The International Journal of Nautical Archaeology (2003) **32.2**: 246–251
doi:10.1016/j.ijna.2003.04.002

An assessment of quality in underwater archaeological surveys using tape measurements

Peter Holt

6 Honcray, Oreston, Plymouth, Devon, PL9 7RP, UK

The quality of an underwater archaeological survey using 3D trilateration with fibreglass tape measures was established on an underwater test site. A precision of 25 mm was calculated for tape measurements giving a position accuracy of 43 mm. Of the 304 measurements which were made during the tests, 20% were found to be in error.

Key words: Survey; quality; accuracy; trilateration.

Introduction

The aim of this work was to determine the quality of an underwater archaeological survey using three-dimensional trilateration with fibreglass tape measures. To achieve this we need to determine the accuracy, precision and reliability of the position of any point in the survey. The assessment of quality measures the sizes and nature of undetected errors, which exist in computed positions.

Three-dimensional trilateration was chosen for this test as the required quality information is produced as a by-product of processing the measurements. This form of trilateration is often called the 'Direct Survey Method' and was popularized in marine archaeology by Rule (1989). The method used for processing distance measurements is similar to that used by Global Positioning System receivers (UKOOA, 1994) and underwater acoustic positioning systems (Kelland, 1973, 1994).

Terminology

Accuracy is considered to be an overall estimate of the errors present in measurements including systematic errors. Accurate measurements are those which are close to the true value. Where systematic errors have been removed accuracy is the same as precision.

Precision is a term used to describe the quality of a position with respect to random errors. Thus a very precise position fix is one where the random errors are small. As the random errors cannot be determined, precision is usually measured by means of a standard deviation.

Standard Deviation (SD) is a measure of the spread of the random errors in any measurement, so the larger the SD the larger the random errors.

Reliability is used to describe the quality of a position with respect to outliers (mistakes or blunders), so in a highly reliable position fix even quite small outliers will be detected.

Precision of a position is determined by assessing the standard deviations of measurements used to calculate that position and computing their propagation through a best-fit process. The standard best-fit process is called least squares: this gives the most desirable result (the one with the highest precision) and is very simple (Cross, 2002; Uren & Price, 1985; Atkinson *et al.*, 1988; Bannister *et al.*, 1994).

Method

To calculate the quality metrics we needed a sample data set of multiple sets of tape measurements made between a number of fixed and rigid points on a typical site underwater. No corrections were applied to the measurements for temperature, sag or tension so the results would be close to those

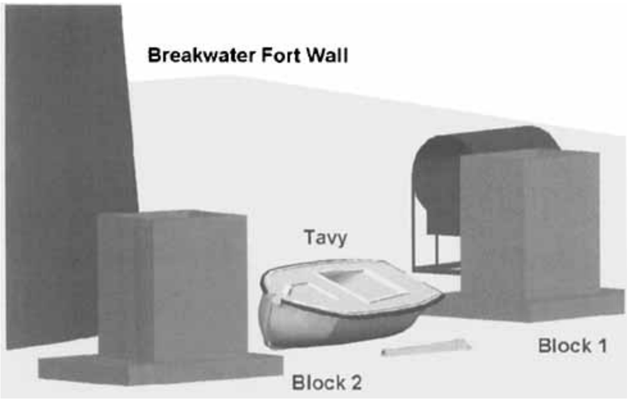

Figure 1. Three-dimensional model of the Breakwater Fort site.

achieved on a typical underwater site. As a suitable site was not available a test site was set up at the base of the Fort behind the Breakwater in Plymouth Sound. This site was used by the Fort Bovisand Underwater Centre as a training ground for commercial divers as it was sheltered, only 10 m deep, had minimal current and had underwater visibility between 2 m and 5 m. The chosen site contained a number of fixed and rigid structures suitable for recording. These included two large concrete blocks, the wall of the Fort itself, and a 7 m long ex-pilot cutter called *Tavy* (Fig. 1). A network of 21 control points was installed on the structures (Fig. 2). The shape was designed to give a large amount of redundancy and minimal sensitivity to depth errors. The control points installed on the structures were 5 mm galvanised coach bolts cemented into pre-drilled holes. The diameter of the control point bolts was accounted for within the processing programme.

Over a period of a year many teams of divers had recorded to 1 mm resolution a pre-defined set of measurements between the control points. The same set of tape measures was used for each exercise. As a check for systematic errors, all tape measures were calibrated against a steel tape measure at 5 m, 10 m and 20 m distances and any tape measures with more than 5 mm in error were not used. To minimise transcription errors, standard recording forms were used and the data was transferred from the form straight into a computer spreadsheet for analysis. Measurements were exported from the spreadsheet directly into the processing program. A set of tape measurements was made on a single baseline on land for comparison. Data from other sites was collected and compared with the results from the test site.

Results

The data set

A total of 32 baselines with distances between 2 m and 13 m were measured more than five times, comprising a total of 178 measurements. Another 85 baselines which had fewer than five measurements each were included, giving 304 measurements in total.

Outliers in repeat measurements

The majority of the baselines showed one or more gross errors even though some lengths were less than 3 m. Fig. 3 shows the residuals for 12 measurements of a 1·7 m long baseline where 11 of the measurements have residuals less than 20 mm but one is in error by nearly 140 mm.

Average standard deviation

With the obvious outliers removed, the mean and standard deviation of the measurements was calculated for each baseline. Together the baselines gave an average standard deviation of 25 mm where the minimum was 8 mm and the maximum was 60 mm. This value for precision of 25 mm was used as the starting point for subsequent analysis.

Adjustment

The measurements were adjusted in one large least-squares network adjustment using the Site Surveyor computer program. The resulting computed positions of the points are the best estimate based on all of the measurements, giving the most likely positions. The program also computes the residual for each measurement, the difference between the measured value and the value computed from the positions of the points.

Fig. 4 shows the residuals plotted in size order with a vertical bar showing the approximate point of separation between valid measurements to the left and outliers to the right. Many large outliers existed in the data set, but there was no clear distinction between a valid measurement and a small outlier. Two methods were used to identify outliers, the first was manually to reject any measurements over a given limit. A rejection limit of three standard deviations (99·7%) was used with our previously computed precision of 25 mm, so any outlier larger than 75 mm was rejected. The second method was to use an automatic rejection process based on the Delft method recommended

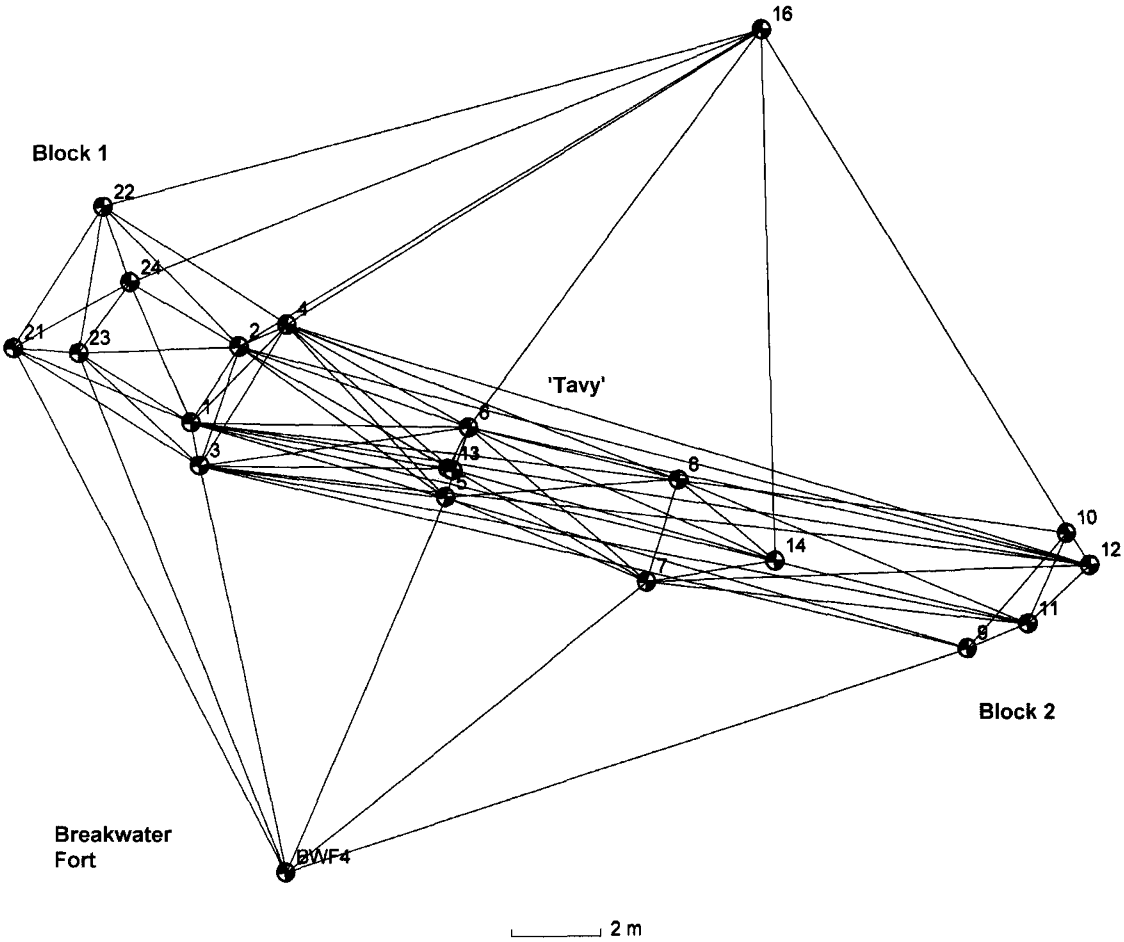

Figure 2. Measurements between the 21 control points.

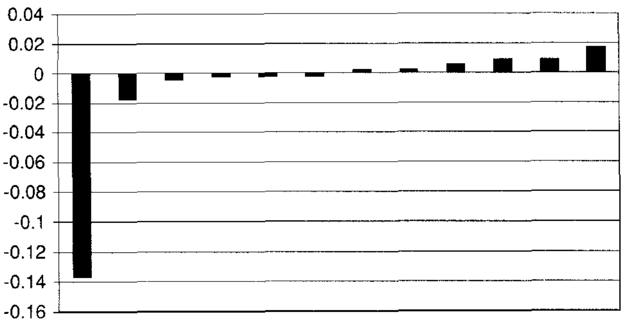

Figure 3. Measurement residuals for baseline 5-6.

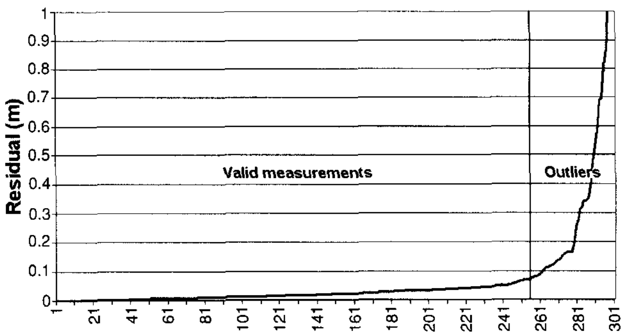

Figure 4. Measurement residuals in size order.

for processing GPS measurements (Table 1). The automatic rejection method is an iterative process that rejects the measurement with the highest w-statistic or normalised residual after adjustment. The process stops when all remaining w-statistic values lie below 2·576 or 99%. The normalised residual is obtained by dividing a residual by its standard deviation so both the manual and automatic methods are driven by our estimate of precision.

Outliers

Of the 304 observations, 168 (55%) were smaller than the defined measurement standard deviation

Table 1. *Post adjustment results*

Rejection	None	3 S.D.	Automatic
RMS of the residuals	143 mm	30 mm	27 mm
Measurements	304	304	304
Measurements used	304	247	240
Measurements rejected	0	57	64
Measurements rejected	0%	18.8%	21.0%

and 57 were larger than 3 times the standard deviation. The Root-Mean-Square (RMS) of residuals gives an idea of how well all the measurements fit together. In the case where no outliers were rejected the RMS value was 143 mm showing that outliers existed in the data set. In the cases where outliers were rejected the RMS of residuals becomes close to the expected value of 25 mm, our nominated precision for the measurements. It was expected that longer measurements would be more likely to have larger residuals. However the data shows that this is not the case. Interestingly, the set of residuals shows no correlation between measurement length and residual (Fig. 5).

Three types of dive teams were used to collect the measurements: commercial diver trainees; marine archaeology students (NAS Part II level); and divers experienced in underwater recording. On average, the trainees and students made 30% mistakes while the experienced divers made only 8%.

Typical point accuracy

The accuracy of the position of a point is expressed as error ellipse based on both the precision of the associated distance measurements and the position of the point within the control network. The ellipse is an approximate graphical representation of the horizontal accuracy in all directions. Error ellipses are commonly shown at 2·447 times their 1 SD values and are then referred to as 95% confidence regions. The sizes of the error ellipses computed by the adjustment program are directly related to the size of the standard deviation defined for the tape measurements. The standard deviation also sets the maximum acceptable residual, so a small standard deviation gives a small size of error ellipse but requires better quality measurements to achieve it. A simulated network of four fixed control points was set up with a test point in the centre, one on a baseline between two control points and a third outside the control network (Fig. 6). The accuracy of the position of a set of points was then calculated based on the given precision of 25 mm (Table 2). The position error ellipses are shown 20 times full scale. This assumes the fixed station positions to be perfectly known: if the co-ordinates are in error then the fix will be thought to be of better quality than it actually is.

Other sites

Measurements from a number of other similar underwater sites were processed in order to obtain an estimate of the distance measurement precision and the percentage of outliers in the data set (Table 3). In each case the automatic rejection tool in Site Surveyor was used. The RMS residual values of between 10 mm to 25 mm are typical for shallow water sites.

Measurements made on land

A set of 12 measurements made on land by trainees over a 12·3 m baseline in sheltered conditions with the tape unsupported gave a standard deviation of 6 mm with no outliers.

Conclusion

Under the test conditions, a standard deviation of 25 mm is valid for measurements made using fibreglass tape measures over distances up to 20 m underwater. This value can be used as a typical figure for tape measurements under similar conditions. The single data set on land showed a standard deviation of 6 mm over a comparable length. The cause of the difference between land and underwater measurements is most likely to be a combination of the effects of any water current on the tape and inability to maintain the correct tension on the tape when underwater. Repeating the tests using a steel-cored tape rather than fibreglass would most probably show an increase in precision as the stretch is less so a higher tension can be used.

The number of outliers in the test data set was approximately 20%, considerably larger than is usually assumed for survey work underwater and larger than achieved at the other sites. The percentage of mistakes made by the different types of divers is a likely cause of the discrepancy. The data in this set was collected by a mixed set of experienced and trainee divers whereas the data for the other sites was collected by experienced site

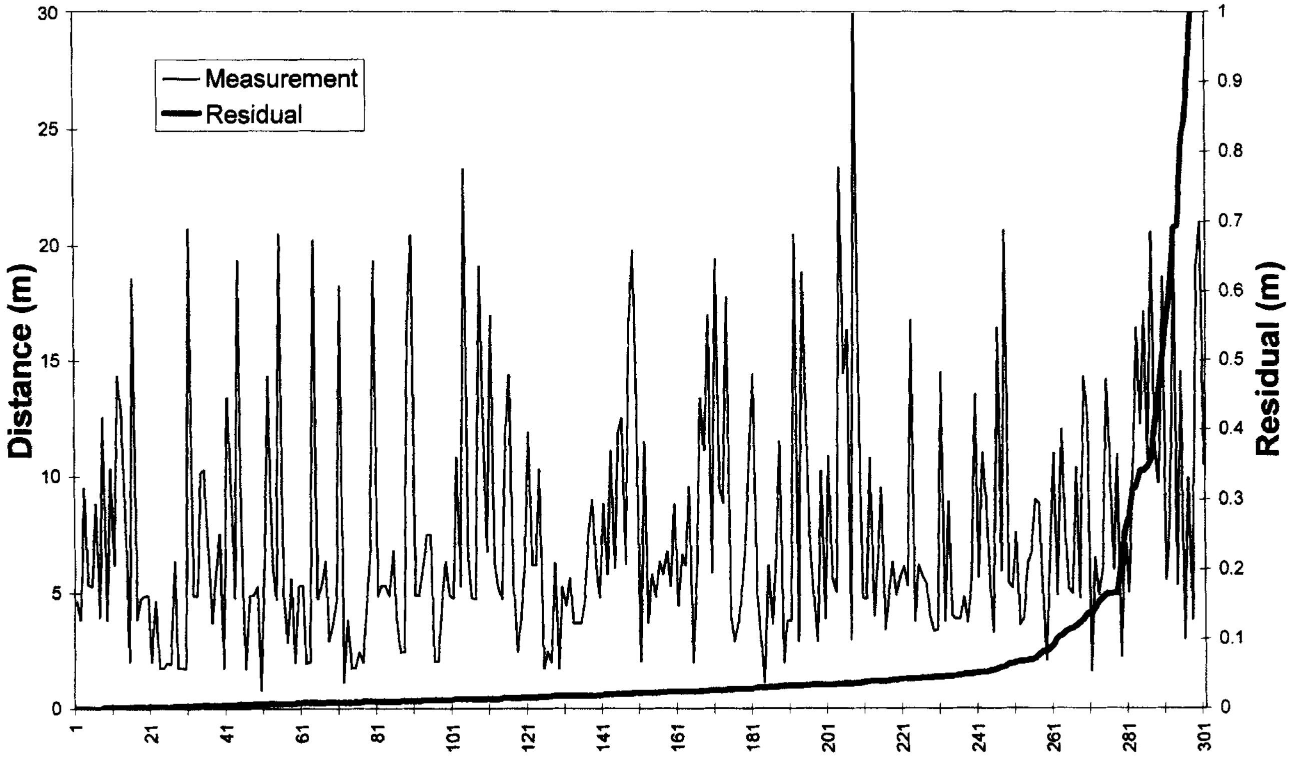

Figure 5. Measurements and residuals in residual size order.

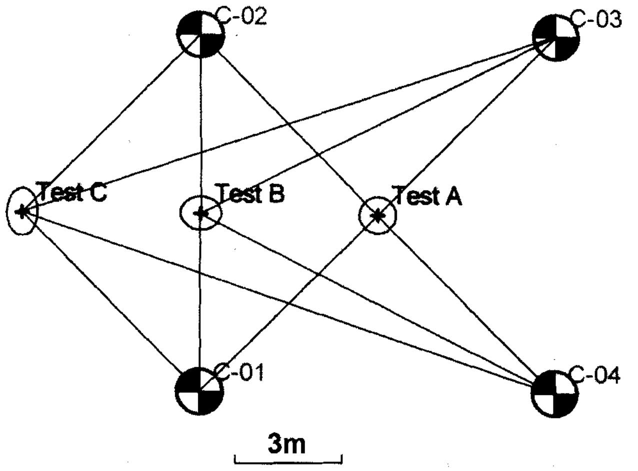

Figure 6. Position precision test.

recorders. Further research is required to produce valid error rates as there is not enough data in this set to form any definite conclusions.

There were outliers even in short baselines where the tape was supported along its whole length. It is unlikely that stretching in the tape caused these outliers, the more likely cause was mis-reading of the tape measure or transcription errors when transferring measurements to the recording forms or from the forms to the computer.

Table 2. *Point accuracy results*

	Semi-Major (95%)	Semi-Minor (95%)	Note
Test A	43 mm	43 mm	Centre of the network
Test B	48 mm	39 mm	Between two control points
Test C	56 mm	37 mm	Outside the control points

Table 3. *Residuals and outliers from other sites*

	Control points	RMS Residuals	Outliers
Hazardous	12	8 mm	8 of 52 15%
Boyne	21	15 mm	8 of 101 8%
Resurgam	12	17 mm	5 of 45 11%
Coronation	12	8 mm	0 of 39 0%
Colossus	18	23 mm	10 of 125 8%
Alum Bay	14	10 mm	9 of 77 12%

There was no obvious correlation between size of outlier and measurement length, so large outliers are as likely to appear in short measurements as long ones. This result was unexpected and hints that a significant proportion of the outliers came from mis-reading or transcription errors. This is important, as it means that such problems can be minimized with diver training and by reducing the number of transcriptions. Where a diver is in voice communication to the surface, fewer mistakes will probably be made if the surface team rather than the diver record the measurements. The best method appears to be to process measurements on the computer as they are made, allowing the immediate identification of outliers, as this minimises the amount of work to be re-done.

The high percentage of outliers emphasises the need for good training, for the collection of redundant measurements, and for appropriate survey data processing techniques to identify and eliminate the outliers. The tests were done using 3D trilateration measurements; however, it is safe to assume that the same numbers of outliers will occur when positioning using offsets and ties as the same measurement procedures are used.

The post-computed position confidence regions for points inside the control network can be approximated to circles 40 mm in radius. This means a typical point positioned using 3D trilateration with tape measures will be accurate to ±40 mm at 95% confidence. This figure can be used as a baseline standard for comparison with other methods of positioning underwater such as acoustic or optical systems under similar conditions.

These tests were done under almost ideal conditions for UK waters and the precision achieved is likely to be the highest achievable using divers of mixed ability. Further research is needed to be able better to quantify the effect of training and environment on measurement accuracy and error rates.

Acknowledgements

I am grateful to Paul Dart for getting the measurements made, POP Fort Bovisand and the Queen's Harbour master, Plymouth, for allowing us to sink the *Tavy* along with Nigel Boston for donating the *Tavy* and the use of his ship to put it on site. My thanks go to Garry Momber at the Hants and Wight Trust for Maritime Archaeology, Anthony Firth at Wessex Archaeology and Kevin Camidge for survey data from other sites. Thanks also go to Jeremy Green at the Western Australian Maritime Museum for his comments on this paper.

References

Atkinson, K., Duncan, A. & Green, J., 1988, The application of a least squares adjustment program to underwater survey, *IJNA,* **17**.2: 119–131.

Bannister, A., Raymond, S. & Baker, R., 1994, *Surveying.* London.

Cross, P. A., 2002, *Advanced Least Squares Applied to Position Fixing.* University of East London Working Paper 6, London.

Kelland, N., 1973, Assessment Trials of Underwater Acoustic Triangulation Equipment, *IJNA,* **2**.1: 163–176.

Kelland, N., 1994, Developments in Integrated Underwater Acoustic Positioning, *Hydrographic Journal,* **71**: 19–27.

Rule, N., 1989, The Direct Survey Method (DSM) of underwater survey, and its application underwater, *IJNA,* **18**.2: 157–162.

UKOOA, 1994, *The Use of Differential GPS in Offshore Surveying.* London.

Uren, J. & Price, W., 1985, *Surveying for Engineers.* London.