

Geographic Information System Software to Remodel Population Data Using Dasymetric Mapping Methods

Techniques and Methods 11–C2

This page intentionally left blank

Geographic Information System Software to Remodel Population Data Using Dasymetric Mapping Methods

By Rachel Sleeter and Michael Gould

Techniques and Methods 11–C2

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
DIRK KEMPTHORNE, Secretary

U.S. Geological Survey
Mark D. Myers, Director

U.S. Geological Survey, Reston, Virginia 2007

This report and any updates to it are available online at:
<http://pubs.usgs.gov/tm/tm11c2>

For product and ordering information:
World Wide Web: <http://www.usgs.gov/pubprod>
Telephone: 1-888-ASK-USGS

For more information on the USGS—the Federal source for science about the Earth,
its natural and living resources, natural hazards, and the environment:
World Wide Web: <http://www.usgs.gov>
Telephone: 1-888-ASK-USGS

Any use of trade, product, or firm names is for descriptive purposes only and does not imply
endorsement by the U.S. Government.

Although this report is in the public domain, permission must be secured from the individual copyright owners to
reproduce any copyrighted material contained within this report.

Cataloging-in-publication data are on file with the Library of Congress (<http://www.loc.gov/>).

Produced in the Western Region, Menlo Park, California
Manuscript approved for publication, February 19, 2008
Text edited by James W. Hendley II
Layout and design by Judy Weathers

Contents

Introduction.....	1
Purpose and Scope	1
Review of Areal-Interpolation and Dasymetric-Mapping Approaches.....	1
Case Study—Clatsop County, Oregon.....	2
Dasymetric-Mapping Extension.....	4
Overview.....	4
Installation.....	4
Enabling and Operating the Dasymetric-Mapping Extension	6
Input Data and Parameters.....	6
Output Files and Symbolization	9
Output Maps	10
Summary.....	10
Acknowledgments.....	11
References Cited.....	14
Appendix—Equations for Dasymetric-Mapping Extension (Mennis, 2003).....	15

Figures

1. Diagram showing how dasymetric mapping redistributes ancillary information to better represent population distribution.....	2
2. Diagram showing three fundamental approaches to mapping population distribution.....	3
3. Screen image of “Welcome to the Dasymetric-Mapping-Tool Setup Wizard” dialog box.....	5
4. Screen image of “Select Installation Folder” dialog box	5
5. Screen image of “Confirm Installation” dialog box.....	6
6. Screen image of “Installation Complete” dialog box	6
7. Screen image of the “Customize” dialog box to enable ArcMap extension.....	7
8. Screen image of the “Dasymetric-Mapping Extension” user form	8
9. Screen image of the “Processing Complete” dialog box.....	9
10. Screen images showing four steps to symbolizing dasymetric-mapping output data	11
11. Choropleth map (top) and dasymetric map (bottom) of Astoria, Oregon	12
12. Choropleth map (left) and dasymetric map (right) of Seaside, Oregon	13

Tables

1. Dasymetric-Mapping Extension output file names with descriptions.....	10
--	----

Geographic Information System Software to Remodel Population Data Using Dasymetric Mapping Methods

By Rachel Sleeter and Michael Gould

Introduction

The U.S. Census Bureau provides decadal demographic data collected at the household level and aggregated to larger enumeration units for anonymity purposes. Although this system is appropriate for the dissemination of large amounts of national demographic data, often the boundaries of the enumeration units do not reflect the distribution of the underlying statistical phenomena. Conventional mapping methods such as choropleth mapping, are primarily employed due to their ease of use. However, the analytical drawbacks of choropleth methods are well known ranging from (1) the artificial transition of population at the boundaries of mapping units to (2) the assumption that the phenomena is evenly distributed across the enumeration unit (when in actuality there can be significant variation).

Many methods to map population distribution have been practiced in geographic information systems (GIS) and remote sensing fields. Many cartographers prefer dasymetric mapping to map population because of its ability to more accurately distribute data over geographic space. Similar to "choropleth maps", a dasymetric map utilizes standardized data (for example, census data). However, rather than using arbitrary enumeration zones to symbolize population distribution, a dasymetric approach introduces ancillary information to redistribute the standardized data into zones relative to land use and land cover (LULC), taking into consideration actual changing densities within the boundaries of the enumeration unit (Langford and Unwin, 1994) (fig. 1). Thus, new zones are created that correlate to the function of the map, capturing spatial variations in population density. The transfer of data from census enumeration units to ancillary-driven homogenous zones is performed by a process called areal interpolation.

Purpose and Scope

An extension module to add specialized functionality to a GIS was created to automate the process of transferring data from source population zones (census enumeration districts) to overlaying target zones (LULC). Here, the dasymetric mapping technique explores an areal interpolation scheme adopted and developed from Mennis (2003) that practices empiri-

cal sampling and areal weighting. This report reviews the extension installation, the dasymetric mapping methodologies behind each point-and-click feature, and the benefits of using this method and associated extension by examining the output files. The ArcMap extension module developed and modified by Michael Gould of the U.S. Geological Survey (USGS), utilizes at its core, original Visual Basic for Applications (VBA) programming written by Jeremy Mennis and Torrin Hultgren in 2004 at the University of Colorado, Boulder, Geography Department.

Even though the nature of a population distribution is more realistically represented with a dasymetric map and adequate research has been done on the utility and function of dasymetric mapping, the complexity of this method compared to the choropleth method often deters cartographers from using the dasymetric mapping method (Charpentier, 1997). The ArcMap extension discussed in this report is designed to facilitate the use of dasymetric mapping by automating the areal interpolation process within a GIS framework. Creating an extension that overcomes the obstacles inherent in dasymetric mapping, the method becomes transferable to a wider spectrum of users.

Review of Areal-Interpolation and Dasymetric-Mapping Approaches

Spatial analysis often requires the transferring of data from one set of geographic zones to another set of overlapping zones. For example, data may exist on the number of people living within census enumeration units, but there is a need to estimate the number of people living within watershed boundaries. Areal interpolation is the process of transforming spatial-data values from a set of source zones (the original unit of data aggregation) to target zones (the intersecting zone). Areal-based functions practice volume-preserving properties, what Tobler (1979) refers to as the pycnophylactic property, that is, no data is lost or created during the transformation. All areal-interpolation methods are based upon the geometric intersection of the source and target zones, where some may or may not use ancillary information to add more intelligence to the redistribution. Simple areal weighting does not use ancillary information and bases interpolation on the areal geometric properties of the source and target

2 Geographic Information System Software to Remodel Population Data Using Dasymetric Mapping Methods

zones. The dasymetric mapping method uses ancillary information to generate parameters used in areal interpolation, while maintaining volume-preserving properties. The most common form of ancillary data used for population estimation is LULC classifications derived from remotely sensed imagery, topographic maps, or land parcel databases (Wright, 1936; Yuan and others, 1997; Mennis, 2003; Hawley and Mollering, 2005). The areal-interpolation performance has a direct correlation with the quality and accuracy of the LULC-derived data.

The most straightforward dasymetric mapping approach using LULC data is the binary division approach in which LULC is classified as “populated” or “unpopulated” and census populations are simply redistributed to “populated” areas (Wright, 1936; Langford and Unwin, 1994; Fisher and Langford, 1996; Holt and others, 2004; and Wu and others, 2005). Holloway and others (1997) combined multiple ancillary datasets to detect and remove uninhabited lands from the area of analysis. Four types of LULC were ruled out, including census blocks with zero population, lands owned by local, State, or Federal government, corporate timberlands, water and wetlands, and open and wooded areas with terrain slopes greater than 15 percent. Holloway and others (1997) assigned a fixed proportion of the total population to each specific LULC (Urban 80 percent, Open 10 percent, Agriculture and Forested 5 percent, and Uninhabited 0 percent). These proportions are determined subjectively with prior knowledge of the area and field observation. One critical drawback to assigning fixed proportions to designated LULC is that the spatial area of each LULC zone within each source zone is not accounted for, so if there is a source

zone with a very small spatial area that is classified urban, that urban zone still receives 80 percent of the source population.

Mennis (2003) used a combination of areal weighting and a three-tier, urban-classification method, using satellite-derived thematic data to break down census tracts into homogenous residential zones. To account for the relative densities in each urban class, empirical sampling provides a proportional density fraction used as a weighted value representative of each urban class. To deduce the necessary urban class population density values, source zones that are entirely occupied by a single inhabited land-cover class are isolated and the population density is calculated by dividing total population by total area. These density ratios are extrapolated to all source zones that have a combination of inhabited LULC classes. Area ratios are then calculated to adjust the population density fraction by the percentage of that block group’s total area that the urban class occupies. The combination of an empirical sampling and areal-based weighting makes this particular dasymetric method a good choice because the method addresses the weaknesses noted in Holloway and others (1997). A simple graphic, figure 2, shows the urban 3-class method with concepts of areal weighting and the relative difference in density between LULC classes.

Case Study—Clatsop County, Oregon

Local communities potentially affected by natural disasters are taking strides to minimize loss in the event of a catastrophic event. Scientific analysis and technological advances have led to

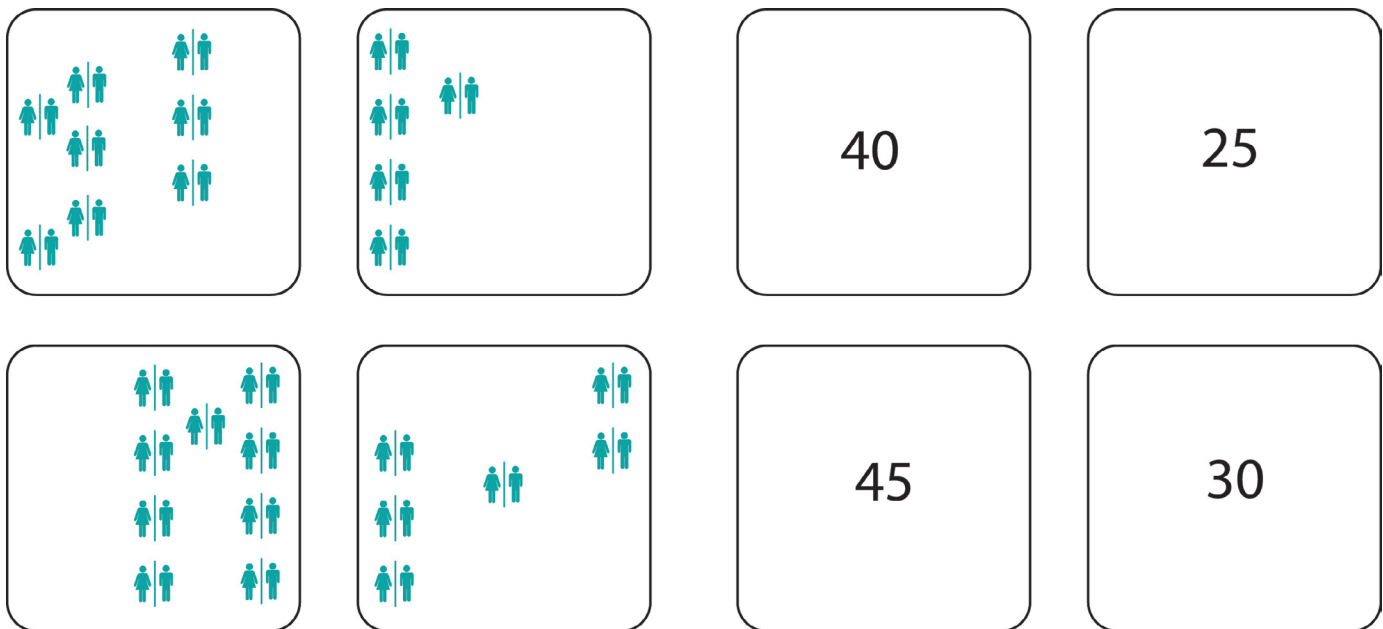


Figure 1. Diagram showing how dasymetric mapping redistributes ancillary information to better represent population distribution. Data may be collected for individual households, as shown by the four blocks on the left. In this example, assume that every male/female pair represents a household of five persons. However, due to confidentiality, data is often aggregated into larger units of analysis, as shown by the four blocks on the right. Dasymetric mapping attempts to disaggregate the four figures on the right into their original state on the left.

the growth of geographic tools that have the capacity to aid local jurisdictions in emergency planning. The use of a GIS allows users to visually and statistically characterize populations at risk within hazard zones. The goal of this case study is to develop a more accurate representation of population distribution along the Oregon Coast, estimating daytime versus nighttime fluctuations for use in emergency, preparedness planning for a Cascadia Subduction Zone (CSZ) tsunami. The implementation of a micro-scale map of population density in the daytime and nighttime provides a realistic snapshot of who is on the ground when a tsunami event occurs, highlighting higher risk areas during response and recovery efforts. By using a dasymetric mapping technique to manipulate U.S. Census Bureau enumeration units, population values are distributed to 10-m grid cells with the input of a tax parcel dataset acting as the primary ancillary indicator of inhabited LULC.

A critical first step in managing these threats is to understand who is exposed to these hazards. In the United States,

emergency managers have access to decadal census data that summarizes residential populations. However, in many coastal communities, there are typically greater numbers of nonresidents (for example, employees, tourists) in high tsunami hazard zones (Wood and Good, 2004). Emergency managers need to estimate the number of residents, employees, and visitors who could potentially be exposed to possible hazards if they are to develop realistic risk reduction and preparedness strategies. Datasets like LandScan, developed as part of the Oak Ridge National Laboratory (ORNL) Global Population Project (Oak Ridge National Laboratory, 2005), focus on ambient population and have been used for post-disaster assessments at the global scale. Ambient population is defined as the number of people likely to be in an area as opposed to residents alone, with population counts for each 1-km² cell being based on probability coefficients. Verified to be accurate to within 10 percent of the given values, these coefficients are based on road proximity,

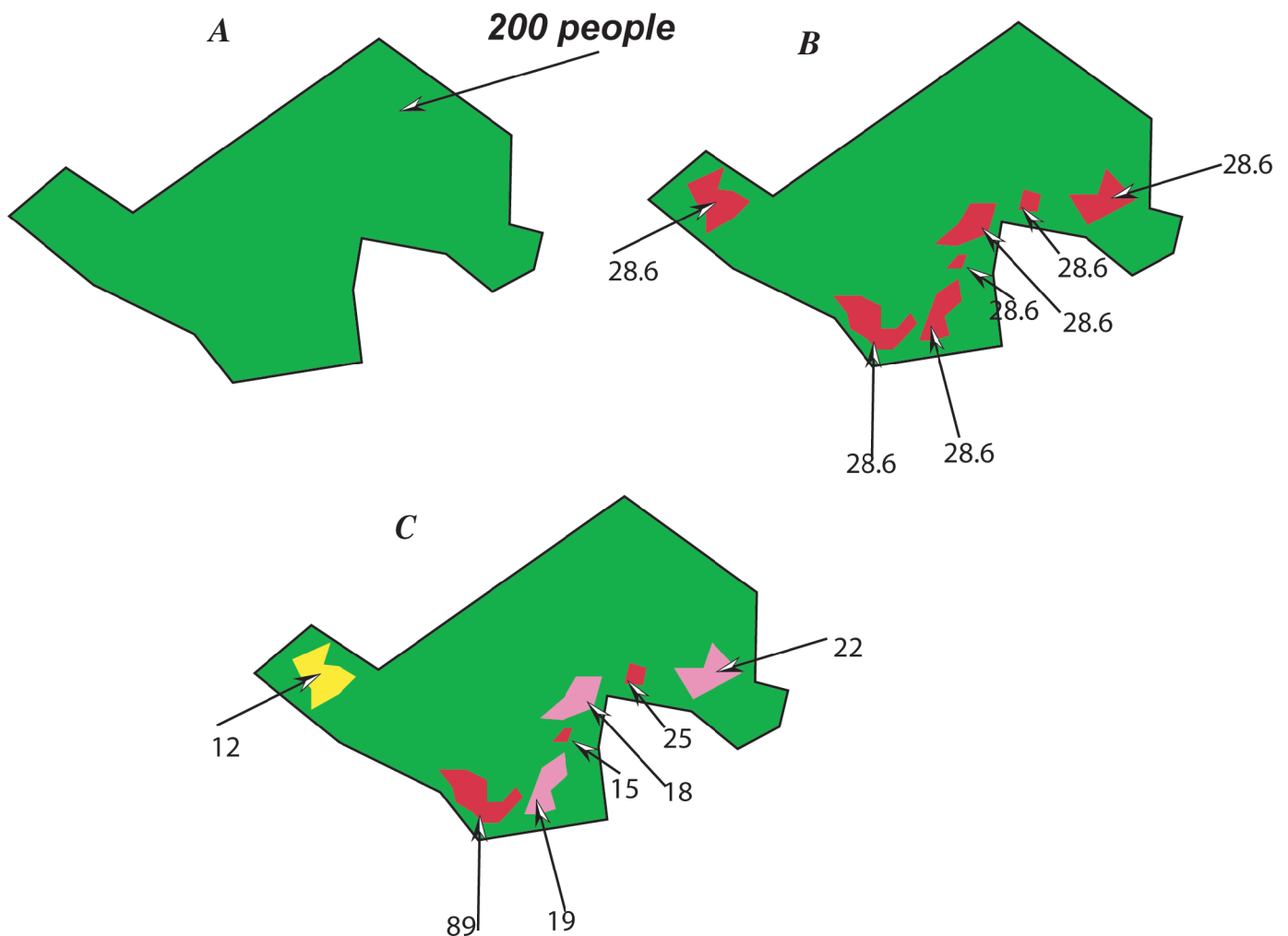


Figure 2. Diagram showing three fundamental approaches to mapping population distribution. *A*, Total number of people, aggregated by census delineated unit; *B*, represents inhabited (red) versus uninhabited (green) with populations evenly distributed within the inhabited land use; and *C*, represents an urban 3-class method, where populations are distributed in high (red), medium (pink), and low (yellow) population class based on LULC-class code and areal weighting.

slope, land cover, and nighttime lights. However, with a 1-km² cell size, Landsat is considered too coarse for local planning. The research conducted at the USGS focused on the dasymetric mapping methods for mapping daytime and nighttime population density at the local level with intentions of applying these methods to other local communities involved with hazard planning and response. For demonstration purposes in this report, we only focus on nighttime population values.

The quality of the interpolation estimates relies heavily on (1) how source zone and target zone boundaries are defined, (2) the LULC classification accuracy, (3) the degree of generalization in the interpolation process, and (4) the characteristics of the partitioned surface (Lam, 1983). We use the census-block group as the source zones of known population values and a tax-parcel database as the target zones. The parcel database has many attributes that allow for hierarchical density weighting relative to building type, land-use, and description. For example, an apartment complex indicates high density, a single family home indicates medium density, and a large acreage estate indicates low density. The parcel database was divided into four density classes (high, medium, low, and exclusion) to act as a hierarchical measure within the interpolation process. There were many incentives for using parcel data for the Clatsop County application. The creation of a methodology that can be transferred to other hazard-prone regions was necessary for future work, which required the data inputs to be accessible, affordable, and uniformly structured. All local governments have land-parcel information for their areas, and about 70 percent of all national parcels have been converted to a geospatial format (National Academy of Sciences, 2007). Also, the scale, resolution, and accuracy derived from parcel data are significantly better than most remote-sensing-derived LULC products, such as the National Land Cover Dataset, created at a 30-m grid resolution (Vogelmann and others, 2001). To capture land-cover classifications at such a high resolution, the more common approach would be to use remotely sensed imagery with high spatial resolution (for example, IKONOS—4-m multispectral imagery). However, due to the spatial and spectral heterogeneity of the urban environment, inconsistencies become prevalent when trying to characterize built-up structures mixed in with vegetation cover (Herold and others, 2002). Consequently, parcel data becomes a more practical option for delineating residential neighborhoods into density classes.

The approach we used to map nighttime population distribution for Clatsop County, Oregon, draws from a combination of previous methods (Mennis 2003; Holloway and others, 1997) leading to the use of four total density classes [three urban-residential classes and one exclusion class (representing all uninhabited areas)]. The parcel attributes describe and designate the land-uses, building types and density parameters present on each parcel of land, which are used as primary indicators to develop four density classes: high density, medium density, low density, and uninhabited. The newly created LULC layer is converted to raster format at a 10-m resolution acting as the target zones for the interpolation of census-block groups. The ArcGIS extension module automates the

areal interpolation using the principles of areal weighting and empirical sampling of relative densities which result in a three-tier distribution of population density at a 10-m pixel resolution. The output maps are shown after the demonstration of the ArcGIS extension (see figs. 11 and 12 later in this report).

Dasymetric-Mapping Extension (DME)

Overview

The GIS extension we developed is summarized in this report to assist potential users with conceptual background of dasymetric mapping relative to the functionality of the extension. The Dasymetric-Mapping Extension (DME) operates in ArcGIS 9.1 or 9.2 as a point-and-click dialog box prompted by a button on the main toolbar. The main objective of this extension is to automate the process of taking population data from census enumeration units and transferring the data values to overlaying homogenous zones while (1) maintaining volume preserving properties and (2) using an empirical sampling method for determining relative densities for each homogenous zone. The outcome is a raster representation of population per pixel.

The input data required for the operability of the DME is twofold. First, the user is required to obtain U.S. Census Bureau data in a shapefile format, at any geographic level (block, block-group, tract, county). Second, the ancillary data needs to represent LULC density parameters in a raster ArcGRID format. For example, the ancillary data used, whether one or multiple datasets, must be recoded into a single ESRI ArcGRID format with four density classes representing high, medium, and low population density, as well as an uninhabited class where zero population will be assigned. Refer to the section “Input Data and Parameters” for detailed guidelines.

Installation

Before installing the DME, the following programs must be loaded on the host computer in order for the extension to work properly:

- ArcGIS 9.1 or 9.2 (must have the Spatial Analyst Extension installed and enabled).
- ArcObjects InterOp libraries [If installed, these libraries would be located in a folder called “DotNet” in the ArcGIS install directory (that is, C:\Program Files\ArcGIS\DotNet). If the folder is missing, run the ArcGIS installation disc in modify mode to install the libraries for .NET.]; and,
- Microsoft .NET Framework 1.1 (To check if your system has this, go to your desktop taskbar and click Start->Control Panel->Add/Remove Programs and look for an entry for the .NET Framework 1.1.).
Note: If you see that Microsoft .NET Framework 2.0 has been installed, the application will still run properly, as long as Microsoft .NET Framework 1.1 is also installed.

The installation package for the DME is loaded into a WinZip file that can be accessed at <http://geography.wr.usgs.gov/science/dasymetric/data.htm>. The WinZip file includes a setup.exe, a setup.ini and a setup.msi file. To guarantee that all

necessary files and linkages are installed, users should install the extension by opening the “setup.exe” file. On opening the “setup.exe” file, the Dasymetric-Mapping-Extension Setup Wizard will be launched (figs. 3-6).

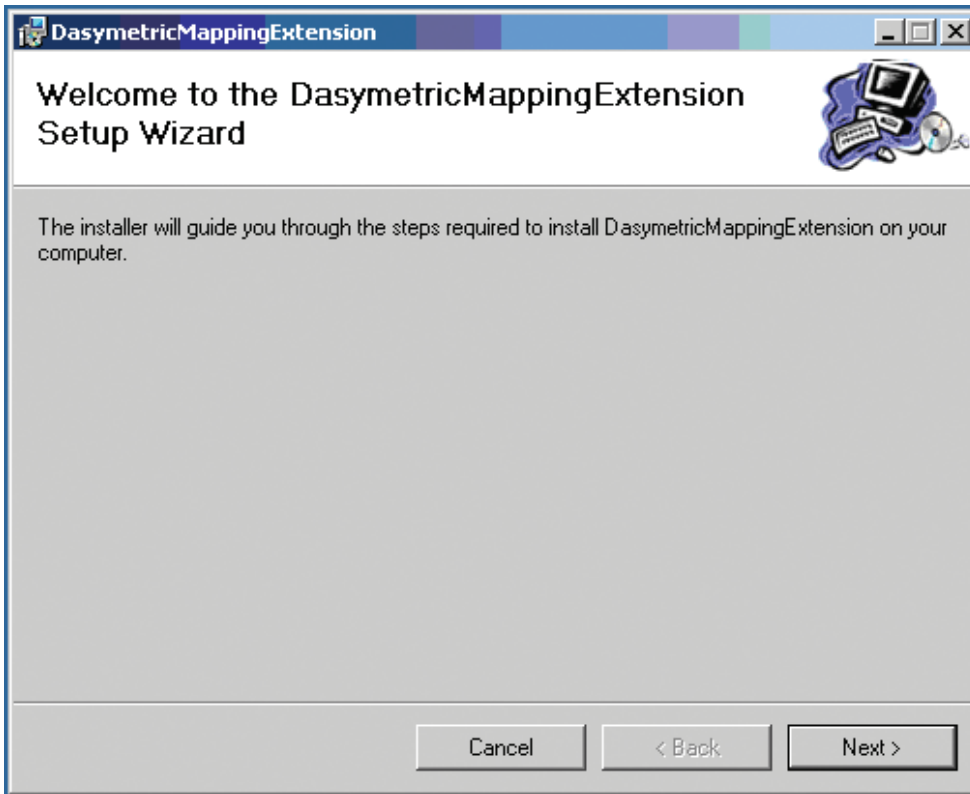


Figure 3. Screen image of the “Welcome to the Dasymetric-Mapping-Extension Setup Wizard” dialog box. Clicking the “Next” button in this screen brings up the “Select Installation Folder” dialog box (fig.4). The user should install the folder to the default directory, which is “C:\Program Files\,” to ensure proper installation and functioning. The user is also able to determine here whether the extension is accessible to “just me” or “everyone” who uses the computer.

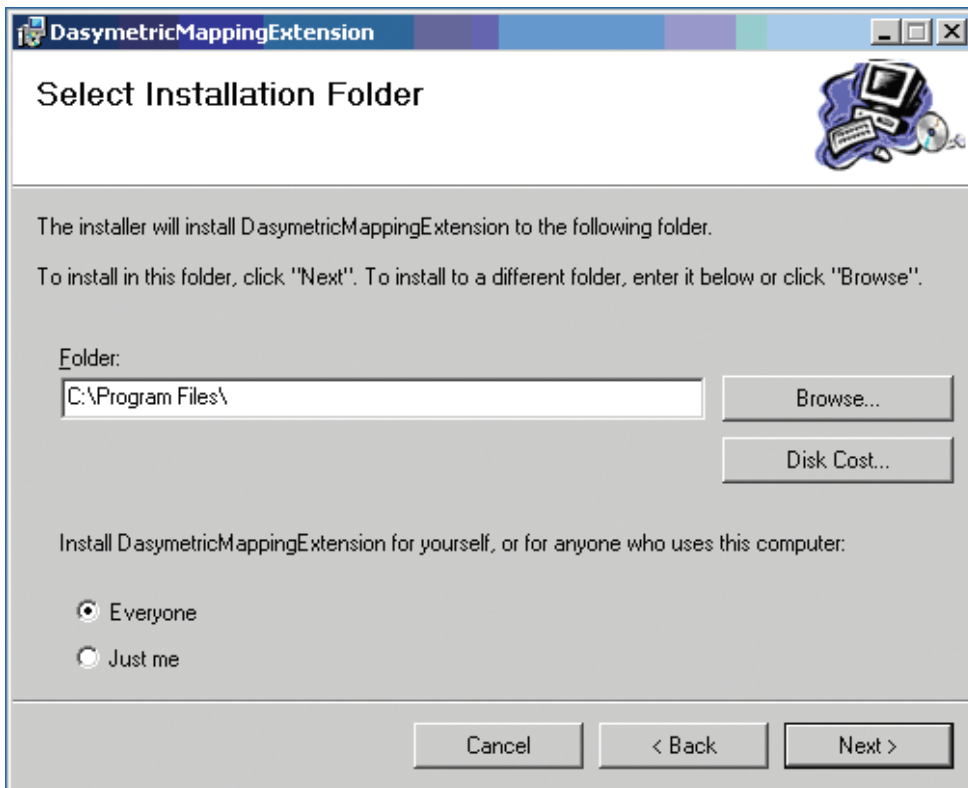


Figure 4. Screen image of the “Select Installation Folder” dialog box. Clicking the “Next” button in the “Select Installation Folder” dialog brings up a “Confirm Installation” dialog box (fig. 5). The user clicks “Next” to begin the extension installation.

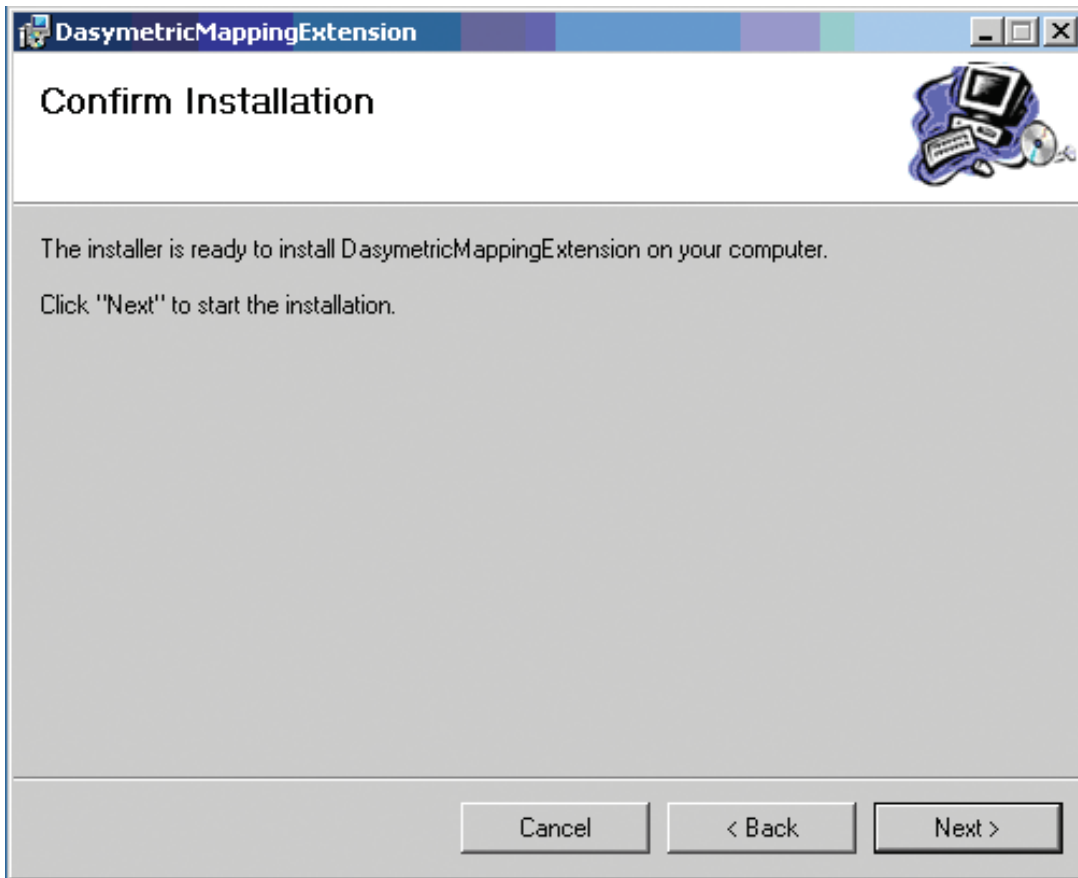


Figure 5. Screen image of the "Confirm Installation" dialog box. If the Dasymetric-Mapping Extension is successfully installed, an "Installation Complete" dialog box will appear (fig. 6). The user should then click the "Close" button to close the installation dialog box.

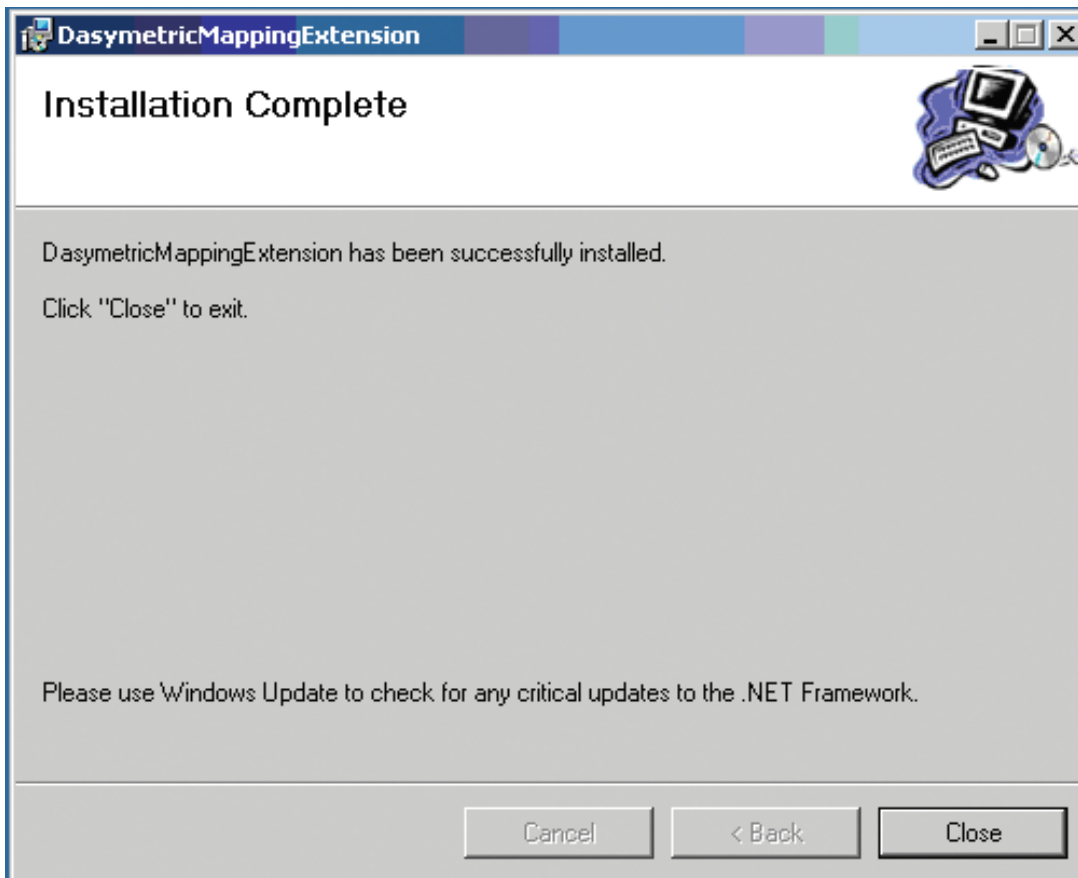


Figure 6. Screen image of the "Installation Complete" dialog box. Note that if the "setup.exe" fails to install the extension, the user must ensure that the host computer has all the necessary permissions and required configurations described earlier in this section.

Enabling and Operating the Dasymetric-Mapping Extension

Once the downloadable extension has been properly installed, the user can enable the extension within ArcGIS 9.1 or 9.2. To do this, the extension will register under the Tools/Customize/Commands menus in ArcMap (fig.7). Simply drag and drop the tool titled, “DasymetricMapping VB.NET” onto an open space in the ArcMap toolbar and the graphical icon will appear as the operating button. In order to open the DME’s main dialog box via the map button on the main toolbar, the user must add their input data files into the ArcMap document by using the “Add Data” function.

Input Data and Parameters

With the intention of attracting a wide range of users and applications, the DME attempts to simplify the dasymetric

mapping process to a point-and-click interface. There are two primary datasets necessary for the DME to function properly; a U.S. Census Bureau population-data shapefile, and a LULC-derived raster layer developed into four data classes representing the inhabited/uninhabited areas as a density stratification (for example, (1) high-density residential, (2) low-density residential, (3) non-urban inhabited, (4) uninhabited). The parameter settings relate to the data fields needed to run the DME and the principles used in the areal interpolation. The numbered steps in the text give a detailed explanation to operating the DME; however, the appendix contains simple mathematical equations derived from Mennis (2003) used as the programming backbone.

- (1) **Enter an output path of folder:** User must browse to choose output destination folder. Output destination folder must be empty when running the software application. Program will not overwrite or add to existing files in output folder, resulting in an error.

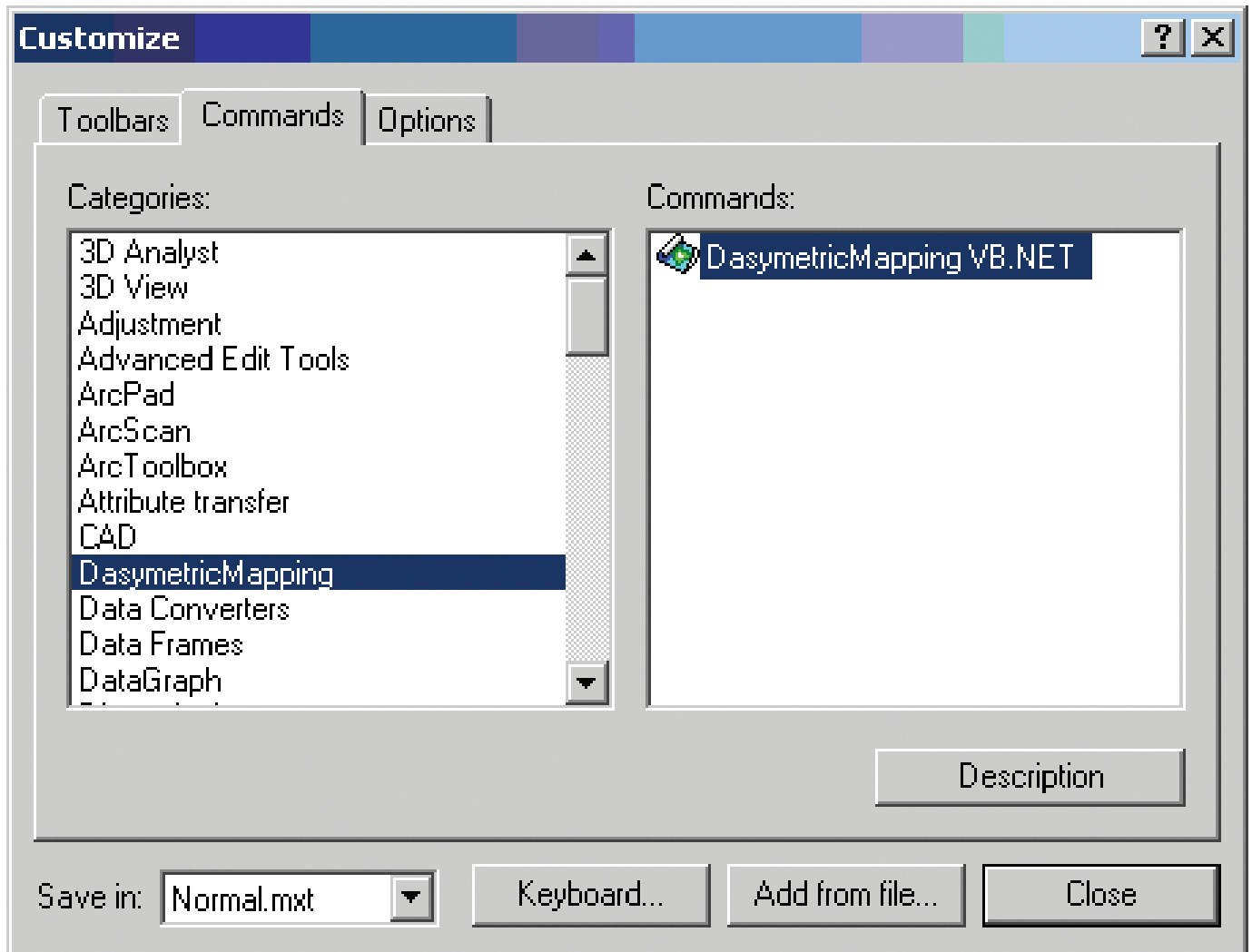


Figure 7. Screen image of the “Customize” dialog box to enable ArcMap extension. By navigating in ArcMap, Tools/Customize/Commands, the Dasymetric-Mapping Extension becomes available. Drag and drop onto your main ArcMap toolbar to enable the extension.

- (2) **Select population layer:** Any demographic data in a geospatial format that has one population value representing each areal unit (polygon) (example, U.S. Census Bureau Data by block, block-group, tract, county.) User must select the population file available in the drop-down list. Population data must be loaded into the ArcMAP project in order for the data to be available in the drop-down list 2 of the user-form.
- (3) **Select ancillary layer:** A land use or land cover-derived raster layer that has been reclassified into four classes representing inhabited/uninhabited areas into density stratification (for example, (1) high-density residential, (2) low-density residential, (3) non-urban inhabited, (4) uninhabited). User must select the ancillary layer from drop-down list 3 on the user-form. The grid cell size can be determined by the user and is dependent on the scale and the input data available. There is no limit to how small the grid cell size can be, however computation time may be affected.
- (4) **Select field in your population dataset that contains population count data:** Drop-down list will be populated with all of the fields present in the population layer.
- (5) **Select percent of source unit that must be covered by the ancillary class:** To model the relative difference in population density between the “inhabited classes” we use an empirical sampling approach. The grid cell with the higher inhabited class has a higher population density than a grid cell with a low or non inhabited class code based the empirical measurement. In order to determine this relative difference between classes, the population density values are sampled for each “inhabited class”. The sampling process selects all block-groups that have a specified “percent cover” (this threshold is set by the user within the tool. for example, 80 percent) by a single “inhabited class” that is equal to or exceeds that threshold. The total population and area are calculated to find the aggregated density of the

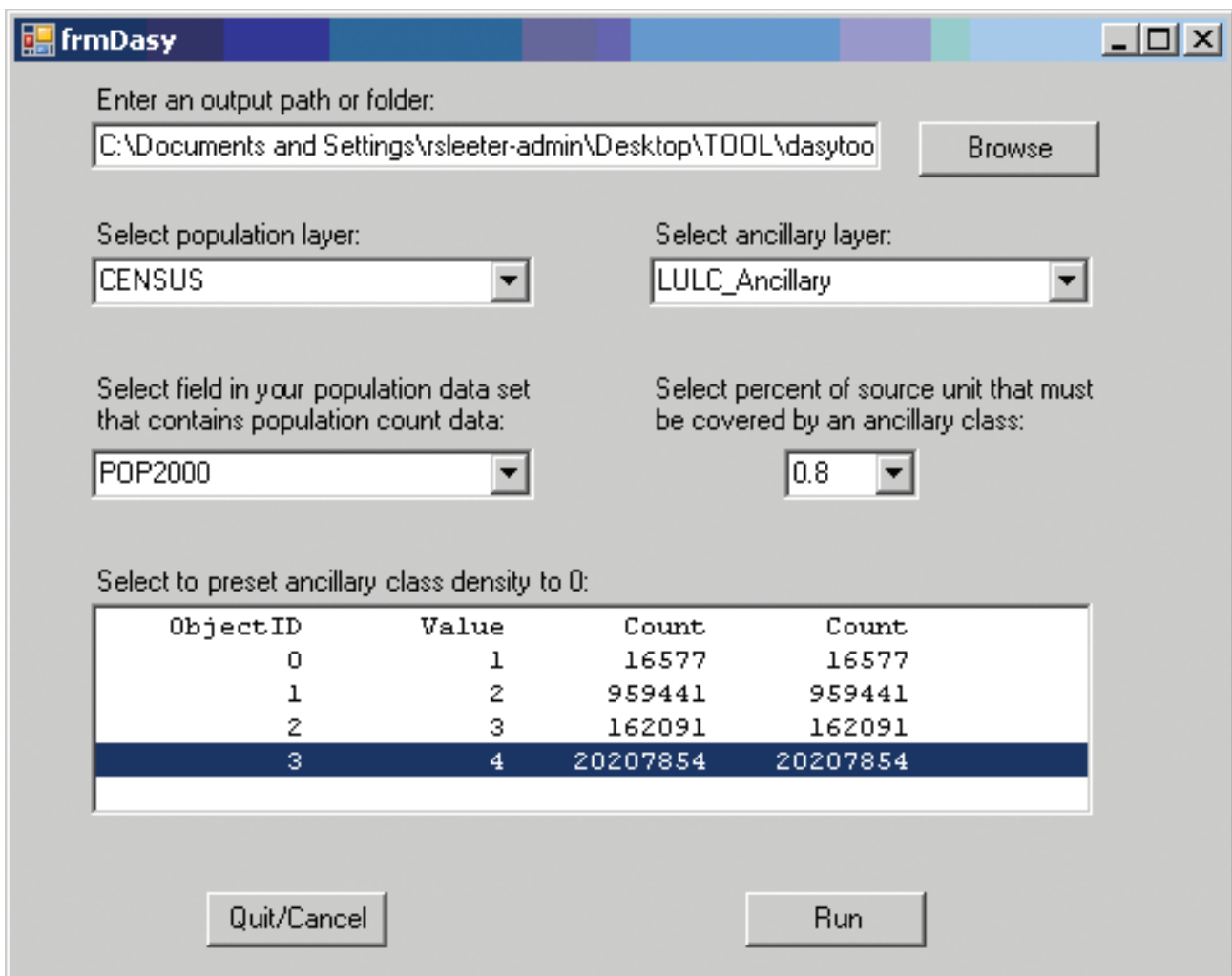


Figure 8. Screen image of the “Dasymetric-Mapping Extension” user form. Numbers 1 through 7 relate to the different steps needed to operate the extension correctly. The steps are explained in the text.

samples. This sampling technique provides a value that can be used as a “population density fraction,” which is the percentage of a block-group’s total population that should be assigned to each “inhabited class” within the block-group.

- (6) **Select preset ancillary class to zero:** If you have an ancillary class that you know is “uninhabited” (for example, water, non-residential developed, uninhabited wetlands, or forest) you may “preset” these classes to “0” so no data may be distributed to these zones.
- (7) **Click run**

Output Files and Symbolization

After successful completion of the DME, the “Processing Complete” dialog box will appear indicating no errors in computation performance (fig. 9.). The DME outputs an ESRI ArcGRID named “Dasymetric Result” representing population/grid cell. The actual data output files have been written to the output directory specified by the user and have file names different than what appears in the “Layers” view with the completion of the DME (table 1). The “Dasymetric Result” file appearing in the view following the comple-

tion of the DME is the result of a tabular join between the “dasy_rast” (ArcGRID) and the “dasytable” (*.dbf). The join is performed by appending all of the data in the “dasytable.dbf” to the “dasy_rast” based on the “Value” fields in both files. We have automated this join to facilitate manipulation of the output data; however, the join will not be saved once the ArcMap document has been closed.

In most GIS applications, the final result is a map that needs to convey information to the user in a manner that can be easily interpreted and related to reality. The effectiveness and use of GIS results can be compromised if the incorrect cartographic symbolization is chosen (Ramroop, 1998). Upon completion of the DME, the symbolization of the output data must be adjusted in order to demonstrate meaningful results. The user may symbolize the output data in any representation desired; however, we recommend the following to maximize data comprehension. In figure 10 we outline four steps within the ArcGIS “Layer Properties” and “Classification” dialogs to help achieve meaningful presentation of the spatial information. The ArcGRID named “Dasymetric Result” has a field (“NEW DENSITY”) that contains all of the new density values produced by the DME and must be modified for proper visualization of the results:

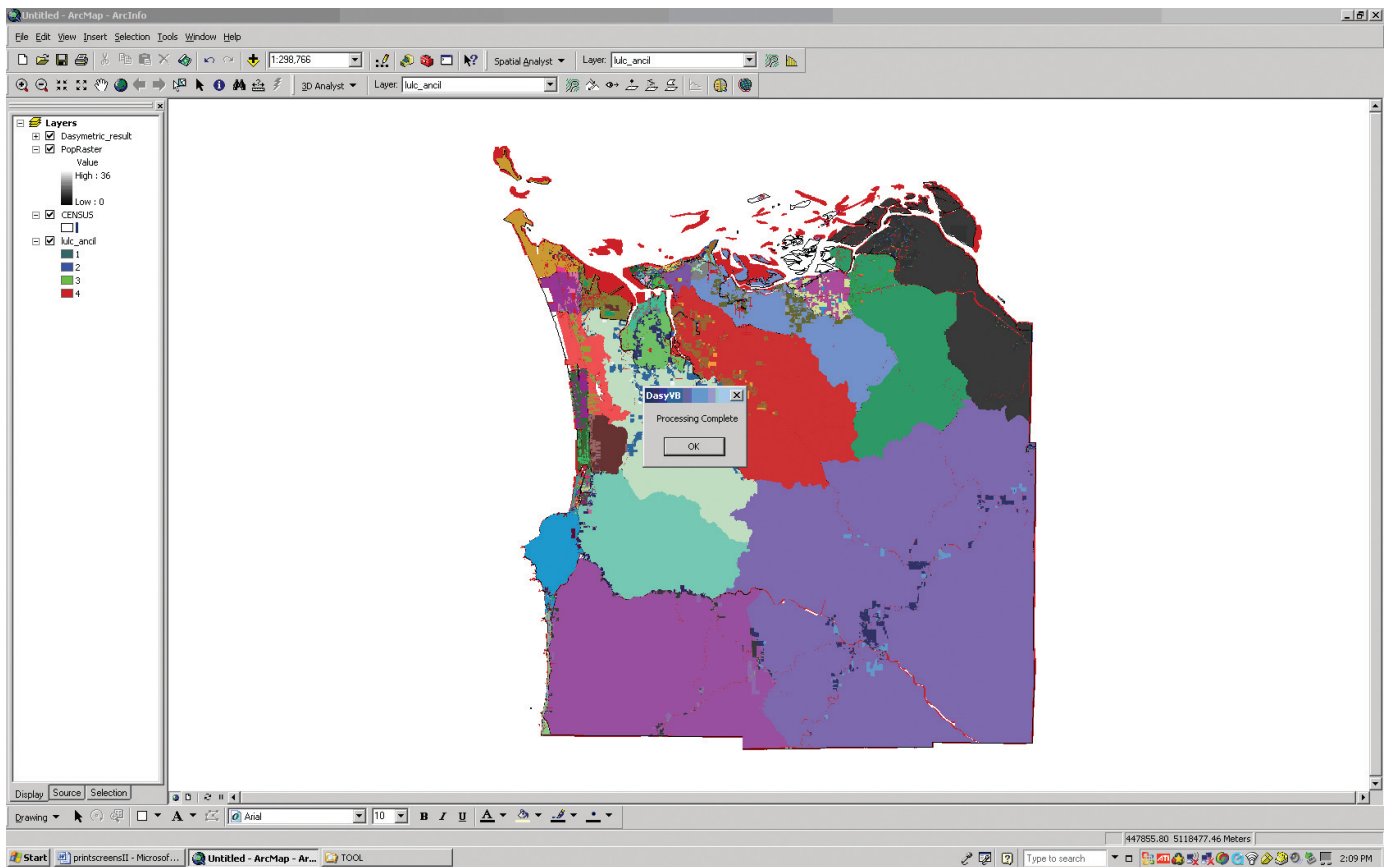


Figure 9. Screen image of the “Processing Complete” dialog box. The output data file will look skewed until the user adjusts the symbolization of the results.

Table 1. Dasymetric-Mapping Extension output file names with descriptions.

DME Output File Names	Format	Description
dasy_rast	ESRI GRID	GRID file of intersection of inputs (LULC and CENSUS) w/ pixel counts and unique ID
popraster	ESRI GRID	conversion of vector census file to raster GRID file
info	ESRI GRID- support files	support file for ESRI GRID - user must keep this file in the same location as GRID file
dasytable	*.dbf	summary table of all results per grid cell (populaton/grid cell)
Dasymetric_Stats	*.dbf	run-time summary for DME

- (1) Right click on the layer named “Dasymetric Result” and go to “Properties”. This will bring up “Layer Properties.” Click the “Symbolization” tab. The recommended symbolization is the “Classified” data range with the desired amount of classes within a natural breaks classification method. Under “Fields - Value,” choose the field named “DasyTable.NEWDENSITY.” Before selecting a color ramp, the user must adjust the histogram to exclude all zeros from the data range. Under “Classification,” click “Classify.”
- (2) In the “Classification” dialog, click “Exclusion” to bring up “Data Exclusion Properties.” Enter “0” into the excluded values field.
- (3) Click on the “Legend” tab. Choose legend properties for all zero values by checking the box “Show color for excluded data” (for example, choose white to symbolize all “0” values). Click OK until returning back to “Layer Properties.”
- (4) After adjusting the histogram of data values to exclude all pixels with zero population, the user should notice that the minimum value is now greater than zero. The final recommended step is to choose a color ramp that signifies population density. Under Color Ramp, there is a pull down menu of choices. Cartographically, density on a color ramp should be symbolized using either dark to light or warm to cold hues. We have chosen the progression from red to green, a warm to cold transition indicating high density in reds, gradually shifting to medium density in yellows, and low density in greens. The zero population values are symbolized in white to indicate the absence of the phenomena. Black would be another wise choice for the zero population values; however, black also implies “no data” within raster environments.

Output Maps

The following maps of Astoria and Seaside (figs. 11 and 12), two coastal cities in Clatsop County, Oregon, are output representations of the DME. The dasymetric maps have been symbolized using the steps outlined in the section above. In these figures, choropleth maps have been made to show the con-

trast in visualization between the two methods. The choropleth map assumes that population is distributed evenly throughout the census block-group, when in actuality there are many areas that are vacant or uninhabited. Due to the ease of use, choropleth maps have been the chosen approach for mapping population density, regardless of some of the inaccuracies associated with aggregated data. For the case study of Clatsop County, where there is a significant tsunami hazard risk, the dasymetric mapping approach provides detailed information indicating which areas are most vulnerable. In comparison to the choropleth maps shown in figures 11 and 12, the emergency responder is able to take the dasymetric map and focus on preparedness strategies down to the parcel level.

In a GIS, there are other functions that can be used to manipulate the output data from the DME. The ArcGRID is a seamless raster file enabling the user to select an area of interest, or a cluster of pixels, and quantify the population of that particular area of interest, regardless of source zone boundaries. This capability is available because the DME creates a surface representation of population density, rather than discrete polygons where one aggregated value represents the entire polygon. The output structure of the DME provides a superior dataset to the choropleth map for overlaying population density with other relative datasets. The areas that have zero population can be symbolized with no color and transparency can be used for the populated pixels to model a much more realistic view as to where people reside.

Summary

The objectives of this GIS software extension are (1) to automate the 3-class dasymetric-mapping method within a GIS environment, (2) to present the point-and click functionality of the DME, (3) highlight the options for optimal symbolization of the results, and (4) relate this software package to a real world application demonstrated in Clatsop County, Oregon. The software extension available here could increase the productivity, ease, and efficiency of researchers/analysts who work with U.S. Census Bureau data and spatial analysis. We have used this methodology to map the coastal

areas of Clatsop County, Oregon in an effort to reveal (1) a superior visualization of the population distribution, (2) areas that have the highest density in relation to the tsunami hazard zones, and (3) what areas need the most focus during a catastrophic event.

There has been significant research conducted on areal-interpolation and dasymetric-mapping methodologies. In the past, some major drawbacks for the use of dasymetric mapping methods were the time-intensive preparation of multiple datasets and transferring the methodology to function smoothly within a GIS environment. With these obstacles in mind, many researchers resort to using standardized U.S. Census Bureau data. By creating geospatial tool or extension that streamlines the areal interpolation process, the user has a cutting edge utility for remodeling population data with very easy point-and-click features. We continue to research the input data and its overall accuracy for the dasymetric-mapping technique, with the intention to help local communities prone to natural disasters. By generating

a downloadable tool with affordable data layers as inputs, this method will be an attractive alternative to the standard aggregated census data that limit geospatial research for emergency management and response.

Acknowledgments

We thank the USGS Geographic Analysis and Monitoring Program for funding this effort. Richard Taketa, Amy Mathie, Mara Tongue, and Susan Benjamin provided insightful comments and technical edits of this manuscript. Michael Gould (*mgould@usgs.gov*) is the chief software developer for the Dasymetric-Mapping Extension and is the primary contact for questions and comments on the extension software functionality. We also thank Jeremy Mennis and Torrin Hultgren for developing the original ArcObjects script based on methods of Mennis (2003).

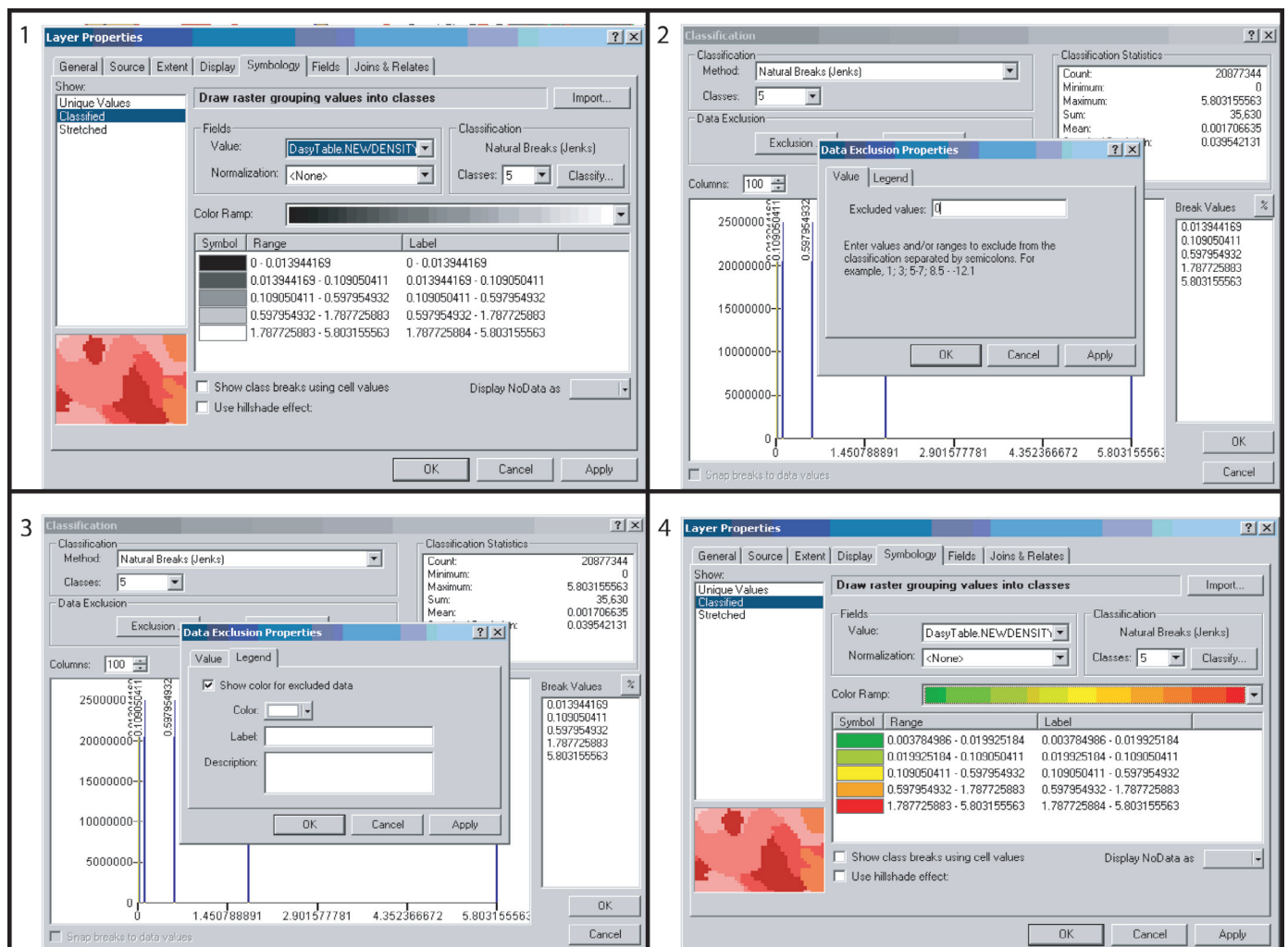


Figure 10. Screen images showing four steps to symbolizing dasymetric-mapping output data.

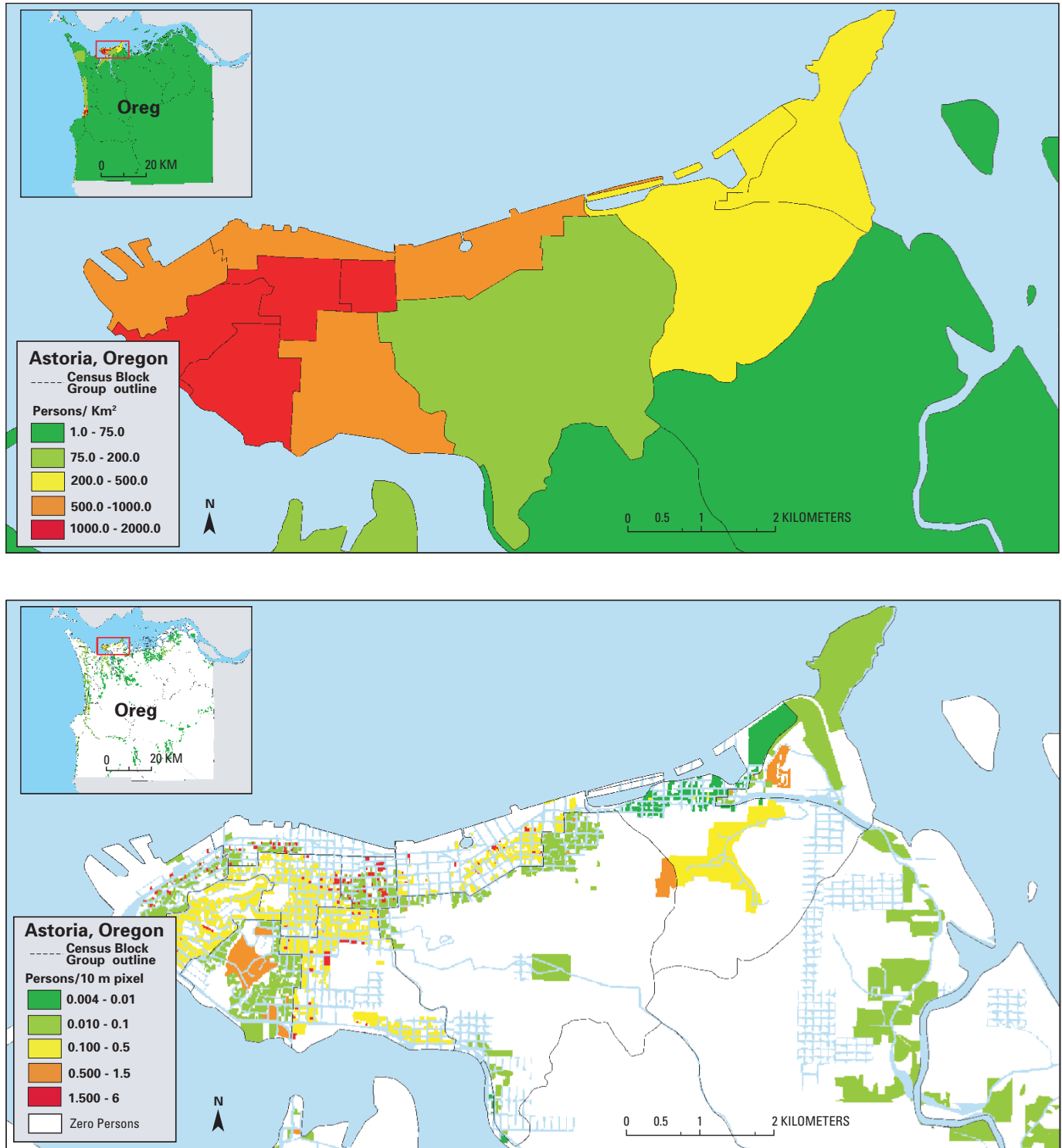


Figure 11. Choropleth map (top) and dasymetric map (bottom) of Astoria, Oregon.

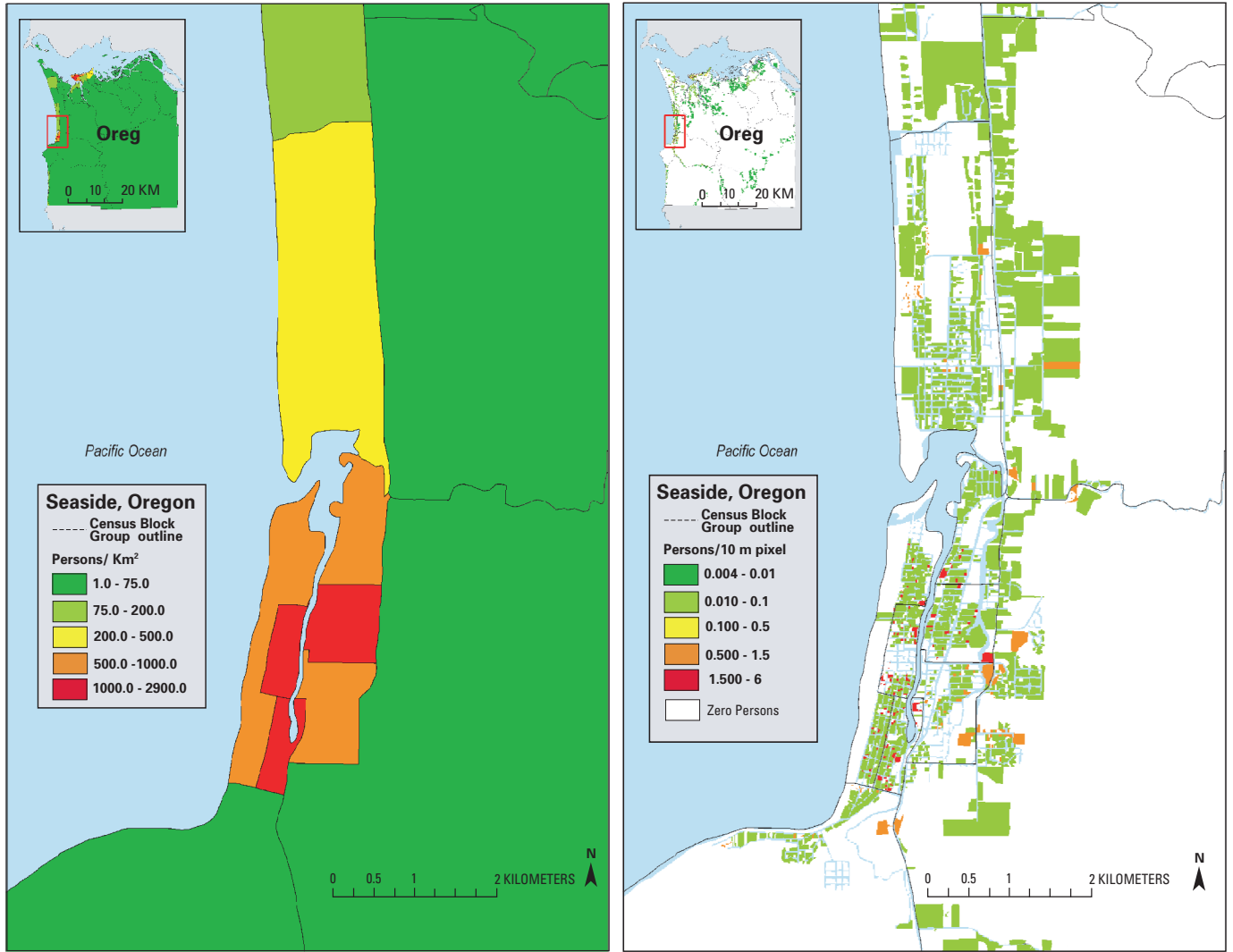


Figure 12. Choropleth map (left) and dasymetric map (right) of Seaside, Oregon.

References Cited

- Charpentier, M., 1997, The dasymetric method with comments on its use in GIS—Paper presented at the 1997 University Consortium for Geographic Information Science annual assembly and summer retreat, June 15-21, Bar Harbor, Maine.
- Eicher, C.L., and Brewer, C.A 2001, Dasymetric mapping and areal interpolation—implementation and evaluation: *Cartography and Geographic Information Science* v. 28, no. 2, p. 125-138.
- Fisher, P.F. and Langford, M., 1996, Modeling sensitivity to accuracy in classified imagery—A study of areal interpolation and dasymetric mapping: *Professional Geographer*. v. 48, no. 3, p. 299-309.
- Flowerdew, Robin and Green, Mick, 1994, Areal interpolation and types of data: In *Spatial Analysis and GIS*, Stewart Fotheringham and Peter Rogerson, (eds.): London, Bristol: Taylor & Francis Ltd., p. 121-146.
- Hawley, K., and Moellering, Harold, 2005, A comparative analysis of areal interpolation methods: *Cartography and Geographic Information Science*, v. 32, no. 4, p. 411-423.
- Herold, M., Müller, A., Gunther, S. and Scepan, J., 2002, Object-oriented mapping and analysis of urban land use/land cover using IKONOS data: *Proceedings of the 22nd EARSEL symposium*, Prague, June 2002.
- Holt, J.B., Lo, C.P., and Hodler, T.W., 2004, Dasymetric estimation of population density areal interpolation of census data: *Cartography and Geographic Information Science*, v.31, no. 2, p. 101-121.
- Holloway, S., Schumacher, J., and Redmond, R., 1997, People & Place: Dasymetric mapping using Arc/Info, in *Cartographic Design Using ArcView and Arc/Info*: University of Montana, Missoula, Wildlife Spatial Analysis Lab.
- Lam, N.S., 1983, Spatial interpolation methods—A review: *The American Cartographer*, v. 10, no. 2, p. 129-149.
- Langford, M., and Unwin, D.J., 1994, Generating and mapping population density surfaces within a geographical information system: *The Cartographic Journal*, v. 31, no. 1, p. 21-26.
- Mennis, J., 2003, Generating surface models of population using dasymetric mapping: *The Professional Geographer*, v. 55, no. 1, p. 31-42.
- National Academy of Sciences, 2007, National land parcel data—A vision for the future: Washington, D.C. [http://dels.nas.edu/dels/rpt_briefs/Land_Parcel_Data_Final.pdf], last accessed February 17, 2008].
- Oak Ridge National Laboratory, 2005, LandScan Global Population Database: Oak Ridge, Tenn. [<http://www.ornl.gov/gist/>], last accessed February 17, 2008].
- Ramroop, S., 1998, Appropriate selection of cartographic symbols in a GIS environment: *Proceedings of the Spatial Information Research Centre's 10th Colloquium*, University of Otago, New Zealand, November, 16-19, 1998.
- Tobler, W., 1979, Smooth pycnophylactic interpolation for geographical regions: *Journal of the American Statistical Association*, v. 74, no. 367, p. 519-536.
- Vogelmann, J.E., Howard, S.M., Yang, L., Larson, C.R., Wylie, B.K., and van Driel, N., 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*, v. 67, p. 650-662.
- Wright, J.K., 1936, A method of mapping densities of population with Cape Cod as an example: *Geographical Review* 26.
- Wood, N., and Good, J., 2004, Vulnerability of a port and harbor community to earthquake and tsunami hazards—the use of GIS in community hazard planning: *Coastal Management*, v. 32, no. 3, p. 243-269.
- Wu, S., Qiu, X., and Wang, L., 2005, Population estimation methods in GIS and remote sensing—A review: *GIScience and Remote Sensing*, v. 42, n. 1, p. 80-96.
- Yuan, Y., Smith, R. and Limp, W.F., 1997, Remodeling census population with spatial information from Landsat TM Imagery: *Computers, Environment, and Urban Systems*, v. 21, no. 3-4, p. 245-258.

Appendix—Equations for Dasymetric-Mapping Extension (Mennis, 2003)

$$(A) d_{uc} = P_{uc} / (P_{bc} + P_{lc} + P_{nc}), \quad (1)$$

where d_{uc} = population-density fraction of urbanization class u in county c , P_{uc} = population density (persons/900 m²) of urbanization class u in county c ,

P_{bc} = population density (persons/900 m²) of urbanization class h (high) in county c ,

P_{lc} = population density (persons/900 m²) of urbanization class l (low) in county c , and

P_{nc} = population density (persons/900 m²) of urbanization class n (nonurban) in county c .

$$(B) ba_{ub} = (n_{ub}/n_b)/0.33, \quad (2)$$

where a_{ub} = area ratio of urbanization class u in block-group b ,

n_{ub} = number of grid cells of urbanization class u in block-group b , and

n_b = number of grid cells in block-group b .

$$(C) f_{ubc} = (d_{uc} * a_{ub}) / [(d_{hc} * a_{hb}) + (d_{lc} * a_{lb}) + (d_{nc} * a_{nb})], \quad (3)$$

where f_{ubc} = total fraction of urbanization class u in block-group b and in county c ,

d_{uc} = population-density fraction of urbanization class u in county c , a_{ub} = area ratio of urbanization class u in block-group b ,

d_{hc} = population-density fraction of urbanization class h (high) in county c , d_{lc} = population-density fraction of urbanization class l (low) in county c , d_{nc} = population-density fraction of urbanization class n (nonurban) in county c ,

a_{hb} = area ratio of urbanization class h (high) in block-group b ,

a_{lb} = area ratio of urbanization class l (low) in block-group b , and

a_{nb} = area ratio of urbanization class n (nonurban) in block-group b .

$$(D) pop_{ubc} = (f_{ubc} * pop_b) / n_{ub} \quad (4)$$

where pop_{ubc} = population assigned to one grid cell of urbanization class u in block-group b and in county c ,

f_{ubc} = total fraction for urbanization class u in block-group b and in county c , pop_b = population of block-group b , and n_{ub} = number of grid cells of urbanization class u in block-group b .

This page intentionally left blank

