Archaeological Prospection

WILEY

RESEARCH ARTICLE OPEN ACCESS

How Does Visibility Count? An Open Data-Driven Approach to Compare the Use of Ground Visibility in Archaeological Field-Walking Surveys in the Mediterranean Region

Néhémie Strupler[1,2]

[1]Institute für Wissensgeschichte des Altertums, Freie Universität Berlin, Berlin, Germany | [2]Department of Archaeology, Durham University, Durham, UK

Correspondence: Néhémie Strupler (nehemie.strupler@fu-berlin.de)

Received: 13 June 2023 | **Revised:** 9 September 2024 | **Accepted:** 14 September 2024

Funding: The research leading to these results received initial funding from the Schweizerischer Nationalfonds zur Förderung der wissenschaftlichen Forschung (SNSF) Early Postdoc. Mobility P2SKP1_181338.

Keywords: archaeological survey | cross-comparison | open data | site detection | survey method

ABSTRACT

There is a common agreement among archaeologists that assessing visibility in the field is essential to measure the accuracy of their observations. Archaeologists widely expect that low visibility negatively impacts the recovery rate of artefacts and sites during field-walking surveys. However, they hold fundamentally divergent opinions on using recorded visibility values and on whether or how to weight the results. In this paper, I undertake a review and comparison of ground visibility assessments from three archaeological field-walking surveys conducted in the eastern Mediterranean, all of which have published their data. Capitalizing on the availability of open data, I recode and analyse the algorithms employed in these surveys. The results highlight the impacts of weighting techniques, and I compare the maps produced with and without weighting. In all cases, the corrections do not substantially change the interpretations of the results at the scale of site identification. As such, this data-driven experiment contributes to the ongoing debate on how to compare effectively and integrate data from various survey projects to study regional trends.

1 | Introduction

Field survey is one of the principal methods to document archaeological resources on the earth's surface (Banning 2002; Knodell et al. 2023). Archaeology plays a key role in understanding socioecological processes and their transformative impacts on people at varying spatiotemporal scales by offering a vast record of human interactions. As the availability of open datasets is steadily increasing, there is a rising potential for conducting comparative research into human occupation of past landscapes. Surveys and especially intensive pedestrian surveys are at the forefront of comparing raw data from fieldwork. However, even after over two decades, there has been widespread consensus on the promising results of comparing different surveys, yet few attempts have moved beyond theoretical preliminaries. (e.g., papers regrouped in Alcock and Cherry 2004; or among others, J. L. Bintliff 1978; Witcher 2008). Archaeologists have conducted relatively few investigations on the methodology to compare results from different survey projects. In this article, I will look at a question that has been repeatedly addressed for 30 years, examining the varied ways it has been approached: How to cope with the impact of the ground visibility on the recovery rate of surface artefacts? This paper uses 'legacy' data collected by previous large-scale, intensive field surveys that have been published online to test the validity of two competing models: what I call the 'weighting model' that implements an algorithm to transform the data and accounts for visibility biases and the 'cautious model', which addresses visibility but does not introduce an explicit weighting of the data. Consequently, this article demonstrates the significant potential of data science to shed new light on long-standing problems in archaeology.

I will largely write the discussion from my point of view as someone who participated in a survey in western Anatolia (Strupler and Wilkinson 2017) and lived in Turkey at the time of conceiving this article. Therefore, it refers to examples and literature from the eastern Mediterranean region because it is the field I know better and from which publications were more easily accessible. Specifically, it is based on the data from three surveys, namely, the Pyla-Koutsopetria Archaeological Project (PKAP), the Sydney Cyprus Survey Project (SCSP) and the Antikythera Survey Project (ASP), which form three study cases (Figure 1 and Table 1). Therefore, some techniques, which are not used in Mediterranean survey because of permit restriction, such as the so-called shovel test, will not be addressed. Finally, in this context, I am primarily interested in the diachronic identification of sites or zones of human occupation to provide a sufficient framework for historical interpretation, rather than evaluating the exact number of sherds or any other artefacts at any given time.

2 | Assessing Visibility in Intensive Field-Walking Survey

Archaeologists have developed methods for intensive field-walking surveys and adopted systematic procedures for recording sites and off-site artefact densities. They widely employ line transect sampling as a method for estimating the abundance of surface artefacts (Banning 2002; Dieudonné 1989). The basic idea of this approach is that observers move along parallel lines regularly dispersed through a study area, looking ahead and, to their left and right for artefacts on the soil surface. When an artefact is detected by an observer, it is recorded and specifically or systematically collected for further study. Surveyors mostly regroup lines into units according to different criteria (e.g., team constitution, existing agricultural fields, geomorphology, or geographic grid) and record a series of data for each unit, such as external conditions (time of day, weather), specific descriptors (topography, soil coverage, modern land use, geomorphology, slope, artefact density) and ground visibility (Banning 2002; J. Bintliff 2013; W. R. Caraher, Nakassis, and Pettegrew 2006). Transect sampling is not unique to archaeology, nor is it the unique sampling method of survey (Stek and Waagen 2022), but it is one of the most common. Researchers in other disciplines such as ecology also use this method extensively to estimate biological populations (Buckland et al. 2015; Manly and Alberto 2015).

Visibility of the ground surface has long been understood as one of the principal variables that strongly influence recovery rates of artefacts and therefore site detection (Schiffer, Sullivan, and Klinger 1978). In daily circumstances, visibility, for example, '500 m of visibility', means the greatest distance at which objects can be identified with the naked eye. In the context of this article and in archaeological surveys, visibility is not measured in metres but as a percentage. Visibility, 'ground coverage' or 'surface exposure', is the degree to which objects on the soil surface can be seen because of vegetation or other elements covering the ground surface but excluding geological processes. It is a way to express how much of the naked topsoil could be seen in a unit area (Gruškovnjak 2019, 71–73). Obviously, visibility is not the only factor that influences survey results, and the impact of many other factors was already demonstrated, such as variability in walkers' perceptions (Hawkins, Stewart, and Banning 2003; Schon 2002), width of surveyed line transect (Banning, Hawkins, and Stewart 2006),

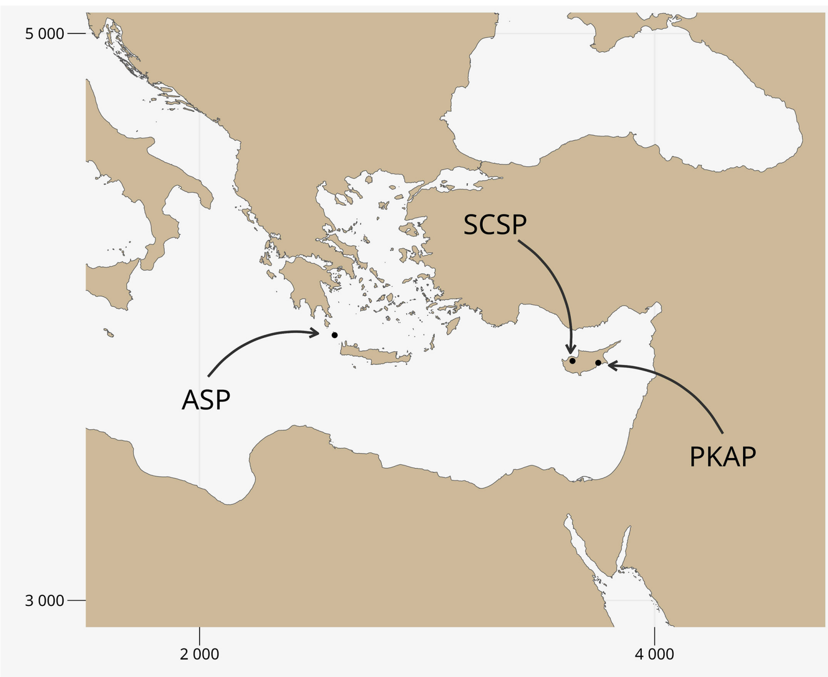

FIGURE 1 | Location of the three survey projects in the eastern Mediterranean region discussed in this paper: ASP, PKAP and SCSP. See Table 1 for further details about the surveys. Map projection: WGS 84/World Equidistant Cylindrical, north-oriented (EPSG: 4087), units of measure in kilometres. Data: Eurostat/GISCO geographical data for the administrative boundaries. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 | Draft layout proposal.

Case	Acronym	Name	Time frame	Survey type	Area surveyed	Dist between walkers	Visibility	Weighting method	Comments
1	PKAP	Pyla-Koutsopetria Archaeological Project	2004–2011	Transects	1 km^2	10 m	Estimated in percentage, ranging from 0% to 100% in 10% intervals	Inverse of visibility	The weighting method was only implemented in a specific chapter.
2	SCSP	Sydney Cyprus Survey Project	1992–1997	Transects	6.54 km^2	5 m	Estimated visibility in percentages between 75% and 100%, along with three categories of background confusion	Inverse of background confusion	The weighting method was implemented consistently.
3	ASP	Antikythera Survey Project	2005–2007	Transects	20 km^2	15 m	Estimated in percentage, ranging from 0% to 100% in 5% intervals	Not implemented	The problem of visibility is discussed, there is but no correction of the data.

walking pace (Banning, Hawkins, and Stewart 2011), artefact size or background confusion. Visibility, however, is considered one of the most influential factors. It has been widely recorded by different projects, thus allowing cross-analysis of different surveys.

Two approaches have been used to test the influence of the visibility: first, resurvey of some units, to evaluate the consistency and reliability of the results (e.g., Ammerman and Feldman 1978; Hirth 1978; Lock, Bell, and Lloyd 1999; Pettegrew 2014; Shennan 1985; see also Banning et al. 2017, 474–476) and experimental approaches (sometimes designated as 'seeding experiments') where artificial units are created to compare known total counts with recovery counts and consequently have the possibility to test the accuracy of survey results (Banning, Hawkins, and Stewart 2006; Clark and Schofield 1991; Meyer and Schon 2003; Schon 2002, 2013; Wandsnider and Camilli 1992). Researchers generally accept that bad visibility conditions lower the recovery rate and can lead to wrong density patterns and consequently to bias in the final patterns of finds repartition. Not only does this affect recovery rate, but poor visibility also impacts other parts of the interpretation, such as the importance of diachronic changes (W. R. Caraher, Nakassis, and Pettegrew 2006).

To account for visibility, in the Mediterranean world (and elsewhere), archaeologists started recording ground surface visibility in several surveys by the mid-1980s (Knodell et al. 2023). For example, the Boeotia Survey adopted a rating from 1 to 10 (J. L. Bintliff 1985), and other projects implemented a percentage scale (i.e., Bevan and Conolly 2013; W. R. Caraher, Moore, and Pettegrew 2014; Cherry, Davis, and Mantzourani 1991; Davis et al. 1997; see also Mattingly 2000, for other references). Although researchers quickly made recording ground visibility a standard procedure, they continue to debate what to do with this value. There are two main positions: (1) ground visibility influences the data, but unpredictably. Transforming data using this value may add more distortion to the general picture than improve it, or (2) ground visibility must be used as a proxy to weight the number of artefacts counted and to analyse the data; without correcting for visibility, we do not know what we are looking at.

In one of the first attempts to study the relationship between ground visibility and site recovery rate, The Keos Survey researchers found that, with higher ground visibility, the recovery rate rises, but they concluded that this was 'not overwhelming' (Cherry, Davis, and Mantzourani 1991, 45, figure 3.6; for an opposite interpretation in a review, see Ammerman 1993), and therefore, no transformation of the data was implemented. The wish to find a correlation function based on the visibility is certainly a position that is widespread among survey practitioners and frequently stated. Lock et al. have surveyed nine fields before and after ploughing (i.e., with better visibility), and although they did not find any direct correlation between the number of collected sherds and visibility, they conclude that 'this is a too small sample to base conclusions on. With further re-walkings we may arrive at a better understanding of the underlying complexities and be able to determine a ground visibility correction function' (1999, 60). On a discussion of survey methodology, Martin Millet insists that even if archaeologists employ corrections, they must always present raw data, and then as research continues 'different and more sensitive 'filters' will be developed to apply better 'corrections' for variation in visibility' (Millet 2000, 93). Similarly, 10 years after, in a seminar on survey in cities, Whitelaw discusses the problem of visibility based on the data from the Knossos Urban Landscape Project. He shows that visibility has a strong impact on artefact recovery, but that the influence of visibility is different according to the artefact types and their size. He concludes that 'what we cannot do, however, is ignore such biases, and rely on raw distribution maps, hoping it somehow "averages out". [...]; if we do, we really do not know what our raw distribution maps are actually telling us' (Whitelaw 2013, 96).

Multiple publications attempted to show that the correlation of visibility and artefact recovery is predictable and that a correlation may be found (Schon 2002, 2013; Stark and Garraty 2008). Archaeologist developed different techniques to use visibility for weighting the counted artefacts, especially pottery sherds. In this article, I implement in the first study case (PKAP) one of the earliest and widest used methods that consists of dividing the number of sherds by the percentage of ground visibility (e.g., J. L. Bintliff and Howard 1999, 53; J. L. Bintliff, Howard, and Snodgrass 2007, 20; Gillings and Sbonias 1999, 36; J. Bintliff et al. 2017, 30). This method brought diverse critiques such as not having been experimentally tested (Barker et al. 2000; Stewart 2020), or being an oversimplifying response to a complex problem. To constrain subjectivity, some surveys assigned the visibility estimate based on a reference (Burgers, Attema, and van Leusen 2004; Van Leusen and Attema 2001, 408). Among others, Robert Schon criticized the one-to-one inverse relationship that is used as a multiplier factor. To address this problem, he conducted a series of experiments to find a more accurate correlation. He did a series of seeding experiments in different survey projects, notably at the Corinthia Project (Schon 2002), at the SCSP (Meyer and Schon 2003) and at the Troodos Archaeological and Environmental Survey Project (Schon 2013). Schon's seminal work provides insightful details and data to better understand how archaeologists record data, what they collect and how representative the sample of a survey is from the actual population. Based on his experiments, he made multiple propositions to calibrate the data. He looked for a correlation between recovery and visibility and found a linear regression that explained well the correlation between visibility and recovery rate (Figure 2, Schon 2002, 149–157). One of the strengths of Schon's work is that it shows if a survey wants to make any transformation based on visibility, it is fundamental to have a protocol to establish visibility in a reliable way. Moreover, Schon makes the convincing point that for any transformation, it is necessary to calibrate the transformation based on experimental results rather than a direct inverse proportion, which has not been validated. However, in my opinion, the regression may be overfitting the data; as it is based on three observations, it may be necessary to be cautious about the strength of this specific correlation. (From a statistical viewpoint, making a regression with so few points is misleading, as only one outlier would strongly influence the result. It is said—as rule-of-thumb—that a minimum of 10 points per variable are required to make a regression; see also Bevan and Conolly 2013, 48–49, for another critique.)

On the other side, there has always been opposition to any data transformation, claiming that the transformation of data may worsen the situation. Mattingly expresses early concerns on the reliability and consistency of the recorded value of visibility and therefore considers that using visibility to modify the pottery count is 'likely to produce distorted "corrected" maps' (Mattingly 2000, 12; a problem addressed later by Schon 2002, as mentioned above). Although everyone recognizes a general tendency to count more artefacts in the landscape when the visibility is higher than when the visibility is low, the implications have been differently interpreted. Bevan and Conolly chose to avoid transforming the data of the Antikythera Survey because, they argue, the impact of visibility is not predictable: There is no linear correlation between individual artefact densities and visibility at the global level or at the unit level, and visibility has little predictive influence on site discovery (Bevan and Conolly 2004, 127–128; see also Barton et al. 2002). Moreover, many covariates should be investigated, and not only visibility (Bevan and Conolly 2013, 48–49).

If researchers discussed when and how to weight data, it is surprising that they mostly set aside the impact of weighting on the interpretation of results. Starting from this observation, this article seeks to offer an evaluation of the impact on the interpretation of the results, especially regarding site identifications.

3 | An Open Data-Driven Exploratory Analysis (DEA)

To take another look at the debate surrounding survey data weighting, instead of focusing on a particular dataset, I will frame the problem of visibility by testing the more common manipulation of visibility in three datasets, whose digital data are published online and freely accessible. It is disappointing to point out that 'free data' was a sufficient discriminant criterion to select datasets from eastern Mediterranean surveys. As far as I am aware, despite growing calls for FAIR data (Attema et al. 2020; de Haas and van Leusen 2020), and projects signalling their intention to make data open (Knodell et al. 2023, 271), only a fraction have been published and are free to study, (re)use, modify and share.

One of the earliest attempts to compare surveys is the work of Davis (2004). However, he only compares surveys with which he was involved and does not deliver the data. Only recently have there been more works that undertake to compare data (Casarotto et al. 2021; Stek and Waagen 2022; Strupler 2021).

Having open archaeological survey datasets holds significant scientific and societal importance (Marwick et al. 2017). Open datasets allow other researchers to replicate and validate findings and promote transparency. FAIR datasets are a valuable resource for teaching, enabling students to have a concrete insight into survey data. Crucially for this article, open datasets facilitate testing of hypotheses, fostering a more comprehensive and rigorous examination of archaeological interpretation (Strupler and Wilkinson 2017). The main databases assembling information on Mediterranean surveys consist of the Fasti Survey (http://www.fastionline.org/survey), MAGIS (http://cgma.depauw.edu/MAGIS/) and, the most

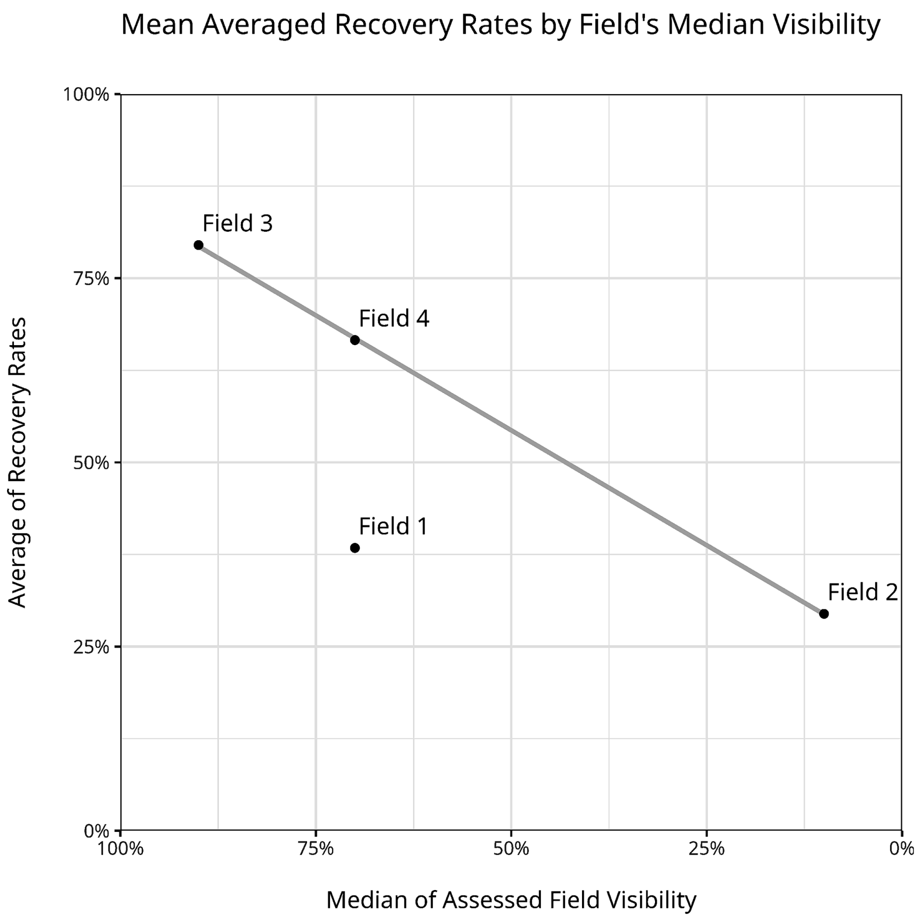

FIGURE 2 | Reproduction of Schon's graphics showing the average recovery rates against the median assessed field visibility (Schon 2002, Chart 5.5, 151). In his dissertation, Schon does not plot or take into account the data from Field 1 and discards it from his regression because of its high background disturbance (Schon 2002, 150). Computational Code to reproduce the graphic: Strupler (2023).

recent, The Fieldwalker.org (https://www.fieldwalker.org/), which incorporates direct links to data if available (Knodell et al. 2023, 271). Online repositories where scholars can publish or access digital data generated by archaeological projects include Open Context (https://opencontext.org/), the Digital Archaeological Record (https://www.tdar.org/), Archaeology Data Service (https://archaeologydataservice.ac.uk/), Zenodo (https://zenodo.org/), IANUS Datenportal (https://datenportal.ianus-fdz.de/), iDAI.geoserver (https://geoserver.dainst.org/) or DANS Data Station Archaeology (https://archaeology.datastations.nl/), among others.

DEA, as a specific branch of data science, seeks to check for assumptions and test hypotheses, often using visual methods, such as maps, and, as such, plays an important role in archaeological surveys. DEA helps to look at data to better understand patterns within the data, detect outliers or find interesting relations among the variables. Maps are arguably the main type of data representations that archaeologists use for spatial data. Archaeologists use maps to interpret the data by displaying specific information and identifying concrete patterns. However, maps have a dual role. They are not only a research tool to detect potential patterns but also serve as evidence. As visual representation, maps are communicative devices that exert a major impact on publishing results and framing arguments (Gillings, Hacıgüzeller, and Lock 2018; Grunwald et al. 2018). Many publications of archaeological surveys use maps as the primary evidence for interpretation, and findings are not necessarily supported by statistical tests. Furthermore, computational analysis of surveys is only rarely reproducible (Strupler 2021). Even today, maps are mainly created as static, printer-friendly versions that do not easily accommodate changes in the statistical analysis. In this article, I argue that the inclusion or exclusion of weighting methods before mapping should influence the patterns visible on published maps, because only in this case, this will lead to different conclusions compared with maps without weighting methods. Conversely, if a statistical manipulation does not significantly alter the distribution patterns (i.e., the appearance of the maps), the interpretation will remain the same as without the weighting method, rendering it superfluous for site identification.

This raises the question: Which data are being mapped? Regional surveys are often classified between 'intensive' and 'extensive' categories. The latter refers to the mapping of sites, which defines the unit of analysis, using methods to detect sites rather than unique artefacts in the landscape (such as satellite imagery or local knowledge). Intensive surveys define artefacts as their unit of analysis and employ methods to systematically recover artefacts from an area, such as transect-based pedestrian field-walking (Knodell et al. 2023, 266–268; Meyer 2023, 143–146). All three analysed surveys are intensive in their techniques, even though the level of intensity varies because of differences in the covered area and the distance between walking lines (see Table 1). Because of their shared geography and their focus on *longue durée* history, these surveys potentially exhibit similar artefact scatters, primarily influenced by the large number of artefacts dating from the Greek and Roman Periods. For example, it is commonplace to discover abundant surface material from antique rural sites, the remains of which are now situated under agricultural fields and are continuously exposed through modern mechanical agricultural ploughing (Knodell et al. 2023, 266). They also share common research questions, typically extending beyond the spatial scope of individual sites to understand changes in human–environment relations over multiple periods. However, the interpretation of data recorded during an intensive survey in a Mediterranean context is a topic of intense debate (Meyer 2023). For some projects, the primary goal is to discover and identify sites, with artefact densities serving as a means to achieve that objective. Other projects emphasize the importance of recording and interpreting areas with lower artefact densities (referred to as 'siteless'), which has led some to recognize the necessity of accurately estimating density.

As well as identifying sites, the re-analysed project in this article, the SCSP, the PKAP and the ASP project sought to understand how people related to the land across the landscape. The projects (as well as the present author) consider the sample to be representative of the original artefact distributions, even if the original amount of artefacts is (and remains) unknown. Specifically, SCSP asserts that the incorporation of pottery data (density and distribution) into maps can potentially facilitate the interpretation of evolving human activities and the utilization of the landscape: 'The study of pottery, then, was the key analytical aspect in assessing the meaning and significance of survey units, in the broadest sense' (Knapp and Given 2003, *Overview and Methodology*). Members of the PKAP team consider that the variation of densities over the landscape provides a coarse approximation of the areas of least and most intensive human activity through time (W. Caraher, Moore, and Pettegrew 2013, p. 35). Members of the ASP team argue that 'the discrepancies between our material expectations and the observed record provide the real basis for any further interpretation' (Bevan and Conolly 2013, 50). Therefore, the distribution maps of artefacts, especially pottery, to recognize sites and zones of occupation play a central role in the interpretation of these surveys, and their appearance has a direct influence on the meaning ascribed to the surveys. It is important to keep in mind that the evidence of presence and activities, sites and their categorization are the focus of these projects, as well as for most Mediterranean archaeologists (Meyer 2023). In this geographical area, few archaeologists are interested in the exact count—a single or five nondiagnostic sherds are likely to not be treated in any great detail. The primary interest is in the diachronic changes in occupation and intensity. Therefore, I will focus on surveys that align with these objectives and not on 'siteless' surveys, which have different goals.

To see how strong weighting techniques change the appearance of maps, I created maps with raw sherd counts side by side with a plot of the weighted data. I use choroplethic maps based on Jenks classification (Jenks and Caspall 1971), that is, a classification that strives to reduce variation within a class and maximize variation among classes. I set a maximum of 10 classes wherever possible, or the maximum that avoids redundant classes. Perception studies indicate that humans are hardly able to discriminate between patterns when more than 10 classes are used. Against intuition, simple maps with few classes convey more information than complex maps with many classes (Darkes 2017). Before classification, I aggregated pottery counts on regular grids (lattices) to allow for a standardized comparison between the different projects. I selected hexagonal grids because this type of lattice is suitable for cluster analysis (Birch, Oom, and Beecham 2007), a common technique for interpreting archaeological spatial data from field survey. Moreover, the projects implemented different collection strategies to date some of the counted artefacts. To assess the potential impact of weighting methods on specific questions, I selected two periods with different distributions from each project and represented them as point patterns. Additionally, to compare the generated maps with and without weighting, a third map was created to illustrate cells with more than one class of difference (i.e., a significant colour variation) after applying Jenks classification (additional figures are provided in the Supporting Information). I use the comparison of class changes to quantify the impact of the weighting techniques, expressed in the number of cells with significant class variation and the corresponding percentage. I generated the maps and the diagrams with *R* (R Core Team 2023), the libraries *sf* (Pebesma 2018), *classInt* (Bivand 2022) and *ggplot2* (Wickham 2016). See Strupler (2023) for the computational code to reproduce the figures.

3.1 | Case 1: The PKAP

At the inception of this project, I planned to use the data from the Beotia Project (J. L. Bintliff, Howard, and Snodgrass 2007), to analyse the first example to implement a weighting method. However, the results I obtained with the data did not match the results from the publication. This is due to some glitches in the published data (Strupler 2021), and the latest publication does not include spatial data, but only the pottery databases and other annexes (J. Bintliff et al. 2017). Therefore, I looked for a project that provides a discussion of a similar weighting method. The PKAP took place on the southern cost of Cyprus and started in 2004. A key component of the research program was an intensive pedestrian survey over an area of 100 ha. The archaeologists divided the area into grid units of 40 m on each side. Teams of four field-walkers positioned at 10-m intervals surveyed each grid. The researchers postulated that each walker covers a 2-m-wide swath through the unit, thus representing a sampling of 20% of the surface (W. R. Caraher, Moore, and Pettegrew 2013, 2014; for a critique of this assumption, see Banning, Hawkins, and Stewart 2011). Pettegrew discusses the problem of weighting the counted artefacts in a specific chapter on the methodology and the validity of the results (Pettegrew 2014). (Even if this chapter discusses other kinds of bias in sampling, such as collection and experience of observers, I will concentrate only on the aspect of visibility. The team also tested the accuracy of their sampling techniques by intensive hoovering the small surface of an already sampled unit and collected all the artefacts to compare the results.) The PKAP coined the term 'pedestrian estimated total', a concept that is equivalent to the weighting of the Beotia Project (J. L. Bintliff, Howard, and Snodgrass 2007). It represents the weighted total count of sherds by using a weighting factor equal to the inverse of the proportion of ground visibility. For example, a total count of 10 sherds associated with 50% visibility is transformed into a pedestrian estimated total of 20 (10*100/50). This dataset therefore represents an excellent foundation to compare the influence of weighting by visibility implementing the first historical, simple weighting factor similar to the Beotia project. It is important to mention that, although weighting is discussed in the publication, it was not implemented in the analysis of the artefact distribution in the publication (W. Caraher, Moore and Pettegrew 2013). Eventually, central for the weighting is how the visibility was assessed and recorded. Surface visibility was recorded at 10% intervals (W. R. Caraher, Moore and Pettegrew 2014, 26), based on the covering of the surface: '50% visibility indicates only half the area of the unit was visible' (Caraher, Moore and Pettegrew 2014, 52).

Using the data and re-encoding the algorithm (for background details, see Strupler 2021), I created two maps, one with a raw density of artefact count (Figure 3, left) and one with the weighted estimated density (Figure 3, centre). A comparison of the Jenks classification (Figure 3, right) shows that 95% of the grid (528 out of 555 cells) has less than a class difference. This highlights that the pattern of distribution is highly similar. The zones of distribution of artefacts and their respective intensity are closely similar in both cases. We can admit that, based on either map being 95% similar, the interpretation would be similar, leading to delimit high clustering and low artefact densities in the same areas.

To address diachronique changes, the project collected the artefacts to be dated using the chronotype system in batches, described as

> Fieldwalker should collect a maximum of one rim, base, handle, and body sherd of each chronotype in his or her transect. If a walker has already collected a combed-ware body sherd and an ARS Form 50 rim, for example, she would not collect additional examples of combed-ware and ARS Form 50 rims found in the tract, but would count them as part of the total count and only collect additional examples of grooved body sherds of different thickness, color, and fabric. If four fieldwalkers walking at 10 m intervals in a 40 × 40 m square were to collect the unique objects visible in their swaths, each unit should produce as many as 16 examples of a single chronotype, corresponding to 4 rims, 4 bases, 4 handles, and 4 bodysherds of the same kind of pottery. (W. R. Caraher, Moore and Pettegrew 2014, 37).

Does visibility impact the visualization of sherds from specific periods using this collection strategy? I selected two periods, Late Bronze Age (henceforth LBA) and Early Roman (ER), which display different patterns. The artefacts dated to the LBA period are strongly clustered (Figure 4). The artefacts dated to the ER period are more evenly spread (Figure 5). In both cases, the clustering and dispersion of artefacts are analogous with or without weighting. If we plot these data aggregated on a hexagonal grid and use colours based on a Jenks classification, the similarity amounts to 99% for the LBA period (551 out of 555 cells; Figure S1) and 96% for the Roman period (534 out of 555 cells, Figure S2). For the LBA period, we note only some changes in the distribution in the area where the clustering is obvious (Figure 4, right). No significant changes are observed. Archaeologists would identify the areas of occupation by using the maps on the left with raw data or the maps in the centre with weighted data. An inspection of the raw numbers behind the maps shows that only the

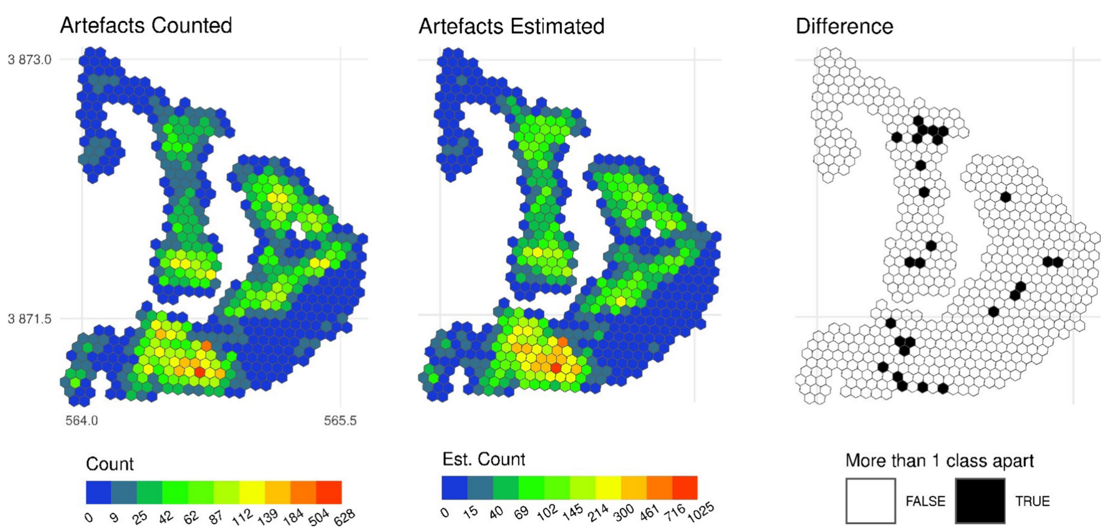

FIGURE 3 | Collected (left) and estimated (centre) artefact distribution densities coloured according to a Jenks classification and their difference (right) from the PKAP Project. Map projection: WGS 84/UTM zone 35N, north-oriented, units of measure in kilometres (EPSG: 32635). [Colour figure can be viewed at wileyonlinelibrary.com]

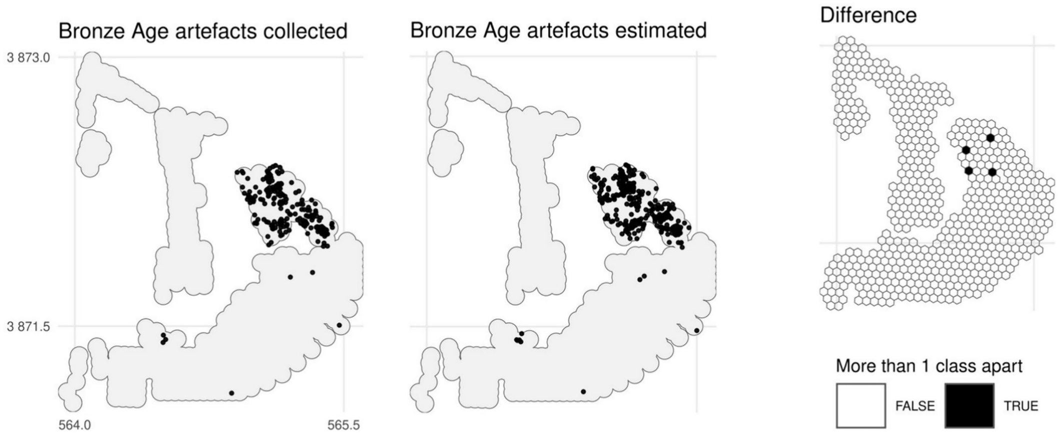

FIGURE 4 | Collected (left) and estimated (centre) Late Bronze Age artefacts. Each point represents a collected (left) or estimated (right) artefact. The difference (right) was computed according to a Jenks classification (Figure S1). Map projection: WGS 84/UTM zone 35N, north-oriented, units of measure in kilometres (EPSG: 32635).

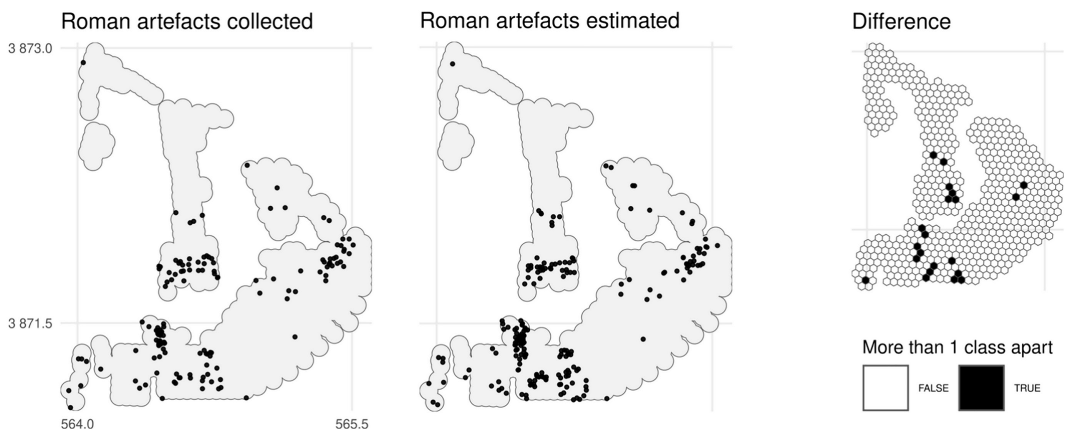

FIGURE 5 | Collected (left) and estimated (centre) Early Roman artefacts. Each point represents a collected (left) or estimated (right) artefact. The difference (right) was computed according to a Jenks classification (Figure S2). Map projection: WGS 84/UTM zone 35N, north-oriented, units of measure in kilometres (EPSG: 32635).

counts associated with visibility below 30% are substantially influenced by the weighting. However, these three groups of visibility (from 0% to 10%, from 10% to 20%, and from 20% to 30%) represent a small part of the units defined in the survey (respectively 3, 9 and 44 units out of 465, totalling to 12%). It is tempting to conclude that weighting by the inverse of visibility would not change the occupation patterns that Mediterranean archaeologists would typically identify.

3.2 | Case 2: The SCSP

The SCSP was an intensive archaeological survey in the northern Troodos Mountains in Cyprus (Given and Knapp 2003; Knapp and Given 2003). Between 1992 and 1997, the SCSP surveyed intensively 654 ha, with a 5-m spacing between walkers. The team divided the area into circa 1550 units distributed in six noncontiguous zones called special interest areas (SIAs). This project also collected artefacts according to the chronotype system, but it implemented a weighting factor for the artefacts based on experimental data. Meyer and Schon propose an equation that takes three values (pottery count, background visibility and background confusion). This was based on their experimentation of five fields artificially seeded and surveyed (Meyer and Schon 2003). They consider that the variable, visibility, does not play an important role compared with the variable background confusion. Background confusion is defined as 'the extent to which red or gray stones, for example, made it more difficult to identify sherds' (Given and Knapp 2003, 34). Banning (2017, 472) refers to a similar effect under the name 'artifact obtrusiveness', a term already used in the 1970s (Schiffer, Sullivan, and Klinger 1978, 6). Although 'surface visibility' and 'artifact obtrusiveness' are two different factors, the first recording the capacity to see the ground and the second capturing the capacity to distinguish artefacts from the soil matrix, the way they implemented the correction factor makes it almost identical to visibility when we compare the impact of weighting. Before implementing their equation, they rescaled visibility to range between 75% and 100% rather than from 0% to 100%. (I could not find any documentation about the rescaling of the visibility.

A close look at the data to understand it shows that it is not strictly linear, but close to the equation $f(x) = 75 + 0.25x$ (for further details, see Strupler 2021.) The purpose of this rescaling was to prevent visibility from having a strong influence on the data. Even with the lowest recorded visibility in the field (1%), it would raise the pottery count by a maximum of 6.7% (for further details, see Strupler 2021). Most of the weighting factor is influenced by 'background confusion', which was recorded in three categories (low, moderate and high confusion). Eventually, their weighting equation for the pottery count is the inverse of the background confusion (Meyer and Schon 2003). This provides a good dataset to compare the effect of implemented weighting in survey projects) though not the difference between visibility and obtrusiveness.

In order to address the impact of the weighting algorithm on the interpretation, I randomly selected one specific area (SIA 7) to produce detailed maps and compare them with the publication and the interpretation of the data (Figure 6). In this case, I recoded the weighting methods from the project and compared it with raw densities. From the global distribution of the sherds (Figure 6, left), we observe some slight differences. The transformation (Figure 6, centre) reduces some higher densities in the north, which are less marked when weighted. However, the differences are minimal. The Jenks classification shows that 99.8% of the grid (958 out of 959 cells) has less than a class difference, confirming that the difference is negligible (Figure 6, right). The areas with notable concentrations remain the same. Although the comparison of specific periods in the PKAP project did not show obvious changes (case 1), does the specific weighting method of the SCSP project make a bigger impact when comparing specific periods? Plotting side by side the density of raw and weighted data for two periods, namely, the 'Geometric to Classical' (GAC) and 'Hellenistic to Late Roman' (HLR) periods, illustrates that only minor differences can be noticed (Figures 7 and 8). If we plot these data aggregated on a hexagonal grid and use colours based on a Jenks classification, the similarity amounts to 98% for the Geometric to Classical period (938 out of 959 cells; Figure S3) and 99% for the Hellenistic to Late Roman period (947 out of 959 cells; Figure S4). Using the map with raw or weighted densities does not change the interpretation of zones of occupation and the identification site for a specific period.

3.3 | Case 3: The ASP

The ASP was a survey focused on the 20-km^2 Greek island of Antikythera, situated between the Peloponnese and Crete. Archaeologists surveyed the island over 4 years (2005–2008) and completed the artefact study by 2010. Following the publication of a series of articles, the researchers made the data available online (Bevan and Conolly 2014) and also published a data paper (Bevan and Conolly 2012), which served as the basis for their book synthesizing the results (Bevan and Conolly 2013). The team conducted the general survey using line transect sampling, with a spacing of 15 m between the walkers, divided into tracts. Each member counted all artefacts but collected only the 'feature' potsherds and all other finds requiring further study and dating (Bevan and Conolly 2013, 13–15).

In their book, the authors discuss the problem of visibility and detail why they prefer not to weight the data (Bevan and Conolly 2009, 2013, 47–50). They reveal that, depending on the number of classes for the visibility, a different relation between density and visibility would be considered. The visibility was recorded as a percentage estimated between 0% and 100%. By using five even classes to classify the percentage of visibility, we would find a linear relation between artefacts counts and visibility (Figure 9), but this is no longer evident by using 10 classes (Figure 10). They stress that they do not believe in a simple relation between visibility and artefacts count. They consider that many interdependent factors

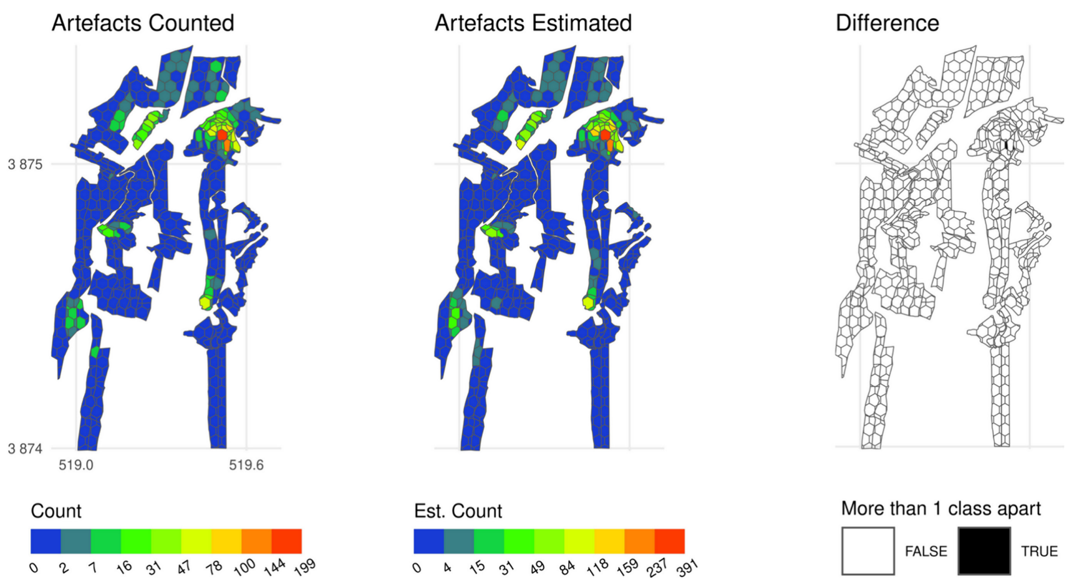

FIGURE 6 | Overall pottery density of the SIA 7 from SCSP, without weighting (left), with weighting (centre) and difference (right). Density is coloured according to a Jenks classification. Map projection: WGS 84/UTM zone 36N, north-oriented, units of measure in kilometres (EPSG: 32636). [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 7 | Location of counted pottery (left) and estimated location of weighted pottery (centre) of the SIA 7 from SCSP. The difference (right) was computed according to a Jenks classification (Figure S3). Pottery estimated according to the project's weighting factor for the Geometric to Classical (1050–312 BC) Period. Map projection: WGS 84/UTM zone 36N, north-oriented, units of measure in kilometres (EPSG: 32636).

FIGURE 8 | Location of counted pottery (left) and estimated location of weighted pottery (centre) of the SIA 7 from SCSP. The difference (right) was computed according to a Jenks classification (Figure S4). Pottery estimated according to the project's weighting factor for the Hellenistic to Late Roman (312 BC–750 AD) Period. Map projection: WGS 84/UTM zone 36N, north-oriented, units of measure in kilometres (EPSG: 32636).

exist (such as the slope, soil types and interaction of humans with the landscape), which may have consequently reduced the amount of vegetation cover, favouring over the long-term the same places over and over. They do not interpret the data with a weighting factor, so I implemented a weighting according to the inverse of the visibility. (In order to have comparable results, I changed the value of visibility to 10% when it was set to 0% to avoid division by zero when dividing by visibility.) This allows to compare the maps and to investigate what impact there may have been on the interpretation, had the counts been weighted.

Plotting side by side the global distribution of pottery without and with weighting by a factor of the inverse of visibility produces similar maps (Figure 11). The Jenks classification reveals that 98% of the grid (1074 out of 1092 cells) shows less than one class difference, confirming that the variation is negligible. The areas with notable concentrations remain unchanged. Even if the relationship between pottery count and visibility is not strictly linear, the results are very similar. A similar inference can be made if we select two periods and look at the distribution of dated artefacts, such as the 'Geometric to

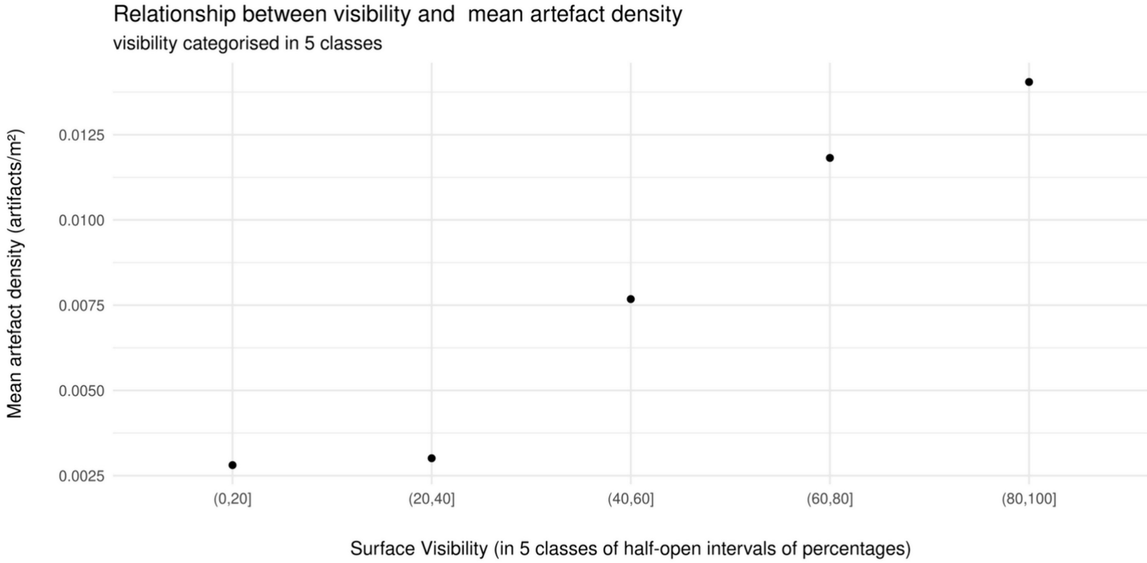

FIGURE 9 | Relationship between ground surface visibility and artefact density for potsherds when visibility is classified into five classes. The different visibility records were classified into half-open intervals, such as (0, 20], indicating an interval from 0 (exclusive) to 20 (inclusive). The graphic replicates (Bevan and Conolly 2013, 49 figure 4.1a), and according to the authors, this could indicate a linear relationship.

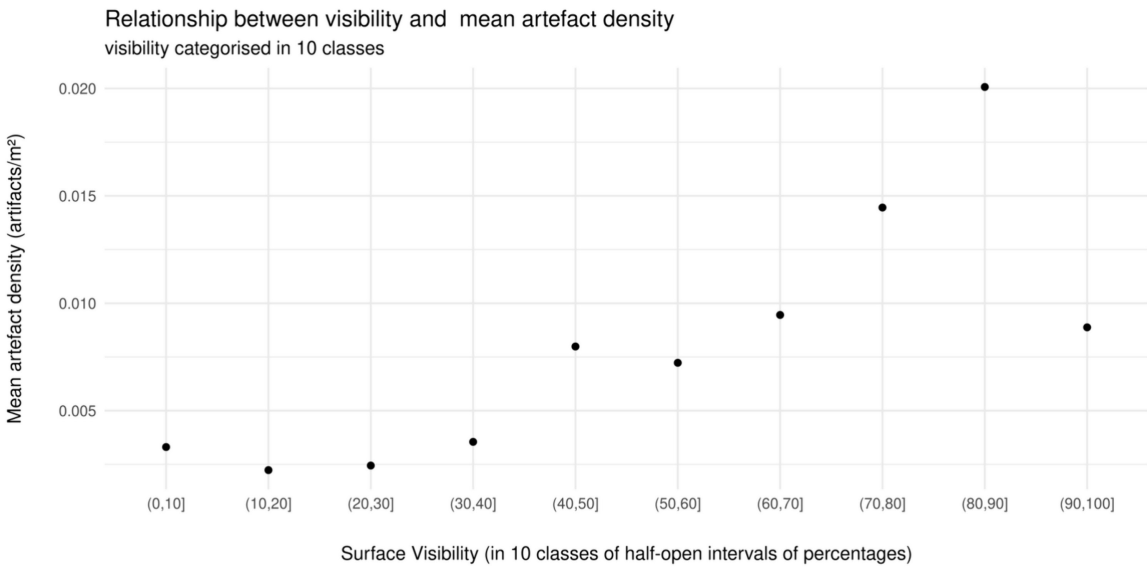

FIGURE 10 | Relationship between ground surface visibility and artefact density for potsherds when visibility is classified into 10 classes. The graphic replicates (Bevan and Conolly 2013, 49 figure 4.1b), and according to the authors, in this representation, the same data would not indicate a linear relationship.

Classical' (900–325 BC) Period and 'Late Roman' (350–650 AD) Period. These periods have been selected because their repartitions are dissimilar, with one cluster in the first period and a broader distribution in the second. The pattern of distribution of the Geometric to Classical pottery is highly similar with our without weighting (Figure 12). If we plot these data aggregated on a hexagonal grid using colours according to a Jenks classification, they are identical (Figure S5). The Hellenistic–Late Roman pottery distribution shows also an identical distribution pattern with our without weighting (Figure 13). If we were to plot using colours according to a Jenks classification, the similarity would be 99% for the period (1081 out of 1092 cells; Figure S6).

4 | Discussion: To Weight or Not to Weight

Visibility is rightly recognized as an influential factor. Who would argue that the cover on the ground surface does not influence what archaeologists see? The three presented survey projects

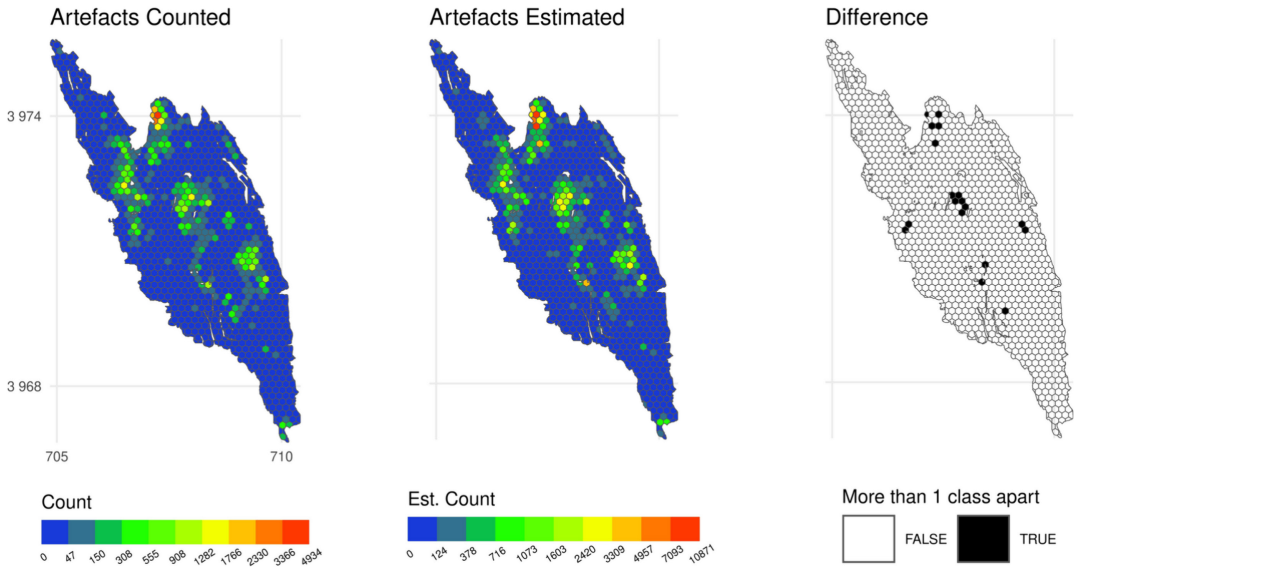

FIGURE 11 | Overall pottery repartition from the ASP without (left), with weighting (right) and difference (right). Pottery density is coloured according to a Jenks classification. Map projection: WGS 84/UTM zone 34N, north-oriented, units of measure in kilometres (EPSG: 32634). [Colour figure can be viewed at wileyonlinelibrary.com]

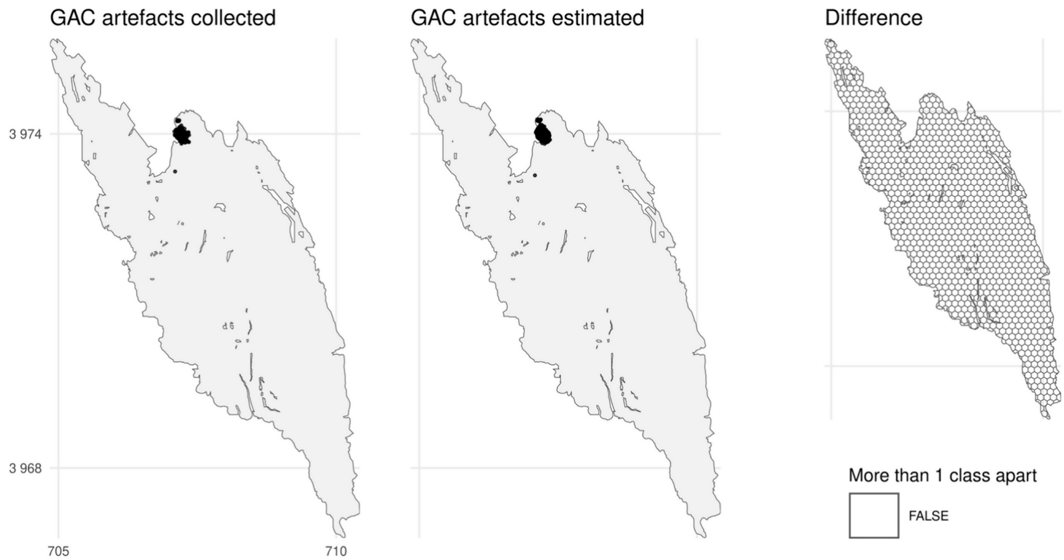

FIGURE 12 | Pottery density (left) and estimated pottery density weighted according to the inverse of visibility (centre) of the Geometric to Classical Period (900–325 BC). The difference (right) was computed according to a Jenks classification (Figure S5). Map projection: WGS 84/UTM zone 34N, north-oriented, units of measure in kilometres (EPSG: 32634).

(PKAP, SCSP and ASP) recorded the visibility for each surveyed unit. However, they all differ in the way they account for it. It was treated separately in a single chapter (PKAP), integrated into the whole interpretation (SCSP) or evaluated but not individually used in the analyses (ASP). They demonstrate concerns over interpreting 'raw' data from the field without looking at the biases introduced by external modern conditions. In this, they account for the definitive move from a purely positivist approach of the earlier years of field survey (see Banning 2002). The influence of the transformation—or its absence—on the interpretation was mainly tackled at the scale of each research project. But what does this mean, in general, and how are we to merge data to answer questions on a bigger scale than the single research project? Few projects that implemented a correction, as well as a few of those that did not, demonstrate how different approaches may have influenced the interpretation of the data (Newhard 2005).

Until now, most implemented weighting techniques involved applying a linear transformation to the whole dataset. Addressing two different weighting methods, our results demonstrate that implementing these transformations may slightly change the data but will not significantly influence the identification of sites or zones of activity based on maps. These straightforward procedures modify the data using an

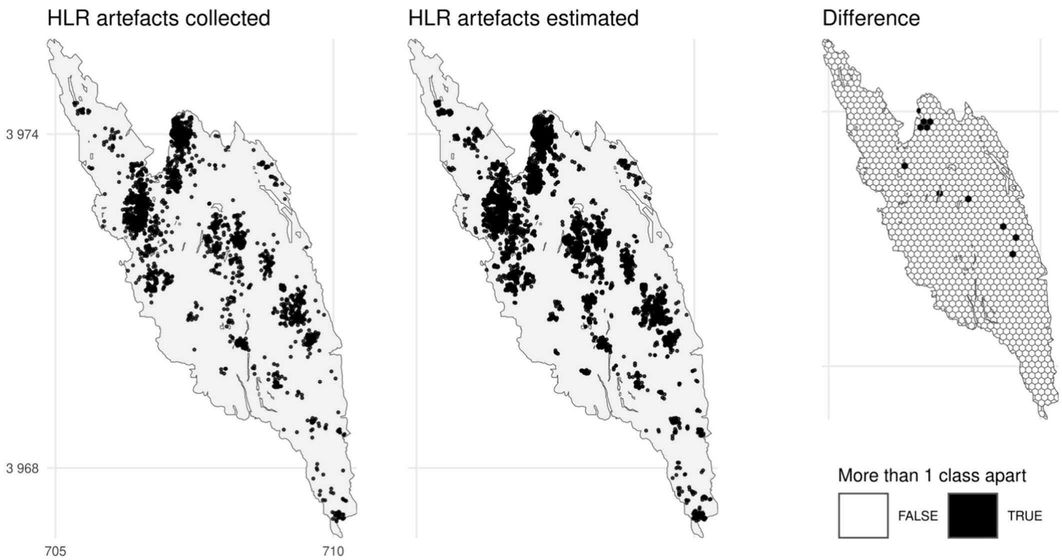

FIGURE 13 | Pottery density (left) and estimated pottery density weighted according to the inverse of visibility (centre) of the Hellenistic–Late Roman Period (350 BC–650 AD). The difference (right) was computed according to a Jenks classification (Figure S6). Map projection: WGS 84/UTM zone 34N, north-oriented, units of measure in kilometres (EPSG: 32634).

TABLE 2 | Summary of class changes when artefacts are weighted.

Case	Project	Collection	Spatial repartition	No. of cells	No. class changes > 1	Percentage of change
1	PKAP	All	Overall	555	27	5%
1	PKAP	Late Bronze Age	Clustered	555	4	1%
1	PKAP	Roman	Dispersed	555	21	4%
2	SCSP	All	Overall	959	1	0%
2	SCSP	Geometric to Classical	Dispersed	959	21	2%
2	SCSP	Hellenistic Late Roman	Clustered	959	12	1%
3	ASP	All	Overall	1092	18	2%
3	ASP	Geometric to Classical	Clustered	1092	0	0%
3	ASP	Hellenistic Late Roman	Dispersed	1092	11	1%

easily comprehensible method, allowing everyone to understand the logic behind it, but they fail to substantially affect the results of site identification (Table 2).

Analysing the three projects left me wondering if a nonlinear transformation might be more successful in significantly transforming the data. Indeed, only a more 'local' or 'zonal' transformation, which considers additional parameters, could potentially adequately weight the data to influence the interpretation of zones of occupation or to define sites. Unlike a transformation that affects the dataset uniformly, a transformation to better identify sites may be more successful if it is based on more highly localized features. Nevertheless, I consider that this would necessitate testing with different data to ascertain its robustness and avoid the risk of overfitting one particular dataset with a transformation.

It is possible to add endless complexity and refinement in evaluating the data and adjust the raw data according to different possible factors (visibility, soil, slope, walker 'characters', etc.). Researchers have already proposed implementations for some of these factors (Banning, Hawkins, and Stewart 2011; Banning et al. 2017; Casarotto et al. 2018; Stark and Garraty 2008). Surveys that may not have experimental data to implement and validate a weighting method, a visibility map, as previously suggested by Terrenato and Ammerman (Terrenato and Ammerman 1996; recently Casarotto et al. 2018), appear to be an easy-to-produce and reliable method to present the ground surface variable. This method clearly presents visibility, geological or other biasing information to determine whether the recognized patterns are attributable to the earth surface conditions. For example, in their publication, Bevan and Connolly plotted the pottery density and vegetation cover on top of each other to give a 'rough impression of surface pottery and surface visibility' (Bevan and Conolly 2013, Plate 3); Casarotto et al. use it to 'asses the constraints that may affect the preservation and the recording of sites' (Casarotto et al. 2018, 188).

5 | Conclusion

It seems ill-advised and oddly positivist to analyse raw results and take their validity for granted. Archaeologists can easily understand the rationale for implementing methods to balance the influence of modern factors such as visibility. Despite multiple claims about the necessity and difficulty of transforming the data, it is striking to observe that the production of corrected maps is scarcer. The data science approach employed in this article argues that projects should test and show the impact of data transformation, independently of the chosen method, and assess how this would modify the interpretation. This is especially valid for projects weighting their results and including it in their pipeline.

In this article, we selected two different time settings to compare the influence of weighting: a timescale including all the recorded artefacts and two cases focused on more sharply defined periods. The former provides a general overview of recurrently more intensely used areas and indicates places where results may have been biased because of later surface changes; the latter is more familiar in interpreting diachronic changes. The analysis of the three projects shows that the weighting method would not influence the identification of sites or zones of occupation.

I suggest that to continue exploring the influence of visibility, it would be beneficial for projects to implement an open DEA to set visibility weighting in a larger frame and control the validity of weighting methods (Banning et al. 2017). It is advisable to use available FAIR data online to assess the significance of the transformation and to keep in mind that data are not only project-relevant but also have more potential if they are integrated. Analysing how each project's data fits into a broader framework and integrating it with Data Science practices can advance regional cross-comparison.

Directly analysing the data, rather than reinterpretation of survey results, is indispensable for merging archaeological data with other disciplines. This approach will significantly improve communication of survey results to scientists from other disciplines and encourage them to take archaeological data more seriously. This is crucial if we are to seriously address the challenges of current climate change (Kerr 2020; Kintigh et al. 2014; Ortman 2019; Smith 2021).

Acknowledgements

I am grateful to the editorial team of Archaeological Prospection and the four anonymous reviewers for the invaluable feedback they provided. In particular, I would like to express my sincere gratitude to Reviewer 3 for the meticulous reading of the manuscript and for taking the time to discuss many aspects of the work. Their insightful comments and suggestions have been invaluable in helping me strengthen my arguments and improve the overall quality of the paper. Open Access funding enabled and organized by Projekt DEAL.

Conflicts of Interest

The author declares that there is no conflict of interest. In the interests of transparency, the author declares to be a member of the Panormos Archaeological Survey (Wilkinson and Slawisch 2017).

Data Availability Statement

The data that support the findings of this study are available for the following: (1) Pyla-Koutsopetria Archaeological Project I at Open Context: https://opencontext.org/projects/3f6dcd13-a476-488e-ed10-47d25513fcb2 (CC BY 4.0); (2) The Antikythera Survey Project at Archaeology Data Service: https://doi.org/10.5284/1024569 (CC BY 4.0); and (3) The Sydney Cyprus Survey Project at Archaeology Data Service: https://doi.org/10.5284/1000208 (ADS Terms of Use and Access). The source code support the findings of this study is available: Strupler, N. (2023). Computational notebook to compare ground visibility in archaeological field-walking surveys (Version 1.0.0) [Computer software]. https://doi.org/10.5281/zenodo.10213978.

References

Alcock, S. E., and J. F. Cherry, eds. 2004. *Side-By-Side Survey: Comparative Regional Studies in the Mediterranean World*. Oxford: Oxbow Books.

Ammerman, A. J. 1993. "Review of J.F. Cherry, J.L. Davis, and E. Mantzourani "Landscape Archaeology as Long-Term History: Northern Keos in the Cycladic Islands"." *Journal of Field Archaeology* 20, no. 3: 367–372.

Ammerman, A. J., and M. W. Feldman. 1978. "Replicated Collection of Site Surfaces." *American Antiquity* 43, no. 4: 734–740. http://www.jstor.org/stable/279505.

Attema, P., J. Bintliff, M. van Leusen, et al. 2020. "A Guide to Good Practice in Mediterranean Surface Survey Projects." *Journal of Greek Archaeology* 5: 1–62. https://doi.org/10.32028/9781789697926-2.

Banning, E. B. 2002. *Archaeological Survey*. New York: Kluwer Academic.

Banning, E. B., A. L. Hawkins, and S. T. Stewart. 2006. "Detection Functions for Archaeological Survey." *American Antiquity* 71, no. 4: 723–742.

Banning, E. B., A. L. Hawkins, and S. T. Stewart. 2011. "Sweep Widths and the Detection of Artifacts in Archaeological Survey." *Journal of Archaeological Science* 38, no. 12: 3447–3458. https://doi.org/10.1016/j.jas.2011.08.007.

Banning, E. B., A. L. Hawkins, S. T. Stewart, P. Hitchings, and S. Edwards. 2017. "Quality Assurance in Archaeological Survey." *Journal of Archaeological Method and Theory* 24, no. 2: 466–488. https://doi.org/10.1007/s10816-016-9274-2.

Barker, G., C. Mee, W. Cavanagh, R. Schon, and S. M. Thompson. 2000. "Responses to 'The Hidden Landscape of Prehistoric Greece', by J.L. Bintliff, P. Howard, and A.M. Snodgrass." *Journal of Mediterranean Archaeology* 13, no. 1: 100–123. https://doi.org/10.1558/jmea.v13i1.29909.

Barton, C. M., J. Bernabeu, J. E. Aura, O. Garcia, and N. La Roca. 2002. "Dynamic Landscapes, Artifact Taphonomy, and Landuse Modeling in the Western Mediterranean." *Geoarchaeology* 17, no. 2: 155–190. https://doi.org/10.1002/gea.10008.

Bevan, A., and J. Conolly. 2004. "GIS, Archaeological Survey, and Landscape Archaeology on the Island of Kythera, Greece." *Journal of Field Archaeology* 29, no. 1–2: 123–138. https://doi.org/10.1179/jfa.2004.29.1-2.123.

Bevan, A., and J. Conolly. 2009. "Modelling Spatial Heterogeneity and Nonstationarity in Artifact-Rich Landscapes." *Journal of Archaeological Science* 36, no. 4: 956–964. https://doi.org/10.1016/j.jas.2008-11-023.

Bevan, A., and J. Conolly. 2012. "Intensive Survey Data From Antikythera, Greece." *Journal of Open Archaeology Data* 1: e3. https://doi.org/10.5334/4f3bcb3f7f21d.

Bevan, A., and J. Conolly. 2013. *Mediterranean Islands, Fragile Communities and Persistent Landscapes: Antikythera in Long-Term Perspective*. Cambridge: Cambridge University Press.

Bevan, A., and J. Conolly. 2014. *The Antikythera Survey Project (Updated 2014)*. York: Archaeological Data Service. https://doi.org/10.5284/1024569.

Bintliff, J. L. 1978. "Off-Site Pottery Distributions. A Regional and Interregional Perspective." *Current Anthropology* 29, no. 3: 506–513.

Bintliff, J. L. 1985. "The Boeotia Survey." In *Archaeological Field Survey in Britain and Abroad*, edited by S. Macready, 196–216. London: Society of Antiquaries.

Bintliff, J. 2013. "*Intra-Site Artefact Surveys*." In *Good Practice in Archaeological Diagnostics*, 193–207. New York: Springer. https://doi.org/10.1007/978-3-319-01784-6_11.

Bintliff, J., E. Farinetti, B. Slapšak, and A. Snodgrass. 2017. *The City of Thespiai: Survey at a Complex Urban Site*. Cambridge: McDonald Institute Monographs.

Bintliff, J. L., and P. Howard. 1999. "Studying Needles in Haystacks—Surface Survey and the Rural Landscape of Central Greece in Roman Times." *Pharos Journal of the Netherlands Institute in Athens* 7: 51–91.

Bintliff, J. L., P. Howard, and A. Snodgrass, eds. 2007. *Testing the Hinterland: The Work of the Boeotia Survey (1989–1991) in the Southern Approaches to the City of Thespiai*. Cambridge: McDonald Institute Monographs.

Birch, C. P. D., S. P. Oom, and J. A. Beecham. 2007. "Rectangular and Hexagonal Grids Used for Observation, Experiment and Simulation in Ecology." *Ecological Modelling* 206, no. 3: 347–359. https://doi.org/10.1016/j.ecolmodel.2007.

Bivand, R. 2022. *classInt: Choose Univariate Class Intervals*. Vienna: R Foundation for Statistical Computing. https://CRAN.R-project.org/package=classInt.

Buckland, S. T., E. A. Rexstad, T. A. Marques, and C. S. Oedekoven. 2015. *Distance Sampling: Methods and Applications*. New York: Springer International Publishing. https://doi.org/10.1007/978-3-319-19219-2.

Burgers, G.-J., P. A. J. Attema, and P. M. van Leusen. 2004. "Walking the 'murge': Interim Report of the Ostuni Field Survey (Apulia, Southern Italy)." *Studi Di Antichità* 11: 257–282.

Caraher, W. R., R. S. Moore, and D. K. Pettegrew. 2013. *Pyla-Koutsopetria Archaeological Project*. Open Context. California: Alexandria Archive Institute. https://doi.org/10.6078/M7B56GNS.

Caraher, W. R., R. S. Moore, and D. K. Pettegrew. 2014. *Pyla-Koutsopetria I: Archaeological Survey of an Ancient Coastal Town*. Boston: American Schools of Oriental Research.

Caraher, W. R., D. Nakassis, and D. K. Pettegrew. 2006. "Siteless Survey and Intensive Data Collection in an Artifact-Rich Environment: Case Studies From the Eastern Corinthia, Greece." *Journal of Mediterranean Archaeology* 19, no. 1: 7–43.

Casarotto, A., J. Garcia Sanchez, T. Stek, and J. Pelgrom. 2021. "Surface Visibility and the Validity of Settlement Patterns in Legacy Survey Datasets." *Fasti Online Documents and Research* 17. http://www.fastionline.org/docs/FOLDER-sur-2021-17.pdf.

Casarotto, A., T. D. Stek, J. Pelgrom, R. H. van Otterloo, and J. Sevink. 2018. "Assessing Visibility and Geomorphological Biases in Regional Field Surveys: The Case of Roman Aesernia." *Geoarchaeology* 33: 177–192. https://doi.org/10.1002/gea.21627.

Cherry, J. F., J. L. Davis, and E. Mantzourani. 1991. *Landscape Archaeology as Long-Term History: Northern Keos in the Cycladic Islands From Earliest Settlement Until Modern Times*. Los Angeles: University of California.

Clark, R. H., and A. J. Schofield. 1991. "By Experiment and Calibration: An Integrated Approach to Archaeology of the Plough Soil." In *Interpreting Artefact Scatters: Contributions to Ploughzone Archaeology*, edited by A. Schofield, 93–105. Oxford: Oxbow Books.

Darkes, G. 2017. "An Introduction to map Design." In *The Routledge Handbook of Mapping and Cartography*. London: Routledge. https://doi.org/10.4324/9781315736822.ch21.

Davis, J. L. 2004. "Are the Landscapes of Greek Prehistory Hidden? A Comparative Approach: Comparative Regional Studies in the Mediterranean World." In *Side-By-Side Survey*, edited by S. E. Alcock and J. F. Cherry, 19–32. Oxford: Oxbow Books.

Davis, J. L., S. E. Alcock, J. Bennet, Y. G. Lolos, and C. W. Shelmerdine. 1997. "The Pylos Regional Archaeological Project Part I: Overview and the Archaeological Survey." *Hesperia: The Journal of the American School of Classical Studies at Athens* 66, no. 3: 391–494.

Dieudonné, N. 1989. "La prospection au sol: Étude bibliographique." *Revue Archéologique Du Centre de La France* 28, no. 2: 217–228. https://doi.org/10.3406/racf.1989.2597.

Gillings, M., P. Hacıgüzeller, and G. Lock, eds. 2018. *Re-Mapping Archaeology: Critical Perspectives, Alternative Mappings*. London: Routledge. https://doi.org/10.4324/9781351267724.

Gillings, M., and K. Sbonias. 1999. "Regional Survey and GIS: The Boeotia Project." In *Geographical Information Systems and Landscape Archaeology*, edited by M. Gillings, D. Mattingly, and J. van Dalen, 35–54. Oxford: Oxbow Books.

Given, M., and A. B. Knapp, eds. 2003. *The Sydney Cyprus Survey Project: Social Approaches to Regional Archaeological Survey*. Los Angeles: Cotsen Institute of Archaeology.

Grunwald, S., K. P. Hofmann, D. A. Werning, and F. Wiedemann, eds. 2018. *Mapping Ancient Identities: Methodisch-kritische reflexionen zu kartierungspraktiken*. Vol. 55. Berlin: Edition Topoi.

Gruškovnjak, L. 2019. "Visibility of Archaeological Record on the Surface." In *Proceedings From the 5th Scientific Conference Methodology and Archaeometry*, edited by I. Miloglav, 57–79. Zagreb: Sveučilište u Zagrebu, Filozofski fakultet, Odsjek za arheologiju. https://doi.org/10.17234/9789531757799.5.

de Haas, T., and M. van Leusen. 2020. "FAIR Survey: Improving Documentation and Archiving Practices Archaeological Field Survey Through CIDOC CRM." *Fasti Online Documents and Research* 12. https://www.fastionline.org/docs/FOLDER-sur-2020-12.pdf.

Hawkins, A. L., S. T. Stewart, and E. B. Banning. 2003. "Interobserver Bias in Enumerated Data From Archaeological Survey." *Journal of Archaeological Science* 30, no. 11: 1503–1512. https://doi.org/10.1016/S0305-4403(03)00051-7.

Hirth, K. G. 1978. "Problems in Data Recovery and Measurement in Settlement Archaeology." *Journal of Field Archaeology* 5, no. 2: 125–131. https://doi.org/10.1179/009346978791489871.

Jenks, G. F., and F. C. Caspall. 1971. "Error on Choroplethic Maps: Definition, Measurement, Reduction." *Annals of the Association of American Geographers* 61, no. 2: 217–244.

Kerr, S. 2020. "The Future of Archaeology, Interdisciplinarity and Global Challenges." *Antiquity* 94, no. 377: 1–12.

Kintigh, K. W., J. H. Altschul, M. C. Beaudry, et al. 2014. "Grand Challenges for Archaeology." *Proceedings of the National Academy of Sciences* 111, no. 3: 879–880. https://doi.org/10.1073/pnas.1324000111.

Knapp, A. B., and M. Given, eds. 2003. *Archive: The Sydney Cyprus Survey Project*. York: Archaeological Data Service. https://doi.org/10.5284/1000208.

Knodell, A. R., T. C. Wilkinson, T. P. Leppard, and H. A. Orengo. 2023. "Survey Archaeology in the Mediterranean World: Regional Traditions and Contributions to Long-Term History." *Journal of*

Archaeological Research 31, no. 2: 263–329. https://doi.org/10.1007/s10814-022-09175-7.

Lock, G., T. Bell, and J. Lloyd. 1999. "Towards a Methodology for Modelling Surface Survey Data: The Sangro Valley Project." In *Geographical Information Systems and Landscape Archaeology*, edited by M. Gillings, D. Mattingly, and J. van Dalen, 35–54. Oxford: Oxbow Books.

Manly, B. F. J., and J. A. N. Alberto. 2015. *Introduction to Ecological Sampling*. New York: CRC Press.

Marwick, B., J. d.'A. Guedes, C. M. Barton, et al. 2017. "Open Science in Archaeology." *SAA Archaeological Record* 17, no. 4: 8–14.

Mattingly, D. 2000. "Methods of Collection, Recording and Quantification." In *Extracting Meaning From Ploughsoil Assemblages*, edited by R. Francovich, 5–15. Oxford: Oxbow Books.

Meyer, N. 2023. "Finding Sites in Mediterranean Survey." *Journal of Mediterranean Archaeology* 35, no. 2: 143–168. https://doi.org/10.1558/jma.25520.

Meyer, N., and R. Schon. 2003. *"Experimental Data."* In *The Sydney Cyprus Survey Project: Social Approaches to Regional Archaeological Survey*, edited by M. Given and A. B. Knapp, 52–57. Los Angeles: Cotsen Institute of Archaeology.

Millet, M. 2000. "Discussion." In *Extracting Meaning From Ploughsoil Assemblages*, edited by R. Francovich, 92–94. Oxford: Oxbow Books.

Newhard, J. M. L. 2005. "Review of the Sydney Cyprus Survey Project: Social Approaches to Regional Archaeological Survey." *Bryn Mawr Classical Review* 05, no. 04 http://bmcr.brynmawr.edu/2005/2005-05-04.html.

Ortman, S. G. 2019. "A New Kind of Relevance for Archaeology." *Frontiers in Digital Humanities* 6. https://doi.org/10.3389/fdigh.2019.00016.

Pebesma, E. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data." *The R Journal* 10, no. 1: 439–446. https://doi.org/10.32614/RJ-2018-009.

Pettegrew, D. K. 2014. "Survey Data and Experiments in Sampling: Archaeological Survey of an Ancient Coastal Town." In *Pyla-Koutsopetria I: Archaeological Survey of an Ancient Coastal Town*, edited by W. R. Caraher, R. S. Moore, and D. K. Pettegrew, 49–68. Boston: American Schools of Oriental Research.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Schiffer, M. B., A. P. Sullivan, and T. C. Klinger. 1978. "The Design of Archaeological Surveys." *World Archaeology* 10, no. 1: 1–28. https://doi.org/10.1080/00438243.1978.9979712.

Schon, R. 2002. *Seeding the Landscape: Experimental Contributions to Regional Survey Methodology*. Pennsylvania: Bryn Mawr College.

Schon, R. 2013. "Seeding Experiments." In *Landscape and Interaction. The Troodos Archaeological and Environmental Survey Project, Cyprus. Volume 1. Methodology, Analysis and Interpretation*, edited by M. Given and H. Corley, 35–39. Oxford: Oxbow Books.

Shennan, S. 1985. *Experiments in the Collection and Analysis of Archaeological Survey Data: The East Hampshire Survey*. Shefield: University of Sheffield.

Smith, M. E. 2021. "Why Archaeology's Relevance to Global Challenges Has Not Been Recognised." *Antiquity* 95, no. 382: 1061–1069.

Stark, B. L., and C. P. Garraty. 2008. "Parallel Archaeological and Visibility Survey in the Western Lower Papaloapan Basin, Veracruz, Mexico." *Journal of Field Archaeology* 33, no. 2: 177–196. https://doi.org/10.1179/009346908791071286.

Stek, T. D., and J. Waagen. 2022. "Scratching the Surface: Integrating Low-Visibility Zones and Large Rural Sites in Landscape Archaeology Using Point Sampling." *Journal of Field Archaeology* 47, no. 4: 228–244. https://doi.org/10.1080/00934690.2022.2031466.

Stewart, D. 2020. "Book Review of Boeotia Project, vol. II: The City of Thespiai. Survey at a Complex Urban Site, edited by John Bintliff, Emeri Farinetti, Božidar Slapšak, and Anthony Snodgrass." *American Journal of Archaeology* 124, no. 2. https://doi.org/10.3764/ajaonline1242.Stewart.

Strupler, N. 2021. "Re-Discovering Archaeological Discoveries. Experiments With Reproducing Archaeological Survey Analysis." *Internet Archaeology* 56. https://doi.org/10.11141/ia.56.6.

Strupler, N. 2023. *Computational Notebook to Compare Ground Visibility in Archaeological Field-Walking Surveys* (Version 1.0.0). Geneva: Zenodo. https://doi.org/10.5281/zenodo.10213978.

Strupler, N., and T. C. Wilkinson. 2017. "Reproducibility in the Field: Transparency, Version Control and Collaboration on the Project Panormos Survey." *Open Archaeology* 3, no. 1: 279–304. https://doi.org/10.1515/opar-2017-0019.

Terrenato, N., and A. J. Ammerman. 1996. "Visibility and Site Recovery in the Cecina Valley Survey, Italy." *Journal of Field Archaeology* 23, no. 1: 91–109. https://doi.org/10.1179/009346996791973990.

Van Leusen, M., and P. Attema. 2001. "Regional Archaeological Patterns in the Sibaritide: Preliminary Results of the RPC Field Survey Campaign 2000." *Palaeohistoria (Haarlem)* 43-44: 397–416. https://ugp.rug.nl/Palaeohistoria/article/view/25118.

Wandsnider, L., and E. L. Camilli. 1992. "The Character of Surface Archaeological Deposits and Its Influence on Survey Accuracy." *Journal of Field Archaeology* 19, no. 2: 169–188. https://doi.org/10.1179/009346992791548914.

Whitelaw, T. 2013. In *Collecting Cities: Some Problems and Prospects*, edited by P. Johnson and M. Millet, 70–106. Oxford: Oxbow Boks.

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer. https://ggplot2.tidyverse.org.

Wilkinson, T. C., and A. Slawisch. 2017. "Panormos 2017: Intensive Survey on the Milesian Peninsula." *Heritage Turkey* 7: 32–33.

Witcher, R. E. 2008. "(Re)surveying Mediterranean Rural Landscapes: GIS and Legacy Survey Data." *Internet Archaeology* 24. https://doi.org/10.11141/ia.24.2.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.