9.7

EVALUATION OF THE ACCURACY OF NATIONAL LANDUSE DATASETS FOR THE DEVELOPMENT OF URBAN CANOPY PARAMETERIZATIONS

Timothy N. McPherson[1], Michael J. Brown[1], Steven J. Burian[2]
[1]Los Alamos National Laboratory, Los Alamos, NM
[2]University of Utah, Salt Lake City, UT

1. Introduction

Describing urban terrain and land use/land cover (LULC) characteristics accurately is an essential element in environmental modeling, assessment and management. Urban terrain and LULC information are especially important in meteorological modeling applications, such as simulations of atmospheric flow over cities, defining surface energy budgets, and fate and transport studies of urban contaminants. Urban terrain data is difficult to acquire and many researchers rely on LULC data to develop urban canopy parameterizations for mesoscale meteorological models. Although urban canopy parameterizations are an extremely useful tool in mesoscale meteorological models, they are often applied using out of date or poorly attributed land use data. The two major sources of national land use data are the U.S. Geological Survey (USGS) land use/land cover dataset and the National Land Cover Database (NLCD). The USGS database was developed based on aerial photographs from circa 1970 and may be significantly underestimating urban areas due to the extensive urban development that has occurred in the United States since the time of the imagery. The NLCD was constructed from more recent LANDSAT TM imagery (circa 1990), but it has limited attribution of urban classes. The age of the data and level of detail could be significantly impacting urban canopy parameterizations and their associated mesoscale meteorological models and the sufficiency of these two datasets for developing urban canopy parameterizations needs to be assessed.

The objective of this investigation is to compare qualitatively and quantitatively the USGS dataset and the NLCD with more detailed and recent datasets. We characterize the shortcomings of the USGS land use and NLCD, especially for environmental modeling applications. First, we analyze the similarities and differences between the urban footprint estimated from the USGS land use dataset, the NLCD, and more detailed land use datasets for select cities. And then, we present a new method for deriving urban footprints for the US.

2. BACKGROUND

In the past several decades multiple models simulating urban transport and dispersion have been developed. These models rely on a wide range of data including urban land use. Urban land use type, e.g., residential, commercial, and industrial, is used in in mesoscale meteorological models to assign building morphological parameters and other surface characteristics to the underlying terrain. These parameters affect plume transport and dispersion, impacting the wind speed, wind direction, and turbulent mixing. Accurate urban land use definition is therefore an important component in mesoscale modeling efforts.

There are two major land use datasets available for the United States: the USGS Land Use/Land Cover (LULC) and the USGS-USEPA National Land Cover Database (NLCD 92). These two datasets, however, have significant deficiencies for urban applications (e.g., Burian and Brown, 2001). In many cases, they include a drastic under estimation of city coverage and/or an over simplification of the urban land use types. Either of these deficiencies could significantly impact transport and dispersion calculations for particular scenarios. In this study, we evaluate the urban footprint (coverage area) of the two major urban land use datasets by comparison to special urban land use datasets obtained directly from city governments.

Urban land use data are derived using one of three methods: 1) computer automation of land use classification from remotely sensed multi-spectral imagery, 2) manual classification of satellite imagery or aerial photographs, 3) or ground inspections. These different methods create land use data with varying levels of spatial resolution and feature classification. Accuracy is also variable between the methods due the inherent strengths and weaknesses in each technique. Numerous land use datasets for local, metropolitan, and national areas have been constructed using these methods. Mesoscale modeling efforts will be improved if a consistent, accurate, and easily obtainable land use dataset for urban areas in the United States is acquired or developed.

Land use/land cover data for areas in the United States can be obtained from several sources including federal agencies, state natural resource agencies, regional planning associations, or commercial vendors. At present, the two national land use databases with complete national coverage are the US Geological Survey (USGS) Land Use/Land Cover (LULC) dataset and the joint 1992 USGS-US Environmental Protection Agency (USEPA) National Land Cover Dataset (NLCD 92). These are available at no cost to the public and can be downloaded via the web.

The LULC dataset was derived by the USGS from National Aeronautics and Space Administration (NASA) high-altitude aerial photographs and National High-Altitude Photography (NHAP) program photographs collected in the 1970s and early 1980s.

The scales of these photographs were typically less than 1:60,000 (USGS, 1990). Land use/land cover was defined in the LULC using an Anderson Level II classification system (Anderson et al., 1976). The LULC is available in vector and raster formats. In the vector format, land use is defined by polygons that represent contiguous areas of homogeneous land use. The minimum polygon size for land use areas is either 4 or 16 hectares (40,000 or 160,000 m^2, respectively). Urban or built-up land and urban transitional areas in the LULC use the 4 hectares minimum polygon size with a minimum feature width of 200 m. For categories other than urban or built-up land and water, the 16-hectare minimum polygon size is used with a minimum feature width of 400 m. The raster format is termed Composite Theme Grid. These data are derived from the polygon files using a regular point sample of the quad. The center point of each cell is 200 meters apart from other center points in adjacent cells (USGS, 1990).

The NLCD 92 (Vogelmann et al., 1998) is a raster-based dataset produced in a cooperative project between the USGS and USEPA. The goal of the NLCD 92 development project was to produce a consistent land cover dataset for the conterminous U.S. based on 30-meter Landsat 5 thematic mapper data supplemented with ancillary data such as census population data. The classification system used for the 30-meter NLCD 92 was a modified form of the Anderson classification system. The analysis and interpretation of the satellite imagery was performed using multi-state image mosaics with ground-truthing performed from a relatively small number of aerial photographs. The accuracy of these data is greatest at the regional level. Local analyses using the NLCD 92 are not recommended without an assessment of the applicability of the data (Vogelmann, et al. 2001).

Both the LULC and NLCD 92 datasets have been used in a number of urban analyses (e.g., Akbari and Rose, 2001a; Akbari and Rose, 2001b) without significant assessment of their accuracy or the implications of that accuracy on the respective study. Burian and Brown (2001) found that these two databases can have significant deficiencies in urban areas including a drastic under estimation of city coverage and/or an over simplification of the urban land use types.

Table 1 lists the urban land uses within the LULC and NLCD 92. Only four of the thirty land uses in the NLCD 92 are urban. The NLCD 92 has two residential land uses, one land use that is a composite of commercial, industrial and transportation uses, and another that represents vegetated areas in urban regions such as parks. The USGS dataset has 46 land use/cover classes, but only seven can be classified as urban. Several USGS LULC land use classes are also composites of multiple uses. The lack of sufficient differentiation in urban land classes in both the NLCD 92 and LULC may reduce the applicability of those datasets in many urban studies. For example, Burian, *et al.* (2001) found that differences between the LULC and a more detailed metropolitan land use dataset affected pollutant loading estimates in a watershed-based non-point source pollution loading model.

Table 1. Urban classification system used in the national land use datasets.

NLCD 92	LULC
Low-Intensity Residential	Residential
High-Intensity Residential	Commercial and Services
Commercial/Industrial/ Transportation	Industrial
Urban/Recreational Grasses	Transportation, Communications and Utilities
	Industrial and Commercial Complexes
	Mixed Urban or Built-up Land
	Other Urban or Built-up Land

In addition to the lack of sufficient urban land use differentiation, both the LULC and NLCD 92 were based on data acquired in the 1970s and 1990s, respectively, and they may not completely define the area of urbanization in given regions. The use of high altitude aerial photography and satellite imagery to derive those datasets could also be causing spatial errors in the definition of urban area. The poor differentiation of urban land uses and errors in the spatial coverage of urban area can significantly impact transport and dispersion calculations for particular scenarios.

3. Methodology

The goal of this study is to evaluate the magnitude of the errors in city coverage of the USGS-USEPA NLCD 92 and USGS LULC for four large metropolitan areas in the southwestern United States. Each national dataset was compared to four higher resolution metropolitan land use datasets in Los Angeles, Phoenix, San Diego, and Denver. These cities were chosen because of the availability of high-resolution metropolitan land use datasets. In this report, each of the two national land use datasets will be evaluated for the precision of their city coverage relative to metropolitan land use datasets that are known to have greater accuracy in urban area classification. The cities studied and the sources of the metropolitan land use data are listed in Table 2. Although the metropolitan datasets are used as a reference in this analysis, they are not ground truth for current land use patterns. The metropolitan datasets are 7 - 12 years old and significant land use alteration could have occurred in the intervening time period. Even though the metropolitan datasets are not ground truth for current conditions, each of these datasets is at a minimum as new as the two national land use datasets and therefore can be used to evaluate the accuracy of those national land use datasets.

Table 2: Study cities & sources of high fidelity urban land use datasets

City	Data Source	Year
Los Angeles	Southern California Association of Governments	1993
San Diego	San Diego Association of Governments	1994/1995
Phoenix	Maricopa County Association of Governments	1995
Denver	Front Range Information Resources Project	1990

3.1 Landuse datasets

Southern California Association of Governments (SCAG)
The SCAG LULC dataset is based on low-altitude aerial photos shot during 1993. The SCAG LULC dataset is classified according to a modified Anderson level III/IV classification system, with 108 LULC classes, 74 of which are urban. The SCAG dataset has a spatial resolution of approximately 0.25 ha (2500 m2).

Maricopa Association of Governments (MAG)
MAG land use land cover data was collected in 1995 using ground truthing, aerial photography and telephone surveys of industrial and commercial properties. The spatial accuracy of the data is 6 meters in developed regions and approximately 60 meters in undeveloped regions. The dataset consists of 25 land uses, 18 of which are urban.

Colorado Front Range Infrastructure Resources Project
The Denver LULC dataset used in this study was generated by the Front Range Infrastructure Resources Project (FRIRP) in 1990. This land use dataset covers from Ft. Collins to the southern portion of the Denver metropolitan area. The FRIRP data was collected from 1-meter resolution digital orthophotographs at a minimum mapping unit of 1 hectare with a minimum polygon width of 38 meters. The FRIRP has been classified using an Anderson hierarchical classification system with 4 to 5 classification levels. The FRIRP dataset contains 20 land use types, 9 of which are urban.

San Diego Association of Governments
The SANDAG LULC dataset is based on 1995 color infrared satellite imagery, 1994/1995 black and white digital orthophotographs and the SanGIS landbase. The SANDAG LULC dataset is classified according to a modified Anderson level II classification system at a spatial resolution of approximately 1:24,000. The SANDAG dataset has 20 LULC classes, 11 of which are urban.

Each of the metropolitan land use classifications and the USGS LULC data were converted to raster format using the Arcview Spatial Analyst GIS software. The grid cell size selected for the new raster-formatted files was equivalent to the grid cell size used in the NLCD 92. This grid cell size was chosen to facilitate spatial analyses between the three datasets and to minimize potential polygon to raster conversion errors as noted by Congalton (1997). The resulting raster files were then used to calculate the area of urban agreement and differences between the LULC and NLCD 92 relative to the metropolitan land use dataset for each city. Following assessment of the accuracy of the LULC and NLCD 92, these datasets were combined using map algebra to determine if the combined dataset more accurately defines the urban footprint. We used the combined dataset to calculate the area of urban agreement and differences between the combined LULC-NLCD 92 dataset relative to the metropolitan land use dataset for each city.

4. Results

4.1 Comparison of metropolitan land use datasets and national datasets

Figures 1-4 display the comparison between the National Land Cover Dataset (NLCD 92) and the high fidelity metropolitan land use datasets. Figures 5-8 display the comparison between the USGS LULC and the metropolitan land use datasets. Each plot contains a map of the urban footprint as defined by both the metropolitan dataset and the respective national land use dataset, a map of the agreement

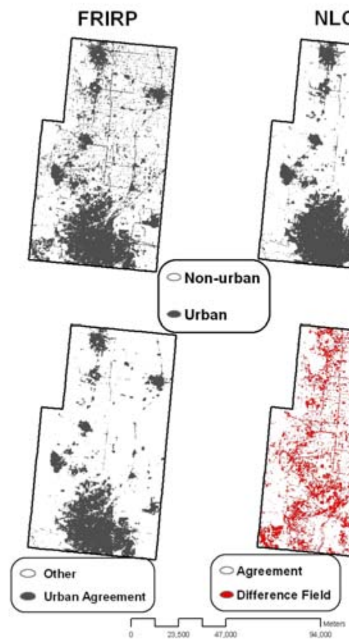

Figure 1. Comparison of the NLCD to the FRIRP dataset in the Denver region.

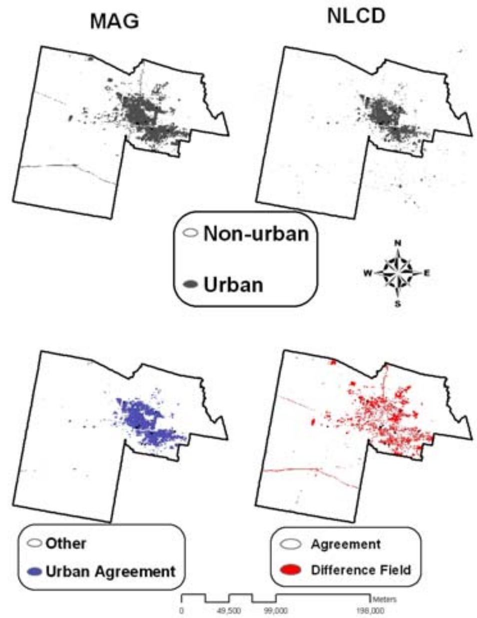

Figure 3. Comparison of the NLCD to the MAG dataset in the Phoenix region.

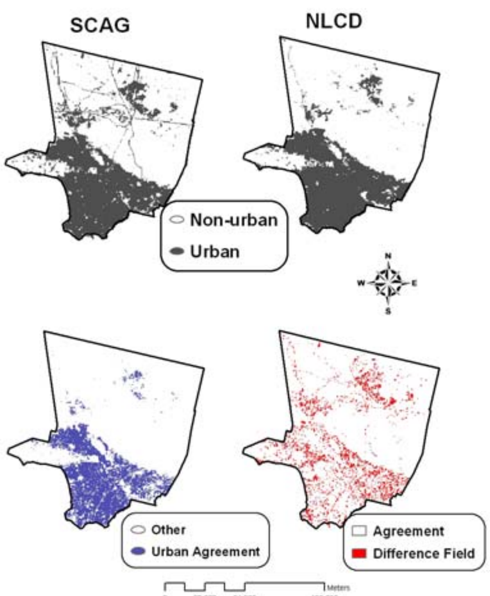

Figure 2. Comparison of the NLCD to the SCAG dataset in the Los Angeles region.

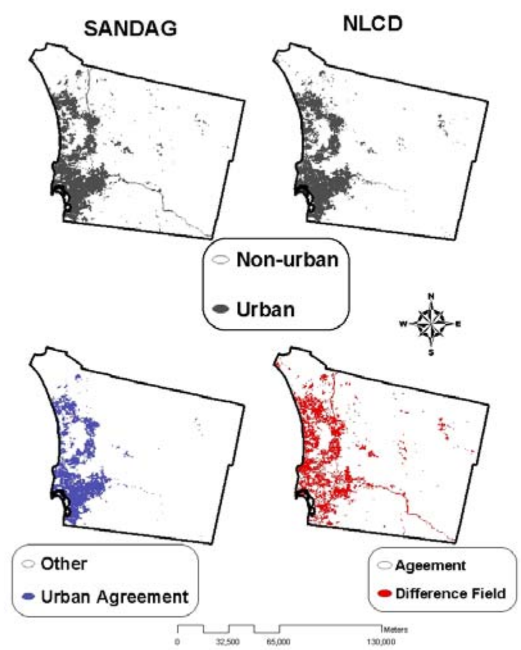

Figure 4. Comparison of the NLCD to the SANDAG dataset in the San Diego region.

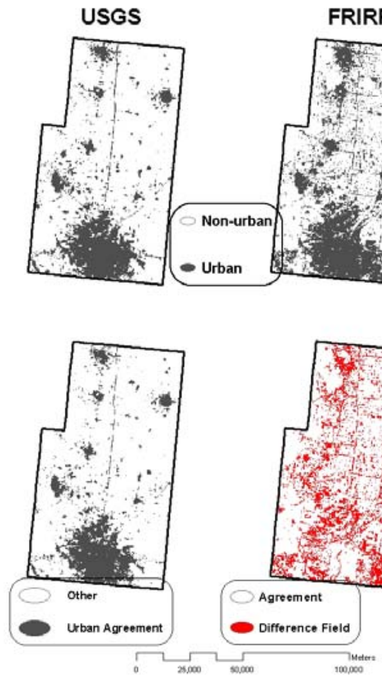

Figure 5. Comparison of the USGS LULC to the FRIRP dataset in the Denver region.

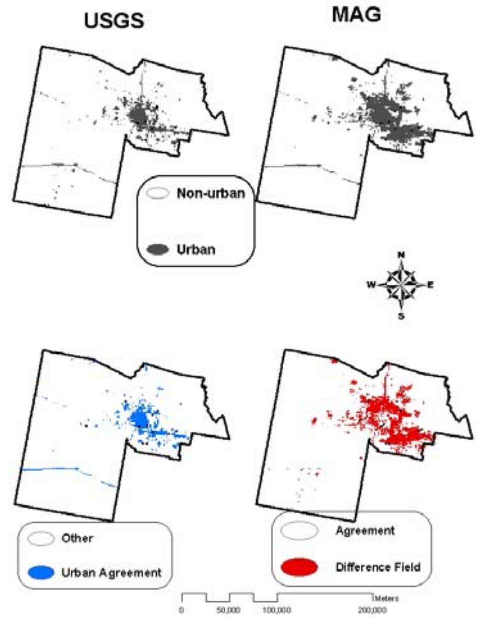

Figure 7. Comparison of the USGS LULC to the MAG dataset in the Phoenix region.

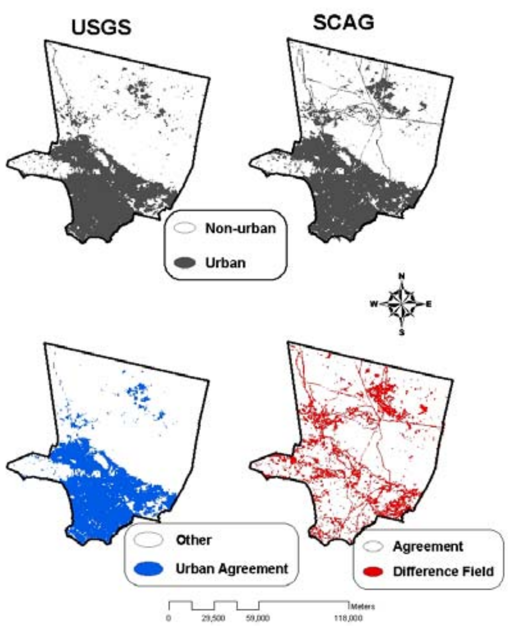

Figure 6. Comparison of the USGS LULC to the SCAG dataset in the Los Angeles region.

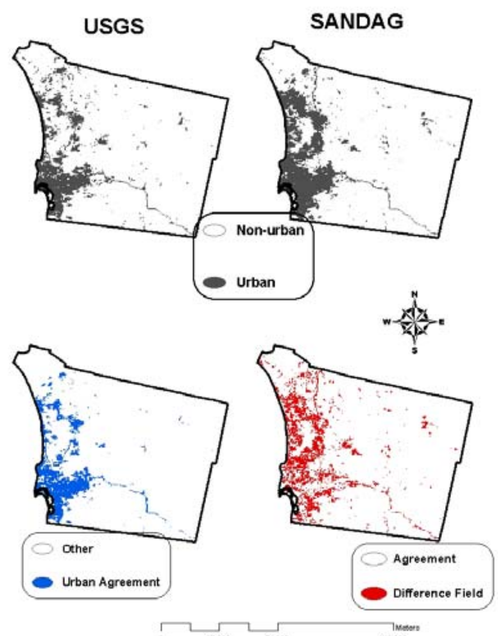

Figure 8. Comparison of the USGS LULC to the SANDAG dataset in the San Diego region.

area for urban land use in the two datasets, and a map of the differences in the classification of urban area. The difference field map contains both errors of omission (i.e., failure to identify an existing urban area) and errors of commission (i.e., falsely identifying a non-urban area as urban) for the NLCD 92 and USGS LULC data.

A cursory view of the top panels in Figs. 1-8 indicate reasonable agreement between the two datasets in urban land use coverage . For the most part, the NLCD 92 and LULC coverages appear to capture the extent of the urbanized area and the core urbanized zones fairly well. However, the difference fields in the bottom right panels reveal that there is significant disagreement, mostly on the fringes, but also in the interior. Table 3 lists the total urban area defined by each dataset. In all cases, the urban area defined by the national datasets are smaller than those defined by the high fidelity metropolitan datasets. The urban area defined in the NLCD 92 is 79 to 85 percent of the urban area defined by the metropolitan land use datasets. For the LULC data, Los Angeles and San Diego have urban areas within 10% of the urban area defined by the metropolitan land use datasets. The accuracy of the LULC, derived from 1970's images, becomes considerably worse in cities where rapid urbanization is a more recent phenomenon such as in Phoenix.

Table 3. Total urban area in each land use dataset

City	NLCD 92 Urban Area (m^2)	LULC Urban Area (m^2)	Metro Urban Area "truth" (m^2)	NLCD 92 to Metro ratio*	LULC to Metro ratio*
Los Angeles	2.43E+09	2.86E+09	2.96E+09	0.82	0.97
Phoenix	1.40E+09	8.16E+08	1.65E+09	0.85	0.50
Denver	1.23E+09	1.08E+09	1.50E+09	0.82	0.72
San Diego	8.63E+08	9.68E+08	1.09E+09	0.79	0.89

* Calculated as National Land use Urban Area/Metropolitan Land use Urban Area

Population change since 1970 helps to explain the ratio of the total urban areas defined in the LULC and metropolitan datasets for Phoenix and Los Angeles. Figure 9 depicts the change in the decadal census population in each of these cities normalized to the 1970 population. Phoenix's population has more than tripled since 1970. This rapid population change was accompanied by rapid urbanization, which explains the lower urban area total in the 1970s era LULC as compared to the metropolitan and NLCD 92 datasets for Phoenix. On the other hand, the LULC represents the total urban area well in Los Angeles. As shown in Figure 9, the population in Los Angeles has not risen considerably relative to the 1970 population and there has been less new urbanization in that time period. The use of 1970's imagery in the making of the USGS LULC is therefore adequate for estimation of the urban area in Los Angeles.

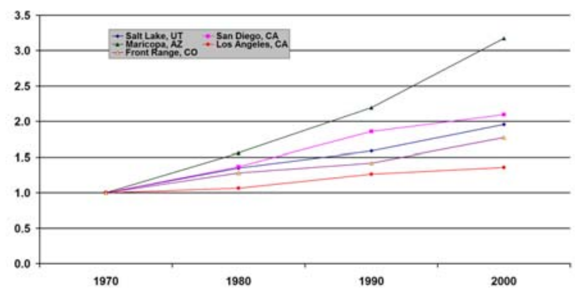

Figure 9. Population growth in select cities. All population numbers are normalized to 1970 population levels for the respective metropolitan region.

In Denver and San Diego, population change alone cannot be used to explain the differences between the urban area defined by the LULC and metropolitan datasets. Population increases by a factor of two in San Diego, and increases by a factor of 1.75 in Denver. This perhaps explains the underestimation of the urban area by the LULC for these two cities, but one would not expect the LULC-to-Metropolitan ratio to be lower for Denver as has been computed (Table 3). Two possible scenarios could explain this peculiarity. First, the elevated population change in San Diego could have occurred in areas that were already urban and hence resulted in little new urbanization. Second, the population increase in Denver could have been dominated by a high percentage of single-family homes, which would result in more new urban area as compared to homes in high density residential areas. Further historical research on the development of these two cities would be needed to determine the adequacy of these explanations.

Although the NLCD 92 and LULC datasets appear to capture much of the urban area of the cities studied, closer inspection indicates both the NLCD 92 and LULC are misclassifying large numbers of grid cells as either urban area when they are non-urban or as non-urban areas when they are urban. That is, the magnitude of the urban areas may be similar, but their locations are different. These dispersed misclassification errors represent a significant overall error when summed. In the next three tables, we look at the spatial correlation of the urban areas in the land use datasets.

Table 4 contains the areas of urban agreement between the LULC and metropolitan datasets and the NLCD 92 and metropolitan datasets (i.e., the urban agreement defines co-located areas where both datasets classify the land use as urban). For the four cities studied here, the NLCD 92 urban areas overlap only from 60-70 percent of the urban area as defined by the metropolitan land use datasets. This result is a bit surprising, as we expected the NLCD 92, derived relatively recently, to more accurately represent the urban spatial coverage. From 30 to 40 percent of the urban area in the metropolitan datasets are labeled as non-urban by the NLCD 92. These errors could be due to classification errors in the NLCD 92 or resolution differences between the metropolitan datasets and the NLCD 92. In urban regions, LANDSAT Thematic Mapper pixels (upon which the NLCD 92 is based) often contain multiple land uses, which can confound land use classification. For example, Epstein *et al.*, (2002) found vegetative canopies often obscured underlying urban structures resulting in land use misclassification when using remotely sensed imagery and unsupervised classification algorithms. The sub-pixel differences in land use can produce a land use dataset that misclassifies a considerable quantity of land on aggregate. Smith *et al.* (2002) found that land cover heterogeneity and small patch size (i.e., low numbers of contiguous pixels classified as the same land cover) increased the classification error in the NLCD 92. Urban areas are often characterized by high land cover heterogeneity and small patch sizes.

Resolution differences between the metropolitan datasets and the NLCD 92 could also account for the high error in urban area classification. The metropolitan datasets are polygon-based with minimum size requirements that are often greater than the 30 meter grid cell size in the NLCD 92. Small vegetated or barren areas within the city (e.g., big lawns, small parks, or vacant lots) that are on the order of the 30 m grid cell size are often denoted to be non-urban by the automated routines used to define the NLCD 92 land use types. In the metropolitan datasets, these small vegetated or barren areas are likely lumped into an urban land use polygon. That is, land use in the metropolitan datasets may have been determined using majority filters, i.e., the land use was defined using the predominant land use in the polygon. Therefore, large polygons of urban land use in the metropolitan datasets may contain several 30 m grid cells that are non-urban in the NLCD 92. These differences may add up to a significant area, which could account for the large differences noted in Table 4.

Table 4 also shows that the LULC intersects, on average, a similar amount of the urban area as the NLCD 92 dataset. However, there is significant variation from city to city, with urban agreement ranging from 39 to 82 percent. The larger range is due to the high errors of omission in the LULC in some cities. This high error is not due to any inherent flaw in the methods used to generate the LULC, but is the result of the age of the LULC data. The USGS LULC is based on high-altitude aerial photographs from the 1970s and early 1980s. In cities, where urbanization has been a recent trend, e.g., Phoenix and Denver, the LULC represents the urban area poorly. In Los Angeles, where much of the development pre-dates 1970-1980, the LULC performs moderately better than the NLCD 92. This may be due to the aforementioned resolution differences between the metropolitan datasets and the NLCD 92 coupled with differences in the classification methods for the underlying land uses. The polygon-based LULC may not have these resolution and classification differences. The LULC, like the metropolitan datasets, uses minimum polygon sizes and therefore small vegetated or barren areas within urban areas would be considered part of a larger urban land use polygon.

Table 4. Area of urban overlap between metropolitan and regional datasets.

City	Metropolitan Dataset: Total Urban Area (m2)	NLCD 92 Area of Urban Agreement (m2)	LULC: Area of Urban Agreement (m2)	NLCD 92:% Urban Agreement	LULC: % Urban Agreement
Los Angeles	2.96E+09	2.08E+09	2.43E+09	70	82
Phoenix	1.65E+09	1.13E+09	6.47E+08	69	39
Denver	1.50E+09	1.01E+09	8.83E+08	67	59
San Diego	1.09E+09	6.69E+08	6.91E+08	61	63
Mean				67	61

* Calculated as the area of the co-located urban area in the National Land use data and the Metropolitan Land use data/Urban area in the Metropolitan data

Tables 5 and 6 lists the commission and omission error rates in the LULC and NLCD 92 datasets relative to the metropolitan dataset. The commission error fraction and the omission error fraction are calculated according to equations 1 and 2, respectively,

$$A_C / A_{Ur} \qquad (1)$$

$$A_O / A_{Um} \qquad (2)$$

where A_c is the area of commission error between the national dataset and the metropolitan dataset, A_o is the area of omission error between the national dataset and the metropolitan dataset, A_{ur} is the urban area in the national dataset, and A_{um} is the urban area in the metropolitan dataset. The commission error is akin to a Type II error or a false positive, i.e., the national dataset is denoting an area as urban, when it is, in fact, non-urban. The omission error is akin to a Type I error or a false negative, i.e., the national dataset is denoting an area as non-urban, when it is, in fact, urban. The development of a consistent, accurate, and easily obtainable land use dataset for urban areas in the United States requires the minimization of both the commission error and omission error.

Table 5. Urban land use error fraction in the LULC dataset.

City	LULC Commission error	LULC Omission error
Denver	0.18	0.41
Los Angeles	0.14	0.18
Phoenix	0.21	0.61
San Diego	0.28	0.37

As noted in Table 5, the omission error in the LULC is consistently greater than the commission error. This indicates that users of the LULC for urban level studies will be excluding significant segments (18-61%) of the city from their analyses. The commission error shows that in the LULC the probability of an area defined as urban actually being non-urban is 0.14 – 0.28. In tandem, these errors could significantly affect transport and dispersion calculations for particular simulations.

Table 6 indicates the omission rate is also greater than the commission rate in NLCD 92 data. Users of the NLCD 92 will also be excluding large areas of the city from their analyses and misclassifying 14-22% of the urban area.

Table 6. Urban land use error fraction in the NLCD 92.

City	NLCD 92 Commission error	NLCD 92 Omission error
Denver	0.18	0.33
Los Angeles	0.14	0.30
Phoenix	0.19	0.31
San Diego	0.22	0.39

As noted above, there are considerable errors in city coverage in the NLCD 92 and LULC datasets. In general, the NLCD 92 defines the urban footprint better than the LULC except in cities where the majority of the urban development occurred prior to 1970. The differences between the two datasets are most likely due to errors in the analytical methods used to create the datasets as well as the age of the data used in the analysis. In this section, we compare the two national datasets and investigate a mechanism for minimizing the error in the urban coverage in the two datasets.

Figures 10-13 show the comparison of the LULC and NLCD 92 for the cities studied. Each plot contains a map of the area of urban agreement with errors of commission and omission between the two datasets. In these figures, the LULC was assessed relative to the NLCD 92, so an error of omission is a region defined as urban in the NLCD 92 and non-urban in the LULC and an error of commission is a region defined as non-urban in the NLCD 92 and urban in the LULC. In each city, the LULC and NLCD 92 show good agreement in the definition of the older sections of the city, but considerable areas within and on the outskirts of these cities are defined differently in the two datasets. The differences between each dataset indicate the potential drawback of using the USGS LULC due to its age. At the outskirts of each city, except Los Angeles, there are large areas of omission error most likely due to the inability of the USGS LULC to capture recent development. This effect is particularly evident in Phoenix and Denver.

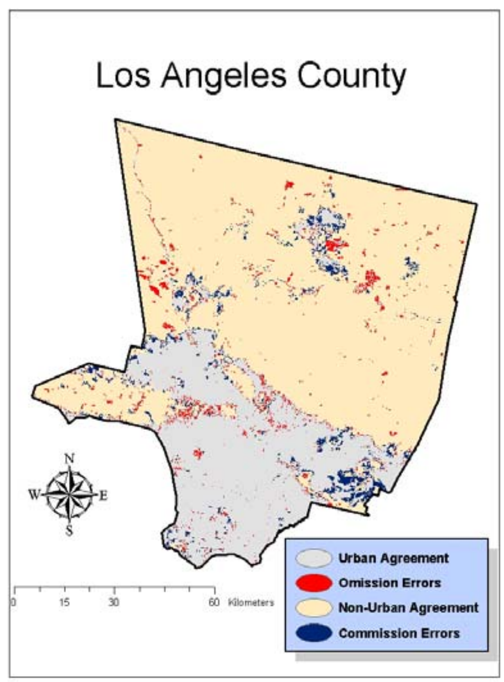

Figure 10. Comparison of NLCD 92 data to the USGS LULC in Los Angeles. Omission Errors – NLCD 92 urban and LULC non-urban; Commission Errors – NLCD 92 non-urban and LULC urban.

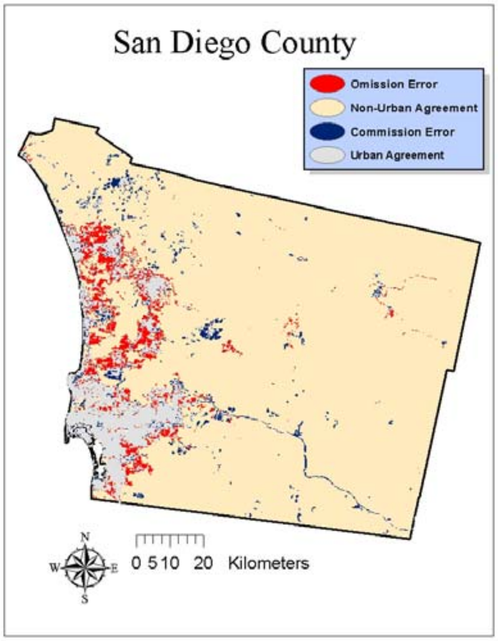

Figure 11. Comparison of NLCD 92 data to the USGS LULC in San Diego. Omission Errors – NLCD 92 urban and LULC non-urban; Commission Errors – NLCD 92 non-urban and LULC urban.

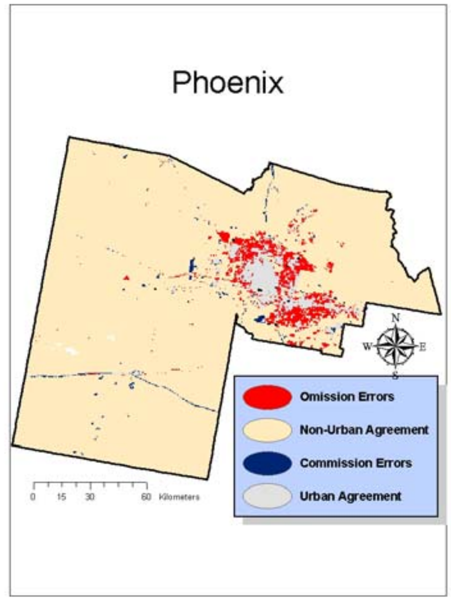

Figure 13. Comparison of NLCD 92 data to the USGS LULC in Phoenix. Omission Errors – NLCD 92 urban and LULC non-urban; Commission Errors – NLCD 92 non-urban and LULC urban.

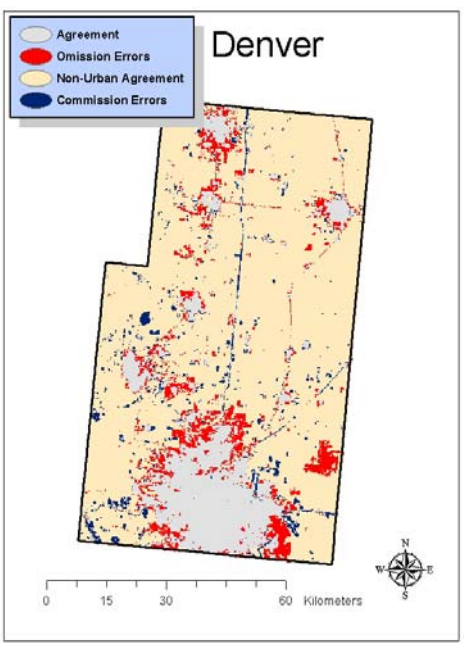

Figure 12. Comparison of NLCD 92 data to the USGS LULC in Denver. Omission Errors – NLCD 92 urban and LULC non-urban; Commission Errors – NLCD 92 non-urban and LULC urban.

Although each dataset has errors that make their application in urban level analyses problematic, the combined strength of the datasets may define the urban footprint more accurately. To test this hypothesis, the NLCD 92 urban footprint and the USGS LULC urban footprint for each city were combined and then compared to their respective metropolitan land use dataset. Table 7 displays the area of the urban footprint from each metropolitan land use dataset and the percentage of that area captured by the NLCD 92, the LULC and the union of the NLCD 92 and LULC. The combined regional dataset intersects, on average, 10 to 15% more of the metropolitan urban footprint than each regional dataset individually. This improvement in urban area agreement is due to a reduction in the omission error of urban land use in the combined dataset.

Table 7. Overlap between Metropolitan urban area and the NLCD 92, LULC, and NLCD 92 + LULC urban areas.

City	Metro Land use Urban Footprint Area	% NLCD 92 Urban Agreement	% USGS LULC Urban Agreement	% Combined NLCD 92-USGS Urban Agreement
Denver	1.50E+09	67	59	76
Los Angeles	2.96E+09	70	82	87
Phoenix	1.65E+09	69	39	74
San Diego	1.09E+09	61	63	81
Mean		67	61	79

Table 8 displays the omission error fraction for the LULC, NLCD 92 and LULC+NLCD 92 relative to the metropolitan datasets. In each city, the combined dataset has a lower omission error than either the NLCD 92 or LULC. Regretfully, the improved capture of urban area in the combined dataset comes at a price of higher rates of commission error as noted in Table 9. In future research, we will attempt to minimize the commission error rate by evaluating the specific urban land uses that contribute to it.

Table 8. Fractions of Urban Omission in the LULC, NLCD 92, and LULC+NLCD 92 datasets.

City	LULC-NLCD 92 Omission error	LULC Omission error	NLCD 92 Omission error
Denver	0.24	0.41	0.33
Los Angeles	0.13	0.18	0.30
Phoenix	0.26	0.61	0.31
San Diego	0.19	0.37	0.39

Table 9. Fractions of Urban Commission in the LULC, NLCD 92, and LULC+NLCD 92 datasets.

City	LULC-NLCD 92 Commission error	LULC Commission error	NLCD 92 Commission error
Denver	0.23	0.18	0.18
Los Angeles	0.17	0.14	0.14
Phoenix	0.23	0.21	0.19
San Diego	0.30	0.28	0.22

7. Summary

The NLCD 92, derived from 1990s satellite imagery, predicted the urban area on average better than the LULC dataset, which was based on 1970s aerial photos. The NLCD 92 was also more consistent in its predictions accounting for 61-70% of the urban area in the metropolitan datasets, while the LULC predicted 39 - 82% of the urban area in the metropolitan datasets. Although the NLCD 92 performed better than the LULC on average, the LULC was more accurate for older cities such as Los Angeles. The accuracy of the LULC was low in cities that have undergone recent urbanization. The combination of the urban area defined by the two national land use datasets was found to improve the accuracy of the urban footprint by reducing the omission error (i.e., reducing the amount of land classified as non-urban when it was in fact urban) in the two datasets. While the omission error was reduced by the combination of the two datasets, the commission error (amount of land incorrectly classified as urban when it was non-urban) was increased. Users of this approach should therefore weigh the relative importance of the two error types prior to using this technique to define the urban footprint.

8. References

Akbari H. and L.S. Rose, 2001a. Characterizing the Fabric of the Urban Environment: A Case Study of Salt Lake City, Utah. LBNL-47851. Heat Island Group, Environmental Energy Technologies Division, Lawrence Berkeley National Laboratory.

Akbari H. and L.S. Rose, 2001b. Characterizing the Fabric of the Urban Environment: A Case Study of Metropolitan Chicago. LBNL-49275. Heat Island Group, Environmental Energy Technologies Division, Lawrence Berkeley National Laboratory.

Anderson, J. R., Hardy, E. E., Roach, J. T. and R.E. Witmer, 1976. A land use and land cover classification system for use with remote sensor data. USGS Professional Paper 964, U.S. Geological Survey.

Burian, S. and M. Brown. 2001. Summary of Los Angeles, California Building and Land Use/Land Cover Data, LA-UR-01-4055, 46 pp.

Burian, S.J., M.J. Brown, and T.N. McPherson, 2001. Evaluation of land use/land cover datasets for urban watershed modeling. Water Science and Technology, 45(9): 269-276.

Congalton, R. 1997. Exploring and Evaluating the Consequences of Vector-to-Raster and Raster-to-Vector Conversion. Photogrammetric Engineering and Remote Sensing 63, 4, 425-434.

Epstein, J., K. Payne, and E. Kramer. 2002. Techniques for Mapping Suburban Sprawl. Photogrammetric Engineering and Remote Sensing 63, 9, 913-918.

Smith, J.H., J.D. Wickham, S.V. Stehman, and L. Yang. 2002. Impacts of Patch Size and Land-Cover Heterogeneity on Thematic Image Classification Accuracy. Photogrammetric Engineering and Remote Sensing 63, 1, 65-70.

USGS (U.S. Geological Survey). 1990. Data users guides. Department of the Interior, Reston, VA.

Vogelmann, J.E., T.L. Sohl, P.V. Campbell, and D.M. Shaw, 1998. Regional land cover characterization using Landsat Thematic Mapper data and ancillary data sources. Environmental Monitoring and Assessment, 51: 415-428.

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.