Assessment and management of ‘at risk’ populations using a novel GIS based UK population database tool

H. Balmforth¹, H. McManus¹ & A. Fowler²

¹Health and Safety Laboratory, Buxton, U.K.
²Health and Safety Executive, Bootle, Merseyside, U.K.

Abstract

This paper presents a study of how a novel tool for identifying and estimating population levels, developed for the UK’s Health and Safety Executive (HSE), has been used for risk assessment and management. The HSE’s National Population Database (NPD) tool provides population data from a local to a national level for England, Scotland and Wales. It is a Geographical Information System (GIS) based tool, which uses a number of different source data sets to produce detailed estimates of a variety of populations. The data sets incorporated in the NPD include tables from the UK census, national mapping on a range of scales, address and transport network data and commercial and business data sets. The methodology developed for the tool involves identifying and locating a range of features, including building types, infrastructure and land use areas, to which population multipliers are applied to produce estimates of the different populations. The population multipliers provide differentiation between population levels at different times of day and between populations of different sensitivity. The NPD was initially developed for deriving populations for major accident hazard modelling on a wide range of scales. It can provide population estimates on a 100m grid or on an individual building basis, for some population types. The populations that are available from the tool include residential, workplace, retail, transport system, leisure and sensitive populations, such as schools, hospitals, care homes or prisons. The tool has been used for a range of tasks including providing estimates of the population around hazardous installations for societal risk calculations and identification of at risk populations and scenario planning. This paper will discuss the NPD tool and present a selection of case studies to illustrate how this new tool has aided the whole risk assessment process from risk policy and development to regulatory and enforcement duties.

Keywords: risk assessment, GIS databases, population estimation, societal risk.
1 Introduction

The UK Health and Safety Commission (HSC) and the Health and Safety Executive (HSE) are the bodies responsible for regulating most of the risks to health and safety arising from work activity in Britain. HSE has the day-to-day responsibility for enforcing health and safety legislation and its regulatory concerns range from nuclear installations and chemical plants, through to mines, factories, farms and many other workplaces.

The Health and Safety Laboratory (HSL) provide scientific support and research for these duties, which is an agency of HSE. One area of expertise supplied by HSL relates to risk assessment and the methodologies and techniques associated with it. Risk assessment research at HSL includes the development of Quantified Risk Assessment (QRA) methodologies and tools and the integration of these tools with Geographical Information Systems (GIS) and databases. A GIS is a set of software tools, which can be used to input, store, analyse, manipulate and display data with a spatial component, it particularly leads itself to integration with QRA tools and population estimation.

A risk assessment is the estimation of levels of risk and the evaluation of the significance of that risk in order to inform decisions about risk management. The risk is to people's health and safety and is expressed in terms of the likelihood of a certain level of harm being suffered. A risk assessment may be either qualitative or quantitative and may provide an indication of individual risk or societal risk, depending on the circumstances and the issues of concern. HSE’s risk assessments are used to examine, amongst other things, the risks to the public associated with the chemical process industries, the bulk transport of toxic and flammable chemicals [1] and to help inform land use planning decisions around major hazard sites [2].

To help inform HSE’s societal risk policy development and the use of societal risk measures, it has become essential that HSE has access to estimates of the population that potentially may be at risk from a major hazard event [3]. This is of particular importance when societal risk is of concern rather than individual risk. With this in mind HSE commissioned the development of the National Population Database (NPD) tool. The NPD tool extends a methodology developed by Staffordshire University [4], [5] to produce a sophisticated tool for deriving population data, from a local to national level for England, Scotland and Wales. The NPD is a GIS based tool that incorporates multiple data sets to provide estimates of a range of populations. It has been developed by Staffordshire University for HSE, who own the tool, and it is hosted and run by HSL.

2 NPD tool methodology

The modelling of major hazard accident events requires estimates of the populations potentially at risk from these types of incidents to be made. In the past, the data HSE has used to make estimates of the spatial distribution of the populations at risk has been generalised and of varying quality.
2.1 NPD structure

The recent development of the NPD has allowed a robust and auditable methodology for population estimation to be produced. The NPD is a GIS based tool, which uses the ESRI ArcGIS software as a platform. The NPD uses a number of different source data sets to produce the estimates of populations. The methodology behind the NPD involves identifying and locating a range of features, including building types, infrastructure and land use areas, to which population multipliers are then applied to produce estimates of the different populations. The population multipliers aim to provide differentiation between population levels at different times of day and between populations of different sensitivity. The data sets used include tables from the 2001 UK census, a range of UK Ordnance Survey (OS) digital mapping and addressing products, plus other commercial datasets. The majority of the locational data used within the NPD come from the OS. Most importantly this includes the OS AdressPoint data set, which is a high accuracy spatially referenced dataset listing all addressed locations in Great Britain [6]. This provides a national grid point, a unique reference and metadata details for each postal address in Great Britain. The 2001 UK census provides the main source of the population data within the NPD. The 2001 census has been produced to a new geographical framework based on postcode geography and includes employment data for the first time [7]. Of particular interest for the NPD development are census tables detailing population age structure, residential, workplace and daytime populations and also household types and average household size.

Table 1: Content and structure of the NPD.

<table>
<thead>
<tr>
<th>Feature Dataset</th>
<th>Layer</th>
<th>Differentiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residential</td>
<td>Residential</td>
<td>Usual or Night Time</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Daytime Term Time / Non-Term Time</td>
</tr>
<tr>
<td>Transport</td>
<td>Roads (major)</td>
<td>Average Daily Flow, Peak Flow, Maximum Capacity</td>
</tr>
<tr>
<td></td>
<td>Railway Stations</td>
<td>Location only</td>
</tr>
<tr>
<td></td>
<td>Ports</td>
<td>Location only</td>
</tr>
<tr>
<td></td>
<td>Airports</td>
<td>Location only</td>
</tr>
<tr>
<td>Sensitive and Communal</td>
<td>Schools</td>
<td>Daytime</td>
</tr>
<tr>
<td>Establishments</td>
<td>Boarding Schools</td>
<td>Night time</td>
</tr>
<tr>
<td></td>
<td>Care Homes</td>
<td>Maximum Capacity</td>
</tr>
<tr>
<td></td>
<td>Hospitals</td>
<td>Maximum Capacity</td>
</tr>
<tr>
<td></td>
<td>Prisons</td>
<td>Maximum Capacity</td>
</tr>
<tr>
<td>Workplace</td>
<td>Workplace Populations</td>
<td>Total Workplace Population</td>
</tr>
<tr>
<td>Retail</td>
<td>Retail Populations</td>
<td>Core Retail Centre, Town Centre, Retail Park</td>
</tr>
<tr>
<td>Leisure Facilities</td>
<td>Stadia</td>
<td>Maximum Capacity</td>
</tr>
<tr>
<td></td>
<td>Camp Sites</td>
<td>(Location only) Camp Sites, Caravan Sites</td>
</tr>
<tr>
<td></td>
<td>Public Attractions</td>
<td>(Location only) Aquarium, Historic House, Motor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Racing Circuit, Racecourse, Theme Park, Wildlife</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Centre, Zoo</td>
</tr>
</tbody>
</table>
2.2 NPD populations

Taking account of these newly available datasets the NPD has enabled populations located within residential, workplace, retail, transport, leisure and communal establishments, particularly sensitive populations, to be derived. Sensitive populations are of particular concern to HSE, as these populations may include people more susceptible to harm, such as school children and people in hospitals or care homes, or people that may be hard to evacuate such as prison populations. A summary of the contents and structure of the NPD populations is shown in Table 1.

The population estimations are available on two different scales, either on a 100m by 100m grid, or on an individual point location basis for some population types. The 100m grid provides populations generalised to the grid centre point and is generally used for larger areas of interest. The individual point location populations represent features or populations at their actual location, which is usually to 1 metre accuracy, although this depends on the source data set being used. Further details of the NPD and a full discussion of the data sets used to create it can be found in the HSE Research Report number 297, available on the HSE website [8].

2.3 Assumptions and testing

The NPD is based on a sophisticated methodology for deriving population estimates, however the database should still be approached as a representation of patterns of potential occupations, rather than a precise measurement of populations. The database was initially developed to represent where people might be at any time when a major hazard accident might occur. This includes trying to represent potentially large concentrations of people that may build up at certain times of the day in locations that at other times would be almost empty. A national database prohibits reliance on locally, intensively collected data, however the NPD does allow a user to supplement or substitute data with local knowledge. This enables the NPD output to be tailored for certain scenarios.

To test the accuracy of the NPD population estimates and to provide confidence in the NPD results, a large-scale ground truthing exercise took place on a sample area during the NPD development. A number of sensitivity tests were also undertaken and measures to avoid double counting were adopted. Further details of the assumptions and the testing regime can be found in the HSE Research Report number 297 [8].

3 Population estimation case studies using the NPD tool

3.1 Case study 1 – populations around major hazard sites

HSE is required to provide advice to Local Authorities on land use planning proposals in the vicinity of major hazard sites. This includes developments in the vicinity of existing installations and also the siting of new major hazard plants.
HSE sets consultation distances (CD) around hazardous sites, within which it consults on land use planning issues.

The NPD tool has been used to examine the numbers and types of population around major hazard sites. For this case study the CD from a site has been digitised and added into the NPD in order to calculate population numbers within this zone. The residential population associated with the individual location points (Residential AddressPoint layer) has been used to provide high-resolution population data within the CD. The nighttime occupation figure for every house within the CD has been used to give a worst-case scenario.

![Figure 1: Case study illustrating the structure of the population within the consultation distance of a major hazard site.](image)

The sensitive populations layer within the NPD has been used to give a further breakdown of the type of populations within the CD. The schools data provides the number and maximum population of each of the schools within the CD. The sensitive layer also provides information related to the number and population of hospitals (number of beds). Within the current CD there are no hospitals, however adding the sensitive layer into the map window allows the user to quickly identify the proximity of any hospitals in the area surrounding the CD. In this case the CD would not have to be increased by a large amount for the populations affected to include 2 hospitals and many more schools. This would considerably increase the sensitive population within the CD.

Details of the retail population is available from the NPD at 100m grid level, this layer identifies areas classed as retail, such as town centres, retail cores, small town centres etc. The layer only populates areas classed as larger retail areas. These are expected to draw visitors in from a wide area, rather than the
immediate local vicinity, this is an attempt to avoid double counting local populations. The smaller retail areas are marked in the layer but have no population data associated with them. The pink dots in Figure 1 show the locations of retail areas and the larger orange dots show those areas that have been populated. This can be used to provide an estimation of non-local populations that may be present and also outside in an area, which may be of concern for emergency planning and evacuation considerations.

![Scenario planning using the NPD tool.](image)

**Figure 2:** Scenario planning using the NPD tool.

### 3.2 Case study 2 – scenario planning

The NPD has been used to gain accurate estimates of population figures that may be affected by a release of a hazardous substance. In this case study plume dimensions have been calculated using the PHAST consequence modelling...
software package and then inserted into the NPD tool. The plume can be rotated to take into account different wind directions, and moved around the edge of the major hazard site to cover the maximum number of residential points and other sensitive populations, in order to give a worst-case estimate of the populations likely to be affected.

Figure 2 clearly shows the location of a school (red dot) within the plume. The NPD layer for schools gives a figure of 202 for the population of this school. This sensitive population can be used alongside the residential population to build a more accurate picture of the populations likely to be affected by an incident at this site.

3.3 Case study 3 – consequence analysis

The NPD has also been used for consequence analysis and in this case study has been used to examine the consequences of 3 different hazardous substance release scenarios. The map in Figure 3 shows the hazard plumes associated with a large hole in a chlorine tank and the catastrophic failure of a 20te chlorine road tanker. The plumes relate to the distance that could be covered by the release due to the hole in a tank and the failure of a road tanker, both of which are on site. The sizes of the plumes are calculated according to the type and quantity of substance stored and varying harm criteria. The plumes can be added into the NPD and rotated according to wind direction or to find the location and orientation affecting the largest numbers of population.

![Figure 3: Population estimation for consequence analysis.](image_url)
Also shown on the map in figure 3 are the individual risk contours around the site, these are set by HSE and used for land use planning advice. The risk contours are constructed by considering the likelihood of a consequence happening; the outer contour, shown in black is the CD for the site. This example illustrates that certain events can have an effect on populations beyond the CD for the site, although the likelihood of these events happening have not been included in the plume calculations.

This example also shows the NPD transport layer, with data for roads shown by the green dots. This layer provides population data for the network of larger roads (motorways, and dual and single carriageway A roads) and gives figures for average daily flow, peak flow (i.e. rush hour traffic) and bumper to bumper traffic scenarios. By using the NPD the populations that may be affected beyond the CD can be identified and estimated for certain scenarios.

Figure 4: Societal risk estimation using populations from the NPD.

3.4 Case study 4 – societal risk estimation

The NPD has also been used to examine potential high consequence, low frequency incidents to help examine the contribution these incidents make to societal risk estimations. In these cases the NPD output has been further girded and inserted into an additional software programme to calculate societal risk FN curves. The population figures were obtained using the NPD residential layers in the localised point or 100m grid format and then extracted in a secondary grid format, imported into an Excel table and input into the societal risk software to obtain FN curves. The resolution of the NPD data has been adapted in this way.
as a very broad picture of the population around a site was required. This is the case when considering scenarios involving a hazardous substance release likely to travel a large distance away from the site. The example above shows a grid covering an 8km radius from the site.

4 Future developments

Future developments of the NPD will be largely governed by the availability of new data sets and by the usage and requirements of the tool by HSE and the wider community. Future use of the NPD could include other point source risk studies, for example with nuclear and environmental risk assessment. It could also be used for linear risk study work, such as that associated with pipelines or transport routes. There is also potential to use the NPD for larger scale risk studies and to examine the impact of health and safety policy on populations. The wide range of data sets used within the NPD will have to be examined carefully when considering updating and maintenance for the tool. Due to the complex nature of the tool, it is important that updating the data sets and also adding in new layer has a proven benefit. This will be determined by the current use of the tool and also any future uses. The task and benefits of updating the tool have to be weighed up against the fact that the tool has to be kept current and relevant to be of full use.

The contents of this paper are © British Crown Copyright.

References