Discussion and Debate:
A Necessary Debate: The Quality and Role of Artifact Data from Archaeological Field Survey

Nathan Meyer

The British School at Athens, Odos Souidias 52, GR 106 76, Athens, Greece
E-mail: nmeyer@bsc.ac.uk

Introduction
The rigor and vitality of a discipline, indeed the ability to make socially meaningful contributions, is founded not solely on the accumulation of observational data but also on principled and reasoned debate concerning theory and method. For that reason, I wish to thank the editorial team at JMA for inviting me to respond to the paper by de Haas, Leppard, Waagen and Wilkinson. To the degree that the authors have missed or misconstrued several of my arguments is no doubt due in part to my own failings in presenting them. Significantly, however, it is also the result of differing intellectual commitments. I address this key, latter point first, and then respond to some of what I consider more minor points.

Artifact Density Data
Simply stated, the respondents (hereafter the Rs) have an intellectual commitment to artifact density data—and to its suitability for aggregation across projects—that I do not share. There are two reasons for my view: (1) I do not think data generated by intensive survey are reliable or comparable, and (2) I do not agree that such data are required to address the research questions that interest the Rs.

On this second point, the Rs echo a point I made in my paper (Meyer 2022: 153): human activity is fluid across the landscape. Whereas I argue this truism need not distract our attention from human place making—the past activity that results in what we call sites—the Rs essentially argue that this is the theoretical justification for measuring artifact density. They argue further that artifact density measures best address certain research questions, for example the expansion and intensification of agricultural production. As they acknowledge, however, my definition of site is ‘any areal expanse with an arrangement or aspect of material culture or environmental features that are likely indicative of—and with the potential to be interpretable as—past human activity, intentional or not’ (Meyer 2022: 152). Thus, there is nothing stopping survey archaeologists from developing a site model that addresses evidence for agricultural production and defining where in the landscape to search for it. In fact, in my approach, such a definition accounts more fully for local conditions. Rather than the building up of sites from density data that are presumed to be accurate and meaningful across disparate locales, researchers would be able to define site types that represent better how intensification and expansion of agricultural production might look in their search area.

Addressing data reliability, the proposed shift in methods is founded entirely on my belief that artifact counts resulting from standard pedestrian survey (i.e. where individuals are spaced several meters apart and visually inspect...
the ground to either side of them) are fundamentally problematic and should only be used internally to a project, and then with extreme care. While acknowledging that there are problems with the data, the Rs assert that we lose too much by giving up on artifact density measures and that instead we should improve our methods, thus making artifact density data more reliable. I won’t reiterate or justify further my position, but instead respond to just one specific comment. The Rs assert that biased data are not necessarily bad data and of course this is true, at least when you are able to understand the degree and direction of the bias. Unfortunately, for reasons I argued (Meyer 2022: 148), we never know—for any given survey unit and for each class of artifact in that unit—how far wrong we are and in what direction. Here, I also register my doubt whether Bayesian exchangeability, as the Rs suggest, applies to this problem.

The Rs are also concerned that my proposal will result in an ‘abdication of duty towards “off-site” finds’ and the ‘discarding’ of high-resolution survey data (not literally to throw away evidence but simply not to have encountered it in the first place) as well as ethical issues caused by ‘ignoring more recent archaeological phases.’ Yet my proposal does not suggest that the archaeological evidence encountered—modern or otherwise—should be treated irresponsibly. That misunderstanding aside, I assume that the Rs believe that by failing to carry out an artifact density survey at high intensity uniformly across the permitted survey area, archaeologists will not encounter the fullness of human activity in the landscape. But surely no survey can claim to capture all human activity in the landscape; the Rs themselves reference the ‘taphonomic variability’ that renders this impossible. Moreover, what is the ‘right intensity’ to perform this pedestrian survey: 15 m spacing, 10 m, 5 m? Does a survey at 15 m likewise result in an abdication of duty when compared to one at 5 m? Doesn’t the right intensity correspond somehow to discovery objectives and locally variable confounding factors? It is exactly such issues as these that lead me to a different intellectual commitment. Quantitative data must surely play a role internal to the project (e.g., in bounding space, in suggesting the presence, intensity and duration of varying human activities). My argument, however, is that such data are supplementary to qualitative data gathered through adaptive sampling. Furthermore, I would argue that artifact density data gathered by consistent intensity across highly variable surfaces are simply not good data.

The Rs do offer a scenario where they believe density data prove their merits: specifically, evaluating the scale of the ceramic-producing economies as indexed by the density of surface ceramic data. However, in my view it is questionable whether such an analysis can be arrived at only by this route, whether that analysis warrants having all projects assume the cost of producing such data and indeed whether the data actually produced by all these projects are sufficiently reliable for such analysis.

To summarize our different intellectual commitments: in the Rs’ preferred model, the archaeologist would commit further to gathering and publishing density data across the survey universe. The fact that 10 m spacing in one location will have a significantly different discovery potential than that same spacing in another location is a matter they believe can be resolved by improving our methods. In my model, in contrast, density data are simply not reliable and thus findings would always be presented in terms of changes in site types and counts. My position is founded on the belief that while I cannot reliably evaluate your density data and vice versa, we can reliably evaluate each other’s site definition models, search techniques and quality assurance programs.

Quality Assurance

Regarding quality assurance, the Rs believe that, with effort, we can overcome problems with
artifact density data. Within the limits of their response, they are not fully able to explain how this might happen; even if it could, however, it would only be through much more rigorous quality assurance programs. Although the Rs allow that ‘there should be more attention to quality assurance in archaeological field survey’, they do not acknowledge that this is not actually happening. In my view, the goal of quality assurance is not to overcome the factors that make density data problematic, as this is simply not possible. Instead, quality assurance is needed to ensure that our search for a predefined target is rigorous and, likewise, to prevent us from making unsupported arguments from the absence of evidence (a point that I made [Meyer 2022: 150] but that is not addressed by the Rs). In my proposed method, quality assurance is not merely something to which we should pay more attention; rather, it is fundamental to executing a rigorous search.

**Empowering Teams**

Executing a rigorous search necessarily relies upon adaptive methods drawn from a quality assurance program and deployed as conditions dictate on a search-by-search basis. As the Rs point out, this requires decision making in the field by the field team—a prospect which the Rs believe would ‘likely put too heavy a strain on most survey team leaders.’

By way of analogy, I would point out that a similar argument was made about the chronotype collection method (Meyer et al. 2003; Gregory 2004), namely that it ‘force[s] very high-level interpretative decisions (e.g., about what exact kinds of pottery styles and fabrics are present) to be made very early on in the field’ (Bevan et al. 2013: 53). Experience with this method, however, has shown that in terms of the quality of chronological and functional data, the chronotype system is efficient and robust (Caraher et al. 2014: 68; Tartaron et al. 2006: 487). Concerns that field walkers cannot collect accurately using the one-of-each-kind-method simply do not hold up to scrutiny; when properly trained, student volunteers are fully capable of collecting in this manner.

Any collection technique can be successful when the survey teams that use it are trained on how to execute it. While adaptive sampling is more complex than standard fieldwalking, removing burden from our field teams should not be accomplished by removing opportunity for engaged participation. Survey leadership does not need to rely on ‘hope’ (as the Rs suggest I do) but can instead provide adequate training and tools. It is a maxim of our age that diversity and inclusivity improve overall team performance; we should be fully invested in making that a reality on field projects.

By way of summary up to this point, all the above can be said to apply even if archaeologists chose not to proceed from a predictive model but instead to use somewhat standard pedestrian survey techniques. Archaeologists can improve the discovery potential of pedestrian survey by conducting quality assurance. They might also provide training and tools to field staff so that adaptive sampling rather than ‘site revisits’ is the primary (though not exclusive) method for gathering more information at locations of interest. Still, I would argue, the artifact density data thus generated are not reliable and thus not the correct external deliverable.

**Other Misunderstandings**

There are other misunderstandings—perhaps stemming from my own presentation—that merit clearing up. The Rs argue that I conflate high-intensity survey with siteless survey methods and seem to think I am against high-intensity survey and its benefits (e.g., disentangling functional and chronological components of heavily occupied sites, understanding better how depositional process may be affecting a site). This is not the case, and my proposed method calls for high-intensity survey where
it matters. I am, however, against the uniform application of high-intensity survey across the survey universe in the belief that accurate density data may result. My argument is that we need to shed the concept of consistent intensity—whether extensive or intensive—and instead move to flexible, iterative, multi-stage methods of exploring a region using the lowest-cost variable intensity.

In some cases, the Rs ascribe to me arguments that I was not attempting to make. For example, I do not ‘assume that “intensive” survey necessitates moving away from defining sites in the landscape.’ In fact, I note the paradox of using the artifact as the unit of discovery and yet still having survey publications replete with sites (Meyer 2022: 152). Likewise, they suggest that I argue for ‘intuiting [sites] on the basis of visual inspection’, whereas there is nothing in my proposed methodology that advocates any such intuition. Rather, the proposed method argues for rigorous exploration that, if needed, can be inspected fully by colleagues. This is the true basis of comparability.

In other cases, the Rs appear to believe I have a ‘positivist’ understanding of the complexity of field survey and the difficulties in identifying sites, or that my approach to sites is contradictory. For example, they assert that I am ‘overconfident’ in the use of site data for comparative purposes. What the Rs miss in this case, however, is that I am not arguing solely for comparing site data. I am arguing for the superior comparability of a coherent and testable package of method and data (site models, quality assurance programs, data recovered from carefully applied search techniques).

The Rs also seem to believe that I am abandoning a ‘formal process’ for defining sites and that I seem to be equating sites simplistically with settlements. On the contrary: what I argue for is the rigorous definition of a site model—settlements and otherwise—using consistent and explicit definitional logic (Meyer 2022: 152-53); this is exactly the kind of work that is needed to help us deal with complexity in comparing sites. Nothing in my proposed method is going to unburden the archaeologist from the hard work of drawing sound conclusions from complex evidence. What it would do, however, is make those conclusions more transparent and thus comparable.

Additionally, the Rs seem to believe that my proposal implies ‘that regionally distinct sampling and visibility biases could be disregarded.’ This is puzzling to me, as I argue on the one hand that these confounding factors render density data suspect, and, on the other, for carefully tuned, variable search techniques designed to overcome visibility issues.

Finally, sites are not, as the Rs state at one point, ‘simply places around which archaeologists fixate because of a concentration of residues’. Rather, sites are, as they state at another point, ‘a crucial unit of analysis that helps us to simplify the complexity of past remains.’ As I wrote (Meyer 2022: 152), the actual unit of discovery conditions what we observe. My proposal is that we put a site model at the heart of our methodology, not individual artifacts.

Conclusion

I am pleased that the respondents recognize that Mediterranean survey archaeologists have provided ‘remarkably little explicit response’ to external critiques and, further, that the purpose of my paper was to invigorate the discussion of survey objectives and methods. I take their comments to have the same purpose and I welcome the debate. Most especially, I trust that the clarifications I offer here have removed any confusion regarding abdication of duty and possibly unethical practices.

To end on a note of agreement, in their article the Rs take the opportunity to advocate for FAIR data. The FAIR data concept was developed to ensure that scholarly data are accessible and usable both by the individual researcher and by machines (Wilkinson et al. 2016). FAIR
data thus underpin future efforts to combine data from multiple projects. To be sure, the Rs’ advocacy for FAIR data lies partially in the hope that archaeologists can aggregate density data from multiple projects and apply quantitative tools. I too would advocate for FAIR data, certainly so that archaeologists can aggregate site data but additionally so that colleagues can test whether conclusions drawn from the site model and the data collected are in fact defensible and repeatable (as cited for the Sydney Cyprus Survey Project [Given and Knapp 2003], on which I served as database designer and GIS analyst—Strupler 2021). Thus, on this important point regarding FAIR data, we are in full agreement: all data and necessary metadata must be published on open-access platforms.

References

Bevan, A., J. Conolly, S. Colledge, C. Frederick, C. Palmer, R. Siddall and A. Stellatou

Caraher, W.R., R.S. Moore and D.K. Pettegrew

Given, M., and A.B. Knapp

Gregory, T.E.

Meyer, N.

Meyer, N., T.E. Gregory and R. Schon

Strupler, N.
2021 Re-discovering archaeological discoveries. Experiments with reproducing archaeological survey analysis. *Internet Archaeology* 56. Online. https://doi.org/10.11141/ia.56.6

Tartaron, T.F., T.E. Gregory, D.J. Pullen … W.R. Caraher and D.K. Pettegrew

Wilkinson, M., M. Dumontier, I. Aalbersberg … J. Zhao and B. Mons