A new tool to increase the robustness of archaeological field survey

Jitte Waagen [ID] *based on peer reviews by* *Tymon de Haas* *and* *Philip Verhagen*

Cite this recommendation as:
Waagen, J. (2024) A new tool to increase the robustness of archaeological field survey. *Peer Community in Archaeology*, 100351. 10.24072/pci.archaeo.100351

This well-written and interesting paper 'Survey Planning, Allocation, Costing and Evaluation (SPACE) Project: Developing a Tool to Help Archaeologists Conduct More Effective Surveys' deals with the development of a 'modular, accessible, and simple web-based platform for survey planning and quality assurance' in the area of pedestrian field survey methods (Banning et al. 2024).

Although there have been excellent treatments of statistics in archaeological field survey (among which various by the first author: Banning 2020, 2021), and there is continuous methodological debate on platforms such as the International Mediterranean Survey Workshop (IMSW), in papers dealing with the current development and state of the field (Knodell et al. 2023), good practices (Attema et al. 2020) or the merits of a quantifying approach to archaeological densities (cf. de Haas et al. 2023), this paper rightfully addresses the lack of rigorous statistical approaches in archaeological field survey. As argued by several scholars such as Orton (2000), this mainly appears the result of lack of knowledge/familiarity/resources to bring in the required expertise etc. with the application of seemingly intricate statistics (cf. Waagen 2022). In this context this paper presents a welcome contribution to the feasibility of a robust archaeological field survey design.

The SPACE application, under development by the authors, is introduced in this paper. It is a software tool that aims to provide different modules to assist archaeologists to make calculations for sample size, coverage, stratification, etc. under the conditions of survey goals and available resources. In the end, the goal is to ensure archaeological field surveys will attain their objectives effectively and permit more confidence in the eventual outcomes. The module concerning Sweep Widths, an issue introduced by the main author in 2006 (Banning 2006) is finished; the sweep width assessment is a methodology to calibrate one's survey project for artefact types, landscape, visibility and person-bound performance, eventually increasing the quality (comparability) of the collected samples. This is by now a well-known calibration technique, yet little used, so this effort to

make that more accessible is certainly laudable. An excellent idea, and another aim of this project, is indeed to build up a database with calibration data, so applying sweep-width corrections will become easier accessible to practitioners who lack time to set up calibration exercises.

It will be very interesting to have a closer look at the eventual platform and to see if, and how, it will be adapted by the larger archaeological field survey community, both from an academic research perspective as from a heritage management point of view. I happily recommend this paper and all debate relating to it, including the excellent peer reviews of the manuscript by Philip Verhagen and Tymon de Haas (available as part of this PCI recommendation procedure), to any practitioner of archaeological field survey.

References:

Attema, P., Bintliff, J., Van Leusen, P.M., Bes, P., de Haas, T., Donev, D., Jongman, W., Kaptijn, E., Mayoral, V., Menchelli, S., Pasquinucci, M., Rosen, S., García Sánchez, J., Luis Gutierrez Soler, L., Stone, D., Tol, G., Vermeulen, F., and Vionis. A. 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., Alicia L. Hawkins, S.T. Stewart, Sweep widths and the detection of artifacts in archaeological survey, Journal of Archaeological Science, Volume 38, Issue 12, 2011, Pages 3447-3458. `https://doi.org/10.1016/j.jas.2011.08.007`

Banning, E.B. 2020. Spatial Sampling. In: Gillings, M., Hacıgüzeller, P., Lock, G. (eds.) Archaeological Spatial Analysis. A Methodological Guide. Routledge.

Banning, E.B. 2021. Sampled to Death? The Rise and Fall of Probability Sampling in Archaeology. American Antiquity, 86(1), 43-60. `https://doi.org/10.1017/aaq.2020.39`

Banning, E. B. Steven Edwards, & Isaac Ullah. (2024). Survey Planning, Allocation, Costing and Evaluation (SPACE) Project: Developing a Tool to Help Archaeologists Conduct More Effective Surveys. Zenodo, 8072178, ver. 9 peer-reviewed and recommended by Peer Community in Archaeology. `https://doi.org/10.5281/zenodo.8072178`

Knodell, A.R., Wilkinson, T.C., Leppard, T.P. et al. 2023. Survey Archaeology in the Mediterranean World: Regional Traditions and Contributions to Long-Term History. J Archaeol Res 31, 263–329 (2023). `https://doi.org/10.1007/s10814-022-09175-7`

Orton, C. 2000. Sampling in Archaeology. Cambridge University Press. `https://doi.org/10.1017/CBO9781139163996`

Waagen, J. 2022. Sampling past landscapes. Methodological inquiries into the bias problems of recording archaeological surface assemblages. PhD-Thesis. `https://hdl.handle.net/11245.1/e9cb922c-c7e4-40a1-b648-7b8065c46880`

de Haas, T., Leppard, T. P., Waagen, J., & Wilkinson, T. (2023). Myopic Misunderstandings? A Reply to Meyer (JMA 35(2), 2022). Journal of Mediterranean Archaeology, 36(1), 127-137. `https://doi.org/10.1558/jma.27148`

Reviews

Evaluation round #2

DOI or URL of the preprint: `https://doi.org/10.5281/zenodo.10681567`
Version of the preprint: 7

Authors' reply, 21 March 2024

If I understand correctly, this round it was just necessary to add the short sections at the end indicating the repository and the link to the methods section, which we've now put separately on Zenodo.

The latest version is now on Zenodo.

Decision by Jitte Waagen ⬤, posted 05 March 2024, validated 06 March 2024

minor textual revisions

Dear authors,

As you can see, both reviewers responded positively to your adaptations. Reviewer 2 suggests some minor textual revisions. If you process these, I will take the paper into final consideration for recommendation.

Best wishes,

Jitte

Reviewed by Philip Verhagen, 01 March 2024

I thank the authors for their thoughtful consideration of my feedback, most of which has been incorporated in the revised version of the paper. Where it regards the points of long-term sustainability and the broader debate on how to convince archaeologists to include these approaches in their survey design, I understand that the limitations of the paper do not allow for more elaborate discussion - though it might still be worth considering in more detail in a different paper. I therefore recommend publication of the revised version as is.

Reviewed by Tymon de Haas, 05 March 2024

In this version of the paper, the authors have addressed several of the issues highligted in my first review (moving the section on why, some modifications and additions to the text). In some cases they have addressed my questions in their written response but not in the text (for example my question regarding the recommended sweep width being in meters - I understand now that this does not mean 1m, 2m, ... but I could still imagine some other readers misunderstand this in the same way that I did). even so, the edits have improved the manuscript and I only have a few additional textual edits:

-lines 24 and 25 might be connected more fluetly by slightly altering line 25 "The aim of the Survey Planning, Allocation, Costing and Evaluation (SPACE) Project is to create **this** online platform whose interacting modules will....

-lines 44-48 could indeed well go into a footnote.

-line 182 "we can use these either provide" seems odd

-line 229: , "making is usable"? seems odd

.

Evaluation round #1

DOI or URL of the preprint: `https://doi.org/10.5281/zenodo.8087238`

Version of the preprint: 3

Authors' reply, 20 February 2024

Download author's reply

Reviewed by Philip Verhagen, 21 July 2023

This paper discusses the implementation of an easy-to-use interface for statistically supported archaeological survey design. As pointed out in the introduction, despite a substantial body of literature and empirical evidence regarding the importance of statistical theory in survey design, this is still insufficiently applied in practice which can lead to undesired outcomes, both in terms of archaeological efficacy as well as in efficient budget allocation. Any project that contributes to improving current practice is thus laudable, and the SPACE project seems to be the most ambitious effort that I have come across in this respect.

I will start my review with some feedback on the paper's structure: because of its substantial introduction to the SPACE project in lines 1-68, it is not immediately clear to the reader that the paper in fact mainly discusses a single module that is still under development - the title also seems to imply that the paper reports on the full project. It is of course necessary to provide some background, but the two flowcharts in Figure 1 and 2 make it seem as if there already is a fully developed system, whereas in practice there is a blueprint, and a first module. I would therefore suggest to discuss the outlook of the project in more detail in a discussion section at the end, and to clearly state the aims of the paper in the introduction.

In the introduction, the term 'survey' is introduced without proper definition. I assume the SPACE project is supposed to cover all types of archaeological survey (field walking, core sampling, test pitting, perhaps even trial trenching?), but the paper is only concerned with field walking, so make sure to keep this consistent in the text.

The motivation for the project is provided by stating that existing knowledge is not used to advantage because it is 'too difficult, makes too little difference, or requires math'. I would love to see more evidence for this, also because the 'too difficult' and 'requiring math' are clearly related. However, this may not be the only issue involved, since archaeologists are more than happy to use other 'difficult' techniques and tools that they don't understand in full detail when they perceive them to be useful. In my own experience in the Netherlands, we encounter similar issues with evidence-based guidelines for core sampling and trial trenching not being used effectively (https://www.sikb.nl/doc/archeo/leidraden/KNA%20Leidraad%20IVO%20karterend%20booronderzoek%20definitief_04122012%20v%202.0.pdf, https://www.sikb.nl/doc/archeo/leidraden/KNA%20Leidraad%20proefsleuvenonderzoek%20definitief_04122012%20v%201.02.pdf). A recent evaluation of these guidelines (that basically provide a decision tree for finding the optimal survey strategy) showed that these are considered to be insufficiently realistic in practice, too idealistic, and needlessly restricting archaeologists when designing their surveys. So, there probably is a bit more to say about the psychological aspects of this. It is not just about providing easy-to-use tools, but also about whether using these for decision-making is considered acceptable. In that sense, I am curious to know what will be the strategy of

the project of achieve its aims - you mention two major target groups in line 203-210, but history shows that it needs a lot of effort to convince people to actually use these tools, so how are you going to approach this?

The core of the paper then discusses the module for calculating optimal sweep width in field walking survey. This I found really interesting, not having read Banning et al.'s 2011 paper before, since it based on the systematic collection of reference data to feed the detection model, which has always been a weak point in published models (see for example this article I published in 2013: 10.1016/j.jas.2012.05.041).

The topic is introduced with a discussion of the mock surveys before explaining the exponential model discussed in lines 110-146. Please add a few lines after line 76 to explain why such an empirical approach is needed. In lines 89-92 you mention a number of field data collection apps, it would be good to provide links to these. Then, crucially, you fail to clearly explain the concept of the elipses (lines 96-100), while this is central to the approach. Please make sure that this is explained in more detail (why not circles, what is the size of the elipses and why?). The example given in Figure 3 is based on a forensic search case study - is this one that you carried out yourselves? While the principles are of course the same, it would be nicer to show the results for 'real' archaeological artifacts. Finally, it was not completely clear to me if the exponential function was chosen on the basis of the data collection results or not. The detection function itself uses the k parameter as a catch-all for effects influencing detection detection probability - I assume that this is calibrated on the basis of the results of the mock surveys in Jordan and Cyprus, and not calculated on the basis of each of the contributing factors?

The idea to set up an Open Access database for further reference data collection is really good. However, how are you going to convince colleagues to set up these experiments all over the world? Can you be a bit more specific about what such a mock survey set up would entail in terms of effort? Otherwise, you might run into more complaints about the approach not being realistic in practice.

Finally, you provide a strategy for sustainability of the tools by relying on GitHub facilities, but long-term Open Access is not necessarily the same as long-term maintenance. In this context it may also be useful to refer to the Dig It, Design It and Dig It, Check It tools developed Amy Mosig Way, since only five years after publication these already seem to be offline (10.1016/j.jasrep.2018.06.034, 10.1016/j.jasrep.2018.07.007, 10.1016/j.dib.2018.08.131). I would appreciate some more of your thoughts on this aspect.

All in all, I found this an interesting paper, generally well written, but I think it will profit from extending the discussion on the necessity and implementation of the tools in practice, as well as on the long-term strategy for maintenance and further development. Also, the explanation of the mock survey analysis needs a bit more detail. And, as indicated above, I recommend to adapt the paper's structure somewhat.

I found no language errors in the text, but the paper structure is not completely following the PCI guidelines - personally, I don't have any strong issues with this, but it may be good to give this a final check.

Reviewed by Tymon de Haas, 19 July 2023

The paper introduces the aims of the SPACE-project; as this project is on-going, the paper presents only a preliminary discussion of one of its parts, the Sweep Width module that can aid survey designers to efficiently set-up their field methods through calibrating for field surveyor performance/artefact retrieval rates. While I fully agree on the usefulness of this module (and the wider project),I think the arguments to 'sell' the use (need?) for this kind of tool in designing surveys can be strenghtened further:

1) in the introduction it is stated that archaeologists have for various reasons not used the available statistical theory to aid designing surveys. it would be useful to explain/argue why this is problematic (financially, logistically, scientifically, ...). perhaps by reference to some practical examples of high cost surveys, cases where aims could not be fully realised, ...? I note that the very short Why do this section could actually be embedded in the introduction to strengthen the rationale there.

2) where it comes to quality control over survey data and more realistic approximation of artefact density estimations, the discussion of the Sweep width module is convincing. It does not yet convincingly argue the suggested use of the module for adapting the general survey strategy: isn't the kind of archaeology one will

map, especially in surveys that are primarily interested in ceramic sites (ADABS, POSIs, ...) more dependent on the interwalker distance than on the sweep width? And if the module will propose a sweep width rounded of to Meters (if I read lines 136/137 correctly), the question also is whether the variations in sweep width are such as to really significantly deviate from current practices (which assume a sweep width of, say, 2 meters). In order to clarify these points and/or to better 'sell' the benefits I would suggest to elaborate further on these points and perhaps illustrate the use of the module/through discussion of one or two examples from the author's work test on Cyprus and in Jordan: how have the actual survey practices (sweep widths/walker spacings) been modified based on the calibrations?

3) a practical point: for surveyors to adopt the calibration procedure much will depend on the necessary time investment in relation to the time available for the survey as a whole. How much time did the actual calibration tests take in the field?

4) The suggestion to have a database of calibration data (lines 170-176) available will probably for many be the preferred (less time-consuming) option. Do you foresee that with more of such data you could provide a set of more general recommendations regarding Sweep width in relation to typical survey conditions?

a few other points:

-should the calibration besides accounting for variations in land use/visibility not also use artefact density as a variable, as overall density will variably affect the capacity of wlakers to obsereve different kinds of artefacts?

-The SPACE project needs a slightly more elaborate introduction: explain in the text what the acronym stands for, who are participating, for how long does the project run, ...

-fig 3 heading (2. locations....) should be removed and caption reproduces sections of text; fig 6 caption refers to "c" that is not included.