J9.80142 THE GEOGRAPHIC AND ENVIRONMENTAL DATABASE INFORMATION SYSTEM (GEDIS) AS A TOOL FOR URBAN DISPERSION MODELING

Richard N. Fry, Jr.
Defense Threat Reduction Agency, Fort Belvoir, Virginia

Ian Griffiths*
Dstl Porton Down, Salisbury, UK

Iain Crawford, Tim Dudman, Matthew Gilbert and David Stewart
RiskAware Ltd, Bristol, UK

1. INTRODUCTION

The Defense Threat Reduction Agency (DTRA) supports US Department of Defense (DoD) counterproliferation and counterforce operations and homeland defense through development of a wide variety of chemical, biological, radiological and nuclear (CBRN) hazard assessment tools. The increasing possibility that military operations or terrorist activities will occur in urban areas has led DTRA to develop enhancements to its widely used Hazard Prediction and Assessment Capability (HPAC) suite of models to enable the modeling of CBRN events within the urban environment. The latest HPAC release includes the UK Ministry of Defence's (MOD) Urban Dispersion Model (UDM), developed by the Defence Science and Technology Laboratory (Dstl), and the Urban Windfield Module (UWM), developed by the Titan Corporation. This "Urban HPAC" was used successfully as part of DTRA's support to the Athens, Greece, Summer Olympic Games, the Boston Democratic National Convention, and the New York City Republican National Convention.

Of critical importance to these modeling tools are the urban building and morphology data. Data can come from a myriad of sources and in varying formats, resolutions, and sizes. The processing, archival, interpretation and display of these data are a complex and necessary task.

The DTRA Urban Modeling Program has been working with the UK MOD / Dstl in developing an urban Geographic and Environmental Database Information System (GEDIS). GEDIS can be used as a tool for rapidly accessing large urban datasets, which may include many millions of buildings. It contains several applications, such as an automated urban morphology calculation, automated quality assurance, and allows importing additional urban datasets.

GEDIS provides the backbone data structure for the HPAC urban modeling capability. DTRA is actively acquiring building data for any and all major metropolitan areas so that urban atmospheric transport calculations can be provided for any location anywhere in the world. By integrating these datasets into GEDIS, HPAC will potentially have access to building datasets for several hundred cities worldwide.

2. GEDIS

GEDIS is a database system that stores spatial and environmental information for use by a range of hazard assessment and prediction models. For the benefit of various dispersion models, for example, it stores building information for many cities around the world. The data is stored in such a way that it is quickly and efficiently retrievable, even though there may be data for hundreds of thousands or even millions of buildings.

GEDIS provides a range of functionality, including:

- The import and export of a variety of data with some automated quality assurance.
- The storage and rapid access to buildings and urban areas, also with editing capabilities.
- The storage and access to imagery data and also gridded data, including population, terrain and land usage.

---

* *Corresponding author addresses*: Ian Griffiths, Dstl Porton Down, Salsbury, SP4 0JQ, UK; email: ihgriffiths@taz.dstl.gov.uk.

- The calculation of statistical properties of user-specified urban regions.
- The ability to provide a graphical representation of the data for use in external systems, e.g. as an urban overlay in a geographic information system (GIS).

Further details on these are given later in this paper.

GEDIS is used in multiple applications. It has been developed primarily for DTRA's HPAC and for the UK MOD's IMPACT CBRN modeling engine. It is being developed by the Dstl, which is part of the UK MOD. Currently the applications using GEDIS are:

- DTRA's HPAC
  Within HPAC GEDIS is used at run time to provide urban data to the Urban Dispersion Model (UDM) and area density information to the Urban Wind Module. It also supplies high-resolution terrain, roughness and canopy height data to the other parts of the HPAC calculation engine and supplies an urban overlay for display in the HPAC Java graphical user interface (GUI). The Micro-SWIFT Spray (MSS) model also uses GEDIS to provide it with ShapeFiles as an input for its urban wind flow calculations.
- Dstl's IMPACT
  IMPACT is a CBRN hazard modeling engine being developed for use in the UK MOD's prototype NBC Battlespace Information System Application (BISA) warning and reporting system. IMPACT is also used for modeling and simulation exercises and training. IMPACT obtains its geographical information for its modeling components through GEDIS.

- Other Dstl systems and tools
  GEDIS is also used in the CB Virtual Battlespace operational research tool, being developed for the UK MOD and DTRA (under the name REASON), as well as the UK Sensor Placement Tool.

2.1 *Components and Architecture*

GEDIS has a modular design as shown in figure 1. There are a number of components within the GEDIS dynamic linked library (DLL) and more in the interface layer. Some of these will be described in greater detail below. Applications can call the individual components within the interface layer (ShapeFile import/export, data import/export, Urban Morphology Extractor, graphic overlay provider and the Urban Database Calculator) or

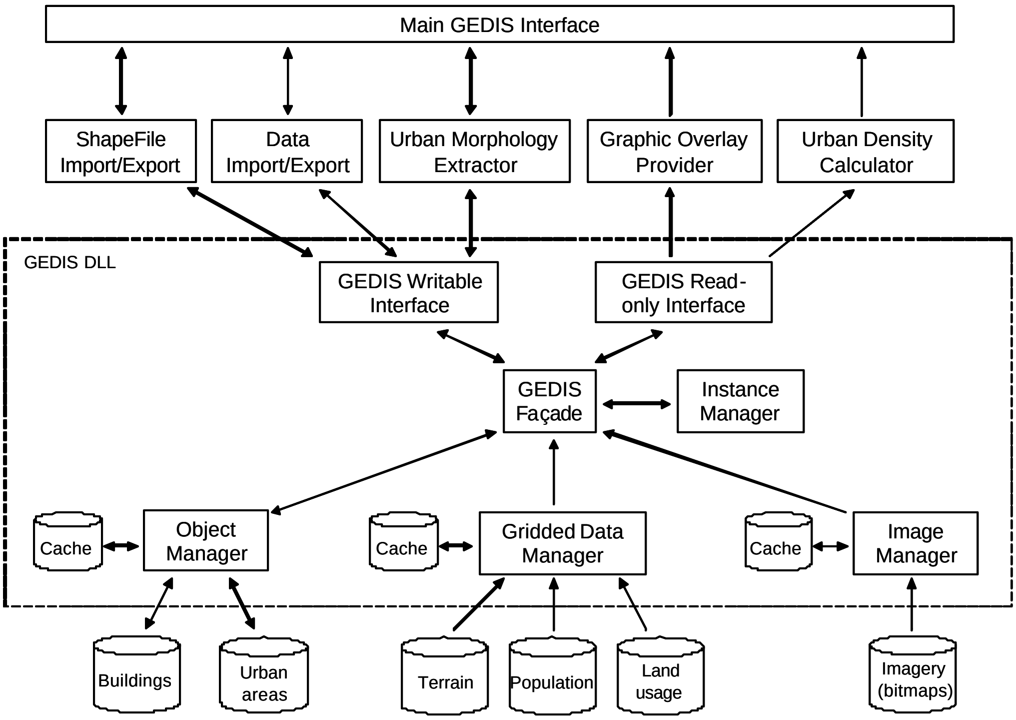

Figure 1: GEDIS architecture overview and components

use the single main GEDIS interface, which is being developed to bundle up all of the GEDIS functionality. GEDIS is written in ANSI C++ and, although currently a Windows application, it could be modified to work on other operating systems.

2.2 *Data Storage and Access*

GEDIS stores a variety of data, including buildings and urban areas, grids of terrain heights, land usage and population, and imagery. These are all accessed by specially developed managers that are optimized for the fastest access to the specific types of data and the sorts of calls made for that data. Each of these managers has an internal memory cache for fast data access. This is supported by the Instance Manager, which controls access to the individual GEDIS databases enabling more efficient caching.

a) Building data and urban area data

GEDIS has the ability to import buildings from data in ESRI ShapeFile format, which is a commonly used format. It is also possible to enter data and view the results using the GEDIS database viewer and editor. Positional data can be stored in either UTM or geodetic format. At the moment a building is stored a closed footprint with a single height value that represents a constant height for the obstacle. It is realized that specific models require varying degrees of individual building detail. Higher fidelity models will require information on the composition of component shapes of varying height to be retained whereas others will require these to be merged to a single obstacle. Some may require the removal of inner spaces such as courtyards while others may need internal structure data such as rooms. Although not yet fully implemented, GEDIS will use a Level of Detail (LOD) hierarchical model in order to store building information in all the versions required by its client applications.

GEDIS stores data on ground areas, each of which is an area that has certain generic properties in common. These can be basic ground areas describing, an area of water, desert, park land, forest, etc. or they can have additional properties for Urban Areas such as average width/length based binned on direction, mean building height etc. Each of these land-usage areas thus contains environmental information that will affect the dispersion of plumes passing over it, for example. These are use extensively by the UDM. Each ground area should have the following attributes defined:

- Albedo for each season
- Bowen ratio for each season
- Canopy height for each season
- Displacement for each season
- Surface roughness, z_0, for each season
- Canopy attenuation coefficient for each season

If the ground area has a building density occupancy of greater than 5%, the area is deemed to be an urban area and it must also additionally have the following values defined:

- Building occupancy density
- Mean building height and height variability
- Mean building width. This value is recorded with respect to 12 compass points.
- Mean building length. This value is recorded with respect to 12 compass points.
- The number of street alignment angles and these angles
- The surface roughness, z_0, owing to the buildings present. This value is recorded with respect to 12 compass points if known for each. (Note, the surface roughness for the ground area above is still provided but this represents the z_0 for the space between the buildings.)

GEDIS database objects have their attribute information split between index data and object data in order to optimize data searching and retrieval performance. The index data is much faster to search then the object data. The data is stored internally within zones and sub zones to again speed up data access, and each sub zone that is requested is then also cached in order to dramatically speed up future searches. GEDIS can sort the retrieved data based on sort criteria passed to the interface. This sort criterion could be any object property, so for example GEDIS could sort all the buildings in a specified region in height order. Similarly the retrieved data can be filtered based on any object property. For example GEDIS could return all those buildings with a footprint area greater than a specified value. Objects can be requested by refining a search on a retrieved dataset (from a previous request), by asking for them by their GEDIS identification keys or by their proximity to a specific location.

The ability to add multiple objects and delete multiple objects is available from an editing interface. Any changes made to the database are not committed until the appropriate API function to save the edits has been called. During the save process if any problems are encountered, GEDIS will roll back all the changes made so that the database cannot become corrupted. (Even if the process is terminated half way through the save, the next time the database is opened it will finish the roll back.) Where there has been no call to save changes, the changed data will be available in memory until the database is closed, after which the changes will be discarded.

Work is currently on-going to implement a more efficient data store format. Rather than partitioning the objects as described above, a spatial tree indexing system will be used. This will allow much faster searches for the obstacles, improving the obstacle access time by an estimated order of magnitude. It should also allow GEDIS to scale up to handle databases of tens of millions of buildings.

b) Gridded data

GEDIS currently handles terrain heights. The method for dealing with gridded data is currently under redevelopment with the aim of making it generic for all types of gridded data. This will make it very easy to either add the ability to access new types of gridded data or to obtain gridded data from new file formats. This Gridded Data Manager (GDM) will also have an independent cache for each of the object types enabling much faster data re-access times. The GDM will return values at individual or specified grids of points on the Earth from the underlying data using the interpolation strategy selected by the calling application.

GEDIS already supports various different types of terrain data and selects the data with the highest resolution from which to retrieve the data. The formats supported are 'dted0' to 'dted3', an HPAC specific format with the same resolution as dted0 and an internal bespoke format for storing data at any required resolution. Further formats can be easily supported if required. This functionality is being migrated to the emerging GDM. Also being added is the handling of gridded population and land usage data.

2.3 *Tools and Other Functionality*

GEDIS contains a range of additional functionality that can be either accessed directly through the relevant tool or through the main GEDIS interface. These include:

a) The Urban Density Calculator (UDC)

The UDC returns gridded data concerning database objects as follows:

- The UDC can return two and three dimensional gridded data concerning the density of the buildings within a specified spatial area. In the three dimensional case, for example, it indicates whether there is actually a building at the specified point on the Earth's surface and at a particular height above ground.

- Gridded ground area attributes, such as canopy height, surface roughness or other bulk properties. These are calculated from the underlying obstacle data where possible and the UDM urban areas discussed in section 2.3a) are not used. Other properties such as terrain height can also be accessed in this way as GEDIS often also has access to high resolution terrain data in regions where building data is available.

Currently the UDC is used by the Urban Wind Module and by the HPAC SCIPUFF dispersion model where data is available.

b) The Urban Morphology Extractor (UME)

The UDM requires a city to be split up into urban areas which are roughly homogeneous in terms of their urban properties. That is most of the buildings within the area are generally of a similar size and orientation, they are mostly aligned similarly, and the building occupancy density is roughly constant over the area. For a town or city of tens of thousands of buildings or even hundreds of thousands of buildings, subdividing the urban region into these distinct roughly homogeneous areas can be excessively time consuming and exceptionally difficult even if there were sufficient time. This is because of the number of differentiating parameters that need to be considered simultaneously. The UME is being developed to help automate this process.

The forerunner of the UME exists and was developed by Dstl for DTRA in order to obtain urban areas for Salt Lake City for use in DTRA's support for the 2002 Winter Olympics. It used a supervised classification algorithm with a simple but slightly inflexible area identification scheme, which was ideal for Salt Lake City with its regular grid pattern arranged closely with the cardinal compass directions. The algorithm works as follows:

(i) An initial grid is created for the entire region being considered. The resolution of this grid is calculated so that there are typically 200 or so buildings in each grid cell.

(ii) The initial grid cells are classified as either non-urban, if the number of obstacles in the cell and area density of the cell both fall under certain user specified thresholds, or urban otherwise. For urban cells the following properties are calculated: the building occupancy density, mean height, mean width, obstacle orientation and the number of identifiable axes of alignments (none, one or two) between the buildings such as would be created by streets and the angles of these axes of alignments. Once these properties have been calculated for a grid cell, its band values can be determined. The user can define for each property a number of bands and the range of property values each band represents. For example, for building occupancy density there may be four bands: 0-10%, 10-25%, 25-50%, 50-100%. Each cell will be assigned band values for each of the parameters being considered.

(iii) The next stage is an adaptive subdivision of the urban areas. Each urban cell will be adaptively subdivided into four quadrants if any of the following conditions is false for any one of the properties being considered (e.g. height, building occupancy density, etc): there is a low variability of the property for the obstacle in the cell, there is a low variability of the property for each of the quadrants, the average value for the property for the cell and the quadrants are roughly similar. For all those cells such that any of these conditions is false for any property, the cell will be subdivided into quadrants. This process can be repeated on the newly formed urban cells. In general it has been found that two repeats of subdivisions are appropriate. However, the tool allows further subdivisions of specified regions to be carried out.

(iv) Finally, cells (at any resolution) are merged when they lie adjacent to one another and if all the selected properties for the cells lie in the same bands.

This algorithm is extremely fast and can process the approximately 60,000 buildings in the Salt Lake City urban database in a couple of minutes. Figure 2 shows a sample output for Salt Lake City. It has been found not to deal as well with more complex city layouts, those that are less grid-like for example. This is to be expected given that the tool was targeted very much at Salt Lake City.

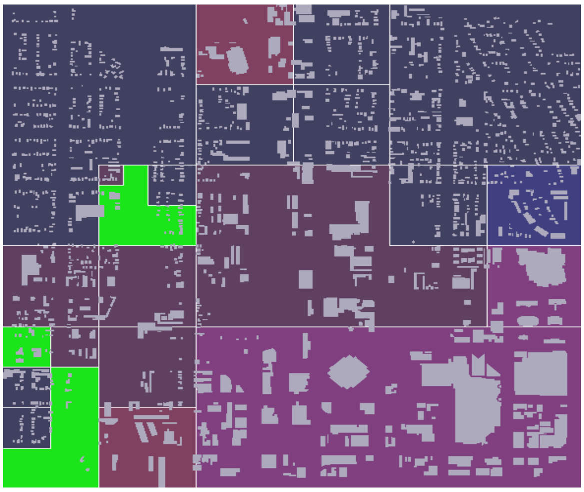

Figure 2: Ground and urban areas calculated for Salt Lake City

Work is now underway to redevelop the tool as the UME, which will be more generally applicable to cities of all types. There are both high resolution (area delineation) and low resolution (large areas are desirable) requirements for a successful algorithm for this problem. It would therefore seem that a region growing algorithm would be most appropriate. In essence the algorithm under development operates as follows:

The algorithm first generates small clusters or regions of buildings based on a variant of a nearest neighbour algorithm. The algorithm then examines each pair of neighbouring regions to assess their similarity for each classification property:

- A sorted list of heights of buildings within the region.
- Binned orientations from each building within the region.
- Binned mean effective widths from each building within the region.
- Street orientation.
- A sorted list of the building mean separation distance to its neighbors within the region.

If the statistical likelihood that the two regions actually constitute parts of a single merged region falls above a user-set threshold for each classification property, the two regions are deemed to be candidates for merging. The pair deemed to be statistically most similar is then merged and the region attributes updated. The algorithm continues to loop around this process until no two neighbouring regions meet the criteria for merging. After some tidying up, the urban areas are deduced from the final regions by using a convex hull drawing algorithm.

c) The Graphic Overlay Provider (GOP)

The GOP is a module written in OpenGL. It accesses the data held within GEDIS to create either two-dimensional or three dimensional representations of specified region. The data displayed can include buildings, urban and other ground areas, terrain and imagery, as well as other data types as they are added to GEDIS (population, land usage, etc). HPAC4.04 uses two-dimensional output from the GOP as the urban overlay for display within its GIS interface. Figure 3 shows a sample three-dimensional output of the GOP.

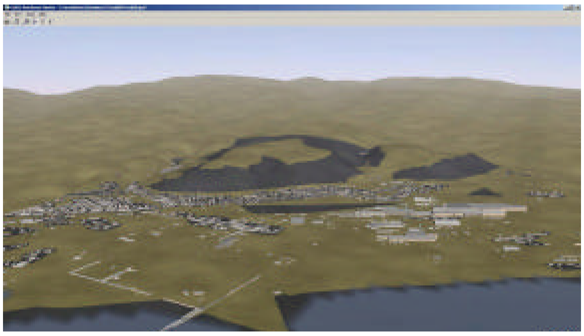

Figure 3: 3-D display of Port Talbot, South Wales, from the Graphical Overlay Provider module

3. DTRA'S URBAN DATABASES

DTRA relies on other US government agencies and their contractors as providers of urban data. Past sources of data have been the National Ground Intelligence Center (NGIC), the National Geospatial-Intelligence Agency (NGA), and the Army's Joint Precision Strike Demonstration – Rapid Terrain Visualization Program (RTV). These entities have obtained urban data through various means, such as stereoscopic satellite imagery and airborne LIDAR. Processing of the raw data has been conducted using products produced by the Harris Corporation (satellite data) and SAIC (LIDAR data).

NGA is conducting a collaborative effort with the US Geological Survey (USGS) called the National Map, which will collect data for approximately 133 major cities and metropolitan areas in the United States. For OCONUS applications, data is more difficult to produce. Stereoscopic satellite imagery is one method that has been employed when airborne LIDAR has not been feasible. Other commercial data vendors have also been explored, but their data has often been price-prohibitive. DTRA is not the lead federal agency for data collection and distribution, so large investments to obtain urban data are generally not defensible.

Once the processed data, in ESRI ShapeFile format, is obtained by DTRA, it can be further processed into GEDIS and added to the HPAC database. This database is envisioned to be accessible by any of the DTRA hazard modeling or consequence assessment tools. Future model additions to the HPAC suite, such as the SAIC Micro Swift/Spray and CFD Research Corp. CFD-Urban models, are being designed to take advantage of the GEDIS architecture.

4. SUMMARY

The development of GEDIS has significantly enhanced the DTRA urban modeling capability. Its ability to efficiently handle enormous sets of data will allow users to access and manipulate information on a global scale. The addition of automated data QC and QA has significantly simplified a once laborious and expensive process. And the streamlined method with which GEDIS accesses its data has drastically increased computation and graphical rendering.

GEDIS is a powerful urban data toolset and database system with a flexibility that will allow it to grow as modeling requirements change. New data types can be added to accommodate many other modeling systems. Linkages to other databases, such as the internal building modeling system COMIS data, will allow interoperability and user interface between models of different scales.