9.4 ASSESSMENT OF METHODS FOR CREATING A NATIONAL BUILDING STATISTICS DATABASE FOR ATMOSPHERIC DISPERSION MODELING

Srinivas Pradeep Velugubantla[1], Steven J. Burian[2], Michael J. Brown[1]
WooSuk Han[2], Andrew T. McKinnon[2], and Timothy N. McPherson[1]

[1] Energy and Environmental Analysis Group, LANL, Los Alamos, NM
[2] Department of Civil and Environmental Engineering, University of Utah, Salt Lake City, UT

1. INTRODUCTION

Mesoscale meteorological codes and transport and dispersion models are increasingly being applied in urban areas. Representing urban terrain characteristics in these models is critical for accurate predictions of air flow, heating and cooling, and airborne contaminant concentrations in cities. A key component of urban terrain characterization is the description of building morphology (e.g., height, plan area, frontal area) and derived properties (e.g., roughness length). Methods to determine building morphological statistics range from manual field surveys to automated processing of digital building databases.

In order to improve the quality and consistency of mesoscale meteorological and atmospheric dispersion modeling, a national dataset of building morphological statistics is needed. Currently, due to the expense and logistics of conducting detailed field surveys, building statistics have been derived for only small sections of a few cities. In most other cities, modeling projects rely on building statistics estimated using intuition and best guesses. There has been increasing emphasis in recent years to derive building statistics using digital building data or other data sources as a proxy for those data. Although there is a current expansion in public and private sector development of digital building data, at present there is insufficient data to derive a national building statistics database using automated analysis tools. Too many cities lack digital data on building footprints and heights and many of the cities having such data do so for only small areas.

Due to the lack of sufficient digital building data, other datasets are used to estimate building statistics. Land use often serves as means to provide building statistics for a model domain, but the strength and consistency of the relationship between land use and building morphology is largely uncertain. In this paper, we investigate whether building statistics can be correlated to the underlying land use. If a reasonable correlation exists, then a national building statistics database could be created since land use is available for the entire U.S.

Digital datasets of building footprint and height information have been obtained, validated and analyzed for eight western U.S. cities covering areas ranging from 6 km^2 to 1653 km^2. Building morphological statistics (including mean and standard deviation of building height, plan area fraction and density, rooftop area density, frontal area index and density, building-to-plan area ratio, complete aspect ratio, height-to-width ratio, roughness length, displacement height, and sky view factor) have been computed for each city at 250-m resolution and are being correlated to underlying land use type. This paper will summarize the building statistics from the eight cites focusing on the variability within each city and between cities as a function of land use.

**Corresponding author address*: Michael Brown, MS-F604, LANL, Los Alamos, NM 87545; E-mail: mbrown@lanl.gov

2. BACKGROUND

Mesoscale meteorological models and many atmospheric dispersion models do not have the spatial resolution to simulate the fluid dynamics and thermodynamics near and around buildings and other sub-grid urban land features. Urban canopy parameterizations are one approach to represent the effects of buildings and urban land features on drag, turbulence production, heating, and radiation trapping (Brown 2000). Urban canopy parameterizations require the high-resolution description of the urban canopy using building (and in some cases tree) morphological statistics and surface cover physical properties. Determining building geometric parameters (e.g., height, footprint area, width) can be accomplished by:

1. performing detailed field surveys and measuring building height and geometry data,

2. visually analyzing high-resolution aerial photographs to estimate building height and geometry data,

3. analyzing stereographic paired digital photographs to extract building height and geometry data, and

4. collecting airborne Light Detection and Ranging (LIDAR) digital elevation models and extracting building height and geometry data.

Ground surveys are time consuming and require *in situ* measurements (i.e., the analyst must be present in the city of interest). Depending on the level of detail recorded and the care taken in collecting measurements a relatively high level of accuracy is possible. However, due to time requirements ground surveys are only feasible for very small sections of cities. Similar to ground surveys, visual analysis of aerial photographs is also time consuming and only feasible for small areas of cities. Accuracy of building statistics will likely be lower than in ground surveys because all geometry and height information is estimated from visual interpretation of aerial photographs (e.g., estimating number of floors and multiplying by a standard height per floor). Analysis of stereographic images will be more efficient than ground surveys and visual photo interpretation, but requires stereo-paired images and special hardware and software to perform the analysis. Accuracy of building statistics from the photogrammetric analysis should be higher than visual interpretation of photographs. Creation of a full-feature digital elevation model (DEM) from airborne LIDAR data is rapidly becoming a popular method for collecting urban 3D data. Primary advantages over other techniques include cost-effective automated collection of large amounts of data in a format that can be easily processed and a relatively high degree of measurement accuracy.

Building morphological statistics can be derived from building height and geometry information using one of the following approaches:

- hand calculations using height and geometry data for a small number of buildings,
- simplified calculations using mean heights and geometric parameters for a larger number of buildings, or
- automated or semi-automated analysis of full-feature DEM or vector representations of buildings using geographic information system (GIS) or image processing software.

Automated analysis of a DEM or digital building database is by far the most efficient means to obtain building statistics for large areas and the results can be accurate if the raw building data being processed is accurate. The major difficulty of computational processing of building data is obtaining (or creating) the building dataset using one of the methods previously described.

Several methods using analysis of aerial photographs and field surveys have been introduced to inventory, estimate, or calculate building morphological statistics (e.g., Ellefsen 1990/1991; Theurer 1993, 1999; Ellefsen and Cionco 2002). Several researchers have also presented GIS and image processing approaches to compute building morphological statistics (e.g., Grimmond and Souch 1994; Ratti and Richens 1999; Burian et al. 2002a; and Long et al. 2002). Using an automated GIS or image processing approach to derive a national database is computationally feasible, the primary problem is lack of a nationally consistent building dataset to process. Consequently, the development of the first generation National Building Statistics Database must rely on existing nationally consistent datasets to produce the national coverage of building statistics.

There are several sources and types of data that may be used in constructing a national database of building statistics. One of the data types commonly related to building morphology, and that also has national coverage, is land use. In this research, we are attempting to affirm the use of land use data in the extrapolation of building statistics beyond areas with building data by assessing the strength of the relationship between land use and building statistics. We used building statistics derived from high-resolution building data and the highest resolution and quality land use datasets available for the cities studied. The following section describes the building and land use datasets used.

3. BUILDING AND LAND USE DATABASES

The preliminary correlation between building statistics and land use and population is being based on a set of building statistics derived for eight cities in the western U.S. using the approach of Burian et al. (2002a). Figure 1 displays the locations of the eight cities and Table 1 lists the characteristics of the building databases. Reports describing these datasets and containing building morphological statistics are listed in the References section.

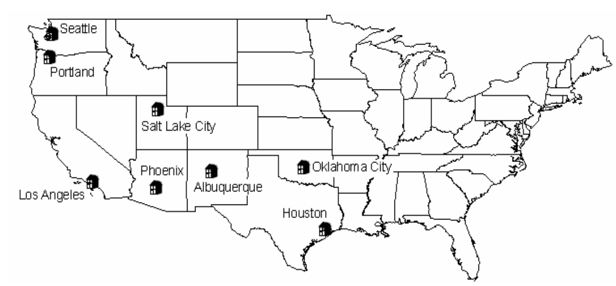

Figure 1. U.S. map showing locations of eight cities for which we have derived building morphological statistics.

Table 1. Characteristics of building databases for the eight western U.S. cities include in the study.

City	Area (km^2)	Number of Buildings	Data Source
Albuquerque	48.5	22,662	I-cubed
Houston	1653.0	664,861	COH*
Los Angeles	12.0	3,353	Aerotopia
Oklahoma City	27.0	6,333	JPSD*
Phoenix	16.7	7,997	Vexcel
Portland	9.5	2,000	I-cubed
Salt Lake City	6.1	2,891	UDS*
Seattle	41.0	35,971	COS*

* COH = City of Houston; JPSD = Joint Precision Strike Demonstration; UDS = Urban Data Solutions; COS = City of Seattle

One concern with the data is the different coverage areas used to derive the building statistics for the eight cities. Analysis areas for all cities include the downtown core area, but the total coverage areas range from 6 km^2 to 1653 km^2. This will play an important part for the downtown core areas, where the Commercial & Services land use predominates and the building statistics will be much different than outlying parts of the city. For other land uses, the amounts in the downtown core area will be relatively small and the building characteristics will change relatively little as a function of distance from downtown compared to Commercial & Services. Therefore, representing Commercial & Services land use with a single building statistic is not appropriate. In the future, we will reevaluate the data with the downtown core areas called out separate.

Building morphological statistics can be derived as a function of urban land use type by integrating the analysis of building databases with land use datasets in a GIS or image processing software package. For this analysis, the two datasets were intersected in GIS and subsequent morphological analyses conducted. Special land use databases were used for each city owing to limitations of the nationally-available National Land Cover Dataset (NLCD) and U.S. Geological Survey (USGS) Land Use/Land Cover (LULC) data (McPherson et al. 2004). Databases were obtained directly from the cities or regional governing/planning entities (Los Angeles, Phoenix, Albuquerque, Oklahoma City, and Seattle) or created using high-resolution digital orthophotos (Salt Lake City, Portland, and Houston). The databases were aggregated into the Anderson Level 2 scheme (Anderson et al. 1976) with the following seven urban land use types: Residential, Commercial & Services, Industrial, Transportation/Communication/Utilities, Mixed Industrial & Commercial, Mixed Urban or Built-up, Other Urban or Built-up.

4. METHODOLOGY

Building morphological statistics were derived as a function of urban land use type by integrating the analysis of building databases with land use datasets in GIS. For this analysis, the two datasets were intersected in GIS and the following building morphological statistics were computed for each homogeneous land use polygon in the land use datasets for the either cities listed previously:

- Mean and standard deviation of building height
- Plan-area-weighted mean building height
- Building height histograms
- Plan area density (λ_p)
- Frontal area density (λ_F)
- Wall-to-plan area ratio
- Complete aspect ratio (λ_C)
- Height-to-width ratio (λ_s)
- Sky view factor (Ψ_{svf})
- Roughness length (z_o)
- Displacement height (z_d)

The roughness length and displacement height were computed using three approaches: (1) an estimate of roughness length and displacement height as 0.1 and 0.5, respectively, times the mean building height (abbreviated Rule in results), (2) using morphometric equations introduced by Raupach (1994) (abbreviated Ra in results), and (3) using morphometric equations introduced by Macdonald et al. (1998) (abbreviated Mac in results). Additional description of each of these statistics and their relevance for dispersion modeling is provided in Burian et al. (2002a).

The focus of our building statistic-land use correlation analysis is to:

1. determine the relative variability of building statistics across land use types for each city, individually,

2. determine the relative variability of building statistics across the eight cities for each land use type, and

3. divide the eight cities into groups with similar building statistics.

The first objective is seeking to identify building statistics that do not vary significantly across land use types for a city. Statistics that have this quality can then be defined, based on the mean value for many cities, independent of land use for extrapolation to other cities with identified similar physical or geographic characteristics. If the low variability across land uses for a statistic is noted for all (or most) cities then it might be possible to consider the building statistic to essentially be constant.

The second objective is seeking to identify building statistics that do not vary significantly across cities for a given land use type. This differs from the first objective in that the building statistics could vary over land use types. Means of the building statistics that do not vary between cities for a single land use type could then be used in an extrapolation process.

The third objective is trying to find groups of cities, or clusters, for which building statistics will have less variability with land use, population, and other datasets. Common physical characteristics or geographic features of the clusters will then be catalogued. If clusters can be identified, then during extrapolation cities without building data will be classified into a cluster and assigned building statistics based on extrapolation of the cluster building statistics.

5. RESULTS

Our preliminary analyses involved compiling the building statistics in tables and creating specialized plots to visually explore the variability of the statistics. Table 2 displays one example of the many tables created. The mean building height is seen to vary significantly between cities. The largest variability occurs for the Commercial & Services category. As noted in McPherson et al. (2004) and earlier in this

Table 2. Mean building heights (m) for eight western U.S. cities shown by urban land use class.

Land Use Class	*Houston, TX*	*Albuquerque, NM*	*Phoenix, AZ*	*Oklahoma City, OK*	*Los Angeles, CA*	*Salt Lake City, UT*	*Portland, OR*	*Seattle, WA*
Residential	5.5	4.3	3.8	4.8	6.4	9.6	10.0	6.0
Commercial & Services	6.0	6.1	8.5	6.0	24.5	17.9	14.1	11.7
Industrial	6.0	5.5	5.1	5.7	6.3	10.8	8.0	7.6
Transportation/ Communications/Utilities	4.6	5.6	0.0	4.2	7.9	---	7.5	8.2
Mixed Industrial & Commercial	5.0	---	---	---	7.5	---	---	--
Mixed Urban or Built-up	5.7	---	---	---	12.0	11.2	---	9.6
Other Urban	4.9	6.1	5.4	7.2	7.4	13.8	8.7	7.5

paper, Commercial & Services is made up of widely varying land use types, from high-rise areas to strip malls to business centers and should be separated into sub-categories based on morphological characteristics. Overall, we note mean building height has little variability across land use types for several of the cities (e.g., Houston, Albuquerque, and Oklahoma City) suggesting the mean building height could be defined outside of the downtown core area for these cities using a single mean value.

For other building morphological parameters, clear relationships with land use were, in general, not found for the eight cities. A high degree of variability among building statistics was found, with the Commercial & Services land use category showing the greatest variability. However, we discovered that many of the building statistic values were similar if the cities were clustered into two groups: the first being cities located in the southwest U.S. (Houston, Albuquerque, Phoenix, and Oklahoma City) and the second cluster consisting of cities located in coastal or mountain regions (Los Angeles, Salt Lake City, Portland, and Seattle). Figures 2 and 3 show the clustering of the building statistics into two clusters for Residential and Commercial & Services land uses, respectively. The plots depict the percent difference of the building statistic for each city from the mean value for all cities. For most all parameters shown there is a clear grouping of southwest cities (warm colors) below the mean value and the coastal/mountain cities (cool colors) above the mean. For both Residential and Commercial & Services land use, the southwest cities appear to cluster more tightly when looking at all building parameters.

Figure 2 indicates that for Residential land use in the four southwest cities most all building statistics cluster to within 50% of one another. Complete aspect ratio is nearly the same for all four cities. Note that the rule-of-thumb roughness length and displacement height are computer directly from mean building height and therefore have the same clustering properties.

Although the overall variability appears greater for the four coastal/mountain cities, for Residential land use there are specific building statistics that have less variability

- Complete Aspect Ratio
- Building to Surface Area
- Height to Width Ratio
- Frontal Area Index

For several variables (height, roughness length, displacement height) Portland and Seattle track together, so do Los Angeles and Salt Lake City.

Figure 3 indicates that among the southwest cities, Houston, Albuquerque and Oklahoma City have very similar building parameter values, while Phoenix followed the same trend but had slightly higher values. The building parameters for the four coastal/mountain cities show very large variability for Commercial& Services land use, except for complete aspect ratio. Los Angeles building statistics were furthest from the mean in most all cases except for height-to-width ratio, plan area fraction, and standard deviation of building height where Seattle exhibited the largest positive differences from the mean. Interestingly, Portland shows up as an average city when considering all eight cities with most percentage differences being close to zero.

Individual building statistics with consistently low and high variability were identified from Figures 2 and 3. The complete aspect ratio (λ_C) was found to have the lowest variability for all land uses, while the roughness length (z_o) had the highest variability. The low variability of the λ_c can be partially attributed to its definition. It is computed as the ratio of the sum of surface area of buildings (walls and rooftops) and exposed ground to the total plan area:

$$\lambda_C = \frac{A_C}{A_T} = \frac{A_W + A_R + A_G}{A_T} \quad (1)$$

where A_C is the complete surface area, A_W is the total wall surface area, A_R is the total rooftop surface area,

A_G is the total exposed ground surface area, and A_T is the total surface area of the site. The values of complete aspect ratio for all land uses ranged from 1.0 to 1.8. A possible reason for the relatively low variability might be due to the sum A_R and A_G being approximately the same as A_T, while A_w may vary. In general outside of the downtown core area, A_T will be much larger than A_W. Although A_W was found to vary considerably between cities in this study and others (e.g., Ellefsen 1990/1991), the effect of the variability of A_W is diminished by its small amount relative to A_T.

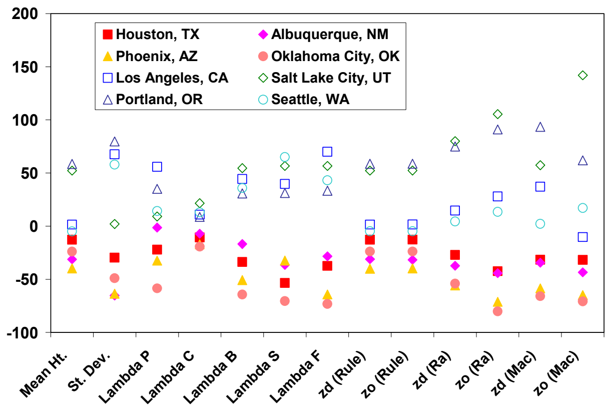

Figure 2. Percent difference from overall mean building statistic value for eight cities for Residential land use.

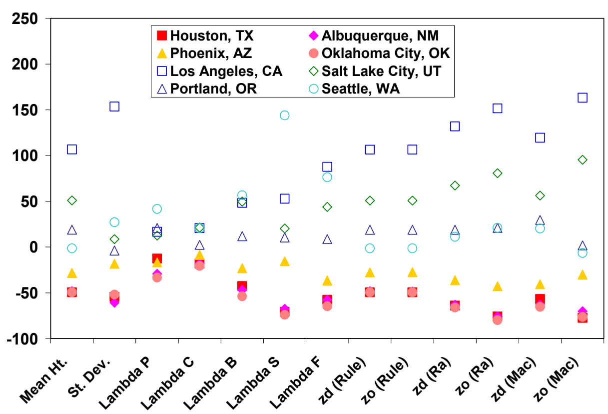

Figure 3. Percent difference from overall mean building statistic value for eight cities for Commercial & Services land use.

Additional data plots were created to focus on four urban land use types: Residential, Commercial & Services, Industrial, and Other Urban or Built-up. These four land uses comprise a major fraction of the urban land use in the eight cities and contained nearly all the buildings. Thus it is important to understand the variability of building statistics for these land uses. Figure 4 shows the plot of mean building height for the four selected urban land use classes for the eight cities. The data plotted in this form does not necessarily corroborate the clustering of cities into two groups. The southwest cities (Houston, Albuquerque, Oklahoma City, and Phoenix) do have relatively little variability for all land use types (range of mean heights from 3 to 8 m), while the coastal/mountain cities have considerable variability (from ~5 m to more than 23 m). But, the value for Los Angeles and Seattle would fit in the range of Southwest cities for all land uses except Commercial & Services. Overall, the Commercial & Services land use category has the highest variability of mean building height from city to city, while Industrial and Residential have the lowest. Thus, using an overall mean value per land use type to extrapolate mean building height to other cities without building data would be reasonable for Industrial and Residential land uses.

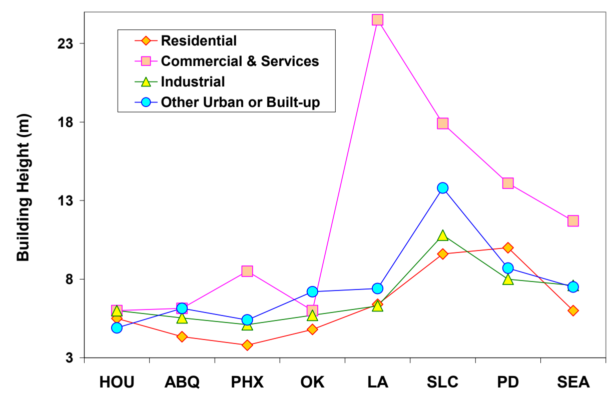

Figure 4. Mean building height for four urban land use types for all eight cities.

Figure 5 depicts the plot of standard deviation of building height for the four urban land use classes for all eight cities. The standard deviation of building height for Residential, Industrial, and Other Urban or Built-up land use types is relatively small for all cities and thus an overall mean value for all cities may be useful for extrapolation to other cities. The Commercial & Services land use category, on the other hand, has considerable variability from city to city and could not be represented accurately using an overall mean value for all cities. Similar to the mean building height, the two clusters of cities are not apparent.

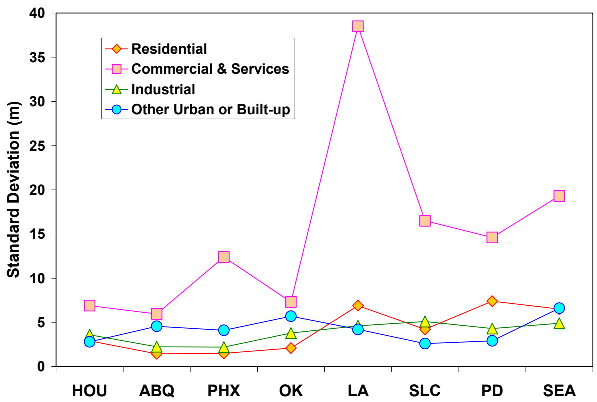

Figure 5. Standard deviation of building height for four urban land use types for all eight cities.

Figure 6 indicates that the southwest cluster of cities (left four cities) has less building parameter variability than the cluster of coastal/mountain cities (right four cities). Other urban land use has the lowest value of height-to-width ratio in all the cities except Oklahoma City. This may be partially attributed to the vague definition of Other Urban or Built-up such that it contains a wide variety of land uses that have relatively sparse building densities including urban parks, educational institutions, government buildings, museums. The low variability for Industrial and Other Urban or Built-up land uses for all eight cities suggests using an overall mean value for extrapolation for these two land uses may be appropriate. The Residential land use has a low amount of variability for the southwest cluster of cities and for the coastal/mountain cities when considered as separate groups. This indicates that the mean height-to-width ratios for the southwest and coastal/mountain clusters may be appropriate for other cities that can be classified in these clusters.

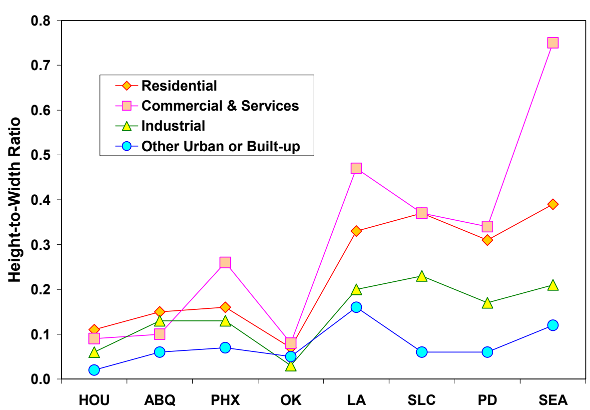

Figure 6. Building height-to-width ratio for four urban land use types for all eight cities.

Further analyses were performed for the other building statistics and significant variability continued to be observed. No general relationships could be established that worked for all parameters, all cities, all land uses. This discovery was somewhat discouraging and suggests using land use as the extrapolation medium by itself will produce fair to poor accuracy for a National Building Statistics Database. Extrapolation of a few building statistics using mean values or mean values for clusters as a function of land use may be possible, but most likely additional information will need to be incorporated into the extrapolation. Results from the building statistic – land use correlation analysis will be summarized and decisions regarding the extrapolation will be reported in the presentation.

6. SUMMARY & FUTURE WORK

This paper summarized preliminary efforts to develop a National Building Statistics Database. Data from more than eight cities have been analyzed and coverage of gridded building statistics for those cities is currently compiled in the form of a national database. Relationships between building statistics and land use have been explored and found to vary significantly. A few parameters were identified as having relatively low variability across the eight cities and these may be represented by mean values and extrapolated without significant error to other urban areas to fill in the national database coverage. However, most parameters were found to be too variable for such an extrapolation. These parameters are being further studied for correlation to daytime and nighttime population and land use. Results from building statistics-population analysis will be highlighted in the presentation.

Acknowledgements

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

7. REFERENCES

Anderson, J. R., E. E. Hardy, J. T. Roach, R. E. Witner, 1976: *A land use and land cover classification system for use with remote sensor data*. USGS Prof. Paper 964.

Brown, M. J., 2000: Urban parameterizations for mesoscale meteorological models. *Mesoscale Atmospheric Dispersion*. Ed., Z. Boybeyi., pp. 193-255.

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

Burian, S. J., S. R. K. Maddula, S. P. Velugubantla, and M. J. Brown, 2004: *Morphological analyses using 3D urban databases: Seattle, WA,* LA-UR-XX-XXXX, Los Alamos National Laboratory, Los Alamos, NM.

Burian, S. J., S. R. K. Maddula, S. P. Velugubantla, and M. J. Brown, 2003a: *Morphological analyses using 3D urban databases: Oklahoma City, OK,* LA-UR-XX-XXXX, Los Alamos National Laboratory, Los Alamos, NM.

Burian, S. J., S. R. K. Maddula, S. P. Velugubantla, and M. J. Brown, 2003b: *Morphological analyses using 3D urban databases: Houston, TX,* LA-UR-XX-XXXX, Los Alamos National Laboratory, Los Alamos, NM.

Burian, S. J., S. R. K. Maddula, S. P. Velugubantla, and M. J. Brown, 2002b: *Morphological analyses using 3D urban databases: Albuquerque, NM,* LA-UR-03-6198, Los Alamos National Laboratory, Los Alamos, NM.

Burian, S. J., S. P. Velugubantla, and M. J. Brown, 2002c: *Morphological analyses using 3D urban databases: Phoenix, Arizona,* LA-UR-02-6726, Los Alamos National Laboratory, Los Alamos, NM.

Burian, S. J., S. P. Velugubantla, K. Chittineni, S. R. K. Maddula, and M. J. Brown, 2002d: *Morphological analyses using 3D urban databases: Portland, Oregon,* LA-UR-02-6725, Los Alamos National Laboratory, Los Alamos, NM.

Burian, S. J., S. P. Velugubantla, and M. J. Brown, 2002e: *Morphological analyses using 3D urban databases: Salt Lake City, Utah,* LA-UR-03-6197, Los Alamos National Laboratory, Los Alamos, NM.

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.

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

Long, N., Mestayer, P. G., and C. Kergomard, 2002: Development of a software to describe the city morphology and to compute aerodynamic parameters from an urban database. Preprints*, Fourth Symposium on the Urban Environment*, AMS, Norfolk, Virginia.

McPherson, T. N. and M. J. Brown, 2004: Estimating daytime and nighttime population distributions in U.S. cities for emergency response activities. Preprints, *84th AMS Annual Meeting*, AMS, Seattle, WA.

McPherson, T. N., M. J. Brown, and S. J. Burian, 2004: Evaluation of the accuracy of national landuse datasets for the development of urban canopy parameterizations. Preprints, *Fifth Symposium on the Urban Environment*, AMS, Vancouver, BC.

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