

Predictive modelling for archaeological sites: Ashokan edicts from the Indian subcontinent

Thomas W. Gillespie*, Monica L. Smith, Scott Barron, Kanika Kalra and Corey Rovzar

This article focuses on the stone inscriptions ascribed to Ashoka, the 3rd century BC ruler of the Mauryan dynasty in ancient India. The locations of 29 known inscriptions and 8 environmental predictors at 1 km pixel resolution were entered into a species distribution model, that reliably predicted the distribution of known Ashokan edicts (AUC score 0.934). Geologic substrate (33%), population density in AD 200 (21%), and slope (13%) explained majority of the variance in the Ashokan edict locations. We have identified 121 possible locations in the Indian subcontinent that conform to the same criteria where yet undiscovered inscriptions may be found.

Keywords: Archaeology, edict locations, environmental metrics, species distribution models, stone inscriptions.

TWO factors have combined that could increase the use of modelling in archaeology: the availability of global landscape geographic information system (GIS) datasets, and the need to identify and protect sites in areas jeopardized by development and other human impacts. An important potential contributor to this process is species distribution modelling, which has been increasingly used across a variety of fields, including biogeography, ecology, conservation biology, and climate change science to identify metrics that define and predict species and ecosystems ranges¹⁻⁴.

Mapping the past, current and future distributions of species and ecosystems has developed at a rapid pace over the last ten years^{4,5}. Species distribution modelling, also referred to as ecological niche or habitat suitability modelling, allows us to map the current distribution of a species and its potential past and future niches⁶. These distribution models have provided hypotheses concerning the location of species to an extent that new species and populations have been discovered^{7,8}.

There are a number of theoretical and empirical similarities between biological species and archaeological sites. Humans, like animals and plants, preferentially target favourable locations for their activities, which in the case of humans includes economic, social and ritual activities. Other factors conditioning the successful emplacement of human investments in the landscape include elevation, topography, climate and the geologic substrate^{9,10}. Sites are fixed-place locations that represent the physical remains of human activities in the past, in which archaeologically recovered features and artefacts

provide confirmation that people found the locations suitable and worthy of investment.

The predictability of archaeological site location is based on a variety of criteria, not all of which are immediately apparent. For example, our earliest ancestors hundreds of thousands of years ago utilized caves as shelters prior to the development of built architecture, but not all known caves were occupied within a given region due to additional factors of selection, such as the preference for microclimates, resource availability, multiple adjacent caves, or proximity to water¹¹.

This article focuses on the stone inscriptions ascribed to Ashoka, the 3rd century BC ruler of the Mauryan dynasty of northern India. Known as 'edicts', the texts are carved into the living rock and onto shaped stone pillars found in the present-day countries of India, Pakistan and Afghanistan (Figure 1). The inscriptions contain pronouncements about kingship, administrative duty and religion, with some sections repeated verbatim from one location to the next. The Ashokan edicts are significant for three reasons. First, they constitute the first decipherable written documents in the Indian subcontinent coincident with the development of urbanism. Secondly, their emplacement as an act of royal proclamation throughout such a large area is interpreted as the evidence for the first substantial unifying political regime of the subcontinent^{12,13}. Finally, the edicts are the first tangible expression of religious practices related to Buddhism, a ritual practice that started in the sixth century BC but only increased in visibility with the imperial sanction provided by Ashoka's proclamations^{14,15}. As documents of national and international significance, the Ashokan edicts are also a focal point of heritage management and preservation. Newly found inscriptions are opportunistically discovered about once a decade, generating considerable public acclaim and scholarly visibility¹⁶⁻¹⁹.

Thomas W. Gillespie, Scott Barron and Corey Rovzar are in the Department of Geography and Monica L. Smith and Kanika Kalra are in the Cotsen Institute of Archaeology and Department of Anthropology, University of California, Los Angeles, CA 90095-1553, USA

*For correspondence. (e-mail: tg@geog.ucla.edu)

This study has three primary objectives. First, we test whether or not a species distribution modelling approach can be used to successfully identify known Ashokan edict locations. Second, we identify the environmental metrics that best explain the distribution of known Ashokan edicts. Third, we identify regions and locations with a high probability of yet undiscovered Ashokan edicts within the Indian subcontinent.

Research design

Species distribution modelling programs require two sets of inputs in order to create distribution maps of the targeted species within landscapes: species locality data from prior research, and environmental predictors in a GIS format^{2,20}.

Ashokan edict locations

Known Ashokan edict locations were mapped on the basis of information found in the descriptive summaries of Allchin and Norman²¹ and Falk²², and on the basis of recent discoveries¹⁸. Map locations of latitude and

longitude were georeferenced in Google Earth as point locations, saved as KMZ files, and converted to ArcGIS shape files. This resulted in the placement of all known living-rock inscriptions to within 100 m as shown in Figure 2. The locations of edicts on moveable stone pillars also are shown in the figure, although the pillar locations were not utilized in the predictive model because the pillars are known to have been moved in both medieval and modern times. The pillars themselves have been subject to reuse, including as road-rollers or religious symbols that are now completely encased or painted over in such a way that the original stone surface can no longer be seen²¹.

Global GIS datasets

GIS predictors were drawn from global datasets on geology, population, climate and topography to ascertain the non-random patterns of distribution in which multiple criteria for placement can be assessed through predictive modelling and the use of spatially explicit GIS datasets.

We downloaded geological data in a vector map format for South Asia from the United States Geological Survey (USGS)²³. This map was designed to help assess global oil and gas reserves as part of the USGS World Energy

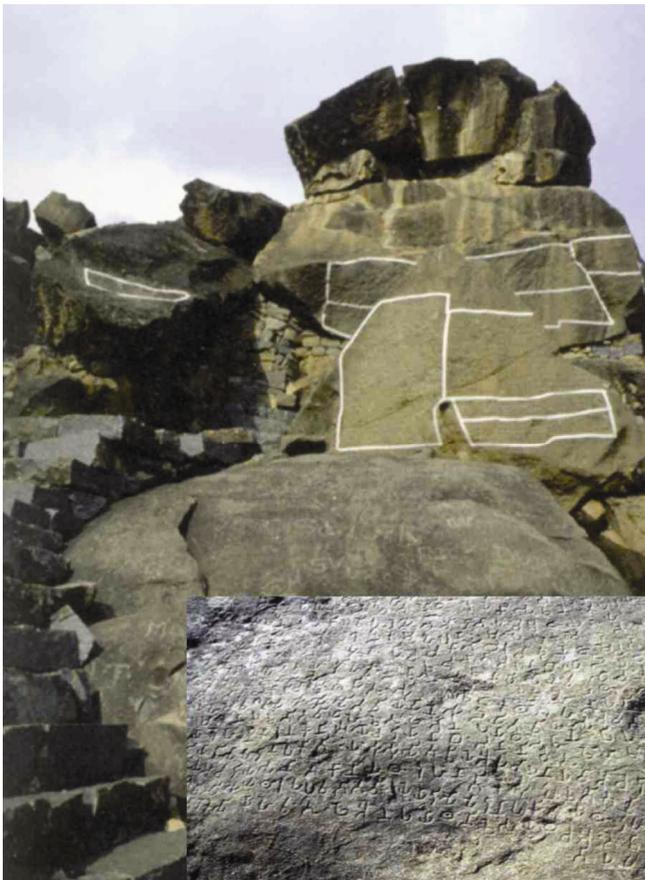


Figure 1. Example of Ashokan edict at Erragudi. (Inset) Script. Source: Falk²².

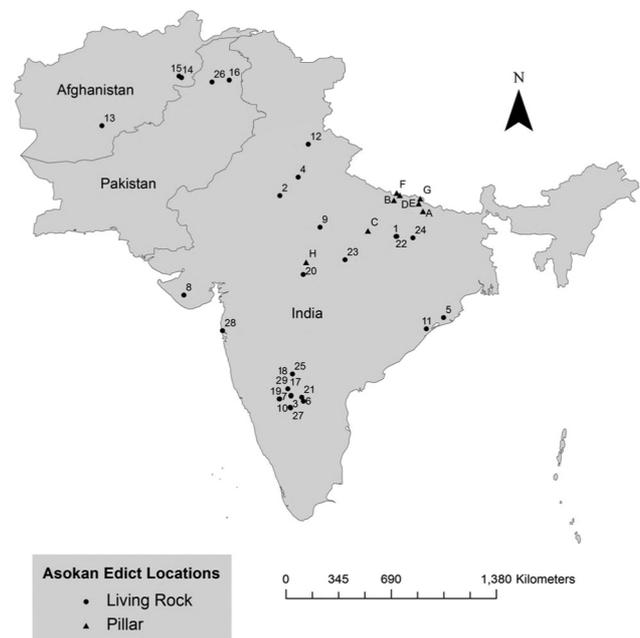


Figure 2. Location of known Ashokan edicts in the Indian subcontinent (numbers denote living-rock edicts; letters denote pillar edicts). 1, Ahraura; 2, Bairat; 3, Brahmagir; 4, Delhi; 5, Dhauli; 6, Erragudi; 7, Gavimath; 8, Girnar; 9, Gujjarra; 10, Jatinga-Ramesvara; 11, Jaugada; 12, Kalsi; 13, Kandahar I and II; 14, Lampaka (Pul-i Darunta); 15, Laghman I and II; 16, Mansehra; 17, Maski; 18, Nittur; 19, Palkigundu; 20, Panguraria; 21, Rajula-Mandagiri; 22, Ratanpurwa; 23, Rupnath; 24, Sahasram; 25, Sannati; 26, Shahbazgarhi; 27, Siddapur; 28, Sopara; 29, Udegolam; A, Araraj; B, Bansi; C, Kausambi; D, Lumbini; E, Nandargarh; F, Nigliva; G, Rampurva; H, Sanchi.

Project, and was created by compiling multiple UNESCO geologic maps primarily from 1976 and 1990 (ref. 23). This map was converted from vector format into 1 km raster format and classified by the dominant geologic formation within each pixel. This resulted in 38 geologic substrates in the study region.

The recent development of the History Database of the Global Environment (HYDE 3.1) provides a spatially explicit database of human-induced global land-use change over the past 12,000 years^{24,25}. We used population density data from the available timescales from the HYDE 3.1 dataset that bracket the period of interest: 1000 BC, AD 0, and AD 200 (refs 24, 25). The data were resampled to 1 km pixel resolution.

There have been significant advances in past, current and future climatic datasets at 1 km resolution^{20,26}. We obtained current climate data from WorldClim (2013) version 1.4 (ref. 27), which includes a set of global climate layers derived from weather station monthly mean temperature and precipitation data²⁶. WorldClim contains 19 derived bioclimatic metrics that represent biologically meaningful climate conditions that have been used to identify the climatic niche of species¹. A subset of metrics hypothesized to be associated with human habitation and agriculture was selected to maximize metric contribution to models: annual mean temperature and annual mean precipitation.

We downloaded elevation data for the regions from the Consortium for Spatial Information (CGIAR-CSI), which provides $5^\circ \times 5^\circ$ mosaicked tiles of 90 m elevation grids. These grids were derived from elevation data, originally collected by NASA's Shuttle Radar Topography Mission (SRTM) in 2003. We resampled the elevation data to 1 km spatial resolution and calculated slope using ArcGIS 10.0 (ESRI, Redlands, CA, USA).

Modelling approach

We used Maxent (version 3.3.3a), a maximum entropy algorithm, to model the current relationship for known Ashokan edict sites and mapped suitable habitat on the basis of climate, topography, geology and population²⁸. Maxent is an algorithm tailored for presence-only species data that have found wide use in modelling current and hindcasting species distribution, with a high performance on presence data that are both limited and spatially biased^{29,30}. Statistics of model performance were calculated using a ten-fold bootstrap replication. This method was chosen due to the small sample size and involves random sampling of the dataset with replacement followed by an analysis of the mean and range from the bootstrap samples to validate the model³⁰. The Maxent output consists of a gridded distribution map with each cell having a logistic index of suitability, or probability of presence between 0 and 1. We used a minimum training presence

threshold which identifies the minimum predicted area possible by locating pixels at least as suitable as the known localities^{30,31}. Model predictions were visualized in ARCMAP 10.0. All predictor metrics were resampled to 1 km pixel size and models run using data from all sites.

Overall model performance was evaluated using the area under the receiving operator characteristics curve (AUC). When using presence-only data, the AUC represents the ability of the model to classify presence more accurately than random prediction, and ranges from 0.5 (random prediction) to 1.0 (perfect prediction). An AUC value greater than 0.75 would suggest that the model is potentially useful for predicting distributions²⁹. We used a Wilcoxon rank sum test to evaluate if the model AUC values were significantly greater than that value for random prediction (0.5).

Landscape analysis

Utilizing *Google Earth* imagery, we conducted a visual search of the Maxent-derived high-probability grids to identify topographical locations (outcrops) and geologic (rock type) attributes similar to known Ashokan edict location as photographed in Falk²² and based on personal observations (Figure 3). One kilometre pixels with a probability $\geq 75\%$ of containing edicts were overlaid on *Google Earth* and searched along a north-to-south gradient. Each pixel was assigned a unique identification number and classified as high likelihood or low likelihood. Locations designated as high likelihood were those that looked similar to known edict locations in Falk²² and generally satisfied two additional criteria: (1) located on raised, lone outcrops, and (2) containing many visible boulders. Low likelihood areas were places without stones, stone outcrops or that contained minimal topography variation. Other criteria also were applied to push

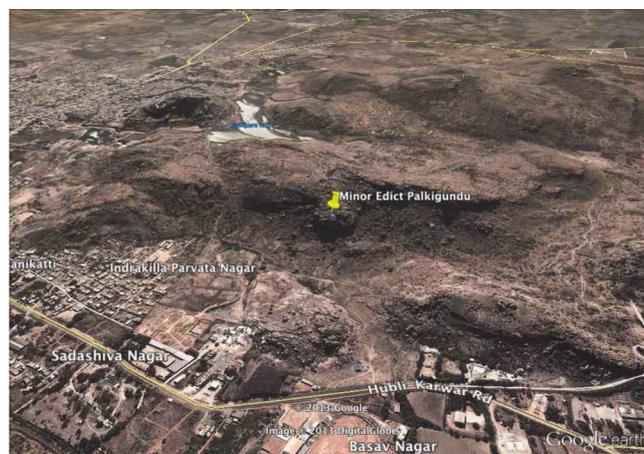


Figure 3. *Google Earth* imagery of known Ashokan edict location at Palkigundu.

Maxent-derived grids into the low likelihood category, such as locations in large cities based on the logic that if an outcrop was located in the area, it is likely that it has already been explored by the surrounding population and does not contain any new edicts.

Results

Ashokan edict distribution models

When distribution models were run based on the 29 known locations, they were successful in reliably predicting the distribution of Ashokan edicts in the Indian subcontinent (Figure 4). Compared to the random prediction (0.5), the AUC score was highly statistically significant (AUC = 0.934, $P < 0.001$, one-tailed Wilcoxon rank sum test of AUC, S.D. = 0.022). This suggests that the model is potentially useful for predicting the distribution of Ashokan edicts and that the environmental metrics have a discernible effect on Ashokan edict regional distribution.

Variable contribution

The environmental variables of geologic substrate (33%), population density in AD 200 (21%), and slope (13%) explained majority of the variance in known Ashokan edict locations (Table 1). Geologic substrate was the metric most associated with edict location and was found to have the highest gain when used in isolation, suggesting

that it has the most useful information (Table 1). Additionally, geologic substrate had the greatest decrease in gain when omitted from the model, suggesting that it has the most information not present in the other variables. Elevation (9%), annual mean precipitation (8%), annual mean temperature (7%), population in AD 0 (5.6%), and population in 1000 BC (3.4%) contributed to the location of the edicts, but explained less of the variance.

Undiscovered Ashokan edicts

The Maxent-derived predicted suitability models identified 2725 1 km pixels with a probability $\geq 75\%$ of containing Ashokan edicts (see Supporting Material A online). After examining each 1 km pixel visually within *Google Earth*, 121 pixels were classified as high likelihood and matched known Ashokan edict attributes (see Supporting Material B online). Examples include points of interest 1973 and 2024 located on small, isolated outcroppings that rise higher than the surrounding region, a configuration similar to the known edict locations of Dhauli, Jaugadh, and Palkigundu (Figure 5). These locations have various boulder-sized rocks, as seen in the images that would make a suitable place for inscriptions.

Discussion

Can species distribution models be applied to Ashokan edicts?

It is clear that known Ashokan edicts have non-random patterns of distribution. Compared to species distribution models of plants and animals with a similar number of point locations, our model has a similar if not slightly higher AUC score of accuracy^{7,8,20,32}. This suggests that species distribution modelling can be used for select type of archaeological sites that are widely distributed across a region or landscape and contain over 30 point locations,

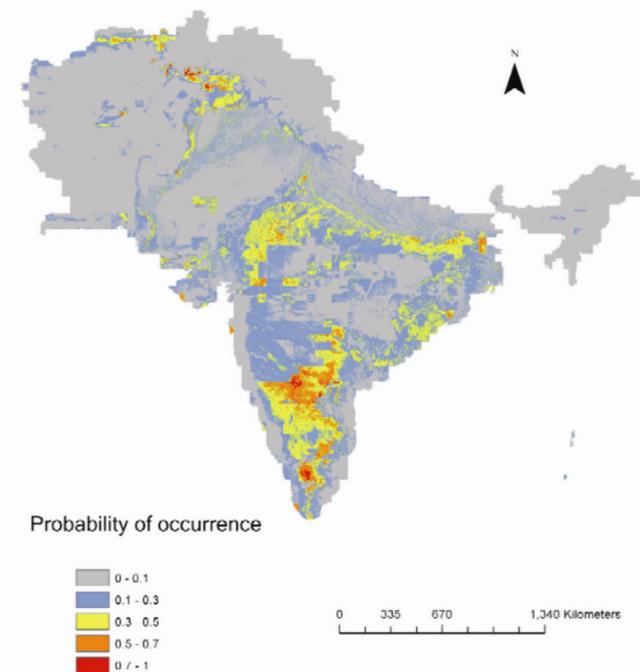


Figure 4. Predicted distribution model of Ashokan edicts on the basis of geology, population, climate and topography.

Table 1. Ashokan sample size (train/test), model mean test AUC, and mean percentage contribution and permutation importance of each variable to the model from ten-fold cross validation

Metrics	Percentage contribution	Permutation importance
No. sites	29	
AUC	0.934	
Geology	32.7	16.2
Population AD 200	20.7	18.2
Slope	13.1	20.2
Elevation	9.1	23.8
Annual precipitation	8.2	7.8
Annual temperature	7.2	5.3
Population AD 0	5.6	6
Population 1000 BC	3.4	2.4

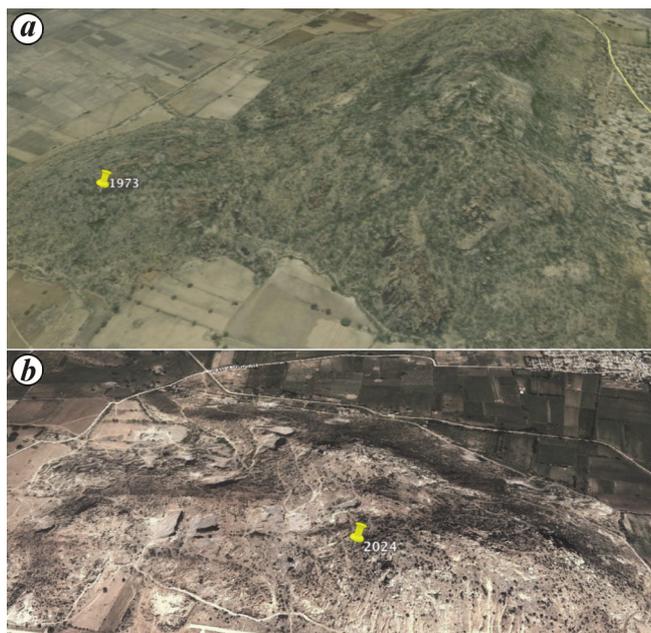


Figure 5. Examples of high likelihood location filtered from the Maxent and landscape refinements within pixel 1973 (a) and 2024 (b).

thus applicable to other types of archaeological sites such as Palaeolithic caves, megaliths and ritual sites. The power of species distribution modelling is that it can identify the amount of variance explained by different environmental metrics and provide a hypothesis concerning the predicted distribution of a site that can be tested and accepted or rejected like any other in science³³.

GIS datasets on geologic substrate, topography and human population explained the most variance in Ashokan edict distributions. Geology was the variable most associated with edict location. Based on geologic substrate, 16 edicts were located on undivided Precambrian rocks, 9 on Quaternary sediments, 2 on Tertiary and Cretaceous volcanic rocks, 2 on Paleozoic rocks, and 1 on Tertiary igneous rocks. This suggests that edict inscribers may have had a preference for certain geologic substrates that occur in the landscape.

Slope and elevation were significant environmental metrics with an average of 0.49° slope for all edicts at 1 km pixel resolution. This is consistent with the rocky outcrops that currently contain Ashokan edicts, which stand out vertically over the landscape; however, variation in the slope is minor at 1 km resolution. It may have been the case that whoever inscribed the edicts preferred places that were slightly higher than the surrounding landscape, but not at the absolute summit of hills. This factor characterizes most of the known edict locations^{21,23}, although there is some variability, such as the Delhi edict that is located on a very low rise and the edict that is located essentially flat on the ground at Rajula-Mandagiri.

The significance of population in AD 200 may not initially seem logical given that the edicts were emplaced about half a millennium earlier. However, the dating of archaeological sites for this time period in the subcontinent is relatively imprecise; in any case it is not surprising that the location of edicts would be placed in areas highly suited to human habitation that were likely to have grown in size over time resulting in their greater archaeological visibility by AD 200. The HYDE database indicators of population in 1000 BC (3.4%) and AD 0 (5.6%) show that incremental population growth did contribute to edict location. It also may be the case that some edicts were carved after the third century BC³⁴. What is clear is that historic GIS datasets, like HYDE 3.1, should be useful in the future of GIS modelling of archaeological sites at 1 km spatial resolution.

Additional future inputs should be sought; for example, time-series data at 1 km on political influence, ethnicity and linguistics would be extremely valuable both as a predictive variable and to assess the impacts of the unifying nature of Ashoka's proclamations. These variables could be created as GIS polygons by experts in the field for different time periods that correspond to the HYDE dataset and would be important additions to future versions of the HYDE dataset.

Finding undiscovered edicts

The use of a predictive GIS model serves a practical and cost-effective function by identifying areas where targeted surveys are most likely to reveal the presence of previously unknown sites at a time when there is increased pressure on landscapes for development. Species distribution models provide at least two novel ways to systematically search for archaeological sites within a landscape or region. First, using species distribution models over the landscape, one can systematically check areas based on those with the highest probability (e.g. 95%). Second, one can visually interpret each location to identify potential locations in the field. As computer scientists have noted, humans often are more effective than computer algorithms in searching for patterns and anomalies in imagery or visual landscapes^{35,36}. Humans also are able to more easily include weighted relevance beyond simple presence-absence determinations^{36,37}. Both techniques add a new dimension to traditional search methods used to discover new archaeological sites.

We believe that a ground-truthing search for edicts in any of these locations would yield a higher-than-average probability of finding new inscriptions. The development of a purpose-specific research project to examine each of the 121 high likelihood areas represents an ideal opportunity for distributed research projects based in local academic institutions in the subcontinent, utilizing the principles of volunteered geographic information (VGI)

and citizen science^{38–40}. We therefore seek to release information about the location of high likelihood areas into the public domain for use by in-country scholars and students.

1. Pearson, R. G. and Dawson, T. P., Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecol. Biogeogr.*, 2003, **12**, 361–371.
2. Guisan, A. and Thuiller, W., Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.*, 2005, **8**, 993–1009.
3. Gillespie, T. W., Foody, G. M., Rocchini, D., Giorgi, A. P. and Saatchi, S., Measuring and modelling biodiversity from space. *Prog. Phys. Geogr.*, 2008, **32**, 203–221.
4. Laffan, S. W., Skidmore, A. K. and Franklin, J., Species distribution and diversity, habitat selection and connectivity. *Int. J. Geogr. Inf. Sci.*, 2014, **28**, 1527–1530.
5. Wiens, J. A., Stralberg, D., Jongsomjit, D., Howell, C. A. and Snyder, M. A., Niches, models, and climate change: assessing the assumptions and uncertainties. *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 19729–19736.
6. Soberón, J., Grinnellian and Eltonian niches and geographic distributions of species. *Ecol. Lett.*, 2007, **10**, 1115–1123.
7. Raxworthy, C. J. *et al.*, Predicting distributions of known and unknown reptile species in Madagascar. *Nature*, 2003, **426**, 837–841.
8. Williams, J. N. *et al.*, Using species distribution models to predict new occurrences for rare plants. *Divers. Distrib.*, 2009, **15**, 565–576.
9. Rockman, M. and Steele, J., *Colonization of Unfamiliar Landscapes: The Archaeology of Adaptation*, Routledge, London, 2003.
10. Carballo, D. M. and Pluckhahn, T., Transportation corridors and political evolution in highland Mesoamerica: settlement analyses incorporating GIS for northern Tlaxcala, Mexico. *J. Anthropol. Archaeol.*, 2007, **26**, 607–629.
11. Dirks, P. H. and Berger, L., Hominin-bearing caves and landscape dynamics in the Cradle of Humankind, South Africa. *J. Afr. Earth Sci.*, 2013, **78**, 109–131.
12. Thapar, R., *Aśoka and the Decline of the Mauryas: with a New Afterword*, Oxford University Press, New Delhi, 1997, 1961.
13. Singh, U., Governing the state and the self: political philosophy and practice in the edicts of Aśoka. *S. Asian Stud.*, 2012, **28**, 131–145.
14. Reat, N. R., *Buddhism: A History*, Asian Humanities Press, Berkeley, 2014.
15. MacGregor, N., *A History of the World in 100 Objects*, Viking, London, 2008.
16. Joshi, M. C. and Pande, B. M., A newly discovered inscription of Aśoka at Bahapur, Delhi. *J. R. Asiatic Soc.*, 1967, **99**, 96–98.
17. Norman, K. R., Aśokan inscriptions from Sannati. *S. Asian Stud.*, 1991, **7**, 101–110.
18. Thaplyal, K. K., *A New Asokan Inscription from Ratanpurwa, Jnana Pravaha*, Varanasi, 2009.
19. Biagi, P., The archaeological sites of the Rohri Hills (Sindh, Pakistan): the way they are being destroyed. *Web J. Cult. Heritage*, 2006, **1**, 77–95.
20. Graham, C. H., Ferrier, S., Huettman, F., Moritz, C. and Peterson, A. T., New developments in museum-based informatics and application in biodiversity analysis. *Trends Ecol. Evol.*, 2004, **19**, 497–503.
21. Allchin, F. R. and Norman, K. R., Guide to the Asokan inscriptions. *S. Asian Stud.*, 1985, **1**, 43–50.
22. Falk, H., *Aśokan Sites and Artefacts: A Source-Book with Bibliography*, Von Zabern, Mainz, 2006.
23. United States Geological Survey, World Geologic Maps, 2013; <http://energy.usgs.gov/OilGas/AssessmentsData/WorldPetroleumAssessment/WorldGeologicMaps.aspx> (accessed on 2 June 2014).
24. Klein Goldewijk, K., Beusen, A. and Janssen, P., Long-term dynamic modelling of global population and built-up area in a spatially-explicit way: HYDE 3.1. *Holocene*, 2010, **20**, 565–573.
25. Klein Goldewijk, K., Beusen, A., Van Drecht, G. and De Vos, M., The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years *Global. Ecol. Biogeogr.*, 2011, **20**, 73–86.
26. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. and Jarvis, A., Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climate*, 2005, **25**, 1965–1978.
27. WorldClim, Global climate data, 2013; <http://www.worldclim.org/> (retrieved on 2 June 2014).
28. Phillips, S. J., Anderson, R. P. and Schapire, R. E., Maximum entropy modelling of species geographic distributions. *Ecol. Model.*, 2006, **190**, 231–259.
29. Elith, J. *et al.*, Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 2006, **29**, 129–151.
30. Pearson, R. G., Raxworthy, C. J., Nakamura, M. and Townsend Peterson, A., Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *J. Biogeogr.*, 2007, **34**, 102–117.
31. Liu, C., Berry, P. M., Dawson, T. P. and Pearson, R. G., Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 2005, **28**, 385–393.
32. Araújo, M. B. and Guisan, A., Five (or so) challenges for species distribution modelling. *J. Biogeogr.*, 2006, **33**, 1677–1688.
33. Wheatley, D. and Gillings, M., *Spatial Technology and Archaeology: The Archaeological Applications of GIS*, CRC Press, London, 2003.
34. Smith, M. L., Networks, territories and the cartography of ancient states. *Ann. Assoc. Am. Geogr.*, 2005, **95**, 832–849.
35. Moravec, H., When will computer hardware match the human brain? *J. Evol. Technol.*, 1998, **1**, 10.
36. Llobera, M., Extending GIS-based visual analysis: the concept of visualsapes. *Int. J. Geogr. Inf. Sci.*, 2010, **17**, 25–48.
37. Thomee, B. and Lew, M. S., Interactive search in image retrieval: a survey. *Int. J. Multimedia Inf. Retrieval*, 2012, **1**, 71–86.
38. Goodchild, M. F., Citizens as sensors: the world of volunteered geography. *GeoJournal*, 2007, **69**, 211–221.
39. Bonney, R. *et al.*, Next steps for citizen science. *Science*, 2014, **343**, 1436–1437.
40. Smith, M. L., Citizen Science in archaeology. *Am. Antiq.*, 2014, **79**, 749–762.

ACKNOWLEDGEMENTS. We thank Rabindra Kumar Mohanty, Tosabanta Padhan, Namita Sugandhi, Sarah Wilkinson, and Edward Wilkinson for discussions on the Ashokan inscriptions.

Received 9 May 2015; revised accepted 7 January 2016

doi: 10.18520/cs/v110/i10/1916-1921