ABS GEOGRAPHY POPULATION WEIGHTED GRID CORRESPONDENCE DOCUMENTATION

Introduction

The ABS has developed a suite of geographical correspondences, primarily to assist users make comparisons and maintain time series between the previous Australian Standard Geographical Classification (ASGC) and the new Australian Statistical Geography Standard (ASGS). Correspondences are a mathematical method of reassigning data from one geographic region to another geographic region.

These correspondences have been created using a new methodology. It uses a population weighted grid, representing either 2006 Census Collection Districts (CD) or 2011 Mesh Block (MB) population data, which creates far more accurate correspondences than have been previously available.

While the ABS recognises that in many cases a correspondence is the only option available when attempting to convert data from one geographic region to another, caution should always be used when assessing the results of corresponded data, as they may not reflect the actual characteristics of a region. Issues surrounding the use of correspondences are discussed in the ABS publication: Information Paper: Converting Data to the Australian Statistical Geography Standard, 2011 (cat. no. 1216.0.55.004).

To assist users with making a determination of how well a correspondence may or may not convert data, the ABS has developed a quality indicator which is supplied with each correspondence.

This document details how the population weighted grid method produces correspondences, and provides a description of how the quality indicator is calculated.

Population Weighted Grid Correspondences

The population weighted grid method that the ABS has adopted is essentially a series of grid points that represent the underlying geographical distribution of the weighting unit, most often CD or Mesh Block population. Each grid point is then assigned a value based on this weighting. This is demonstrated in the example below.
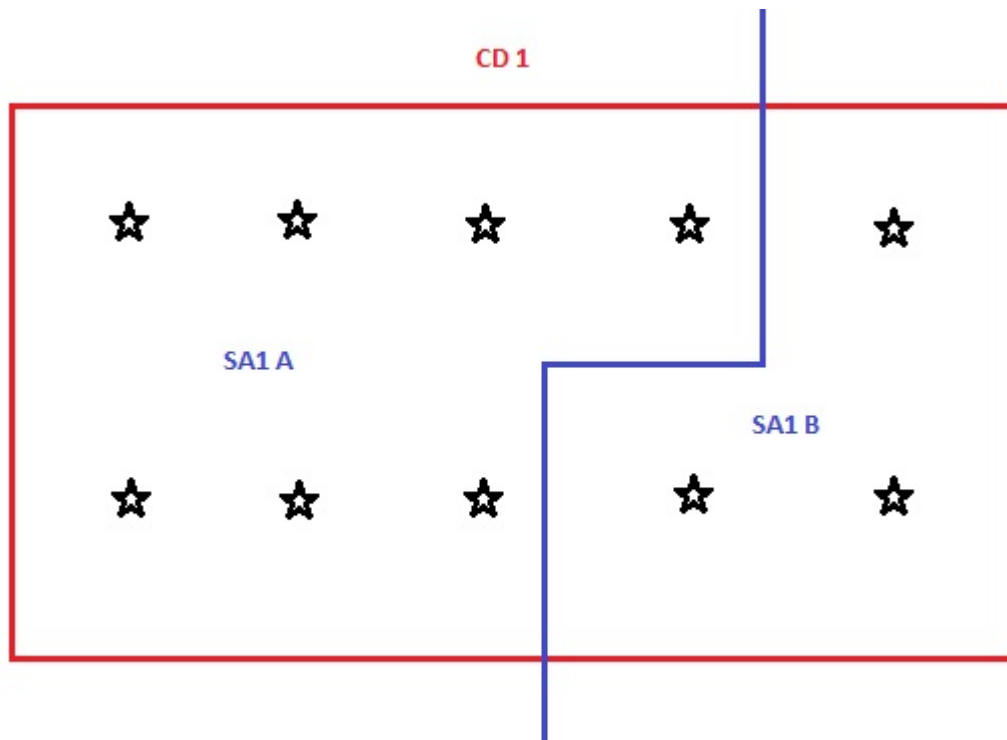
Figure 1: Example of a CD with population grid points and SA1 regions.

The correspondence in this example is from CD (the FROM region) to Statistical Area Level 1 (SA1) (the TO region), CD population is the weighting. The hypothetical CD above contains 200 persons. This population is represented by ten evenly distributed grid points, each grid point representing 20 persons.

The next step in the correspondence generation process is to determine the proportion that the CD, as the FROM unit, is donating to the respective SA1 TO units. As can be seen in the diagram above there are 7 grid points in SA1 A, and three in SA1 B. Given that each grid point represents 20 persons,140 persons are located in SA1 A and 60 in SA1 B. The proportion is then calculated by dividing the population found in each of the TO regions by the total population of the FROM region. Therefore the proportions are as follow:

- SA1 A: 140 / 200 which gives a ratio of 0.7 or 70 per cent.
- SA1 B: 60 / 200 which gives a ratio of 0.3 or 30 per cent.

So the result is that the CD in question is donating 70 per cent of its data to SA1 A, and 30 per cent of its data to SA1 B.

The benefit of using this method is that any two sets of geographic regions can have a correspondence generated for them, and that any attribute value can be distributed across the grid to be used as the weighting unit.

Quality Indicator

The change in geographical classification, with the ABS moving from the ASGC to the ASGS, has resulted in an increase in demand for correspondences to convert past data to the new ASGS. As a result the ABS conducted an investigation to determine how accurately correspondences converted data. This found that while some correspondences converted data well, there were many cases where the converted data did not reflect the actual characteristics of some geographical regions. Based on these findings a quality indicator was developed to inform data users of where the converted data values are likely to be accurate, and where caution will be needed to be used when assessing the results.

The method that has been developed to generate the quality indicator involves a number of steps. Firstly it looks at the value that a FROM region donates to a TO region as a ratio of the whole FROM region. The next step is to examine the value that the FROM region donates to the TO region as a ratio of the whole TO region. These two values are then multiplied together to provide the component for that FROM region. This process is then repeated for each donating FROM region, with the component values then added to provide the overall score for the TO region. Based on the score returned, a textual description is then applied as to how well the ABS expects data to be converted to the TO region. This is highlighted in the example below.
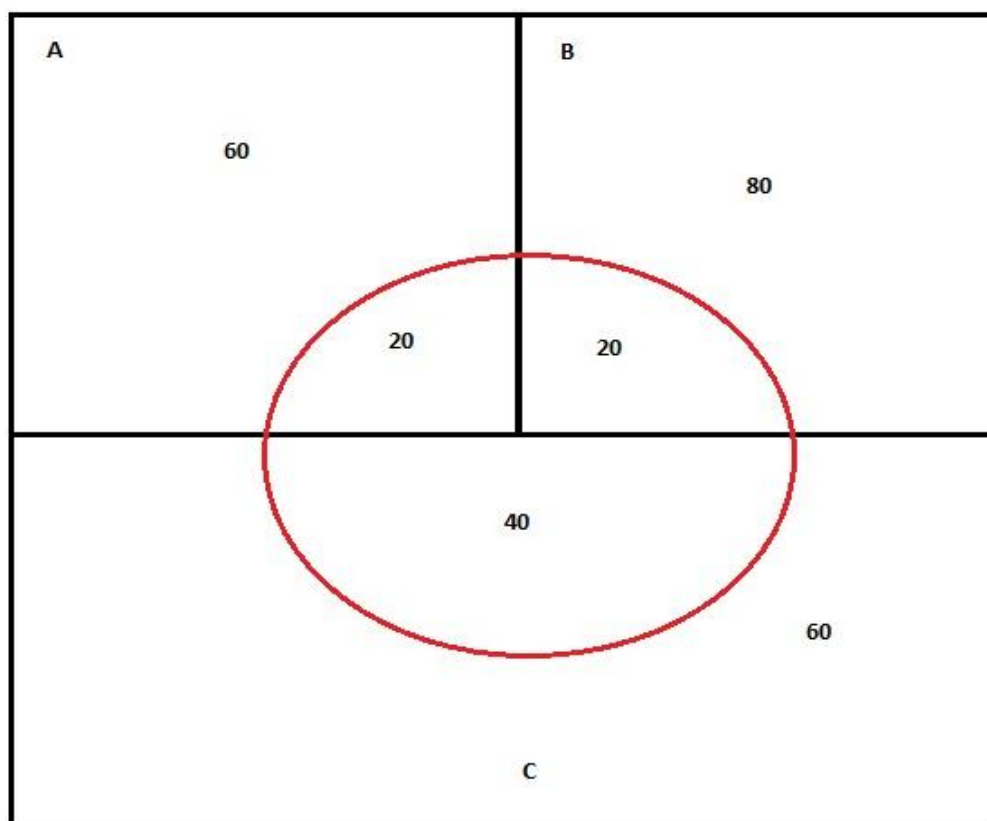

Figure 2: Illustration of 3 FROM regions to 1 TO region.

In this example there are three FROM regions A, B and C are represented by the black boundaries. The TO region is represented by the red ellipse.

Region A donates 20 persons to the TO region, while there are a further 60 people in FROM Region A that are not donated to the TO region. Therefore the ratio of FROM region A is 20 / 80, or 0.25. The next step is to look at the value that is being donated from Region A compared to the total value of the TO region. Region A donates 20 persons, and the total population is 80. So in this case the ratio is 20 / 80, or 0.25. Region A's component score is then calculated by multiplying 0.25 x 0.25 giving Region A a component score of 0.0625.

The same process is then applied to FROM Regions B and C. Region B donates 20 persons with a further 80 persons in the remainder of the FROM region. Therefore its ratio is 20 / 100 or 0.2. Region B donates 20 persons and the total population of the TO region is 80 so the ratio is 20 / 80 or 0.25. Region B's component score is therefore 0.2 x 0.25 or 0.05.

Again Region C donates 40 persons with another 60 in the remainder of FROM Region C. The ratio is 40 / 100 or 0.4. The 40 persons donated are then compared against the total population of the TO region of 80, so the ratio is 40 / 80 or 0.5. This results in the component score for From Region C being 0.4 x 0.5 or 0.2.

The final step is to add the three component scores. In this case:

- Region A = 0.0625
- Region B = 0.05
- Region C = 0.2

The final result is that the TO region in this example would have a quality indicator score of 0.3125, a score that the ABS would regard as being poor, meaning that caution would have to be used when using the results of data converted to the TO region.

The textual descriptions and there definitions that will be supplied for each TO region in a correspondence are as follows.

Good – The ABS expects that for this TO region the correspondence will convert data to a high degree of accuracy and users can expect the converted data will reflect the actual characteristics of the geographic regions involved.

Acceptable – The ABS expects that for this TO region the correspondence will convert data to a reasonable degree of accuracy, though caution needs to be applied as the quality of the converted data will vary and may differ from the actual characteristics of the geographic regions involved.

Poor – The ABS expects that for this TO region there is a high likelihood the correspondence will not convert data accurately and that the converted data should be used with caution as it may not reflect the actual characteristics of many of the geographic regions involved.

Overall Quality Indicator

An overall quality indicator is given to each correspondence. The aim of this is to provide users with a reasonable idea of how well the correspondence will convert data across the whole of the correspondence.

The overall quality indicator is derived from multiplying the population of each TO region with that TO regions quality indicator score, based on the methodology described above. The values produced by this multiplication for each TO region are then added together. This aggregated value is then divided by the total population of the TO regions. This will return a result similar to the individual quality indicator scores. Similar textual descriptions are then applied.

> **Good** – The ABS expects that the correspondence will convert data overall to a high degree of accuracy and users can expect the converted data will reflect the actual characteristics of the geographic regions involved.
>
> **Acceptable** – The ABS expects that the correspondence will convert data overall to a reasonable degree of accuracy, though caution needs to be applied as the quality of the converted data will vary and may differ in parts from the actual characteristics of the geographic regions involved.
>
> **Poor** – The ABS expects there is a high likelihood the correspondence will not convert data overall accurately and that the converted data should be used with caution as it may not reflect the actual characteristics of many of the geographic regions involved.

File Format

The correspondences will be supplied in Microsoft Excel format. Within each Microsoft Excel file there may be several Worksheets along with a Contents page and Explanatory Notes.

These Worksheets are as follows:

QI_MEASURE
This Worksheet contains the overall quality indicator in textual description. This Worksheet will always be supplied with correspondences.

QI_INDICATOR
This Worksheet contains the individual quality indicator in textual descriptions for every TO region.This Worksheet will always be supplied with correspondences.

CORRESPONDENCE
This Worksheet contains the main correspondence and the majority of the records. This Worksheet will always be supplied with correspondences.

NULL_TO_OR_FROM_FIELD

This Worksheet contains records where a FROM region does not have a corresponding TO region, or vice versa. An example of when this may occur is when one geography dataset contain islands which are not included in the other dataset. This Worksheet will only be supplied if records fall in to this category.

BELOW_MINIMUM_OUTPUT_SIZE

This Worksheet contains records that are below a pre-set minimum output size (typically below 0.01). These are records where the proportion of the FROM region that is being donated is very small and is deemed as being statistically insignificant. This Worksheet will only be supplied if there are records that fall in to this category.

MISSING_TO_UNITS

Contains records where the TO unit is not represented elsewhere in the correspondence. This is due to the TO unit being very small relative to the FROM unit and, as a result, a grid point is not associated with the TO unit. In cases where this occurs, documentation will be included with the affected correspondence as well as a list of the TO units that are not represented in the other Worksheets.

File Naming Convention For Grid Based Correspondences

Correspondence File Name

Grid based correspondences supplied by the ABS have a standard naming convention applied. The examples below relates to a correspondence where 2011 Statistical Areas Level 2 (SA2) are being corresponded to 2011 Local Government Areas (LGA).

File name:
Statistical Area Level 2 2011 TO Local Government Area 2011

and

CG_SA2_2011_LGA_2011.xls

Table 1: Character and meaning of the file name.

Character	Meaning
C	Correspondence
G	Grid based correspondence
SA2	Represents the name of the FROM region, in this case Statistical Area Level 2
2011	The year that this version of the FROM region was released
LGA	Represents the name of the TO region, in this case Local Government Area
2011	The year that this version of the TO region was released
.xls	The format that the file is being supplied, Microsoft Excel format

Correspondence Workbook and Field Definitions

Below is an example of the content for each of the Worksheets in the correspondence Microsoft Excel Workbook files provided in this publication. Definition of the fields in the Worksheets is also provided with the examples.

The QI_MEASURE Worksheet

Table 2: An example of the overall quality indicator of a grid based correspondence file.

QI_MEASURE
Good

In the above example the field name and descriptions are:

QI_MEASURE
The overall quality indicator for the entire correspondence.

The QI_INDICATOR Worksheet

Table 3: An example of the quality indicator of a grid based correspondence file for each TO region.

LGA_CODE_2011	LGA_NAME_2011	QI_INDICATOR
57630	Sandstone (S)	Poor
56160	Murchison (S)	Poor
34830	Mapoon (S)	Poor
37570	Wujal Wujal (S)	Poor
70540	Belyuen (S)	Poor
58470	Upper Gascoyne (S)	Poor
52380	Cue (S)	Poor
56860	Nungarin (S)	Poor
34570	Lockhart River (S)	Poor
55390	Menzies (S)	Poor

In the above example the field names and descriptions are as follows:

LGA_CODE_2011
This is a unique code associated with each TO region, to which a textual description of quality is supplied. In this case it is the unique Local Government Area code.

LGA_NAME_2011
This is the name of the Local Government Area which in this example is the TO region to which a textual description of quality is supplied.

QI_INDICATOR
This is the textual description of quality that is supplied for each TO region of the correspondence.

The CORRESPONDENCE Worksheet

Table 4: An example of a grid based correspondence file.

SA2_MAINCODE_2011	SA2_NAME_2011	LGA_CODE_2011	LGA_NAME_2011	RATIO	PERCENT
101011001	Goulburn	13310	Goulburn Mulwaree (A)	1	100
101011002	Goulburn Region	13310	Goulburn Mulwaree (A)	0.4690172	46.9017245
101011002	Goulburn Region	16180	Palerang (A)	0.0113619	1.1361892
101011002	Goulburn Region	17640	Upper Lachlan Shire (A)	0.5180712	51.8071226
101011002	Goulburn Region	18350	Wingecarribee (A)	0.0003445	0.0344547
101011002	Goulburn Region	18710	Yass Valley (A)	0.0012051	0.120509
101011003	Yass	18710	Yass Valley (A)	1	100
101011004	Yass Region	11050	Boorowa (A)	0.0229001	2.2900082
101011004	Yass Region	13700	Harden (A)	0.0290942	2.9094239
101011004	Yass Region	17640	Upper Lachlan Shire (A)	0.1131549	11.3154917
101011004	Yass Region	18710	Yass Valley (A)	0.8348507	83.4850664
101011005	Young	13700	Harden (A)	0.0108006	1.0800613
101011005	Young	18750	Young (A)	0.9891994	98.9199387
101011006	Young Region	11050	Boorowa (A)	0.2921527	29.2152719
101011006	Young Region	13700	Harden (A)	0.4283191	42.8319116
101011006	Young Region	18750	Young (A)	0.2795282	27.9528165

In the above example the field names and descriptions are as follows:

SA2_MAINCODE_2011
This is the unique numerical code representing the FROM region and in this case, the unique 2011 Statistical Area Level 2 code.

SA2_NAME_2011
This is a textual label associated with the unique code of the FROM region, in this case it is the textual label for each 2011 Statistical Area Level 2.

LGA_CODE_2011
This is the unique numerical code representing the TO region, in this case it is the unique

2011 Local Government Area code.

LGA_NAME_2011
This is a textual label associated with the unique code of the TO region, in this case it is the textual label for each 2011 Local Government Area.

RATIO
This field describes the Ratio of the FROM region that is being donated to the TO region. The Ratio is a figure between 0 and 1.

PERCENTAGE
This field describes the Percentage of the FROM region that is being donated to the TO region. The Percentage is the Ratio multiplied by 100.

The NULL_TO_OR_FROM_FIELD Worksheet

Table 5: An example of a table identifying NULL areas in either the TO or FROM region in a grid based correspondence.

SA2_MAINCODE_2011	SA2_NAME_2011	LGA_CODE_2011	LGA_NAME_2011	RATIO	PERCENT
801061066	Lake Burley Griffin			1	100

In the above example the field names and descriptions are as follows:

SA2_MAINCODE_2011
This is the unique numerical code representing the FROM region, in this case it is the unique 2011 Statistical Area Level 2 code.

SA2_NAME_2011
This is a textual label associated with the unique code of the FROM region, in this case it is the textual label for each 2011 Statistical Area Level 2.

LGA_CODE_2011
This is the unique numerical code representing the TO region, in this case it is the unique 2011 Local Government Area code. In the example above there is no LGA Code listed which indicates that the SA2 of Lake Burley Griffin does not correspond with any LGA.

LGA_NAME_2011
This is a textual label associated with the unique code of the TO region, in this case it is the textual label for each 2011 Local Government Area.

RATIO
This field describes the Ratio of the FROM region that is being donated to the TO region. The Ratio is a figure between 0 and 1.

PERCENTAGE

This field describes the Percentage of the FROM region that is being donated to the TO region. The Percentage is the Ratio multiplied by 100.

The BELOW_MINIMUM_OUTPUT_SIZE Worksheet

Table 6: An example of a table identifying ratios and percents of a TO region that is below minimum output size.

SA2_MAINCODE_2011	SA2_NAME_2011	LGA_CODE_2011	LGA_NAME_2011	RATIO	PERCENT
101011004	Yass Region	17500	Tumut Shire (A)	1.00E-07	9.80E-06
101031013	Bombala	10550	Bega Valley (A)	1.20E-06	0.0001241
101031015	Cooma Region	10550	Bega Valley (A)	2.20E-06	0.0002205
101041022	Deua - Wadbilliga	12050	Cooma-Monaro (A)	1.70E-05	0.0017011

In the above example the field names and descriptions are as follows:

SA2_MAINCODE_2011
This is the unique numerical code representing the FROM region, in this case it is the unique 2011 Statistical Area Level 2 code.

SA2_NAME_2011
This is a textual label associated with the unique code of the FROM region, in this case it is the textual label for each 2011 Statistical Area Level 2.

LGA_CODE_2011
This is the unique numerical code representing the TO region, in this case it is the unique 2011 Local Government Area code.

LGA_NAME_2011
This is a textual label associated with the unique code of the TO region, in this case it is the textual label for each 2011 Local Government Area.

RATIO
This field describes the Ratio of the FROM region that is being donated to the TO region. The Ratio is a figure between 0 and 1. In many cases, as can be seen in the example above, the amount that a FROM region is donating to a TO region is very small and is expressed as an exponential value.

PERCENTAGE
This field describes the Percentage of the FROM region that is being donated to the TO region. The Percentage is the Ratio multiplied by 100. In many cases, as can be seen in the example above, the amount that a FROM region is donating to a TO region is very small and is expressed as an exponential value.

The MISSING_TO_UNITS Worksheet

There may be cases where a TO unit is not represented in a correspondence file. This is due to the TO unit being very small relative to the FROM unit, and as a result a grid point is not associated with the TO unit. In cases where this occurs, an additional worksheet will be included with the affected correspondence file. It will consist of a list of the TO units that are not represented in any of the other Worksheets listed above.

Further Information

More information on the ASGS and ABS Statistical Geography can be found by visiting the ABS website: http://www.abs.gov.au/geography

Some Population Weighted Grid Correspondences are available in the following product:

Australian Statistical Geography Standard (ASGS): Correspondences, July 2011 (cat no. 1270.0.55.006)

Any questions, comments or further correspondence requests can be emailed to **geography@abs.gov.au**.