Mapping Urban Areas by Fusing Multiple Sources of Coarse Resolution Remotely Sensed Data

Annemarie Schneider, Mark A. Friedl, Douglas K. McIver, and Curtis E. Woodcock

Abstract
In recent decades, rapid rates of population growth and urban expansion have led to widespread conversion of natural ecosystems and agricultural lands to urban land cover. The amount and rate of this land conversion affects local and regional ecosystems, climate, biogeochemistry, as well as food production. The main objective of the research described in this paper is to improve understanding of the methodological and validation requirements for mapping urban land cover over large areas from coarse resolution remotely sensed data. A technique called boosting is used to improve supervised classification accuracy and provides a means to integrate MODIS data with the DMSP nighttime lights data set and gridded population data. Results for North America indicate that fusion of these three data types improves urban classification results by resolving confusion between urban and other classes that occurs when any one of the data sets is used by itself. Traditional measures of accuracy assessment as well as new, maplet-based methods demonstrate the effectiveness of the methodology for creating maps of cities at continental scales.

Introduction
Urban populations have exploded in the last three decades, with nearly 50 percent (3 billion) of the Earth’s inhabitants currently living in cities (United Nations, 2000). Further, while urban land cover presently accounts for less than one percent of the Earth’s land area, this proportion is growing rapidly as more and more cities expand into natural ecosystems and agricultural areas (Miller, 1988; Douglas, 1994). While urbanization cannot be halted, identifying and anticipating the location, size, and growth rate of urban areas is an important component to understanding and mitigating many aspects of global population growth and, by extension, global change.

Previous studies of urban areas from remote sensing have consistently relied on fine resolution data (Landsat, SPOT), which limits these studies to small areas (Jensen and Toll, 1982; Paola and Schowengerdt, 1995; Seto et al., 2001). The only existing continental scale maps of urban areas are the Digital Chart of the World (DCW) urban layer (Danko, 1992), and maps derived from the Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS) “low light” sensor, or nighttime lights data (Elvidge et al., 1996). While the DCW urban data are extremely valuable, they were compiled from Operational Navigation Chart data from the 1960s and no longer provide an accurate representation of the current size of many cities. The nighttime lights data have produced stunning images of urban areas for many years. However, for mapping purposes, the data possess coarse (2.7 km) resolution, are poorly registered, and exhibit “blooming” effects that inflate city boundaries (Elvidge et al., 1996; Imhoff et al., 1997).

Recent coarse resolution land-cover mapping efforts from remote sensing data have not included cities because of the limitations of Advanced Very High Resolution Radiometer (AVHRR) imagery, lack of quality training data, and the inadequacies of traditional classification algorithms (Loveland et al., 2000; Hansen et al., 2000; Friedl et al., 2002). Data from NASA’s recently launched Terra platform offer a new opportunity for mapping urban areas, providing the potential to transform such studies from local to global scales. In particular, Terra’s Moderate Resolution Imaging Spectroradiometer (MODIS) acquires multispectral data with high temporal frequency, and possesses seven spectral bands designed specifically to monitor land processes and properties at regional to global scales. Unfortunately, classifications based on MODIS data alone result in confusion between urban and barren areas. This paper describes ongoing efforts to produce reliable representations of urban areas at 1-km resolution, as part of a larger project mapping global land cover from MODIS data.

The primary goal of this paper is to describe and assess a methodology for mapping urban land cover at a 1-km spatial resolution by fusing multiple sources of coarse resolution data. Two major tasks were involved in this study. First, a supervised decision tree classification method was developed by fusing 1-km MODIS data and two ancillary sources: the nighttime lights data (Elvidge et al., 1999) and gridded population density data (Töbner et al., 1995; Deichmann et al., 2001). The second task was to establish the best means for evaluating the accuracy of urban land-cover maps produced over large regions, an issue that is especially problematic when the class of interest is a small fraction of the total area mapped.

Background
Remotely sensed data offer advantages over traditional methods for mapping cities, studying urban morphology and growth, and understanding the effects of human activity on the environment. In the last decade, several studies have used remote sensing as a basis for mapping vegetated land cover at continental scales (e.g., Loveland et al., 1999). However, none of these investigations have derived urban land cover from remote sensing sources. The most commonly used map of urban
areas is the Digital Chart of the World populated places data (DCW) (Danko, 1992). Unfortunately, this data set is inconsistent globally, and often does not include newer cities or new growth from recent decades.

A variety of efforts have used nighttime lights data to map urban areas (Welch, 1980; Elvidge et al., 1996; Imhoff et al., 1997). Recently, a calibrated radiance data set was created using atmospherically corrected DMSP-OLS data acquired from 1996 to 1997. This product uses a new variable gain control to accommodate the wide range of light intensities encountered around the globe (Elvidge et al., 1999). While the data have not been validated for accuracy (Elvidge et al., 1999; D. Stutzer, pers. comm.), the images provide a striking depiction of human activity on the planet. It is important to recognize that maps derived from DMSP-OLS data provide a representation of light, but do not necessarily represent the built environment or settlement patterns. In particular, brightly lit agricultural areas and non-urban light sources such as gas flares and fires are captured in these data sets. Despite these problems, the data have received much attention from both researchers and the media, because they provide one of the only recent attempts to map cities on a global scale (Sutton, 1997; Owen et al., 1998; Doll et al., 2000).

As large area mapping with Landsat data becomes increasingly common, a third type of continental-scale map has become available. The CORINE (Coordination of Information on the Environment) Programme completed a remote-sensing-based, fine scale map of land cover for the European Union (CORINE, 1994). Similarly, the MultiResolution Land Characteristics Consortium recently completed an unsupervised classification of Landsat data for the U.S. known as the National Land Cover Data (NLC data) (Vogelmann et al., 1998; Vogelmann et al., 2001; Yang et al., 2001). Both of these data sets include cities, but three key problems arise for map users. First, global studies are impossible because the maps are limited to a few countries. Second, class definitions and map quality are not consistent. Third, the maps are produced at fine spatial scales, sacrificing manageability of the data set at continental scales.

Methods

Overview

In the sections that follow, we describe a methodology for creating continental scale maps of urban areas using multiple sources of input data. For this work we limit our analyses to data from North America. The method involved three main steps, as shown in Figure 1. In the first step, the nighttime lights data and gridded population density data were combined in a logistic regression model to produce a probability surface for urban areas. In the second step, a decision tree algorithm was trained using a global set of training sites for 17 land-cover classes (including urban) defined by the International Geosphere-Biosphere Program (IGBP), and the trained tree was applied to the MODIS data. The output from this stage provided a map of per-pixel probabilities for each of 17 classes. The class probabilities and the probability surface were then used as input to the third step, where Bayes’ Rule was applied at every pixel. To do this, the probabilities of urban areas derived from the logistic regression were used as a \textit{prior} probabilities, and the final pixel label was assigned based on the maximum likelihood derived from the \textit{a posteriori} probabilities. In this way, information from all three data sources was fused to create a final map of urban areas. Details regarding each of the three steps are provided below.

Data

The first step, creating the probability surface, required three sets of data: (1) nighttime lights data, (2) gridded population density data, and (3) NLC data. The second step, deriving classification probabilities, relied on two sets of data: (1) MODIS imagery and (2) training sites (both urban and non-urban) selected by the MODIS Landcover team at Boston University.

The 1-km resolution nighttime lights data (Elvidge et al., 1999) and the gridded population data at a 2.5-minute (about five kilometers) resolution (Tobler et al., 1995; Deichmann et al., 2001), resampled to one kilometer (see Figure 2) were used as inputs to a logistic regression model to create a map of “prior probability” for urban areas. To reduce blooming effects and lessen confusion in the logistic model, the nighttime lights data were thresholded using a fixed radiance value. Based on extensive visual inspection of North American cities, a value of $89.4 \times 10^{-10}$ watts/cm$^2$/sr/$\mu$m was identified as the most effective threshold for discriminating less urbanized, less populated areas outside cities from urban areas. Note that, for areas with radiance values above this threshold, continuous values were retained. Beyond this threshold, the lit area becomes fragmented and urban areas are excluded.

A sample of 4000 1-km$^2$ pixels were selected randomly from the nighttime lights and gridded population density data, and a binary urban/non-urban class map was produced from the 30-m-resolution NLC data (Vogelmann et al., 2001). This latter data set served as the dependent variable to train the logistic model. The urban class defined in the NLC maps was produced by manual digitization of Landsat data and aerial photos (Vogelmann et al., 1998), and is considered a reliable and consistent source of data for cities in the U.S. This database provides three categories of urban land cover: (1) commercial/industrial, (2) high intensity residential, and (3) low intensity residential. For this work these classes were aggregated into a single urban class to ease comparison.
The classification stage in the second step employed a full year of MODIS imagery acquired between 15 October 2000 and 15 October 2001. These data included seven bands of cloud cleared, atmospherically corrected Nadir BRDF-Adjusted Reflectance values (Lucht et al., 2000), and the MODIS Enhanced Vegetation Index (EVI) composited over 16-day periods. The MODIS classification methodology was designed to utilize a full year of data to exploit temporal differences in land-cover types (Strahler et al., 1999; Friedl et al., 2002).

The decision tree algorithm was trained using 1700 training sites ranging from 1 to 100 km² in area, classified according to the IGCB system (Belward and Loveland, 1997). Each site was obtained by manual interpretation of Landsat data, and was assessed for interpretation errors by two or more analysts. Urban training sites were selected from cities and suburbs using the criteria that the area in question must be composed of greater than 50 percent urban land cover.

**Estimating a Probability Surface Using Logistic Regression**

To estimate the *a priori* probabilities for the presence of urban land cover, three logistic regression models were constructed. The first model used the gridded population data as the sole predictor variable, the second used the nighttime lights data, and the third included both ancillary data sources as predictor variables. The model of urban/non-urban areas constructed in this way possessed the form

\[
P_{\text{urban}} = \frac{\exp(U_i)}{1 + \exp(U_i)}
\]

where \( P_{\text{urban}} \) is the probability that a pixel is urban or non-urban. The value of \( U_i \) is provided by a conventional linear regression model. For example, the model using both the nighttime lights and the gridded population data to predict urban/non-urban is

\[
U_i = \beta_0 + \beta_1 \text{population}_i + \beta_2 \text{lights}_i + e_i
\]

where \( \text{population}_i \) is the log of the value of the population density for the \( i \)th pixel, \( \text{lights}_i \) is the value of the nighttime lights data at the \( i \)th pixel, the \( \beta \) coefficients are determined empirically using the logistic regression subroutine available in the Splus Design/Hmisc library (Alzola and Harrell, 2003), and \( e_i \) is the random error term.

Logistic regression model results were evaluated using a variety of diagnostic statistics. The statistical significance of the regression coefficients was assessed by testing the null hypothesis \( \beta = 0 \) using a *t* statistic. Values that exceeded a critical threshold (\( p < 0.05 \)) indicated that the coefficient was statistically significant. The statistical significance of the model was analyzed using a likelihood ratio statistic (\( \omega \)), which is \( \chi^2 \) distributed with \( k - 1 \) degrees of freedom (Hastie and Tibshirani, 1990; Hosmer and Lemeshow, 2000). This test statistic is given by

\[
\omega = -2(\log L_0 - \log L_1)
\]

where \( L_1 \) is the value of the likelihood function for the full model (estimated by Equation 1) and \( L_0 \) is the value of the intercept. An additional test statistic, the pseudo R² (Estrella, 1998), was used to provide a measure of fit: i.e.,

\[
\text{pseudo R}^2 = 1 - \frac{\log L_0}{\log L_1}
\]

where \( L_0 \) and \( L_1 \) are defined above.

The results for the three models show that each of the coefficients was significant at \( p < 0.05 \). The final model was chosen based on the lowest likelihood ratio statistic, and the highest pseudo R², as shown in Table 1. The difference between the various models was not large, although the model that combined both the nighttime lights and the population information had the lowest log likelihood score and the highest pseudo R² (model three). Using this model, a probability surface of urban/non-urban land cover was estimated for each 1-km pixel in North America.

**Supervised Classification**

**Decision Tree Classification**

The method employed to compute conditional probabilities of IGCB land-cover classes relies on a supervised decision tree classification algorithm (C4.5) that has been widely used in the machine learning community (Quinlan, 1993). Decision trees have recently received increased attention for remote sensing problems, and have been especially effective for coarse resolution classifications because they are able to handle noisy or missing data, they require no *a priori* assumptions regarding the distribution of the input data, and they can handle complex, nonlinear relations between features and classes (Fayyad and Irani, 1992; Friedl and Brodley, 1997).

Decision tree algorithms employ a hierarchical classification procedure that recursively partitions a data set into smaller subdivisions. The “tree” is composed of a root node, a set of internal nodes, and a set of terminal or leaf nodes. The splits defined at each internal node are estimated from training data using a statistical “learning” procedure that creates more homogeneous subsets of the data. In C4.5, each split is determined by a metric called the information gain ratio, which measures the reduction in entropy in the data produced by a split. Using this metric, decisions are made at the split that maximize the reduction in entropy of the training data in the descendent nodes. (For detailed information, please refer to Quinlan (1993).) The splitting process continues until each leaf node contains only observations from a single class or no gain in information results from further splitting. The class label is then assigned to the terminal leaf node into which the observation falls.

---

**Table 1. A Comparison of Results from Three Logistic Regression Models Used to Estimate an Urban/Non-Urban Probability Surface**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model One p Value</th>
<th>Model Two p Value</th>
<th>Model Three p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 )</td>
<td>-3.45 &lt; 0.001</td>
<td>-2.11 &lt; 0.001</td>
<td>-2.58 &lt; 0.001</td>
</tr>
<tr>
<td>( \beta_{\text{pop}} )</td>
<td>0.86 &lt; 0.001</td>
<td>0.11 &lt; 0.001</td>
<td></td>
</tr>
<tr>
<td>( \beta_{\text{lights}} )</td>
<td>0.12 &lt; 0.001</td>
<td>0.18 &lt; 0.001</td>
<td></td>
</tr>
<tr>
<td>log likelihood</td>
<td>2295.00 &lt; 0.001</td>
<td>1910.10 &lt; 0.001</td>
<td>1580.50 &lt; 0.001</td>
</tr>
<tr>
<td>pseudo R²</td>
<td>0.08</td>
<td>0.29</td>
<td>0.45</td>
</tr>
</tbody>
</table>

Figure 2. Examples of the population density data and the nighttime lights data for the Baltimore-Washington, D.C. area. White areas indicate high population density and high radiance levels, respectively.
A crucial step in the use of a decision tree is to correct the tree for overfitting by pruning the leaf nodes. Training data often contain noise, which decision trees split into leaf nodes (Mingers, 1989; Quinlan, 1987). This results in overfitting to noise in the data. The C4.5 algorithm uses error-based pruning to minimize this effect (Quinlan, 1993).

**Boosting**

A relatively new technique known as “boosting” has been widely shown to increase classification accuracies using supervised algorithms. Recent work in the remote sensing community has confirmed that boosting is effective for land-cover mapping with remotely sensed data, because it reduces misclassification errors and is resistant to overfitting (Friedl et al., 1999; McIver and Friedl, 2001). Boosting improves classification accuracy by estimating multiple classifiers while systematically varying the training sample. At each iteration of the classifier, the training sample is modified to focus the classification algorithm on examples that were difficult to classify in the previous iteration. The final classification is then produced by an accuracy weighted vote across all of the classifications (Quinlan, 1996). Previous work in the remote sensing domain suggests that accuracy tends to stabilize by about ten boosting iterations (Friedl et al., 1999; McIver and Friedl, 2001).

Recently, a statistical examination of boosting has shown that this technique is equivalent to a form of additive logistic regression (Friedman et al., 2000; Collins et al., 2002). As a result, probabilities of class membership can be assigned for each class at every pixel. Using these probabilities, it is possible to calculate a posteriori probabilities of urban areas with the aid of ancillary information. Ancillary data are especially helpful in coarse resolution classification problems, where landscapes may contain high levels of within-class variability, or where the remotely sensed data provide relatively weak separability among classes (McIver and Friedl, 2002).

The theoretical basis for computing a posteriori probabilities is Bayes’ Rule, which is derived from the definition of conditional probability (Robert, 1997). For this work, Bayes’ Rule was applied at every pixel to combine the probability surface (a priori information) derived from the nighttime lights and population data with the conditional probabilities derived from the boosted decision trees.

**Results**

Representative results for urban areas in North America created by fusing nighttime lights and population data with the MODIS decision tree results (hereafter, the fusion map) are presented in Figure 3. While it is difficult to assess the quality of the results at continental scales, regional views of the Pacific Northwest, the East Coast, California, and Texas show that the sizes and shapes of cities are in good agreement with the expected urban morphology of each region.

Plates 1 and 2 each present five panels showing the urban class at the scale of individual cities. The two urban areas presented in these figures, Baltimore-Washington, D.C. and the San Francisco Bay area, each possess different regional land-cover characteristics.

---

**Figure 3.** Four regional views of the fusion map of urban areas, shown in grey. From the upper left is (a) the Pacific Northwest showing Vancouver, Seattle, and Portland, (b) the East Coast urban corridor, (c) southern California showing the Los Angeles-San Diego conurbation, and (d) southeastern Texas, which includes the sprawling cities of Dallas and Houston.
Plate 1. The Baltimore—Washington, D.C. area showing (a) fine resolution Landsat TM imagery (urban areas appear purple), (b) 1-km MODIS data (urban also appears purple), (c) the urban data from the DCW, (d) the nighttime lights data (orange boundary represents threshold), and (e) the fusion map of urban areas.

Plate 2. The San Francisco Bay area showing (a) fine resolution Landsat TM imagery (urban areas appear purple), (b) 1-km MODIS data (urban also appears purple), (c) the urban data from the DCW (note misregistration), (d) the nighttime lights data (orange boundary represents threshold), and (e) the fusion map of urban areas.
characteristics, as well as areas of new growth not captured by the DCW. In both figures, Panel (a) presents Landsat TM imagery for each area, where settlement patterns can be seen intermixed with forested and agricultural areas. Similar patterns can be seen in the MODIS imagery in Panel (b), but with substantially less spatial detail. Panel (c) presents the DCW urban boundary layer. The DCW data do a good job of characterizing the core urban area, but fail to capture expanses of new growth and suburbs. Panel (d) presents the nighttime lights data set. While maps from the thresholded DMSP-OLS data appear reasonable at continental scales, Panel (d) clearly illustrates how this data set leads to overestimation of urban areas. Finally, Panel (e) presents the results achieved by fusing MODIS data with nighttime lights and population density data.

Visual inspection suggests that the decision tree/prior probability methodology works quite well. At local scales, the classified maps (Plates 1e and 2e) are in general agreement with the DCW maps. However, some differences are apparent. For example, low density residential areas that are not present in the DCW maps southwest of Baltimore are clearly visible in the new map (Plates 1a, 1c, and 1e). Similarly, in the San Francisco Bay area, the fusion map provides a depiction of urban areas that includes areas missing in the DCW (both near the bay and to the east), but which does not have the extensive blooming associated with the nighttime lights data alone (Plates 2a, 2c, and 2e).

**Accuracy Assessment: Issues for Coarse Resolution Maps**

Effective use of classified maps requires thorough analysis and quantification of map quality. The conventional approach to accuracy assessment is to compare an independent, random sample of “ground truth” points to the classified maps (Card, 1982; Congalton, 1991). Results are then tabulated in an error matrix, which yields overall map accuracy. User’s and producer’s accuracies, which measure errors of commission or omission, respectively, are often estimated (Congalton, 1991; Stehman, 1997).

For this work, we used an independent test set comprised of approximately 400 1-km² pixels. Practical considerations precluded the use of a formal sample design (Fitzpatrick-Lins, 1981; Curran and Williamson, 1985; Stehman, 2001), thereby reducing the reliability of the results. Producer’s accuracy was used instead of an overall accuracy estimate because the urban class represents only one percent of the total land area. Results show the degree to which the a priori information corrects errors in the urban class. In particular, when the a priori probabilities derived from the nighttime lights and population data were not used, the producer’s accuracy was below 30 percent. When a priori probabilities were included, the producer’s accuracy increased to 79 percent.

The fusion map of urban land cover was also compared against the DCW urban data (Danko, 1992), the radiance calibrated nighttime lights data (Elvidge et al., 1999), and a regional, fine resolution map of urban areas: the NLC data (Vogelmann et al., 2001). For consistency and to ease comparison, the same threshold was applied to the lights product. Three maplet-based methods were used to supplement the overall accuracy statistics (Stoms, 1996; Cihlar et al., 2000). While overall measures are still important, the maplet-based methods provide a complementary means of assessment. Specifically, the maplet methods used here are designed to check for errors in both city size and city location, highlighting the strengths and weaknesses of each data set by representing current urban land cover.

**Per-Pixel Comparison**

First, the per-pixel agreement between the NLC data (aggregated to 1 km using a simple majority rule) and the three maps of urban land cover was examined at the level of individual cities. Visual inspection of the maps revealed that, at least qualitatively, many of the same spatial patterns present in the high-resolution NLC data appear in the fusion map (Plate 3). In particular, Plate 3d shows the agreement when the map produced in this work was combined with the NLC data, with errors of commission in pink and errors of omission in black. Assessment of a sample of cities across the U.S. revealed similar results. The urban cores were mapped correctly, but misclassified areas along city boundaries suggest that the urban perimeter is difficult to map using coarse resolution data. Reprojection problems and issues associated with the aggregation of the NLC data to 1 km may explain a significant amount of this disagreement. The aggregated map (Plate 3c) is especially fragmented compared to the finer resolution map, illustrating that some of the “errors of commission” in the fusion map may be considered urban by conventional definitions.

To provide a more general assessment, error matrices between the NLC data and the three maps were computed for ten cities, which were chosen to reflect a range of sizes and geographic distribution. Overall percent agreement was calculated for two classes, urban and non-urban (Figure 4). As expected, agreement between the DCW and the NLC data is similar to that for the fusion map, ranging from 80 to 95 percent. The DCW data show slightly higher agreement in nearly all cities relative to the fusion map. Note, however, that this result is somewhat misleading because the out-of-date DCW systematically underestimates urban areas, and no DCW urban pixels fall outside the more current boundaries defined by the NLC data. Visual comparison of the DCW and the NLC data (Plates 1 and 2) supports this conclusion, where areas of new urban growth are missing in the DCW data, but are captured in the fusion map. The fusion map, on the other hand, captures city size and pattern more closely, but the effect of mixed pixels at the fringe causes urban pixels to fall outside the true boundary, thereby lowering the overall agreement estimate.

The lowest agreement in Figure 4 is between the NLC data and the nighttime lights data. These results indicate that, while the lights data are able to capture all of the urban pixels in the NLC data, large areas identified as being urban in the nighttime lights data are outside of the urban areas identified in the NLC data. Again, this result suggests that the nighttime lights data overestimate city size, despite using a threshold to reduce blooming effects.

![Figure 4. Overall per-pixel agreement between NLC data and large area maps of urban land cover for ten cities selected in the United States.](image-url)
Data from the ten cities were aggregated to compute a single error matrix for each map, shown in Tables 2, 3, and 4 (the columns are the predicted class while the rows are the NLC "truth data"). While the overall proportion of correctly classified pixels differs only slightly among the three large area maps, inspection of patterns in user’s and producer’s accuracies provide useful insights.

The user’s accuracy represents the error of commission; pixels that are non-urban but which the map mislabels as urban (total correct per class divided by the column total). The producer’s accuracy measures errors of omission; pixels that are truly urban but which the map represents otherwise (total correct per class divided by the row total). Because the data used to compile the DCW are outdated, the user’s and producer’s accuracies are high for the non-urban class of this map (Table 2). Conversely, the producer’s accuracy for the urban class is lower (50.8 percent) than the user’s accuracy (66.2 percent). This result occurs because the DCW map misses many areas of new growth and suburbs. Note that, while the DCW may provide a reasonable depiction of urban areas in North America, this source is not reliable on other continents or in locations where urban growth has been more rapid in the last three decades.

For the nighttime lights data (Table 3) the user’s accuracy for the urban class is substantially lower (37.9 percent) than the producer’s accuracy (88.5 percent). Because of the nature of the nighttime lights data (i.e., they are a depiction of light sources and not urban land cover), many pixels fall outside the true city boundaries. This results in a low user’s accuracy. On the other hand, the nighttime lights data have no problem capturing the “true” urban pixels correctly, which results in a high producer’s accuracy. The same effect can be seen in the user’s and producer’s accuracies for the non-urban class: the producer’s accuracy is low (19.0 percent), while the user’s accuracy is high (98.2 percent). Finally, the error matrix for the fusion map (Table 4) shows results similar to those of the DCW data (Table 2). Further, this table illustrates that the user’s and producer’s accuracies are quite similar for the fusion map, a result that is desirable (Stehman, 1997).

Evaluating City Size
A second measure of map quality can be obtained from estimates of city size. This method compares the area mapped as urban for a given city at a 1-km resolution against the NLC data at a 30-m resolution. Figure 5 compares estimates of city size obtained from the NLC data against estimates from the DCW, the nighttime lights data, and the fusion map. The DCW map (Figure 5a) slightly underestimates city size, as expected. The nighttime lights data (Figure 5b) systematically overestimates city size because of the blooming problem (Elvidge et al., 1996). In the fusion map (Figure 5c) the sizes of the 42 cities are estimated without bias. The root-mean-square error for each of these plots corroborate this result, with values of 324.0, 1527.7, and 204.7 km² for the DCW, the nighttime lights data, and the fusion map, respectively.

Figure 5. A comparison of city size in continental scale maps of urban areas (vertical axis), using the urban areas defined by the 30-m NLC data to represent “true” city size (horizontal axis). The DCW data systematically underestimate city size (a), the nighttime lights data systematically overestimate city size (b), and the map produced from the combination of MODIS, nighttime lights, and population data (c) achieves the closest and least biased fit to the “true” city size.
Assessing City Location

While estimates of city size provide information regarding the overall agreement between each of the maps and the NLC data, such estimates provide no information regarding the geographic pattern (if any) of these errors. To explore the nature of these errors, a series of one-kilometer buffers around each city was used to calculate the spatial distribution of errors relative to the urban core. This analysis was performed at the state level. Each one-kilometer buffer was overlaid on each of the maps, and the number of urban pixels falling within each ring was calculated. This procedure was repeated (i.e., the buffer was expanded) until all non-urban areas within the state were considered (Figure 6).

Results from this analysis are shown for three states: Michigan, Ohio, and Washington (Figure 7). In this figure, the number of urban pixels falling within each buffer is reported as a function of the distance from the urban core. The number of urban pixels falling in these rings is graphed as a function of the distance from the urban core, normalized by the amount of urban area for the state.

Figure 6. Demonstration of the buffer process used to assess the location of errors. The figure shows the urban core in black with three 1-km buffers.

Figure 7. A comparison of locational accuracy in three different states for the three large area maps of urban land cover. Locational error is represented as a function of the distance of incorrectly labeled pixels from the “true” city boundary determined by the National Land Cover Data, normalized by the amount of urban area for the state.
of urban pixels in each buffer is normalized by the area of the urban class in the state, thereby accounting for differing levels of urbanization in each state. Figure 7 shows that, for the fusion map, the majority of the misclassified pixels fall in the first buffer zone outside of the “true” city boundary. Beyond the first two buffer zones (2 km), the amount of error is negligible. The results are quite similar for the DCW map, as might be expected.

For the nighttime lights data, results reveal the extent of the blooming problem: in each state, the amount of error in the first two buffer zones is substantial, representing from 50 to 100 percent of the total urban area as error. This result reinforces the difficulty in using the nighttime lights data to map urban land cover.

Discussion
Results from this research show that, by fusing MODIS data with the nighttime lights and population density data, it was possible to map urban areas on a continental scale with reasonable accuracy. Considering the frequency of missing values in the MODIS data, the small amount of urban training data utilized, and the expected loss of information due to aggregation and the mixed pixel effect, the results are quite reasonable and promising.

This research also pursued a variety of accuracy assessment approaches, including methods that compare regional and local subsets of the fusion map to the DCW, the nighttime lights product, and the NLC urban data. These assessments reveal substantial information regarding the nature and quantity of map errors that complement traditional accuracy measures. For example, our analysis showed that 99 percent of incorrectly classified pixels in the fusion map were located within two kilometers of the NLC city boundary. This result provides evidence that much of the apparent error in urban land cover can be attributed to registration uncertainty and issues associated with sub-pixel heterogeneity in coarse resolution data. Also, it is important to note that aggregation of the NLC data may have introduced errors and bias. Errors of commission in the fusion map may reflect pixels that in reality contain substantially built-up areas, but which are mislabeled by the aggregation method. This suggests the need for further investigation of urban-class definitions, especially along the urban fringe. In addition, time discrepancies between the NLC data (1992) and the MODIS data (2000–2001) may introduce errors during accuracy assessment. The effect of this issue should be minor, however, considering that the majority of cities in North America have not expanded by several kilometers in the last five to ten years.

One particularly important result from this work is that the nighttime lights data and gridded population data possess useful information regarding the spatial extent of urban areas that can be integrated with the multispectral and multitemporal data available from MODIS. However, any of the three data sets, when used by themselves, provided inferior results relative to the fusion map. These data sources should therefore be used with caution.

The logical extension of this work will be to develop a generalization of the methodology that can be applied globally. Doing so will require a variety of modifications, including calibration of the logistic regression model for each continent, reexamination of the nighttime lights threshold, and possible use of alternative ancillary data sets for \textit{a priori} information. In particular, urban radiances conveyed by the nighttime lights data are quite different with respect to urban land cover across the globe (e.g., areas of India, North Korea, and parts of Africa with equivalent population densities have cities that are substantially less bright compared to industrialized cities), and the population density data has coarser resolution for countries without detailed population data.

Conclusions
The main objective of this research was to improve understanding of the methodological and validation requirements for mapping urban areas over large areas. Urban areas were mapped on a continental scale by fusing MODIS data and two ancillary sources: the nighttime lights and the gridded population density data. Decision trees used in association with \textit{a priori} probabilities were particularly effective for resolving confusion between urban areas and other land-cover types. This result reflects a trend towards increased fusion of different data types to provide more representative characterization of land-cover qualities (Benediktsson and Sveinsson, 2000; Ranchin et al., 2001). This is especially true for urban areas, where high levels of within-region and between-region variability make classification especially difficult.

Finally, this research explored a variety of avenues for map accuracy and assessment. While conventional methods are difficult to apply because of the coarse resolution of the data and the small areal extent of urban land cover, the use of alternative accuracy approaches can provide valuable indicators regarding map quality. The maps of urbanized areas produced by the fusion of MODIS, nighttime lights, and gridded population data were at least as accurate as the DCW, and superior to the nighttime lights product alone. Understanding these methodological and validation requirements is critical for future urban mapping at regional, continental, and global scales.

Acknowledgments
This work was supported by NASA grants NAG5-31369 and NAG5-30401. The authors gratefully acknowledge Marc Imhoff and David Stitzer for use of the radiance calibrated nighttime lights data set, as well as three anonymous reviewers for their helpful suggestions.

References


Fitzpatrick-Lins, K., 1981. Comparison of sampling procedures and data analysis for a land-use and land-cover map, Photogrammetric Engineering & Remote Sensing, 47:343–351.


(Received 29 May 2002; revised and accepted 16 January 2003)