MANCHESTER 1824

The University of Manchester

Cathie Marsh Centre for Census and Survey Research

Provide custom geography conversion lookup tables that enable area data disseminated for the 1991 Census wards to be compatible with the 2001 Census ward definition

CCSR Working Paper 2005-07

Paul Norman and Ludi Simpson
Paul.Norman@manchester.ac.uk, Ludi@manchester.ac.uk>

www.ccsr.ac.uk

Provide custom geography conversion lookup tables that enable area data disseminated for the 1991 Census wards to be compatible with the 2001 Census ward definitions

Consultancy work carried out by the Cathie Marsh Centre for Census and Survey Research (CCSR) on behalf of Ralph Smith, Associate Director of GIS and Computing, West Midlands Cancer Intelligence Unit, The Public Health Building, Birmingham, B15 2TT

Introduction

In England and Wales, between the 1991 and 2001 Censuses there have been minor changes to the ward boundaries in many locations and substantial restructuring of the ward definitions within various local authorities, in particular the new Unitary Authorities (UAs). Simpson (2002) and Norman *et al.* (2003) have devised methods to convert between different geographical systems (e.g. census and postal) and to establish time-series data on a consistent geographical basis in the face of (ward) boundary changes over time. These authors used lookup table (LUT) directories which link postcodes to other geographical areas to develop Geographical Conversion Tables (GCTs) which can be used to adjust socio-demographic data from the 'source' geography for which the data pre-exist, to the 'target' geography, the zonal system for which the data are needed.

In the GCT approaches referred to above, input data are disaggregated to sets of unit postcodes that comprise each source geography unit and then reaggregated into the target geography units using the sets of postcodes that comprise the wards in that boundary system. The assumption is that the distribution of residential postcodes are a proxy for population distribution. Since people are not evenly distributed across postcodes (with an urban-rural gradient, for example) the disaggregations are enhanced by the use of additional attribute counts at each postcode. To date, the disaggregation weights have been calculated using address and household counts that fall in each intersection between the source and target geographies.

The Small Area Health Statistics Unit (SAHSU) make available ward level population estimates by age and sex for use in health research. These population estimates are disseminated for the 1991 ward geography for each year from the early 1980s to the late 1990s. This new consultancy being reported here is needed because population data are required for the time-series but based on the 2001 ward definitions so that analyses are relevant to a contemporary geography.

There are various facets to the work that has been carried out:

- Create a geography conversion table (GCT) from all 1991 Census wards in England and Wales to 2001 Census wards based on recently released 2001 information
- Create a further GCT from the 1991 Census wards to the 2001 Census wards based on the 1991 postcode-household count distribution
- Provide a computer program to automatically convert files of data disseminated for the 1991 wards to the 2001 boundary definitions
- Validate the GCTs and outputs and revise the above accordingly.

These are described in this report. Program user instructions and computer code are given in an appendix. Apart from the custom written conversion program, the majority of the data preparation and checking has been carried out using SPSS, ArcView and Excel.

Create a geography conversion table (GCT) from all 1991 Census wards in England and Wales to 2001 Census wards

To operationalise the conversion between 1991 Census wards and 2001 Census wards, the GCTs need to be assembled in which every residential postcode has the following information attached:

- A reference to the source 1991 ward with which the postcode is associated
- A reference to the target 2001 ward with which the postcode is associated
- A count of a further attribute, such as population or households, at each postcode

For clarity, it is worth noting that in 1991 the geographical Census hierarchy of National > Standard Region > Local Authority (LA) > Ward > Enumeration District (ED) existed. In 2001 the equivalent Census hierarchy is National > Government Office Region > Local Authority (LA) (including the new UAs) > Ward > Output Area (OA). Relevant to this application, pre-2001 directories associated postcodes with EDs and 2001 directories associate postcodes with OAs. Since EDs and OAs perfectly aggregate to the wards in the respective census years, much of the data preparation for the GCTs will be at these sub-ward areas.

Two GCTs have been developed. The first is based on recently released 2001 Census postcode-headcounts. A reason to use this information is because counts of persons, males, females and households are available. Simpson (2002) recommends that the GCT weighting criterion should be correlated with the data being converted and, since the distribution of persons, males or females may be different to the distribution of households, it is anticipated that the use of these will enhance the previous use of just

address or household counts. The second GCT is based on the 1991 Postcode-Enumeration District Directory, a file which contains household counts for each postcode-ED intersection. The development of these GCTs is described below.

Creation of a 1991 to 2001 ward GCT with 2001 information as the starting point

During 2003 the Office for National Statistics (ONS) released LUT directories relating to the 2001 Census. First released was the Postcode-OA (PCOA) directory followed by the Postcode Headcount file (PCHC). The latter contains counts of persons, males, females and households at each postcode and, since also included is a reference to the OA with which each postcode is associated, the former file is redundant. The missing information for this application is a link back in time to the 1991 EDs and this needs to be attached.

The All Fields Postcode Directory (AFPD) with postcode information for 1999 is the most recent file available to the academic community that potentially contains the required information, however the file is becoming dated. West Midlands Cancer Intelligence Unit (WMCIU) made available to this work an extract from the National Health Service Postcode Directory (NHSPD) dated as close to the 2001 Census as possible (version 2001/A). For each postcode, this file contains links to various administrative zones including 1991 Census EDs. Dates of postcode introduction and termination along with indicators of user type and PO Box allow the selection of geographical, residential postcodes existing during the early part of 2001.

The cleaned NHSPD for 2001 was joined to the PCHC using the postcodes as the common item. This was carried out to add the 1991 ED information to the 2001 headcounts and OA locations. Of the 1,300,523 records in the PCHC file, all but 21,596 (i.e. 1.66%) were matched. The matched postcodes were annotated as a ‘Type 1’ match.

The unmatched 21,596 postcodes extracted from the PCHC file were matched with the AFPD to add the 1991 ED information wherever possible. The result of this was that a further 9,667 postcodes had the 1991 information added (a ‘Type 2’ match). 11,930 (0.92% of the original total) remained unmatched. The list of unmatched postcodes was found to contain 1,144 records in which the headcount file indicated both zero population and zero households. Since these could have no influence on the source-target overlap weights, these records were removed. To match the remainder, the OA population weighted centroids were joined from the digital boundary files supplied by ONS to the OA with which each postcode is associated. The GIS point in polygon technique was used to determine which ED the centroid lay within.

The remaining 10,786 records had the 1991 ED codes added in this way and were termed a ‘Type 3’ match.

The distribution of the 20,453 postcode matches types 2 and 3 (Fig. 1) shows no apparent geographical bias that is different to the distribution of all other postcodes. Whilst the links between geographies may be in error for the smallest geographical areas, it is anticipated that this should not unduly affect the data conversions between wards.

Figure 1: Distribution of match 2 (crosses) and match 3 (circles) postcode locations

The resulting file (pc-ed91-oa01-pchc-based.sav) contains 1,299,380 postcode records each of which has fields for the 1991 ED and 2001 OA associated with each postcode. 1991 and 2001 ward codes were derived from the ED and OA codes. This file can be aggregated into unique combinations of ward 1991 and ward 2001 codes along with the headcounts of total persons, males, females and households. This information can be used to provide weights in the GCT for overlaps represented by the population proportion of the 1991 source ward intersecting a 2001 Census target ward.

Creation of a 1991 to 2001 ward GCT with 1991 information as the starting point

The Postcode-Enumeration District Directory (PCED) for 1991 has, for each postcode record, the ED with which each postcode is associated, a count of households present at each postcode and the eastings and northings of the National Grid Reference (NGR) location. Linkages to the 2001 OA geography can be made using the GIS point in polygon technique where a point is the postcode's NGR and a polygon is the digital OA boundary.

To achieve links to the 2001 OAs a GIS point file of the PCED residential postcode locations was constructed. The eastings (x) and northings (y) were all adjusted to be consistent with the map units of the ONS OA boundary file and 50m was added to both the x and y coordinates, a procedure recommended by Gatrell *et al.* (1991) if linkages between geographies are to be established by GIS point in polygon operations.

After the postcode point in OA polygon links had been carried out, the resulting file (pc-ed91-oa01-pced-based.sav) has 1,744,476 records. The excess of records over the 'PCHC-based' file is because some postcodes crossed ED and ward boundaries in 1991 with the household counts apportioned accordingly. Each record has the 1991 ED and 2001 OA with which each postcode is associated along with the total number of households located at each postcode. After deriving 1991 and 2001 ward codes from the ED and OA codes the file could be aggregated to provide household counts in the intersections between the ward source and target geographies.

Development of a computer program to automatically convert files of data disseminated for the 1991 wards to the 2001 boundary definitions

A computer program has been written in the Fortran 90/95 language. This is an initial program development and allows various outputs to be compared. The final version in the light of results is supplied. The program structure is as follows:

Information required from the user

The user is offered the choice of the 1991- or the 2001-based GCTs. If the latter is selected there is a choice of indicator variable to used for the conversion weight: persons, males, females or households.

The number variables that are to be converted must be entered, the maximum number of variables that can be processed during each run of the program is 100. (This can be altered by the program developer if required.

Input files

The GCTs are read in from a subdirectory, as are ward code references and the source geography input data to be converted. The user needs to insert the data to be converted into the file Wards-1991.csv.

Disaggregating the input data

The counts of the indicator variable for each intersection are summed for every source ward. These 1991 ward totals are written out to a file, wd91-counts.csv.

The disaggregation weight is calculated as the count of the indicator variable in each source-target intersection divided by the total of that variable in the source ward. These disaggregation weights are written out to a file, Disagg-weights.csv.

Converting the input data

The source ward geography input data are disaggregated to the overlaps between source and target ward geographies by multiplying each variable by the disaggregation weight for each intersection. The estimates for each intersection are summed using the references to the target wards. These are written out to the file Converted-data.csv.

The data conversion process carried out in the program and described above can be formally defined as follows. Let:

D = count of a socio-demographic variable to be converted into a new geography,

x = the index for the year associated with the source geography,

i = the index for the wards as defined in the source geography,

j = the index for the wards as defined in the target geography,

p = the index for the postcodes as defined in the source geography,

H = the weight (the number of postcodes × indicator variable in each source-target intersection),

$p \in i$ = postcode p is a member of a set of postcodes associated with source ward i, and

$p \in j$ = postcode p is a member of a set of postcodes associated with target ward j.

The source year x ward variable $D^i(x)$ is disaggregated using the weight $H^p(x)$ divided by the total of the weights in each source ward i:

$$D^p(x) = D^i(x)\left(\frac{H^p(x)}{\sum_{p \in i} H^p(x)}\right)$$

The estimated postcode variable $D^p(x)$ is reaggregated into the target wards j:

$$D^j(x) = \sum_{p \in j} D^p(x)$$

Once the source data $D^i(x)$ have been adjusted, the variable $D^j(x)$ is an *estimate* of the original data made for the target geography.

Validation of the GCTs

The fundamental information required for the GCTs are reference codes to both the source geography in which the data pre-exist and the target geography into which the data are to be converted, together with a weight to indicate the extent of overlap between each zone in the source geography and each zone in the target geography. In this application the weights are derived from postcode counts enhanced by population

or household counts. The conversion weights taking a value of more than zero but less than or equal to one, will sum across intersections to one for each source area and will therefore be 'exhaustive' with no data lost during the conversion process (Simpson 2002: 73). As described above, two GCTs have been developed and these will be referred to as the 1991-based GCT and the 2001-based GCT since the postcode directories which formed their basis relates to these years.

The source-target overlap weights are an output of the Fortran program and the weights for the 2001-based and 1991-based GCTs can be compared. Using the 'persons' variable from which to calculate the disaggregation weights, the 2001-based GCT includes 38,335 (non-zero population) records and the 1991-based GCT has 17,453 records. Weights of one indicate that a 1991 source ward fits wholly within a 2001 target wards (this may well be because of no boundary change between censuses). The 2001-based GCT has 531 source-target weights of one whereas the 1991-based GCT has 4,112. Weights of less than one exist where the GCT apportions source data to different 2001 wards. The GCT based on the postcode-household counts in the 1991 PCED suggests fewer fully contained 1991 wards or no boundary changes than the postcode-headcount of persons.

Simpson (2002) defined the 'degree of hierarchy' to assess the data conversion process between pairs of geographies. The degree of hierarchy is the proportion of source units that are contained within a single target unit. If the degree of hierarchy is 100%, the source units fully nest within the target units (such as the relationship between electoral wards and LA districts) or, more relevant to this application, either that there is no boundary change from 1991 to 2001 or that all of the 1991 ward data should be allocated to a 2001 ward with possible contributions from other source wards. The degree of hierarchy is defined as:

$$100 \times \left(\frac{\sum_{ij} (w_{ij}, if\,(w_{ij} = 1))}{S} \right),$$

where

i = a source geography unit,

j = a target geography unit

w_{ij} = a weight, the proportion of the source geography that overlaps the target geography, and

S = the number of source geography units.

The 2001 GCT and 1991 GCT have degrees of hierarchy of around 6% and 43% respectively. Evidently, they are telling different stories about the relationship between 1991 and 2001 ward boundary definitions. A further measure defined by Simpson (2002), the 'degree of fit', can be used to indicate how differently the GCTs may be estimating the 1991 data by the 2001 geography. The degree of fit sums the largest weight from each source unit and expresses this as a proportion of the number of source units. The degree of fit shows how closely on average the number of postcodes × indicator variable in a source geography unit matches the number of postcodes × indicator variable in its largest overlap with a target unit. This approach recognises that source unit boundaries may be close to the target unit boundaries even when they do not fit them exactly and is an important measure of uncertainty. The degree of fit is defined as:

$$100 \times \left(\frac{\sum_i (\max w_{st})}{S} \right),$$

where

$\max w_{ij}$ = the maximum disaggregation weight for a source *i* to target *j* overlap.

Although the degrees of hierarchy are quite different, the degrees of fit are much closer. For the 1991-based GCT the degree of fit is 88.42% and for the 2001-based GCT it is 85.48%. Although the PCED-based GCT results in a better fit, this suggests that data converted using the GCTs will be similar.

There is a need to check whether boundary changes are being identified in locations where wards have changed and that boundary changes are not being indicated in locations where wards have not changed. The digital ward boundaries for 1991 and 2001 have been obtained and overlaid. GIS intersections of these boundary sets are a means by which to indicate boundary change but a vast number (millions) of sliver differences are found where digitising differences occur and where legal definitions alter (particularly around coastlines, estuaries and inland water courses) even though the boundary commission/ONS would indicate no official boundary change. To see whether a weight of 1 is appropriate, sample locations have been identified and examined.

Areas of no change

The Swansea area in Wales, for example, has conversion weights in the GCTs that disagree. Table 1 below has information from the GCTs derived from the 1991 and 2001 postcode-attribute counts. According to the 2001 GCT, only one out of the five wards listed is fully contained in a 2001 ward, whereas according to the 1991 GCT only data for one ward should be apportioned across the later boundaries. Inspection of the digital boundaries indicates that 1991 GCT would be correct (on an area

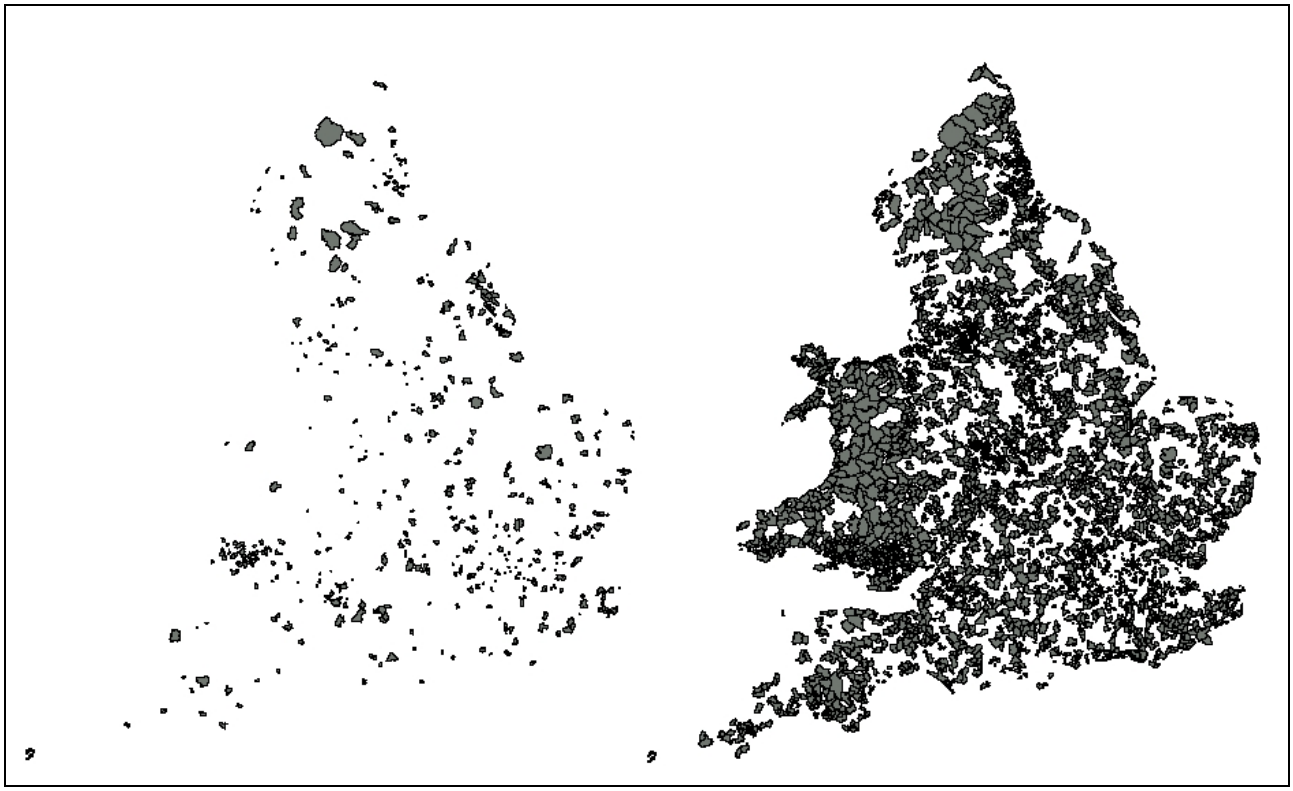
basis) in apportioning 1991 ward reference number 9505 roughly 7:3 to wards 3032 and 3039. All other wards have not changed. If the 2001-based GCT is used, whilst the majority of the data will be allocated to the correct ward (and similar degrees of fit), small amounts of data are likely to be incorrectly allocated to nearby wards. Potentially the estimated matches of postcodes to EDs described above (the Type 2 and 3 matches) may be in error, however, this has not been found to be the case since incorrect linking exists through the Type 1 matches. The situation shown for Swansea is typical and accounts for the much larger number of records in the 2001-based GCT and low degree of hierarchy/no boundary change indication.

Table 1: Extract from the 2001- and 1991-based GCTs

2001-based			1991-based		
Source	**Weight**	**Target**	**Source**	**Weight**	**Target**
9505	0.6768	3032	9505	0.7140	3032
9505	0.0006	3038	9505	0.2860	3039
9505	0.2681	3039	**(Sum**	**1.0000)**	
9505	0.0533	3043			
9505	0.0012	3059			
(Sum	**1.0000)**				
9514	1.0000	3051	9514	1.0000	3051
9515	0.0002	2896	9515	1.0000	3052
9515	0.0007	3046			
9515	0.0477	3047			
9515	0.0002	3050			
9515	0.9423	3052			
9515	0.0090	3053			
(Sum	**1.0000)**				
9517	0.0183	3032	9517	1.0000	3054
9517	0.9739	3054			
9517	0.0078	3067			
(Sum	**1.0000)**				
9527	0.0105	3034	9527	1.0000	3067
9527	0.0018	3051			
9527	0.0014	3054			
9527	0.0005	3055			
9527	0.9858	3067			
(Sum	**1.0000)**				

The maps in Figure 2 show the 1991 wards in England and Wales estimated by each GCT to have a weight of 1 to allocate data between 1991 and 2001. Evidently, the extent of complete allocation of 1991 data is much wider when the PCED-based information is used.

Figure 2: Wards with a disaggregation weight of 1 as indicated by the 2001 (left) and 1991 (right)-based GCTs



Areas of change

Are the GCTs identifying boundary changes where they do occur, or are the changes missed? The postcodes which constitute a ward in 1991 and 2001 as indicated by the 1991-based and 2001-based GCTs can be inspected to determine whether the extent of boundary changes are being identified sufficiently. In England 46 Unitary Authorities (UAs) were created during the inter-censal period and previous research (Norman *et al.,* 2003) identified that the sub-district ward structures in these locations can change appreciably. These locations have been focussed on.

Peterborough UA, for example, experienced considerable structural change to its constituent wards (Fig. 3). The sets of postcodes that constitute wards in the source and target years have been examined to see whether change and no change are being identified correctly. For the GCT based on 2001 information (Fig. 4), whilst the vast majority of postcodes are correctly associated in space with the wards in 1991 and 2001, a number of postcodes are associated with the 'wrong' ward in either or both of the source and target geographies. This will result in data being incorrectly allocated and accounts for the large number of records in the GCT with very small weights and the relatively small number of weights of 1. In contrast, the GCT based on PCED information for 1991 (Fig. 5) has sets of postcodes that more reliably can be taken to comprise a ward in each year. Other locations are consistent with the situation illustrated here for Peterborough.

Figure 3: Ward boundaries in Peterborough UA 1991 and 2001

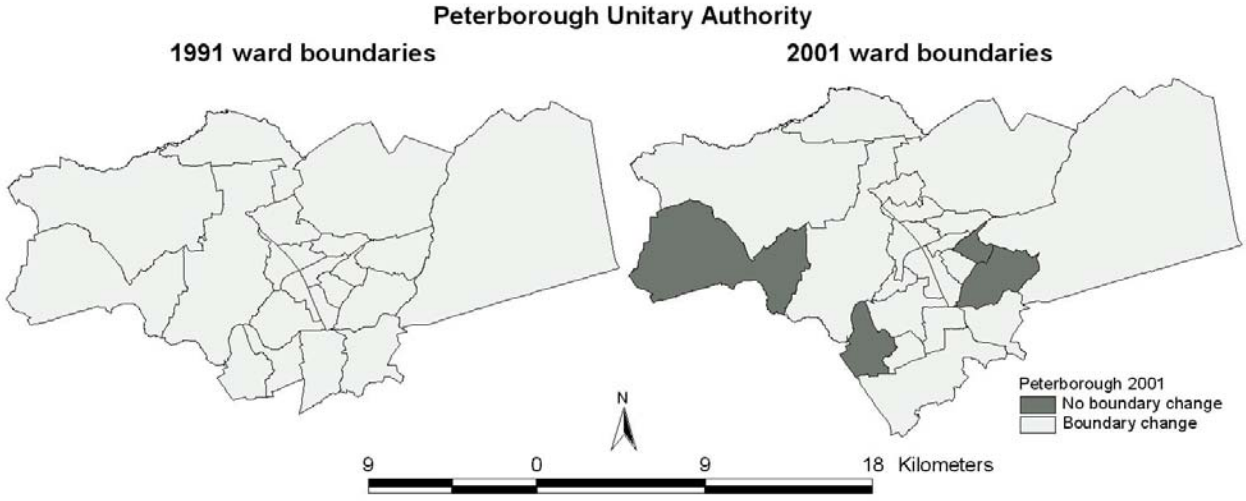


Figure 4: 2001-based GCT, postcodes comprising various 1991 wards (above) and 2001 wards (below) in Peterborough *(best viewed in colour)*

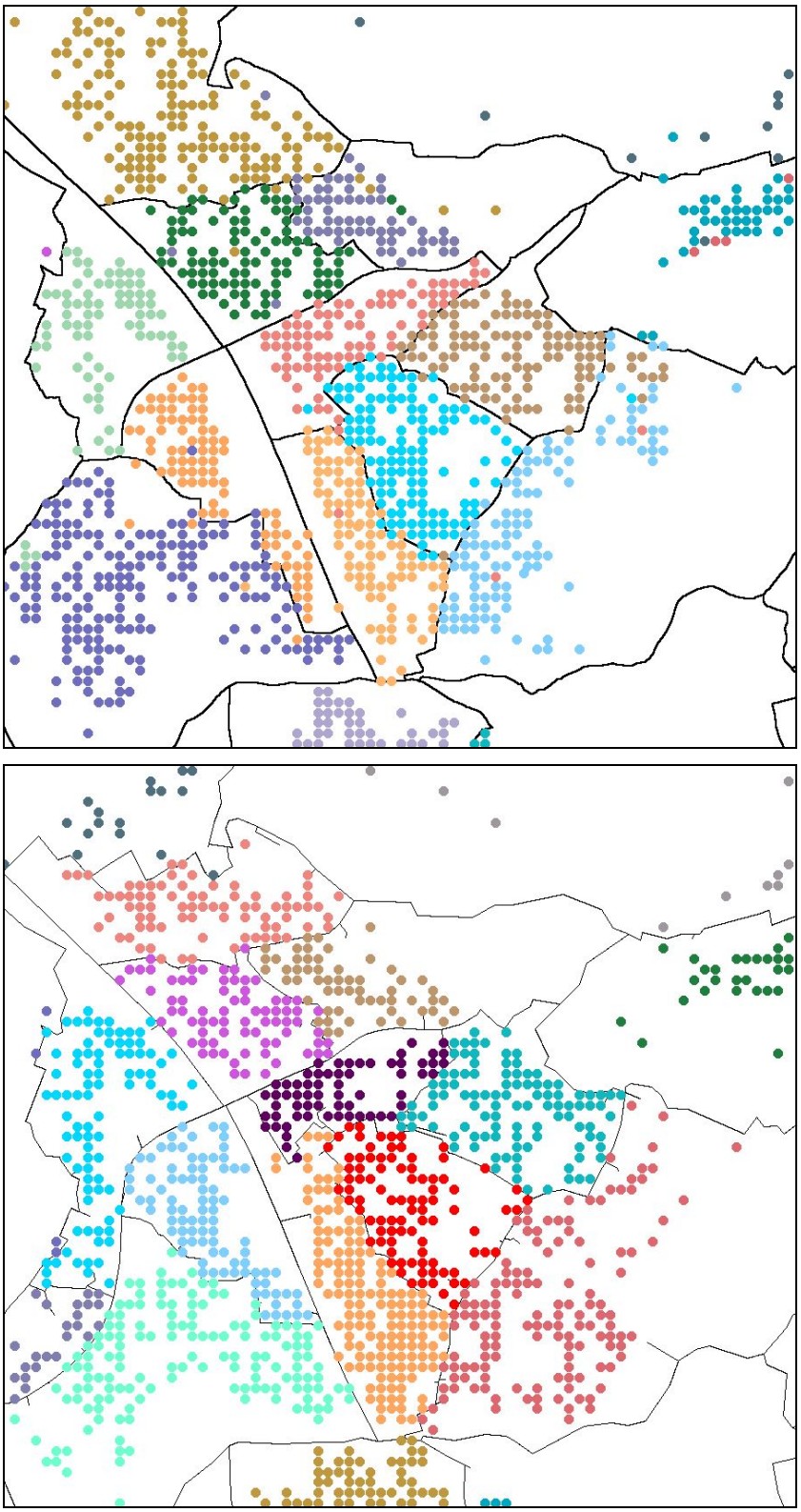
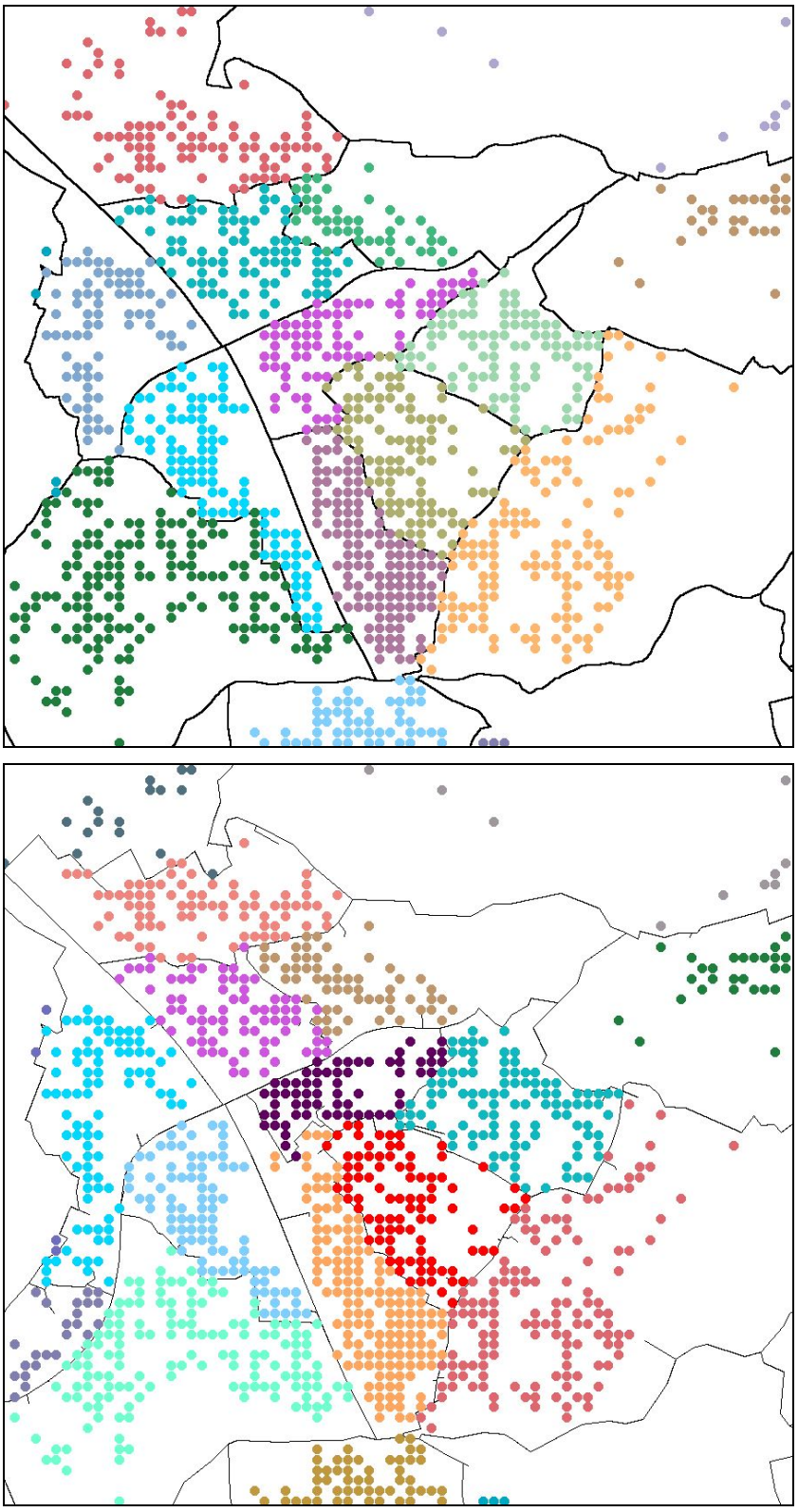


Figure 5: 1991-based GCT, postcodes comprising various 1991 wards (above) and 2001 wards (below) in Peterborough *(best viewed in colour)*



Are the converted outputs plausible?

Sample data from CASWEB for total population in each 1991 ward have been converted to the 2001 geography using the GCTs. A problem was identified with the 1991-based GCT in that there was a shortfall in the output data that amounted to 0.02% of the input data. This situation did not exist in the 2001-based GCT. To allow for this a routine was added to the Fortran program to constrain the sum of each variable of the converted data to be consistent with the sum of each initial input variable. Detailed investigation of the PCED-based GCT revealed two 1991 wards for which there were no source ward references so the population data for these wards were lost from the conversion process. The GCT was manually amended and the problem solved. The constraints routine has been retained to correct possible rounding errors in the conversion process.

Since it is a reasonable expectation that the distribution of population in 1991 will largely resemble the distribution 10 years later in 2001, total population converted to the 2001 wards can be compared with populations obtained from the 2001 Census (Fig. 6). The 1991-based GCT results in population estimated for the 2001 ward definitions which less closely represent the 2001 Census information than the GCTs based on 2001 postcode headcounts, but it is an arguable point that it necessarily should do. Using the 1991-based GCT some very small 1991 populations are estimated in a few 2001 locations relative to the 2001 populations. These were found to be locations where large scale postcode changes/introductions have taken place, suggesting considerable new build. For example, in Swindon (Fig. 7), not only did the local authority become a Unitary Authority with changes to the constituent wards, but there were changes to the postcode conventions in the area from, for example, SN2 3JG to SN254YL. Many new postcodes exist here. It is understandable that the conversions based on postcode-person counts in 2001 appear to perform better in these locations.

It may well be that where postcodes have been introduced or changed since 1991 there are linkage problems to the 1991 EDs. This is because in the post-1991 directories released by ONS the allocation of postcodes to EDs has invariably been automatically imputed from the closest postcode alphabetically. These postcodes may not necessarily be located next to each other in space, however.

Figure 6: Comparisons of 2001 Census populations with 1991 populations estimated for the 2001 ward geography by 1991-based GCT (above) and 2001-based GCT (below)

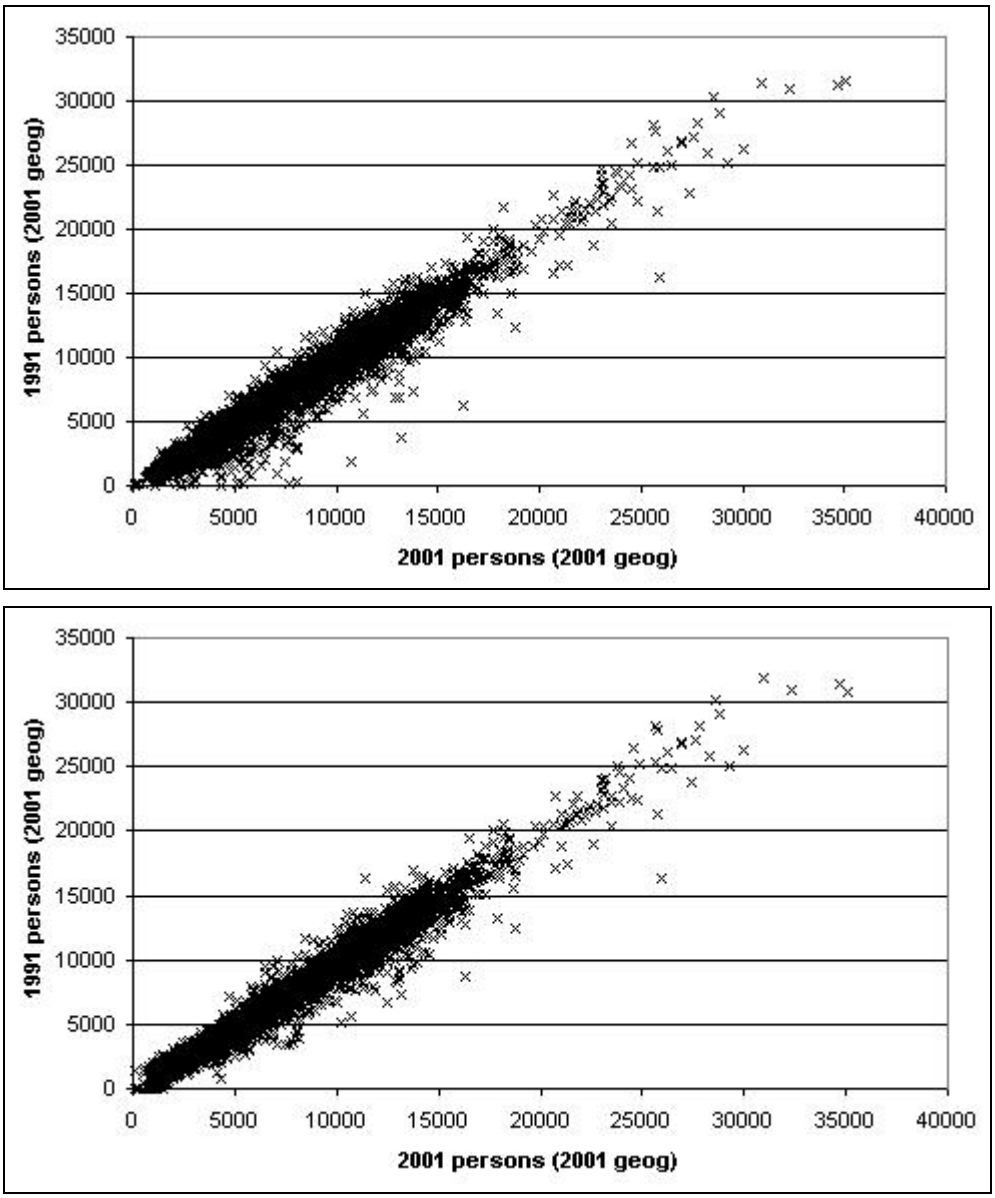
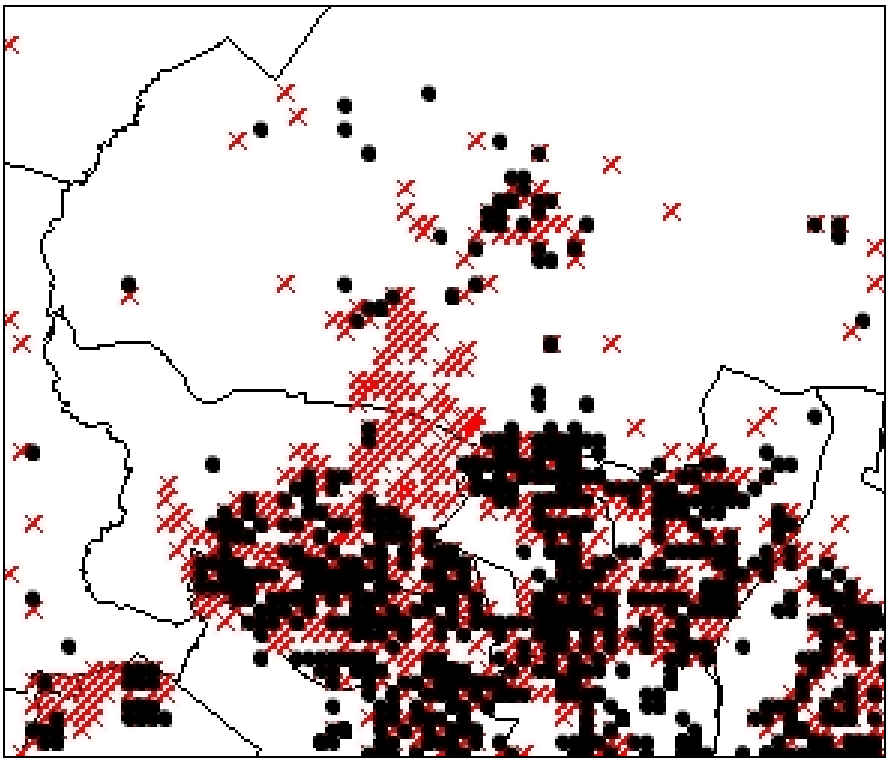


Figure 7: 1991 PCED-based postcode distribution (•) and 2001 NHS postcode directory distribution (x) in part of Swindon



Conversions for wards in the City of London and the Isles of Scilly present difficulties given relatively small population numbers, small numbers of households in the postcode directories and suppressed data giving nil populations in some wards and data transferred to nearby wards. Whilst input data with 1991 ward references are converted to the 2001 ward geography, it is recommended that data in these locations are combined to local authority level once converted between the ward geographies. Similarly, in other locations if the Census Local Based Statistics (LBS) has zero counts for suppressed wards, zero counts will be transferred in the conversion process. If 1991 Census variables are to be converted, it may be a better strategy either: i.) to obtain ward data from the Small Area Statistics (SAS) (which has less detail and lower confidentiality constraints) or ii.) to obtain ED level data from the SAS, aggregate these to the wards they comprise and then convert to the 2001 ward geography (in both cases after deleting the Shipping wards).

Due to nil counts of households and/or lack of postcode linkage 1991 to 2001, data for the wards listed in Table 2 will not be converted. If data in these wards need converting: the City of London wards can be combined with data for 01AAFQ, Cripplegate; data for Bryher ward in the Isles of Scilly can be combined with 16FDFB, Agnes.

Table 2: 1991 wards data for which are not converted

Local Authority	1991 Ward code	Ward name
City of London	01AAFF	Bread Street
	01AAFJ	Candlewick
	01AAFP	Cornhill
Isles of Scilly	16FDFA	Bryher

Various converted data are listed in Table 3 of LBS and SAS total ward populations using the 1991 and 2001-based GCTs. The converted 1991 SAS populations (S020001) have a greater correspondence with the 2001 distribution than the converted LBS populations (L020001). The small populations estimated for various 2001 wards are the result of postcodes being associated with the 'wrong' ward. It can be seen that this occurs more frequently when the 2001-based GCT is used, emphasizing the more 'fuzzy' nature of the conversions using this file.

Table 3: 1991 LBS and SAS total ward populations converted to 2001 ward geography

	1991-PCED-based GCT		2001-PCHC-based GCT		2001 Census
WD01CD1	I020001	S020001	I020001	S020001	Persons 2001
00AAFE	4	4	10	10	106
00AAFS	0	140	0	140	181
00AAFT	31	558	28	473	687
00AAFY	2	81	0	79	282
00AAFZ	0	112	0	75	141
00NCQA	0	910	210	959	1019
00NCQE	0	802	227	885	901
00NCQQ	0	871	90	890	860
00NCQU	0	947	274	997	1360
00NCRD	0	1299	181	1282	1277
00NCRX	0	934	45	934	834
00NNQH	0	838	46	832	936
00NNQJ	0	1364	63	1103	1441
00NNQM	0	1213	67	1118	1217
00NNQP	0	901	92	947	918
00NNRH	46	945	86	843	1057
00NNRQ	0	995	293	1132	1561
00NNRT	0	922	37	904	1055
00NNRZ	0	986	0	988	1005
00NNTE	0	917	157	1049	1167
00NNTK	0	926	152	961	1104
00NQPZ	0	914	279	1029	1115
15UHFB	0	83	0	161	165
15UHFC	0	111	0	111	142
15UHFE	0	165	0	165	180
16UFHH	0	889	3	883	928
20UHGB	0	669	0	663	685
20UHGC	0	692	0	672	754
20UHGJ	0	657	62	713	670
20UHGQ	0	718	0	691	713
29UBHT	14	14	435	436	1124
34UBFS	64	65	1551	1552	2345
35UBGD	0	876	58	895	875
35UFGN	0	980	5	972	936
35UFHG	0	891	0	889	908
35UFHN	0	872	3	891	974
38UBGU	27	27	1962	1963	4371
39UFGF	8	987	71	989	1118
39UFGM	0	876	84	721	914
39UFGS	61	930	248	1059	1086
40UFFZ	0	826	25	841	859
40UFGD	0	846	15	828	854
40UFGE	13	808	30	809	867

Summary

To enable the conversion of data disseminated for the 1991 ward geography to the 2001 boundary definitions, geographical conversion tables have been developed based on 1991 and 2001 postcode directories.

The 2001 Postcode Headcount file potentially offers the advantage of disaggregation weights being calculated using counts of persons, males, females or households in each source-target intersection. However, the linkages of postcodes to ward geographies has been shown to distribute data outside the correct target ward. No details have been given, but ONS have recently withdrawn the supply of the Postcode-Headcount file because of unreliability. In the future, as and when a reliable file is released, along with the All Fields Postcode Directory for 2001 that includes ED and OA references, the 2001-based GCT can be redeveloped. The intention then would be to enable conversions using postcode headcounts of persons, males or females to convert ward totals of demographic variables and to use postcode counts of household to convert each ward's household data (e.g. tenure, overcrowding, etc.).

The GCT based on the 1991 Postcode-Enumeration District Directory, whilst only offering counts of households at each postcode through which to derive the disaggregation weights, results in more reliable conversion of data between the 1991 and 2001 Census ward geographies. This is the method that is to be recommended. In the light of these conclusions the Fortran program has been revised to include just the GCT based on 1991 disaggregation information the code for this program and user instructions are in the appendix. Apart from user choices of GCT, the program structure and algebra given above remain the same.

Paul Norman and Ludi Simpson, November 2003

References

Gatrell A C, Dunn C E and Boyle P J (1991) The relative utility of the Central Postcode Directory and Pinpoint Address Code in applications of geographical information systems, *Environment and Planning A*, 23: 1447-1458

Norman P, Rees P and Boyle P (2003) Achieving data compatibility over space and time: creating consistent geographical zones, *International Journal of Population Geography*. Vol 9, Issue 5, September-October 2003: 365-386

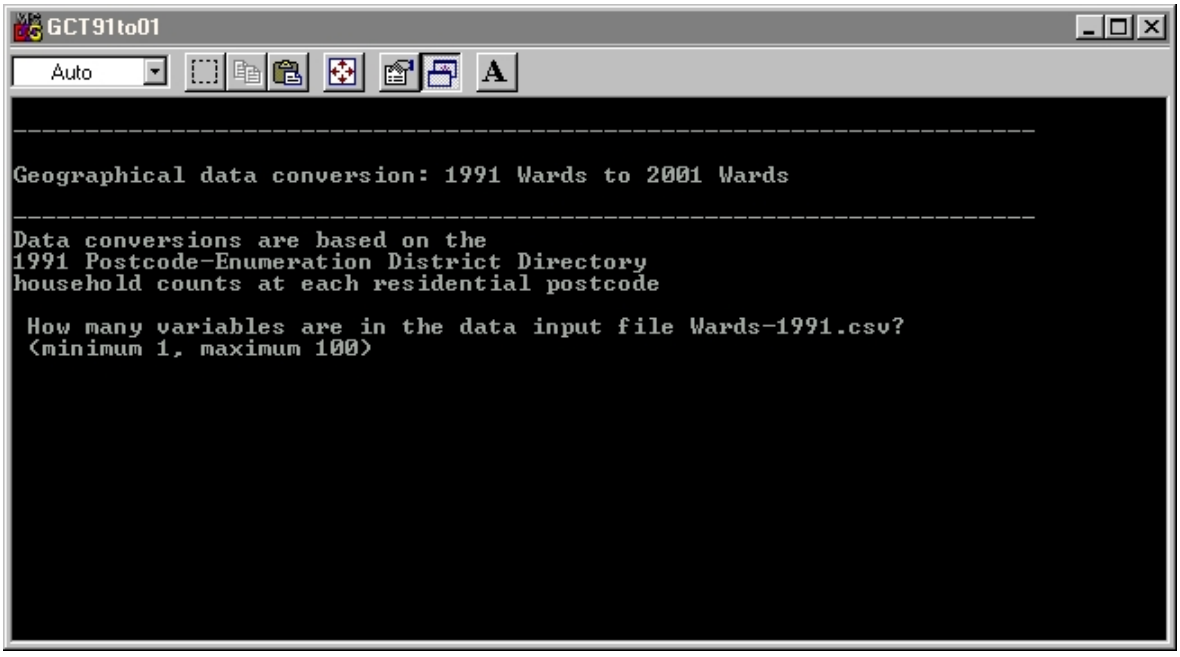
Simpson L (2002) Geography conversion tables: a framework for conversion of data between geographical units, *International Journal of Population Geography* 8: 69-82

Appendix 1: Converting 1991 ward data to 2001 ward geography: GCT91to01.exe program user instructions

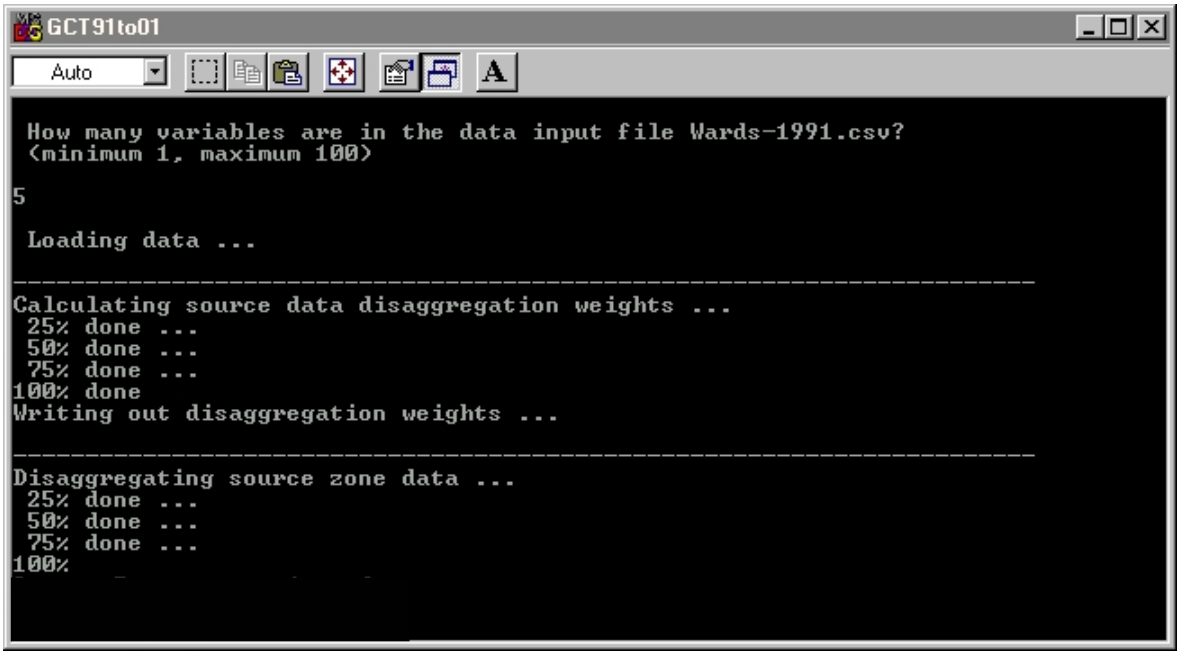
GCT91to01.exe will convert data for 1991 wards to the 2001 ward geography. All file paths and file names are hard coded, so copy all files and subdirectories to the computer's C drive.

To convert data, navigate to the subdirectory C:\GCT91to01\Input-data and open the file Wards-1991.csv in Excel. The first three columns in this file have ward references that must be retained. The user's data, to a maximum of 100 variables, must be pasted into column D onwards and the file saved replacing the existing file. The wards listed are in the same order as a download of ward data from CASWEB but with the Shipping wards removed. The user's data must be in the same order.

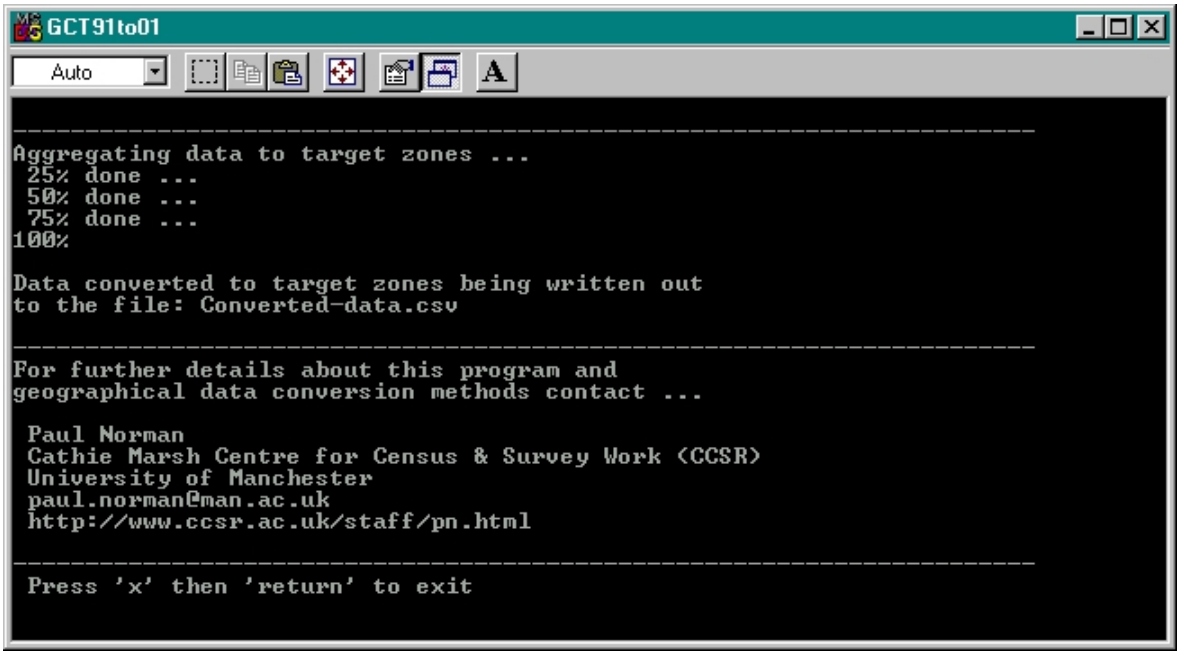
Double click on GCT91to01.exe to start the program. A DOS box opens (see below). The user is asked 'How many variables are in the data input file Wards-1991.csv?'. Type the number and hit return. Assuming the number of variables is between 1 and 100, the program runs.



The input data and geographical conversion table are then loaded, the program calculates the data disaggregation weights and uses these to disaggregate the input data to the intersection overlaps between the source and target geographies. Program progress is reported (see below).



The disaggregated data are then summed into the target 2001 wards with the results written out to the file Converted-data.csv to be found in the subdirectory C:\GCT91to01\Output-data. Also in this folder are files which give counts of households in each source ward and, most usefully, the source to target disaggregation weights that have been calculated.



N.B. To re-run the program, new data must be saved to the input file Wards-1991.csv. Any previous data in the output file Converted-data.csv will be over-written. Neither of these files must be open when the conversion program runs.

If the hard coding of file paths to the C drive presents a user with a problem, get in touch, it is no problem to change. Similarly, if the upper limit of 100 variables is restrictive, this can be increased but for practical reasons is worth keeping within the limits of Excel (i.e. no more than 250 columns of data to be converted).

Appendix 2: Fortran program code

```
program GCT91to01
      implicit none

!Source geography is 1991 Census wards, Target geography is 2001 Census wards

integer,parameter::intersections=17453 !Number of source-target intersections
integer,parameter::source_zones=9527 !Number of wards in 1991 source geography
integer,parameter::target_zones=8850 !Number of wards in 2001 target geography
integer::variables !Number of variables/columns to be converted
integer::a !Number of variables/columns to be converted
character(1)::control

!GCT intersection file
integer*4::i
integer*4::source_zone_ref(17453)
integer*4::target_zone_ref(17453)
real*8::hhs91(17453) !Count of attribute at each intersection

!Source zone variables (1991 wards)
integer*4::j
integer*4::source_num(9527) !Zone ref number
real*8::source_hhs91(9527) !Total households in each source zone
real*8::source_data(9527,100) !Data to be adjusted by source zone and number of variables
character(6)::wd91cd1(9527)
character(4)::wd91cd2(9527)
character(8)::variable_codes(100)

!Target geography variables (2001 wards)
integer*4::k
integer*4::target_num(8850) !Zone ref number
real*8::target_data(8850,100) !Final adjusted data by target zone and variables
character(7)::title1 !2001 ward refs/codes
character(7)::title2 !2001 ward refs/codes
character(7)::title3 !2001 ward refs/codes
character(6)::wd01cd1(8850) !2001 ward refs/codes
character(4)::wd01cd2(8850) !2001 ward refs/codes

!1991/2001 intersection disaggregation
real*8::disagg_weight(41308) !Household adjusted weight
real*8::intersection_data(41308,100) !Data allocated to each intersection

real*8::adj(100) !Scale factor to adjust results
real*8::s_tot(100) !Source total to adjust results
real*8::t_tot(100) !Target total to adjust results

!Format information for file I/O
character(1)::comma="," !Used for writing out files in csv format
101 format (i4,4(a1,f10.0))
104 format (a7,a1,a7,a1,a7,100(a1,a8))
105 format (i4,a1,a6,a1,a4,100(a1,f25.15))
```

```
print *, " "
print *, “--------------------------------------------------------------------------------"
print *, " "
print *, "Geographical data conversion: 1991 Wards to 2001 Wards"
print *, " "
print *, “--------------------------------------------------------------------------------"

!Intitialise variables (not strictly necessary in Fortran 90/95)
source_zone_ref = 0
target_zone_ref = 0
hhs91 = 1
source_num = 0
source_hhs91 = 0
source_data = 0
target_num = 0
target_data = 0
disagg_weight = 0
intersection_data = 0
variable_codes = 0
variables = 0
adj = 0
s_tot = 0
t_tot = 0

print *, "Data conversions are based on the"
print *, "1991 Postcode-Enumeration District Directory"
print *, "household counts at each residential postcode"

do
      print *," "
      print *," How many variables are in the data input file Wards-1991.csv?"
      print *," (minimum 1, maximum 100)"
      print *," "
      read *, variables
      if ((variables .GT. 100) .OR. (variables .LT. 1)) then
            print *, " The maximum number of variables is 100"
            print *, " The minimum number is 1"
            print *, " Please re-enter ..."
            print *," "
      end if
      if ((variables .GT. 0) .AND. (variables .LT. 101)) then
            print *," "
            print *, " Loading data ... "
            exit
      else
            cycle
      end if
end do

!Read in intersection GCT based on 1991 information
open (unit = 50, file = "C:\GCT91to01\LUTs\wd91-wd01-pced-based.csv", status = 'old')
read (unit = 50, fmt= *) !Ignores file header
do i=1,intersections
      read (unit=50, fmt=*) source_zone_ref(i),target_zone_ref(i),hhs91(i)
end do
close (unit = 50)
```

```
!Read in source geography data
open (unit = 60, file = "C:\GCT91to01\Input-data\wards-1991.csv", status = 'old')
read (unit = 60, fmt= *) title1, title2, title3, (variable_codes(a),a=1,variables)
do j=1,source_zones
      read (unit=60, fmt=*) source_num(j),wd91cd1(j),wd91cd2(j), (source_data (j,a),a=1,variables)
end do
close (unit=60)

!Read in references to target geography codes
open (unit = 70, file = "C:\GCT91to01\LUTs\2001-wards.csv", status = 'old')
read (unit = 70, fmt =*) title1, title2, title3
do k=1,target_zones
      read (unit=70, fmt=*) target_num(k),wd01cd1(k),wd01cd2(k)
end do
close (unit = 70)

!Count number of persons/males/females/households in each source zone
!This sums the headcounts by intersection to the source ward geography
do i=1,intersections
      source_hhs91(source_zone_ref(i)) = source_hhs91(source_zone_ref(i)) + hhs91(i)
end do

!Write out counts of persons/males/females/households in each source zone
open (unit = 80, file = "C:\GCT91to01\Output-data\wd91-counts.csv", status = 'unknown')
      write (unit=80, fmt=*) "Wardnum,Households" !Titles for columns in output
do j=1,source_zones
      write (unit=80, fmt=*) source_num(j),comma,source_hhs91(j)
end do
close (unit=80)

! Calculating disaggregation weights by dividing households counts in each
! source-target intersection by total households in each source zone
print *, " "
 print *, “--------------------------------------------------------------------------------"
print *, "Calculating source data disaggregation weights ..."
do i=1,intersections
      do j=1,source_zones
            if (source_zone_ref(i) .EQ. source_num(j)) then
                  if (hhs91(i) .GT. 0)then
                        disagg_weight(i) = hhs91(i)/source_hhs91(j)
                  end if
            end if
      end do
      if (i .EQ. 4000) then
            print *, " 25% done ..."
      end if
      if (i .EQ. 8000) then
            print *, " 50% done ..."
      end if
      if (i .EQ. 12000) then
            print *, " 75% done ..."
      end if
end do
print *, "100% done "
```

```
!Write out disaggregation weights
print *, "Writing out disaggregation weights ..."
open (unit=90, file="C:\GCT91to01\Output-data\Disagg-weights.csv", status='unknown')
do i=1,intersections
      write (unit=90, fmt=*) source_zone_ref(i),comma,disagg_weight(i),comma,target_zone_ref(i)
end do
close (unit=90)

!Disaggregating source zone data
print *, " "
 print *, “--------------------------------------------------------------------------------"
print *, "Disaggregating source zone data ..."
do i=1,intersections
      do j=1,source_zones
            if(source_zone_ref(i) .EQ. source_num(j))then
                  do a=1,variables
                        intersection_data(i,a) = source_data(j,a) * disagg_weight(i)
                  end do
            end if
      end do
      if (i .EQ. 4000) then
            print *, " 25% done ..."
      end if
      if (i .EQ. 8000) then
            print *, " 50% done ..."
      end if
      if (i .EQ. 12000) then
            print *, " 75% done ..."
      end if
end do
print *, "100% "

! Aggregate data to target zones
print *, " "
 print *, “--------------------------------------------------------------------------------"
print *, "Aggregating data to target zones ..."
do i=1,intersections
      do a=1,variables
            target_data(target_zone_ref(i),a) = target_data(target_zone_ref(i),a) + intersection_data(i,a)
      end do
      if (i .EQ. 4000) then
            print *, " 25% done ..."
      end if
      if (i .EQ. 8000) then
            print *, " 50% done ..."
      end if
      if (i .EQ. 12000) then
            print *, " 75% done ..."
      end if
end do
print *, "100% "
```

```
!Constraining converted data to original total for each variable
do a=1,variables
	do j=1,source_zones
		s_tot(a) = s_tot(a) + source_data(j,a)
	end do
	do k=1,target_zones
		t_tot(a) = t_tot(a) + target_data(k,a)
	end do
	adj(a) = s_tot(a)/t_tot(a)
end do

do k=1,target_zones
	do a=1,variables
		target_data(k,a) = target_data(k,a) * adj(a)
	end do
end do

!Write out data aggregated to target geography
print *, " "
print *, "Data converted to target zones being written out "
print *, "to the file: Converted-data.csv"
print *, " "
open (unit = 95, file = "C:\GCT91to01\Output-data\Converted-data.csv", status = 'unknown')
write (unit=95, fmt=104) title1,comma,title2,comma,title3,(comma,variable_codes(a),a=1,variables)
do k=1,target_zones
	write (unit=95, fmt=105)
target_num(k),comma,wd01cd1(k),comma,wd01cd2(k),(comma,target_data(k,a),a=1,variables)
end do
close (unit=95)

print *, "------------------------------------------------------------------------"
print *, "For further details about this program and"
print *, "geographical data conversion methods contact ..."
print *, " "
print *, " Paul Norman"
print *, " Cathie Marsh Centre for Census & Survey Work (CCSR)"
print *, " University of Manchester"
print *, " paul.norman@man.ac.uk"
print *, " http://www.ccsr.ac.uk/staff/pn.html"
print *, " "

do
	 print *, "------------------------------------------------------------------------"
	print *, " Press 'x' then 'return' to exit"
	print *, " "
	read *, control
	if (control == 'x') then !Allows user to quit
		exit
	else
		cycle
	end if
end do

end program GCT91to01
```