

# SurfaceBuilder247Py: User Guide

V2 March 2025

## Purpose

[\*SurfaceBuilder247Py\*](#) is software which produces gridded population distribution models for specific times and dates. *SurfaceBuilder247Py* implements the Population24/7 spatiotemporal modelling framework first introduced in Martin et al (2015). It processes a series of input centroid locations, such as census Output Area (OA) centroids, schools, hospitals, etc. with associated population counts and time profiles, and models these onto a regular geographical grid for a specified target time and date. A background layer allows weighting of locations according to their probability of containing population and may be used to represent the transportation network and areas to be excluded from the modelling such as the sea. The program proceeds by redistributing the total population represented by a series of “origin” centroids onto the grid, taking account of relevant “destination” centroid locations, time profiles, catchment areas and the background layer, so as to best reflect the population demanded by each location for the specific time being modelled. Unlike many conventional population mapping tools, which assume all population to be present only at residential locations (which is, in effect, just a simplified “night time” population representation), *SurfaceBuilder247Py* allows the modelling of any desired time of day for any date, provided that suitable input data are available.

Further information about the Population24/7 modelling framework, downloadable datasets, and relevant papers, can be obtained from the [Population24/7](#) website. As described below, use of *SurfaceBuilder247Py* requires preparation of an extensive input data library. A selection of exemplar input datasets and sample modelled outputs for England and Wales for 2011 are available from the UK Data Service at <https://reshare.ukdataservice.ac.uk/853950/> or via the [Population 24/7 Download](#) web service. These can be used in both the VB .NET and Python versions of *SurfaceBuilder247*.

## Version history

*SurfaceBuilder247Py* was developed by University of Southampton GeoData Institute, on behalf of the Office for National Statistics (ONS). It is an open source, freely available, Python (3.x and above) redevelopment of the previous Visual Basic .NET *SurfaceBuilder247* code developed by University of Southampton, which was itself a very substantial development of the algorithm used in the original *SurfaceBuilder* software, written by David Martin. *SurfaceBuilder247Py* (and *SurfaceBuilder247*) produce time-specific estimates of population distribution, whereas *SurfaceBuilder* only produced spatial estimates. The VB .NET version of *SurfaceBuilder247* was a principal output from the ESRC-funded “Population 24/7: space-time specific population surface modelling” project (Award RES-062-23-1811), further developed under the ESRC “Near real-time spatiotemporal population estimates for health, emergency response and national security” project (Award ES/P010768/1). Development of the current *SurfaceBuilder247Py* version was funded by ONS.

## User guide

### Download, license and installation

*SurfaceBuilder247Py* has been written in Python and can be downloaded from <https://github.com/geo-data/SurfaceBuilder247py> or imported as a package from the official Python Package Index (PyPI) (<https://pypi.org/project/SurfaceBuilder247/>). It is provided under an [MIT license](#).

*SurfaceBuilder247Py* can be run on any platform and in any Python (v3 and above) environment. Once the source code is downloaded or the Package installed and suitable data downloaded, a *main.py* script (an example is included in the github repository) is run to carry out all of the steps needed to load data, run the model and save outputs. The only source code dependency which may require installation on a new Python setup is NumPy, a widely available Python library for numerical computing. The software has been tested by the project team on PCs and servers running Windows, Linux and within Docker. There are no minimum resource requirements specified, however, as the application is computationally intensive, processor speed and available memory have a significant impact on modelling performance for large datasets. Large runs may take several hours and users are strongly advised to test the entire data structure and modelling process using a very small area first: run length is broadly proportional to the number and density of centroids in the study area.

### Folder structure

A folder structure should be set up as follows in order that the various files required by the program can be correctly located. The entire system should be contained in a folder in a suitable working location e.g. C:/Pop247. The program is best run from this location. Inside this folder, four subfolders are recommended:

- *Docs* can be used to hold relevant documentation such as this User Guide, relevant papers, metadata etc.
- *src* contains the program files noted above, which can be obtained as described above.
- *Data* This folder is essential as it contains the data library which the program uses to produce the population estimates.

### Contents of *src/* folder

Once downloaded, the ***src/*** folder will contain the files shown in Table 1.

**Table 1 Contents of *src/* folder for *SurfaceBuilder247Py***

<b><i>src/</i> -</b>	
<i>sb247_modelRun.txt</i>	Pseudocode descriptions of the algorithms implemented
<i>sb247_python.md</i>	Documentation of the Python structure and code
<i>main.py</i>	A sample main Python entry program
<i>sb247.py</i>	The main SB247 Class, the container for all functionality and data
<i>projectParams.py</i>	Stores all Project parameters
<i>modelRun.py</i>	Run the model including the main algorithms
<i>locationIndex.py</i>	Structures and methods to rapidly identify data within a bounding area
<i>gridCreate.py</i>	Create gridded data from model run outputs
<i>unit_tests.py</i>	Perform a series of validations (requires sample Data)

**sb247\_python.md** and **sb247\_modelRun.txt** provide an overview of the Python code structure and detailed pseudocode descriptions of the algorithms that implement the tool. Users wishing to understand the underlying workings of the model should start with these. Alternatively a user can get started straight away by editing and running the sample **main.py** provided. In order to verify the tool is working and the sample data has been installed correctly, users can run the **unit\_tests.py** initially. Detailed instructions are included in the *Running the program* section below.

### *Contents of Data folder*

The **Data** folder requires further subfolders which contain the different file types either required or generated by the program as follows:

- *BckGrnds* contains pre-prepared background map layers
- *DataLogs* contains data log files generated by *SurfaceBuilder247Py*
- *Dests* contains pre-prepared destination centroid data files
- *Origins* contains pre-prepared origin centroid data files
- *Results* contains results files generated by *SurfaceBuilder247Py*
- *RunLogs* contains run logs generated by *SurfaceBuilder247Py*
- *SessionParas* contains files recording session parameters for *SurfaceBuilder247Py* runs. These may be specified directly in *main.py* or loaded in from this folder.
- *TimeSeries* contains the pre-prepared timeseries library files

### *Preparing a data library*

Preparing for a *SurfaceBuilder247Py* run requires the user to assemble an extensive library of data files in the *Bckgrnds*, *Dests*, *Origins* and *TimeSeries* folders. This is the most time-consuming part of the work and requires extensive planning and a full understanding of the spatiotemporal modelling concepts explained in Martin et al (2015) and below.

### *Background layers*

The background map layers are ascii (.txt) files set out in the asciigrid format used by ESRI's ArcGIS software, at the same cell resolution as the desired output grids and covering the entire area to be modelled. These grids contain weighting values to indicate the relative likelihood of population in the transportation system being located in each cell and nodata values to indicate cells in which no population should be located. Such a layer may typically be created in a GIS by rasterising a transportation network with associated capacity values and overlaying a layer representing areas of sea, open water or other areas from which modelled population must be excluded. The file contains a header in the form:

```
ncols      3500
nrows      3500
xllcorner  0
yllcorner  0
cellsize   200
NODATA_value -9999
(Followed by individual cell values in row primary order...)
```

A background file should be chosen which is appropriate to the scenario to be modelled, thus a weighted transportation model suitable for a morning rush hour would look very different to that for a Sunday evening and it is up to the user to select suitable background assumptions.

### *Origins and destinations*

The main inputs to SurfaceBuilder247 modelling are the files representing population **origins** and **destinations**. These essentially represent all locations which may be considered “containers” of population at some time within the reference frame of the modelling. Examples of **origins** are census OA centroids, whose total population counts sum to equal the total population of the model. Examples of **destinations** are any locations which have no resident population but at some time may contain population such as workplaces, educational establishments, health care facilities, etc. The levels of detail of origin and destination centroids may be expanded as far as available data and modelling requirements allow but it is important that within a single data library the entire population is only accounted for once within the origin centroid set and that destinations similarly avoid the possibility of double counting. For example, origin populations may be referenced to census OAs or to postcode locations, but only one of these referencing systems should be used within the same geographic extent, such that the entire population is accounted for once and once only. Similarly, a destination file may contain workplaces referenced by OA centroids, or may contain every known workplace separately identified by exact grid coordinates, but it is essential that within a single data library each workforce is only represented once, hence there should be no possibility that the employees of a given workplace are also included within the workforce total for some higher level geographical unit.

The definition of centroid files requires the specification of several basic parameters relating to the data. Each centroid has a name (such as “St. George’s Hospital”) and unique identifier (such as a census OA code or postcode), x and y grid coordinates and a total population count. The total population count is then subdivided into a number of **sub-groups** (e.g. aged 0-4, aged, 5-9, higher education students, etc.) which must sum to the exact total and should have a standard definition within a single data library. The sub-groups are used to differentiate the behaviour of different populations and will be reflected in the available destination data, for example age groups which match divisions in known educational and employment activities. Further parameters are declared for the overall centroid data file as discussed in the following paragraph, but these are default values and may be overridden by exact values provided for any individual centroid.

The name of a **time profile** (or **timeseries**) containing profiles of population activity must be declared in a destination centroid file. For example, time profiles of working hours would be associated with workplace centroids. Within the centroid file, subsets of workplaces may be associated with more specific time profiles. A **local area dispersion (LAD) function** is defined for each centroid, which controls the distance decay algorithm to be used to spread population around the centroid. At present, only the Cressman function has been implemented, although others could be implemented in the future. A local dispersion parameter defines the spatial extent of a centroid – for example a small site like a primary school might be set to 100m, whereas a large site such as a university campus may be set to 1000m. As with the other parameters, the header sets default values which may be overridden if exact information is available in relation to individual centroid records. A **wide area dispersion (WAD) function** describes the structure of distance decay in the travel of population to a given destination centroid and is expressed as a series of distance bands

with associated percentages of the centroid's population. This can be conceptualised as the catchment area of the destination e.g. a school catchment area or the distance travelled to work by workers in an industry sector. Additionally, **major flows** may be explicitly recorded if specific population movements are known to exist between geographical units in the input data, for example a commuting flow from a particular ward to a specific workplace. This structure allows such known population movements to be incorporated into the model before distance decay functions are used to estimate the remaining interactions.

Providing these basic structural requirements are met, the user has considerable flexibility in describing any locations which may at some time contain population. It is advisable to keep all centroids of different types (workplaces, educational establishments, etc.) in separate files to aid data management. Origin and destination centroid files are constructed in comma separated (.csv) format and should conform to the layout described in **Annex A** for the multi-line header structure which defines various dataset parameters and the field layout of the individual records which follow. In the data block, each centroid is represented by a single line which conforms to the structure defined in the header. For destination centroids, each will need to be associated with a timeseries which indicates the pattern of population occupation of that location relative to its total capacity. More parameters are required for destination than origin centroids. For origin files only, the proportion of each age group which is **mobile** (i.e. available for spatial reallocation) is defined, allowing "immobile" populations such as prisoners or elderly persons in permanent residential care to be retained at their origin locations.

### *Time profiles*

**Timeseries** is the final essential input data type and takes the form of a time profile library file, in .xlsx format, which is placed in the timeseries folder with the name **timeseries.xls**. The timeseries file contains all the time profiles which are identified in the corresponding destination centroid files. The following example shows one simple time profile.

Prim04-11		
InTravel		
00:00:00		0
08:30:00		100
09:00:00		10
15:00:00		90
16:00:00		5
17:00:00		5
17:30:00		0
OnSite		
00:00:00		0
09:00:00		90
10:00:00		100
15:00:00		10
16:00:00		5
17:00:00		0

A time profile comprises two parts, an **InTravel** and an **OnSite** component, and is always associated with a destination centroid, such that the **percentages are expressed relative to the declared population capacity** of that destination. The first block represents the times at which given percentages of the destination population are expected to be in the transportation system and the second the times at which they are expected to be present at the destination. The example time profile above is called Prim04-11 and describes a day during which 0% of the population in question travel between midnight and 08:30 and 100% travel between 08:30 and 09:00, etc. Any number of time divisions may be used within the day. The second block of time profile information contains percentages of the total population capacity actually present at the destination at different times of day. Further time profiles with variant names may be used to describe the behaviour of the same destination on different days of the week, which might typically be included in different files. Multiple time profiles may be included within a single timeseries file, by adding further pairs of columns. It is not necessary for all the time profiles within a single file to be of the same length.

### *Analysis area and study area*

The **analysis area** defines the geographical area for which results will be provided. The analysis area must not exceed the dimensions of the background file. A buffer distance is also selected. This is a distance surrounding the analysis area and modelling covers this entire extended area, known as the **study area**. Input data should cover the entire study area. The study area must be set large enough so that the results in the analysis area are not unduly influenced by population movements from outside the study area, to avoid edge effects. For example, if there is a major settlement 20km beyond the edge of the analysis area which interacts with the analysis area through overlapping travel to work areas, service catchment areas, etc. the buffer should be set greater than 20km. When using UK census data, a buffer of at least 50km is recommended.

### *Running the program*

*SurfaceBuilder247Py* should only be run when a complete data library has been assembled as described above.

As outlined in the ‘Download, license and installation’ section above, the user will have first downloaded (or cloned) the source code from the github repository, or installed the python package from PyPI.

Examples of each method are:

```
https://github.com/geo-data/SurfaceBuilder247py/archive/refs/tags/v1.1.zip
```

```
git clone https://github.com/geo-data/SurfaceBuilder247py
```

```
pip install SurfaceBuilder247
```

Assuming the data library is now available within a ‘Data/’ subdirectory of the src folder, then the Unit tests may be run to verify correct installation:

```
python unit_tests.py
```

The output from this should indicate whether the tests have been passed successfully.

The user will then copy or edit the main.py script using a suitable text editor in order to set up the parameters for their own run of the program, executed, when ready, as:

```
python main.py
```

The sample main.py provided with the source code includes ready to run examples of a model run for a single set of parameters and age group (function main() ), or a sample run looping through all age groups within the selected origin data file (function main\_allages() ). The user will select the preferred first run by editing the end of the script which selects the function to be run by default.

```
246     # Executed when the program runs
247
248     if __name__ == '__main__':
249         # main() # Single age band
250
251         main_allages() # Loop through multiple age bands
```

Within the selected function, the script includes, with clear comments, each step of the process of identifying the location of data, specifying run parameters from a Python dictionary (or session file in the previous tool format), loading and checking the data.

```
32     # instantiate an object of the SB247 class
33     # project directory parameter will prefix all file location references
34     sb = SB247('./Data')
35
36     # instance variables of the class are accessible as required
37     print('Project directory: ' + sb.projDir)
38
39     # Project parameters can be loaded from a Dictionary or (VB compatible) file
40
41     # Load Project parameters from a dictionary
42
43     proj_dict = {
44         'analysisarray': [373000, 160000, 40, 40, 200], # BL_E, BL_N, nrows, ncols, cellsize
45         'buffer': 8000,
46         'background': 'rastp6_monfri_10_16_2011.txt',
47         'timeseries': 'TimeSeries.xls',
48         'origin': 'Origin_Eng_OTT_2011.csv',
49         'destarray': ['Dest_Eng_Accom_OTT_2011.csv',
50                     'Dest_Eng_Agri+Fish_OTT_2011.csv',
51                     'Dest_Eng_Education_UniPGR_OTT_2011.csv',
52                     'Dest_Eng_Healthcare_OTT_2011.csv',
53                     'Dest_Eng_Mine+Transp_OTT_2011.csv',
54                     'Dest_Eng_Public+Office_OTT_2011.csv',
55                     'Dest_Eng_Retail+Arts_OTT_2011.csv',
56                     'Dest_Eng_Service_OTT_2011.csv']
57     }
58
59     sb.loadProjectParamsFromDict(proj_dict)
60     --
```

Once the required data is loaded, the selected script function includes code to run the model (for a single specified age group, or looping through all ages), creating and saving gridded (or csv) data after each run.

The model run function can be supplied with parameters to speed up runs. This may be useful for initial testing before running the model fully. Specifying a `sample_rate` greater than 1 will process a sampled subset of origin and destination data rows.

```

103
104     sb.runSBModel(ageband, run_date, run_time,
105                  # optional testing parameters, for quick model run, remove or set to 1 for normal operation
106                  destination_sample_rate = 1, # process 1 in N rows of each destination dataset
107                  origin_sample_rate = 1      # process 1 in N rows of origin data
108                  )
109

```

Specifying a threshold to ignore very small weighting values in the background layer can also significantly speed up a run. This is done in the initial data load portion of the script. Setting the threshold to 0.0001 will decrease the run time, but warnings may be raised in the log when destinations are not near a major road and have nowhere to distribute their in transit population on to.

```

69     # Load the Background from an Ascii grid file
70     # Optional threshold parameter selection of lowest value to use for inTravel dispersion
71     # 0 includes all data, 0.0001 will speed things up for large areas, but can cause problems with dispersing population
72     # in areas with few transport links
73
74     sb.loadBackgroundFromFile('BckGrnds/' + sb.projParams.background,
75                             threshold = 0)
76

```

The script contains other options which enable the user to choose which output files are written and in what file formats, with sensible defaults specified; the user need only edit these settings if needed.

The tool uses the standard Python logging module to write useful information during all stages of session parameter and data loading and model run. The supplied `main.py` script sets default values to write log information to the standard output and include INFO level messages and above. The code illustrates other logging options – writing to a file or showing additional DEBUG, WARNING or ERROR messages only.

```

17     # Set up the destination, level and style of logging (INFO, WARNING, ERROR)
18     #
19     # logging.basicConfig(filename='log_sb2472py.txt', level=logging.DEBUG, format='%(asctime)s - %(levelname)s - %(message)s')
20     # logging.basicConfig(level=logging.WARNING, format='%(message)s')
21
22     logging.basicConfig(level=logging.INFO, format='%(message)s')

```

## Summary program operation

During a program run, *SurfaceBuilder247Py* processes all the input centroid data relating to the specified population sub-group within the study area, one centroid at a time, by working through all the specified origin files. Any part of the population sub-group which is not mobile is directly transferred to the same location in the output. The sum of the population counts in these input files defines the population available for modelling. Modelling then proceeds by working through all the destination centroids falling within the study area and considering their time profiles in relation to the target time and date set for the modelling run. Any centroid which has no associated population

activity for the sub-group being modelled at the time and date in question is discarded. All remaining centroids therefore have some level of population “demand” at the target time and population counts will be transferred to destinations from origins falling within the wide area dispersion function around each destination. This population is split between those who are OnSite and those InTravel. Thus, the time profile may suggest that a workplace is occupied by 90% of its capacity number of employees of 200 at 09:30. 180 employees from the relevant population sub-group will be transferred away from origin centroids falling within the wide area dispersion function, which is typically derived from census travel to work data or travel survey data. This function will determine what proportions of the 180 employees are obtained from specified distance bands of the destination. If insufficient population supply is available in any band, the search will be widened. Of these 180 employees, the time profile information may suggest that 80% are actually at the place of employment and a further 20% are in the transportation system. The OnSite populations will be assigned to the workplace using the local spread parameter and the 20% in the transportation system will be spread across the background layer using the weights provided by the background file, which will prevent allocation of population into areas such as the sea and will concentrate moving populations onto the transportation network. When all allocations to origins and destinations are complete, the original SurfaceBuilder redistribution algorithm is used to spread these counts across local cells as specified by the LAD parameter, for example spreading residential populations from a census OA centroid into the surrounding residential areas.

### *Output files*

A variety of default and optional output files are provided and are written to the Results folder by default. Filenames are automatically generated using the file prefix specified in main.py, together with the time modelled, and the relevant output descriptor. Outputs are written in ASCII Grid format at the specified resolution for the specified analysis area, enabling easy import into GIS and other software.

Table 2 shows the standard files produced for each run (for a single population subgroup and time):

**Table 2 Standard files provided for each run**

Filename	Details
Origins_Immob	Immobile population at origins
Origins_Remain	Population remaining at origins after redistribution to destinations on site and in transit
Origins_All	Total population at origins (sum of origins_immob and origins_remain)
Dest_OnSite	Population on site at destinations
Dest_InTravel	Population in transit to or from destinations
Total	Total population (sum of Origins_Immobile + Origins_Remain + Dest_OnSite + Dest_InTravel)

Further optional outputs may be requested by setting the relevant parameters to 'True' in main.py. For example:

- **Breakdown by destination type:** by default, the ASCII grids are produced for *all* destination types summed, but outputs may also be requested for *each* destination type separately e.g. education, healthcare etc.
- **Non-LD versions:** by default, the ASCII grids are produced *after* the Local Area Dispersion function has been applied (indicated by \_LD in their filename), but outputs may optionally be requested for the population on site at origins and destinations prior to application of the Local Area Dispersion function.
- **Results.csv:** additional output in a bespoke .csv format.
- **Sum for all population subgroups:** In order to obtain a model for the entire population, the same modelling scenario must be re-run for each population subgroup and the results summed. As described above, an example function (main\_allages) in main.py illustrates how to do this. By default main\_allages reads in the population subgroups to be summed from the origin file, but equally the user can specify specific subgroups for inclusion.

## Contacts

*SurfaceBuilder247Py* forms part of the broader University of Southampton Population24/7 programme of research. The Population24/7 team are continually developing their spatiotemporal modelling approaches, the *SurfaceBuilder247* software and their applications within various sectors. They are always keen to hear from other researchers or practitioners who intend to either use the program independently or are interested in contributing to its further development. The team can be contacted as follows:

Email: [Pop247@geodata.soton.ac.uk](mailto:Pop247@geodata.soton.ac.uk)

X / Twitter: @pop247

Samantha Cockings: [s.cockings@soton.ac.uk](mailto:s.cockings@soton.ac.uk)

Current and previous members of the Population24/7 team include: [Samantha Cockings](#), [David Martin](#), [Andrew Harfoot](#), [Jason Sadler](#), Hugh Darrah, Samuel Leung, Alan Smith, Becky Alexis-Martin.

Geography and Environmental Science, University of Southampton, Southampton, SO17 1BJ, UK

## References

Martin D, Cockings S, Leung Y (2015) Developing a flexible framework for spatiotemporal population modelling. *Annals of the Association of American Geographers* 105(4) 754-772 [doi: 10.1080/00045608.2015.1022089](https://doi.org/10.1080/00045608.2015.1022089)

## Annex A: SurfaceBuilder247Py Origin and destination data file formats

This is a comma separated values file which begins with 20 header records.

Line	Orig	Dest	Keyword	Content
1	Y	Y	Type	Dataset type: "Orig" or "Dest"
2	Y	Y	Title	Dataset title
3	Y	Y	Comment	Comment on data
4	Y	Y	Data block	N1, N2, N3 N1 = row containing data column header N2 = first row of data records N3 = number of data records
5	Y	Y	UniqueID	Data column containing area code (e.g. Postcode, OACode)
6	Y	Y	X	Data column containing X grid coordinate
7	Y	Y	Y	Data column containing Y grid coordinate
8	Y	Y	PopTotal	Data column containing total population for each record
9	Y	Y	PopSubGroups	G1 = Number of population subgroups used
10	Y	Y	PopSubGroupsData	{default values x G1},{columns x G1} {default values} = default % values for each of G1 sub-groups (sum = 100) {columns} = data columns containing each population subgroup value
11	N	Y	TimeProfile	Code, N1 Code = Default time profile scheme N1 = Data column containing specific time profiles
12	N	N	LocalDispersionType	Code, N1 Code = Used to define the Local dispersion function – currently not read as Cressman is used by default for all Origins and Dests (with spatial extent as defined in row 12 (LocalDispersion)) N1 = column defining centroid-specific LocalDispersion Type, if any
13	N	Y	LocalDispersionParameter	N1, N2 N1 = Default spatial extent of locations N2 = Data column containing specific spatial extents
14	N	Y	WideAreaDispersion	Spatial range distribution, N1 e.g. 100>5 2000>80 5000>5 10000>3 20000>2 5 where: 5% pop work from home (narrow radius of 100 specified), 80% pop >100 to 2000m, 5% >2000-5000m, 3% >5000 to 10000m and 2% >10000 to 20000 and finally 5% > 20000 – the radius of the final band is based on a uniform distribution density of all populations (in above example, 95% is within 20000, so radius of final band is 20000/sqrt(.95)) N1 = Data column containing spatial range distribution
15	N	Y	MajorFlows	{Columns x G} – only for Dests e.g. 00MSNE>5 00MSMY>5 00MSNA>2 – this represents 3 major flows for the dest – 5% of its population comes from origins with IdentifyingCode beginning with the characters '00MSNE', 5% from origins with IdentifyingCode beginning with the characters '00MSMY' and 2% from origins with IdentifyingCode beginning with the characters '00MSNA'. Note that only origins within the study area are included.
16	Y	Y	DOA	Reference date, N1 Reference date = textual description of reference data

				N1 = Data column containing specific reference dates
17	Y	N	Mobility	{default values x G1},{columns x G1} Only for Origins Mobility is a percentage and must be between 0 and 100 (0 = entire population subgroup is immobile; 100 = entire population subgroup is mobile) {default values} = Mobility for each PopSubGroup {columns} = data columns containing each population subgroup value
18	Y	N	RegionalIdentifier	Default region code (e.g. UK), N1 Only for Origins N1 = Data column containing specific RegionalIdentifiers (e.g. SW, SE)
19	Y	N	RegionalAdj	{values x G} Only for Origins A vector of RegionalIdentifiers and percentages (e.g. SW>109 SE>96 UK>100 = original value*109% in Southwest; original value*96% in Southeast; original value*100% in UK)
20	Blank line			
21	Blank line			

Line 22 (defined as N1 in data block field above) contains column numbers for ease of reference

Line 23 (defined as N2 in data block field above) is the first data record

Data block continues for N3 lines (as defined in data block field above)