Geography Best-fit Matrix
Summary Report

Kate Trafford, GIS Analyst
Geographic Information Science and Analysis Team
Scottish Government

1. Introduction

Scottish Neighbourhood Statistics (SNS) is the Scottish Government's ongoing programme to improve the availability, consistency, and accessibility of small area statistics in Scotland. Information from SNS is used to inform the Scottish Government and other public sector agencies approach to delivering services and improving the quality of life for people living in Scotland. To disseminate information, SNS relies on the data zone geography as a building block to aggregate up to higher geographies. An exact match is achieved with geographies such as intermediate zones, however, in most cases the fit between data zones and other geographies is a 'best fit' match.

The Geography Best-fit Matrix provides a measure of how well the different SNS geographies match based upon the 'percentage fit' of populations covered by each geographical unit. The matrix allows policy makers and researchers to assess the viability of linking various geographically-based datasets, and to estimate the level of impact this may have on the results. The matrix is classified in terms of acceptability of fit in order to offer guidance on when matching geographies is or is not appropriate.

An MS Access database application was developed to analyse the core SNS geographies (summarised in Table 1.1) and a range of non-standard geographies for which SNS provides specialist reports. The overall fit is assessed in terms of postcode population matches between areas. The application can also be used to generate lookup tables between geographies.

Table 1.1: Core SNS geographies and their sources

Code	Geography	Source	Units	Year	Scotland Wide
PCI	GROS Postcodes	GROS	139,045	2001	Y
COA	Census Output Areas	GROS	42,604	2001	Y
DZN	Data Zones	Scottish Government	6,505	2001	Y
IZN	Intermediate Zones	Scottish Government	1,235	2001	Y
MMW	Multi Member Electoral Wards	OS Boundary Line	353	2007	Y
SPC	Scottish Parliamentary Constituencies	OS Boundary Line	73	2007	Y
CHP	Community Health Partnerships	Scottish Government	40	2008	Y
LA	Local Authorities	OS Boundary Line	32	2007	Y
HBA	Health Board Areas	Scottish Government	14	2008	Y
UR6	Urban/Rural Classification, 6 Fold	Scottish Government	6	2008	Y

2. Methodology

A population-based methodology is used to describe how well different geographies match. The 'percentage fit' between the initial ('from') geography and the aggregate ('to') geography is calculated as the percentage of the population in a single unit of the aggregate geography covered by associated units in the initial geography. For example, if linking data zones to multi-member wards, the fit would be calculated as the percentage of the population in each multi-member ward who are covered by data zones linked to that ward. This process is illustrated by Figure 2.1.

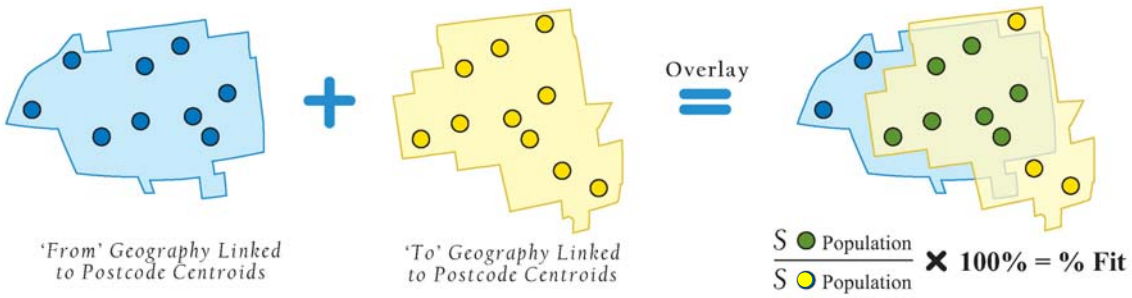


Figure 2.1: Population based method for measuring geography fit

As the available 2001 Census population figures are by postcode, the first step in the matching process is to create spatial joins between postcode centroids and each of the geographies in question. Next, an MS Access application is used to compare the postcodes in each initial ('from') geographical unit with the postcodes in the dominant unit of the aggregate ('to') geography. ('Dominant' means the aggregate unit which contains the majority of the population.) Populations are then summarised and the percentage population match between the two geographies is calculated. An overall measure of fit is obtained by averaging the percentage fit for each unit of the aggregate geography.

3. Output

The application produces several output tables for each pair of geographies compared, most importantly:

1. A lookup table linking each feature in the initial geography to a single unit in the aggregate geography (along with the percentage of population in each unit that falls in this dominant unit of the aggregate geography).
2. A table showing the percentage fit for each unit in the aggregate geography (along with a catagorisation of the quality of fit).
3. A summary table showing the overall percentage fit across the entire geography.

An example of one of the result tables (described in item 1 above) is shown in Table 3.1. Table 3.1 compares the fit of Data Zones to Agricultural Parishes. Columns prefixed by "FR_" show the geography type, code, name and population of the intitial from geography (data zones).

Table 3.1: Results table comparing Data Zones to Agricultural Parishes

RESULT_FILTERED_FR : Table

FR_TYPE	FR_CODE	FR_NAME	FR_POP	TO_TYPE	TO_CODE	TO_NAME	MATCH_POP	FR_PCT_FIT
DZN	S01000011	S01000011	918	PAR	11	Peterculter	918	100
DZN	S01000012	S01000012	874	PAR	473	Nigg	467	53.4
DZN	S01000013	S01000013	898	PAR	11	Peterculter	898	100
DZN	S01000014	S01000014	919	PAR	473	Nigg	539	58.7
DZN	S01000015	S01000015	655	PAR	473	Nigg	655	100
DZN	S01000016	S01000016	564	PAR	11	Peterculter	564	100
DZN	S01000017	S01000017	937	PAR	1	Aberdeen	472	50.4
DZN	S01000018	S01000018	577	PAR	1	Aberdeen	352	61
DZN	S01000019	S01000019	588	PAR	11	Peterculter	588	100

Record: 11 of 6505

The remaining columns prefixed by "TO_" show the type, code, and name of the aggregate geography (parishes), along with the matching population between the initial and aggregate geographies and the percentage fit for each unit. These results are then averaged to give an overall measure of fit, and classified to guide users in whether or not matching between the geographies in question is appropriate. The categories of fit are:

97 - 100 %	Recommended Fit	90 - 96.9 %	Acceptable Fit	Below 90 %	Unacceptable Fit

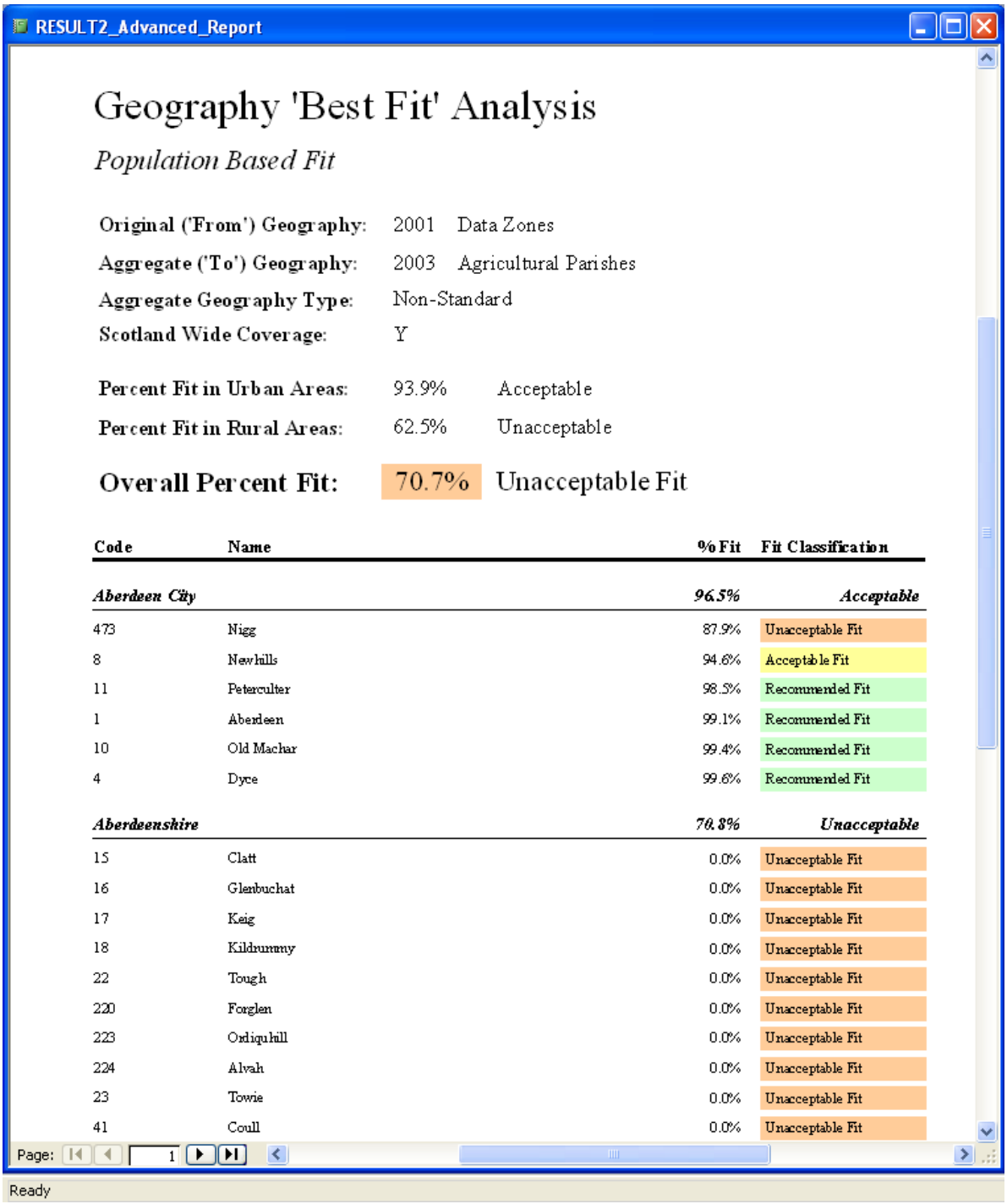
For matches of 97% and above, it is recommended that the geographies can be matched for data dissemination or analysis. Acceptable fit means that matching of geographies is appropriate though there is some level of acceptable error. Matches below 90% are deemed to be unacceptable and should not be used for dissemination or analysis. Note that the matches are Scotland-wide averages; locally there may be a much better or much worse fit. Analysis for all the core SNS geographies has been completed and results are shown in Table 3.2. Figures represent overall averages and in some cases, rounding has occurred (e.g. a value of 99.999 will appear as 100.0%). Values of 100% in bold text denote where an exact match has been achieved, and figures have not been rounded up.

Table 3.2: Geography Best-fit Matrix and Classification, for Core SNS Geographies

	Aggregate 'To' Geography								
	COA	**DZN**	**IZN**	**MMW**	**SPC**	**CHP**	**LA**	**HBA**	**UR6**
COA		**100.0%**	**100.0%**	99.4%	99.8%	100.0%	100.0%	100.0%	99.4%
DZN			**100.0%**	96.9%	99.0%	99.9%	100.0%	100.0%	97.4%
IZN				89.9%	97.6%	99.6%	100.0%	100.0%	94.2%
MMW					92.3%	99.4%	**100.0%**	100.0%	76.9%
SPC						89.1%	92.7%	98.8%	46.5%
CHP							**100.0%**	99.5%	43.7%
LA								99.5%	44.0%
HBA									34.6%

Example: Scotland-wide, the match from data zone to CHP is 99.9%, which is considered a 'recommended fit'.

The Best Fit Matrix has also been used by GI-SAT to create reports for more in-depth analysis of geographies. These reports allow the user to see whether geographies where the fit is poor are randomly spread across Scotland, or if there are just certain regions where the fit is very poor. An example report in Figure 3.1 shows results for the data zone to agricultural parish comparison. Results are broken down by Local Authority, and summary results tabulated for urban and rural areas.


RESULT2_Advanced_Report

Geography 'Best Fit' Analysis

Population Based Fit

Original ('From') Geography: 2001 Data Zones
Aggregate ('To') Geography: 2003 Agricultural Parishes
Aggregate Geography Type: Non-Standard
Scotland Wide Coverage: Y

Percent Fit in Urban Areas: 93.9% Acceptable
Percent Fit in Rural Areas: 62.5% Unacceptable

Overall Percent Fit: 70.7% Unacceptable Fit

Code	Name	% Fit	Fit Classification
Aberdeen City		***96.5%***	***Acceptable***
473	Nigg	87.9%	Unacceptable Fit
8	Newhills	94.6%	Acceptable Fit
11	Peterculter	98.5%	Recommended Fit
1	Aberdeen	99.1%	Recommended Fit
10	Old Machar	99.4%	Recommended Fit
4	Dyce	99.6%	Recommended Fit
Aberdeenshire		***70.8%***	***Unacceptable***
15	Clatt	0.0%	Unacceptable Fit
16	Glenbuchat	0.0%	Unacceptable Fit
17	Keig	0.0%	Unacceptable Fit
18	Kildrummy	0.0%	Unacceptable Fit
22	Tough	0.0%	Unacceptable Fit
220	Forglen	0.0%	Unacceptable Fit
223	Ordiquhill	0.0%	Unacceptable Fit
224	Alvah	0.0%	Unacceptable Fit
23	Towie	0.0%	Unacceptable Fit
41	Coull	0.0%	Unacceptable Fit

Page: 1

Ready

Figure 3.1: Sample report produced for Data Zone to Parish comparison.