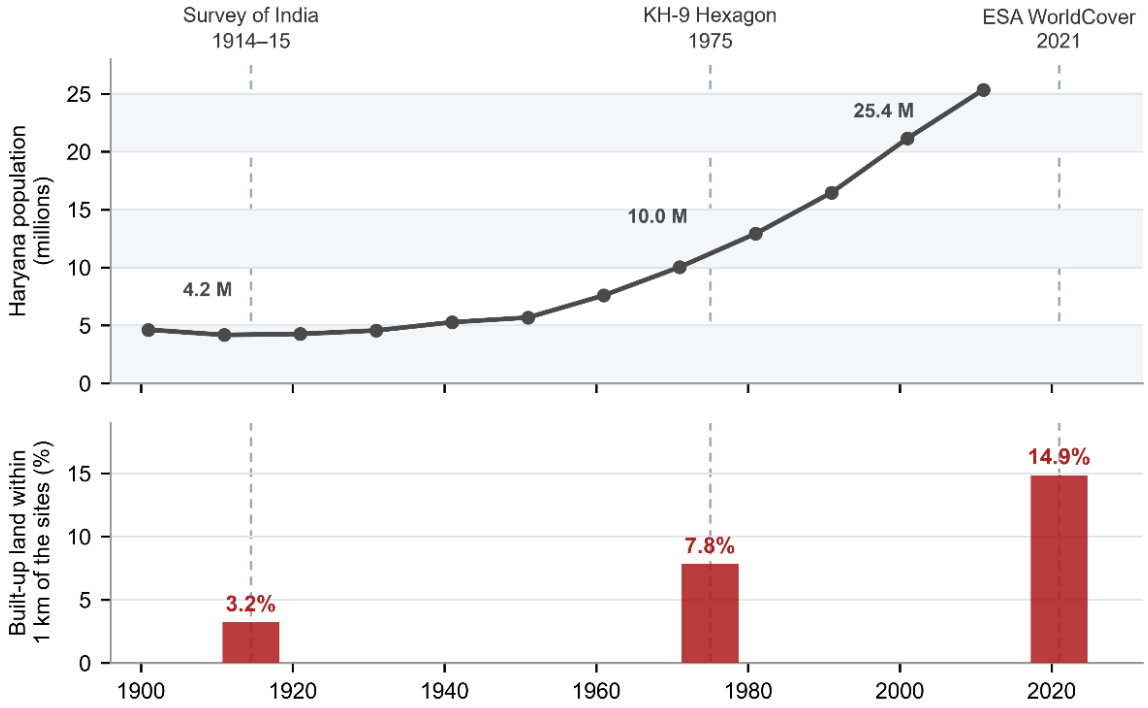


Figure 2. Three-epoch heritage-monitoring workflow using SoI, Hexagon, along with Sunspot, and ESA WorldCover.

2.1 Study area and archaeological background

Figure 3. Showing the population and average built-up land increase around the twelve study sites in Haryana. Above:



state population at each decennial census, 1901-2011, with the three research epochs marked. Below: built-up land as a share of observed ground within 1 km of the sites at each epoch, pooled (area-weighted) across the twelve sites; values for individual sites are given in Table 3 and Figure 7. Haryana was constituted in 1966; earlier figures relate to the territory that now forms the state [42].

The state of Haryana occupies some 44,200 km² of the northwestern Indo-Gangetic plain, bounded by the Yamuna River to the east and grading westward into the semi-arid margin of the Thar desert. Its landscape is dominated by the alluvial plains of the Ghaggar-Hakra and its relict paleochannels, along which much of the region's early settlement was distributed [43–45]. Haryana is among the most intensively cultivated states in India, and it combines high archaeological density with severe development pressure. Figure 3 shows that the population of the territory now forming the state had risen almost sixfold over the interval spanned by our three epochs, from approximately 4.2 million when the SoI sheets were surveyed to 25.4 million at the most recent census (2011). Even the rate of that increase changed during that period, with decadal growth above a quarter in every decade from 1961 to 2001 [42,46,47]. Figure 3, furthermore, compares that number against our estimated built-up of land immediately around the twelve study sites at each of our three epochs. It shows that for Haryana, as is likely the case for areas with similar development patterns, the sharpest demographic growth precedes the sharpest expansion of settlements near the sites. Haryana was constituted in 1966, so figures for earlier censuses relate to the territory that now

forms the state. How that growth was expressed in the landscape, first through cultivation and later through settlement, is considered in Section 3.1.

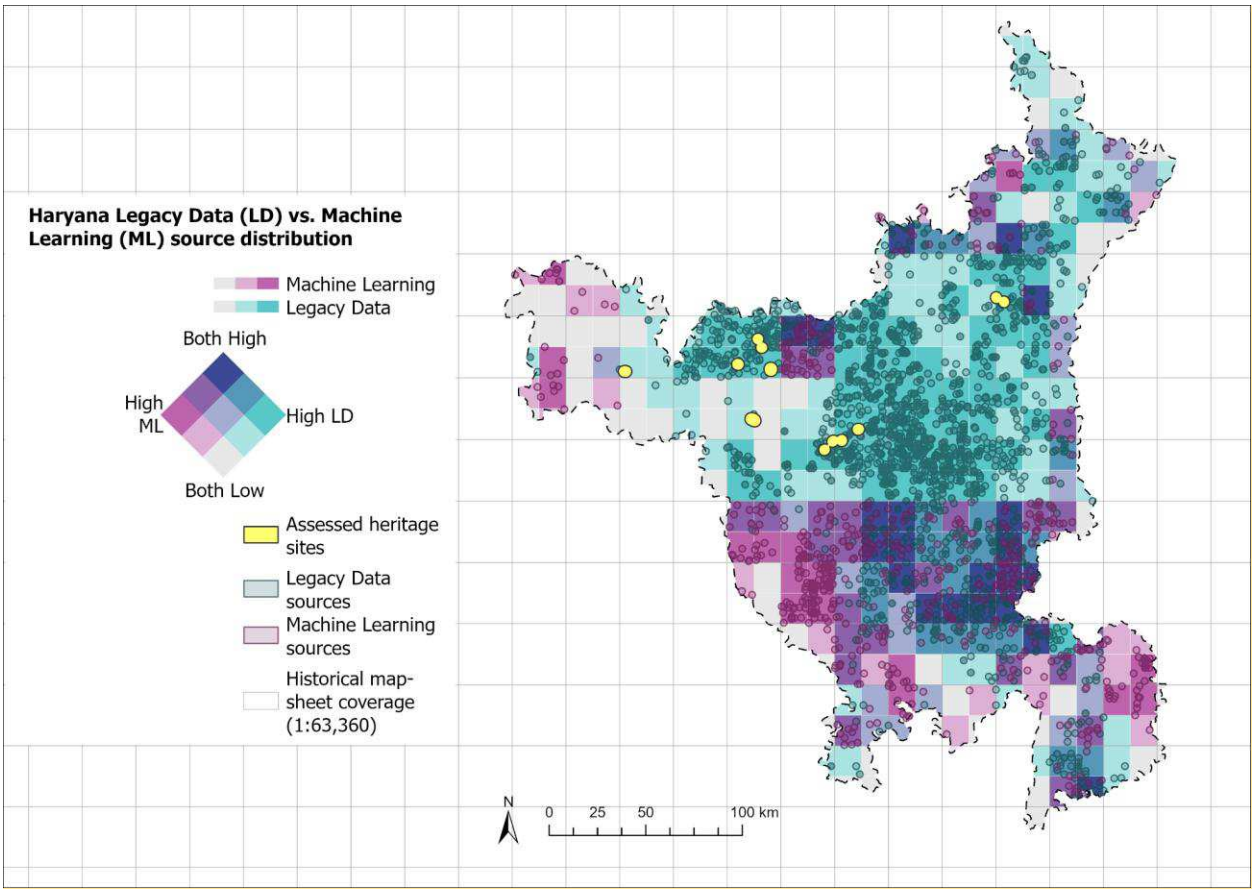


Figure 4. Bivariate map of legacy-data (LD) records versus machine-learning (ML) mound detections per grid cell across Haryana (site concentration), showing the complementary coverage of the two datasets conducted in the MAHSA project. Note that the lack of ML points in the center of the state is likely due to the earlier image quality of the SoI Maps used in the previous ML workflow and is to be improved [16]. Legacy data: [6,10,12,48–50].

Background information about Haryana's rich heritage sites draws from several legacy data sources, such as SoI, but also archaeological surveys that have taken place in the region for over six decades of separate survey campaigns [10,50]. Parts of this dataset were compiled and curated by earlier research projects (Land, Water, and Settlement; TwoRains) [6,48–51]. The MAHSA project is working towards consolidating these efforts, together with analyzing published gazetteers and recent systematic fieldwork to ground-truth remotely sensed archaeological sites. To date, MAHSA's open-access database holds close to 3,000 known site locations in the Haryana state (https://databasemahsa.org), although these are currently not (yet) released to the general public while a secure permissions structure is developed within the database in order to enable the protection of vulnerable sites (Figure 4). MAHSA has complemented these legacy records with

machine learning, using automated mound detection on the SoI map series to add several hundred further candidate sites in Haryana alone (684 mounds; [16]). The two evidence streams have complementary geographies (Figure 4). Location precision varies widely in these legacy sources due to survey methods and the availability of different technologies at the time of recording: surveys conducted before GPS was widely available often recorded locations only to degrees and minutes, which translates into errors from several hundred meters up to ten kilometers in some cases [7,8]. The individual case-study sites (Section 2.2.5) were drawn from this database, and the site-anchored design (Section 2.4) is, in part, a response to this kind of positional uncertainty.

2.2 Data acquisition

Three epochs represented through historical imagery were considered in this research, together with the MAHSA heritage site records. Each source and platform for acquisition is described in this section.

2.2.1 Survey of India (SoI) map sheets

The SoI one-inch-to-one-mile map series (1:63,360 scale) represents a systematic and high-resolution documentation of the South Asian landscape, with topographic maps published from the early 19th century onwards [15]. Their initial use served imperial administrative, military, and revenue objectives [7,8]. However, for MAHSA and archaeological research, these high-quality geographic records incidentally preserved a wealth of information regarding thousands of elevated mound features that were visible prior to the intensive agricultural and urban transformations of the mid-to-late 20th century [7,12]. To some degree, the maps capture a "vanished" landscape, documenting archaeological sites that were often first identified by SoI survey teams [8].

In this research, seventeen digitized and georeferenced SoI one-inch-to-one-mile sheets (map grids: 044K14; 044O02, O03, O06, O07, O10, O11, O12, O14, O15, O16; 053C03, C04, C09, C10, C13, C14), surveyed in 1914 and 1915 and scanned in color by *the British Library and Cambridge University Library*, provide the earliest baseline in this analysis, all from the year 1914-1915. For eleven of the twelve sites, we have complete sheet coverage of their analysis area, and the twelfth, Sirsa, has 90%. One further sheet covering part of the Sirsa area, 044K15 was available, but we excluded it because it was a 1946 revision drawn in a different cartographic style and carrying very little area-fill color for crop and grasslands. Furthermore, its later publication date would mean it would both mix a mid-century land surface into an early-century epoch and enlarge the observed area without contributing meaningfully classified land cover.

The SoI maps have a good separation of the visualization of land use and land cover classification on these map series. For the purpose of this research, we identified the labels associated with land development, such as cultivated and grassland, as well as water, built-up areas, and vegetation. The legend on Figure 5 highlights these distinct classes.

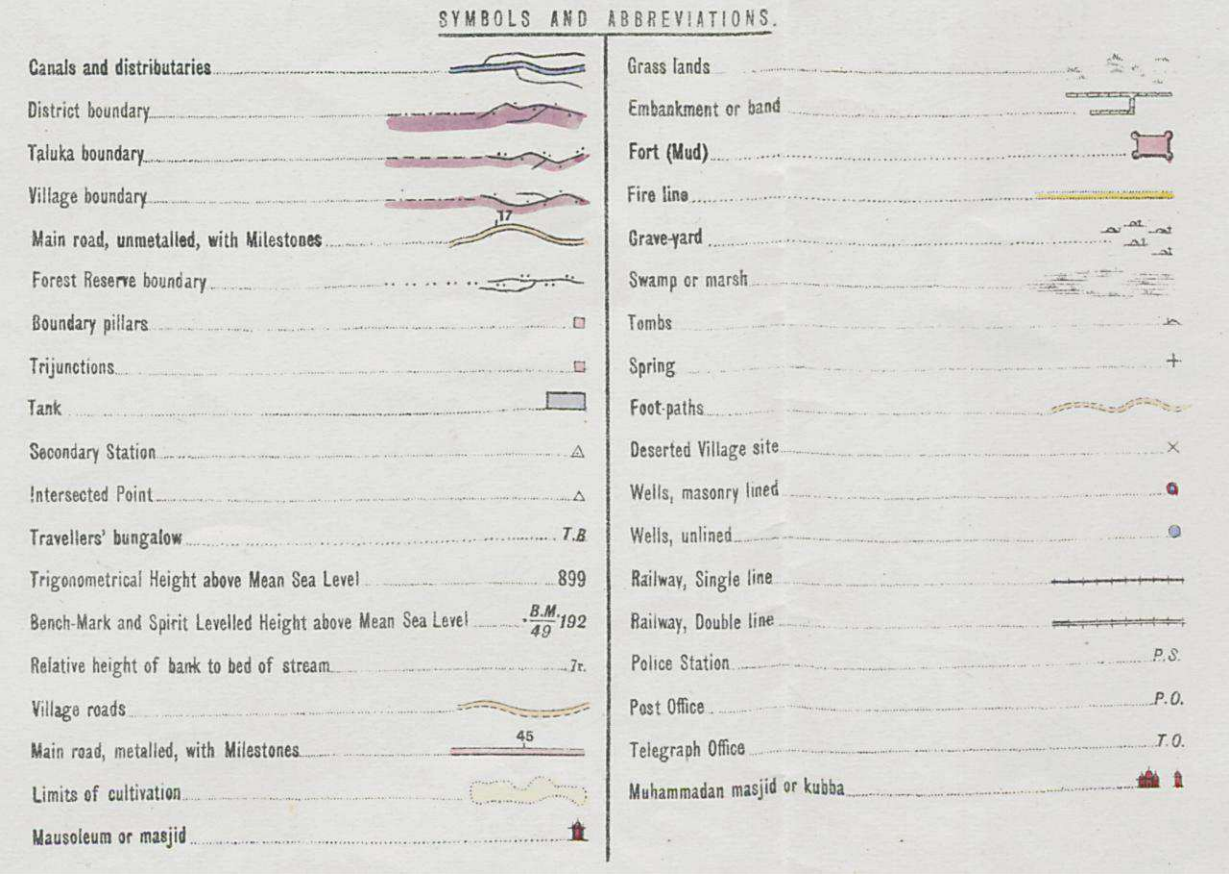


Figure 5. Survey of India legend, showing the main administrative, settlement, and land use boundaries frequently drawn on the historical map collection. In this research, "limits of cultivation" and "grasslands" were of particular interest, as well as the built-up area. Map source: British Library, 1 inch 1 mile, shelfmark: IOR_X_9875_1_47_1906_001, URL: https://databasemahsa.org/en/report/463607f2-3fb4-43df-8df1-375e0027ca64

2.2.2 The Sunspot platform

Sunspot (https://sunspot.cast.uark.edu) is an open-source geospatial referencing tool for historical imagery and has been essential in bringing together the Haryana research methodology. Sunspot's initial use focused on correcting the distortions associated with Corona panoramic (KH-4A/4B) imagery [28,29]. Sunspot's ability to ease the modeling of Corona imagery has led to dozens of publications and the corresponding thousands of archaeological sites discovered (e.g., [19,20,52]). The associated Corona Atlas, which served as a visual mapping tool for any of the Corona imagery that was processed through Sunspot (see [29,41]), quantified the extent of impact: 2,214 images covering 7.7 million sq. km, or 13% of the Earth's land surface.

Sunspot has been recently updated to accommodate not only Corona panoramic cameras but also Hexagon (KH-9) panoramic and mapping (frame) cameras. Users retrieve Corona and Hexagon images directly in Sunspot; they can measure ground control points using a contemporary image-based reference map; create a sensor model that maps pixels to geographic coordinates; and export an image that is suitable for both 2D (orthorectified images) and 3D (oriented images) analysis using the same methods and in the same software applications used for modern aerial and satellite photogrammetric analysis. This process is described in the workflow developed by Vidyarthi and colleagues [53] and was used in this analysis. Provided the ground control points have good distribution, users can take advantage of the entirety of the image, be it locations near nadir or at the edges of a coverage area.

Sunspot also has capabilities for collaborative work. Multiple registered users can be added to and collaborate on the same project, which can include any number of images the user selects. The same user can have multiple projects, with different sets of imagery in each. Besides unique names for the projects and short descriptions of the contents, the images can be renamed or have additional descriptions associated with them. This allows for better project organization, not least the option of collaborative working to find the image(s) of interest for a specific research question, which may be important for projects that may have many images to sort through or prioritize. The models produced from ground control points are also sorted by user, adding an element of quality control for the project owner. Lastly, Sunspot includes a machine-to-machine (M2M) interface with EarthExplorer for the direct retrieval of scanned images, removing bandwidth-intensive downloading and uploading of imagery during the processing stages. All these capabilities were used in the preparation of the Haryana dataset, which allowed the MAHSA team, based on multiple continents and working on different computers, to coordinate and contribute collectively to the project.

2.2.3 Hexagon KH-9 imagery

For the Haryana methodology, we selected six forward-camera frames from Hexagon Mission 1211 (D3C1211-100048F018-F023), obtained through USGS EarthExplorer and processed in Sunspot [53], to provide the mid-century epoch. Each is a panoramic strip of roughly 330 x 15 km at sub-meter nominal resolution.

The Hexagon missions span the period between 1971 and 1984 and were collected from 19 reconnaissance satellites that were flown in low earth orbit [54,55]. Both panoramic camera and mapping camera imagery are available through EarthExplorer. The panoramic camera imagery used in this analysis included two 60-inch focal cameras, which could be up to 580 x ~24 km in footprint [56]. The lure of the panoramic camera—its resolution of 1-3 feet, its stereoscopic capabilities, and its repeat, dense coverage that is more cloud-free than Corona—provides an advantage over Corona imagery or becomes another historical source if both Corona and Hexagon are considered together [37,40,56]. Hexagon Mission 1211, which flew over a cloud-free Haryana on December 11, 1975, falls almost exactly in the middle of this period between the SoI maps and modern imagery and therefore supplies an additional, important observation for heritage monitoring. As with Corona, Hexagon is low-cost, at $30 for the initial scan and free thereafter.

An emerging set of publications focuses on Hexagon panoramic camera imagery and its potential for scientific research (e.g., [21,25,36–40]). Case studies that have been more successful at Hexagon modeling have generally either done so at small areas (e.g., [38,57,58]), involved time-intensive, high numbers of ground control points, or required command-line knowledge (e.g., [31,34,59]). For Hexagon imagery to be incorporated broadly into heritage management, especially at regional or country-level scales, approaches need to accommodate full imagery analysis with minimal processing time and more novice users. Sunspot fits well into this niche and needs.

2.2.4 ESA WorldCover

To establish a contemporary baseline for long-term heritage monitoring analysis, researchers increasingly rely on standardized, global datasets that facilitate integration with historical remote sensing archives [60,61]. Using a shared land-cover classification, such as ESA WorldCover, makes modern multispectral data directly comparable with historical reconstructions from panchromatic CORONA and HEXAGON imagery [24]. This comparability is critical for long-term analysis, as it allows for the precise quantification of multi-decadal transformations. Using established global products also makes deep-learning models more transferable and avoids the substantial computational effort required to classify contemporary sensor data manually [61,62].

Within this methodological framework, the modern epoch is provided by the ESA WorldCover 10 m land-cover product for 2021 [63], chosen because its classes map cleanly onto the historical

ones and because it spares the project from having to classify modern imagery at all. The product and its accuracy are described in Section 2.3.3.

2.2.5 MAHSA Arches database and site selection

Site locations and mapped perimeters were drawn from the MAHSA database, which provides both a modern or recent anchor point for the heritage locations and, where the extent has been digitized, a site perimeter. These provide geometry around which the analysis zones are built (Section 2.4).

We selected 12 sites for quantitative analysis (Figure 1), spanning the region's chronological range and its environmental gradient, from the canal-irrigated east to the semi-arid west. These include the sites of Rakhigarhi, Masudpur I, Masudpur VII, Lohari Ragho I, Bhirrana, Kunal, Burj Pirwala, Bhuna, Agroha, Sirsa, Ankaar, and Bahola. The sites represent the range of settlement sizes in the region, including, for example, one of the largest Indus cities (Rakhigarhi), as well as excavated Early Harappan villages (Kunal, Bhirrana), Early Historic and Medieval towns (Agroha, Sirsa), and small rural mounds (the Masudpur-Lohari Ragho cluster) [7,10,12,48–51]. They also provide a test for site visibility in imagery. Working with this smaller set allowed us to refine a workflow that can later be applied across the full database of sites and to establish what Hexagon images can enable at multiple scales, be it site-specific or across the state and across a large, varied landscape.

2.3 Data processing

2.3.1 SoI maps, 1914-1915

The SoI sheets use a consistent color-coded symbology: settlements are represented in solid red, water in blue, and cropland by a strong yellow wash and grassland by a pale, broken white one. Map legends, however, may appear differently per map sheet depending on what is depicted. This standardized convention permits rule-based feature extraction from the map’s printed color-coded symbology, rather than inference of land cover from a satellite image surface. Comparable approaches have been used to extract settlement features from the same map series [12,64].

Map land cover was classified using color thresholds in HSV color space. Rules were applied in a fixed priority sequence: built-up, water, vegetation, cultivated land, and grassland, so that pixels meeting more than one criterion were assigned consistently. Morphological opening was then applied to remove narrow linear features, including red road casings, while retaining larger built-up blocks.

Reference class	Pixels	Built	Water	Crop land	Grassl and	Vegeta tion	No fill	Recall (producer's accuracy)	Precision (user's accuracy)
Built-up	19,889	**91.9**	0.0	0.0	0.0	0.0	8.1	**91.9**	**69.3**
Water	39,442	18.1	**53.5**	0.1	0.2	0.0	28.1	**53.5**	**99.6**
Cropland	277,469	0.0	0.0	**97.0**	2.0	0.2	0.7	**97.0**	**96.0**
Grassland	102,883	0.1	0.1	2.6	**88.9**	0.1	8.3	**88.9**	**79.7**
Vegetation	22,019	0.2	0.0	0.4	0.0	**93.5**	5.9	**93.5**	**96.7**
No areal fill	30,027	2.3	0.0	28.1	58.7	0.0	**10.8**	**10.8**	**11.7**

Table 1. Land-cover classification accuracy for the SoI sheets. Every pixel is scored and calibrated on the sheets that exclude its own class. Producer accuracy is the proportion of a reference class correctly classified; user accuracy is the proportion of pixels assigned to a class that belong to it. All figures are percentages except the pixel counts. Macro accuracy 72.6%, pixel accuracy 86.2%.

Sheets were mosaicked at their native resolution, and the classification was reprojected to UTM Zone 43N. Nearest-neighbor resampling was used at both stages. Only extent-clipped sheets were used, as the printed margin, title block, and legend would otherwise be classified as map content, and overlaps between adjacent sheets were resolved on a first sheet-wins basis (unless proved problematic). No normalization was applied between sheets: per-sheet saturation normalization was tested as a means of absorbing differences in printing and scanning between series but not used, because it widened the cross-sheet spread for five of the six classes. Morphological opening used 3x3 four-connected structuring elements, applied for three iterations for built-up, cropland, and grassland; two for vegetation; and one for water. By applying several iterations of morphological opening, features narrower than 2n + 1 pixels are removed. In the end, this left us with six different classes: built-up, water, vegetation, cropland, grassland, and ground carrying no area-fill color.

Thresholds for the classes were calibrated using 51 manually hand-digitized reference polygons distributed across five sheets representing the different map numbers. Transferability between sheets was assessed using leave-one-sheet-out cross-validation. The classifications assigned for this analysis were overall quite good: built-up at 91.9% producer accuracy, cropland at 97.0%, grassland at 88.9%, and vegetation at 93.5% (Table 1). Two classes are weak. Water reaches only 53.5% of producer accuracy, though at 99.6% user accuracy is almost never claimed falsely, and it occupies under 0.3% of the analysis area. Ground carrying no area-fill color is recovered at 10.8%, with 58.7% of it absorbed into grassland; this class has no positive rule and is whatever falls through the others, so widening the thematic classes necessarily costs it. Neither weakness

affects built-up fractions, since both classes sit inside the observed denominator. Our initial calibration based on a single sheet performed well internally but transferred poorly to other sheets because the saturation ranges used for cropland and grassland left an unclassified gap. This led us to expand the sample data to more sheets.

Two limitations originated from the source material itself. First, bare ground and grassland overlap in saturation within individual sheets and cannot be separated reliably using the rule-based color segmentation alone. This potentially affected estimates of cultivated extent but not estimates of built-up land. Pixels without an area fill were therefore retained in the denominator for built-up fractions but excluded from comparisons of cultivated land. Ground at the very edge of a sheet carries no area-fill color and is classified as no fill; unfortunately, sometimes a narrow band along each sheet boundary is recorded as grassland. Measured, however, against the sheet index, this contributes under 0.2 percentage points to any site's grassland fraction and is therefore negligible. Second, the maps depict the archaeological mounds themselves at Rakhigarhi, including annotated mound heights, providing an independent documentary record of site conditions in the early twentieth century.

2.3.2 Hexagon KH-9, 1975

Each Hexagon frame was georeferenced in Sunspot with 21-27 GCPs placed on stable features (e.g. the corners of buildings in the center of settlements, canal junctions, and road and railway crossings) spread along the full length of the strip and across both edges. That is one to two orders of magnitude fewer points than conventional polynomial georeferencing of panoramic film would need at this scale. Sunspot's machine-to-machine (M2M) link with EarthExplorer also means that no raw film scan (8-25 GB per frame) needed downloading or uploading to the users' machines, a real consideration wherever bandwidth is limited. Only the output files, like the orthorectified imagery, had to be downloaded in order to visualize and further analyze the results in QGIS.

Accuracy measures of our outputs were estimated by leave-one-out cross-validation (LOO CV) at the GCPs. The six frames came out at median accuracies of 181-496 m (Table 2): a two-order-of-magnitude improvement on the raw film geometry, from about two dozen GCPs per frame, and sufficient for grid-cell statistics on ~14 km cells (with the site-anchored design adding further protection).

KH-9 Frame	GCPs used / placed	LOO CV median (m)
F018	20 / 21	358
F019	25 / 26	181
F020	20 / 23	496
F021	22 / 27	216
F022	24 / 24	263
F023	24 / 27	212

Table 2. Georeferencing accuracy of the six Hexagon Mission 1211 frames (leave-one-out cross-validation at ground control points).

One important limitation and potential difficulty in working with large Hexagon imagery is that films of this age sometimes bring non-uniform deformation (film shrinkage) that a georeferencing model has difficulty transforming. To address this, we implemented an additional refinement step directly following the model export from Sunspot for each frame. First, a baseline orthomosaic was generated from the exported model. The geographic coordinates of all ground control points (GCPs) were then retrieved from this initial export to compute frame-specific polynomial corrections, which were subsequently fitted and executed as a final image warping transformation.

To classify the 1975 land cover, we utilized HexaLCSeg [61], an established benchmark comprising KH-9 panoramic imagery paired with land-cover categories standardized to ESA WorldCover classes (derived from 10,000 image patches across Türkiye and Bulgaria) alongside pre-trained deep-learning models (https://github.com/RSandAI/HexaLCSeg). We applied its published U-Net++ model to the Haryana KH-9 frames. Having this benchmark available addresses a common constraint for heritage research: getting quantitative land cover out of historical imagery without a labor-intensive labeling process to create training data. We applied the best-performing model variant without further fine-tuning and similarly prepared the imagery, matching the resolution and tiling on which the model was trained. Reusing a published model in a new region, however, often has limitations in terms of its transferability. The model's reported accuracy, 88.5%, belongs to the landscape it was trained and tested on, in this case Türkiye and Bulgaria, and does not necessarily transfer particularly well to a different region. We therefore considered that the approach also would work on the Hexagon film of different regions, but did not presume that we would reach the same accuracy in our region. Given this, we set out our specific model alterations below.

Two slight adaptations to the source KH-9 image inputs were necessary. First, pixel values of zero on KH-9 film are valid image information, since deep shadow and dark roofing in village cores scan to zero, so treating zero as no data, as is conventional, deletes parts of the built-up classification that was required for our analysis. We instead shifted valid zeros that were within the image frame, to a value of one; so that zero does align with the no data value. Second, as a consequence of the former, for the geometric test needed to determine inner and outer areas of the

frame, we had to distinguish genuinely dark areas from the dark serrated edge of the film itself. Where the frame border is admitted into the valid area, the model reads it as tree cover, and in our initial implementation this accounted for between a quarter and a half of all vegetation recorded in the outermost analysis zone. Validity, therefore, was now determined by filling enclosed dark areas rather than by morphological closing, which retains interior dark pixels while excluding the border. Built-up fractions are essentially unchanged by this correction; the 1975 vegetation figures are not and are reported after it. Sites falling in the narrow gaps between adjacent strips were classified from the overlapping neighboring frame, selected automatically by coverage.

Two patterns follow from the difference between the training landscape and Haryana rather than from a fault in the model. In densely settled areas it can read urban fabric as vegetation, an error most common at the largest settlements; preparing the imagery as the model's authors did reduces this substantially, and at Sirsa, a district capital in 1975, the town core then resolves as a single built-up mass rather than as scattered fragments. In open country the opposite occurs, and thin linear features such as roads and canal banks can be read as built-up. Because these run throughout the landscape rather than clustering near the sites, built-up land in 1975 is somewhat less concentrated around the sites than in either the map or the modern epoch. No cross-epoch correction is applied: each epoch rests on its own source.

2.3.3 ESA WorldCover, 2021

The modern epoch is provided by ESA WorldCover v200, a global land-cover product at 10 m resolution derived from combined Sentinel-1 radar and Sentinel-2 optical imagery for the calendar year 2021 [63]. It distinguishes eleven classes and reports a global overall accuracy of 76.7%. Two properties make it well suited to this analysis. Its class definitions are the scheme on which the HexaLCSeg labels were modeled, so the 1975 and 2021 epochs harmonize without arbitrary reassignment, and it is an already validated product, which removes the need to classify modern imagery ourselves. Its limitations matter equally: at 10 m, individual rural structures fall below the minimum mapping unit, so potentially built-up land in sparsely settled areas is underrepresented relative to what the sub-meter Hexagon film can resolve, like individual buildings [58]. Modern built-up fractions should therefore be read as somewhat conservative in rural settings.

2.3.4 Harmonization for long-term analysis

All three classifications were mapped to a common six-class scheme (built-up, cropland, grassland, natural or vegetated, bare or other, and water), with a seventh category, observed but unclassified, applying only to the SoI map epoch (Figure 6). The three sources divide non-cultivated ground differently, and the difference is instructive. The Survey of India records grassland as a land-use judgment and makes less use of a separate natural-vegetation category, while the two later epochs separate herbaceous from woody cover on appearance. Across the twelve sites the 1975 and 2021 classifications agree well on how much ground is vegetated but not cultivated once their grassland and natural-vegetation classes are taken together ($r = 0.68$, $n = 12$), but neither corresponds to the map epoch's grassland. We therefore report grassland for the map epoch only, and compare the latter two on the combined category rather than on the split. For each site and epoch, we determined two things, always within that epoch's own anchored geometry: land-cover fractions in each of four concentric zones extending to 0-1, 1-2, 2-5, and 5-10 km from the site anchor; and the distance from the anchor to the nearest built-up land. Areas were measured in a projected system (UTM 43N).

Harmonised class	Survey of India, 1914–15	KH-9 Hexagon, 1975 (HexaLCSeg)	ESA WorldCover, 2021
Built-up	built-up (red fill)	built-up	built-up
Cropland	fields (strong w ash)	crop, permanent crop	cropland
Grassland	pale w ash ("Grass lands")	grass	grassland
Natural / vegetated	vegetation tint	tree, shrub	tree cover, shrubland
Bare / other	—	bare	bare / sparse
Water	w ater	w ater	permanent w ater
Observed, unclassified	no areal colour	—	—

Figure 6. Harmonization of the three source classifications onto the a common scheme. Each row gives one harmonized class and the native classes that enter it from the Survey of India sheets, the HexaLCSeg labeling of the KH-9 frames, and ESA WorldCover. Native classes not listed are the products' own no-data values and are carried through as no data. Two asymmetries should be mentioned: "observed, unclassified" is surveyed ground carrying no area-fill color and exists only for the map epoch, and "bare or other" has no Survey of India source. Grassland is reported for the map epoch alone: the Survey of India legend carries "Grasslands" as a land-use category, listed separately from the limit of cultivation, while the two later epochs classify herbaceous cover on appearance as grassland.

2.4 Analysis

All change statistics around each site are summarized to four buffer zones, extending to 0-1, 1-2, 2-5, and 5-10 km from each site point. The zones are defined on the site anchors and are therefore independent of map-sheet geometry, which matters because the three epochs are drawn from sources whose extents have nothing in common. Reporting by distance rather than within a single cell also allows the analysis to describe not only how much development surrounds a site but also how that development is distributed with respect to it.

The mapped site perimeter is subtracted from every zone (based on the modern estimated site extent) so that a site's own area is never counted as its land cover, and encroachment always refers to development on ground that was not already the site. Where the zones of neighboring sites overlap, the overlap is retained for each site, but the area of every other site perimeter is likewise excluded.

Comparing three epochs raised an obvious problem: each dataset carries its own positional error, a few hundred meters for the corrected Hexagon frames and similar or larger for the georeferenced SoI sheets. Overlaying them pixel by pixel would turn those errors into false change, and a standard change detection was therefore not possible. Therefore, instead of aiming for the smallest possible pixel accuracy, which is unrealistic for panoramic film at this scale and age, we fixed the analysis to the sites themselves by period. For every site the same ground feature was marked twice, once in modern reference imagery and once in the corrected Hexagon imagery. The offsets between these paired points ranged from 179 to 673 m, in line with the georeferencing accuracy reported above. Each site's zones were then constructed from the anchor appropriate to that epoch so that the site sits in the same position relative to its zones in every geometry. Every epoch is classified and measured in its own geometry, and only zone-level fractions and anchor-centered distances, statistics that shrug off residual misregistration, are compared across epochs. The SoI sheets proved close enough to the modern geometry to share the modern anchor.

The design was tested directly by displacing each anchor by the full magnitude of its registration error and recomputing. Built-up fractions change by a median of 1.21 percentage points in the innermost zone and 0.08 percentage points in the outermost, an order of magnitude smaller than the differences the analysis reports.

2.5 Statistics and reproducibility

The land-cover fractions and distance-to-nearest-built metrics reported in this study are descriptive quantities, computed for each site, zone, and epoch; no inferential statistical tests were applied to them. Classification accuracy for the SoI sheets was assessed by leave-one-sheet-out cross-validation (Section 2.3.1; Table 1, which gives the reference pixels per class) and the RMSE values provided in Sunspot, and georeferencing accuracy for the Hexagon frames by leave-one-out cross-validation at the ground control points (Section 2.3.2; Table 2, which gives the GCPs used per frame).

3. Results

3.1 Quantifying a century of landscape risk

Incorporating Hexagon imagery into the comparison between the SoI maps and modern imagery offered a substantially improved analysis of landscape change in Haryana and refined the methodology on which further sites in the region may later be identified. In particular, the higher resolution of Hexagon provides a much better-quality assessment capability than was possible before. While there are alternative data sources for the 1970s available (e.g., GHS-BUILT-S [65], Landsat 1-5 MSS [66]), none of these can provide the spatial level of detail that Hexagon imagery can provide, and digitizing these at the scale conducted in this research would not have been possible without the Sunspot platform.

Before, the Haryana timeline of landscape change in this region came in two slices: SoI and present day. Adding the Hexagon epoch separates two processes that a map-to-satellite comparison necessarily conflates. The cultivated landscape around these sites was not yet closed in 1915: cultivation covered 64% of observed ground on the SoI sheets, with a further 20% classed as grass and uncultivated land, and by 1975 the Hexagon classification shows a cropland matrix of 84%, reaching 92% in 2021. Because a drawn map and a photograph record ground in different ways, a measure less sensitive to how each source draws the cultivated boundary is the land that is not cropland under any reading, and that falls from 35% of observed ground in 1915 to 14% in 1975 and 4% in 2021. The Green Revolution therefore arrived at a landscape still being enclosed and consolidated, a field system that the preceding half-century had been assembling.

That closure was not uniform with respect to the sites. In 1915 the ground within one kilometer of the mounds was the least cultivated in the landscape, at 47% cropland against 65% between five and ten kilometers, and carried the most grass, at 30% against 19%. These mounds stood in grassland rather than in fields: elevated, cleared, but not cultivated.

The main change is in built-up land, and this seems to come later in the century. Within one kilometer of the sites, built-up cover rises from 3.2% in 1915 to 7.8% in 1975 and reaches 14.9% by 2021. Annualized, because the intervals differ in length, growth runs 0.08 percentage points per year to 1975 and 0.15 after it, so the pace roughly doubles once the field system has closed. The Hexagon epoch places the inflection between the two: the field system had closed by 1975, while settlement growth was still largely ahead. Its spatial pattern changed as well. In 1915 built-up land was already strongly related to distance from the sites, standing at 3.2% within one kilometer against 0.8% between five and ten kilometers, a ratio of more than four to one. By 2021, the innermost kilometer carried 14.9% against 2.9% in the outermost zone, a ratio of five to one. The relative concentration has therefore held roughly constant across the century, at four to five times the background density, while the absolute gap has widened from 2.4 to 12.0 percentage points. Settlement has always favored these places, and for a reason that is not accidental: mounds are elevated, dry, already cleared, and, on the 1915 evidence, held as common rather than cultivated ground, which makes them preferred building land. What has changed is not the preference but its volume. The pressure on these sites is therefore structural rather than incidental and will not resolve of its own accord, so that risk to archaeological heritage is rising faster than regional development figures alone would suggest.

Read against the demographic record introduced in Section 2.1 (Figure 3), the sequence is informative. Cultivation covered 64% of observed ground in 1914-15 and had reached 84% by 1975, after which it continued to rise only slowly; built-up land within a kilometer of the sites, by contrast, more than doubled before 1975 and then doubled again. The earlier phase of demographic growth was therefore absorbed by the extension of cultivation onto ground that had not carried it, and the later phase by settlement.

The two stages threaten heritage differently: the first levels mounds within an expanding field system, the successive phase buries or seals them under settlement. Knowing at which stage a site has passed through informs the remote assessment of its condition and the planning of ground truthing, because a site leveled and plowed before being built on is likely to retain less of its original context than one built on directly. Urban and agricultural development are unavoidable necessities, since people need to eat and be housed, but informed decisions about how to avoid further damage can support compromises that balance development against the protection of invaluable archaeological heritage.

3.2 Differentiated risk at the site level

The twelve sites selected for additional consideration allow for localized risk assessments of the sites of highest concern. At the scale of these twelve individual sites, encroachment is strongly differentiated. The varying local historical trajectories are represented in Table 3 and Figure 7 and would impact policy recommendations.

At one extreme, Sirsa, an Early Historic and Medieval urban center in the semi-arid west of the state, has become urban-engulfed: 75% of the land within 1 km of the site was built up by 2021, against 31% in 1975 and 18% on the SoI sheets. The site sits within a district capital that has grown outward around it, and the sequence shows that the decisive interval was the half-century after 1975 rather than the half-century before. Bhuna, a medieval mound now enclosed by a substantial village, follows the same trajectory one step behind, rising from 1.5% in 1915 to 12% in 1975 and 38% in 2021. The comparison between the two is instructive: at Sirsa the surrounding land is effectively committed, whereas at Bhuna the process is still underway and the outcome is not yet fixed.

At Rakhigarhi, among the largest Indus cities, the modern village is expanding directly over the mound complex: built-up cover within 1 km rises from 4% on the SoI sheets to 19% in 1975 and 24% in 2021, and built-up land has adjoined the site in every epoch. Most of that growth predates the 1975 film, whereas at Sirsa most of it came afterwards; comparing the SoI sheets with present-day land cover alone could not distinguish these two sequences. Bhirrana shows the same accelerating profile at lower levels, while at Agroha almost all of the growth falls after 1975. At the other extreme, the rural Harappan cluster of Masudpur I, Masudpur VII, and Lohari Ragho I registers essentially no direct built-up encroachment in any epoch, never exceeding 1% within 1 km. Their main risk is agricultural: all three now sit inside a continuous cropland matrix that had already reached them by 1975. Ankaar, Bahola, Kunal, and Burj Pirwala fall in between.

Site	1914–15	1975	2021	Pressure class
Sirsa	17.7	31.1	75.1	Urban-engulfed
Bhuna	1.5	11.8	37.8	Heavy encroachment
Rakhigarhi	3.9	19.5	23.7	Heavy encroachment
Bhirrana	9.2	11.7	20.1	Heavy encroachment
Agroha	1.1	1.5	14.0	Accelerating
Ankaar	0.4	6.9	5.6	Moderate
Kunal	0.8	2.2	4.6	Moderate
Burj Pirwala	0.8	2.9	4.4	Moderate
Bahola	1.9	6.0	2.7	Low
Masudpur I	0.1	0.6	0.1	Agricultural only
Masudpur VII	0.7	0.3	0.1	Agricultural only
Lohari Ragho I	0.1	0.4	0.0	Agricultural only

Table 3. Built-up land by epoch and site, as a percentage of observed ground within 1 km of the site anchor. The mapped site perimeter is excluded, so the figure measures development beside the site rather than on it. Each epoch is classified from its own source with no cross-epoch correction. At five of the twelve sites the 1975 figure exceeds the 2021 one; all five are small rural sites where the 10 m WorldCover product cannot resolve dispersed structures that the sub-meter film records.

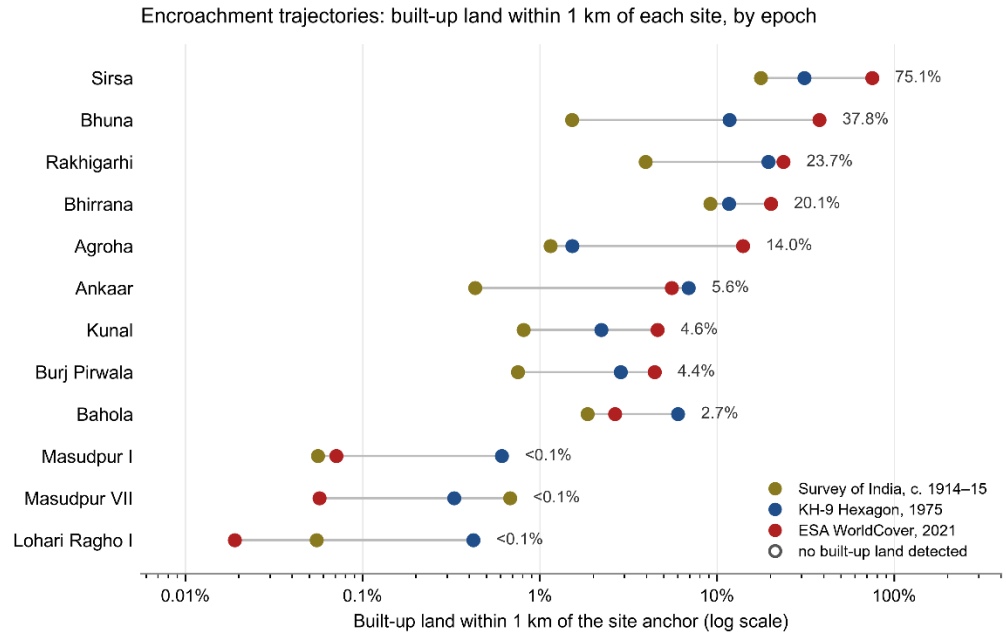


Figure 7. Encroachment trajectories at the twelve study sites: built-up land within 1 km of each site anchor, by epoch (SoI, 1914-1915; Hexagon, December 1975; ESA WorldCover, 2021). Sites are ordered by 2021 pressure; the horizontal axis is logarithmic. Each row traces a century of change, from the urbanizing trajectories of Sirsa and Bhuna to the near-zero values of the Masudpur-Lohari Ragho cluster at every epoch.

3.3 Contributions of Hexagon Imagery to the Haryana Case Study

The three-epoch imagery approach (SoI, Hexagon, and ESA WorldCover) demonstrates the value of Hexagon imagery in large-scale and long-term heritage management. The Hexagon layer repeatedly records conditions that neither bounding epoch does. The resolution is excellent at a site level, both for larger urban centers and rural areas, and to have this across a large geographic area can be tremendously helpful. It also furthers our understanding of the timing and impact of the Green Revolution and can therefore help refine the rate of change, both broadly and at the site level. For example, the village near Rakhigarhi had already advanced over the mounds by 1975. However, certain open areas have since been built over today. Because those areas can be dated to the interval after 1975, they can be distinguished from ground lost before the mound complex was ever formally recorded, and the parts of the site that remain open are identifiable rather than assumed. Around the Masudpur sites, it shows that the enclosing field system was complete by 1975, dating the leveling risk to the first interval and indicating that the dominant present-day threat there is continued cultivation rather than settlement growth. At Sirsa, Hexagon captures the last moment when the site's surroundings were still visible as a distinct historic town core, which constrains the original extent of the settlement in a way that no modern image can and provides a basis for targeting any future salvage work. For heritage risk management, the metrics of all these examples translate directly into monitoring priorities (Table 3). Sites in the heavy encroachment class call for plot-level monitoring of the built-up frontier, for which the Hexagon-modern image pair is the natural baseline. The agricultural-only cluster calls for monitoring mound integrity under continued cultivation rather than settlement growth. Refined encroachment trajectories, enabled by Hexagon, fit into heritage management that can be deployed throughout the state. Figure 8 shows the full three-epoch record for three contrasting sites, with the analysis zones overlaid.

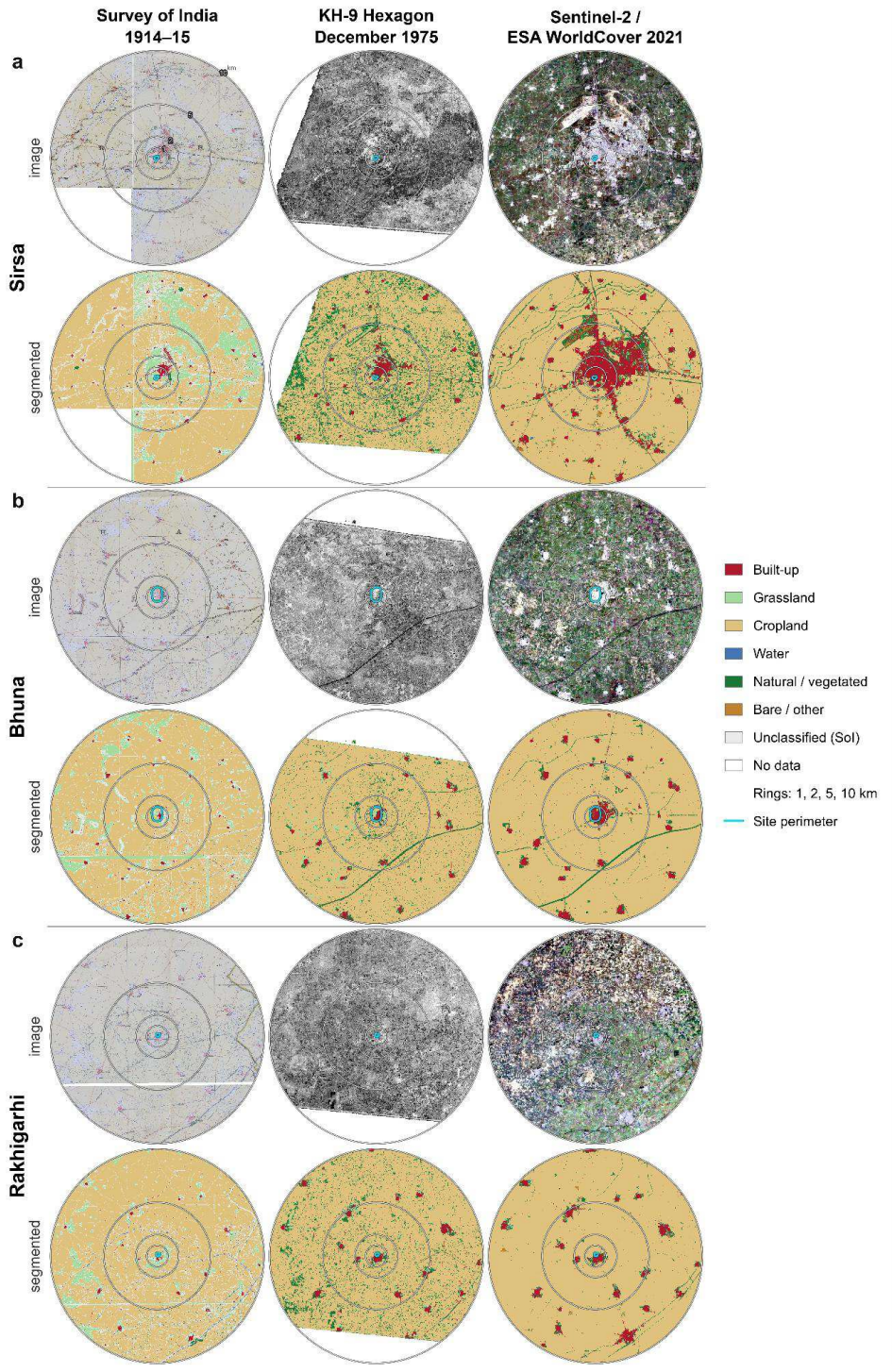


Figure 8. The three-epoch record for three contrasting sites (a. Sirsa, b. Bhuna, c. Rakhigarhi). For each epoch the source data are shown beside the harmonized classification over the 10 km analysis area, with the ring zones at 1, 2, 5, and 10 km and the mapped site perimeter in cyan. 2021 imagery: modified Copernicus Sentinel-2 data 2021, processed by ESA. Our web map can be used to view a more detailed overview of the segmentation and output results: https://arcg.is/1rmKfL5.

4. Discussion

In South Asia, rapid agricultural and urban development pose threats to many archaeological sites. Many of these sites are being leveled and encroached upon at an alarming rate [8]. Our Haryana case study demonstrates how historical maps and imagery, alongside legacy data, can help distinguish the timing and character of these pressures. Adding the 1975 Hexagon epoch reveals differences in the trajectories of agricultural expansion and settlement growth that would be misidentified by a comparison of the historical maps and contemporary land-cover data alone. These differences provide a basis for more targeted long-term heritage monitoring and planning decisions.

Hexagon's contribution lies in providing a high-resolution historical observation between the SoI maps and contemporary land-cover data. In many locations in the Global South, it may provide coverage for periods poorly represented by maps, aerial photographs, or other satellite imagery. CAST's updates to the Sunspot georeferencing tool make this record more accessible for investigating the pace of land use and land cover change. In Haryana, the ground around the study sites was a mosaic of fields and mapped grassland when the SoI sheets were surveyed and a largely continuous field system by 1975. This brackets a period of agricultural transformation that includes the Green Revolution, although the three observations cannot isolate its contribution from earlier changes. Settlement growth followed different trajectories: much of the measured increase around Rakhigarhi preceded 1975, whereas the larger increases at Sirsa and Agroha occurred subsequently. Around the Masudpur sites, extensive cultivation coexists with little mapped built-up development. These contrasts show why monitoring priorities need to distinguish settlement expansion from continued agricultural pressure.

Applying this approach across larger areas also depends on making historical imagery practical to process. The Haryana case study demonstrates how Sunspot can support landscape-scale analysis by multi-user teams, including those with limited internet bandwidth. The main advantages for the MAHSA team were threefold: 1) much of the processing took place on the Sunspot platform, avoiding the need to download and upload large raw film scans; 2) the six frames were modeled using 20–25 ground control points per frame, and 3) shared projects enabled teams working in different locations, including heritage practitioners with limited knowledge of geospatial methods, to incorporate Hexagon imagery into their workflows. These practical benefits are particularly relevant to MAHSA, whose teams work across countries with differing access to computing resources and internet bandwidth. Shared processing and collaborative workflows support the extension of this approach from individual case studies to larger programs of heritage documentation and monitoring.

The value of this combined record also reflects the purposes for which its sources were originally created. Hexagon was born from geopolitical competition, just as the SoI maps were created to support British Imperial control in South Asia [8,12,15]. As with other archives of aerial

photography and satellite imagery (e.g., [20,67,68]), these sources have since found uses in conservation that were not foreseen at the time of recording (e.g., [69]). The distinction between cultivated ground and grassland, important to revenue assessment, provides evidence of the earlier landscape around the mounds through the maps' printed symbology. Reconnaissance film provides detailed observations of subsequent conditions during a period for which dedicated heritage records are limited. Both archives recorded land in order to control it; together, they now help trace the transformation of those landscapes and heritage sites.

Beyond the Haryana case study, Hexagon offers considerable potential for other forms of heritage and landscape monitoring. Archaeologists have demonstrated the range of surface features visible in the imagery, including individual houses, courtyard complexes, tells, alleyways, canals, and pastoralist campsites [38,40]. Machine learning has supported feature extraction from historical maps (e.g., [64]) and mound detection using recent multisensor, multitemporal satellite data [71]. Here, we tested the existing HexaLCSeg benchmark dataset [61] for automated land-cover classification of Hexagon imagery in Haryana, applying it to heritage monitoring. As with other historical imagery (e.g., [8,70]), comparisons with modern records can support the assessment of landscape change. Wider applications include investigating changes in ecosystem extent and structure and human impacts such as forest loss [21]. Other historical imagery studies illustrate the relevance of this longer perspective to coastal erosion [72], while Hexagon has also been used to quantify urban expansion [58]. Such records can help investigate past heritage loss associated with settlement growth, changing waterways, and agricultural development, as well as identify pressures on surviving sites [19].

These applications nevertheless depend on the coverage, resolution, and classification characteristics of each source. The site-anchored approach reduces sensitivity to positional offsets, but it does not remove differences between map symbology, Hexagon classification, and WorldCover. These differences are particularly relevant to small changes in rural built-up cover, where classification errors and differences in spatial resolution may affect the apparent trajectory. The twelve selected sites demonstrate contrasting development histories rather than a representative estimate of their prevalence across Haryana. Surrounding land-cover change provides an indicator of development pressure; field verification remains necessary to establish the extent and character of archaeological damage. These considerations make the workflow most useful for identifying priorities for closer investigation and monitoring.

Bringing together SoI maps, Hexagon imagery, and contemporary land-cover data provides a century-long perspective on development pressure around archaeological sites in Haryana. The intermediate Hexagon epoch helps distinguish the timing and character of agricultural expansion and settlement growth, revealing different trajectories that matter for how individual sites are investigated and monitored. Sunspot makes this historical imagery more practical to incorporate into collaborative workflows, providing a basis for extending the approach to other regions where suitable coverage is available. In Haryana, as well as in its larger research area, MAHSA's existing

partnerships and fieldwork offer a route for using these findings to prioritize verification and monitoring. Together, these contributions show how historical landscape records can inform present-day heritage management: understanding how development pressure has accumulated around a site provides a stronger basis for deciding where closer investigation and conservation attention are needed.

5. Data Availability

ESA WorldCover v200 is publicly available [63]. Raw Hexagon film scans are available from the USGS EarthExplorer archive, and the HexaLCSeg benchmark and model weights from Sertel et al. [61]. The segmentation results for all three epochs can be viewed as an online web map at https://arcg.is/1rmKfL5 . A subset of the segmentation results on the Survey of India maps is available at Zenodo (https://doi.org/10.5281/zenodo.23407161). Zone-level land-cover fractions and distance-to-built metrics are available from the corresponding author on request. Site locations and perimeters from the MAHSA database (databasemahsa.org), the orthorectified Hexagon frames, and the harmonized land-cover rasters are not publicly available, to protect vulnerable archaeological sites, but are available from the corresponding author on reasonable request.

6. Code Availability

The code used for the SoI map classification is available at Zenodo (https://doi.org/10.5281/zenodo.23407161). The custom code used for the KH-9 preprocessing and refinement and for the zone-level analysis is publicly available and is available to the Editors and referees, and to qualified researchers on reasonable request from the corresponding author. Sunspot is available at https://sunspot.cast.uark.edu.

7. Acknowledgements

The authors would like to acknowledge the ongoing work and contributions of our respective teams. The documentation work of MAHSA has been generously supported by Arcadia, under grant numbers 4180 and 5143. Work on Sunspot has been generously funded over many years by a number of sponsors, including ACLS, NASA, NEH, and USGS. Funding for the updated version of Sunspot was provided by USGS. The workshop that brought together the University of Arkansas and University of Cambridge teams to discuss Hexagon imagery and Sunspot in June 2024, and initiated the discussions, was supported by the Leica Geosystems Endowment at the Center for Advanced Spatial Technologies and hosted by the MAHSA team as well as its sister project, Mapping Africa's Endangered Archaeological Sites and Monuments (MAEASaM), Project Manager Stefania Merlo, PI Paul Lane, also funded by Arcadia [4179 and 5144]. The funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript. We would like to extend our gratitude to the workshop participants for their contributions during and after the event: Junaid Abdul Jabbar, Philipp Barthelme, Rosie Campbell, Afifa Sana Khan, Kuili Suganya, Jack Tomaney, Azadeh Vafadari, Dana AlSalamin, Akinbowale Akintayo, and others who joined online. We would also like to extend our thanks to the other current and former CAST staff that have contributed to Sunspot and the Corona Atlas

(https://cast.uark.edu/directory/). We also thank Max Fancourt and other colleagues from the *Barcelona Supercomputing Center* (BSC) for their helpful feedback during the preparation of this work. The MAHSA project is building on foundations created by the *Land, Water and Settlement* (2007-2015) and *TwoRains* (2015-2021) projects, which were collaborations between the University of Cambridge and Banaras Hindu University, under the direction of Cameron Petrie and R.N. Singh. In India the MAHSA project continues its collaboration with Banaras Hindu University, and is also collaborating with colleagues from the University of Kerala, IIT Gandhinagar and NIAS in Bangalore (https://www.mahsa.arch.cam.ac.uk/about/project-team).

8. Author Contributions

Conceptualization: CK, PJG, RR; Formal analysis: PJG; Funding Acquisition: SW, JC, RR, CAP; Investigation: PJG, CK; Methodology: PJG, CK; Project administration: CK, RR; Resources: SW, HH, CAP; Software: SW, HH, PJG; Supervision: CK, RR, CAP; Validation: PJG, VV, MS, JT; Visualization: PJG, JT; Writing – original draft: PJG, CK, RR; Writing – review and editing: CK, PJG, RR, SW, JC, MS, JT, VV, CAP. All authors read and approved the final manuscript.

9. Competing Interests

All authors declare no financial or non-financial competing interests.

10. References

[1] ICCROM. *Annual Report 2016*. International Centre for the Study of the Preservation and Restoration of Cultural Property, Rome. https://www.iccrom.org/report/annual-report-2016-0 (2016).

[2] Historic England. *Heritage at Risk: Findings*. https://historicengland.org.uk/listing/heritage-at-risk/findings/ (2025).

[3] Rayne, L., Rouhani, B. & Makovics, J. L. Remote sensing and disaster risk management for cultural heritage. in *Managing Disaster Risks to Cultural Heritage* (eds Rouhani, B. & Romao, X.) https://doi.org/10.4324/9781003263647 (Routledge, 2023).

[4] Sardella, A. et al. Risk mapping for the sustainable protection of cultural heritage in extreme changing environments. *Atmosphere* **11,** 700. https://doi.org/10.3390/atmos11070700 (2020).

[5] Petrie, C. A. South Asia. in *The Oxford Handbook of Cities in World History* (ed. Clark, P.) 83–104 (Oxford Univ. Press, 2013).

[6] Green, A. S. & Petrie, C. A. Landscapes of urbanization and de-urbanization: a large-scale approach to investigating the Indus Civilization's settlement distributions in Northwest India. *J. Field Archaeol.* **43,** 284–299. https://doi.org/10.1080/00934690.2018.1464332 (2018).

[7] Petrie, C. A. et al. Mapping archaeology while mapping an empire: using historical maps to reconstruct ancient settlement landscapes in modern India and Pakistan. *Geosciences* **9,** 11. https://doi.org/10.3390/geosciences9010011 (2018).

[8] Petrie, C. A. et al. Hidden in plain sight: the unrecognized contribution of the Survey of India in the documentation of ancient settlements in Pakistan and India. *J. Field Archaeol.* **51,** 107–128. https://doi.org/10.1080/00934690.2025.2572881 (2026).

[9] Das, S. K. & Tripathi, H. India's Green Revolution: fact and fallacy. *Int. J. Bio-Resour. Stress Manag.* **5,** 153. https://doi.org/10.5958/j.0976-4038.5.1.030 (2014).

[10] Singh, R. N. et al. Changing patterns of settlement in the rise and fall of Harappan urbanism and beyond: a preliminary report on the Rakhigarhi Hinterland Survey 2009. *Man Environ.* **35,** 37–53 (2010).

[11] Lindsay, I. & Mkrtchyan, A. Free and low-cost aerial remote sensing in archaeology: an overview of data sources and recent applications in the South Caucasus. *Adv. Archaeol. Pract.* **11,** 164–183. https://doi.org/10.1017/aap.2023.3 (2023).

[12] Green, A. S. 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,** 2089. https://doi.org/10.3390/rs11182089 (2019).

[13] Nicu, I. C. Tracking natural and anthropic risks from historical maps as a tool for cultural heritage assessment: a case study. *Environ. Earth Sci.* **76,** 330. https://doi.org/10.1007/s12665-017-6656-z (2017).

[14] Malaperdas, G., Grethe, R. & Zacharias, N. Depicting the past: The value of old maps and topographic diagrams in cultural heritage through GIS. *J. Archaeol. Sci. Rep.* **52,** 104276. https://doi.org/10.1016/j.jasrep.2023.104276 (2023).

[15] Edney, M. H. *Mapping an Empire: The Geographical Construction of British India, 1765–1843* (Univ. Chicago Press, 1999).

[16] Berganzo-Besga, I. et al. Curriculum learning-based strategy for low-density archaeological mound detection from historical maps in India and Pakistan. *Sci. Rep.* **13,** 11257. https://doi.org/10.1038/s41598-023-38190-x (2023).

[17] Conesa, F. C., Orengo, H. A., Lobo, A. & Petrie, C. A. An algorithm to detect endangered cultural heritage by agricultural expansion in drylands at a global scale. *Remote Sens.* **15,** 53. https://doi.org/10.3390/rs15010053 (2022).

[18] Kour, N., Garcia-Molsosa, A., Orengo, H. & Conesa, F. Mapping the past, preserving the future: integrating remote sensing and geospatial data with local conservation efforts in Jammu, India: a case study of Biyan Tibba. in *Spatial Narratives of India's Heritage: Integrating Geospatial Analysis in the Humanities* (ed. Rajani, M. B.) https://doi.org/10.1007/978-981-95-0117-5_10 (Springer Nature, 2025).

[19] Casana, J., Cothren, J. & Kalayci, T. Swords into ploughshares: archaeological applications of Corona satellite imagery in the Near East. *Internet Archaeol.* **32**. https://doi.org/10.11141/ia.32.2 (2012).

[20] Casana, J. Global-scale archaeological prospection using Corona satellite imagery: automated, crowd-sourced, and expert-led approaches. *J. Field Archaeol.* **45,** S89–S100. https://doi.org/10.1080/00934690.2020.1713285 (2020).

[21] Munteanu, C. et al. The potential of historical spy-satellite imagery to support research in ecology and conservation. *BioScience* **74,** 159–168. https://doi.org/10.1093/biosci/biae002 (2024).

[22] Shahbandeh, M., Kaim, D. & Kozak, J. Using CORONA imagery to study land use and land cover change, a review of applications. *Remote Sens.* **15,** 2793. https://doi.org/10.3390/rs15112793 (2023).

[23] Stratoulias, D. & Kabadayı, M. E. Land cover feature extraction from Corona spy satellite images during the Cold War – 1968. in *IGARSS 2020 – 2020 IEEE International Geoscience and Remote Sensing Symposium* 6069–6072 https://doi.org/10.1109/IGARSS39084.2020.9324485 (IEEE, 2020).

[24] Kabadayı, M. E., Ettehadi Osgouei, P. & Sertel, E. Agricultural land abandonment in Bulgaria: a long-term remote sensing perspective, 1950–1980. *Land* **11,** 1855. https://doi.org/10.3390/land11101855 (2022).

[25] Kabadayı, M. E., Sertel, E., Tumer, I. N. & Sengul, G. S. Analyzing historical land cover change at the interplay of urbanization, rural depopulation, and agricultural land abandonment in Bulgaria and Turkey since the 1950s. in *IGARSS 2024 – 2024 IEEE International Geoscience and Remote Sensing Symposium* 4722–4726 https://doi.org/10.1109/IGARSS53475.2024.10640416 (IEEE, 2024).

[26] Casana, J., Goodman, D. D. & Ferwerda, C. A wall or a road? A remote sensing-based investigation of fortifications on Rome's eastern frontier. *Antiquity* **97,** 1516–1533. https://doi.org/10.15184/aqy.2023.153 (2023).

[27] Casana, J. Satellite imagery-based analysis of archaeological looting in Syria. *Near East. Archaeol.* **78,** 142–152. https://doi.org/10.5615/neareastarch.78.3.0142 (2015).

[28] Casana, J. & Cothren, J. Stereo analysis, DEM extraction and orthorectification of Corona satellite imagery: archaeological applications from the Near East. *Antiquity* **82,** 732–749. https://doi.org/10.1017/S0003598X00097349 (2008).

[29] Casana, J. & Cothren, J. The Corona Atlas project: orthorectification of Corona satellite imagery and regional-scale archaeological exploration in the Near East. in *Mapping Archaeological Landscapes from Space* (SpringerBriefs in Archaeology 5) https://doi.org/10.1007/978-1-4614-6074-9_4 (Springer, 2013).

[30] Ghuffar, S., Bolch, T., Rupnik, E. & Bhattacharya, A. A pipeline for automated processing of declassified Corona KH-4 (1962–1972) stereo imagery. *IEEE Trans. Geosci. Remote Sens.* **60,** 5629614. https://doi.org/10.1109/TGRS.2022.3200151 (2022).

[31] Ghuffar, S., King, O., Guillet, G., Rupnik, E. & Bolch, T. Brief communication: glacier mapping and change estimation using very high-resolution declassified Hexagon KH-9 panoramic stereo imagery (1971–1984). *Cryosphere* **17,** 1299–1306. https://doi.org/10.5194/tc-17-1299-2023 (2023).

[32] Iacone, B., Allington, G. R. H. & Engstrom, R. A methodology for georeferencing and mosaicking Corona imagery in semi-arid environments. *Remote Sens.* **14,** 5395. https://doi.org/10.3390/rs14215395 (2022).

[33] Mihai, B., Nistor, C., Toma, L. & Săvulescu, I. High resolution landscape change analysis with Corona KH-4B imagery. A case study from Iron Gates Reservoir Area. *Procedia Environ. Sci.* **32,** 200–210. https://doi.org/10.1016/j.proenv.2016.03.025 (2016).

[34] Barthelme, P., Darbyshire, E., Spracklen, D. V. & Watmough, G. R. Detecting Vietnam War bomb craters in declassified historical KH-9 satellite imagery. *Sci. Remote Sens.* **10,** 100143. https://doi.org/10.1016/j.srs.2024.100143 (2024).

[35] Klehm, C. et al. Using declassified imagery: issues and approaches. Preprint at SSRN https://doi.org/10.2139/ssrn.7398577 (2026).

[36] Fowler, M. J. F. The archaeological potential of declassified HEXAGON KH-9 panoramic camera satellite photographs. *AARGnews* **53,** 30–36. https://aargonline.com/wp/wp-content/uploads/2022/10/AARGnews53.pdf (2016).

[37] Fowler, M. J. F. The archaeological potential of KH-9 HEXAGON satellite photographs: examples from southern England. *AARGnews* **63,** 12–19 (2021).

[38] Gussone, M., Fowler, M. J. F. & Schneider, A. Al-Hira through the Keyhole – Settlement research in al-Hira/Najaf, Iraq using declassified Hexagon satellite imagery. *AARGnews* **67,** 53–67 (2023).

[39] Marzolff, I. et al. Monitoring dryland trees with remote sensing. Part A: beyond CORONA, historical HEXAGON satellite imagery as a new data source for mapping open-canopy woodlands on the tree level. *Front. Environ. Sci.* **10,** 896702. https://doi.org/10.3389/fenvs.2022.896702 (2022).

[40] Hammer, E., FitzPatrick, M. & Ur, J. Succeeding Corona: declassified Hexagon intelligence imagery for archaeological and historical research. *Antiquity* **96,** 679–695. https://doi.org/10.15184/aqy.2022.22 (2022).

[41] Klehm, C. The use and challenges of spatial data in archaeology. *Adv. Archaeol. Pract.* **11,** 104–110. https://doi.org/10.1017/aap.2022.38 (2023).

[42] Census of India. *Table A-2: Decadal Variation in Population since 1901, Haryana* (Office of the Registrar General & Census Commissioner, Ministry of Home Affairs, Government of India, New Delhi, 2011).

[43] Neogi, S., French, C. A. I., Durcan, J. A., Singh, R. N. & Petrie, C. A. Geoarchaeological insights into the location of Indus settlements on the plains of northwest India. *Quat. Res.* **94,** 137–155. https://doi.org/10.1017/qua.2019.70 (2020).

[44] Orengo, H. A. & Petrie, C. A. 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,** 735. https://doi.org/10.3390/rs9070735 (2017).

[45] Garcia-Molsosa, A., Orengo, H. A. & Petrie, C. A. Reconstructing long-term settlement histories on complex alluvial floodplains by integrating historical map analysis and remote-sensing: an archaeological analysis of the landscape of the Indus River Basin. *Herit. Sci.* **11,** 141. https://doi.org/10.1186/s40494-023-00985-6 (2023).

[46] Government of Haryana. *Statistical Abstract of Haryana 2020–21* (Department of Economic and Statistical Analysis, Government of Haryana, Chandigarh, 2021).

[47] Dyson, T. *A Population History of India: From the First Modern People to the Present Day* (Oxford Univ. Press, 2018).

[48] Petrie, C. A., Singh, R. N. & Singh, A. K. Investigating changing settlement dynamics on the plains: the 2009 survey and excavations at Masudpur (Hissar District, Haryana). *Puratattva* **39,** 38–49 (2009).

[49] Singh, R. N., Petrie, C. A., Singh, A. K. & Singh, M. Excavations at Masudpur (Hissar District, Haryana) – 2009: a preliminary report. *Bhāratī* **33,** 35–49 (2009).

[50] Singh, R. N., Petrie, C. A., Pawar, V., Pandey, A. K. & Parikh, D. New insights into settlement along the Ghaggar and its hinterland: a preliminary report on the Ghaggar Hinterland Survey 2010. *Man Environ.* **36,** 89–106 (2011).

[51] Petrie, C. A. et al. Adaptation to variable environments, resilience to climate change: investigating Land, Water and Settlement in northwest India. *Curr. Anthropol.* **58,** 1–30 (2017).

[52] Kalayci, T. Agricultural production and stability of settlement systems in Upper Mesopotamia during the Early Bronze Age (Third Millennium BCE). PhD thesis, Univ. Arkansas, Fayetteville. https://scholarworks.uark.edu/etd/719 (2013).

[53] Vidyarthi, V., Tomaney, J., Gerrits, P., Khan, A. S. & Sarmah, M. User guide to Sunspot (CAST) (Version 2). *Zenodo* https://doi.org/10.5281/zenodo.19186343 (2026).

[54] Burnett, M. G. *Hexagon (KH-9) Mapping Camera Program and Evolution*. Center for the Study of National Reconnaissance, Chantilly, VA. https://catalog.hathitrust.org/Record/011720473 (2012).

[55] Young, P. *The US Geological Survey, the US Department of Defense, and the US Intelligence Community: 100 Years of Mapping and Remote Sensing Collaboration, 1879–1979*. National Land Imaging Program Circular 1552 (US Geological Survey, Reston, VA, 2025).

[56] Fowler, M. J. F. The archaeological potential of KH-9 HEXAGON satellite photographs: the Roman temporary camps and siege works at Machaerus and Masada. *AARGnews* **64,** 37–53 (2022).

[57] Nita, M. D., Munteanu, C., Gutman, G., Abrudan, I. V. & Radeloff, V. C. Widespread forest cutting in the aftermath of World War II captured by broad-scale historical Corona spy satellite photography. *Remote Sens. Environ.* **204,** 322–332. https://doi.org/10.1016/j.rse.2017.10.021 (2018).

[58] Schug, F. et al. Quantifying multi-decadal urban growth using Hexagon spy satellite imagery and deep learning building detection across four global cities. *Landsc. Urban Plan.* **264,** 105500. https://doi.org/10.1016/j.landurbplan.2025.105500 (2025).

[59] Dehecq, A. et al. Automated processing of declassified KH-9 Hexagon satellite images for global elevation change analysis since the 1970s. *Front. Earth Sci.* **8,** 566802. https://doi.org/10.3389/feart.2020.566802 (2020).

[60] Buławka, N. & Orengo, H. A. Application of multi-temporal and multisource satellite imagery in the study of irrigated landscapes in arid climates. *Remote Sens.* **16,** 1997. https://doi.org/10.3390/rs16111997 (2024).

[61] Sertel, E., Kabadayi, M. E., Sengul, G. S. & Tumer, I. N. HexaLCSeg: a historical benchmark dataset from Hexagon satellite images for land cover segmentation. *IEEE Geosci. Remote Sens. Mag.* **12,** 197–206. https://doi.org/10.1109/MGRS.2024.3394248 (2024).

[62] Shahtahmassebi, A., Shahtahmassebi, G., Moore, N. & Atkinson, P. M. Establishing a historical fine-resolution urban database from KH-9 HEXAGON MCS and PCS imagery including an urban landscape map, stereo anaglyph image and digital surface model. *PFG – J. Photogramm. Remote Sens. Geoinf. Sci.* **93,** 611–644. https://doi.org/10.1007/s41064-025-00362-x (2025).

[63] Zanaga, D. et al. ESA WorldCover 10 m 2021 v200. *Zenodo* https://doi.org/10.5281/zenodo.7254221 (2022).

[64] Garcia-Molsosa, A. et al. Potential of deep learning segmentation for the extraction of archaeological features from historical map series. *Archaeol. Prospect.* **28,** 187–199. https://doi.org/10.1002/arp.1807 (2021).

[65] Pesaresi, M. & Politis, P. GHS-BUILT-S R2023A – GHS built-up surface grid, derived from Sentinel-2 composite and Landsat, multitemporal (1975–2030). *European Commission, Joint Research Centre (JRC)* https://doi.org/10.2905/9F06F36F-4B11-47EC-ABB0-4F8B7B1D72EA (2023).

[66] Earth Resources Observation and Science (EROS) Center. Landsat 1–5 Multispectral Scanner Level-1, Collection 1. *U.S. Geological Survey* https://doi.org/10.5066/F7H994GQ (1972).

[67] Fradley, M. British inter-war aerial photogrammetric mapping in the MENA region: archives, access and research potential. *Levant* **53,** 336–346. https://doi.org/10.1080/00758914.2021.1992879 (2021).

[68] Hammer, E. & Ur, J. Near Eastern Landscapes and Declassified U2 Aerial Imagery. *Adv. Archaeol. Pract.* **7,** 107–126. https://doi.org/10.1017/aap.2018.38 (2019).

[69] Cuca, B., Zaina, F. & Tapete, D. Monitoring of damages to cultural heritage across Europe using remote sensing and Earth Observation: assessment of scientific and grey literature. *Remote Sens.* **15,** 3748. https://doi.org/10.3390/rs15153748 (2023).

[70] Risbøl, O., Briese, C., Doneus, M. & Nesbakken, A. Monitoring cultural heritage by comparing DEMs derived from historical aerial photographs and airborne laser scanning. *J. Cult. Herit.* **16,** 202–209. https://doi.org/10.1016/j.culher.2014.04.002 (2015).

[71] Orengo, H. A. et al. Automated detection of archaeological mounds using machine-learning classification of multisensor and multitemporal satellite data. *Proc. Natl Acad. Sci. USA* **117,** 18240–18250. https://doi.org/10.1073/pnas.2005583117 (2020).

[72] Westley, K., Nikolaus, J., Emrage, A., Flemming, N. & Cooper, A. The impact of coastal erosion on the archaeology of the Cyrenaican coast of Eastern Libya. *PLOS ONE* **18,** e0283703. https://doi.org/10.1371/journal.pone.0283703 (2023).