5.2 EMERGING URBAN DATABASES FOR METEOROLOGICAL AND DISPERSION MODELING

Steven J. Burian[1*], Michael J. Brown[2], Timothy N. McPherson[2], James Hartman[1], WooSuk Han[1], Indumathi Jeyachandran[1], and Johnathan F. Rush[2]

[1]Department of Civil & Environmental Engineering, University of Utah, Salt Lake City, Utah
[2]Systems Engineering and Integration Group, LANL, Los Alamos, NM

1. INTRODUCTION

Mesoscale meteorological codes and transport and dispersion models are increasingly being applied in urban areas. Representing urban terrain characteristics is critical for accurate predictions of air flow, heating and cooling, and airborne contaminant concentrations. We have developed a national database of urban footprints, 'tall building district' footprints, and building statistics and we continue to expand and refine these emerging urban databases. Urban footprints have been derived nationally using a Los Alamos National Laboratory-developed day-night population database. Thresholds in the population dataset that characterize urban areas were determined using urban footprints derived from high resolution regional landuse datasets. These thresholds were applied to the national coverage of day-night population to create a national urban footprint dataset. Tall building districts for the 46 largest metropolitan areas in the US (based on 2000 Census estimates) were digitized from digital orthophotos. In addition, methods are being developed to classify tall building districts using automated analysis of population data, satellite data (e.g., synthetic aperture radar), and airborne lidar and these approaches will be used in the future to derive a nationally consistent coverage for all metropolitan areas. Statistics of building height, geometry, and density (e.g., mean building height, plan area density, frontal area density, sky view factor) have been computed at 250-m resolution from three-dimensional digital building data for parts of 17 metropolitan areas in the US and are being extrapolated to other metropolitan areas. These three core datasets are being geo-referenced to a geographic coordinate system with the North American Datum 1983, rasterized to 250-m resolution, and packaged for use in typical mapping software (e.g., ArcGIS) for importation to modeling systems. This paper describes the current databases, on-going development, and future planned expansions and refinements.

2. BACKGROUND

The recent urbanization of numerical weather and dispersion models has introduced the problem of delineating the urban footprint and describing the urban extent with a set of representative geometric, radiation, thermodynamic, and surface cover parameters. These urban canopy parameters (UCP's) include aerodynamic roughness properties (e.g., roughness length), building height characteristics (e.g., mean height, standard deviation, histograms), building geometry characteristics (e.g., height-to-width ratio, wall-to-plan area ratio, complete aspect ratio), building volume characteristics (e.g. building plan and frontal area densities), radiation trapping parameters (e.g., sky view factor), surface cover properties (e.g., impervious surfaces, albedo), surface material properties (e.g., heat storage capacity, emissivity), vegetation type, height and geometry, and more. We are currently developing approaches to produce national urban databases containing much of this information. In this paper we focus on recent efforts to define at the national level the urban extent and building geometrical properties.

A handful of researchers over the years have pioneered the work on obtaining surface cover and morphological parameters for cities at the micro to neighborhood scale (10^2 to 10^4 m) (e.g., Ellefsen 1990/1991). Grimmond and Souch (1994) were among the first researchers to present a geographic information system (GIS)-based technique for representing surface cover and morphological characteristics of the urban terrain for urban climate studies. Cionco and Ellefsen (1998) and Ellefsen and Cionco (2002) expanded Ellefsen's (1990/1991) original morphological inventorying procedure using a higher resolution (100 m X 100 m) grid cell size (and then a 50 m X 50 m cell size) for use in a high resolution wind flow model and included more characteristics of urban canopy elements in the database.

Grimmond and Oke (1999) reviewed several methods to define aerodynamic characteristics of urban areas using morphometric approaches. The work compared several methods to determine the roughness length, displacement height, depth of roughness sublayer and aerodynamic conductance based on measures of building and tree morphology. GIS databases were developed for 11 sites in seven North American cities and were used to characterize the morphological characteristics of the terrain and, using the morphometric equations, the aerodynamic parameters.

Lack of processing automation limited the early approaches to very small areas of a select number of cities. With recent advances in data collection and management and improved processing capabilities including GIS and image processing tools large areas covered by 3D digital building and tree datasets can now be analyzed automatically to extract morphological information (e.g., building height and geometry characteristics, roughness length). Ratti and Richens

**Corresponding author address*: Steve Burian, 122 S. Central Campus Dr., Suite 104, Salt Lake City, UT 84112; E-mail: burian@eng.utah.edu

(1999), for example, built upon the initial effort of Richens (1997) to implement efficient urban terrain analysis algorithms in an image processing framework built within the MATLAB software package. Ratti et al. (2002) used the image processing approach to compute building plan and frontal area densities, distribution of heights, standard deviation, aerodynamic roughness length, and sky view factor for three European cities (London, Toulouse, and Berlin) and two US cities (Salt Lake City and Los Angeles). The results illustrated the roughness length differences between European and US cities.

Burian et al. (2002) presented an approach using GIS to process 3D building datasets to compute building height characteristics (mean, standard deviation, plan-area-weighted mean, histograms), plan area density, frontal area density, wall-to-plan area ratio, complete aspect ratio, height-to-width ratio, roughness length, and displacement height. This automated GIS approach was used to compute UCP's for Los Angeles, Phoenix, Salt Lake City, Portland, Albuquerque, Oklahoma City, Seattle, and Houston. The GIS approach has recently been expanded to include analysis of 3D vegetation, other 2D GIS datasets (e.g., roads) and multi-spectral imagery to compute an expanded set of parameters including surface cover fractions, impervious surfaces, sky view factor, predominant street orientation, and more (Burian et al. 2003).

Long et al. (2002) developed and tested the DFMap software to process vector building and vegetation data (BDTopo) available from the French National Geographic Institute (IGN). With DFMap, a user can select a cell size and wind direction to compute a series of morphometric and aerodynamic roughness parameters. Long (2003) used the DFMap software to compute morphological statistics and define urban land use/cover types using an unsupervised k-means analysis. The analysis tools and approach were tested using data for the city of Marseille. Long et al. (2003) extended the DFMap application by incorporating the analysis of multi-spectral and panchromatic imagery in an attempt to improve the definition of urban surface cover.

The recent activity in developing urban databases has focused on developing database city by city for specific modeling studies. Specifically, the focus has been the development of tools and techniques to do the processing. And although significant progress has been made and substantial datasets have been compiled, the coverage of urban data in the US remains difficult to obtain and is not in a standardized form available for easy input to a wide variety of modeling systems. This paper describes ongoing efforts to address this problem by deriving urban databases at the national level using a standardized approach and data format.

3. URBAN FOOTPRINTS

The first step in representing urban areas in numerical models is to define their land surface extent. We are developing a nationwide urban footprint database that defines the land surface extent of urban areas. Although other datasets exist which may be used to create urban footprints, these data are subject to error due to their currency and methodology of generation.

Previous footprint studies have used national land use/land cover datasets or nighttime satellite imagery to derive the footprints (Elvidge et al. 1998). The USGS has two complete national land use/land cover datasets, but the most recent one is based on 1992 Landsat imagery, and the extent of many urban areas has increased in the intervening years (Vogelmann et al. 2001). Another technique is to derive footprints by threshholding nighttime city data acquired by the Defense Meteorological Satellite Program's Operational Linescan System (Imhoff et al. 1997). This technique accurately captures most urban areas but also tends to create commission error due to the properties of the sensor on the satellite. The satellite sensor is easily saturated, and the presence of urban areas next to reflective surfaces such as snow or water will create a "blooming" effect, giving the impression that an urban area is larger than it really is. The sensors are also sensitive to temporary sources of light, such as fires and gas flares, and must be corrected. In this research, we are creating an urban footprint database that is built from a LANL developed population database (McPherson and Brown 2003). A footprint product derived from population density will not suffer the problems associated with misclassifications of satellite imagery or non-urban light sources. Furthermore given the population database is based on circa 2000 data, our footprints are more current than many of those derived from satellite imagery or aerial photography.

The urban footprint database is based on a population database developed using 2000 US Census data and 1999 State Business Directory (SBD) data. The population database was created by disaggregating census blockgroup and tract level data into a 250 meter resolution grid built from road data and business demographic data. Using business demographic data allows the working population to be redistributed to places of work during the daytime. The database's spatial accuracy is continually being enhanced with more ancillary data, and future versions will account for the population in school, commuting, or shopping. Figure 1 demonstrates the technique used to create the footprint database. A threshold of population density that corresponded to the presence of urban land cover as defined by high resolution local land use dataset was used to define urban areas. In the figure, red areas represent regions where the population per grid cell is greater 3000, the yellow areas represent regions where the population per grid cell is greater than 300, and green areas represent regions where the population per grid cell is greater than 30. The cross-hatched area is the urban footprint as defined by the local high resolution land use dataset. The plot indicates a reasonable fit to the regional land use dataset for a threshold of 300. This evaluation is based on a visual inspection only. In the future, we will conduct quantified error evaluation for each possible threshold at intervals of 5 people per grid cell.

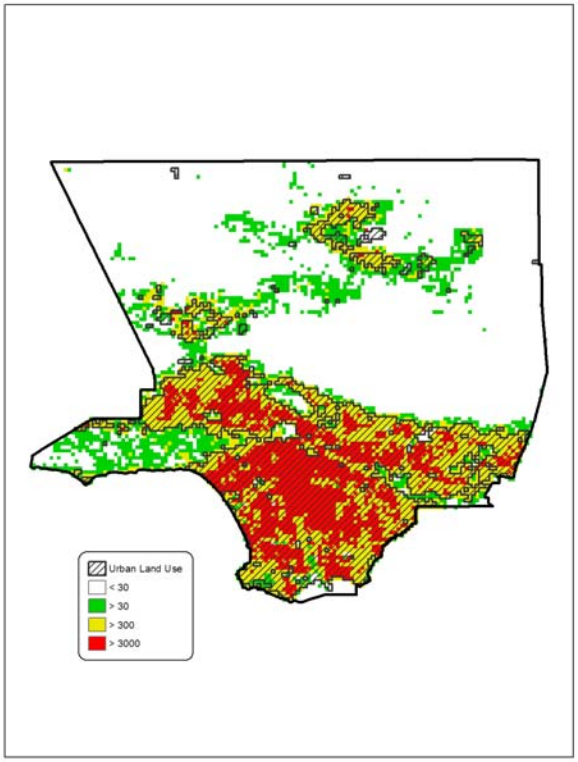

Figure 1. Population thresholds for delineating urban footprints.

The thresholds will be evaluated based the ratio of area to perimeter for given urban entities and the analysis of commission and omission errors with respect to the local land use/land cover obtained from cities and local governments.

Using a threshold of 300, we have created a preliminary national urban footprint dataset (see Figure 2). At the conference we will present quantified comparisons. In the future, we will also evaluate the need for regional scale thresholds due to the possible variability in population density relative to urbanization.

Figure 2. The National Urban Footprint Dataset.

4. TALL BUILDING DISTRICTS

Defining the urban footprint is an important step to differentiate the rural and urban areas and thus accurately represent the spatial locations of rural and urban processes in the models. Equally as important as the differentiation between rural and urban is the representation of the variability of land surface characteristics in the urban environment. A first order representation of the variability can be accomplished by differentiating the so-called tall building districts (TBD's) from the remainder of the urban area. The TBD's by definition consist of the concentration of taller buildings present in the central business districts of cities. The surface roughness and density of structures in these areas is unique compared to other parts of a city. The differences in land surface characteristics are significant for modeling and thus warrant delineation.

The first step in the development of a national database of TBD footprints was the selection of the initial set of metropolitan areas to include and the definition of the criteria used to delineate a TBD. We selected for inclusion in the initial database the 46 largest US metropolitan areas based on population in the 2000 US Census (see Figure 3). It was decided to delineate the initial TBD's using manual digitization of digital orthophotos. This provided a highly accurate approach for the initial set of TBD's, and later it also shall provide a verification dataset to use to validate more complex automated approaches that are currently being developed. The criteria selected to define a TBD were straightforward. The areas were objectively defined to include concentrations of buildings that appeared in the photo to be greater than five stories. The analyst identified the location of the concentration of tall buildings, and then identified the contiguous city blocks within the tall building area that contained at least one tall building defined as having at least five stories. The aggregation of all city blocks identified as containing a tall building in the concentrated zone were defined to be a TBD. The minimum size to constitute a TBD was 0.5 km^2. The first five TBD's defined in this fashion were checked against three dimensional building datasets (see Section 5 below) to verify the use of the visual observation of buildings greater than five stories as the criterion. Figure 2 displays the delineated TBD for Boston, MA.

5. NATIONAL BUILDING STATISTICS DATABASE

The use of the TBD's provides a first-order method to incorporate spatial variability of urban land surface characteristics into models. Clearly the tall building districts will have substantially different mean building heights, plan area fractions, sky view factors, fractions of impervious surfaces, and so on. To provide a more detailed representation of the spatial variability requires a database containing computed or estimated land surface characteristics either on a continuous basis or in discrete grid cells with a very high resolution.

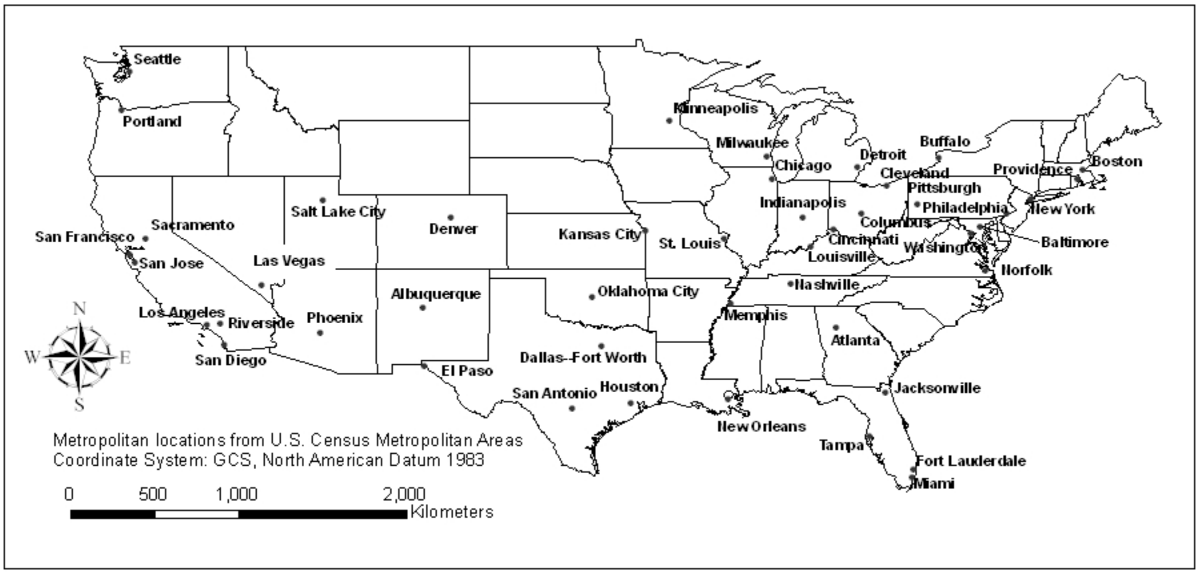

Figure 3. Metropolitan areas included in the tall building district database.

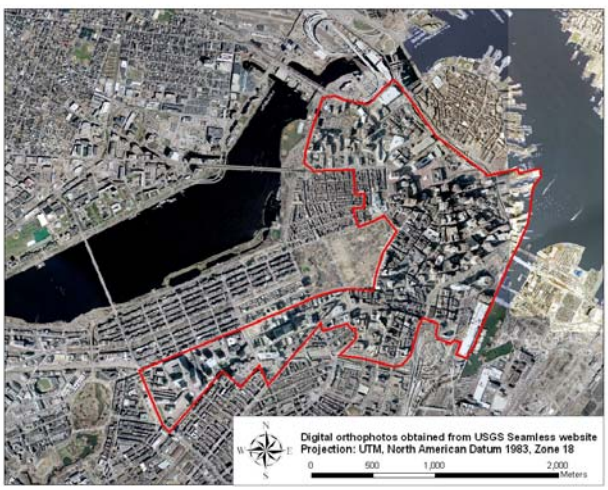

Figure 4. Boston tall building district.

To meet this need, the *first generation* National Building Statistics Database (NBSD) has been computed from a set of three dimensional building datasets for 17 metropolitan areas in the US. Figure 5 shows the distribution of the cities in the NBSD throughout the US. Regionally, five of the cities are on the west coast, two are in the southwest, two are in the mountain west, two are on the gulf coast, two are in the midwest, and three are on the east coast. This is not an ideal distribution, but it is a fairly comprehensive regional coverage with multiple cities per region. Currently additional building datasets are being incorporated into the database and building statistics are being extracted. These will be released with the *second generation* NBSD in the Summer of 2006.

The characteristics of the building datasets used to derive the NBSD are contained in Table 1. All data extents are smaller than the complete metropolitan area, but are centered on the important tall building districts. The majority of the datasets were either obtained from commercial vendors (e.g., i-cubed, Vexcel, Inc., Urban Data Solutions) or extracted from airborne lidar data by the National Geospatial-Intelligence Agency (NGA) using a set of tools created by Science Applications International Corporation (SAIC) in collaboration with the Defense Threat Reduction Agency (DTRA). The Houston dataset is significantly larger than the other datasets. It was derived by University of Utah researchers by modifying an existing building footprint dataset available from the city engineering department. The modification involved comparing the existing footprints to high-resolution digital orthophotos and deleting buildings that did not appear in the orthophoto and digitizing buildings that did appear. The building heights were then derived by overlaying the modified footprint coverage onto a 1-m full-feature DEM produced from airborne lidar data collected by Terrapoint, LLC. Additional details of each building dataset are included in the individual city processing reports included on the NBSD CD.

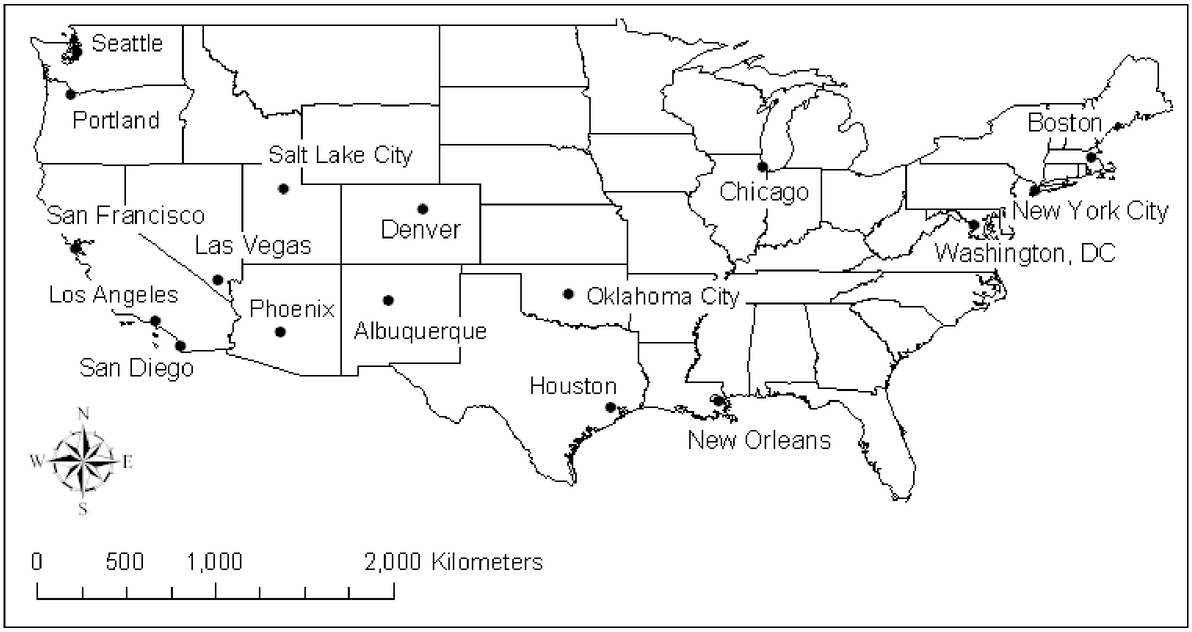

Figure 5. Cities with building data in the National Building Statistics Database.

Table 1. Characteristics of building datasets used to derive NBSD (listed alphabetically).

City	Data Extent (km^2)	Tall Building District Extent (km^2)	Number of Buildings*
Albuquerque, NM	48.5	1.1	22,662
Boston, MA	256.0	3.0	132,007
Chicago, IL	154.1	9.8	47,197
Denver, CO	141.4	5.9	70,209
Houston, TX	1648.6	3.3	664,861
Las Vegas, NV	200.0	10.6	85,030
Los Angeles, CA	12.8	2.5	3,353
New Orleans, LA	26.1	3.7	13,836
New York, NY	321.2	32.5	43,513
Oklahoma City, OK	27.0	0.7	6,333
Phoenix, AZ	16.8	1.7	7,997
Portland, OR	9.5	1.8	2,000
Salt Lake City, UT	140.0	1.6	61,669
San Diego, CA	301.4	3.2	110,432
San Francisco, CA	185.2	4.5	46,935
Seattle, WA	40.8	2.3	35,971
Washington, DC	41.7	13.2	5,756

* The number in this column represents the number of features in the dataset. The data represent in some cases multiple buildings with a single feature or a single building with multiple features. Therefore, the number of buildings listed in the column is an approximation.

The following building statistics are included in the first generation NBSD:

- Mean building height
- Standard deviation of building height
- Plan-area-weighted mean building height
- Height histograms
- Plan area fraction
- Plan area density (at 1-m height increments)
- Building roof area density (at 1-m height increments)
- Frontal area index
- Frontal area density (at 1-m height increments)
- Building surface-to-plan area ratio
- Complete aspect ratio
- Height-to-width ratio
- Sky view factor

To compute these parameters, the Urban Morphological Analysis Processor (UMAP) (Burian et al. 2005) tool developed by researchers at the University of Utah was used. UMAP is a code developed to process three dimensional building datasets to compute urban canopy parameters such as the ones listed above. UMAP is a tool developed for use with the ESRI ArcGIS 9 geographic information system (GIS) software package. UMAP computes building statistics and other land surface parameters selected by a user at a user defined resolution and spatial distribution. UMAP was designed to derive gridded surface parameter datasets corresponding spatially to an atmospheric dispersion modeling domain and has been used in numerous urban morphological studies for various meteorological, dispersion, air quality, and climate modeling activities.

The methods used to compute the building statistics listed above at 250-m spatial resolution are described in detail in the UMAP documentation available from the first author. The results of the *first generation* NBSD are being distributed on a CD and the *second generation* NBSD will be available through a web site and on CD. The distribution CD contains the gridded building statistics in shapefile, raster, Excel, and ascii text format. Figure 6 illustrates a sample gridded dataset containing the computed sky view factor for the San Diego metropolitan area. The quality assurance/quality control documentation accompanies a brief report for each city that describes the data used in the processing, the results, and QA/QC issues.

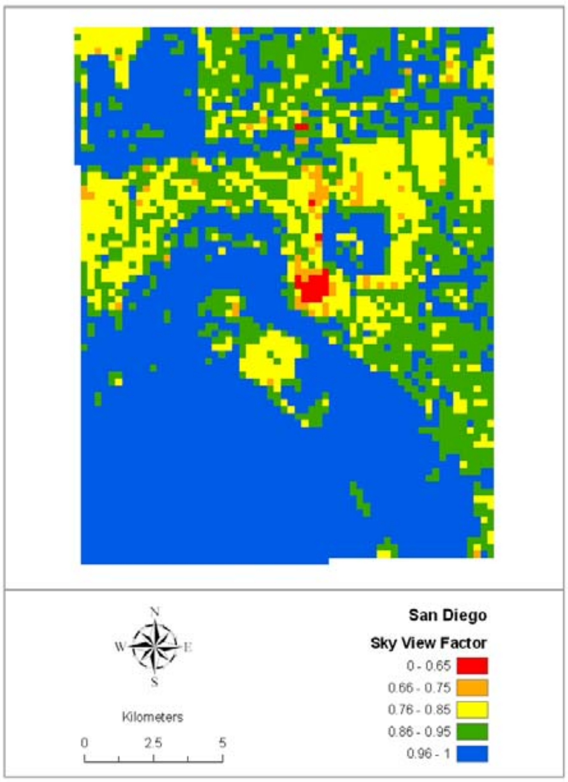

Figure 6. Sky View factor for the San Diego metropolitan area.

6. SUMMARY

A suite of urban databases are being developed by researchers at the University of Utah and Los Alamos National Laboratory. The primary application of the databases is atmospheric transport and dispersion models, but other fine scale modeling applications, e.g., air quality modeling, meteorological modeling, climate modeling, hydrologic modeling are also directly applicable. Currently, the *second generation* NBSD is being created with the following updates to the *first generation* database:

- additional building datasets have been obtained and are being processed
- additional building statistics and urban canopy parameters are being incorporated
- the building statistics are being correlated to underlying population-land use/cover complex and the correlations are being used to extrapolate building statistics to all metropolitan areas in the US
- a web interface is being developed to permit access to the derived building statistics, digital orthophotos, tall building districts, urban footprints, and other emerging urban databases being generated

The *first generation* NBSD CD is available from the authors. Contact the authors to be added to the distribution list.

Acknowledgements

This work was supported by the Department of Homeland Security Biological Countermeasures Program.

7. REFERENCES

Burian, S. J., M. J. Brown, and S. P. Linger, 2002: *Morphological analysis using 3D building databases: Los Angeles*, California. LA-UR-02-0781. Los Alamos National Laboratory, 66 pp.

Burian, S. J., W. Han, S. P. Velegubantla, S. R. K. Maddula, 2003: *Development of gridded fields of urban canopy parameters for Models-3/CMAQ/MM5*. U.S. EPA Internal Report, RTP, North Carolina.

Burian, S. J., A. T. McKinnon, J. Hartman, and W. Han, 2005: *Urban Morphological Analysis Processor (UMAP) Users Manual.* Report Submitted to Los Alamos National Laboratory.

Cionco, R. M. and R. Ellefsen, 1998: High resolution urban morphology data for urban wind flow modeling. *Atm. Env.*, **32**(1), 7-17.

Ellefsen, R., 1990/1991: Mapping and measuring buildings in the canopy boundary layer in ten U.S. cities. *Energy and Buildings*, **15-16**, 1025-1049.

Ellefsen, R. A. and R. M. Cionco, 2002: A method for inventorying urban morphology. Preprints*, Fourth Symposium on the Urban Environment*, AMS, Norfolk, VA.

Elvidge, C. D., K. E. Baugh, et al., 1998: Radiance calibration of DMSP-OLS low-light imaging data of human settlements. *Remote Sensing of Environment*, **68**, 77-88.

Grimmond, C. S. B. and T. Oke, 1999: Aerodynamic properties of urban areas derived from analysis of surface form. *J. Appl. Met.*, **38**, 1262-1292.

Grimmond, C. S. B. and C. Souch, 1994: Surface description for urban climate studies: A GIS based methodology. *Geocarto Int.*, **9**, 47-59.

Imhoff, M. L., W. T. Lawrence, D. C. Stutzer, and C. D. Elvidge, 1997: A technique for using composite DMSP/OLS "City Lights" satellite data to map urban area. *Remote Sensing of Environment*, **61**, 367-370.

Long, N., 2003: *Analyses morphologiques et aérodynamiques du tissue urbain – Application á la climatologie urbaine de Marseille pendant la campagne ESCOMPTE*. Thesis, École Centrale de Nantes, Nantes, France.

Long, N., S. Kermadi, C. Kergomard, P. G. Mestayer, and A. Trébouet, 2003: Urban cover modes and thermodynamic parameters from urban database and satellite data: A comparison for Marseille during ESCOMPTE, Proceedings*, Fifth Int. Conf. Urban Climate*, Łódź, Poland, September 2003.

McPherson, T. N. and M. J. Brown, 2003: Determination of the spatial and temporal distribution of population for air toxics exposure assessments, 5th AMS Urban Env. Conf., Vancouver, B.C., 11 pp. LA-UR-04-4912.

Ratti, C. and P. Richens, 1999: Urban texture analysis with image processing techniques, *Proceedings of the CAADFutures99 Conference*, Atlanta, GA.

Ratti, C., S. Di Sabatino, R. Britter, M. Brown, F. Caton, and S. Burian, 2002: Analysis of 3-D urban databases with respect to pollution dispersion for a number of European and American cities. *Water, Air, and Soil Pollution: Focus*, **2**, 459-469.

Richens, P., 1997: Image processing for urban scale environmental modeling. *Proceedings of the Int. Conf. on Building Simulation '97*, Prague, 163-171.

Vogelmann, J. E., S. M. Howard, L. Yang, C. R. Larson, B. K. Wylie, and N. Van Dreil, 2001: Completion of the 1990s National Land Cover Dataset for the conterminous United States from LANDSAT thematic mapper data and ancillary data sources. *Photogrammetric Engineering and Remote Sensing*, **67**(6), 650-662.