Archaeological field surveys, especially of the non-site or off-site kind, aim to produce a detailed, fair and complete record of the archaeological remains that are present on the land surface. However, all practising survey archaeologists agree that many factors conspire to reduce the representativity of the samples that are collected.

The experimental study of these factors began in the early 1980s, with Stephen Shennan’s attempts in East Hampshire to assess the reliability of his surface collections (Shennan 1985). Two other noteworthy studies, both conducted in the late 1990s, are Robert Schön’s PhD research in Greece (Schön 2002) and Edward Banning’s experimental work on detection curves (Banning 2006).

Unfortunately the lessons of these studies cannot be easily applied to other existing regional field survey databases, such as the ones produced over the past 15 years by the Groningen Institute of Archaeology in central and southern Italy. The main reason for this is the fact that survey protocols (for both sampling and documentation) are not yet sufficiently standardized to allow a direct comparison across projects.

In order to determine the reliability of our own survey datasets, we felt the need to experimentally test some of the basic assumptions and parameters underlying modern extensive and intensive field walking surveys. We believe that experiments such as these will also contribute towards solving the problem of interregional and international comparability of field walking datasets.

Preliminary Results

Q: What proportion of the surface artefact assemblage is picked up during standard surveys?
A: Between 50 and 75% of Ware classes 1, 2 and 4 are picked up, despite Ware 4 being much rarer than 1 and 2. [Data for Ware 5 are irrelevant here]

Q: What proportion of the surface artefact assemblage is picked up as a function of the distance to the central line?
A: Contrary to expectation, this proportion is the highest between 5 and 10 cm to either side of the central line, possibly as a result of the ‘peninsula scan’ used by most walkers.

Q: Do trained and untrained walkers pick up different sizes of artefacts?
A: Yes, untrained walkers pick up higher proportions of all size classes except the largest (size 5 – diameter >10 cm) probably because they walk slightly slower and their mouth width is smaller.

Q: How quickly do untrained walkers acquire sufficient experience to be able to take high-quality samples?
A: The graph below shows one inexperienced walker learning to recognize and avoid two categories (Wc3 and Wc4) of increased difficulty, during one week of survey.

Further Work

The analysis of the data collected in July 2014 has only just begun. We will present a more thorough analysis at the International Mediterranean Survey Workshop, November 7-8, at Groningen University. A further set of desktop experiments will be carried out on existing GIA survey databases by Ms. Witmer to test assumptions about site detection theory, as well as to evaluate strategies for on-site collection and for measuring the variability of site assemblages.

The results will submitted for publication in the Journal of Field Archaeology.