

Scottish Index of Multiple Deprivation 2016

Geographic Access Domain Methodology

Alan Bragg and Lynne Addison
 Geographic Information Science and Analysis Team
 Scottish Government

1. Introduction

The Geographic Access domain was added to existing domains in SIMD 2004, to model the ability of Scottish citizens to reach a number of key services. Unlike the other domains, Geographic Access consists of two sub-domains; public transport journeys and journeys made by private car. This accounts for citizens who will be wholly dependent on public transport, alongside car owners.

The domain measures aspects of access deprivation and is relevant to all people, as it is equally important to be able to access key local services in both rural and urban areas¹. The development of the domain since 2004 is summarised in Table 1.1.

Table 1.1: Development of the Geographic Access Domain, 2006-2016

	<i>SIMD 2004</i>	<i>SIMD 2006, 2009, 2012, 2016</i>
Drive Times to...	<ul style="list-style-type: none"> • GP Surgeries • Fuel Stations • Post Offices • Retail Centres • Primary Schools 	<ul style="list-style-type: none"> • GP Surgeries • Fuel Stations • Post Offices • Retail Centres • Primary Schools • Secondary Schools
Public Transport Times to...	N/A	<ul style="list-style-type: none"> • GP Surgeries • Post Offices • Retail Centres

For SIMD 2004, the Geographic Access domain was based purely on the time taken to drive to basic services. For SIMD 2006 onwards secondary schools were added to the list of destinations. Supermarkets were removed in favour of retail centres; to capture areas with clustered multiple retail outlets such as highstreets.

In addition to changes in service destinations, SIMD 2006 also introduced the analysis of accessibility to basic services by public transport and/or walking. The Office of the Chief Statistician (OCS) has chosen to use the same indicators for SIMD 2009 onwards.

¹ www.scotland.gov.uk/Publications/2004/10/20089/45173

2. Data Sources

The data sources used in calculating the Geographic Access domain for SIMD are summarised in Table 2.1 below.

Table 2.1: Data Sources of the Geographic Access Domain, 2009-2016

	Dataset	SIMD 2009		SIMD 2012		SIMD 2016	
		Source	Records	Source	Records	Source	Records
Origins	Census Output Area Population Weighted Centroid	with Census Total Pop, GROS 2001 & Average School Age Pop, SG 2006-08	42,604	with Census Total Pop, GROS 2001 & Average School Age Pop, SG 2009-11	42,604	with Census Total Pop, NRS 2011 & Average School Age Pop, SG 2013-15	46,351
Destinations	GP Surgeries	OS PointX (December 2008)	1,066	OS PointX (December 2011)	932	Landmark PointX (March 2015)	1,126
	Fuel Stations	Experian Catalist (Q1, 2009)	943	OS PointX (Dec 2011)	891	Landmark PointX (March 2015)	889
	Post Offices	Royal Mail (Q1, 2009)	1,432	Royal Mail (Q1, 2012)	1,328	Royal Mail (2015)	1,276
	Retail Centres	CACI Retail (2008)	402	CACI Retail (2011)	469	CACI Retail (2015)	464
	Primary Schools	OS PointX (December 2008)	2,135	OS PointX (December 2011)	2,064	Landmark PointX (March 2015)	2,052
	Secondary Schools	OS PointX (December 2008)	381	OS PointX (December 2011)	375	Landmark PointX (March 2015)	368
Transport Network	Road Network	OS ITN (March 2009)	N/A	OS ITN 2012	N/A	OS ITN April 2015	N/A
	Road Speed data	N/A	N/A	ITIS Road Speed Data (March 2012).	N/A	& INRIX Road Speed Data (April 2015).	N/A
	Public Transport Network	National Public Transport Data Repository (October 2008)	N/A	National Public Transport Data Repository (October 2011)	N/A	Traveline National Dataset and Associate of Train Operating Companies data (October 2015)	N/A

Origins

Population-weighted Census Output Area centroids were used as origins because they best represent the distribution of the population throughout Scotland. Census Output Area centroids were created by the National Records of Scotland, (NRS) following the 2011

Census, and weighted according to the 'master' postcode grid reference within each output area. The master postcode is the postcode grid reference that is nearest the average, weighted by the household count, of all the postcode grid references for a Census Output Area². This is the first SIMD to use the 2011 Census Output Area centroids; SIMD 2012 used the 2001 centroids.

The population for each centroid, found by summing population totals from the census 2011 at each postcode, are used to population weight the results from Output Areas when aggregating the travel time results to Data Zones. The primary and secondary schools calculations do not use the census population totals, but instead use school age population totals for the years 2013 to 2015. These include students at primary or secondary stage from the Pupil Census from September each year, and are summed for each Output Area

GP Surgeries, Fuel Stations, Primary and Secondary Schools

Locations for GP surgeries, Fuel Stations, Primary Schools and Secondary Schools were taken from the Landmark's PointX product (March 2015 release). PointX includes data for all of Scotland, plus a 10 kilometre buffer into England so that accessibility can be appropriately observed for those who live near the border. PointX provides two sets of coordinates per point; one set representing the physical location of the point and a second set representing the nearest point on Ordnance Survey's Integrated Transport Network (ITN) product, so that the points can be used in network analysis.

The PointX data includes schools with both a primary and secondary department only once, typically as a secondary school. Additionally, junior high schools (schools with a primary department and a secondary department typically reaching only S3/S4) are mostly not included in the data.

Post Offices

Post Office data were obtained from the Royal Mail, which detailed the location and opening hours for all conventional branches and outreach services across Scotland. Any Post Office service provided for at least 6 hours or more per week were included in analysis, including mobile post offices. Home delivery services were not included in the analysis as they cannot be represented by a fixed point.

The coordinates provided by the Royal Mail were not spatially accurate. When checked against background mapping it was found that points were often located off of the road network and sometimes at a distance from the true post office location. To correct this issue, the Post Offices in the Royal Mail data were firstly matched with their record in the PointX Post Office dataset, to find the coordinates on the road network. For 150 post offices, there was no corresponding record in PointX, so the location of the Post Office was manually checked and relocated to the nearest location on ITN.

² <https://data.gov.uk/dataset/2011-census-output-areas>

Retail Centres

Retail center locations were derived from CACI Retail Footprint (2015) data. The Retail Footprint is a gravity model that defines catchments for shopping centers selling comparison goods in Great Britain³. The retail centre locations are based on postcode sectors, weighted by shop locations.

Road Network

For SIMD 2016, Ordnance Survey's Integrated Transport Network (ITN) product was used to model Scotland's roads. ITN consists of road geometry (motorways, pedestrian streets, alleys, etc.) and routing information (e.g. access, banned turns and other restrictions). Each segment of the road is allocated a unique Topological Identifier (TOID), which can be used to join other data to the network.

INRIX Road Speeds

Up to 2009, SIMD calculations were run using default road speed profiles from the software. For SIMD 2012 and 2016, observed road speeds were used for the drive time calculations. The data is compiled by the software and data provider INRIX, and is distributed via Basemap. The data is derived from GPS units and mobile phones, and produces car and van speeds for the road network, with average speeds recorded for rush hour, off peak and night times for both weekdays and weekends. The time periods of collection are shown in Table 2.2.

Table 2.2: Data Collection Periods for INRIX 2015 Data

Peak	Off-Peak	Peak	Night
07:00 to 09:00	09:00 to 16:00	16:00 to 19:00	19:00 to 07:00

For SIMD, most journeys to services would take place between peak and off peak times, and therefore night time journey speeds were not used. Some road segments don't have a corresponding speed recorded in the INRIX data, so their speed is derived from the speeds of surrounding segments. This occurs more often in rural areas (e.g. Western Isles, Orkney and Shetland) where there are fewer vehicles using these routes.

Public Transport Network

Public transport data, compiled from the Traveline National Dataset (TNDS) and Associate of Train Operating Companies (ATOC), were provided by Basemap. This is provided as a quarterly dataset including bus, rail, coach, ferry and tram data. Originally, the intention was to use the April 2015 release, to avoid problems with summer timetables in the July release and to comply with the September 31st 2015 cut-off date for SIMD 2016. This was later adjusted to use the October 2015 release, as there were a considerable number of

³ <http://www.caci.co.uk/products/product/retail-catchments>

stops missing from the April release. Although this makes an improvement to the data, it is believed that the data is not as accurate as the data for SIMD 2012 due to the exclusion of Department for Transport data that had previously been included within the dataset.

To account for a wait time required to make time to board before ferry journeys, the dataset builds an extra ten minutes into the timetabled journey time for ferries. Analysis of the data revealed this was not applied consistently to all ferry routes. Additionally, 10 minutes was not seen as a realistic period of wait time; in SIMD 2012, a wait time of 30 minutes for drive times and 20 minutes for public transport was applied if the journey utilised a ferry. To account for this inconsistency in the data, wait times were adjusted at the processing stage; see *Processing* section for details.

3. Methodology

Software

SIMD 2016 employed Basemap TRACC software V1.2.1 to calculate both drive and public transport journey times for the Geographic Access domain. TRACC uses SQL Server (localdb)\v11.0 databases to perform travel time calculations, process and store results.

Microsoft SQL Server Management Studio 2012 is used alongside TRACC. This allows the user to delete indexes created in TRACC after each public transport calculation. These indexes record public transport timetable and stop parameters, but quickly use up available database space when allowed to accumulate. SQL Server Management Studio was also required to import the road speed data correctly. TRACC automatically adds an 'OSGB' prefix to every TOID when the ITN network data is imported. This prefix prevents INRIX data from joining to the road network correctly so a SQL query was created by Basemap, and run in SQL Server Management Studio, to correctly import the INRIX data.

Accessibility Calculation

TRACC software is used to find the closest of each type of destination to every origin, and its travel time, by both driving and public transport. The travel time calculations must be performed for each Output Area, then the results aggregated to Data Zone level, to provide an average travel time for each Data Zone. Drive times were calculated for each of the six key services, and public transport times were calculated for only three services. Times were not calculated for schools, because these tend to use transport for which there is no available national data, and fuel stations. Key points of the methodology are as follows and the calculations are summarised in Figure 3.1:

- All calculations were performed for a Tuesday, as this was deemed to be the most representative weekday.
- As drive time speeds and public transport availability varies by time of day, travel times were calculated at four time windows to incorporate both peak and off-peak periods and outbound and inbound trips.

- Drive times for schools were calculated for only two time windows, one for the morning peak journey, and another for the evening. This aims to better represent the regular 'school run' pattern.
- Journey times must start and finish within these four hour time periods.
- A maximum journey time of three hours (180 minutes) is imposed.
- Drive time calculations can make use of ferries, providing these are not passenger ferry routes, as private vehicles cannot make use of these services.

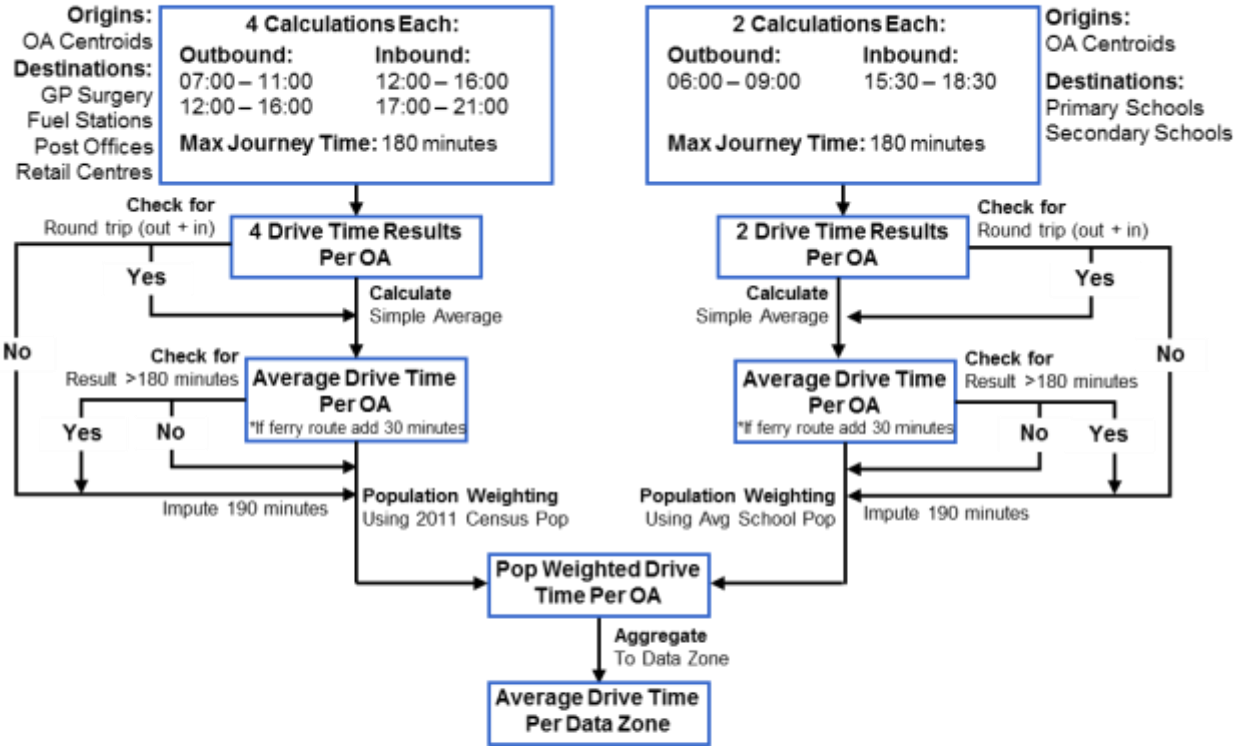
TRACC finds the number of nearest destinations stipulated by the user by performing a crow-flies distance calculation to each destination until it has found the allotted number. Then for each destination a travel time calculation is made. For this reason it is desirable to have more than one nearest destination, as the physically closest destination may be separated from the origin by a geographical barrier, such as open water, whilst the second nearest destination may be reached directly by road.

To calculate the drive times, TRACC routes from each origin to the nearest road network link, then through the road network at varying speeds to the destination using the quickest route. It is possible that the route driven also includes ferry transport. In TRACC (unlike previously, in Accession software), it is not possible to combine public transport and road network calculations in the *road network* mode, so the software has to be 'tricked' into performing the same sort of analysis through other means. So, the *public transport and road network* mode is used, which calculates a travel time as though the sections of road network traversed to and from public transport nodes were done so on foot, allocating a travel speed of 4.8 km/hour to these sections of the journey. Basemap's development team supplied a SQL query that can be run directly in the database of each TRACC project which updates walk speeds with INRIX road speed values, so TRACC will 'walk' through the network at vehicular speeds.

For public transport travel times, the *public transport and road network* mode is used which links each origin to the closest public transport stop, if available, at the 4.8 km/hour walking speed through the road network. This uses only the road network, not footpaths, and excludes walking on motorways. Using as many different modes of public transport as necessary, TRACC then routes each trip through the public transport network to reach the destination. Distances of up to 500m were permitted to interchange between public transport services, and 2km to reach the initial public transport stop. If the destination is within 2km of the origin, the travel time to the nearest destination would be the walk time if this is quicker than the public transport alternatives. If the destination is not within 2km of the origin, then the route taken will be via public transport stops, even if the services of these stops are not utilised.

Drive Time Calculations

GP Surgery, Fuel Stations, Post Offices, Retail Centres, Primary and Secondary Schools



Public Transport Time Calculations

GP Surgery, Post Offices, Retail Centres

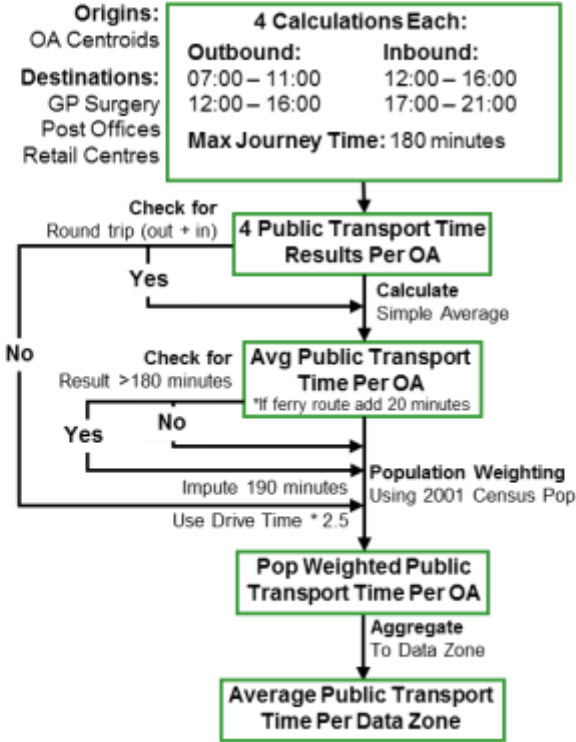


Figure 3.1: Summary of SIMD 2016 calculations and processing workflow

Processing

Once the drive time and public transport time calculations have been performed for each time window, these must be averaged to find overall output area travel times, then aggregated to the data zone geography. Results were processed using Python 2.7, firstly calculating a simple average of results for each output area. When an outward and inward journey were not possible, average values were not calculated for the output area. This occurred for three reasons for drive time calculations: no ferries are available within the time window, the ferry required is passenger only or the journey takes longer than 180 minutes. For public transport, this occurs due to no public transport being possible within the time window, or the journey taking more than 180 minutes. For all these cases, the drive time was adjusted to 190 minutes in an effort to represent the poor accessibility of the area. In previous versions of SIMD, a case occurred where the software could not link the Output Area centroid to the network. These output areas were excluded from further analysis, based upon findings from SIMD 2009 that the actual travel times would align with the calculated average. For public transport calculations, if an average travel time could not be calculated, the average was calculated as 2.5 times the equivalent drive time of this Output Area. Where the drive time is 190 minutes, the public transport time is also imputed as 190 minutes. For both public transport and drive time calculations, if the average journey time for the output area was greater than 180 minutes, the journey time was adjusted to 190 minutes. The imputation of 190 minutes was introduced in SIMD 2012, but was applied retroactively only to drive time results.

Due to the public transport data incorporating a 10 minute wait time into the timetable for some, but not all, ferries, corrections must be made for journeys which use a ferry at the processing stage. This is done by identifying those ferries which have already had a wait time added and cross-referencing this with a list of output areas which should have wait times added due to their requirement to travel by ferry. Then the required period of time is added to the average travel time: 30 minutes for driving (20 if already adjusted) and 20 minutes for public transport (10 if already adjusted). The final results of the output area processing steps are seen in Figure 3.2, showing how many output areas had results altered due to inaccessibility, long travel times, or to account for ferries.

		Fuel	GPs	POs	PS	Retail	SS
Drive Times	Accessible OAs	46344	46344	46350	46336	46323	46339
	Inaccessible OAs	7	7	1	15	28	12
	Inaccessible OAs Adjusted	7	7	1	15	28	12
	High Travel Time OAs Adjusted	0	0	0	0	0	0
	Ferry wait times added	29	16	2	17	55	58
Public Transport Times	Accessible OAs		45231	45283		45176	
	Inaccessible OAs		1120	1068		1175	
	Inaccessible OAs Adjusted		1120	1068		1175	
	High Travel Time OAs Adjusted		3	0		37	
	Ferry wait times added		16	2		55	

Figure 3.2: Summary statistics of output area processing for SIMD 2016

Output area drive times were population weighted using 2011 census figures from NRS, as this is the most recently available population data. For primary and secondary schools, a sum of three years (2013 – 2015) school age population was used. For some output areas, there were no pupils in the output area for each year. In previous versions of SIMD, this multiplier of 0 would result in the exclusion of the output area at the population weighting step. For SIMD 2016 this was altered so that for output areas with no pupils for all three years, a pupil population of 0.5 pupils was imputed. Following population weighting, the results were aggregated to data zone level to create an average drive time and public transport journey time in minutes for each destination from each data zone.

4. Output

Drive time averages in minutes, for each data zone, were provided for fuel stations, GP surgeries, post offices, retail centres, primary and secondary schools. Average journey time by public transport in minutes, for each data zone, were also provided for GP surgeries, post offices and retail centres only. These results were checked to verify that the travel times were reasonable and any changes in travel time compared to 2012 results are valid.