



Accession National Runs Final Report

Report

Prepared by MVA for
Department for Transport

October 2005

Document Control

Project Title:	Accession National Runs Final Report
MVA Project Number:	C3396600
Document Type:	Report
WP Reference:	MB
Temporary Directory & File Name:	C:\Documents and Settings\enielsen\My Documents\WorkingFiles\workspace.mva.co.uk\C3396600 - Accession National Runs Final Report.doc
Final Workspace Location	

Document Approval

Primary Author:	Matt Brown
Other Author(s):	Erik Nielsen
Reviewer(s):	Andrew Smith, Mike Slinn
Formatted by:	EN

Distribution

Issue	Date	Distribution	Comments
1	28/06/2005	MVA	Initial internal draft
2	16/09/05	MVA, DfT	Reviewed by DfT
3	04/10/05	MVA, DfT	Changes following DfT review
4	18/10/05	MVA, DfT	Changes following DfT review

© **MVA 2005** The contents of this proposal remain the intellectual property of MVA Ltd and may only be used in connection with the brief for which it was submitted. It is specifically forbidden to communicate the contents to any third party without prior permission in writing from MVA, and all reasonable precautions must be taken to avoid this occurring.



Contents

1	Introduction	3
1.1	Purpose Of This Document	3
2	Project Outline	5
2.1	Overview	5
2.2	Stage 0 – Information gathering	5
2.3	Stage 1a – Initial System Development	6
2.4	Stage 1b – Report Design And Development, Calculations for Two Areas.	6
2.5	Stage 2 – Remaining Report Development, Calculations For England, Run Reports For England.	6
3	Data Preparation	9
3.1	Overview	9
3.2	Calculation Areas	10
3.3	Origin Data	11
3.4	Destination Data	12
3.5	PT Data	13
3.6	Road	14
3.7	Origins	14
4	Calculations	15
4.1	Introduction	15
4.2	Parameter Choice:	15
4.3	Algorithm Changes To Connection Method	16
4.4	Algorithm Changes To Wait Times	17
4.5	Algorithm Changes To Sampling Interval	17
5	Reports	19
5.1	Introduction	19
5.2	Schools	19
5.3	Threshold Reports	19
5.4	Continuous Reports	20
5.5	Banding Reports	23
5.6	Composite Reports	23
5.7	Testing process	24
5.8	National Runs Current Hardware Setup	25
A.	Appendix: Straight Line Walk Distance Factor Research	27
a.	Summary	27
b.	Method	27
c.	Data description	27
d.	Findings	28
B.	Appendix B: Continuous Cut Off Methods	31



**C. Appendix C: Methods And Improvements Made Available To The
Wider ACCESSION User Community 35**

Tables

Table 3.1 Datasets for National Runs 9
Table 3.2 Origins with duplicate COA codes. 12



1 Introduction

1.1 Purpose Of This Document

- 1.1.1 This document provides a final report on the Accession National Runs project. Its purpose is to provide information on the key decisions taken during the project, reasons for actions taken, review what has been achieved and look at outstanding actions and follow on recommendations. **It should be noted that this document does not intend to explain how Local Authorities can use ACCESSION to reproduce National Runs indicators, although it will provide details of how the core indicators have been calculated, for information.**
- 1.1.2 Where appropriate, this document illustrates how some of the many changes made to ACCESSION in order to undertake the National Runs project have been integrated, free of charge, into the commercially-available versions of ACCESSION for the benefit of the user community.
- 1.1.3 The structure of the document is as follows:
- 1.1.4 Chapter 2 provides an outline of the key project events in order to understand the project workflow.
- 1.1.5 Chapter 3 provides detail on the data preparation tasks.
- 1.1.6 Chapter 4 provides detail on the Accession runs tasks.
- 1.1.7 Chapter 5 provides detail on the Oracle indicator tasks.



2 Project Outline

2.1 Overview

2.1.1 The purpose of the ACCESSION National Runs project was to perform England-wide accessibility calculations (not including London), and to provide a set of key indicators to the Department for Transport.

2.1.2 The project had four main stages:

- Stage 0 – Information Gathering.
- Stage 1a – Hardware setup, system development (calculations), report design.
- Stage 1b – Report design, report development, calculations for 2 areas.
- Stage 2 – Remaining report development, calculations for England, run reports for England.

2.1.3 The project was run in stages to reduce risk to both DfT and to MVA. As the computations being processed were unprecedented in terms of data volume, splitting the project up meant that manageable and achievable pieces of work could be planned at the end of the preceding stage. This avoided making decisions too early that could have a negative effect on processes, resources, and costs.

2.1.4 Throughout each stage, the project decisions were taken with respect to three main technical topics:

- Data preparation;
- Accession runs tasks; and
- Indicator production.

2.1.5 These are discussed in later chapters.

2.1.6 Throughout the project certain changes and improvements to ACCESSION have been made available to the wider user community through new versions, and an overview of these can be found in Appendix C.

2.2 Stage 0 – Information gathering

2.2.1 This stage was run to investigate what would need to be done to facilitate the successful computation of travel times, in a realistic and acceptable timeframe, for the whole of England. This included:

- How to split up the country into manageable chunks;
- Deciding what ACCESSION parameter settings would be appropriate for the computations; and
- What amendments would be needed to ACCESSION's computations.



2.3 Stage 1a – Initial System Development

2.3.1 The purpose of this stage was to set up the hardware for the system, to develop the ACCESSION code, and to design the reports. The key events of this stage were:

- Procuring and setting up of four computers for the project (one server to host Oracle, two high-specification PCs for performing the National Run indicators, and one data preparation computer);
- Making various amendments to the ACCESSION code to allow the runs to complete in a sensible time frame and within the constraints of computer memory (RAM);
- Designing a data repository for the large volume of data that the National-level calculations would produce. This was to be based on an Oracle relational database management system (RDBMS).
- Developing tools to export results from ACCESSION into the Oracle RDBMS.
- Working closely with DfT to refine the report requirements and design the processes that would be necessary to compute the indicators.

2.4 Stage 1b – Report Design And Development, Calculations for Two Areas.

2.4.1 The purpose of this stage was to undertake some report development, to complete the report design and to calculate travel times for two discrete areas. The key events of this stage were:

- Developing ACCESSION's ability to calculate "Threshold" reports on a National level, based on the Oracle RDBMS;
- Refining ideas and designing the "Continuous", "Banding", and "Composite" reports in close collaboration with DfT;
- Using the modified version of ACCESSION to compute travel times for the two largest calculation areas.

2.4.2 Note that this document describes these reports in greater detail in subsequent chapters. **Also, please note that the Continuous, Banding, and Composite reports are not available in the commercial version of the ACCESSION software.**

2.5 Stage 2 – Remaining Report Development, Calculations For England, Run Reports For England.

2.5.1 This final stage of the National Runs project was to complete all remaining development of the remaining National Run indicator reports, and to output all of DfT's required reports. The key events of this stage were:



- Completing the remaining report development, covering “Continuous”, “Banding”, and “Composite” reports;
- Performing all travel time calculations that were required by DfT;
- Running all reports (“Threshold”, “Continuous”, “Banding”, “Composite”) for England;
- Delivering all ACCESSION-based deliverables to DfT and writing this report.



3 Data Preparation

3.1 Overview

3.1.1 There are five different types of data used for the project:

- Origin data;
- Destination data;
- Population data (geo-demographics);
- Road data; and
- PT data.

3.1.2 This chapter outlines the steps and processes used to prepare these datasets.

3.1.3 The main source of data is the “DfT Data Disk”. Information was provided from the following sources: Edubase website, Department for Health, Geobusiness Solutions Retailers, the THALES website, the Office of National Statistics (ONS), and the Ordnance Survey (OS).

3.1.4 Below is a list of datasets, by type, listed with supply sources, geographical coverage, level of aggregation, and the format in which it was received by MVA. **Please note that all data have been used on DfT license agreements and have been used for the National Runs Project.**

Table 3.1 Datasets for National Runs

Datasets for National Runs

	Name	Coverage	Data Scope	Source	Format
Origins	Census Output Areas (COAs)	England	location	ONS	MapInfo
Destinations	Primary Schools	England+Buffer	location	DfT/Edubase	MapInfo
	Secondary Schools	England+Buffer	location	DfT/Edubase	MapInfo
	Tertiary Schools	England+Buffer	location	DfT/Edubase	MapInfo
	SOAs (work)	England+Buffer	location	ONS	MapInfo
	Number of Employed at OA (to be summed at SOA level)	England+Buffer	location	DfT	csv
	GPs	England+Buffer	location	DfT/DH	MapInfo
	Hospitals	England+Buffer	location	DfT/DH	MapInfo
	Supermarkets	England+Buffer	location	DfT/Geobusiness	MapInfo
Geodemographics	Children 5-10	England	COA	DfT / Plasc	csv
	Children 5-10 in receipt of free school meals	England	COA	DfT / Plasc	csv
	Children 11-15	England	COA	DfT / Plasc	csv
	Children 11-15 in receipt of free school meals	England	COA	DfT / Plasc	csv
	Population 16-19	England	COA	2001 Census	csv
	Households	England	COA	2001 Census	csv
	Households without access to a car	England	COA	2001 Census	csv
	2003 Population in receipt of Job seekers' allowance (Emp at Risk)	England	Ward	ONS 2003 Data	csv
	2001 Population of unemployed (in support of pop with JSA)	England	COA	2001 Census	csv
	2001 Population 16-74 (Employment)	England	COA	2001 Census	csv
	2002 Population by LA (in support of employment population)	England	LA	ONS 2002 Data	csv
Roads	Oscar AM	England+Buffer		OS	MapInfo
PT	Naplan Stops	England+Buffer	location	DfT/THALES	.csv
	ATCO CIF Bus	England+Buffer	LTA	DfT/THALES	cif
	- Rail PT	England+Buffer	LTA	DfT/THALES	cif
	- Other PT	England+Buffer	LTA	DfT/THALES	cif

3.1.5 In the above table, “Coverage” is generally either “England” or “England + Buffer”. The “buffer” is a 20km zone around England. It was used to ensure



that the project minimised the risk of border-related issues. For example, a hospital that was geographically located 100 metres into North Yorkshire County (in area 10) would possibly be used by nearby residents in Durham County (in Area 11). Therefore, the buffer zone ensured that the hospital was “included” in the calculation for Area 11 as well as for Area 10.

3.1.6 In terms of data scope, the table above shows the level at which we received the data. Definitions are as follows:

- “Location” means a single point (i.e. easting and northing);
- “LTA” means Local Transport Authority;
- “COA” means Census Output Area;

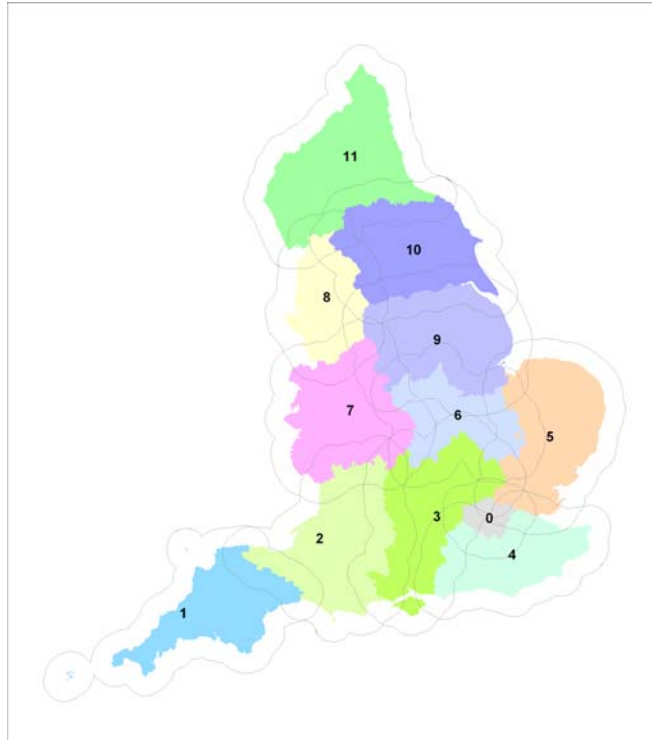
3.2 Calculation Areas

3.2.1 Calculation areas were introduced because in the commercial version of ACCESSION, there is an upper limit to the number of origins and destinations combined with the derived number of links from PT and road networks. This limit is dictated by the architecture on which ACCESSION was specified to run. Specifically:

- RAM (which is the most significant);
- Data size (total data size must not exceed 2GB); and
- CPU time needed to complete calculation.



3.2.2 England was split into 11 calculation areas based on estimated number of stops and services, destinations and origins within each of the areas. These 11 areas did not match with any existing regional structures and were decided on for purely practical reasons. The size of the calculation areas was also influenced by the fact that DfT wanted all services and destinations within 20 kilometres of any origin to be included in the accessibility calculation. The size for the calculation areas were derived by balancing import times and calculation times in Accession.



3.3 Origin Data

3.3.1 The origin data is made up of population weighted Output Area Centroids. These are as supplied by Office Of National Statistics (ONS) for the 2001 census.

3.3.2 In a minority of cases, some OAs have duplicate points, i.e. multiple coordinates for the same OA Code. This can occur where population is disparate. We are aware of 11 instances of this. Since we require that all OAs are unique, we have removed duplicate records from the relevant calculation areas. The point removed was chosen based on a match with an OS population centre map without duplicate points. The full process is described in Information Note 10: Duplicate COA Population Centres.

3.3.3 The OA codes affected were:

Table 3.2 Origins with duplicate COA codes.

ORIGINID	COACODE		ORIGINID	COACODE
63182	00MWPG0008		94805	26UCHB0001
63183	00MWPG0008		94806	26UCHB0001
69263	13UDHE0004		94807	26UCHB0001
69264	13UDHE0004		115044	35UBFT0003
78109	18UGHQ0006		115045	35UBFT0003
78110	18UGHQ0006		132744	43UJFT0010
78630	18UKGL0001		132745	43UJFT0010
78631	18UKGL0001		133238	43ULGF0006
80944	20UEGM0001		133239	43ULGF0006
80945	20UEGM0001		140108	47UCJA0006
85979	22UGGS0010		140109	47UCJA0006
85980	22UGGS0010			

3.4 Destination Data

3.4.1 The destination data came from many sources as described in 3.1.

3.4.2 In principle, all destination data have the same basic attributes:

- Name of location;
- Easting and Northing (x, y coordinates); and
- A "weight". (All destinations have a weight of 1, apart from employment data which has the number of workplaces as weight). The weight is used to balance accessibility indicators for reporting purposes.

3.4.3 This data underwent the following process:

- 1) Collating destinations of a given destination type into a GIS format.
- 2) Assigning a unique ID to each destination.
- 3) Loading destination data into Oracle together with any necessary weighting information.
- 4) The calculation areas were extended using a DfT-specified 20Km buffer. These became known as "clipping areas". Clipping areas were used to split the destinations into calculation areas. Note that one destination may be in several calculation areas (however, an origin could not be in more than one calculation area).
- 5) Loading the destinations into the ACCESSION repositories for the calculations.

3.4.4 In the above list, steps 1, 2, and 4 were performed using a standard GIS tool (MapInfo).



3.5 PT Data

Importing PT Data Into ACCESSION National Runs

3.5.1 ATCO CIF files (timetables) were loaded on a calculation-area basis.

3.5.2 In ACCESSION, the size of the PT network in memory during a calculation is critical. It must be kept to a minimum in order to ensure both sensible processing times and the viability of the calculations themselves. In order to achieve this, a process was devised and utilised during the preparation of the PT networks used for the calculations.

3.5.3 This process is described below:

England-Wide:

- NaPTAN stops on a National level were loaded into an ACCESSION repository. This repository became the foundation PT data for each of the 11 calculation areas.
- Identification of which PT data supply areas intersect with the each of the Calculation Areas plus 20 km buffer.

For each PT Area:

- Import of all ATCO CIF files for the area in question, for each mode in turn;
- Removed all unused PT stops from the repository (made available in version 1.3 of ACCESSION);
- Removed all services and stops that does not fall strictly within the 20km buffer of the Calculation area. (to do this, a strict clipping tool was created for the National Runs version of ACCESSION, to be made available in the next commercial version, ACCESSION 1.4);
- Deleted all records from the StopAlias table in the ACCESSION repository as they are not actively used in the calculations.

Improvements To ACCESSION's Importers

3.5.4 Difficulties were encountered during the Importing process, and a number of amendments were made to the ACCESSION ATCO CIF importer to deal with these. The changes involved fixes to the merging and duplication of trips when performing multiple imports to the same PT network.

3.5.5 In addition to improvements made to the ATCO CIF importer, MVA also made revisions to other areas of ACCESSION to improve performance, stability, and usability. So as to improve the commercially-available ACCESSION product, these changes were implemented and rolled out to our user community in hot fixes 1.2.1 and 1.2.2 and released in full in 1.3.



3.6 Road

- 3.6.1 The OSCAR Asset Manager data set was used as a road link data set for road-based calculations (for the National Runs project, this was only used for cycle journeys).
- 3.6.2 The OSCAR network was “clipped” to the buffered calculation areas using MapInfo and then loaded into ACCESSION using the standard import settings for OSCAR data.

3.7 Origins

- 3.7.1 First, a MapInfo table containing all origins (COA population centres) was created. Each origin was assigned a calculation area identifier. The origins were then imported as one dataset into ACCESSION. This allowed a unique ID to be allocated to each origin. The COA code was stored in the Description field and the CalculationAreaID was stored in the weight field.
- 3.7.2 To maintain the relationship between origins in the ACCESSION repositories and the origins in Oracle, it was important that all origins had a unique identifier across all calculation areas. Therefore the following process was used to create the 11 distinct origin sets needed in ACCESSION. Once the origins (COA centroids) were imported into ACCESSION, 11 records were inserted in the “OriginSet” table, i.e. one record per calculation area. This allowed for the OriginSetID column in the Origin table to be updated to match the ‘CalculationAreaID’ imported with the Origins.
- 3.7.3 This origin dataset was then exported from ACCESSION to the Oracle RDBMS. There are two important identifiers in this table: OriginID and COAcode.
- 3.7.4 The Origin table in Oracle contains information on population for the COAs.



4 Calculations

4.1 Introduction

4.1.1 For the National Runs project, there were essentially two calculations that were carried out for each of the 11 areas:

- PT and Walk Calculation; and
- Road Cycle Calculation.

4.1.2 PT Calculations involve calculating the accessibility to all destinations from all origins using the PT network that was loaded from ATCO CIF files. This calculation was performed using ACCESSION's standard network accessibility calculation module as a **PT and walk (no road network) calculation**.

4.1.3 Road Cycle Calculations involve calculating accessibility to all destinations from all origins using the OSCAR-based road network that was loaded. This calculation was performed using ACCESSION's standard network accessibility calculation module as a **road-only cycle calculation**.

4.2 Parameter Choice:

4.2.1 The parameters used (that affected calculation results) are listed below.

Average Walk and Cycle Speeds

4.2.2 Walk speeds were kept at the default 4.8km/h. Cycle speeds on the network are also the default 16km/h on all roads except on motorways, private roads with restricted access, pedestrian streets, and alleys. For these exceptions, the speed was set at zero, i.e. ACCESSION's default value.

Straight Line Walk Distance Factor

4.2.3 The straight line walk distance factor was set to 1.4 following a brief research study and some testing, as instigated by DfT.

4.2.4 Indications from the study is that the average detour distance for trips shorter than 2.5 km crow flies distance in East Sussex is 1.4 -1.5. However the composition of origins and destination also has an effect on the desired Detour Factor used so the case is not clear cut. Following the research DfT asked MVA to use the factor 1.4 for the calculations leading to the National Indicator values.

4.2.5 A more detailed summary can be found in Appendix A.

Connection And Interchange Distances

4.2.6 The maximum connection distance for national indicator calculations is 2Km. This is to ensure that even relatively remote COAs will connect to the PT and road networks.



- 4.2.7 The maximum interchange distance is set at 0.5 km.
- 4.2.8 The distinction between connection distance and interchange distance is that the connection distance indicates the maximum distance for creating links for origins and destinations to road nodes and stops, where the interchange distance gives the maximum distance for creating links between stops. The introduction of a reduced value for the interchanges improves the memory requirements and calculation times dramatically whilst having no detrimental effect on calculation results.

Time Period

- 4.2.9 The time period is from Tuesday 07:00-09:00, not the originally specified 08:00-09:00 as this would cause problems for the continuous calculations. In Accession 1.2 all journeys have to start and finish within the given time period. If only one hour was allowed for the calculation the results would not represent a fair value for the morning peak with the given thresholds being considered.

Catchments

- 4.2.10 In addition to the above time limitations a maximum journey time of two hours has been set. The primary reason for this is to limit the number of Origin-Destination pairs returned by the cycling calculations. This also reduces the calculation times.

4.3 Algorithm Changes To Connection Method

- 4.3.1 In order to perform the calculations, it was necessary to optimise ACCESSION's memory usage so as to reduce calculation times. Therefore, the connection method was redesigned to provide for optimised memory management. There were two elements addressed: network building times; and reducing the number of connection links.
- 4.3.2 The major time saver was the introduction of a spatial index for the network builder which helps to speed up network building on large networks with large amount of nodes. This addressed both the issue of network building times and reducing the number of connection links. This has been added to the commercially-available version of ACCESSION since version 1.2.
- 4.3.3 To reduce the number of connection links (and therefore save memory), a new method for connection was devised and programmed. This method introduced a limit of 8 links when connecting Road Nodes with Origins, Destinations, and PT Stop Nodes per link type (per Origin, Destination or PT Stop). This has also been added to the commercially-available version since ACCESSION 1.2. The method for selecting the 8 connections splits the connection area of a road node into 45-degree segments and creates a connection to the nearest node in each of the segments.



- 4.3.4 In addition, a method was added that reduced “stop-hopping” by disallowing stops with no services in the time window to be included in the calculation. This has been added to ACCESSION 1.3.

4.4 Algorithm Changes To Wait Times

- 4.4.1 A new method for path finding in time tabled PT networks has been implemented for the ACCESSION National Runs project. The method, broadly involves the elimination of wait time at the start and end of a PT journey. This is different from the way ACCESSION 1.2 handles PT journeys.
- 4.4.2 The issue of wait time at journey ends stems from a combination of the Sampling Interval and an implementation “tweak” to speed up calculations. The calculation tweak means that journeys are calculated from the destination to the origin rather than the other way. The wait time effect would also be present if the journey was calculated from origin to destination but it would appear at the start of the journey instead.
- 4.4.3 This has been implemented (as an option, to allow users to maintain continuity with previous calculation runs) since version 1.3 of ACCESSION for PT-only network accessibility calculations.

4.5 Algorithm Changes To Sampling Interval

- 4.5.1 In order to model travel time by Public Transport, ACCESSION models the best available options at set intervals (e.g. every 10 minutes) throughout the period being assessed. This has the benefit of reducing the amount of data required by the algorithm to compute all journeys. The downside to this is that the wait time for any given journey can vary by anything up to the sampling interval depending upon the time of arrival at the PT stop.
- 4.5.2 The default sampling time in ACCESSION is ten minutes. This sample interval does not provide results of the right granularity for the National Runs indicators. The trade off between accuracy and calculation time was found to be best with a combination of the no wait time approach and an increased sample frequency of five minutes. This doubles the processing time but cuts the average “error” by $\frac{3}{4}$. For National Runs calculations, the sampling interval was therefore lowered to 5 minutes combined with the new method of ignoring wait time at both journey ends. This produces the best measures of journey times.



5 Reports

5.1 Introduction

- 5.1.1 The ACCESSION National Runs project produced four main categories of reports: Threshold; Continuous; Banding; and Composite.
- 5.1.2 MVA have supplied DfT with a document called “Reports Design Document” that defines in full the methodology used to produce the reports. **This section is intended only to provide an overview of the reports and to describe the key decisions taken on report implementation.**
- 5.1.3 The report design process was led by DfT, with input from MVA where appropriate.

5.2 Schools

- 5.2.1 During the development of the ACCESSION National Runs project, it was identified that there were issues with the schools destination data sets. Broadly speaking some infant and junior schools sharing the same location were removed from the primary school dataset, whilst schools with sixth forms were added to the original Further Education dataset.
- 5.2.2 Two steps were taken to correct this:
- An exclusion flag was added to the destination table and the queries were changed so results for these destinations were not included in the report calculations.
 - Where the missed out further education locations coincided with secondary schools already included in the calculations a synthesised calculation result set was created. The new FE destinations were added to the “Destination” table in Oracle. These destinations have IDs starting from 200000. A correspondence list between FE destinations and their coinciding secondary schools destinations was used to copy cycle and PT results and insert them with new destinations thus effectively adding new results to the data sets.
- 5.2.3 The school amendment file was provided by DfT as a list of “DestinationIDs” to be excluded and a list of “DestinationIDs” to be duplicated with a new destination type.

5.3 Threshold Reports

- 5.3.1 Threshold reports measure accessibility by allowing the user to limit the desired results by those paths that fall within the given thresholds e.g. within 15 minutes. Interim calculations generate results for the SOAs, WARDS, LAs and LTAs within specified geographic areas and across calculation types. These are collated to form results for England as a single output.



5.3.2 The look up table used to aggregated data from COA to LSOA, Ward, and LA are the Lookup 1 (OA_LSOA_Ward_LA_lookup_Aug04) available from National Statistics. The LTA level has been added by MVA.

5.3.3 The deliverables for the threshold reports were 4 Excel spreadsheets:

Table 5.4 Threshold file deliverables

Worksheet	Rows	Columns
ThreshLSOAs	27717	112
ThreshWards	7304	111
ThreshLAs	321	110
TheshLTAs	116	109

5.3.4 During the implementation of the Threshold reports, two issues were addressed: zero population; and accuracy of ratios.

Zero Population

5.3.5 Whenever a “division by zero” population has taken place, ACCESSION National Runs replaced the result by a negative number. However, the specification for Threshold reports dictates that none of the results should be negative. Therefore, ACCESSION substitutes negative numbers with a full stop (‘.’) when the reports are written to Excel or to csv files. Thus the ‘.’ value in a report implies that no population was present for the group in discussion for this specific aggregation (LSOA, Ward, LA, or LTA).

Accuracy Of Ratios

5.3.6 The accuracy of ratios shown in Excel and csv files is one decimal place, as requested by DfT. The actual data accuracy is a lot higher, but for reporting purposes, the values are rounded to one decimal place using simple arithmetic rounding.

5.4 Continuous Reports

5.4.1 The purpose of Continuous calculations is to measure accessibility for an origin, but taking into account all relevant destinations. There are broadly three types of continuous calculations:

- Relative Hansen;
- Weighted Relative Hansen;
- Relative Hansen Aggregated Across Modes; and
- Weighted Relative Hansen Aggregated Across Modes.

5.4.2 Relative Hansen and Weighted Relative Hansen values are progressively calculated at the COA level for each Mode and Destination Type combination,



then grouped across the reporting area levels (LSOA, Ward, LA, LTA) to form the final indicator values.

- 5.4.3 During the implementation of the Continuous reports, three issues were addressed: **volume of data**; **selection of relevant destinations**; and **aggregation of values**. The method of addressing these issues is described below:

Volume Of Data

- 5.4.4 The calculation method for continuous reports required the processing of all origin-destination pairs that were produced by ACCESSION's network accessibility calculation. This resulted in a high volume of data. In order to meet timescale deadlines, the continuous calculations needed to be processed using subsets of the entire national dataset.
- 5.4.5 To achieve this, the data set (output of ACCESSION's network accessibility calculation) was split into smaller groups using a number of different fields for categorisation. These groups, ranked by increasing complexity of implementation are: Calculation Area; Mode (for some areas); and Destination Type.
- 5.4.6 The split into smaller groups had no effect on the final results of the Continuous reports.
- 5.4.7 The National Runs version of ACCESSION report screen (that is accessed by the "user") deals, for a given area, with exactly one batch of data. As described above, for certain areas, data were split by mode, effectively creating several "batches" for a given area. In order to circumvent this problem, the standard ACCESSION National Runs report screen was not used for final production of the deliverable reports. Instead, a manual process was invoked whereby the report calculations were done through Oracle (using SQL calls) and the results were merged prior to delivery to DfT.
- 5.4.8 In future, the individual report calculations *could* be done through the ACCESSION National Runs interface, but sufficient time would need to be allocated for doing this.
- 5.4.9 Data for each area were processed to the stage where we had the final Relative Hansen and Weighted Relative Hansen values at COA level for each destination type (e.g. Primary Schools, Secondary Schools, etc.). This process was invoked "manually" by making calls in Oracle to stored procedures (code) written by MVA. At this stage, intermediate results that were used to calculate the Relative Hansen and Weighted Relative Hansen values were deleted from the database. Following calculation of these disaggregated Hansen values, the aggregations for the Continuous reports were produced, again by making calls in Oracle to stored procedures (code)



written by MVA. The Continuous reports submitted to DfT were produced using the ACCESSION National Runs report screens.

- 5.4.10 By splitting the processing into stages, as described above, together with significant tuning of the Oracle database (tablespace allocation, index and primary key management, SGA tuning), we maximised the speed at which reports could be produced.

Selection Of Relevant Destinations

- 5.4.11 This DfT-specified process limits the destinations that are considered during the calculation of the continuous indicators. This selection process for destinations applies to all "Continuous" reports.

- 5.4.12 Two different thresholds were introduced to exclude destinations from contributing to the Weighted Relative Hansen figures:

- Cumulative Cut-Off Value; and
- Minimum Value Threshold.

- 5.4.13 The Cumulative Cut-Off Value is used to restrict the destinations being considered when calculating the indicator value for an origin. The Cumulative Cut-Off Value is computed as a percentage of the sum of all indicator values for the origin and specifies the maximum sum of Hansen value of destinations to be excluded. It was decided (subject to possible future revision) that a value of 5% would be used. Any destinations whose cumulative Hansen value was greater than 5% of the total Hansen value per origin would be included in the calculation of the indicator.

- 5.4.14 The Minimum Value Threshold specifies the minimum value by which a destination must be included in the calculation. If a destination represents more than 4% of the sum of indicator values for all destinations for a given origin this value cannot be excluded, this filter overrides the cumulative cut-off value filter.

- 5.4.15 The two thresholds in conjunction create a flexible filter that allows for both dense and less-dense destination types to be computed efficiently.

- 5.4.16 A full explanation of the threshold and cut-off values can be found in Appendix B of this document.

Aggregation Of Values

- 5.4.17 In the initial report design, indicators were to be either summed or averaged. However, it was subsequently decided that the indicators should be based on a **population weighted average**, as an un-weighted average would be misleading. It would be misleading because thinly populated areas and densely populated areas would be assigning an equal weight. This



becomes particularly problematic when specific population segments are examined (e.g. people at risk in regards to health). Summing indicator values without weighing them by population could result in undesired values. For example, one person with a 10-minute journey to hospital plus 50 people with a 25 minutes journey to hospital would give either 17.5min (arithmetic average) or 24.7min (population-weighted average).

5.4.18 Therefore we do not attempt an aggregation of the un-weighted Relative Hansen measure.

5.5 Banding Reports

5.5.1 The banding reports are the result of a reformatting the results from the Continuous reports and then scoring of the result. The bandings are performed on two levels: LSOA level and Ward level. LSOAs and Wards were ordered and then placed into decile bands according to their indicator scores (Weight Relative Hansen and Aggregated Weighted Relative Hansen). Each decile band received a Decile Band Value of between 1 (“least accessibility”) and 10 (“most accessibility”). The 10 Decile Band Values are as follows:

Table 5.1 Decile Bandings

Band	Lower SOAs	Wards
Decile 1	1 to 2771	1 to 730
Decile 2	2772 to 5543	731 to 1461
Decile 3	5544 to 8314	1462 to 2191
Decile 4	8315 to 11086	2192 to 2922
Decile 5	11087 to 13858	2923 to 3652
Decile 6	13859 to 16630	3653 to 4382
Decile 7	16631 to 19402	4383 to 5113
Decile 8	19403 to 22174	5114 to 5843
Decile 9	22175 to 24946	5844 to 6574
Decile 10	24947 to 27717	6575 to 7304

5.5.2 Sources for the geography lookups are described in 5.3.2.

5.5.3 Decile Bands were provided by DfT for SOA and Ward level. It was decided that banding reports should be limited to these two levels.

5.6 Composite Reports

5.6.1 The Composite Reports are a combination accessibility scores for a variety of destinations and populations. For example, accessibility measures for education establishments (primary schools, secondary schools, and colleges) may be combined to produce an overall accessibility score for “education”, for a given area (e.g. LSOA) and particular population (e.g. “at risk”, “all”).



- 5.6.2 Composite reports required the calculation of a Weighted Corrected Relative Hansen value. This value is computed at COA level, for each destination type, and for each mode (PT, Cycle, both PT and Cycle aggregated). These values were then put through several processes in order to score and rank SOAs, Wards, LAs and LTAs, by sector types (e.g. "education", "health", "all", etc.) and by population (e.g. "at risk", "all").
- 5.6.3 It was clarified that each of the composite Weighted Corrected Relative Hansen indicators should evaluate into a number between zero and 1. The weighting for the different groups are:
- Education All = Primary All (PT) *0.25+ Secondary All (PT / cycle) * 0.25 + Tertiary All (PT / Cycle) *0.5
 - Education Risk = Primary Risk (PT) *0.5 + Secondary Risk (PT / cycle) * 0.5
 - Health All = HospitalAll (PT) *0.5+ GPAll (PT) *0.5
 - ALL = (EducationAll + WorkAll + ShopAll + HealthAll) /4
- 5.6.4 We use one step to calculate all aggregated values. The description in the Report Design Document supplied to DfT (chapter 8.2) describes several different steps which we have reduced to one step in our calculation where aggregated values are calculated.
- 5.6.5 Where the cycle mode is not available for a particular destination type (e.g. hospitals) the modal split for that destination type ensures that the cycle indicator is not included.
- 5.6.6 The different weighting described in the document is achieved in the calculation by using a weight for all destinations (remember that only workplace destinations have a weight different from 1).

5.7 Testing process

- 5.7.1 The testing process for National Runs reports was performed using a test data set consisting of parts of a calculation area.

Testing consisted of taking results from Accession 1.2 which have passed 'Black box' tests as part of the Accession product testing phase and using them as input data for the National Runs Oracle database whilst independently performing the relevant calculations, using MS Excel, to produce hand calculated equivalent results.

- 5.7.2 The hand calculated results were then used as target results for 'black box' testing the results from the Oracle database.



5.7.3 The initial 'Black box' test of the Oracle results led to a further 'White box' test where any discrepancies were found.

5.7.4 The result of the test was a list of corrections to the processes in Oracle that led to fully compliant results to a suitable degree of accuracy considering the inherent differences in the software packages used as well as the level of accuracy needed for the output.

5.8 National Runs Current Hardware Setup

5.8.1 The ACCESSION National Runs hardware setup consists of 4 servers. Listed below are the specifications and configurations of the servers.

"ACCESSIONRUNS" Server

5.8.2 Primary calculation machine, for storing data and performing large calculations over long periods of time. Dual processors allow two calculations to be performed in parallel.

- Dual 3.6GHz Xeon CPUs;
- 2 x 136GB hard disks in RAID-1 producing 136GB of storage; and
- 4GB of RAM.

"ACCESSIONRUNS2" Server

5.8.3 Secondary calculation machine (and backup machine) for ACCESSION National Runs to use while the ACCESSIONRUNS server is in use. Same specification as the ACCESSIONRUNS server.

"ACCESSIONDATA" Server

5.8.4 Data storage server to store large amounts of GIS mapping data and allow data manipulation to prepare for calculations on the ACCESSIONRUNS machines.

- Single 3.4GHz Xeon CPU;
- 2x 74GB hard disks in RAID-0 producing 136GB of storage; and
- 4GB of RAM.

"ACCESSIONORACLE" Server

5.8.5 An Oracle server that handles the data that Accession calculations produce on the ACCESSIONRUNS machines. Multiple databases are stored on this server taking up around half the available disk space.



5.8.6 The server also has a drive which acts as a backup for specific folders on each of the four servers. The backup occurs weekly.

- Single 3.6GHz Xeon CPU;
- 4x 136GB disks in RAID-0 producing 546GB of storage;
- 4GB of RAM; and
- Oracle Standard Edition.



A. Appendix: Straight Line Walk Distance Factor Research

a. Summary

- i. This Appendix describes the method used to create a basis of trip detour factors as a function of the straight line (crow flies) distances between points and the distance travelled through a road network and the conclusions found.
- ii. Indications is that the average detour distance for trips shorter than 2.5 km crow flies distance in East Sussex is 1.4 -1.5. However the composition of origins and destination also has an effect on the desired Detour Factor used so the case is not clear cut.

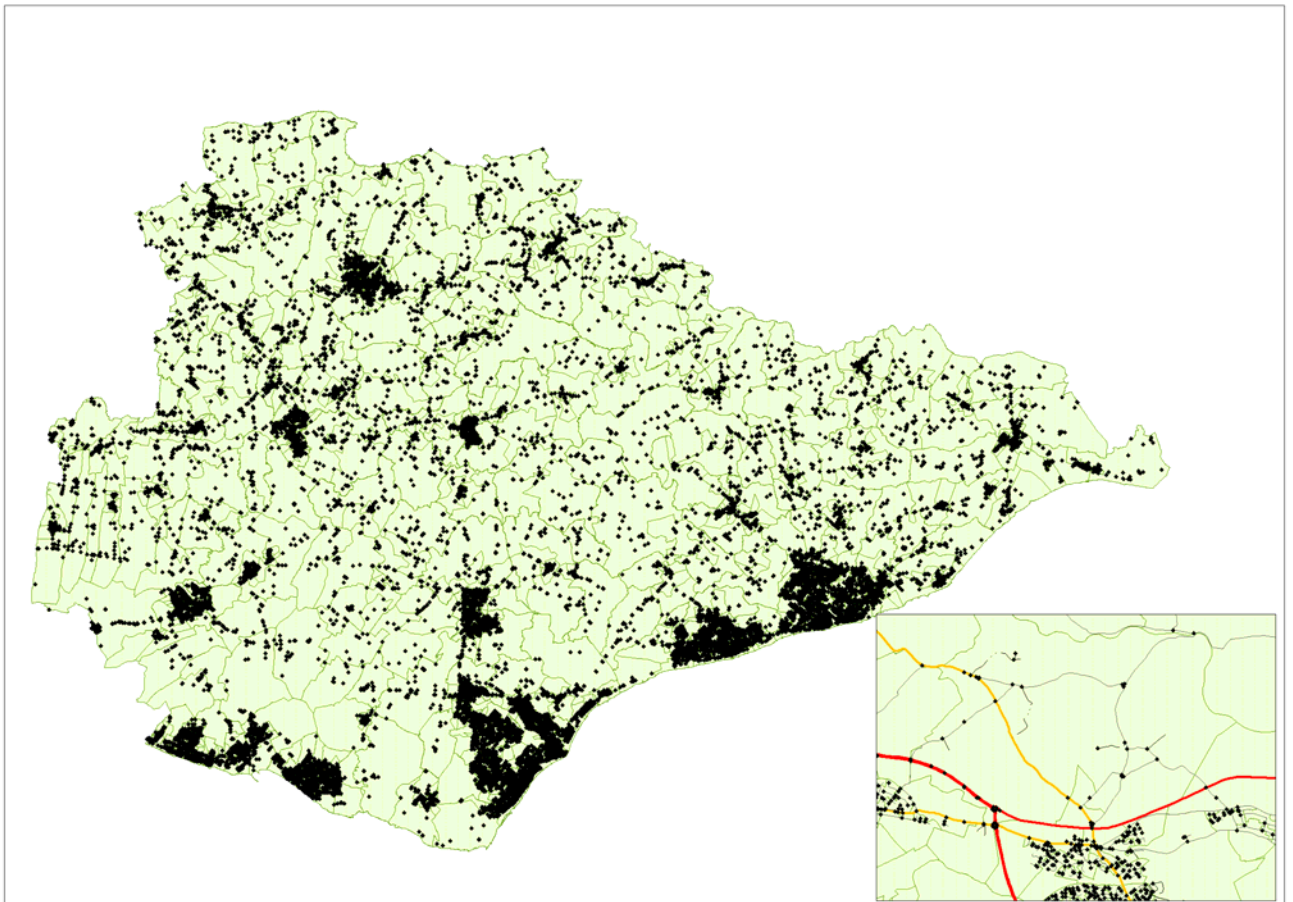
b. Method

- i. Assume that 'paths' used to connect the network nodes (Stops and Junctions) to Origins, Destinations and other network nodes (by interchanges and stop to junction connections) are of a similar geography as the roads included in the OSCAR map.
- ii. If this assumption is correct then an average Detour Factor for journeys can be given as the travel distance by road from any road node to any other road node divided by the Cartesian distance between the two nodes.
- iii. By creating an Origin set and an identical Destination Set based on the starting points of road links in OSCAR a big number of trip distances can be calculated in Accession.
- iv. Since the maximum crow flies connection distance in National Runs is 2km a catchments of 5km (along the road) has been chosen. This enables trips of detour factors up to 2.5 to be included for the longest crowflies distance.

c. Data description

- i. East Sussex is selected as a test case (relatively flat and a good mix of urban and rural areas).
- ii. Around 28,000 road nodes are selected to act as Origins as well as destinations for a road distance based run with a cut off of 5 km. Connection distance will be set to something very low (10m). See map below.



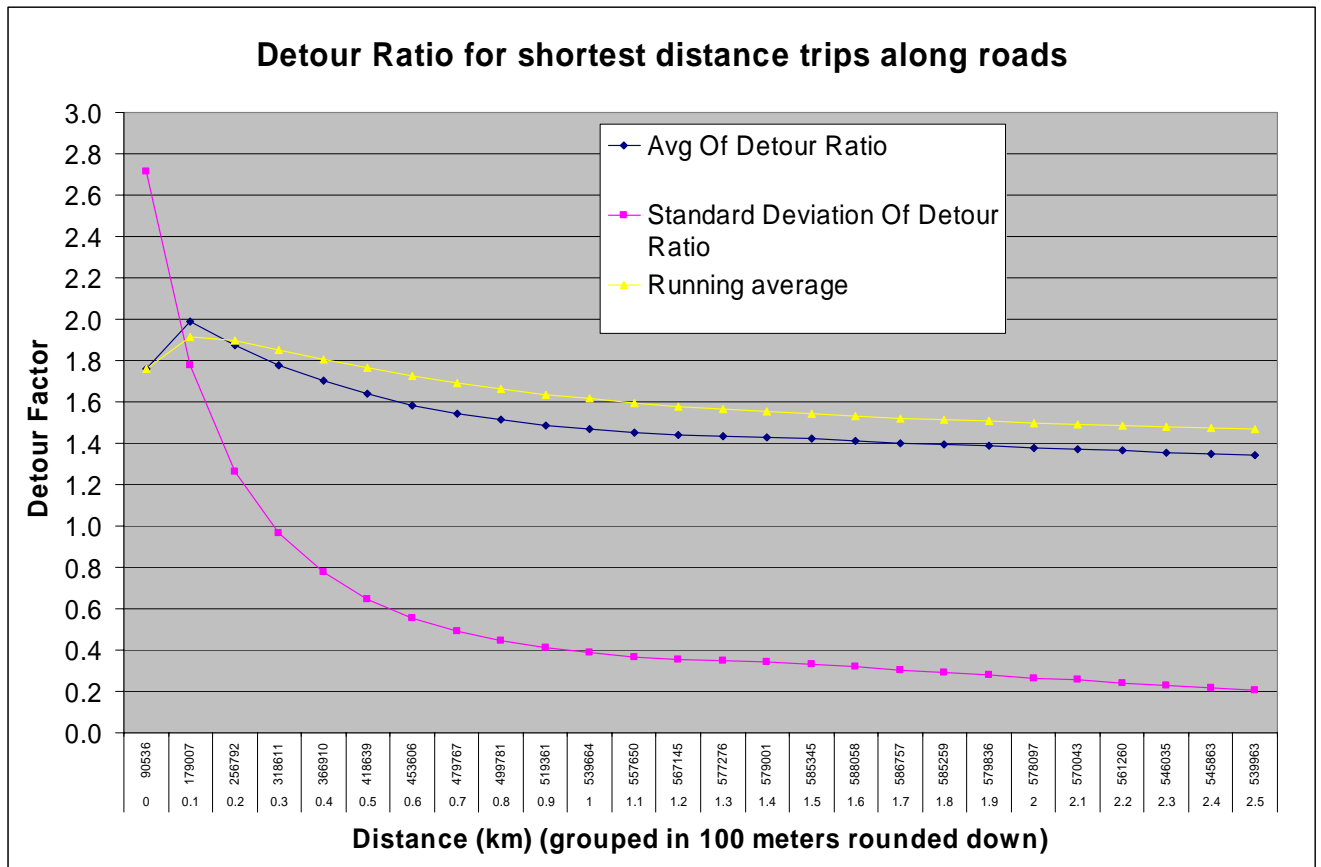


- iii. An Accession network accessibility calculation with the specific selection of Origins and destinations (that matches an Oscar Network Nodes exactly) was carried out.

d. Findings

- i. This study has found (looking at more than 20 million trips) that the average detour factor expectedly varies with the distance travelled. Very short journeys tend to have a higher average detour factor than the longer trips.
- ii. The results have been statistically manipulated to give us the best possible idea of detour factors for trips. All trips Crow flies distance was calculated and grouped into intervals of 100m. The following statistics were recorded:
 - count
 - Average of Crow Flies distance
 - Standard Deviation of Crow Flies Distance
 - Average of Journey Distance
 - Standard deviation of Journey Distance
 - Average Journey Distance/Crow Flies Distance
 - Standard deviation of Journey Distance/Crow Flies Distance.

- iii. In the table below detour factors for trips up to 2.5 km crow flies distance can be considered valid. Trips longer than this are affected more by the catchments of 5km max journey length



and have therefore not been presented.

- iv. The research indicates that a Detour Factor of 1.4-1.5 seems to be a more appropriate figure than the 1.2 currently used.
- v. However, the case is not as clear cut as that because of a number of other factors influencing the measure of accessibility:

- A) Origins are representative centroids of COAs aggregating a number of addresses. This means that depending on the destination location the detour factor may vary for the different ‘inhabitants’ of the zone. Depending on whether we want the minimum, maximum or plain average of journey lengths for the zone inhabitants one can argue a lower detour factor should be applied. This is a very complex problem and it is not in the scope of this study to cover this in any more details.
- B) Destinations are in most cases actual geo-coded addresses (Schools, GPs, Hospitals and Shopping) with the exception of Employment which is based on LSOA centroids. This mix also complicates a clear recommendation.



- C) Interchanges between stops, would probably on average be better represented by a detour factor of 1.5 because these are based on locations and connect to other locations (such as road nodes and other stops). There could be an argument that interchanges should have a separate factor all together in the calculations (also not covered in this research).

- D) The connection creation method is different for Road calculations and PT calculations which mean that the effect of changing the detour factor is also going have a different impact on the two calculation types. Connections to road nodes are more 'realistic' than the connections to PT stops.



B. Appendix B: Continuous Cut Off Methods

The continuous calculations provide a score for each origin based upon the destinations it can reach. The destinations that are considered for each origin need to be carefully controlled to prevent the result being unduly influenced by using large numbers of destinations. This can be controlled by limiting the destinations considered for each origin. This is done by using two techniques - the Cumulative cut off and the Threshold Value cut off. The destinations that are cut off by BOTH of these techniques are not considered for the origin in question.

Cumulative Cut off

- 1) Compute the Hansen indicator for each OD pair for the Origin in question
- 2) Rank the OD pairs and result by the indicator result in ascending order.
- 3) Compute the cumulative sum of the indicator scores for each OD pair in turn.
- 4) Compute the total of the indicator scores
- 5) Compute the cumulative cut off value by multiplying the total indicator score by a specified percentage value e.g. 0.05
- 6) Flag the OD pairs as 'potential exclusions' where its cumulative value is less than the cut off value.

Threshold Value Cut off

- 1) Using the total indicator score computed in step 4 of the cumulative cut off, compute the threshold value by multiplying the total indicator score by a specified percentage value e.g. 0.03
- 2) Flag the OD pairs as 'potential exclusions' where its threshold value is less than the cut off value.

Combination

Exclude any destinations for the origin in question where the OD pair is flagged for exclusion by both the cumulative cut off and the threshold cut off. The tables below show two worked examples of the cut off process. Note that the records in red and annotated 'Yes' in the 'Eliminated' column are those that are not considered as part of the continuous calculation.



Table B1. Two origins, four destinations

Origin	Dest	Time	Hansen Indicator	Cumulative Total	Total Indicator Value	Cumulative Cut Off Value (5%)	Potential Exclusion from Cumulative Cut Off	Threshold Cut Off Value (3%)	Potential Exclusion from Threshold Value Cut Off	Eliminated
1	4	160	0.106459	0.1064585	2.167604214	0.108380211	y	0.065028126	n	
1	3	60	0.431711	0.53816903			n		n	
1	2	30	0.657047	1.19521585			n		n	
1	1	2	0.972388	2.16760421			n		n	
2	1	1000000000	0	0	1.377945636	0.068897282	y	0.041338369	y	Yes
2	2	1000000000	0	0			y		y	Yes
2	3	45	0.532592	0.5325918			n		n	
2	4	12	0.845354	1.37794564			n		n	
O	Sum of included dests	Denom	Rel hansen							
1	2.061146	4	0.515286							
2	1.377946	2	0.688973							

Table B2 One origin, fifteen destinations

Origin	Dest	Time	Hansen Indicator	Cumulative Total	Total Indicator Value	Cumulative Cut Off Value (5%)	Potential Exclusion from Cumulative Cut Off	Threshold Cut Off Value (3%)	Potential Exclusion from Threshold Value Cut Off	Eliminated
1	1	1000000000	0	0	3.751931603	0.18759658	y	0.112557948	y	Yes
1	2	1000000000	0	0			y		y	Yes
1	3	1000000000	0	0			y		y	Yes
1	4	1000000000	0	0			y		y	Yes
1	5	350	0.007447	0.00744658			y		y	Yes
1	6	300	0.014996	0.02244216			y		y	Yes
1	7	250	0.030197	0.05263954			y		y	Yes
1	8	200	0.06081	0.11344961			y		y	Yes
1	9	170	0.092551	0.20600018			n		y	
1	10	100	0.246597	0.45259715			n		n	
1	11	50	0.496585	0.94918245			n		n	
1	12	30	0.657047	1.60622927			n		n	
1	13	27	0.685231	2.29145977			n		n	
1	14	25	0.704688	2.99614786			n		n	
1	15	20	0.755784	3.7519316			n		n	
O	Sum of included dests	Denom	Rel hansen							
1	3.638482	7	0.519783							

C. Appendix C: Methods And Improvements Made Available To The Wider ACCESSION User Community

ACCESSION has seen new methods and improvements on existing methods through out the National Accessibility Indicators project. This appendix aims to highlight these new features.

National Accessibility Indicator functions

ACCESSION has seen enhancements needed to produce the National Accessibility Indicators. To allow ACCESSION users to reproduce these calculations, new settings have been added.

The settings set introduce:

- Higher Straight Line Walk Distance Factor of 1.4 based on research on short journeys by road networks;
- Sampling interval lowered to 5 minutes to create result with the right granularity;
- Connection distance defaults to 2km; and
- An interchange distance of 0.5km.

The setting object should be used in conjunction with the new No First Wait Time option on the Network Accessibility form to produce results that are comparable to the results used as a basis for the National Indicator reports.

ATCO CIF importers (ACCESSION version 1.2 and 1.3)

The ATCO CIF import has been improved to increase speed and accuracy.

Speed

We have improved the way we build PT trips from its PT stops during the import. This has resulted in a much faster import process.

Accuracy

We have improved the way that Accession deals with stops that have duplicate x,y coordinates. If this is detected during an import the stop will only be stored once. The main stop reference will be the reference of the initial stop. Subsequent duplicate stops will have their names and references will be stored as 'Aliases'. These can be seen when examining stop details. The result of this change means that some trips will have less missing stops after an import. Added warnings on missing coordinate data.

Calculations (ACCESSION version 1.2)

The main network algorithm has undergone some changes in order to make accession faster, more efficient and more accurate, whilst providing more user feedback.

Lower memory usage / Faster algorithm

We have lowered the amount of memory usage during calculations. This makes accession run faster, particularly when using a computer that has limited RAM.

New optional connection method for more realistic interchanges

We have provided a new optional connection method when we construct the network for the algorithm. This allows greater control over the make up of the network. Previous versions of accession allow the user to specify a 'Network Connection



Distance'. This dictates the maximum distance across which origins, destinations, road nodes and PT stops are joined to provide viable journey paths. We have now added another option to this to allow the specification of an 'interchange distance'. This allows the user to set the distance a traveller can traverse between PT stops – the 'interchanges'. The advantages of this are threefold:

- The journeys more closely reflect reality as travellers are unlikely to walk far between stops when changing services
- Use of this parameter reduces 'stop hopping', where a traveller can walk large distances PT stop to PT stop, and thereby reach a destination provided that there are PT stops closer together than the network connection distance. Introducing an interchange distance allows the user to specify a larger distance to join and leave the network, then only traverse PT stops that are close enough together.
- Reducing the distances also reduces the number of links in the network. This in turn reduces the memory usage and speeds up the algorithm.

Please note that this may increase the number of 'inaccessible' origins, or increase travel times.

More realistic road connections

We have altered the method through which road 'nodes' are connected to origins, destinations or PT Stops. The previous method joined road nodes to all origins / destinations / PT stops that were within the network connection distance. Accession now joins these road nodes more intelligently by only joining to the nearest road node within each of 8 45 degree segments surrounding the origin / destination / PT stop in question (within the network connection distance).

This means that there is now a maximum of 8 road nodes per origin / destination / PT stop, resulting in smaller networks. This in turn lowers the memory usage and speeds up Accession. See the diagram in the help file for more details.

Sampling

We now allow control over the 'sampling' interval used by Accession to control the assessment of public transport services in the algorithm.

In order to model travel time by Public Transport, Accession models the best available options at set intervals (e.g. every 10 minutes) throughout the period being assessed. This has the benefit of reducing the amount of data required by the algorithm to compute all journeys. The downside to this is that the wait time for any given journey can vary by anything up to the sampling interval depending upon the time of arrival at the PT stop.

Versions 1.0 to 1.1.7 all use a default setting of 10 minutes as a non-editable sampling interval. Version 1.2 represents a further performance improvement in the algorithm and as a result it is now more appropriate for the user to be able to determine the sampling interval they wish to use. Valid values for this interval are between 1 and 10. For maximum accuracy, the sampling interval can be set to 1 (minute).

This will result in possible journeys being modelled every minute. It should be noted however that this will have a dramatic effect on network size in memory and calculation time. Alternatively the value can be left at 10 to retain the current performance. Values in-between e.g. 5 minutes can also be used in order to strike a balance between the two. Note that the lower the sampling interval, the lower the potential travel time between Origin and Destination. This affects PT calculations



only. The sampling interval can be altered within settings. To do this the setting set will need to be the new set as the default setting set cannot be edited.

More Feedback

We have added an additional feedback form to show you progress during a calculation. This should make it easier to plan your work and know when to expect a calculation to be finished.

PT Editing Tools

Remove all unused PT stops from the repository (made available in version 1.3 of ACCESSION);

Strict Clipping tool to remove all services and stops that do not fall strictly within a specified area (to be made available in the next commercial version, ACCESSION 1.4);

