An introduction to the quantification of assemblages of pottery

Clive Orton

1. Introduction

The aim of this paper is to help excavators to decide whether, and how, to quantify pottery from excavations. It does not attempt to say what else they should or should not do. For example, useful evidence can be gained from a study of conjoining sherds (Moorhouse 1986, 88–100), a topic outside the scope of this paper. The aspect of quantification examined is that of saying *how much* pottery of different 'types' is present in various assemblages. Nothing will be said about the definition of 'types'; I shall assume this question has been settled already. The idea of 'type' as used here refers to aspects that relate to a whole vessel (e.g. fabric, form) and not to only part of a vessel (e.g. decoration). In other words, two sherds from the same vessel belong to the same 'type'.

2. Why quantify?

There seem to be three main uses to which archaeologists put quantified information about assemblages:

(a) chronological (e.g. seriation, see for example Carver 1985),
(b) spatial (e.g. for fall-off patterns from a production centre, see Hodder 1974),
(c) functional/social (e.g. for defining 'activity areas', see Millett 1980a, or for examining differences in status within a town, e.g. Redman 1979).

3. What to quantify?

Any assemblage of pottery *could* be quantified, but in practice one should quantify only if there are definite reasons for doing so. We have to decide which assemblages are worth quantifying, and why, bearing in mind that quantification, like any other archaeological task, has its costs. They may be tangible, like the cost of paying someone to do it and providing space and equipment, or less tangible, like the opportunity cost of not doing something else because you are busy quantifying pottery.

The criteria for your decision appear to be:

(a) the size of the assemblage – is it large enough to yield reliable information?

(b) the nature of the assemblage – is it primary or secondary? does it appear to have accumulated over a short or long period?
(c) the need for quantified information – are other assemblages available for comparison?

There are no rigid answers to these questions. Generally, one will want to quantify large primary groups that can be compared with other groups on a local or regional scale, but at times there may be good reasons for looking at smaller or less coherent groups. One's first quantified group is an act of faith.

4. How to quantify

To be able to say how much pottery we have, we need what in mathematical terms is called a *measure* of the quantity of pottery. There are four measures in common use:

(a) sherd count,
(b) weight, or a related measure, such as surface area, displacement volume, the table space occupied by closely-packed, non-overlapping sherds, or the volume of a 'heap' of sherds,
(c) vessels represented, i.e. the number of vessels, sherds from which are present in an assemblage. Because of the difficulty of telling whether non-joining sherds belong to the same vessel, the terms 'minimum vessels' (sherds are assumed to belong to the same vessel unless they can be shown to belong to different ones) and 'maximum vessels' (sherds are assumed to belong to different vessels unless they can be shown to belong to the same one) are sometimes used,
(d) vessel equivalents. Each sherd is a fraction of the vessel from which it comes (e.g. 1%). If we could measure these fractions for each sherd and add them up, we would have a number equivalent to the number of vessels that could be made up if all the sherds belonged to the same vessels.

There are two fundamental differences between measures (a) and (b) on one hand and (c) and (d) on the other:

(i) measures (c) and (d), in different ways, give individual vessels a count of '1', while (a) and (b) take account of the different amounts of material in
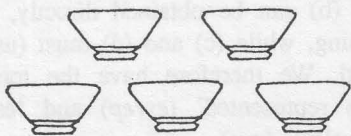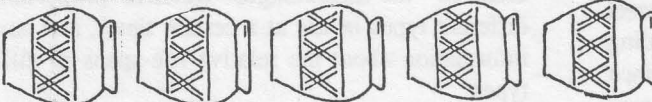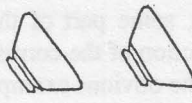
ratio   bias

life-assemblage

death-assemblage

	coarse-ware	fine-ware	ratio	bias
			1:1	
			2.5:1	
1st breakage	each → 10sh	each → 5sh		
	50sh, 5kg	10sh, 0.4kg		
retrieval (i) 50%:	25sh, 2.5kg	5sh, 0.2kg		
(ii) 10%:	5sh, 0.5kg	1sh, 0.04kg		
quantification				
(a) sherd (i)	25	5	5:1	2
count (ii)	5	1	5:1	2
(b) weight (i)	2.5	0.2	12.5:1	5
(ii)	0.5	0.04	12.5:1	5
(c) evreps (i)	5	2	2.5:1	1
(ii)	3	1	3.0:1	1.2
(d) eves (i)	2.5	1	2.5:1	1
(ii)	0.5	0.2	2.5:1	1
2nd breakage	each sh → 10sh	each sh → 5sh		
	500sh, 5kg	50sh, 0.4kg		
retrieval (i) 50%:	250sh, 2.5kg	50sh, 0.2kg		
(ii) 2%:	10sh, 0.1kg	1sh, 0.008kg		
quantification				
(a) sherd (i)	250	25	10:1	4
count (ii)	10	1	10:1	4
(b) weight (i)	2.5	0.2	12.5:1	5
(ii)	0.1	0.008	12.5:1	5
(c) evreps (i)	5	2	2.5:1	1
(ii)	4	1	4:1	1.67
(d) eves (i)	2.5	1	2.5:1	1
(ii)	0.1	0.04	2.5:1	1

Fig. 1: hypothetical `histories' of 5 coarse-ware and 2 fine-ware
pots, from deposition to retrieval and quantification. The `bias'
column is relative to the death assemblage, and 1 = unbiased.
Note the behaviour of the figures for bias:
(i) sherd count figure varies between breakage rates, but not
    within them,
(ii) weight figure does not vary (equals relative weights of
     types),
(iii) evrep figure varies both within and between breakage rates,
      but is unbiased for high retrieval percentages,
(iv) eves figure is unbiased throughout.

different pots, e.g. a large vessel weighs more than a small one, and is likely to break into more sherds,

(ii) measures (a) and (b) can be obtained directly, by counting or weighing, while (c) and (d) must (usually) be estimated. We therefore have the terms 'estimated vessels represented' (*evrep*) and 'estimated vessel equivalent' (*eve*).

To estimate vessel equivalents, some part of the vessel which can be measured as a fraction of the corresponding whole must be used. Rims are the obvious example, since the size of a rim sherd can be expressed as a percentage of a complete rim by the use of a simple chart. Bases can be similarly measured (though often with more difficulty), and other easily recognisable features (e.g. handle roots) can be useful for some sorts of pottery. The choice will depend on the nature of the pottery; it is worth testing for consistency (e.g. between rims and bases) before deciding. My general preference is for an average of rims and bases.

Discussion about the merits of these measures has gone on for at least 25 years (see for example Solheim 1960; Glover 1972; Egloff 1973; Hulthen 1974; Hinton 1977; Millett 1980b). However, much of the work has seemed to look for measures which best agree with the archaeologist's intuitive preconceptions; a more explicit statement of criteria, and objective comparison of measures with them, are needed to advance the debate.

The choice of measure depends (in part) on the use to which it will be put. Most archaeologists seem to be interested firstly in numbers of actual vessels, and secondly in these numbers expressed as percentages of the total of an assemblage. They may be adjusted by certain factors for special purposes, e.g. by the values of different types, by the volume they would occupy in a kiln, or by their estimated life-span (to try to get back to a 'life-assemblage' from an observed 'death-assemblage'). I suggest that, except in unusual situations, absolute numbers are of little use, and the main interest should be in a breakdown of an assemblage by fabric, form or possibly some other factor, with the aim of comparing it with other assemblages. The assemblage is seen as a sample from a population, from which it has been derived *via* a series of transforms (Schiffer 1972). We are really interested in the parameters (e.g. percentages) of the population, which we have to estimate from our sample. The task of estimating the size of the original has recently been approached by Miller (1986); his methods should be seen as a first attempt at a difficult problem.

5. What makes a good measure?

There are (so far) four theoretical criteria which should underly the choice of a measure, as well as practical ones such as time, cost, expertise, etc. They are:

(a) *bias.* – An estimate based on a sample should approach the true parameter in the population as the size of the sample approaches that of the population. If we wish to estimate percentages based on numbers of vessels, then measure based on weight or sherd count will (almost) always be biased in this sense. Also, any measure would be biased if we are interested in 'life-assemblages' (relative proportions of different types in use at a certain time), and have no information about the relative life-spans of different types.

(b) *invariance.* – It could be argued that we could live with bias, provided it was the same in different assemblages. For example, if a measure always exaggerates the proportion of a certain type (perhaps because it is heavy), we can still compare proportions in different assemblages, and say, for example, that there is *relatively* more of it in one assemblage than another. A measure with this property, that its bias is not affected by a transform, is here called *invariant under that transform*. A good example is weight –if a particular type forms a certain proportion of a population, we would expect it to form the same proportion (within limits due to sampling) of a sample, regardless of the intervening processes of breakage and retrieval.

(c) *variability.* – Other things being equal, a measure which gives a estimate with a higher precision (lower standard deviation) is preferable to one with a lower precision.

(d) *interpretability.* – It would be useful to be able to put confidence limits on an estimate of a proportion, or to assess whether the value of a proportion observed in two assemblages differs significantly from one to another. Some measures may lend themselves to this approach more readily than others.

6. Assessment

The four measures set out here were examined in the light of the first two criteria in Orton (1975). All measures are biased relative to a life-assemblage; relative to a death-assemblage, the work showed that (approximately, and using the present terminology)

(a) sherd count was biased, but invariant under changes in retrieval percentage,

(b) weight was biased, but invariant under changes in breakage rate and retrieval percentage,

(c) evreps was biased, and not invariant,

(d) eves was unbiased and invariant.

These results are illustrated in Fig. 1. Although the figures are hypothetical, they show how the measures could be expected to behave in simple cases. It is especially interesting that, when evreps are used, identical assemblages yield *different* estimates of the coarse:fine

ratio, if either the breakage or retrieval rate differed. This measure can create spurious differences and mask real ones.

Further work, involving computer simulation, looked also at the third criterion (Orton 1982). Evreps often scored well here, but not well enough to overcome its drawbacks of bias and lack of invariance. In the present state of knowledge, I recommend the use of measures (b) weight and (d) eves together, so that comparisons between assemblages can be made.

7. Spin-offs

Useful information can be obtained, however, if combinations of measures are available. For example, we can calculate

(a) 'brokenness' = sherds/eve (called 'b')
(b) 'completeness' = eves/evreps (called 'c') (Orton 1986).
(a) *brokenness* depends on both the type of pottery and its context, as it measures the extent to which the pottery has been broken. Assemblages must be compared type-by-type, not overall, since differences in pottery types can mask differences in degree of breakage, and vice versa. This statistic can be used to show how 'secondary' an assemblage is, since it starts at a value of 1 (whole vessels) and decreases each time breakage occurs.
(b) *completeness* depends only on the context, and should be the same for all types in an assemblage. It can tell us something about the nature of the context.

These statistics were used to help interpret the layers of fill of a section of the Devils Ditch at Boxgrove near Chichester (Bedwin and Orton 1984). Here a difference between the lower fills of the ditch (b = 100 sh/eve; c = 10%) and the upper fills (b = 200 sh/eve; c = 5%) suggested a recutting of the ditch.

8. The future

We need to look critically at the assumptions behind this work, particularly into what we mean by a 'sample'. We also need to be able to link this theory to recent statistical advances in the analysis of compositional data (Aitchison 1986), so that these advances can be fully exploited.

Bibliography

Aitchison, J. 1986: *The statistical analysis of compositional data.* London.

Bedwin, O. and Orton, C. R. 1984: The excavation of the eastern terminal of the Devil's Ditch (Chichester Dykes), Boxgrove, West Sussex, 1982. *Sussex Archaeol Collect* 122 63–74 and fiche.

Carver, M. O. H. 1985: Theory and practice in Urban Pottery Seriation. *Journal Archaeol. Science* 12, 353–366.

Dawson, G. J. 1979: Excavations at Guy's Hospital 1967. *Res Vol Surrey Archaeol Soc* 7, 27–65.

Egloff, B. J. 1973: A method for counting ceramic rim sherds. *Amer Antiquity* 38(3), 351–3.

Glover, I. C. 1972: Excavation in Timor. PhD thesis, Australian National University.

Hinton, D. A. 1977: 'Rudely made earthen vessels' of the twelfth to fifteenth centuries AD. In *Ceramics and Early Commerce* (ed. D. P. S. Peacock). London: Academic Press, 221–38.

Hodder, I. 1974: Some marketing models for Romano-British coarse pottery. *Britannia* 5, 340–59.

Hulthen, B. 1974: On choice of element for determination of quantity of pottery. *Norweg Archaeol Rev* 7(1), 1–5.

Miller, G. 1986: Of fish and sherds: a model for estimating vessel populations from minimal vessel counts. *Hist Archaeol* 20(2), 59–85.

Millett, M. 1980a: How much pottery? In *Pottery and the Archaeologist* (ed. M. Millett), Institute of Archaeology Occasional Paper no. 4, 77–80.

Millett, M. 1980b: An approach to the functional interpretation of pottery. In *Pottery and the Archaeoloqist* (ed. M. Millett), Institute of Archaeology Occasional Paper no. 4, 35–48.

Moorhouse, S. 1986: Non-Dating Uses of Medieval Pottery, *Medieval Ceramics* 10, 85–124.

Orton, C. R. 1975: Quantitative pottery studies: some progress, problems and prospects. *Sci Archaeol* 16 30–5.

Orton C. R. 1982: Computer simulation experiments to assess the performance of measures of quantity of pottery. *World Archaeol* 14(1) 1–20.

Orton, C. R. 1986: Two useful parameters for pottery research. *Computer Appl. Archaeol.* 114–20.

Redman, C. L. 1979: Description and inference with late medieval pottery from Qsar es-Seghir, Morocco. *Medieval ceramics* 3 63–79.

Schiffer, M. B. 1972: Archaeological context and systemic context. *Amer Antiquity* 37(2) 156–65.

Solheim, W. G. 1960: The use of sherd weights and counts in handling of archaeological data. *Current Anthropology* 1 325–9.