Distributed Unmetered Load Statistical Sampling Audit Guideline

Guidelines

25 October 2016

Version control

Version	Date amended	Comments
1	25 October 2016	Draft for consultation

Executive summary

This guideline sets out a methodology for statistical sampling of distributed unmetered load (DUML). The main type of DUML is street lighting, but there are other types.

The statistical sampling methodology is intended for use in assessing the accuracy of a participant's DUML database. This assessment is carried out as part of auditing the database.

The guideline provides guidance on how an auditor should:

- design the sampling programme
- carry out the sampling
- analyse the results
- form statistically valid conclusions.

It is not compulsory for an auditor to use statistical sampling. However, Part 15 of the Electricity Industry Participation Code 2010 (Code) requires the accuracy of the DUML database to be assessed. Statistical sampling is one method to provide a meaningful assessment.

Further, if an auditor does use statistical sampling, it is not compulsory for them to use the methodology set out in this guideline. However, if the auditor deviates from this methodology, they need to demonstrate that their alternative approach is statistically valid.

The information in this guideline is not intended to be definitive and should not be used instead of legal advice. If there is any inconsistency between this information and the Code, the Code takes precedence.

Contents

Figures

1 Introduction

What this guideline provides

1.1 This guideline sets out a methodology for statistical sampling of DUML.

1.2 The statistical sampling methodology is intended for use in assessing the accuracy of a participant's DUML database. It seeks to produce robust estimates of the total wattage[1] of DUML in the area covered by the database, which can then be compared with the total wattage shown in the DUML database. If the estimated wattage is significantly higher or lower than the database wattage, then that is evidence that the database is inaccurate.

1.3 The guideline provides guidance on how an auditor should:

(a) design the sampling programme

(b) carry out the sampling

(c) analyse the results

(d) form statistically valid conclusions and present these in their report.

1.4 It is not compulsory for an auditor to use statistical sampling. However, the Code requires the accuracy of the DUML database to be assessed, and statistical sampling is one method of providing a meaningful assessment.

1.5 Further, if an auditor does use statistical sampling, it is not compulsory for them to use the methodology set out in this guideline. However, if the auditor deviates from this methodology, they need to demonstrate that their alternative approach is statistically valid.

1.6 Note that this guideline focuses on street lighting, which makes up the majority of DUML. However, it is also applicable to other types of DUML.

Regulatory context

The Code

1.7 Part 15 of the Code governs reconciliation, including the requirements for audits of reconciliation participants.

1.8 Clause 11 of Schedule 15.3 of the Code provides that, "a retailer must ensure that an up-to-date database is maintained for each type of DUML for which it is responsible". The clause also sets out the requirements a DUML database must satisfy.

1.9 Clause 15.37B requires retailers to arrange for audits in respect of DUML and requires the auditor to verify that information recorded in the retailer's DUML database is complete and accurate.[2]

1.10 This guideline provides a methodology for assessing whether the total wattage in a DUML database is likely to be complete and accurate.

[1] The wattage is a measure of the total amount of power that can be consumed by the DUML, in MW. Another key measure is the *volume* of electricity consumed, in MWh-per-year terms. The volume can be calculated as the product of the wattage and the (assumed) average hours per year of operation.

[2] Clause 15.37B(b).

Other guidelines

1.11 The *Distributed Unmetered Load Audit Guideline* provides a structured approach for audits of DUML that is consistent across all auditors and participants. It also outlines retailers' Code obligations for DUML, and auditors' requirements for carrying out DUML audits.

2 Introduction to some statistical sampling concepts

Sampling

2.1 Sampling is the process of using a random sample to make inferences about the parent population. These terms are explained in the table below:

Term	Meaning
Population	The complete set of units
Units	For eg, items, places, or people
Sample	A subset selected from the population

2.2 In basic terms, spatial sampling means sampling using a map. In spatial sampling, the population is the entire area of interest, while the sample is the smaller subarea that is inspected.

2.3 In order for valid statistical inferences to be drawn, the sample needs to be selected randomly, ie, based on chance, rather than on judgement.

2.4 There are various ways of carrying out random sampling, some more complex than others. In "simple random sampling", each unit in the population can be selected with equal probability.

2.5 Sampling can either be "with replacement", in which a unit can be sampled more than once, or "without replacement", in which a unit can be sampled once at most.

2.6 It is important to select the sample in a way that avoids bias. In this context, "bias" is a problem with the sampling process that causes the results to err in a particular direction.

2.7 "Sampling effort" refers to the total amount of units sampled. It is important to take a large enough sample to have adequate power. The "power" is the ability to detect some characteristic of the population successfully. In other words, if the sample is too small, it is unlikely to yield conclusive and correct answers.

Estimation

2.8 A "point estimate" is a single number which is an estimate of some characteristic of the population. It is common to estimate the population "mean" (average), but other quantities can also be estimated, eg:

(a) the population total

(b) the population "standard deviation" (a measure of variability)

(c) the ratio of the averages of two different quantities

(d) the difference between the averages of two different quantities.

2.9 The formula used to calculate a point estimate depends on the quantity being estimated, the way in which the sample was randomly selected, and the characteristics of the population (which may be known, assumed, or inferred from the characteristics of the sample).

2.10 A point estimate on its own does not show the level of uncertainty. The estimate could be very "precise" (high level of certainty), or very imprecise – it is impossible to tell by looking at the point estimate alone. For this reason, interval estimates are also used.

2.11 An "interval estimate" identifies a range in which the population characteristic is likely to fall. A narrow interval reflects an (apparently) high level of precision; a wider interval shows a lower level of precision.

2.12 A "two-sided" interval indicates that the population characteristic is likely to fall between some lower value and some upper value – for instance, between 8 and 12.

2.13 A "one-sided" interval indicates either that the population characteristic is likely to exceed some lower value, or that it is likely to be less than some upper value. For example, a one-sided interval might be ‘at least 7’, or ‘no more than 13’.

2.14 The most commonly used type of interval estimate is a "confidence interval". Confidence intervals can be one-sided or two-sided.

2.15 Confidence intervals are constructed using a "confidence level" – commonly 90 %, 95 %, or 99 %. (All else being equal, the higher the confidence level, the wider the interval.) By design, an X % interval should have an X % chance of being correct (ie, of containing the true value). For instance, 95 % of correctly calculated 95 % confidence intervals should include the true value.[3]

2.16 If an X % confidence interval does not contain some value Y, then we can say that the true value “is different from Y at the X% level of statistical significance”. Loosely speaking, it is unlikely (but still possible) that Y is the true value.

2.17 For instance, suppose a 95 % confidence interval for the difference between the means (averages) of two quantities is from 2 to 12. This interval does not include 0 (which would be the difference, if the means of the two quantities were the same). Therefore, we can say that ‘the means of the two quantities are different at the 95 % level of statistical significance’. In other words, there is reasonably strong evidence that the averages of the two quantities are not the same.

2.18 The method used to calculate an interval estimate depends on the same kinds of factors as for a point estimate. Generally, calculating interval estimates is more complicated than calculating point estimates.

Stratification

2.19 Stratification refers to dividing the population into groups or segments (the strata), and sampling separately from each group. The strata should cover the entire population, and should not overlap. In other words, each unit should be in exactly one stratum.

[3] Of course, in practice it will be impossible to tell which are the ‘bad’ 5% of intervals that fail to include the true value, since the true value is unknown

2.20 In spatial sampling, strata typically reflect subareas. In other words, the entire area of interest is ‘carved up’ into strata, and units are then sampled from each stratum.

2.21 To add the most value, strata should be chosen so that the units within each stratum are (typically) more similar to each other than they are to the units in the other strata.

2.22 Stratification is best carried out before sampling (pre-stratification), as opposed to after sampling (post-stratification).

2.23 Modified formulas are used in the estimation stage, to take account of the stratification that has been carried out.

2.24 The main objective of stratification is to produce more accurate (point or interval) estimates. When stratification is carried out correctly, it can achieve this goal in two ways:

(a) At the sampling stage, by ensuring that sampling effort is spread appropriately across strata. (Without stratification, it might happen by chance that most or all of the sampling effort is allocated to a particular part of the population – possibly not the most important part – while obtaining little or no information on the other parts.)

(b) At the estimation stage, by leveraging the information that different strata may have different characteristics.

2.25 Stratification can have other benefits for auditors and audited participants, including:

(a) making surveying easier and/or cheaper

(b) producing more accurate estimates of the characteristics of each individual stratum (as opposed to those of the population as a whole).

3 How to carry out statistical sampling of DUML

Overview

3.1 The methodology consists of:

(a) planning the survey, by:

(i) identifying the area of interest (ie, the entire area covered by the DUML database)

(ii) dividing the area of interest into strata (optional)

(iii) breaking down the area of interest into a large number of area units

(iv) randomly selecting a sample of area units (this can be informally thought of as ‘tossing darts at a map’)

(b) visiting each area unit in the sample and observing all DUML (of the type of interest) within its boundaries

(c) using the results of the survey to form a 95 % confidence interval for the difference between total actual wattage and total DUML database wattage, across the entire area of interest

(d) drawing conclusions about the accuracy of the DUML database and documenting these in the auditor’s report.

3.2 The auditor should document the methodology they apply, with particular focus on any deviations from the methodology described here.

3.3 The following subsections detail the steps for clause 4.1(a) to (d) above.

A. Planning the survey

Identify the area of interest

3.4 The auditor's first step is to obtain a map, or maps, which cover the entire area of interest, eg, the area of the city council's responsibility (the city council being the customer for the streetlights).

3.5 Secondly, mark the area of interest on the map(s).

Optionally, divide the area of interest into strata

3.6 The auditor may choose to divide the area of interest into two or more strata, for convenience and/or to produce more accurate estimates.

3.7 If the auditor chooses to do this, then they should use judgement to identify appropriate strata. Some factors for the auditor's consideration are:

(a) the strata should reflect natural divisions in the area of interest

(b) strata should be chosen so that area units in the same stratum are (on average) more similar in attributes than area units in different strata

(c) the strata should be exhaustive, covering all DUML in the area of interest

(d) the strata should not be overlapping, so that each item of DUML should fall into exactly one stratum

(e) each stratum must contain some DUML (according to the DUML database)

(f) as a rule of thumb the auditor should have no more than five strata, otherwise some strata will have very little sampling coverage.

3.8 Some examples of strata that the auditor may to use are:

(a) CBD

(b) suburban

(c) new subdivisions

(d) rural

(e) medium-sized towns

(f) small towns.

3.9 Once the strata are identified, the auditor should:

(a) mark the boundaries between strata on the map

(b) determine which stratum each item of DUML in the database falls into

(c) determine the total wattage of DUML in each stratum, according to the database (these strata totals will be used in the estimation stage).

3.10 To carry out step (c) above (determine the total wattage of DUML in each stratum, according to the database), the auditor will need to be able to identify the location of each item of DUML in the database (eg, as latitude/longitude). **If there is not sufficient information in the database for the auditor to determine the location of each item of DUML, the statistical sampling approach is probably not suitable.**

3.11 The auditor should calculate the total wattages in accordance with the standardised table of wattages published by the Authority.

Subdivide the area of interest into area units

3.12 Break down the area of interest into a large number of area units. Mark the boundaries between area units on the map. Assign each area unit a number, or some other kind of label.

3.13 Create between 100 and 300 area units, depending on how large the area of interest is. If too few area units are created, the auditor will need to inspect a relatively high proportion of the DUML in the area of interest. If too many, the auditor will inspect a relatively low proportion of DUML, which may mean that the statistical power is insufficient to produce robust estimates.

3.14 The auditor should use judgement to decide how best to break down the area of interest, having regard to the practicalities of inspecting the area units. Some other factors for the auditor to consider are:

(a) it is preferable, but not essential, if each area unit contains a roughly similar amount of DUML (of the type of interest)

(b) an area unit that includes no DUML according to the database should not be created, unless the auditor suspects that it actually does contain DUML (eg, an area that appears as if it ought to have streetlights, based on inspection of a map)

(c) the area units should be exhaustive, covering all DUML in the area of interest

(d) the area units should not be overlapping, so that each item of DUML should fall into exactly one area unit

(e) if the auditor uses strata, an area unit should not cross the boundary between two strata.

3.15 The auditor may choose to use different approaches in different strata. For instance, the auditor may choose to:

(a) divide up a CBD area using a regularly spaced grid (Figure 1)

(b) make each new subdivision a separate area unit, with a 'blob-like' shape (Figure 2)

(c) make each small town a separate area unit (Figure 3).

Figure 1: Example – creation of area units using a regularly spaced grid

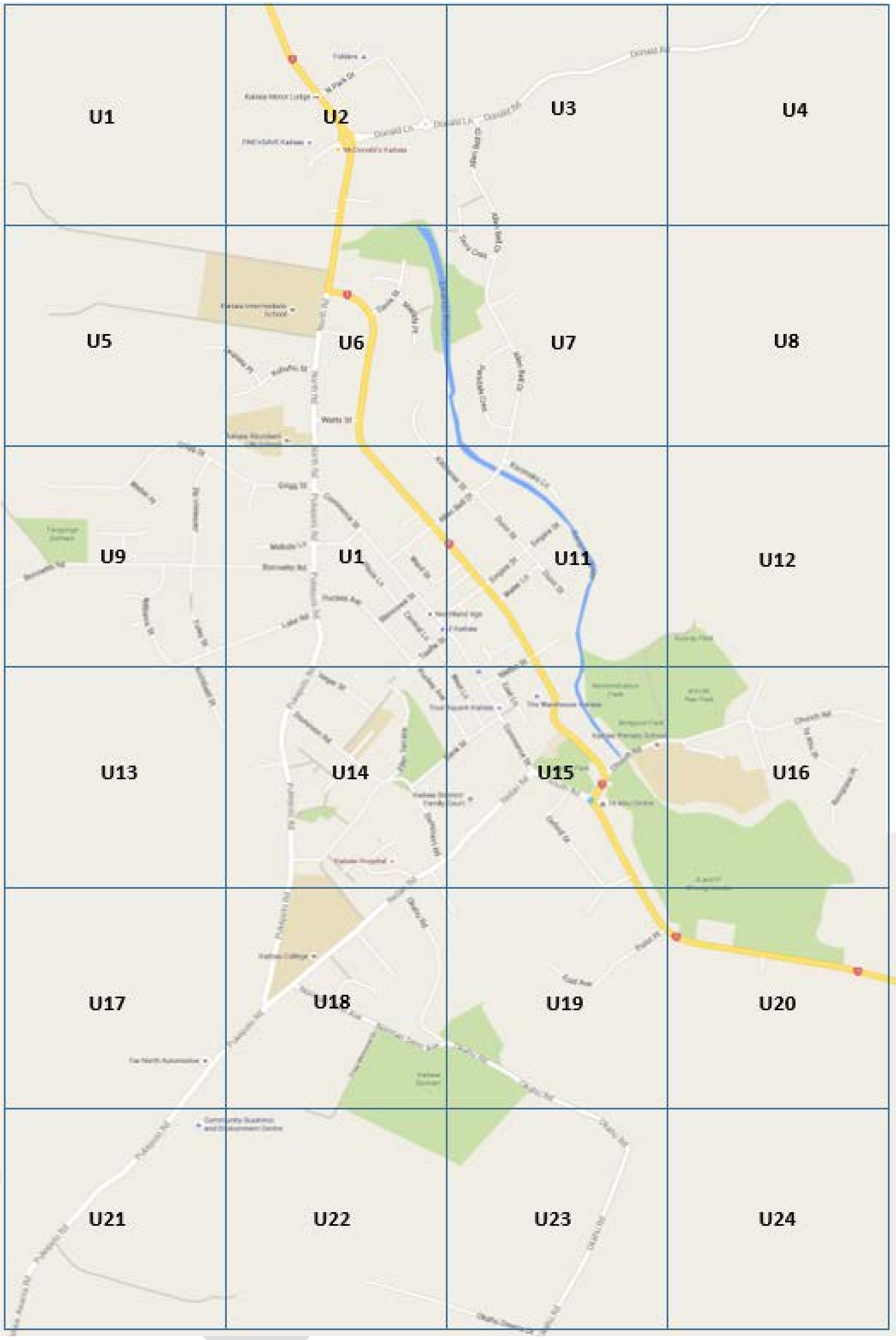
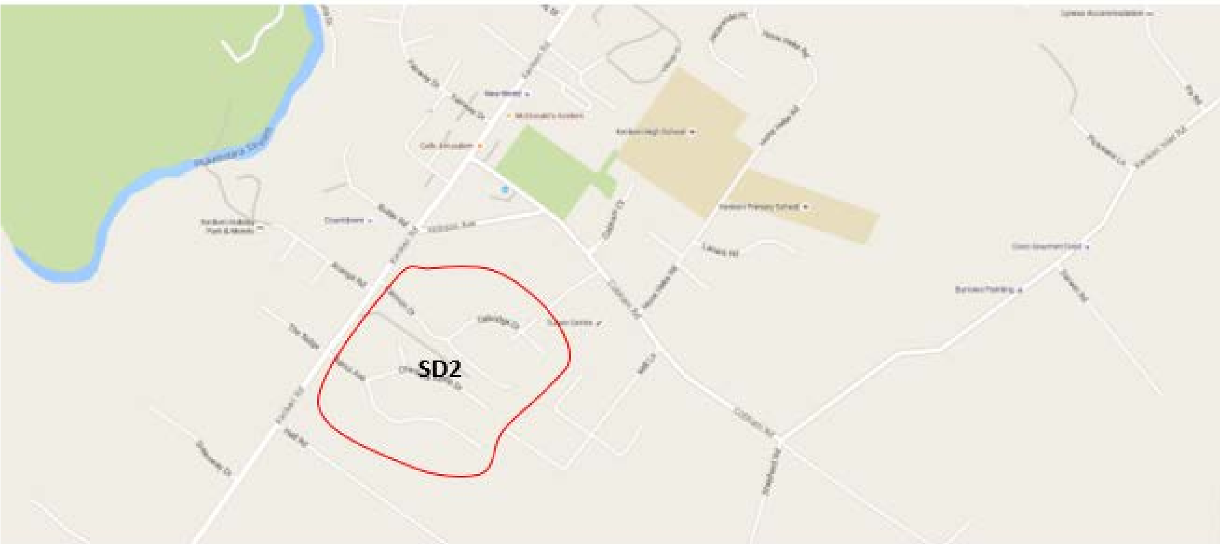


Figure 2: Examples – creation of area units using 'blobs' to delineate new subdivisions



Note that, if these 'blobs' were located in a larger area that was divided up using a regular grid, it would be necessary to 'cut holes' in the grid to make room for the blobs.

No item of DUML should sit both in a blob and a grid square. Each item of DUML should be in exactly one area unit.

Figure 3: Example – creation of area units, where each small town is a separate area unit

Determine the total amount of sampling effort

3.16 Decide *N*, the number of area units that you will inspect.

3.17 The Authority recommends that the auditor inspects at least *N*=20 area units. If the sample size is less than 20, it is less likely that there will be sufficient statistical power to produce robust estimates (Appendix A). This may manifest as very wide confidence intervals, showing a low level of precision.

3.18 The auditor may wish to sample more than *N*=20 area units, to achieve a higher level of precision.

If the auditor uses strata, divide the sampling effort between strata

3.19 The auditor should decide how to break down *N* between strata.

3.20 As a rule of thumb, the amount of sampling effort allocated to each stratum should be roughly proportional to the total DUML wattage in the stratum, except if the DUML database is known or suspected to be relatively inaccurate for a particular stratum, more sampling effort should be allocated to that stratum.

3.21 For instance:

(a) suppose the area of interest has been divided into three strata: urban (containing 50 % of the total DUML wattage), suburban (35 %) and new subdivisions (15 %). The auditor might allocate 10 samples to the urban stratum, 7 to suburban and 3 to new subdivisions

(b) however, if the auditor expected that the DUML database might be incomplete for some new subdivisions, they might double the number of samples to 6 for the new subdivisions stratum – at the cost of allocating just 8 samples to urban and 6 to suburban.

3.22 Each stratum must have at least 2 samples.

Randomly select a sample of area units to inspect

3.23 If not using strata, then randomly select *N* area units. If using strata, then for each stratum, randomly select the chosen number of area units.

3.24 Use simple random sampling without replacement, ie, assign an equal probability to each area unit, and select any given area unit once at most.

3.25 A web-based random number generator may be useful.[4]

[4] See e.g. https://www.random.org/integers/

B. Carrying out the survey

3.26 For each area unit in the sample, the auditor should:

- (a) visit the area unit and:
  - (i) find each item of DUML (of the type of interest) within the area unit, including items which are not present in the database (if any)
  - (ii) record the wattage of each item
  - (iii) sum these wattages to calculate a *total observed wattage* for the area unit
- (b) identify all items of DUML within the area unit, according to the DUML database, and sum their wattages to calculate a "total database wattage" for the area unit.

3.27 Wattages should be calculated in accordance with the standardised table of wattages published by the Authority.

3.28 At the end of this procedure, the auditor should have two numbers for each of the N≥20 area units: a total observed wattage and a total database wattage. These figures will be used in the estimation stage.

3.29 Where there are items of load that are not part of the database (for example private streetlights), the auditor should identify the ICP that the items of load are associated with.

C. Analysing the survey results

3.30 This step focuses on estimating *R*, the ratio of the true total wattage to the true DUML wattage across the entire area of interest.

(a) if *R* is less than 1, then the DUML database overstates the amount of DUML

(b) if *R* is 1, then the DUML database is perfectly accurate (in terms of the amount of DUML)

(c) if *R* is greater than 1, then the DUML database understates the amount of DUML.

3.31 The auditor should calculate a 95 % confidence interval for *R*.

3.32 The auditor should also express this confidence interval in terms of the difference between the true total wattage and the true DUML wattage, both in:

(a) wattage terms (MW)

(b) volume terms (MWh per year).

3.33 The Authority has published a spreadsheet tool that the auditor can use to calculate the above confidence intervals. The spreadsheet includes user instructions covering:

(a) software requirements

(b) how to enter the input data, which are:

(i) a list of strata, with the total DUML database wattage for each

(ii) a list of area units sampled, with the total observed wattage and total database wattage for each

(iii) the assumed 'on time' in hours per year (defaulting to 4,271 hours) which is used to convert from MW to MWh per year

(c) how to generate the confidence intervals.

3.34 Appendix B sets out the statistical methodology used in the spreadsheet.

3.35 If the auditor decides to generate the confidence intervals without using the spreadsheet tool, they should consider that:

(a) if they have used stratification, they should use a type of confidence interval that takes the stratification into account

(b) unless the sample size is substantially larger than N=20, they should use a type of confidence interval such as a bootstrap confidence interval[5] that does not rely on normal distribution assumptions.

[5] The most commonly used kind of confidence intervals are 't' intervals (see e.g. http://www.statisticshowto.com/how-to-find-a-confidence-interval/). These intervals tend to be inaccurate if (a) the sample size is small, and (b) the characteristic of interest is not normally distributed – i.e. does not follow a bell curve. Bootstrap intervals are calculated in a different way that does not rely on normal distribution assumptions (see e.g. https://en.wikipedia.org/wiki/Bootstrapping_%28statistics%29).

D. Interpreting the survey results

3.36 The auditor should include in their report the confidence intervals listed in the previous section.

3.37 Four possible scenarios are listed below. The auditor should indicate which scenario applies.

Scenario 1: Good accuracy, good precision

3.38 This scenario applies if:

(a) the point estimate of *R* is between 0.975 and 1.025

(b) the gap between the upper and lower bounds of the confidence interval for *R* is no wider than 0.05.

3.39 The auditor can conclude from this scenario that:

(a) the best available estimate indicates that the database is accurate within +/- 2.5%[6]

(b) the estimate is sufficiently precise.

3.40 This is the best outcome.

Scenario 2: Apparently good accuracy, but poor precision

3.41 This scenario applies if:

(a) the point estimate of *R* is between 0.975 and 1.025, but

(b) the gap between the upper and lower bounds of the confidence interval for *R* is wider than 0.05.

3.42 The auditor can conclude from this scenario that:

(a) the best available estimate indicates that the database is accurate within +/- 2.5%; but

(b) the estimate is not very precise: there is a significant probability that the database is less accurate than the above statement would suggest.

Scenario 3: Poor accuracy, demonstrated with statistical significance

3.43 This scenario applies if:

(a) the point estimate of *R* is less than 0.975 and the upper bound of the confidence interval for *R* is less than 1; or

(b) the point estimate of *R* is more than 1.025 and the lower bound of the confidence interval for *R* is more than 1.

3.44 The auditor can conclude from this scenario that the database is inaccurate. There is strong evidence to support this finding. In statistical terms, the inaccuracy is statistically significant at the 95 % level.

[6] This level of accuracy would be consistent with the accuracy requirement if the site was CAT1 metered, and is therefore deemed sufficient by the Authority.

Scenario 4: Apparently poor accuracy, but not demonstrated with statistical significance

3.45 This scenario applies if:

(a) the point estimate of *R* is less than 0.975 but the confidence interval for *R* still includes 1; or

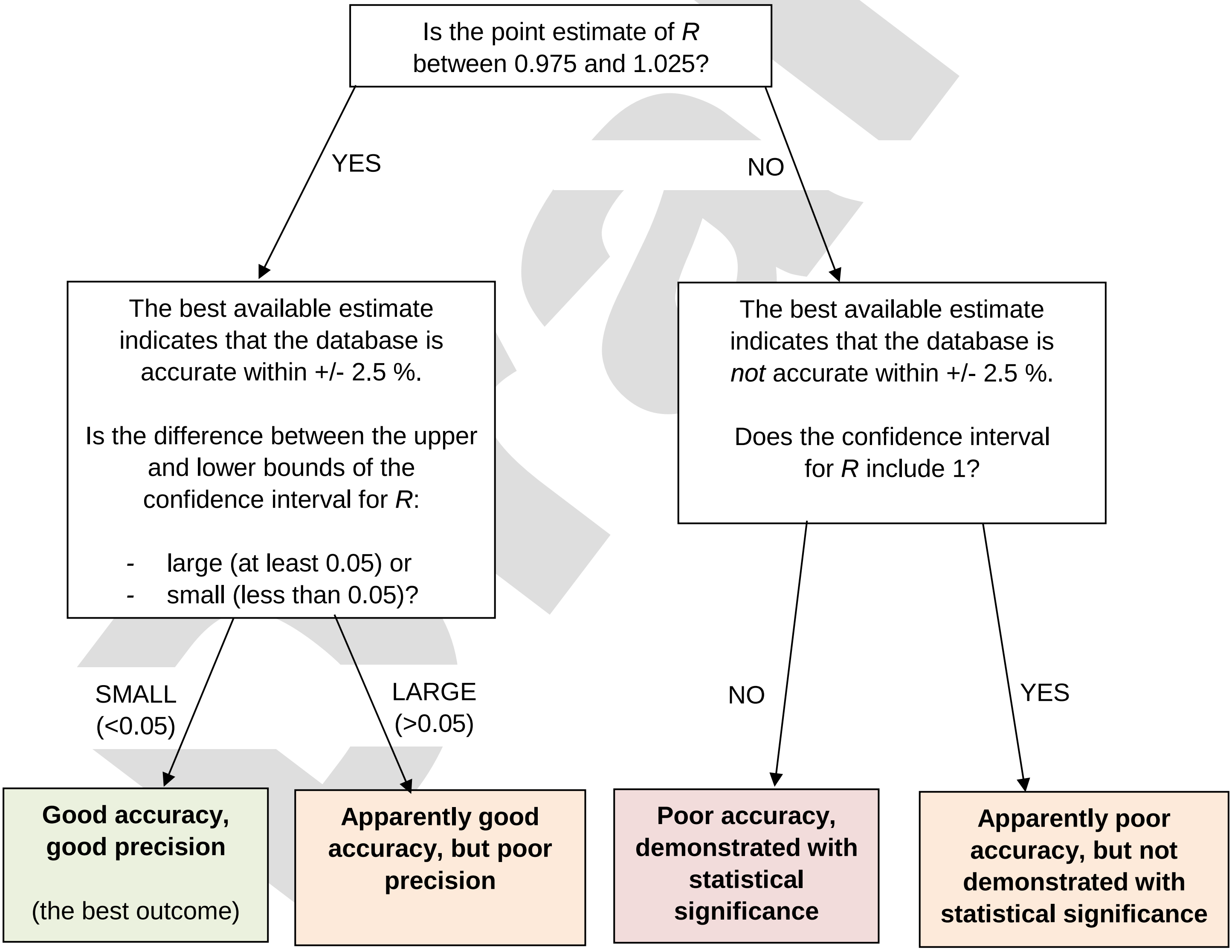
(b) the point estimate of *R* is more than 1.025 but the confidence interval for *R* still includes 1.

3.46 The auditor can conclude from this scenario that:

(a) the best available estimate indicates that the database is inaccurate, but

(b) the estimate is not very precise. It is still a credible possibility that the database is accurate. In statistical terms, the inaccuracy is not statistically significant at the 95 % level.

3.47 The process of determining which scenario applies is illustrated in the flowchart below:



4 Potential implications of the survey results

The reconciliation participant may be required to correct the database

4.1 The Authority expects reconciliation participants to arrange for the contents of the database to be corrected:

(a) if scenario 3 applies (“Poor accuracy, demonstrated with statistical significance”); or

(b) potentially if scenario 4 applies (“Apparently poor accuracy, but not demonstrated with statistical significance”).

4.2 It would not be adequate for the reconciliation participant to correct the inaccuracy by using *R* as a calibration factor (ie, simply multiplying their estimated electricity consumption by the estimated value of *R*). Rather, they should fix the errors in the database relating to specific DUML items.

Subsequent audits

4.3 The next audit date will be determined by a combination of factors, including:

(a) the estimated accuracy of the database, eg, the Authority may push back the next audit date because the database was found to be accurate

(b) the precision of the estimate, eg, the Authority may bring forward the next audit date in order for a more precise estimate to be obtained

(c) the amount of DUML in the database – because inaccuracy will have more impact on the market if there is a substantial volume of DUML

(d) the rate at which the database is changing – because change creates opportunities for inaccuracy

(e) the effectiveness of controls put in place by the database owner – because these controls lessen the likelihood of inaccuracy.

4.4 Based on the auditor’s report, the Authority is more likely to advance the next audit date if:

(a) scenario 2 applies (“Apparently good accuracy, but poor precision”), or

(b) scenario 3 applies (“Poor accuracy, demonstrated with statistical significance”), or

(c) scenario 4 applies (“Apparently poor accuracy, but not demonstrated with statistical significance”).

4.5 The next audit date would depend on the proposed corrective actions (if any) and the timeframe over which these actions were to take place.

4.6 Conversely, the Authority is more likely to postpone the next audit date if scenario 1 applies (“Good accuracy, good precision”).

Appendix A The statistical methodology used in the Authority's spreadsheet tool

Parameters of interest

A.1 The key parameter is *R* – the ratio of the true total wattage to the true DUML wattage across the entire area of interest.

A.2 Given a point or interval estimate of *R*, it is straightforward to convert to an estimate of:

(a) the difference *D* between true total wattage and true DUML wattage:

$\widehat{D} = (\widehat{R} - 1)$ x true DUML wattage

(b) the error *C* in annual energy consumption (MWh):

$\widehat{C} = \widehat{D}$ x assumed hours per year of operation.

Point estimation

A.3 In the absence of stratification, the point estimate of *R* is:

$$\widehat{R} = \frac{\sum_i O_i}{\sum_i D_i}$$

where:

O_i is the total observed wattage in the *i*th sampled area unit

D_i is the total DUML database wattage in the *i*th sampled area unit.

A.4 This is the classic ratio estimator.[7]

A.5 With stratification, the point estimate of *R* is:

$$\widehat{R} = \sum_s \left(T_s \frac{\sum_{i \in s} O_i}{\sum_{i \in s} D_i} \right) \Big/ \sum_s T_s$$

where T_s is the total DUML database wattage in stratum *s*.

A.6 This ratio estimator is known to be biased. Bias-corrected estimators are available, but are not recommended because:

(a) they are more complex and less intuitive

(b) the bias of the classic estimator tends to be small for moderate to large sample sizes

(c) the primary focus of the audit is on the interval estimate below – which includes bias correction – rather than on the point estimate.

[7] See e.g. https://en.wikipedia.org/wiki/Ratio_estimator

Interval estimation

A.7 The interval estimates produced by the spreadsheet tool are two-sided 95 % confidence intervals, produced using the accelerated bias-corrected bootstrap confidence interval of Efron and Tibshirani.[8]

A.8 Standard *t* or *z* confidence intervals would not be appropriate, because they are sensitive to normality assumptions unless the sample size is large. In this application, these assumptions do not hold – the sampling distribution of $\hat{R}$ is not normal and the sample size is unlikely to be large. Therefore the spreadsheet uses bootstrap confidence intervals, which do not rely on normality assumptions.

A.9 Bootstrapping is a common and highly flexible technique for producing confidence intervals (among other things). The basis of bootstrapping is to resample (with replacement) from the sampled data many times, and recalculate the point estimator for each resample. When the resampled point estimates are collated together, they form the bootstrap distribution, which can be used as an approximation to the true sampling distribution of the estimator. This approximation can be used to generate confidence intervals.

A.10 The implementation of bootstrapping in the spreadsheet tool resamples area units within each stratum separately, thus maintaining the distribution of sampling effort across strata.

A.11 By default, the spreadsheet tool carries out 20,000 bootstrap replications.

A.12 The spreadsheet checks for various (certain or potential) data entry errors, and also for the following mathematical problems:

(a) less than two samples in any stratum

(b) null DUML wattage in any stratum

(c) null variability in (O_i / D_i), and hence the bootstrap distribution collapses to a point.

[8] See eg, *An Introduction to the Bootstrap,* Efron, B. and Tibshirani, R.J., 1994, Chapman and Hall/CRC

Appendix B The rationale for recommending a sample size of at least N=20 area units

B.1 The recommendation to use at least N=20 is based on a statistical power calculation, described below.

B.2 The Authority defined two hypothetical scenarios:

(a) the red scenario (very poor DUML database quality) in which the ratio of O_i to D_i is uniformly distributed, in the range of:

(i) 1.1 to 1.2, for 20 % of area units (representing cases in which lights have been mistakenly omitted from the DUML database)

(ii) 1.0 to 1.1, for 27 % of area units

(iii) 1, for 50 % of area units

(iv) 0.9 to 1.0, for just 3 % of area units

(b) the orange scenario (moderately poor DUML database quality) in which the ratio of O_i to D_i is uniformly distributed, in the range of:

(i) 1.05 to 1.15, for 10% of area units

(ii) 1.0 to 1.05, for 27% of area units

(iii) 1, for 60% of area units

(iv) 0.95 to 1.0, for just 3% of area units.

B.3 The Authority sought to determine a sample size for which it would be:

(a) extremely likely that the *null hypothesis*[9] of R=1 would be rejected under the red scenario

(b) highly likely that the *null hypothesis* of R=1 would be rejected under the orange scenario.

B.4 In order to determine the required sample size, the Authority carried out the following steps:

(a) for a range of sample sizes N:

(i) generate 500 simulated samples of size N from the red scenario

(ii) for each of these simulated samples, calculate the 95 % bootstrap confidence interval

(iii) deem the null hypothesis rejected if the lower bound of the bootstrap confidence interval exceeds 1

(iv) calculate the statistical power as the proportion of the 500 simulated samples in which the null hypothesis is rejected

(b) repeat the above process for the orange scenario

---

[9] The null hypothesis is the assumption to be tested using statistics. If there is sufficient evidence opposing the null hypothesis, it is rejected.

(c) determine the minimum sample size *N* for which the statistical power is at least 90 % under the orange scenario and at least 95 % under the red scenario.

B.5 The above process returned a minimum sample size of *N*=20 (though *N*=15 was nearly sufficient).

B.6 This calculation does not take stratification into account. Successful stratification may increase the statistical power beyond the level indicated above.

B.7 The above analysis was carried out in the R programming language[10] using the *boot* library.[11] Source code is available on request.

.

[10] https://www.r-project.org/

[11] https://cran.r-project.org/web/packages/boot/boot.pdf

Glossary of abbreviations and terms

accelerated bias-corrected bootstrap confidence interval	Type of bootstrap confidence interval with good statistical properties, developed by Efron & Tibshirani
area of interest	The entire area covered by the DUML database
area unit	One of a large number of subareas that could potentially be sampled in the statistical sampling process
Authority	Electricity Authority
bias	In this context, a problem with a statistical sampling process that causes the results to err in a particular direction
bootstrap confidence interval	Type of confidence interval based on resampling, that does not require normality assumptions
Code	Electricity Industry Participation Code 2010
confidence interval	A common type of interval estimate. Confidence intervals are constructed using a confidence level – commonly 90%, 95% or 99%. (All else being equal, the higher the confidence level, the wider the interval.) By design, an X% interval should have an X% chance of being correct (i.e. of containing the true value)
DUML	Distributed unmetered load
interval estimate	A range in which the population characteristic is likely to fall
mean	Average
null hypothesis	The null hypothesis is the assumption to be tested using statistics. If there is sufficient evidence opposing the null hypothesis, it is rejected
point estimate	A single number which is an estimate of some characteristic of the population
population	The complete set of units (e.g. items, places or people) of interest
power, statistical power	The ability to detect some characteristic of the population successfully
R	The ratio of the true total wattage to the true DUML wattage across the entire area of interest. (Also the name of a statistical programming language)
random sample	Sample chosen based on chance, rather than on judgement
sampling	The process of using a random sample to make inferences about the parent population
sampling effort, sample size	*N*, the total amount of units sampled

sampling with replacement	Sample in which a unit can be sampled more than once – c.f. **sampling without replacement**, in which a unit can be sampled once at most
simple random sample	Random sample in which each unit in the population can be selected with equal probability
stratification	Dividing the population into groups or segments (the **strata**), and sampling separately from each group
total observed wattage	In a sampled area unit, the total wattage of all units of DUML (of the relevant type) observed within the area unit
total database wattage	In a sampled area unit, the total wattage of all units of DUML (of the relevant type) according to the DUML database
two-sided interval	An interval estimate that indicates that the population characteristic is likely to fall between some lower value and some upper value. C.f. a **one-sided interval**, which indicates either that the population characteristic is likely to exceed some lower value, or that it is likely to be less than some upper value
volume	In this context, the volume of electricity consumed by the DUML (in MWh)
wattage	In this context, the total amount of power that can be consumed by the DUML (in MW)