Quantogram Analysis

Wilfrid Kendall

10-18 January 2017

> The early eleventh-century tract 'Episcopus' enjoins "that no measure-rod (*metegyrd*) be longer than another, but all are to be regulated by the confessor's [i.e. parish priest's] measure (*gemete*); and every measure in his parish (*scriftscire*) and every weight is to be regulated by his direction very correctly; and if there is anything in dispute, the bishop is to arbitrate".
>
> (translated by Blair, 2013)

```r
library(ggplot2) # advanced graphics
library(cowplot) # needed for multiple plots
library(ggmap)   # online maps
library(Hmisc)   # needed for first.word
```

1 Introduction.

Data were provided by Prof. John Blair of Queen's College Oxford, arising as measurements taken from archaeological plans of Anglo-Saxon sites of major buildings. Here we report on an analysis which evaluates the extent to which these data represent evidence for a single unit of measurement underlying dimensions of building plans from Anglo-Saxon England. Measurements were taken in metres to nearest 5 centimetres, and were based on fixed baselines (loosely corresponding to N/S and E/W axes) determined for each building, and grouped in lines of cumulative measurements. The research hypothesis is that the dimensions so measured have a tendency to be multiples of some fixed *quantum* (here measured in metres). I use the term *quantum* in place of "module" both out of filial piety (D.G. Kendall, 1974) and also to emphasize that the results below are purely statistical, and need to be set in a clear scholarly context if they are to have any historical value.

It is useful to bear in mind the empirical but severely non-statistical work of Huggins [1991], who argues for the use of a "short rod" (as modulus) of 4.65 metres. He bases this on visual examination of ground plans of buildings of the period. Blair [2013]} arrives independently at a modulus of 4.57 metres, but considers the difference to be of no great significance.

This paper follows up on preliminary work (Kendall, 2013, an appendix to Blair, 2013). In contrast to Kendall [2013], the following analysis is based on an observation of Huggins [1991] (namely that one should make distance comparisons based on the central lines of walls). This allows us to produce a refined estimate of the underlying quantum or modulus.

2 Concerning the data set

2.1 Data format

Data were presented in `DOCX` format as files `Measurements.Key.docx`, `Measurements.Data.docx`. The first of these, `Measurements.Key.docx`, describes the data as given in Appendix A. Appendix B provides a full listing of the data-set.

2.2 Data acquisition

The contents of `Measurements.Data.docx` were copied into a text file `Measurement.csv`. The resulting data were double-spaced, and contained: identifiers of buildings (lines beginning with alphabetic characters), and comma-separated measurements as decimals (lines beginning with a digit or square brackets). The interpretations of the comma-separated measurements are summarized in the following table.

Identifier	Interpretation
y	distance in metres
s	site number
n	line number
o	orientation (W-E or N-S)
a	aspect (facing baseline or not)
m	order in measurement line

Decimal numbers enclosed in square brackets were actually produced by estimation, but advice was to treat these as accurate. Accordingly the square brackets were stripped from these data

```
column.names <- c("y", "s", "n", "o", "a", "m")
raw.data <- read.table("Data/Measurements.txt", header=FALSE, sep=",",
                       col.names=column.names, colClasses="character", fill=TRUE, skip=2)
head(raw.data)
```

```
##           y                  s  n  o  a  m
## 1 Canterbury  SS Peter and Paul
## 2       0.65                 1  1  1  1  1
## 3       9.15                 1  1  1  0  2
## 4       9.50                 1  1  1  1  3
## 5       0.65                 1  2  1  1  1
## 6       4.10                 1  2  1  0  2
```

Lines naming the five different sites were selected out.

```
is.site.name <- is.na(as.numeric(raw.data$m))
site.names   <- raw.data[is.site.name,]
site.names   <- trimws(paste(site.names[,1], site.names[,2]))
cat(t(site.names), sep=', ')
```

```
## Canterbury  SS Peter and Paul, Canterbury  St Pancras, Hexham, Escomb, Brixworth
```

If we remove these lines then the result can be converted to a data frame of measurements, each row named using the integer values of the corresponding non-measurement variables `n`, `m`, `a`, `o`, `s`. We adjust the values of the `o` field to be `1` and `2` rather than `0` and `1`.

```
strip.brackets <- function(x){as.numeric(sub("\\]", "", sub("\\[", "", x)))}
meas  <- raw.data[!is.site.name,]
meas  <- apply(meas, c(1,2), strip.brackets)
select <- c(3,6,5,4,2)
row.summaries <- apply(meas[,select], 1, paste, collapse="-")
meas  <- as.data.frame(meas,row.names=row.summaries)
meas$o <- meas$o+1
head(meas)
```

```
##              y s n o a m
## 1-1-1-1-1 0.65 1 1 2 1 1
## 1-2-0-1-1 9.15 1 1 2 0 2
## 1-3-1-1-1 9.50 1 1 2 1 3
## 2-1-1-1-1 0.65 1 2 2 1 1
## 2-2-0-1-1 4.10 1 2 2 0 2
## 2-3-1-1-1 4.70 1 2 2 1 3
```

The geocode function from ggmap was used to help generate some geographical metadata, in the form of latitude and longitude for each location. The resulting locations are plotted in Figure 1. It is evident that the sites group as follows:

(a) the southern group (sites 1 and 2) of two separate sites at Canterbury;

(b) the northern group (sides 3 and 4) of two sites in Northumberland;

(c) a single site in the Midlands (site 5) at Brixworth.

```
make.metadata <- function(x) unlist(geocode(first.word(as.character(x))))
metadata <- data.frame(t(mapply(make.metadata,  as.vector(site.names))))
metadata$num <- as.character(1:nrow(metadata))
metadata$num[[2]] <- paste(metadata$num[[1]], ", ", metadata$num[[2]], sep="")
metadata$lat2 = metadata$lat + 0.2
```

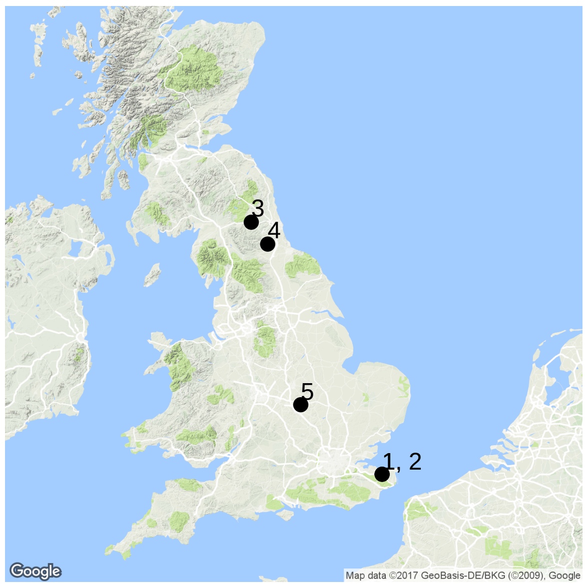

Figure 1: Locations of sites: Canterbury (1,2), Brixworth (5), Escomb (4), Hexham (3).

2.3 Data summary

The numbers of measurement lines per site and measurements per measurement line are given as follows,

	1	2	3	4	5
Number of measurement lines per site:	3	2	3	2	3

	1	2	3	4	5	6	7	8	9	10	11	12	13
Number of measurements per line:	3	3	7	3	3	7	9	4	5	7	10	13	5

The numbers of measurements per site are as follows.

	1	2	3	4	5
Number of measurements per site:	13	6	20	12	28

	1	2	3	4	5

Thus there are 13 lines of measurement and 79 measurements in all.

Inspection shows that the measurement aspects (`a`) alternate along the lines of measurement.

```
head(split(meas$a, meas$n, drop=TRUE))
```

```
## $`1`
## [1] 1 0 1
##
## $`2`
## [1] 1 0 1
##
## $`3`
## [1] 1 0 1 0 1 0 1
##
## $`4`
## [1] 1 0 1
##
## $`5`
## [1] 1 0 1
##
## $`6`
## [1] 1 0 1 0 1 0 1
```

2.4 Data manipulation

We divide the measurements into different lines (values of `n`), successive pairs have to be averaged to locate the centre of each wall, and differences in length must be computed per line. So we must implement the following:

1. Prepend `0.0` to each row (the baseline of the corresponding set of measurements);
2. Split into lists of measurements for each line;
3. Average successive pairs;
4. Difference all entries from the first entry in each row.

There is something strange about Line 10, which commences with a pair of entries of 0.70**.** However I will leave this aside for now.

Lines 8 and 11 have to be treated differently from the others: in these cases the initial aspect variable indicates that measurements are taken from the first measurement rather than the notional baseline at `0.0`. So step 1 has to be omitted for these lines.

In addition lines 6 and 7 contain measurements relating to crypt-like objects. In order for these to be treated properly when computing centre-lines ("averaging successive pairs"), it is necessary to duplicate the relevant observations.

We duplicate second and third measurements in line 6, and fourth and fifth measurements in line 7. (Very little difference arises from deleting these measurements instead, or indeed from leaving them unadjusted.)

```
cat("Duplicate lines ")
```

```
## Duplicate lines
```

```r
meas[31,]
```

```
##              y s n o a m
## 7-5-1-0-3 9.35 3 7 1 1 5
```

```r
meas[30,]
```

```
##           y s n o a m
## 7-4-0-0-3 7 3 7 1 0 4
```

```r
meas[22,]
```

```
##              y s n o a m
## 6-3-1-1-3 5.95 3 6 2 1 3
```

```r
meas[21,]
```

```
##           y s n o a m
## 6-2-0-1-3 2 3 6 2 0 2
```

```r
meas <- rbind(meas[1:21,], meas[21:22,], meas[22:30,], meas[30:31,], meas[31:dim(meas)[1],])
```

We prepend labelled zero measurements, excepting rows 8 and 11.

```r
rmax <- max(meas$n)
cmax <- length(names(meas))
insertion <- array(rep(0, rmax * cmax), dim=c(rmax, cmax))
insertion[,3] <- 1:rmax
insertion <- insertion[-c(8,11), ] # omitting lines 8 and 11
mmeas <- as.data.frame(rbind(as.matrix(insertion), as.matrix(meas)),
                       row.names=names(meas))
```

Now we split the measurements into a ragged array of measurements for each line, making a similar split for site number (`s`), line number (`n`) and orientation (`o`).

```r
line.meas <- split(mmeas$y, mmeas$n)
line.s <- split(mmeas$s, mmeas$n)
line.n <- split(mmeas$n, mmeas$n)
line.o <- split(mmeas$o, mmeas$n)
```

3 Analysis based on centre-lines of walls

We choose to follow Huggins [1991] in assessing measurements in terms of averaged midpoints for each wall (the "centre-line" of the wall). (This is contrary to the preference suggested in discussions with Blair, namely the measurement of distances *between* walls, which was adopted in Kendall, 2013.) Accordingly, we average consecutive pairs for each line, and finally subtract off the initial average. This is used to construct the data frame `data`, composed of differences of wall centre-lines from a baseline.

```r
average.pairs <- function(z) as.vector(matrix(data=c(1,1)/2, ncol=2)
                                        %*% matrix(data=unlist(z), nrow=2))
averages <- lapply(line.meas, average.pairs)
differences <- lapply(averages, function(z) tail(z-z[1], -1))
extract.function1 <- function (z) as.vector(z)[2:((length(z)/2))]
data <- data.frame(x=c(differences, recursive=TRUE),
                   s=c(lapply(line.s, extract.function1), recursive=TRUE),
                   n=c(lapply(line.n, extract.function1), recursive=TRUE),
```

```
                   o=c(lapply(line.o, extract.function1), recursive=TRUE))
head(data)
```

```
##           x s n o
## 1   9.000 1 1 2
## 2   4.075 1 2 2
## 31  4.125 1 3 1
## 32 13.500 1 3 1
## 33 17.600 1 3 1
## 4  13.650 2 4 2
```

3.1 Preliminary inspection

Figure 2(a) displays the differences of centre-lines from a baseline in an index plot, and Figure 2(a) plots them in a histogram. There is visual evidence in the histogram for a regularity of data-spacing such as might have arisen from a quantum of measurement of order 5m. However it is unclear from this preliminary graphical analysis whether the grouping is significant, and (even if this is decided positively) it is not evident how one might determine accuracy of any estimate one might make of the putative quantum from a purely visual assessment.

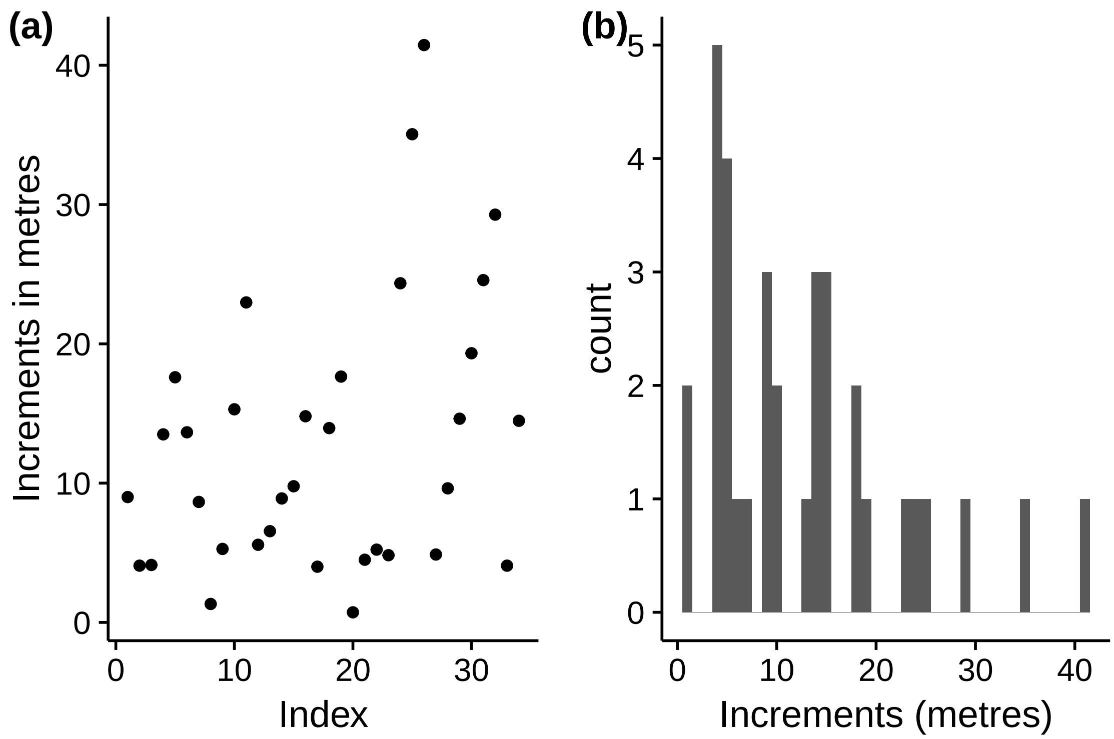

Figure 2: Increments: (a) index plot and (b) histogram (based on consecutive pairs of centre-lines).

Plotting distances of wall faces along the measurement lines against separate measurement lines produces regularities which seem to support the hypothesis of an underlying quantum: see Figure 3.

An even more persuasive picture is obtained if we substitute wall centre-lines for wall faces: see Figure 4. This strongly suggests that our analyses should be based on centre-lines. However we are still left with the questions of

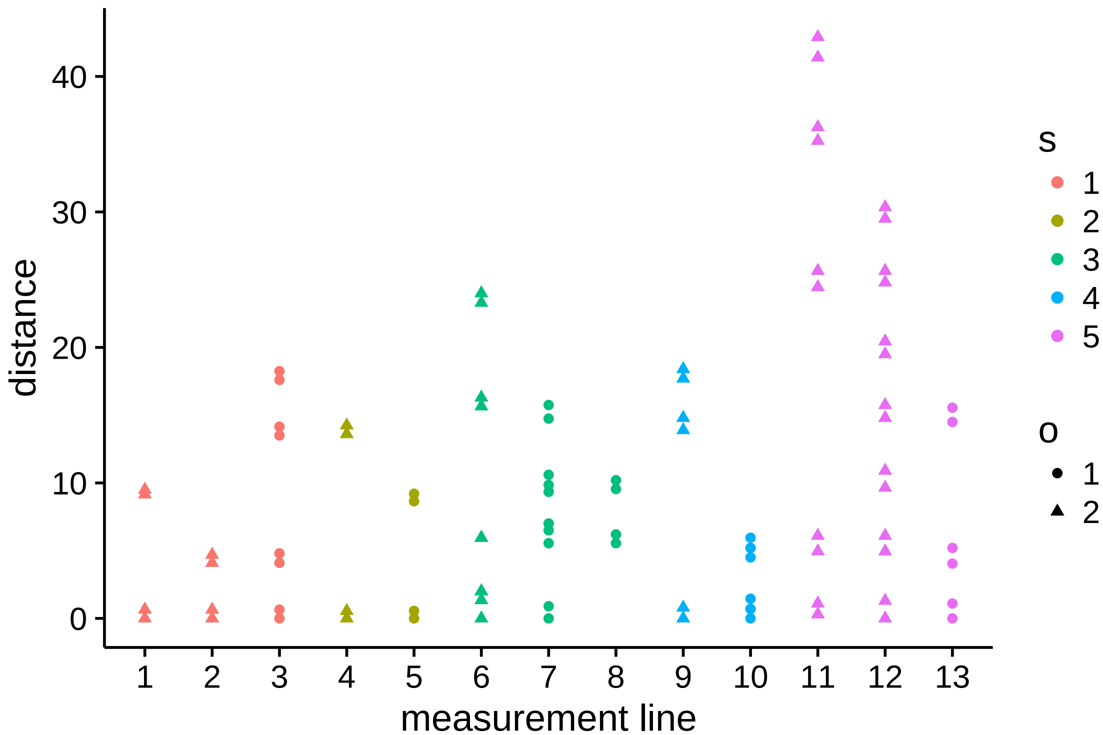

Figure 3: Distances in metres based on measurements of wall faces. Shapes indicate wall orientation, colours indicate sites.

1. confirming statistically that this impression corresponds to evidence for an actual quantum being employed;
2. obtaining a quantitative estimate of the quantum;
3. obtaining a sense of how accurately the quantum is being estimated.

We shall see that we can obtain answers to all these questions.

3.2 Quantogram display

We employ a variation on the (cosine-)quantogram technique of D.G. Kendall [1974]. This technique considers the following statistic (justified by considerations taken from the statistical theory of directional data):

$$\phi(\tau) \quad = \quad \sqrt{\frac{2}{N}} \sum_{j=1}^{N} \cos(2\pi X_j \tau)\,, \tag{1}$$

where $N = 34$ is the number of observations; $X_1, \ldots, X_N$ are the observations themselves (here, the differences from baselines); $\tau = 1/q$ is a "frequency", being the reciprocal of q the proposed quantum length (hence measured in units of metre^{-1}). Thus if all X_j were exactly equal to some multiple of q then all the cosines $\cos(2\pi X_j \tau)$ would be equal to 1, hence ϕ would take on the high value of $\sqrt{2N}$ (in this case $\sqrt{2N} = 8.246$). On the other hand, if q was essentially unrelated to the X_j then an argument based on the central limit theorem leads us to expect ϕ to be essentially normally distributed, of zero mean and unit standard deviation. Figure 5 plots $\phi(\tau)$ as τ varies over reasonable values of τ; high peaks signal the possibility of a quantum. Here the lower frequency bound of $0.01m^{-1}$ corresponds to the absurdly large quantum of $100m$ (the data *do* appear to follow such a quantum, but this corresponds to the uninteresting fact that at this scale all

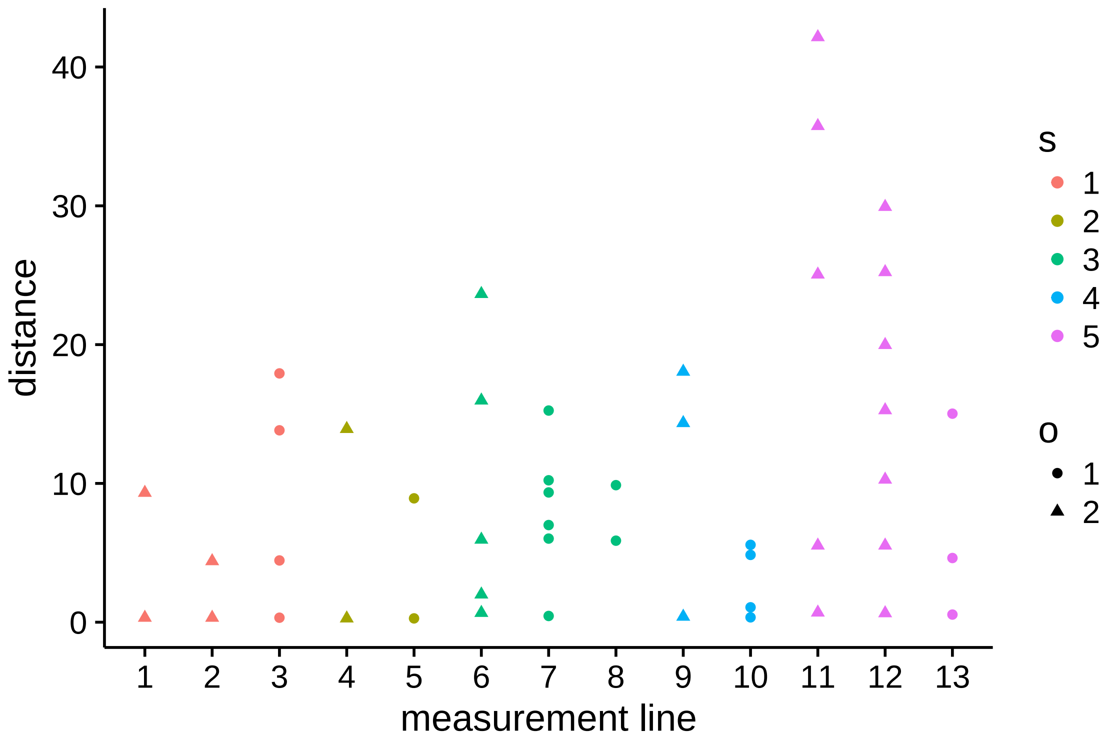

Figure 4: Distances in metres based on measurements of wall centre-lines. Shapes indicate wall orientation, colours indicate sites.

increments are relatively nearly zero! We refer to this peak as the "trivial" peak), while the high frequency bound of $3m^{-1}$ corresponds to a lower bound for the quantum of 0.333 metres. Attention is drawn to the highest non-trivial peak, located towards the left of the figure, at frequency of approximately $0.2m^{-1}$ corresponding to a quantum of around 5 metres.

```
freq <- seq(from=0.01, to=3, length=1000)
cluster <- function(a, b) cos(2 * pi * a / b) # Used only in quantocalc below!
quantocalc <- function (f, x) apply(outer(x, 1/f, cluster), 2, sum)*sqrt(2/length(x))
quantos <- quantocalc(freq, data$x)
```

3.3 Quantogram significance

Assessment of the highest non-trivial peak requires some care. Theory identifies the asymptotical distribution of such quantograms as that of a particular stationary Gaussian process [Kent, 1975, Csörgő, 1980] (using the limit theory of Cramér and Leadbetter, 1967). However *a priori* it is not clear whether the limit theory applies, both because the data do contain some dependencies (measurements sharing the measurement line, possible variation arising from the different geographical locations of the various sites) and also because the peak lies quite near to the extreme left end of the quantogram, which is where one might expect the stationary limiting approximation to break down.

We therefore follow D.G. Kendall [1974], by comparing this quantogram with `simulated.replicates` = 499 quantograms of simulated data, using a smoothing parameter. The seed of the random number generator is set by `set.seed` = 5. We stipulate the "smoothing parameter" to be given by `smooth.sd` = 2.5. The simulated data are generated by randomly perturbing the differences of centre-lines from baselines, using normally distributed perturbations, of zero mean and standard deviation (sd) `smooth.sd`, then taking absolute values of the perturbed data (to avoid encountering negative deviations from the baseline). This produces a dataset

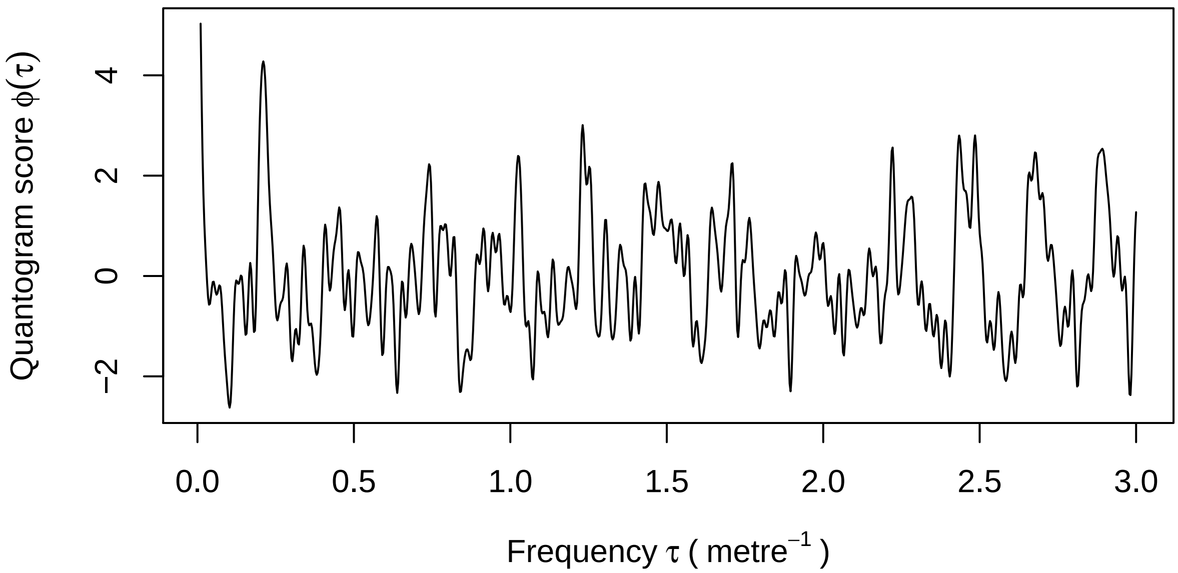

Figure 5: Quantogram of increments (based on consecutive pairs of centre-lines).

whose distribution is similar to the original dataset, but with possible quantum effects smoothed away: see Figure 6. Density plots of original and perturbed data are compared in Figure 7, using the normal probability density of standard deviation `smooth.sd` to convert the data point set into a probability density in each case. The choice of standard deviation is based on resolving the conflict between (a) producing a perturbed dataset which is substantially different in terms of quanta, while (b) requiring the perturbed dataset to be similar in overall terms. Figures 6 and 7 indicate that this has been achieved with our choice above. To use a term introduced by D.G. Kendall, we are employing a "lateral perturbation" of the dataset. Our choice of the smoothing parameter is somewhat *ad hoc*, based on the informal criterion that the upper envelope of the simulated quantograms should not track the actual quantogram too closely. Reassuringly, conclusions appear to be robust under variation of the smoothing parameter. Because of the averaging, the imputed smoothing standard deviation based on wall boundary measurements (rather than centre-line measurements) is $\sqrt{2} * \texttt{smooth.sd} = 3.5355339$.

Note that it could be argued that we should perturb the original dataset of cumulative measurements, since this will give rise to correlated changes in the increments. At present we take the simple approach, especially since the intention is only to construct a comparison dataset, and since there is only a limited amount of correlation resulting from differencing with the zero measurement at the baseline. (However we later will adopt the more complicated approach of perturbing the original dataset, when we wish to consider all possible differences of measurements in each measurement line in Section 3.5.)

Finally, we construct $\texttt{simulated.replicates} = 499$ such laterally perturbed datasets, compute the fifth largest value at each frequency point of the 499 resulting quantograms, and compare this with the original quantogram. The perturbations allow for the fact that measurements are differenced with the wall nearest the baseline, which may itself be perturbed, thus supplying a positive correlation. It is evident that the quantogram is reasonably stationary around the estimated quantum, as suggested in the work of D.G. Kendall [1974] and Kent [1975].

```
result <- c()

for (i in seq(simulated.replicates)) {
  baseperturbation <- rnorm(1, 0, smooth.sd)
  ysim <- abs(data$x + rnorm(data$x, 0, smooth.sd) - baseperturbation)
  result <- rbind(result, quantocalc(freq, ysim))
  }
```

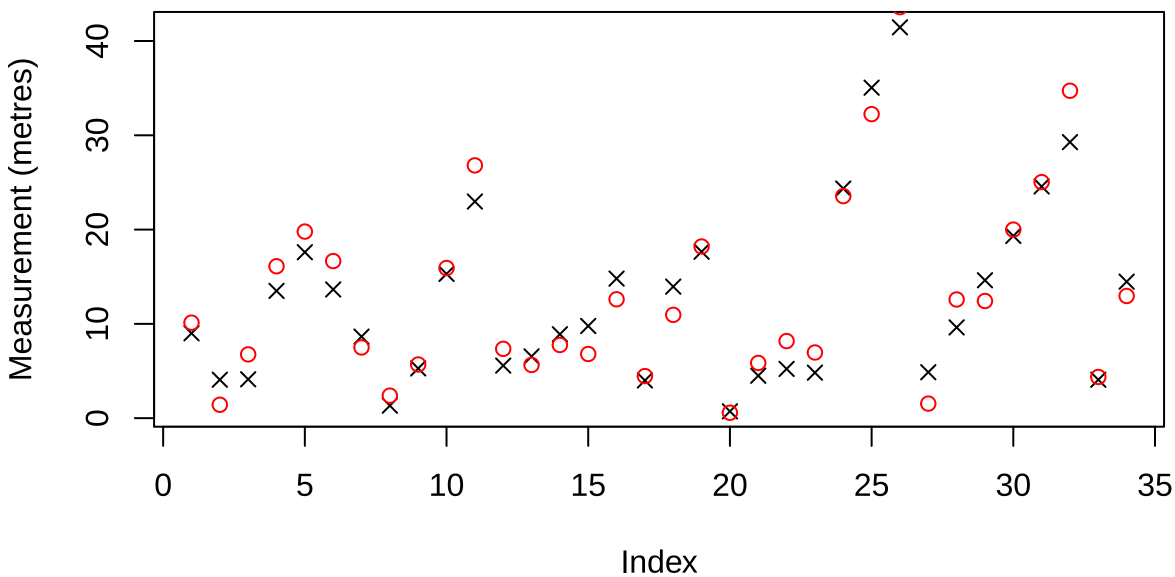

Figure 6: Original dataset (crosses) and laterally perturbed dataset (circles) based on differences of centre-lines from baselines.

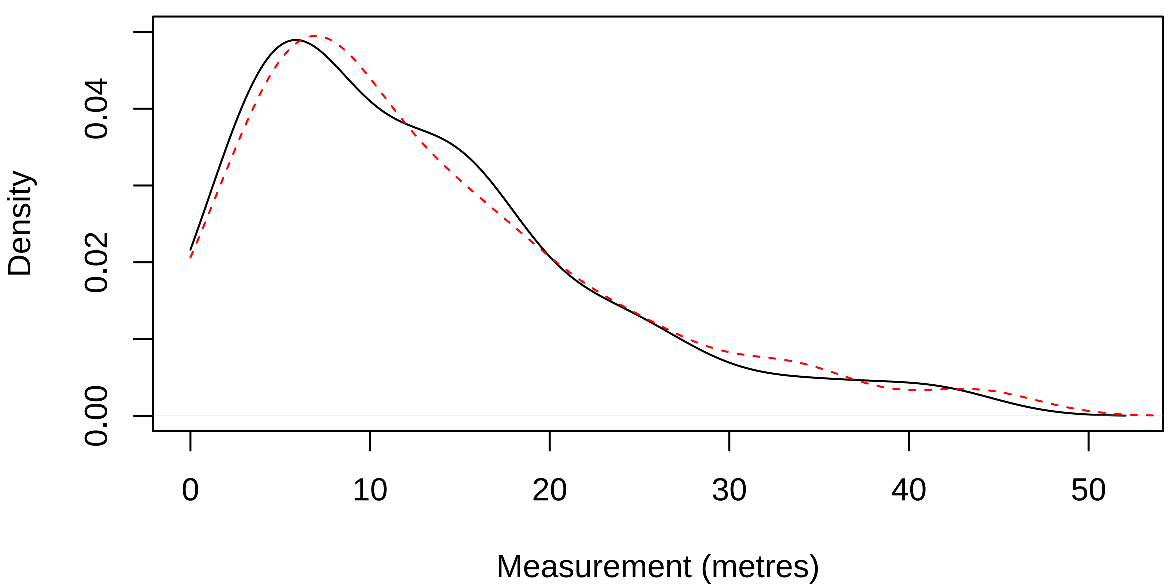

Figure 7: Smoothed density plots for laterally perturbed dataset (red dashed curve) and original (black continuous curve) (based on differences of centre-lines from baselines).

```r
upper <- apply(result, 2, sort)[dim(result)[1]-4,]
```

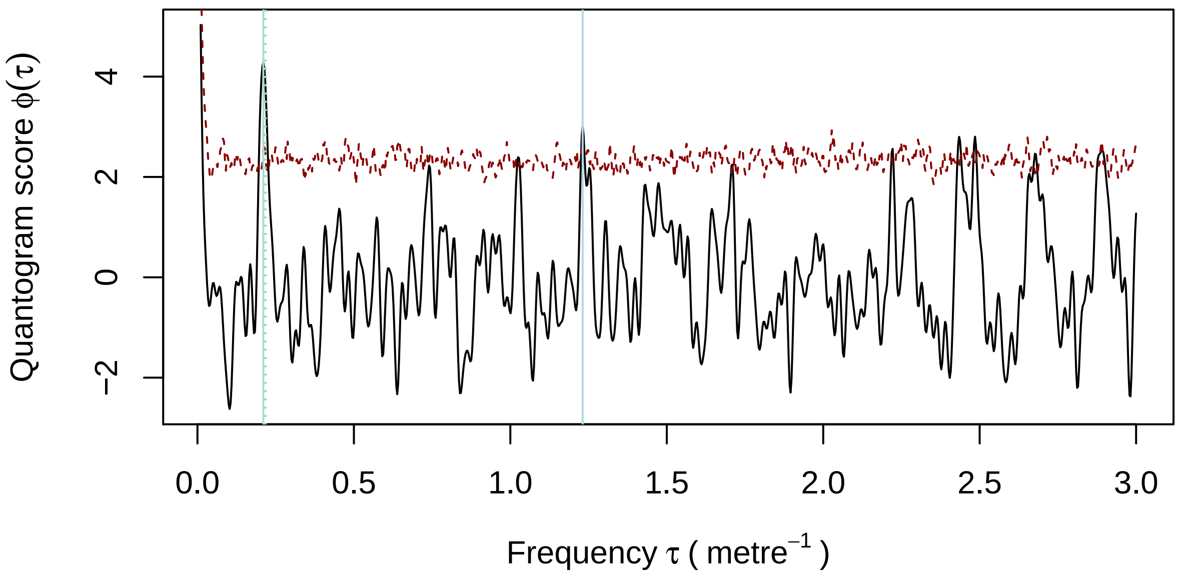

Figure 8: Quantogram with simulation upper envelope (based on differences of centre-lines from baselines).

Inspection suggests that the previously mentioned peak ($\tau \approx 0.2$) is significant, in that it rises well above the upper envelope of the simulated quantograms. There is also a lesser peak at around $\tau = 1.2$ and a corresponding "second harmonic" at $\tau = 2.45$ which we choose to ignore. "Significance" here does not refer to a strict formal statistical significance, which is not appropriate in this exploratory mode unless the comparison between simulations and observed quantogram is carried out at a frequency specified *a priori*. Our interest is confined to peaks which rise substantially above the envelope. (This could be formalized using false discovery rate techniques, but we leave this for another time.)

We locate, and mark with (light blue) solid vertical lines, the two maxima which definitely rise above the simulation envelope, and compute the corresponding quanta. There are also (light green) dotted vertical lines marking frequencies corresponding to quanta of $q = 4.57$ and $q = 4.65$, suggested by the independent ground-plan analyses of Prof. Blair and Huggins [1991]. However these are not visually distinguishable from the peak at $\tau = 0.211$, equivalently $q = 4.75$ (see also the discussion in Section 3.4.2).

```
##  Frequency levels:     0.211 ,     1.23
##  Quantum levels:        4.75 ,    0.812
##  Quantogram heights:    4.28 ,     3.01
##  Envelope heights:      2.52 ,      2.6
```

The quantum at $\tau = 0.211$ metres^{-1}, equivalently $q = 4.75$ metres, metres appears to be of definite significance in this preliminary analysis. The quantum at $\tau = 1.2$ metres^{-1}, equivalently $q = 0.81$ metres, is of lesser significance, and we choose to give no further consideration to this lesser peak.

3.4 Refinements

3.4.1 Construction error

Suppose that measurements of differences between centre-lines were of the form $X = mq + \sigma\varepsilon$, where m is integral and ε is random with standard normal distribution, so that there is an underlying quantum q but actual differences between centre-lines are perturbed randomly. Thus σ could be deemed to be the error introduced at the construction stage, introduced by Saxon builders who based their constructions on multiples of the putative quantum. Then we might expect (1) to be maximized at around $\tau = 1/q$, with mean value

$$\mathbb{E}\left[\phi(1/q)\right] \quad = \quad \sqrt{\frac{2}{N}}\sum_{j=1}^{N}\mathbb{E}\left[\cos(2\pi\sigma\varepsilon/q)\right] \quad = \quad \sqrt{2N}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\cos(2\pi\sigma u/q)e^{-u^2/2}\mathrm{d}u \quad = \quad \sqrt{2N}e^{-2\pi^2\sigma^2/q^2}\,. \tag{2}$$

Hence we can employ the method of moments to arrive at a crude estimate of σ, namely

```
sigma <- quantum * sqrt( log(sqrt(2 * nrow(data)) / quantos[argfreq1]) / (2 * pi**2) )
```

$$\hat{\sigma} \quad = \quad q\sqrt{\frac{1}{2\pi^2}\log\left(\frac{\sqrt{2N}}{4.6088}\right)} \quad = \quad 0.866\,. \tag{3}$$

This should be compared with the indicative estimate of the quantum, namely 4.75, and might be viewed as an estimate of the deviation from quantized increments arising from construction and/or measurement.

3.4.2 Error of the quantum estimate

Note that σ does *not* measure the error of our estimate of the quantum! For that, we follow the appendices in D.G. Kendall [1974]. We first construct a version of the increments sequence which is rounded to give integer multiples of the estimated quantum. Then we construct a simulated version of the dataset using the construction error, and evaluate the quantogram and determine the resulting estimate. This is repeated a sufficient number of times (25 times below), so as to generate an estimate of the standard deviation to be expected. The family of simulated quantograms is graphed in Figure 9; a boxplot of the simulated quanta is given in Figure 10.

```
freqest <- seq(from=0.15, to=0.3, length=500)
quantosshort <- quantocalc(freqest, data$x)
```

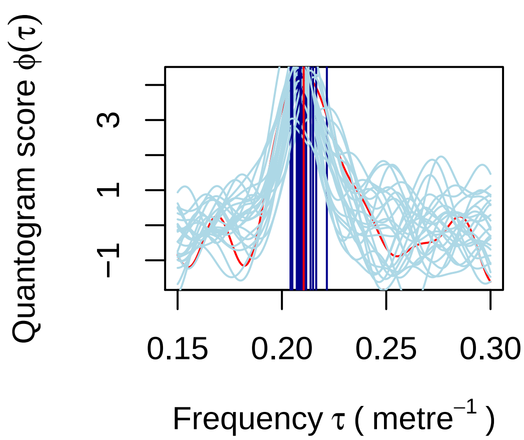

Figure 9: Quantograms of estimated increments (based on differences of centre-lines from baselines).

```
## Values from simulated sample:
##  [1] 4.620370 4.759180 4.800385 4.899362 4.759180 4.725379 4.807322
##  [8] 4.765998 4.513795 4.828254 4.745602 4.821256 4.786571 4.821256
## [15] 4.786571 4.779693 4.793468 4.779693 4.807322 4.793468 4.892157
## [22] 4.652681 4.678856 4.877810 4.786571
```

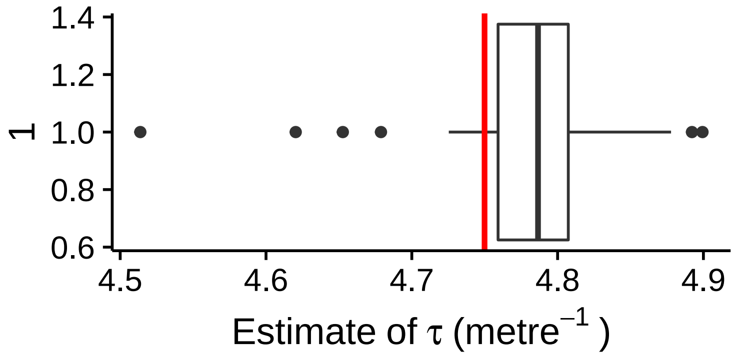

Figure 10: Boxplot of estimates of τ using simulated data. Red vertical line indicates actual estimate. (Based on differences of centre-lines from baselines.)

The estimated quantum is now given in metres by 4.75 ($\pm$0.17) (mean value $\pm$2 sd). Compare Huggins [1991]: $q = 4.65m$. Note that the larger Kendall [2013] error estimate ($\pm$2 sd $\approx$ 0.26 in metres) is based on distances between facing edges of walls; the Kendall [2013] analysis has a different starting point so there is no reason to expect agreement. The lower error estimate derived from centre-lines supports the impression of regularity given by Figures 4 (a) and (b). However these error estimates are best understood as merely being indicative of the order of magnitude of the possible estimation error.

Construction of the simulation sample begs several questions (particularly, whether the model $X = mq + \sigma\varepsilon$ is really a suitable model for the data); the main message of this error analysis is to note that the available measurement data permit an estimate of the quantum to an accuracy no better than $\pm$0.17 metres (2 standard deviations).

3.5 Using all possible differences

It is natural to consider whether it would be preferable to conduct an analysis based on all possible differences of measurements within each measurement line. A disadvantage of this approach is that we will collect disproportionately more differences from measurement lines containing more measurements, since we will form $\binom{m}{2}$ differences from a sequence of m cumulative measurements. On the other hand, we may gain extra evidence concerning the quantum, since there will be more large differences.

To do this, first collect all the differences generated by each set of averages.

```
multiincrement <- function (z) outer(z, z, "-")[lower.tri(mat.or.vec(length(z),length(z)))]
all.differences <- lapply(as.array(averages), multiincrement)
extract.function2 <- function(z) rep.int(as.vector(z)[2], length(z)*(length(z)-2)/8)
full.data <- data.frame(x=c(all.differences, recursive=TRUE),
                        s=c(lapply(line.s, extract.function2), recursive=TRUE),
                        n=c(lapply(line.n, extract.function2), recursive=TRUE),
                        o=c(lapply(line.o, extract.function2), recursive=TRUE))
head(full.data)
```

```
##         x s n o
## 1   9.000 1 1 2
## 2   4.075 1 2 2
## 31  4.125 1 3 1
## 32 13.500 1 3 1
## 33 17.600 1 3 1
## 34  9.375 1 3 1
```

The quantogram is now applied to the vector of differences, disregarding implicit dependences (arising because some centre-line differences will be sums of others).

```
quantos <- quantocalc(freq, full.data$x)
```

Determine the value of the observed quantum as before, based on Figure 14, noting that (as one might expect) there are subsidiary peaks at most integer multiples of the maximizing frequency.

```
argfreq <- max(quantos[freq > 0.1 & freq < 0.5])==quantos

allquantum <- 1 / freq[argfreq]
```

```
##  Frequency level:     0.208
##  Quantum level:        4.82
##  Quantogram height:    5.64
```

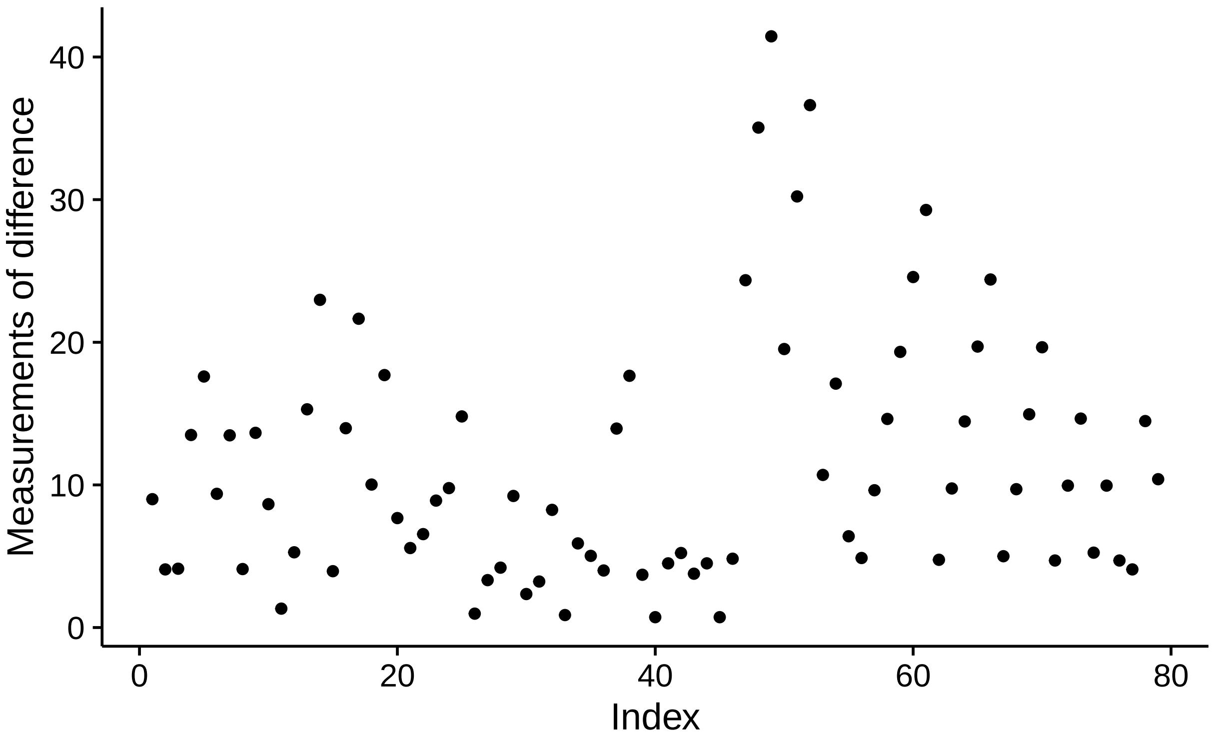

Figure 11: All possible differences within each line of measurement.

Now compute simulation envelopes, *based on perturbing the centre-lines computed from the original data-set.* This leads to a somewhat more laborious computation.

```
result <- c()

perturb <- function (m) m + rnorm(m, 0, smooth.sd)
for (i in seq(simulated.replicates)) {
  xsim <- as.array(lapply(averages, perturb))
  ynewsimall <- c(lapply(xsim, multiincrement), recursive=TRUE)
  quantossim <- quantocalc(freq, ynewsimall)
  result <- rbind(result, quantossim)
  }
upper <- apply(result, 2, sort)[dim(result)[1]-4,]
```

Estimate the “construction error” as before.

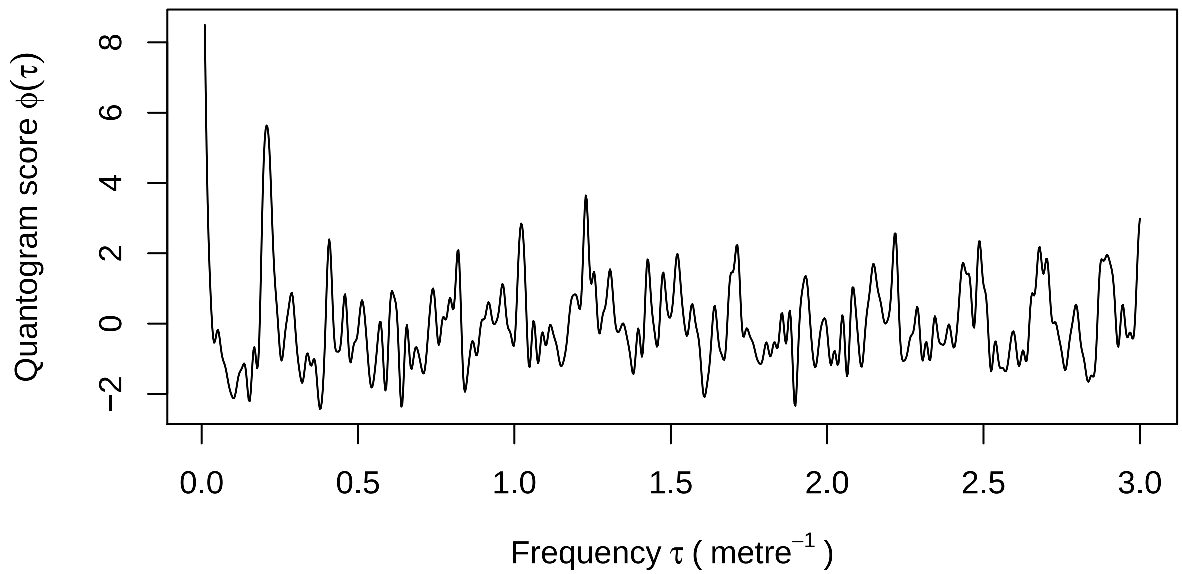

Figure 12: Quantogram (based on all possible pairs of centre-lines).

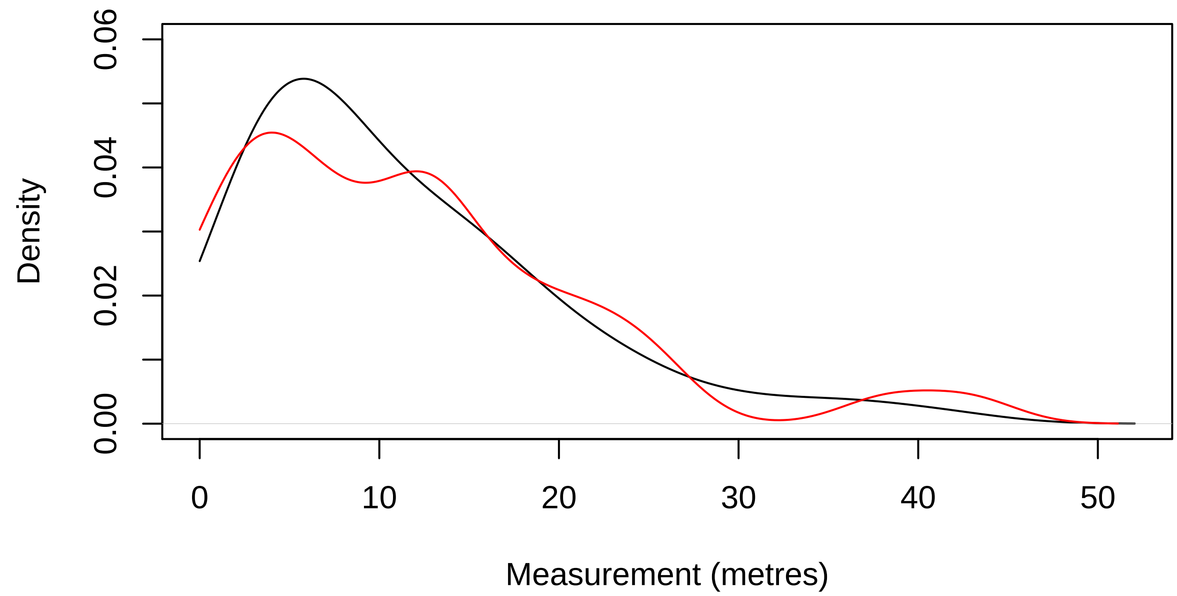

Figure 13: Smoothed density plots for laterally perturbed dataset (red curve) and original (black curve). (Based on all possible pairs of centre-lines.)

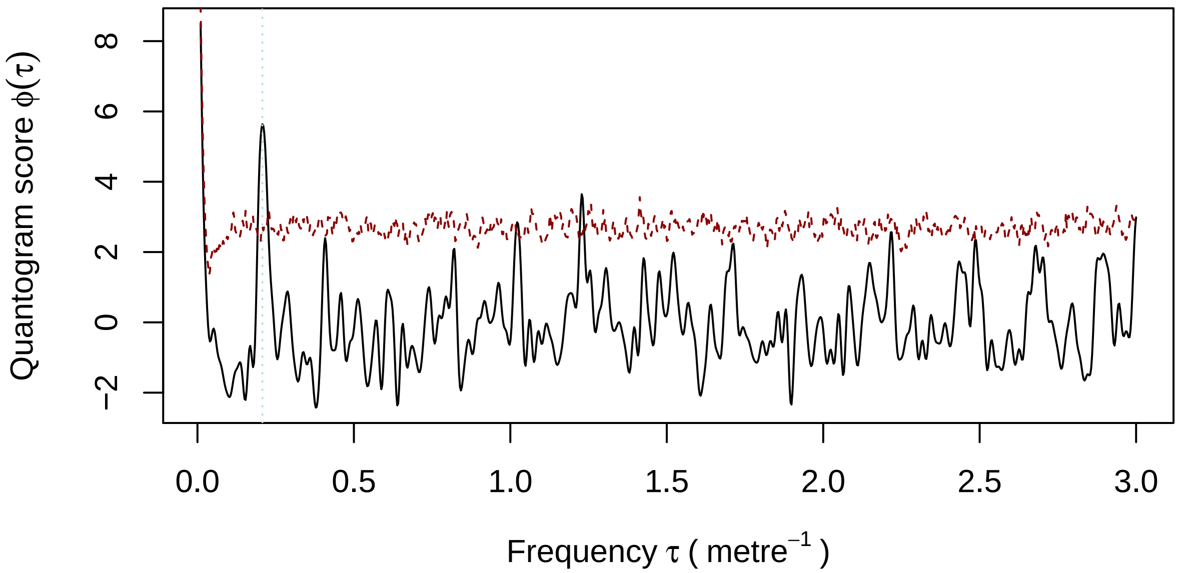

Figure 14: Quantogram with simulation upper envelope (based on all possible pairs of centre-lines).

```
sigma.all <- allquantum * sqrt( log(2 * sqrt(nrow(full.data)) /
                                        quantos[argfreq]) / (2 * pi**2) )
```

This is a little larger, namely 1.16, but it is reassuring that it is of the same order of magnitude as the previous estimate 0.866.

Estimate the error of the estimate for the quantum, once again proceeding as before.

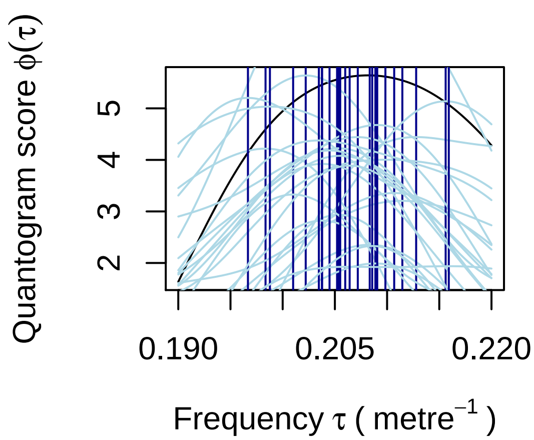

Figure 15: Quantograms of estimated increments near peak (based on all possible pairs of centre-lines).

```
## Values from simulated sample:
```

```
##  [1] 4.728961 4.799923 4.699567 4.854558 5.041423 4.945491 4.765543
##  [8] 4.746504 4.870193 4.637978 4.873047 5.084573 4.787489 4.975075
## [15] 4.907553 4.907553 4.783359 4.826386 4.865919 4.890239 5.030749
## [22] 4.914804 4.794389 4.631520 4.844660
```

The estimated quantum is now given by 4.82 (± 0.23) (mean value ± 2 sd). (Compare $q = 4.65$ from work of Huggins, 1991, and $q = 4.75$ (± 0.17) from the initial analysis given above in Section 3.3.) Note that the error

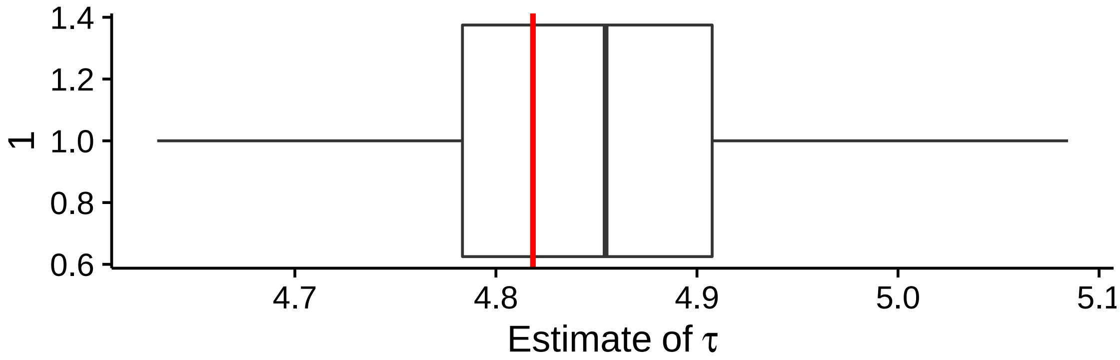

Figure 16: Boxplot of τ using simulated data. Red vertical line indicates actual estimate. (Based on all possible pairs of centre-lines.)

estimate is increased when the analysis is based on all differences in each measurement line. However the error estimate is still comparable to the accuracy with which the measurements were taken (see Appendix A). Generally the estimate does not significantly alter, whether we use all possible pairs or use only differences from the respective baselines.

4 Discussion

We have found some evidence for a quantum, with the most satisfying account being given by an analysis of measurements taken from the centre-lines of walls and analyzing all possible pairs (Section 3.5). The estimated quantum is now given in metres by 4.82 ($\pm$0.23) (mean value $\pm$2 sd). (Compare $q = 4.65$ derived from data found in Huggins, 1991.) Bearing in mind the estimated error, and the preliminary nature of this analysis, this is in general agreement with Huggins' work. Possibly the most useful contribution is to suggest that putative quanta are determined by the archaeological evidence to within a range of (about) $\pm$0.25 metres.

Residuals are computed as differences from fitted values, themselves obtained by rounding to the nearest mutliple of the quantum.

```
full.data$fittedvalues <- abs(round(full.data$x / allquantum) * allquantum)
full.data$residuals <- full.data$x - full.data$fittedvalues
```

We can gain some visual impression of how well the quantum fits by plotting residuals (data minus fitted values) against fitted values: see Figure 17. Horizontal lines indicate the extent of estimated "construction error" measureed by plus-or-minus 2.5×`sigma.all` (using 2.5 rather than 2 to allow for the relative paucity of data). A few measurements seem to lie on the margin of being questionable, bearing out the tentative nature of this analysis.

Boxplots of residuals per site (Figure 18) reflect the uneven-ness of the actual data here: Section 2.3 notes the number of measurements per measurement line; corresponding to this, the numbers of wall centre-liness per site (measuring the amount of information supplied per site) are given by

```
tabulate(data$s)
```

```
## [1]  5  2 10  5 12
```

Note from the geographical metadata that sites 1 and 2 are southern, sites 3 and 4 are nothern, site 5 is midlands. Bearing in mind the small number of measurements for southern sites, and the substantial estimated "construction error" `sigma.all`, it is hard to argue in favour of any geographical effects.

A corresponding set of boxplots of differences (Figure 19) does not suggest a difference of scale parallelling

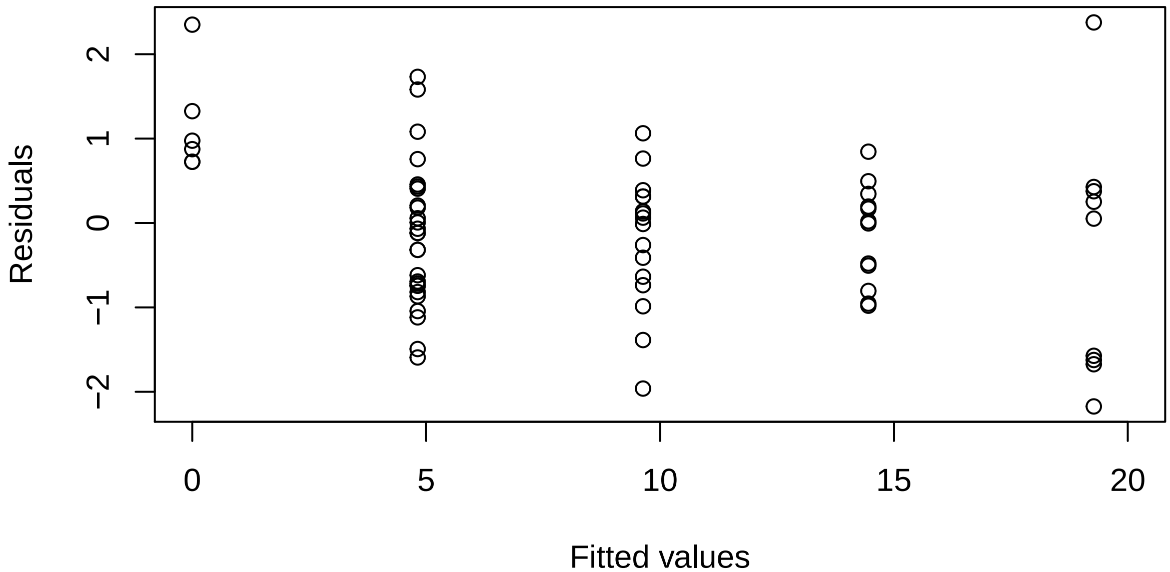

Figure 17: Residuals against absolute fitted values (based on all possible pairs of centre-lines).

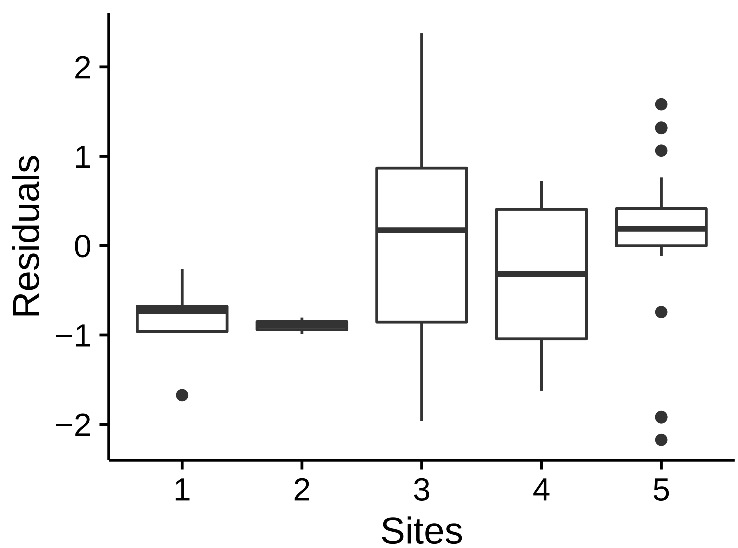

Figure 18: Boxplots of residuals per site (based on all possible pairs of centre-lines).

the southern-northern-midlands divide, though Brixworth (site 5) exhibits a wider range of measurement differences.

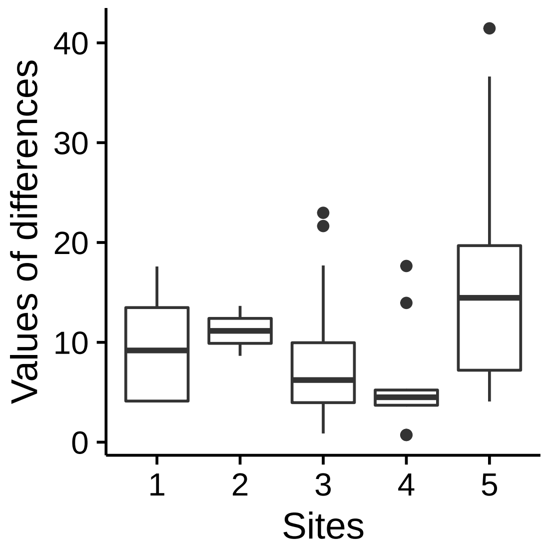

Figure 19: Boxplots of measurement differences per site (based on all possible pairs of centre-lines).

The actual numbers of wall centre-lines per orientation (Figure 20) are more evenly divided (hence yielding better boxplots of residuals). The numbers are given by

```
tabulate(data$o)
```

```
## [1] 15 19
```

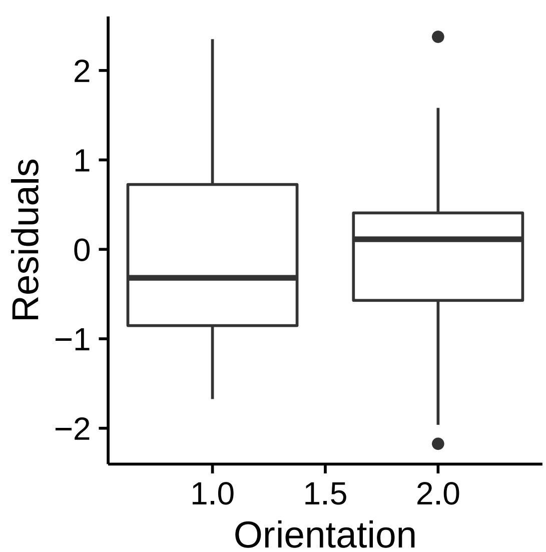

Figure 20: Boxplots of residuals per orientation (N/S *versus* E/W, based on all possible pairs of centre-lines).

Here are several further points which one would like to consider:

1. whether the length itself agrees with evidence from the subject-specific context;
2. whether the effect would be reduced or increased if we used more sophisticated models accounting for associated variables (differences between buildings, between different lines of measurement, between N/S versus E/W, restricting to increments deriving from facing walls, . . .);

3. we would like to consider whether the effect persists when more data is acquired. However data in this particular form (actual measurements from ground plans revealed by excavation) are rare. Some potentially relevant data can be obtained from maps of posthole locations. However there are considerable difficulties in reliably determining the measurements of the walls presumed to underlay these data. Barnes [2015], in investigations resulting in her MSc dissertation, found that the posthole data are more informative in yielding evidence for perpendicularity rather than measurement modulus.

The utility of more sophisticated analyses, for example accounting for effects as in point 2 above, is severely limited by the amount of data available.

References

Clair Barnes. *Statistics in Anglo-Saxon Archaeology.* Msc dissertation, University of Warwick, 2015.

John Blair. Grid-planning in Anglo-Saxon Settlements: the Short Perch and the Four-Perch Module. *Anglo-Saxon Studies in Archaeology and History*, 18:18–61, 2013.

Harald Cramér and M. Ross Leadbetter. *Stationary And Related Stochastic Processes: Sample Function Properties And Their Applications.* John Wiley & Sons Inc., 1967. ISBN 0486438279. URL http://books.google.com/books?id=H6CC1NUwSukC&pgis=1.

Sándor Csörgő. On the quantogram of Kendall and Kent. *Journal of Applied Probability*, 17(2):440–447, 1980. URL http://www.jstor.org/stable/10.2307/3213033.

Peter J Huggins. Anglo-Saxon Timber Building Measurements: Recent Results. *Medieval Archaeology*, 35: 6–28, 1991.

Peter J Huggins, Kirsty Rodwell, and Warwick J Rodwell. Anglo-Saxon and Scandinavian building measurements. In P J Drury, editor, *Structural Reconstruction: Approaches to the Interpretation of the Excavated Remains of Buildings*, page 65 pp. B A R British Series 110, 1982. ISBN 978-0-86054-193-6. URL http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Anglo-Saxon+and+Scandinavian+Building+Measurements#0.

David George Kendall. Hunting Quanta. *Philosophical Transactions of the Royal Society, London, Series A*, 276(1257):231–266, May 1974.

Wilfrid Stephen Kendall. Modules for Anglo-Saxon constructions: Appendix to "Grid-Planning in Anglo-Saxon Settlements: the Short Perch and the Four-Perch Module" by John Blair. *Anglo-Saxon Studies in Archaeology and History*, 18:55–57, 2013.

John T Kent. A weak convergence theorem for the empirical characteristic function. *Journal of Applied Probability*, 12(3):515–523, 1975. URL http://www.jstor.org/stable/10.2307/3212866.

A APPENDIX: description of the data

Here follows a description of the data-set produced by Prof. Blair.

`**KEY TO DATA**`

In each line, items of information are in the following sequence:

[distance], [number of site], [number of line of measurement], [1 = W-E, 0 = N-S], [1 = furthest face of wall from baseline, 0 = nearest face], [number of measurement from baseline in this row]

BUT in the case of the Hexham crypt, I have interpreted 'nearest face' and 'furthest face' as the internal faces of voids.

In the case of Escomb, the pairs of duplicated measurements in line 10 are because the inner faces of the nave walls are exactly in line with the outer faces of the chancel walls.

Occasional measurements in square brackets are cases where that section of wall is not now extant, but can be extrapolated from other evidence: they can be taken as reliable.

Each measurement is to the nearest 5 cm.

There are thirteen lines of measurements:

Canterbury, SS Peter and Paul (site 1).
Line 1: W to E, through S row of porticus.
Line 2: W to E, through nave.
Line 3: N to S, through nave and both rows of porticus

Canterbury, St Pancras (site 2).
Line 4: W to E, through nave
Line 5: N to S, through nave

Hexham (site 3).
Line 6: W to E axially, through crypt and detached eastern chapel.
Line 7: S to N, through crypt and both flanking porticus.
Line 8: S to N, through detached eastern chapel. **Important:** this line has an even number of measurements!

Escomb (site 4).
Line 9: W to E, through nave and chancel.
Line 10: S to N, on line of chancel arch wall.

Brixworth (site 5).
Line 11: W to E, axially through whole church.
Line 12: W to E, through N row of porticus.
Line 13: N to S, through W end of nave and porticus. **Important:** this line has an even number of measurements!

Appendix B provides a full listing of the data-set supplied by Blair.

B APPENDIX: data supplied by Blair

The above `rmarkdown` script uses the following partially cleaned data (which comprise the text file `Data/Measurements.txt`).

```
DATA

Canterbury, SS Peter and Paul

0.65, 1, 1, 1, 1, 1

9.15, 1, 1, 1, 0, 2

9.50, 1, 1, 1, 1, 3

0.65, 1, 2, 1, 1, 1

4.10, 1, 2, 1, 0, 2

4.70, 1, 2, 1, 1, 3

0.65, 1, 3, 0, 1, 1

[4.10], 1, 3, 0, 0, 2

4.80, 1, 3, 0, 1, 3

[13.50], 1, 3, 0, 0, 4

[14.15], 1, 3, 0, 1, 5

17.60, 1, 3, 0, 0, 6

18.25, 1, 3, 0, 1, 7

Canterbury, St Pancras

0.55, 2, 4, 1, 1, 1

13.60, 2, 4, 1, 0, 2

14.25, 2, 4, 1, 1, 3

0.55, 2, 5, 0, 1, 1
```

```
8.65, 2, 5, 0, 0, 2
9.20, 2, 5, 0, 1, 3
Hexham
1.35, 3, 6, 1, 1, 1
2.00, 3, 6, 1, 0, 2
5.95, 3, 6, 1, 1, 3
15.65, 3, 6, 1, 0, 4
16.30, 3, 6, 1, 1, 5
23.30, 3, 6, 1, 0, 6
24.00, 3, 6, 1, 1, 7
0.90, 3, 7, 0, 1, 1
5.55, 3, 7, 0, 0, 2
6.50, 3, 7, 0, 1, 3
7.00, 3, 7, 0, 0, 4
9.35, 3, 7, 0, 1, 5
9.85, 3, 7, 0, 0, 6
10.60, 3, 7, 0, 1, 7
[14.75], 3, 7, 0, 0, 8
[15.75], 3, 7, 0, 1, 9
5.55, 3, 8, 0, 0, 1
6.20, 3, 8, 0, 1, 2
9.55, 3, 8, 0, 0, 3
10.20, 3, 8, 0, 1, 4
Escomb
0.80, 4, 9, 1, 1, 1
13.90, 4, 9, 1, 0, 2
14.80, 4, 9, 1, 1, 3
17.70, 4, 9, 1, 0, 4
18.40, 4, 9, 1, 1, 5
0.70, 4, 10, 0, 1, 1
0.70, 4, 10, 0, 0, 2
1.45, 4, 10, 0, 1, 3
4.50, 4, 10, 0, 0, 4
5.20, 4, 10, 0, 1, 5
5.20, 4, 10, 0, 0, 6
5.95, 4, 10, 0, 1, 7
Brixworth
0.30, 5, 11, 1, 0, 1
1.10, 5, 11, 1, 1, 2
4.95, 5, 11, 1, 0, 3
6.10, 5, 11, 1, 1, 4
24.45, 5, 11, 1, 0, 5
25.65, 5, 11, 1, 1, 6
35.25, 5, 11, 1, 0, 7
36.25, 5, 11, 1, 1, 8
41.40, 5, 11, 1, 0, 9
42.90, 5, 11, 1, 1, 10
1.30, 5, 12, 1, 1, 1
4.95, 5, 12, 1, 0, 2
6.10, 5, 12, 1, 1, 3
9.65, 5, 12, 1, 0, 4
10.90, 5, 12, 1, 1, 5
14.80, 5, 12, 1, 0, 6
15.75, 5, 12, 1, 1, 7
19.50, 5, 12, 1, 0, 8
20.45, 5, 12, 1, 1, 9
24.80, 5, 12, 1, 0, 10
25.65, 5, 12, 1, 1, 11
29.50, 5, 12, 1, 0, 12
30.35, 5, 12, 1, 1, 13
1.10, 5, 13, 0, 1, 1
4.05, 5, 13, 0, 0, 2
5.20, 5, 13, 0, 1, 3
14.50, 5, 13, 0, 0, 4
15.55, 5, 13, 0, 1, 5
```

C APPENDIX: data taken from Huggins *et al*

Section ?? uses further data obtained from Huggins et al. [1982]:

HugginsRodwellRodwell-1982
TABLE 1 Primary sites (Page 28). Note emphasis on ratios!
West Stow Hall 1, 8.3, 4.2
West Stow Hall 2, 10.2, 4.3
West Stow Hall 3, 5.9, 4.2
Chalton, 11.8, 6.7
Cheddar E Hall 1, 33.9, 3.4
Cheddar W Hall 1, 16.8, 8.5
Cheddar Long Hall, 23.6, 5.1, 6.0
Cheddar Bldg N, 8.6, 6.9
Yeavering A4, 25.3, 11.8, 5.0, 3.4
Yeavering B, 11.8, 5.9
Yeavering A3, 30.5, 10.1, 28.4
Rivenhall Bldg 5, 15.0, 5.1
Nazeingbury Church, 5.0
Mucking 20, 9.9, 5.0

TABLE 2 (Page 33) disregarded as measurements are given only as estimates in terms of Northern Rods.

TABLE 3 Thetford (Page 34)
B, 5.44
C, 5.35, 14.13
D, 5.49
E, 5.20
G, 9.40

TABLE 4 Stone buildings in northern Essex (Page 56)
Rivenhall Nave, 15.10, 8.38
Rivenhall Chancel, 6.63, 8.38
Kelvedon Nave, 8.23
Inworth Nave, 6.71
Inworth Chancel, 4.88, 5.33
Alphamstone Nave, 13.11, 6.85
Little Birch Nave, 10.36, 6.71
Great Braxted Nave, 13.41, 8.08
Great Braxted Chancel, 5.03
ALresford Nave, 13.41, 6.63
Layer de la Haye Nave, 13.41, 8.08
Colchester Holy Trinity Tower, 5.03, 5.10
Great Tey Tower, 6.71, 6.71
Wendens Ambo Nave, 11.53, 6.85
Wendens Ambo Tower, 5.03, 5.03
West Mersea Tower, 4.90, 5.13

TABLE 5 Stone Buildings (Pages 58-59)
Silchester (Roman) Nave, 8.38, 4.16
Silchester (Roman) Nave + Aisles, 8.50, 6.70
Icklingham (Roman) Nave, 6.70, 4.20
Canterbury (Roman) Nave, 13.41, 8.68
Bradwell-on-Sea Nave, 16.66, 8.15
Canterbury SS Peter & Paul Nave, 18.29, 18.02, 8.08
Gloucester St Oswald Nave, 21.90, 9.90, 6.60, 6.65
Hadstock Nave, 7.02, 7.02, 5.25, 5.25
Alton Barnes Nave, 8.48, 5.28

```
Raunds phase 2 Nave, 8.50, 5.30
Raunds phase 2 Chancel, 5.30, 5.30
Ledsham Yorks Nave, 15.39, 6.71
Jarrow Durham Nave, 20.70, 6.55
Jarrow Durham Chancel, 6.55, 5.18
Jarrow Durham E Church, 13.11, 5.18
Jarrow Durham Bldg A, 27.00, 7.00
Monkwearmout Nave, 6.85
Monkwearmout Tower, 3.48, 3.35
Morland Tower, 3.40, 6.63, 3.40, 6.63
```