[Supplementary material]

A systematic approach for studying the persistence of settlements in the past

Katherine A. Crawford[1,2,*][ORCID: 0000-0001-7616-6099], Angela C. Huster[1,3][ORCID: 0000-0001-8044-9526], Matthew Peeples[1][ORCID: 0000-0003-4496-623X], Nicholas Gauthier[4][ORCID: 0000-0002-2225-5827], Michael E. Smith[1][ORCID: 0000-0002-0677-8206], José Lobo[5], Abigail M. York[1][ORCID: 0000-0002-2313-9262] & Dan Lawrence[6][ORCID: 0000-0001-5613-1243]

[1] School of Human Evolution and Social Change, Arizona State University, Tempe, USA

[2] Science and Technology in Archaeology and Culture Research Center, The Cyprus Institute, Nicosia, Cyprus

[3] PaleoWest, Phoenix, USA

[4] Florida Museum of Natural History, University of Florida, Gainesville, USA

[5] School of Sustainability, Arizona State University, Tempe, USA

[6] Department of Archaeology, Durham University, UK

[*] Author for correspondence ✉ Katherine.A.Crawford@asu.edu

The online supplementary material details the methods used to derive the different persistence of occupation measures used to create the figures in this paper and provides a description of each dataset.

All data files and code for reproducing the results in this paper can be found on GitHub (https://github.com/mpeeples2008/SettlementPersistence)

Methods

Occupation length calculations

Description of the issue

Archaeological survey datasets often assign sites to chronological periods. Periods are defined by relatively easy-to-observe changes in material culture; most often pottery types, though other materials such as lithics, metal, glass, and architectural characteristics can also be used. However, given the length of periods in many surveys (multiple centuries), it is often unlikely that all sites, what we term occupation locations, dated to a given period were occupied for the entire span of that period. Historically, this has been discussed as the 'contemporaneity problem' within archaeological settlement pattern analysis, where the assumption that all sites dating to a period were simultaneously occupied may give erroneously high population and/or settlement density estimates (Schacht 1984).

There are four factors to consider that affect our understanding of the relationship between the actual duration of occupation at a location during a chronological period and the total length of the chronological period: 1) As period lengths decrease, it becomes less likely that the length of occupation is substantially over-estimated; this can be broadly thought of as the ratio between sites which date only to a single period relative to the number of sites which date the focal period plus an additional period on one side or the other; 2) the degree of material culture change necessary to identify an archaeological period change may not be randomly distributed in time. The causes of changes within material culture, which tend to be used to designate periods, often correspond to periods of change and are likely to be causes for settlement abandonment or foundation. Conversely, an absence of material culture change should often correlate with settlement continuity, where sites are less likely to be abandoned or founded in the middle of a period. This can be seen in Hanson's Roman data (Hanson 2016), where very few city sites are abandoned during the height of the Roman Empire. If there is a strong correlation between site foundation/abandonment and archaeologically visible period transitions, the true length of site occupation will be closer to the total length of the period than would be predicted from the ratio of single and multi-phase sites; 3) as a result of the two prior points, the degree of mismatch between period length and site occupation length is most likely not evenly distributed across a dataset; and 4) the final issue is related to the first and last periods in a dataset. Depending on the temporal span covered by the dataset, the occupation spans of sites in these two periods may be artificially truncated, because the true occupation of the site extended before or after the total span considered by the survey. While this is an issue throughout the dataset it becomes more significant the closer you get to either end of the temporal limits of a dataset. Because most of the surveys that we have used begin with sites dating to the earliest sedentary/agricultural settlement of the region, this issue has generally not applied to the beginning of the sequences under study. On the opposite end, many of the surveys do have internally logical but artificially truncated end dates (such as the conquest of the Americas or the fall of the Roman Empire).

Dewar & Kintigh calculations

Various models have been developed to address the contemporaneity question, with the end goal of estimating how many of a period's sites were occupied simultaneously. Several of these include estimating an average length of site duration as an intermediate step and can be used for our purpose of estimating occupation spans within a period. We applied the methods presented by Dewar (1991) and Kintigh (1994), using the app at https://franklynam.com/dev/dewarmodel/ to run the simulations for the former variant. Both methods use the ratios of different categories of single and multiphase sites to estimate foundation rates, abandonment rates, and average occupation spans for a period. Dewar's variant is simulation-based and considers all sites equivalent. Because this approach is simulation-based, reruns of the data will produce modestly different values. Kintigh's variant is calculation based and considers "B-type" sites (those that extend to both sides of the focal period) to be categorically different to those dating to a single period, or extending to only one side of the focal period. Dewar's variant consistently gives longer estimated occupation spans than Kintigh's. This general approach prioritises issues 1 and 3, at the expense of issue 2. This approach has the advantage of providing an estimate of site occupation that is independent of the length of the degree of chronological resolution available for an area.

Both methods use the same classification of sites in A, B, C, and D types, based on the presence or absence of occupation in the periods immediately before or after the focal period (Table S1). The sum of types A–D should equal the total number of sites that existed during a focal period.

Table S1. Representation of identifying sub-phases within chronological periods.

Site type	**Previous period**	**Focal period**	**Next period**
A	A type: occupied during the focal period and the immediately prior period.		
B	B type: occupied during the focal period and both the immediately prior and immediately following periods.		
C		C type: occupied during the focal period and the immediately following period.	
D		D type: occupied only during the focal period.	

The simulated occupation length values were run using the Dewar app simulation discussed above. The app provides an estimated number of simultaneous occupations and an associated standard deviation. The average span of occupation can be calculated from these as:

$$Average\ Occupation\ Span = 1 \div \left(\frac{\left(\frac{N\ A\ sites + N\ D\ sites}{Period\ Length}\right)}{N. Sim, Occs. given\ by\ App} \right)$$

Using the average of 100 simulation runs for each period, the average occupation span was calculated for each period in a regional sequence. The resulting estimate was divided by the original period length to give the former as a proportion of the latter.

The method given in Kintigh allows for the direct calculation of an estimated occupation span from the number of sites of each type as:

$$\begin{aligned} &Average\ Occupation\ Span \\ &= 0.5\left(\left(Period\ length \frac{N. C\ Sites}{N. C\ Sites + N. D\ Sites}\right) + \left(Period\ Length \frac{N. A\ Sites}{N. A\ Sites + N. D\ Sites}\right)\right) \end{aligned}$$

The average occupation span for each period produced using this method was also divided by the original period length to give the former as a proportion of the latter.

These two estimates of the percentage of the period for which a site was occupied were taken as bracketing estimates for the true average occupation span. The average proportion for each region was calculated as an average estimate for all periods in a regional sequence, and a rounded proportion was selected based on the resulting average. For datasets where both methods indicated that the average site occupation span across all periods was close to or above the associated period length, then the original full period lengths were used for estimating the length of site occupations. For cases where the average occupation span was substantially less than the period length, a fractional value approximately midway between the values produced by the Dewar and Kintigh methods was selected (e.g. half or one-third of the period length) and used for the dataset.

Considering the impact of phase length

In the discussion in the article, we explore the degree to which the coarseness of the chronological phases between cases or within cases through time might impact the overall patterns of persistence we observe. In particular, we evaluated whether sites that were occupied during earlier or longer phases consistently had higher persistence values. As the discussion in the main body of the article outlines, the Santa Valley cases showed some tendency for a relationship between phase length and persistence while the other cases with phase-based estimates (Basin of Mexico, Yautepec Valley and Central Italy) were more variable. In this supplement, we briefly expand upon this with a series of figures that help visualise the relationship between phase length and settlement longevity.

In the figures below (Figures S1–S4), we consider sites in the top 10 per cent of settlement persistence values for each case. The x-axis represents time in years BC/AD with each phase shown as a rectangle marking the beginning and ending dates of that phase. The phases are arrayed along the y-axis in chronological order and the height of each rectangle represents the proportion of sites in the 10 per cent of persistence values for that case that were occupied in the given interval. The rectangles are also colour coded such that yellow represents a low proportion of the highest persistence sites occupied in that interval and red represents a high proportion. Thus, taller and redder rectangles indicate that more of the longest-lived settlements overlap with the interval in question. As these plots show, the Santa Valley (Figure S4) has a greater concentration of early sites in the longest phases in the top persistence values, but we see no such pattern in other cases. Specifically, the highest concentration of top persistence values for Central Italy (Figure S2) are in the Republican and Imperial periods, in the Late Toltec and Aztec periods for the Basin of Mexico (Figure S1), and are generally more evenly distributed among the Formative and Classic periods in the Yautepec Valley (Figure S3). Overall, this suggests that longer and earlier phases are not the driving force behind long-lived settlements in these regions.

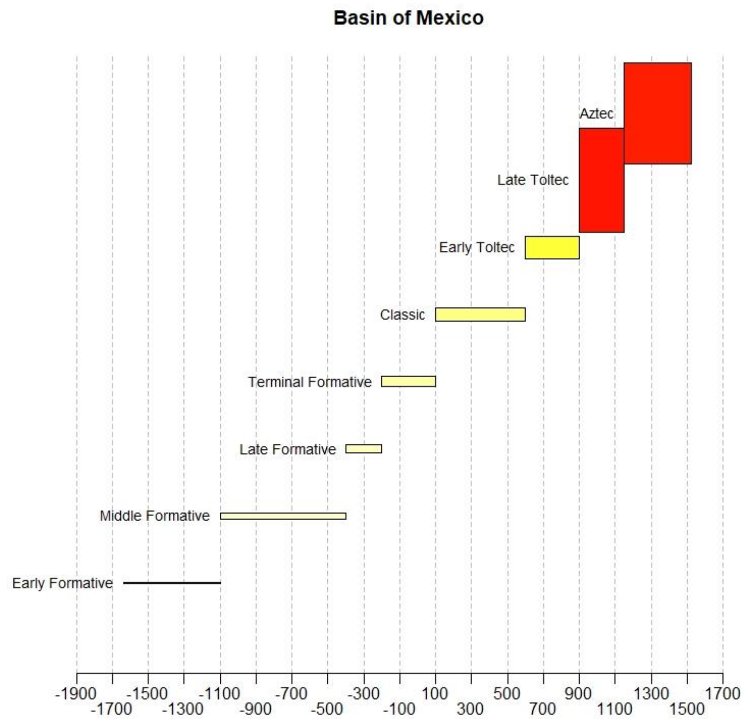

Figure S1. The proportion of the top 10 per cent persistence values for each period interval in the Basin of Mexico dataset (figure by M. Peeples).

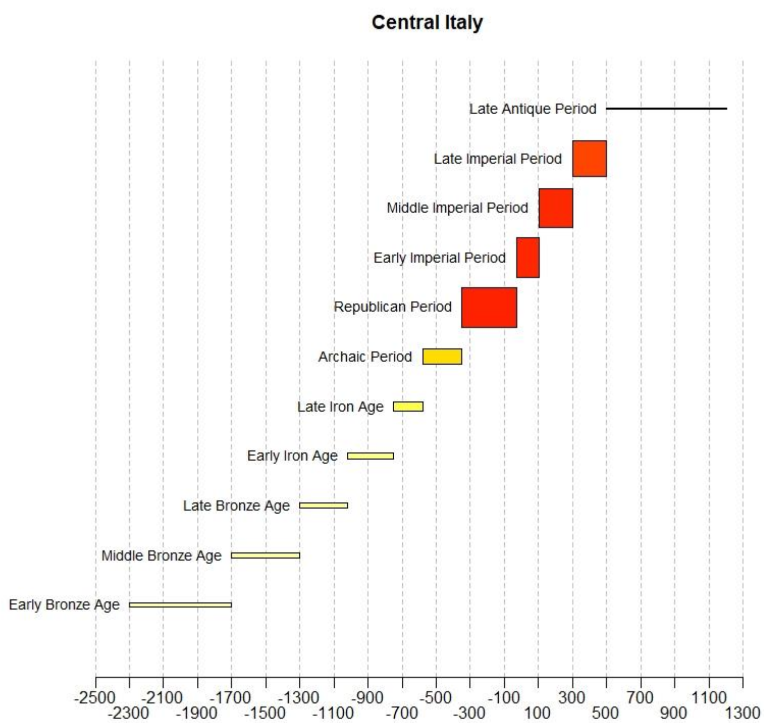

Figure S2. The proportion of the top 10 per cent persistence values for each period interval in the Central Italy dataset (figure by M. Peeples).

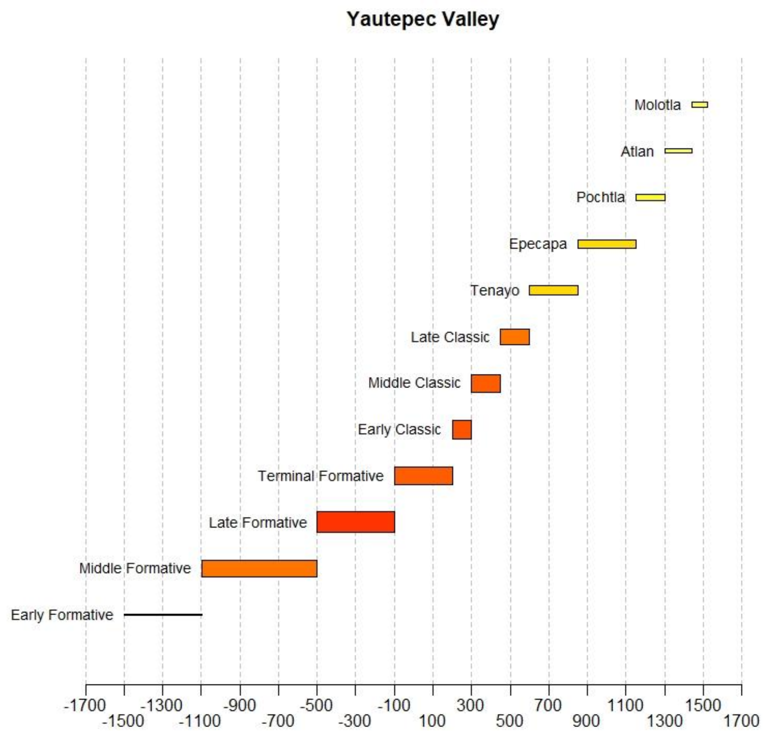

Figure S3. The proportion of the top 10 per cent persistence values for each period interval in the Yautepec Valley dataset (figure by M. Peeples).

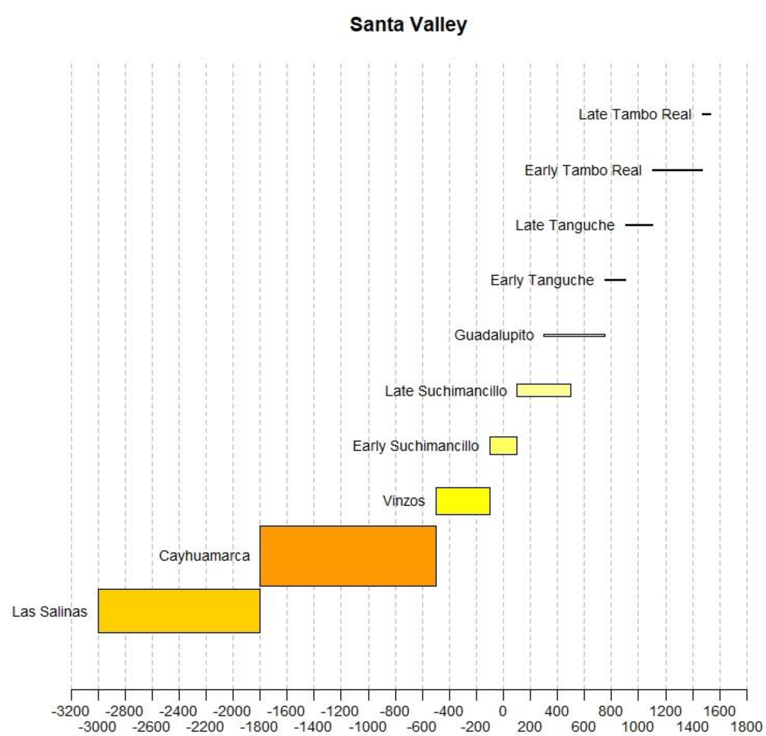

Figure S4. The proportion of the top 10 per cent persistence values for each period interval in the Santa Valley dataset (figure by M. Peeples).

Regional environmental potential calculations

Climatic potential NPP is calculated using the Miami model (Lieth 1973) and modern temperature and precipitation averages from the CHELSA V1.2 dataset (Karger *et al.* 2018), a set of spatially downscaled climate maps that correct for topoclimatic influences such as temperature inversions and orographic precipitation (Karger *et al.* 2017). The result yielded estimates of climatic potential NPP in grams of dry matter per square meter per year, with a theoretical maximum value of 3000 grams $m^{-2}\ yr^{-1}$. We retained the raw NPP values, rather than convert them to expected crop yields using empirical coefficients, to ease comparison among our culturally and geographically distinct case studies. We clipped the resulting spatial NPP estimates to each of the seven regional case studies, using the original survey boundaries where available and present-day administrative boundaries or minimum bounding polygons including all dataset sites where they were not. Although we used present-day climate data from the 1979–2013 period, assuming the rank-ordering of NPP distributions among our regional case studies was roughly the same in the past as the present, future work could incorporate paleoclimate data where available.'

Dataset descriptions

Basin of Mexico dataset

Region description

The Basin of Mexico is a high-altitude, closed hydrological basin surrounding modern Mexico City. During pre-Hispanic periods, the centre of the basin was occupied by a series of interconnected lakes. The initial human occupation of the area occurred during the Paleoindian period, as evidenced by multiple kill sites of mammoths and other megafauna. Agriculture, ceramic use, and permanent settlement began in the Early Formative (1640–1100 BC). The Basin was subsequently the centre of the pre-Hispanic Teotihuacan and Aztec states and remained the political centre of Mexico under Spanish and Mexican rule. Pre-Hispanic agriculture relied heavily on corn, beans, squash, amaranth, and agave, with domestic animals limited to dogs and turkeys.

Data sources and field methods

During the 1960s–1980s, the *Basin of Mexico Survey Project* conducted field-by-field pedestrian surveys of approximately three-quarters of the region; most of the remainder was inaccessible under Mexico City. Archaeologists recorded the spatial extent of sites during each period based on artifact scatters and surface-visible architecture and made collections of diagnostic artifacts at most locations.

The results of the project have been published in two parallel formats. First, there are a series of volumes for each subregion of the survey (except Temascalapa, which is still in progress), with individual text-based site descriptions and various analyses of regional trends (Sanders 1970; Parsons 1971, 2008; Blanton 1972; Parsons & Whalen 1982; Sanders & Gorenflo 2007). Second, there are two volumes of standardised tabular data (site locations, environmental zones, site areas, estimated populations, etc), which collectively cover the entire survey region (Parsons *et al.* 1983; Gorenflo & Sanders 2007). The older of these two volumes has been digitised and uploaded to tDAR and the newer one is accompanied by a CD-ROM with digital versions of the tables. These two data sources were merged into a single database by (Ortman *et al.* 2014) and the authors of that project provided us with a copy.

Data manipulation

Formatting the data for the current analyses required three types of modifications.

First, non-residential sites were removed from the dataset. These were locations identified by the original survey as being: Questionable, Unknown, Ceremonial Center, Isolated Ceremonial Center, Salt Station, Quarry, or Irrigation Canal.

Second, phase-specific occupations were linked to continuous occupations at a given location. The Basin of Mexico Survey used single periods of occupation at a given location as their primary unit of data recording. The textual site descriptions for each period-specific site list its cross-correspondences with overlapping sites dating to other time periods. Two regions (Ixtapalapa and Teotihuacan) have existing tables of site overlaps over time; these tables were used for these two regions. The resulting coding added two variables, “Occupation Site Location” which is shared by all overlapping phase-sites at a location, and “Suboccupation” which divides the former into temporally continuous sets of periods within a location. As an example, a site location with two sequential periods of occupation, followed by a period of abandonment, and then a single period of occupation, would have a single occupation site location and two suboccupations.

Third, the chronological periods were standardised to a uniform eight-period sequence (EF, MF, LF, TF, CL, ET, LT, AZ) that could be applied to all sub-regions of the survey (Table S2). Some subregions of the survey used chronological periods below the level of the primary eight-period sequence for one or more primary periods and assigned separate phase-specific site numbers for each subphase. To avoid double counting such sites, only the subphase with the larger estimated area was retained for analysis.

Table S2. The chronological scheme used for the Basin of Mexico Survey dataset.

Period code	Period name	Start date	End date	Duration (years)
EF	Early Formative	1640 BC	1100 BC	540
MF	Middle Formative	1100 BC	400 BC	700
LF	Late Formative	400 BC	200 BC	200
TF	Terminal Formative	200 BC	AD 100	300
CL	Classic	AD 100	AD 600	500
ET	Early Toltec	AD 600	AD 900	300
LT	Late Toltec	AD 900	AD 1150	250
AZ	Aztec	AD 1150	AD 1520	370
		Dates from Nichols (2016) JAR		
		Early to Middle Formative dates from Stoner and Nichols (2019)		

Occupation length estimations

Minor adjustments were made to the start and end dates of individual periods, relative to those given in the original reports, to bring them up to date with the current understanding of the chronological sequence. The dates used are based on Nichols (2016) and Stoner and Nichols (2019).

Average site occupation spans were estimated for each period, based on the methods described in Dewar (1991) and Kintigh (1994). The differences between these two methods indicated that half of the original period length was generally a reasonable estimate of a site's occupation span, so this fraction was used. Single-period occupations, first periods of occupation, and last periods of occupation were assigned half the length of the original period. Middle periods of an occupation sequence (those with at least one immediately adjacent period on both ends) were assigned the entire length of the original period. In an exception to this general system, sites dating to the final, Aztec, period of occupation were treated as middle periods if they had a prior occupation. The Aztec period was a time of significant population growth, and it was very rare for sites to be abandoned during this period.

Yautepec Valley dataset

Region description

The Yautepec Valley is a north-south oriented river drainage, located in the modern Mexican state of Morelos, south of Mexico City. The area is subtropical, providing a good environment for cotton cultivation during the pre-Hispanic period, and sugarcane cultivation under Spanish rule. Otherwise, the cultural sequence and subsistence base were similar to those previously described for the Basin of Mexico. Two of the empires centred in the Basin of Mexico, the Teotihuacan and Aztec states, controlled the Yautepec Valley.

Data sources and field methods

The Yautepec Valley survey, directed by Michael Smith, Timothy Hare and Lisa Montiel, was conducted during the 1990s. Survey teams fieldwalked the entire valley, recording the extent of artifact scatters and surface visible architecture dating to different time periods. In urban areas, site recording was limited to opportunistic recording of open areas. In addition, the project made both grab-bag collections (of the most diagnostic artifacts) and systematic collections (of all artifacts in a standard area).

The resulting dataset was submitted to the Mexican government as a technical report (Smith *et al.* 2006) and further updates and further details are published in Hare and Montiel's dissertations (Hare 2000; Montiel 2010) and a recent article (Smith *et al.* 2021). Smith provided the authors with a version of the dataset integrating Hare and Montiel's updates for this project. A version of the project dataset, with redacted site locations, can be downloaded from tDAR (Yautepec Archaeological Survey).

Data manipulation

This dataset required three types of editing for use in the current analyses.

First, sites dating after the Spanish conquest of Mexico were removed from the dataset. Sites dating to the Colonial and Republican periods are disproportionately located under modern settlements, making them difficult to identify with archaeological survey methods.

Second, sites that could only be identified to a broader period of time, rather than one of the primary periods used by the project, were removed from the dataset. These were sites assigned to either the Classic period (rather than the Early, Middle, or Late subperiods), or the Middle-Late Postclassic (rather than the Middle, Late-A, or Late-B subperiods). In conjunction with the previous step, this resulted in a standard set of twelve non-overlapping chronological periods for the dataset (Table S3).

Third, phase-specific occupations were linked to continuous occupations at a given location. Similarly to the Basin of Mexico, the Yautepec Valley used single periods of occupation at a given location as their primary unit of data recording. All of the periods of occupation at a given location share a site number, and these sets of sites were subdivided into temporally continuous sets of periods within a location. As an example, a site location with two sequential periods of occupation, followed by a period of abandonment, and then a single period of occupation, would have a single site number and two suboccupations.

Table S3. The chronological scheme used for the Yautepec Valley dataset.

Dates	**Yautepec Valley period**	**Code**	**General central Mexican period**
AD 1440–1520	Molotla	M	Late Postclassic, B
AD 1300–1440	Atlan	A	Late Postclassic, A
AD 1150–1300	Pochtla	P	Middle Postclassic
AD 850–1150	Epecapa	E	Early Postclassic
AD 600–850	Tenayo	T	Epiclassic
AD 450–600	Late Classic	LC	Late Classic
AD 300–450	Middle Classic	MC	Middle Classic
AD 200–300	Early Classic	EC	Early Classic
100 BC–200	Terminal Formative	TF	Terminal Formative
500–100 BC	Late Formative	LF	Late Formative
1100–500 BC	Middle Formative	MF	Middle Formative
1500–1100 BC	Early Formative	EF	Early Formative

Occupation length estimations

The phase fraction estimates produced by the Dewar and Kintigh methods indicated that the whole archaeological period was the most accurate estimate of site duration; i.e. that most sites were occupied for nearly the full extent of the average period. This difference from the Basin of Mexico is most likely due to a combination of shorter phases used by the Yautepec project and a higher degree of settlement persistence in this region.

Southeast US dataset

Region description

This dataset covers mound sites in the US states of Alabama, Georgia, Mississippi, and Tennessee, spanning approximately a thousand years of the Late Woodland and Mississippian periods (~AD 600–1600). During this period, the south-east was characterised by small polities consisting of a central community and smaller surrounding villages or hamlets. Polities formed a rotating mosaic of alliances and higher-level confederations, such as those centred at Moundville. The region was interconnected by trade routes that regularly moved goods over long distances as well as a shared symbolic repertoire (Brown 2004). During the beginning of the study period, the region was transitioning from a reliance on Southern Agricultural Complex crops (sunflower, goosefoot, maygrass, sumpweed, and knotweed) to maize-based agriculture, which characterised the remainder of the study period (Peres 2017).

Data sources and field methods

This dataset is a compilation of known platform mound sites in the states of Alabama, Georgia, Mississippi, and Tennessee, compiled by the Coweeta Long Term Ecological Research Program and Hally (2019). It is available online at: https://portal.edirepository.org/nis/metadataviewer?packageid=knb-lter-cwt.4042.19. The data were compiled from a variety of sources, including historic maps, scholarly publications, CRM reports, and state archaeological site files. The data have been previously analysed and published in Hally and Chamblee (2019). Mound sites are highly visible against the topography of most of the study area, and this dataset should be considered a reasonably comprehensive list of large sites dating to this period in the region. The project created two datasets, one which uses the site as the unit of analysis, and the other that uses the individual mound as the unit of analysis. This project uses the former dataset, which records the Smithsonian trinomial (the standard US site number system), the site name, the number of mounds at the site, the names of phases of occupation (in local chronological sequences), and the start and end dates of any phases of occupation recorded for the site. The chronological assignments for phases in regional chronologies were updated to reflect the current dates for the phase. For security reasons, the publicly available datasets do not include the coordinates for the sites, though these can be requested from the relevant state archaeological offices by qualified archaeologists.

Data manipulation

This dataset required the identification of continuous periods of occupation and discontinuities of occupation at each site. Because of the macroregional nature of this dataset, different sites used different chronological periods, and some sites had periods from multiple chronological sequences included, particularly if they had been subject to archaeological investigations in different decades. Occupation phases were examined by hand and grouped into one or more continuous occupations of a location. Once phases were grouped, each phase in an occupation was coded as Start, End, Middle, Only, or Duplicate. Duplicate phases were those that were completely contained within the time range presented by the other listed phases at the site, usually as a result of multiple chronological sequences being applied to the same site.

Occupation length estimations

Because the sites in the data set do not use a uniform set of chronological periods across the dataset, the methods presented in Dewar and Kintigh cannot be directly applied. Instead, the ratio of single to multiphase sites was compared to the study cases where these methods could be used. For the SE dataset, this ratio falls between that seen for the Basin of Mexico (where half the length of the period was an appropriate estimate) and the Santa Valley (where one-third of the period was an appropriate estimate). As a result, 2/5 of the duration of beginning, end, or only periods of occupation, plus the entire length of middle periods, were used for this analysis, with two exceptions. The first exception is where the only phases given contained significant amounts of overlap (e.g. AD 1000–1200 and 1100–1350). In this case, 2/5 of the total span, from the earliest start to the latest end dates, was used. The second exception is that in cases where the total length of a start, end, or only period was under 100 years, the entire length of the period was used, on the premise that sites with chronological periods of 50 or 75 years were likely already closely approximating the actual dates of use of the site.

US Southwest dataset

Region description

This dataset includes areas in the US Southwest (Arizona, New Mexico, Utah, Colorado and small portions of Nevada, Texas, and California) and small portions of Northwest Mexico (portions of Sonora and Chihuahua) defined within the cyberSW study area (Mills *et al.* 2020). The period considered for this study includes sites dated between AD 800 and 1800 though as noted in the paper for some analyses we limit these data to the pre-contact period before 1540 to avoid measuring the effects of colonial processes. The settlements included in this database were largely occupied by sedentary agricultural populations with sites ranging from 10 people up to thousands of people for the largest sites. The region includes areas that have traditionally been defined as several distinct archaeological culture areas including the Ancestral Pueblo region, the Mogollon region, the Hohokam region, the Patayan region, and portions of the Fremont region. Despite all of this diversity, there is considerable evidence of frequent interaction, population movement, and exchange across this large zone (see Mills *et al.* 2013, 2018).

Data sources and field methods

The data used for this project come from the cyberSW database (Mills *et al.* 2020) and the associated cybersw.org web platform (https://www.cybersw.org). This database represents a major cyberinfrastructure project focused on gathering settlement and material culture from across the US Southwest and Mexican Northwest. The database includes information on a substantial portion of archaeological settlements (>12 rooms) going back to AD 800 as well as a sample of full-coverage surveys and smaller settlements throughout the region. These data were gathered in a series of National Science Foundation supported projects and earlier efforts and now include information on more than 20 000 sites and more than 13.7 million typed ceramic objects. The database also includes information on date assignments for all of the typed ceramic materials included. These data were compiled from over 100 years' worth of academic and cultural resource management research. All of these data are publicly available on the cyberSW web platform along with tools to download the data or analyse them in a web browser.

Data manipulation

For this project, we used all sites in the cyberSW database with at least 20 chronologically sensitive ceramic sherds which represented a total of 5220 sites dating between AD 800 and 1800. For each site, we pooled all systematic typed ceramic collections from various sources in the database and also associated information on site size (room count).

Occupation length estimations

To generate occupation length estimates we used a set of previously published methods (Ortman 2016; Mills *et al.* 2018) as well as tools built directly into the cyberSW web platform. Specifically, our approach has been referred to as Uniform Probability Density Analysis (UPDA; see Ortman 2016 for an extended discussion). This method consists of an empirical Bayesian approach to combining information on the frequency of ceramic types in a given context along with the date ranges assigned to those types. Briefly, this approach first defines the minimum modeling periods based on the overlapping ranges of ceramic types represented at a site. In other words, if a site had two types, one dating from AD 1000–1150 and another dating from AD 1100–1275, we would have three modeling periods (AD 1000–1099, 1100–1149 and 1150–1275). We then create a prior assuming a uniform distribution for each ceramic type such that a given sherd has an equal probability of having been discarded during any year within the date range for that time. This information is then combined with a conditional which assumes that sherds are more likely to be deposited in the portion of their potential ranges defined by greater overlap across all types found at the site. These two things are then combined into new posterior estimates which provide probabilities that a site was occupied in each modeling interval. To avoid unlikely long occupation spans driven by periods with a very low probability of occupation (which could be driven by just one or a few sherds dating to other periods or processes like heir looming) we then further trimmed the probability curve to the inner 95% highest probability interval to obtain the estimated site beginning and end dates (see Mills *et al.* 2018). We used a previously documented R approach (https://github.com/mpeeples2008/UniformProbabilityDensityAnalysis).

Santa Valley dataset

Region description

The Santa Valley is located on the Pacific Coast of Peru; it is one of a series of major drainages which run from the highlands to the coast, providing water for irrigation through otherwise highly arid environments. The area has a very long history of occupation, from initial marine-oriented pre-ceramic settlements on the coast, through the Chimu, Moche, Casma, and Inca cultures and/or states.

Data sources and field methods

The Santa Valley Survey was conducted in 1979–1980 under the direction of David Wilson. It covered approximately 750km^2, including the entire lower Santa Valley and portions of the desert to the north, between the Santa and Chao Valleys. Wilson was a student of Parsons, who directed many of the Basin of Mexico survey segments, so the methods used by the two projects are generally comparable. Fieldworkers walked the region field-by-field and recorded the presence of artifacts and architecture. Artefact collections, especially ceramics, and architectural styles were used to assign sites to phases and to establish the extent of multiphase sites during each period of occupation.

The results of the project, including tabulations of sites and their sizes, probable function, and estimated populations were published in Wilson (1988). The current project digitised the data tables from this volume for the current project.

Data manipulation

This dataset required two modifications for use by the current project. First, non-residential sites were removed from the analysis. In contrast to most of the other datasets used for this project, the Santa Valley survey recorded a significant number of sites where the primary function was not residential. Second, phase-specific occupations were linked to continuous occupations at a given location. The Santa Valley Survey used single periods of occupation at a given location as their primary unit of data recording. Each entry in the data tables in Wilson (1988) lists the cross-correspondences with overlapping sites dating to other time periods. These sequences of occupation at a single location were then subdivided into sets of continuous occupation, where occupations divided by one or more periods of abandonment were considered two distinct occupations.

Occupation length estimation

Wilson defined a series of ten named periods for the Santa Valley, defined primarily by ceramic cross-ties to adjacent valleys (especially the Viru Valley). The relative positions of these phases have remained well supported by later work, but the absolute dates associated with them have shifted significantly (Table S4). The revised dates given here for the first three periods in the sequence are based on (Chamussy & Goepfert 2019); dates for periods 4–7 are based on Chapdelaine's work (Chapdelaine *et al.* 2009; Chapdelaine 2010, 2011; Szpak *et al.* 2020), and dates for the final three periods are based on (Lau 2004). Comparisons with Downey's (2014) revision of the Viru Valley sequence were considered throughout.

Table S4. The chronological scheme used for the Santa Valley dataset.

	Original Wilson chronology				Revised dates		
Pd. No.	**Pd. Name**	**Start**	**End**	**Span**	**Start-Rev**	**End-Rev**	**Span**
10	Late Tambo Real	1350	1532	182	1470	1532	62
9	Early Tambo Real	1150	1350	200	1100	1470	370
8	Late Tanguche	900	1150	250	900	1100	200
7	Early Tanguche	650	900	250	750	900	150
6	Guadalupito	400	650	250	300	750	450
5	Late Suchimancillo	200	400	200	100	500	400
4	Early Suchimancillo	0	200	200	−100	100	200

3	Vinzos	−350	0	350	−500	−100	400
2	Cayhuamarca	−1000	−350	650	−1800	−500	1300
1	Las Salinas	?	−1800	1000+	−3000	−1800	1200

The revised dates resulted in one special case, concerning the boundary between periods 5 and 6. The material culture associated with these two periods seems to have a substantial period of overlap, where the Suchimancillo component represents the continuation of the local tradition, and the Guadalupito (Moche III and IV) material represents an intrusive tradition brought by conquering colonists. This appears in the revised chronology as a 200-year period of overlap between these two phases. For purposes of estimating the duration of occupation, these two periods were treated as their full (overlapping) lengths when they were the beginning or end periods of occupation of a location. When they were “middle” periods, the 200-year overlap was subtracted from the total length of occupation.

Average site occupation spans were estimated for each period, based on the methods described in Dewar (1991) and Kintigh (1994). The differences between these two methods indicated that one third of the original period length was generally a reasonable estimate of a site’s occupation span, so this fraction was used. Single-period occupations, first periods of occupation, and last periods of occupation were assigned one third the length of the original period. Middle periods of an occupation sequence (those with at least one immediately adjacent period on both ends) were assigned the entire length of the original period, with the exception of the previously noted removal of the overlap between periods 5 and 6.

Central Italy dataset

Region description

Central Italy includes the present-day regions of Lazio, including the city of Rome, Tuscany, and a portion of western Umbria. Initial human occupation of central Italy dates to 10 000 BP, based on cave habitations and evidence of lithic assemblages (Palmisano *et al.* 2017). Agricultural and permanent settlements began to form during the Neolithic period (*c*. 6000 BC). The Eneolithic period (3000–2300 BC) saw the beginning growth of settlement size, with many occupation locations growing over 1ha in size. The Bronze Age (2300–1020 BC) and Iron Age (1020–580 BC) were marked by changes in settlement patterning, with large nucleated urban centres forming by the Late Iron Age and Archaic periods (750–480 BC). By the third century BC, the region's settlements began to be unified under the central power of the city of Rome. The number of settlements within the region continued to increase until reaching their peak during the mid-second century AD, following the continued growth of the Roman Empire. The end of the Imperial period saw an overall decrease in population, which continued until the fall of the Roman Empire during the fifth century AD.

Data sources and field methods

The dataset used in this article was drawn from Palmisano and colleagues (2017, 2018). The original data source, available at: https://doi.org/10.14324/000.ds.1575442, was created from 59 separate archaeological surveys and comprises 7386 identified site locations dating from the Late Mesolithic (*c*. 8000 BC) to the Late Imperial period (*c*. AD 500). The dataset ends at the Late Imperial period since this was the focal period compilation.

Data manipulation

The existing data underwent three different forms of manipulation. First, the original dataset was amended to remove occupation locations (sites) dating prior to the Bronze Age since the chronological resolution of earlier periods is inconsistent with the rest of the dataset. Non-residential occupation locations, such as graves, were additionally removed from the dataset. Second, the individual period distinctions were systemised throughout the dataset (Table S5). The following chart shows the period designations which follow those designated by Palmisano and colleagues (2017) in the original dataset. In instances where there is possible deviance between what the start and end date can be, one date was chosen to ensure a consistent methodology for working with the dataset. This is only relevant to the Middle Bronze Age to Late Iron Age periods. For each occupation location, if only one general period entry was given this was expanded to include all periods based on the occupation locations start and end dates. For instance, an occupation location labelled as belonging to the Roman period (350 BC–AD 300) was expanded to include the Republican Period, Early Imperial Period, and Mid-Imperial Period.

Table S5. The chronological scheme used for the central Italy dataset.

Period	Period start date	Period end date
Early Bronze Age	2300 BC	1700 BC
Middle Bronze Age	1700 BC	1300 BC
Late Bronze Age	1300 BC	1020 BC
Early Iron Age	1020 BC	750 BC
Late Iron Age	750 BC	580 BC
Archaic period	580 BC	350 BC
Republican period	350 BC	30 BC
Early Imperial period	30 BC	AD 100
Mid-Imperial period	AD 100	AD 300
Late Imperial period	AD 300	AD 500
Late Antique period	AD 500	AD 1200

Third, phase-specific occupations were linked together into continuous occupations at a given location with a new *SubOcc1* or “suboccupation” variable. Any occupation during a major period (e.g Republican Period) that abuts occupation in an adjacent period (e.g. Archaic and Early Imperial) is considered to be a continuous occupation phase. This allows for consistent time blocks across the entire survey region.

Occupation length estimations

The absolute period length was calculated from an occupation location’s attributed start and end dates for each continuous occupation. The decision to use the absolute period rather than half periods was determined based on the calculation of the Dewar and Dewar-Kintigh estimations for the average occupation span. These calculations were made using the full period length for each period designation. The results give a percentage close to or over 100 per cent of the given occupation span. Therefore, it was determined that absolute periods based on the dataset's given date ranges were most appropriate.

Fertile Crescent dataset

Region description

The Fertile Crescent in Southwest Asia is defined by the arc of mountains which runs from the Mediterranean coast to the Persian Gulf, including the Taurus and Zagros ranges, and the hilly flanks and lowland plains between these upland zones and the steppe-desert of Arabia. Here we use a dataset which covers the northern part of this zone, including the broad, flat plains of inland Syria, south-eastern Turkey and northern Iraq. In contrast to southern Iraq, dry farming predominates in this region throughout the study period, only becoming significant in the later part of the Iron Age (800 BC onwards). The region was settled from the Palaeolithic and was one of the first areas in the world to develop agriculture and sedentary settlement. This was a millennial scale process, beginning in the Pre-Pottery Neolithic (9700–8000 BC). By the Ubaid/Early Chalcolithic period (6000–4500 BC) sedentary settlement was widespread, and archaeological survey shows that most lowland plains would have been settled by a network of village communities, with some settlements reaching 10ha. Nucleated urban centres of 10–120ha emerged during the Late Chalcolithic (4500–3100 BC) and remain part of the settlement system until the present day. Fluctuations in numbers of urban sites are visible, with a peak during the second half of the Early Bronze Age (2600–2000 BC) and a gradual decline in the Middle Bronze Age (2000–1600 BC) and Late Bronze Age (1600–1200 BC). The dataset used here includes the first part of the Iron Age (1200–300 BC) but not the period of major territorial empires which begins with the Neo-Assyrians around 900 BC.

Data sources and field methods

The dataset used in this article was drawn from Lawrence and colleagues (2021), building on that published by Lawrence and colleagues (2016). It includes 131 urban sites divided into 283 occupation phases dating to the period between 4000 and 1000 BC. Urban sites are defined as those exceeding ten hectares in size (Lawrence *et al.* 2016). The dataset was compiled from 43 different surveys, with additional information provided by excavation reports where available. It was also supplemented by analysis of satellite imagery, particularly Corona spy photography.

Reported archaeological phases were converted into years BP using established regional chronologies.

Data manipulation
The dataset was used as provided.

Occupation length estimations
The absolute dates for each continuous occupation were used as provided within the dataset. The Dewar and Kintigh methods were applied using 100 to 300 year intervals since the dataset is presented in years BP. The results indicated that the absolute period dates were most accurate following this methodology.

References

BLANTON, R.E. 1972. *Prehispanic settlement patterns of the Ixtapalapa peninsula region Mexico* (Occasional Papers in Anthropology 6). University Park: Pennsylvania State University.

BROWN, J.A. 2004. Exchange and interaction until 1500, in R.D. Fogelson (ed.) *Southeast, vol. 14*: 677–85 (Handbook of North American Indians 14). Washington, D.C.: Smithsonian Institution.

CHAMUSSY, V. & N. GOEPFERT. 2019. From warless to warlike times in the central Andes: the origins of institutional war between Moche and Casma valleys, northern coast of Peru. *Americae: European Journal of Americanist Archaeology* 4: 7–36.

CHAPDELAINE, C. 2010. Moche and Wari during the Middle Horizon on the North Coast of Peru, in J. Jennings (ed.) *Exploring the nature of Middle Horizon Peru away from Wari centers*: 213–32. Albuquerque: University of New Mexico Press.

– 2011. Recent advances in Moche archaeology. *Journal of Archaeological Research* 19: 191–231. https://doi.org/10.1007/s10814-010-9046-6.

CHAPDELAINE, C., V. PIMENTEL & J. GAMBOA. 2009. Gallinazo cultural identity in the lower Santa Valley: ceramics, architecture, burial patterns, and sociopolitical organization, in J.-F. Millaire & M. Morlion (ed.) *Gallinazo*: 181–206. Los Angeles (CA): Cotsen Institute of Archaeology. https://doi.org/10.2307/j.ctvdjrqbs.18

Coweeta Long Term Ecological Research Program & D.J. HALLY. 2019. *Mound stage data for mounds in Alabama, Georgia, Mississippi, and Tennessee ver 19. Environmental Data Initiative.* Available at: https://portal.edirepository.org/nis/mapbrowse?packageid=knb-lter-cwt.4042.19 (accessed 23 March 2022) https://doi.org/10.6073/PASTA/09C5C68CC096365D09112EBDF7D45F08

DEWAR, R.E. 1991. Incorporating variation in occupation span into settlement-pattern analysis. *American Antiquity* 56: 604–20. https://doi.org/10.2307/281539

DOWNEY, J.T. 2014. Statecraft in the Virú Valley, Peru, in the first millennium AD. Unpublished PhD dissertation, University of Western Ontario.

GORENFLO, L.J. & W.T. SANDERS. 2007. *Archaeological settlement pattern data from the Cuautitlan, Temascalapa, and Teotihuacan regions, Mexico* (Occasional Papers in Anthropology 30). University Park: Pennsylvania State University.

HALLY, D.J. & J.F. CHAMBLEE. 2019. The temporal distribution and duration of Mississippian polities in Alabama, Georgia, Mississippi, and Tennessee. *American Antiquity* 84: 420–37. https://doi.org/10.1017/aaq.2019.31

HANSON, J.W. 2016. *An urban geography of the Roman world, 100 BC to AD 300* (Archaeopress Roman Archaeology 18). Oxford: Archaeopress. https://doi.org/10.2307/j.ctv17db2z4

HARE, T. 2000. Political economy, spatial analysis, and Postclassic states in the Yautepec Valley, Mexico. Unpublished PhD dissertation, SUNY Albany.

KARGER, D.N. *et al*. 2017. Climatologies at high resolution for the Earth's land surface areas. *Scientific Data* 4: 170122. https://doi.org/10.1038/sdata.2017.122

– 2018. Data from: climatologies at high resolution for the Earth's land surface areas (version 1). Dryad. Available at: https://datadryad.org/stash/dataset/doi:10.5061/dryad.kd1d4 (accessed 23 March 2022). https://doi.org/10.5061/DRYAD.KD1D4

KINTIGH, K.W. 1994. Contending with contemporaneity in settlement-pattern studies. *American Antiquity* 59: 143–48. https://doi.org/10.2307/3085507.

LAU, G.F. 2004. The Recuay Culture of Peru's north-central highlands: a reappraisal of chronology and its implications. *Journal of Field Archaeology* 29: 177–202. https://doi.org/10.1179/jfa.2004.29.1-2.177

LAWRENCE, D. *et al*. 2016. Long term population, city size and climate trends in the Fertile Crescent: a first approximation. *PLoS ONE* 11. https://doi.org/10.1371/journal.pone.0152563.

LAWRENCE, D., A. PALMISANO & M.W. DE GRUCHY. 2021. Collapse and continuity: a multi-proxy reconstruction of settlement organization and population trajectories in the northern Fertile Crescent during the 4.2kya rapid climate change event. *PLoS ONE* 16: e0244871. https://doi.org/10.1371/journal.pone.0244871.

LIETH, H. 1973. Primary production: terrestrial ecosystems. *Human Ecology* 1: 303–32. https://doi.org/10.1007/BF01536729

MILLS, B.J. *et al*. 2013. Transformation of social networks in the late pre-Hispanic US Southwest. *Proceedings of the National Academy of Sciences of the USA* 110: 5785–90. https://doi.org/10.1073/pnas.1219966110

– 2018. Evaluating Chaco migration scenarios using dynamic social network analysis. *Antiquity* 92: 922–39. https://doi.org/10.15184/aqy.2018.86

– 2020. CyberSW (version 1.0). Available at: https://cybersw.org (accessed 23 March 2022).

MONTIEL, L. 2010. Teotihuacan imperialism in the Yautepec Valley, Morelos, Mexico. Unpublished PhD dissertation, SUNY Albany.

NICHOLS, D.L. 2016. Teotihuacan. *Journal of Archaeological Research* 24: 1–74. https://doi.org/10.1007/s10814-015-9085-0

ORTMAN, S.G. 2016. Uniform probability density analysis and population history in the northern Rio Grande. *Journal of Archaeological Method and Theory* 23: 95–126. https://doi.org/10.1007/s10816-014-9227-6

ORTMAN, S.G., A. CABANISS, J.O. STUM & L.M.A. BETTENCOURT. 2014. The prehistory of urban scaling. *PLoS ONE* 9: e87902. https://doi.org/10.1371/journal.pone.0087902

PALMISANO, A., A. BEVAN & S. SHENNAN. 2017. Comparing archaeological proxies for long-term population patterns: an example from central Italy. *Journal of Archaeological Science* 87: 59–72. https://doi.org/10.1016/j.jas.2017.10.001

– 2018. Regional demographic trends and settlement patterns in central Italy: archaeological sites and radiocarbon dates. *Journal of Open Archaeology Data* 6: 2. https://doi.org/10.5334/joad.43

PARSONS, J.R. 1971. *Prehistoric settlement patterns in the Texcoco region, Mexico* (Mexico Memoirs of the Museum of Anthropology 3). Ann Arbor: University of Michigan Press. https://doi.org/10.3998/mpub.11396210

– 2008. *Prehispanic settlement patterns in the Northwestern Valley of Mexico: the Zumpango region* (Region Memoirs of the Museum of Anthropology 45). Ann Arbor: University of Michigan Press. https://doi.org/10.3998/mpub.11395688

PARSONS, J.R. & M.E. WHALEN. 1982. *Prehispanic settlement patterns in the Southern Valley of Mexico: the Chalco-Xochimilco region* (Technical Reports 14). Ann Arbor: University of Michigan Press. https://doi.org/10.3998/mpub.11395526

PARSONS, J.R., K.W. KINTIGH & S.A. GREGG. 1983. *Archaeological settlement pattern data from the Chalco, Xochimilco, Ixtapalapa, Texcoco, and Zumpango regions* (Mexico Technical Reports 14). Ann Arbor: University of Michigan Museum of Anthropology. https://doi.org/10.3998/mpub.11394607

PERES, T.M. 2017. Foodways archaeology: a decade of research from the Southeastern United States. *Journal of Archaeological Research* 25: 421–60. https://doi.org/10.1007/s10814-017-9104-4

SANDERS, W.T. (ed.). 1970. The Teotihuacan Valley Project *final report*, Vol. 1–5 (Occasional Papers in Anthropology 3, 10, 13, 15, 19, 20, 21, 24, 25, 26 & 27). University Park: Pennsylvania State University.

SANDERS, W.T. & L.J. GORENFLO (ed.). 2007. *Prehispanic settlement patterns in the Cuautitlan region, Mexico*. University Park: Pennsylvania State University.

SCHACHT, R.M. 1984. The contemporaneity problem. *American Antiquity* 49: 678–95. https://doi.org/10.2307/279736.

SMITH, M.E., T.S. HARE & L.M. MONTIEL. 2006. *Reconocimiento Superficial del Valle de Yautepec, Morelos: Informe Final*. Mexico City: INAH.

SMITH, M.E. *et al*. 2021. Settlement patterns and urbanization in the Yautepec Valley of central Mexico. *Open Archaeology* 7: 378–416. https://doi.org/10.1515/opar-2020-0141

SZPAK, P., J.-F. MILLAIRE & C. CHAPDELAINE. 2020. An Integrated isotopic study of Early Intermediate Period camelid husbandry in the Santa Valley, Perú. *Environmental Archaeology* 25: 279–95. https://doi.org/10.1080/14614103.2019.1583302

STONER, W.D. & D.L. NICHOLS. 2019. The Altica project: reframing the Formative in the Basin of Mexico. *Ancient Mesoamerica* 30: 247–65. https://doi.org/10.1017/S0956536118000421

Yautepec Archaeological Survey. tDAR id: 459479. Available at: https://core.tdar.org/project/459479/yautepec-archaeological-survey (accessed 23 March 2022). https://doi.org/10.48512/XCV8459479

WILSON, D.J. 1988. *Prehispanic settlement patterns in the lower Santa Valley, Peru: a regional perspective on the origins and development of complex North Coast society*. Washington, D.C.: Smithsonian Institution.