LandScan: A Global Population Database for Estimating Populations at Risk

Jerome E. Dobson, Edward A. Bright, Phillip R. Coleman, Richard C. Durfee, and Brian A. Worley

Abstract

The LandScan Global Population Project produced a world-wide 1998 population database at a 30- by 30-second resolution for estimating ambient populations at risk. Best available census counts were distributed to cells based on probability coefficients which, in turn, were based on road proximity, slope, land cover, and nighttime lights. LandScan 1998 has been completed for the entire world. Verification and validation (V&V) studies were conducted routinely for all regions and more extensively for Israel, Germany, and the southwestern United States. Geographic information systems (GIS) were essential for conflation of diverse input variables, computation of probability coefficients, allocation of population to cells, and reconciliation of cell totals with aggregate (usually province) control totals. Remote sensing was an essential source of two input variables—land cover and nighttime lights—and one ancillary database—high-resolution panchromatic imagery—used in V&V of the population model and resulting LandScan database.

Introduction

Natural and manmade disasters place vast populations at risk, often with little or no advance warning. Consider the following examples:

- An industrial plant releases hazardous chemicals into the atmosphere, as a Union Carbide plant did in Bhopal, India in 1984.
- A nuclear power plant releases radiation, as Chernobyl did in 1986.
- Toxic gases spread from a terrorist's bomb, as sarin did in Tokyo, Japan in 1994.
- A volcano erupts, as Mount Vesuvius did in 79 AD and Mount Pinatubo did in 1991, spewing ash and plumes of poisonous gases over populated areas.

These examples represent global threats to local places, and geographic information is essential for quick and effective response. How will the contaminant be dispersed? Where will it go? How many people are at risk? Who are they? Where are they? Emergency response by the United Nations, the United States, and other national and international organizations requires simulation of contaminant transport by air and water *plus* improved estimates of global population distribution.

Air diffusion models available today are capable of estimating contaminant plumes at spatial precisions far exceeding those of most official censuses. For many years, the U. S. Census Bureau has enhanced the precision of global population estimates through a manual procedure designed to allocate rural populations to 20- by 30-minute cells and urban populations to circles centered on major population concentrations. Yet, analysis of most hazardous releases requires data resolutions on the order of 1 square kilometer or even finer. To meet this need, Oak Ridge National Laboratory (ORNL) developed an automated procedure to allocate rural and urban population distributions to 30- by 30-second cells. The resulting population distribution can be used for (1) emergency response to natural disasters; nuclear, biological, and chemical accidents; terrorist incidents; or other threats; (2) humanitarian relief in famines and other long term disasters; (3) protection of civilian populations during regional conflicts; (4) estimation of populations affected by global sea level rise; and (5) numerous other environmental and demographic applications. The database should be of special interest to geographers, environmental scientists, emergency managers, and decision makers at national and international levels of government.

ORNL's Global Population Project, part of a larger global database effort called LandScan, collects best available census counts (usually at province level) for each country, projects aggregate populations to a target year (presently 1998), calculates a probability coefficient for each cell, and applies the coefficients to census counts which are employed as control totals for appropriate areas (usually provinces). Ideally, the polygons associated with aggregate populations are administrative units with accurate census counts, but the procedure will work for any polygon. The probability coefficient is based on slope, proximity to roads, land cover, nighttime lights, and an urban density factor. A geographic information system (GIS) is essential for conflation of diverse input variables, computation of probability coefficients, allocation of population to cells, and reconciliation of cell totals with aggregate (usually province) control totals. Remote sensing is an essential source of two input variables-land cover and nighttime lights-and one ancillary database-high-resolution panchromatic imagery-used in verification and validation (V&V) of the population model and resulting LandScan database.

Ambient Versus Residential Population

The resulting LandScan distribution represents an ambient population which integrates diurnal movements and collective travel habits into a single measure. This is desirable for purposes of emergency response and, fortuitously, is easier to accomplish with currently available global imagery and other geographic data. Consider, for example, the hypothetical case of a cell with a major multilane highway passing through an

Oak Ridge National Laboratory, P. O. Box 2008, MS 6237, Oak Ridge, TN 37831 (jed@ornl.gov).

Photogrammetric Engineering & Remote Sensing
Vol. 66, No. 7, July 2000, pp. 849–857.

0099-1112/00/6607–849$3.00/0

uninhabited desert. If a hazardous release contaminates the cell, many lives will be at risk even though no one lives there. Most official census counts, if available at such fine resolution, would show zero population because most national censuses are concerned with residential population based primarily on where people sleep rather than where they work or travel. In the LandScan procedure, population is apportioned to the cell based on the presence of a highway and, perhaps, on nighttime lights emanating from ambient traffic. Consider another cell containing a large agricultural field and no houses. Most censuses would place farm workers in their village residences and record zero populations for their fields. Yet, a few lives are at risk in the fields, depending on when the hazardous release occurs. Hence, our procedure shows a small population in the crop cell and a slightly reduced population in the village to suggest, albeit imprecisely, the collective time that villagers are in their fields rather than their homes. Even arid grassland cells have sparse populations assigned to simulate the movements of nomads and other herders. We integrate all ambient population into a single value for each cell and do not attempt to distinguish the timing of such movements. The same can be said of factories, airports, and other places of work and travel.

Best Available Population Databases

Census Counts

We reiterate that our purpose is to distribute populations based on their likely ambient locations integrated over a 24-hour period for typical days, weeks, and seasons. In contrast, most censuses count people at their nighttime residences. All census counts, including the official censuses of advanced nations like the United States, are stochastic estimates. Accuracy and precision are limited by census takers' access to homes, their understandings of personal work and travel habits, time of day, frequency of repetition, resources, and sometimes outright manipulation to meet political objectives. Many nations are reluctant to release detailed census counts, and some release only a national total. For most of the world, the best available official census data are at province level (i.e., one administrative division below national) and of varying age, i.e., up to several decades old. A few nations (e.g., Israel) release high-quality census counts for sub-provinces, but only a few release the geometry of sub-province boundaries in digital form (e.g., U. S. Census TIGER files).

Ultimately, ORNL analysts chose a single population count for each nation or province based on careful evaluation of arguments and evidence offered by demographers. Official census counts were acquired from published sources and evaluated skeptically. For most countries, the demographic literature is surprisingly rich; deficiencies are recognized by scholars; and adjustments have been proposed in published articles and reports, many of them available through the Internet. A table of census counts and growth rates by country and province is available from the lead author.

P-95 Circles and Rural Cells

Since 1965, the Geographic Studies Branch of the International Programs Center (IPC) of the U. S. Bureau of the Census has generated the most authoritative and, prior to 1995, the finest spatial resolution population database available for the whole Earth (Leddy, 1994). IPC acquires latest census counts; conducts extensive evaluations; projects total country population growth based on births, deaths, and migrations; distributes country population to small areas; and projects small area populations annually for 12 years. Rural populations are allocated to cells measuring 20 minutes latitude by 30 minutes longitude. In certain areas—such as the United States, western Europe, and Israel—rural populations are allocated to "mini-cells" measuring 5 by 7.5 minutes. Urban agglomerations of 25,000 people or more are covered by one or more circles encompassing at least 95 percent of the population. These features, ranging from 0.3 to 2.0 nautical miles in radius, are known as P-95 circles. Each circle must contain at least 5,000 people, and at least 80 percent of the area covered by large circles (1.1-km radius or greater) must be residential built-up. Smaller circles (0.9-km radius) often are placed on the expanding edge of cities in anticipation of future growth.

Plate 1 displays the distribution of P-95 circles, rural cells, and mini-cells in the San Francisco Bay area. At this latitude each rural cell covers about 1,600 sq km, each mini-cell about 100 sq km, and each LandScan cell about 0.7 sq km.

The Global Demography Project

The Global Demography Project (Tobler *et al.*, 1995), conducted by the National Center for Geographic Information and Analysis (NCGIA), developed a 1994 population database at a 5- by 5-minute resolution for most of the world (57° S to 72° N). This constitutes the finest resolution global population database yet produced. However, its utility is limited due to three factors acknowledged by its authors: (1) census data were obtained from the United Nations Statistical Division, which makes no attempt to evaluate the accuracy of census counts provided by

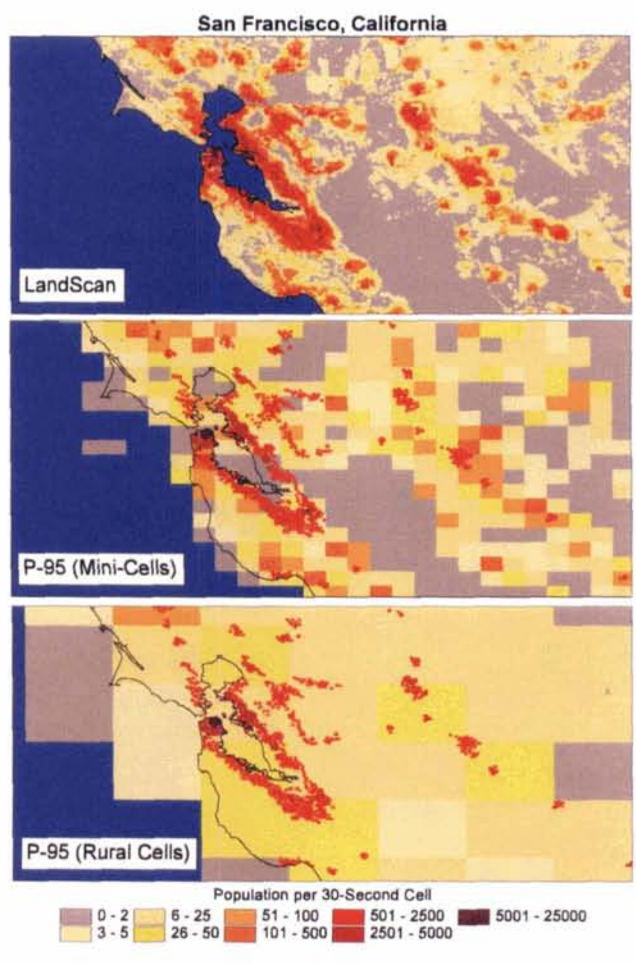

Plate 1. A comparison of LandScan cells (30 arc seconds by 30 arc seconds), P-95 circles and rural cells (20 arc minutes by 30 arc minutes), and P-95 circles and mini-cells (5 arc minutes by 7.5 arc minutes) for the San Francisco Bay area.

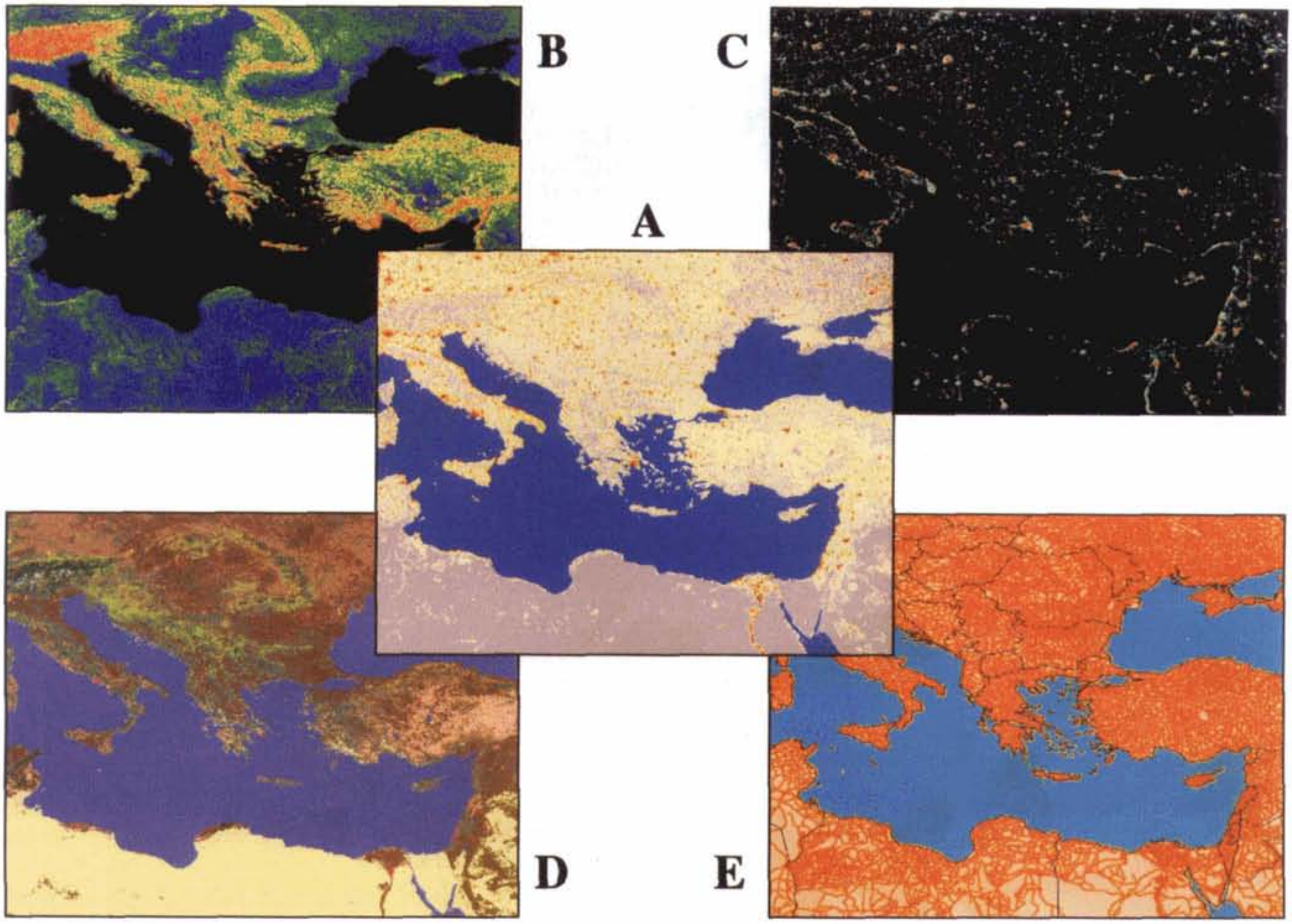

Plate 2. LandScan 1998 population distribution (A) and input data-slope (B), nighttime lights (C), land cover (D), and proximity to roads (E)-for the Mediterranean region at 30-arc-second resolution.

individual nations; (2) census dates, ranging from 1979 to 1994, were projected to 1994 based on annual growth rates by country, also provided by the United Nations; and (3) the algorithm employed to distribute population from administrative units (usually provinces) to cells is purely cartographic and is based on population alone. The authors note certain types of errors resulting from these factors, and suggest that improvement would result from a "smart" interpolation or co-Kriging that incorporates ancillary data such as location and size of towns and cities, roads, railroads, natural features, and nighttime lights.

Input Variables

Calculation of the probability coefficient for each cell depends on publicly available databases offering worldwide coverage of roads, slope, land cover, and nighttime lights at scales of 1:1,000,000 or larger and resolutions of 1 km or finer. The sources and characteristics of current databases are discussed in this section. All data are processed and transformed into a 30- by 30-second latitude/longitude grid cell system.

Roads

Transportation networks (i.e., roads, railroads, airports, and navigable waterways) are primary indicators of population. Roads are especially indicative because of their vital role in human settlements with or without other forms of transport. It would be helpful to know the location of all roads and to calculate road densities as suggestive of population densities, but this is not possible for most of the world. The United States is an exception due to the availability of U. S. Census TIGER files which include the geometry of local roads and even some private driveways and farm roads. The best global coverage of road networks is from the National Imagery and Mapping Agency's (NIMA) Vector Smart Map (VMAP) series (Plate 2). VMAP-Level 0 (formerly Digital Chart of the World) is publicly available and covers the entire world at 1:1,000,000 scale. We consider VMAP-Level 0 a staple source for global coverage of road networks, though we plan to include VMAP-Level 1 data (1:250,000 scale) in future iterations as tiles become available.

Slope

LandScan employs NIMA's Digital Terrain Elevation Data (DTED) Level 0, 30 Arc Second Terrain Data (Plate 2) from which we calculate a single slope gradient for each 1-km LandScan cell. Slope is an important variable in the LandScan population probability coefficient because most human settlements occur on flat to gently sloping terrain. Even in regions noted for hillside settlement, relative measures of slope correspond (inversely) with population density. A better measure of slope would be the area (at resolutions approaching the typical size of individual home sites) in each slope category, expressed as a percentage of LandScan cell area. LandScan's slope resolution is limited, however, by data availability and by the processing burden that would be required for global coverage.

Land Cover

Perhaps the best single indicator of population density is land-cover type. With local knowledge and well-structured *in situ* sampling, one conceivably might determine average population densities per unit of area for each land-cover type which then could be multiplied times the total area occupied by that type. In most regions, population would range from extremely low density in desert, water, wetlands, ice, or tundra land cover to high density in developed land cover associated with urban land use. Arid grasslands, forests, and cultivated lands would range between. Globally, of course, such rigorous *in situ* sampling is infeasible, especially in politically sensitive areas. Alternatively, LandScan analysts assign relative weights to each land-cover type and employ these weights in calculating the probability coefficient for each cell.

Even at a 1-km resolution, land cover can be a good indicator of relative population density, and its efficacy improves as resolution approaches the typical size of individual homesites. For example, the National Oceanic and Atmospheric Administration's (NOAA) Coastal Change Analysis Program (C-CAP) has demonstrated that high intensity developed and low intensity developed land cover can be distinguished reliably for coastal regions of the United States with Landsat Thematic Mapper (TM) imagery at a 30-m resolution (Dobson *et al.*, 1995). Currently, the best land-cover database available worldwide is the U. S. Geological Survey's (USGS) Global Land Cover Characteristics (GLCC) database (Plate 2) derived from Advanced Very High Resolution Radiometry (AVHRR) satellite imagery at a 1-km resolution (Loveland, 1991). Globally, GLCC is the staple land-cover database for calculation of LandScan probability coefficients. Regionally, we found it reasonably reliable for all land-cover types except wetlands and developed lands, but there is considerable variation in accuracy from cell to cell. In test comparisons in the United States, most C-CAP wetlands were recorded as water in GLCC. For all areas we tested in the Middle East, GLCC's developed land-cover category was a rasterized version of VMAP-Level 0's "populated polygons," with attendant limitations which are discussed in the following section.

Global land-cover databases are expected to improve as new satellite data become available. The MODIS (200 to 500 m resolution, 36 spectral bands) satellite likely will replace AVHRR as the staple data source. Landsat Thematic Mapper (TM), Advanced Land Imager (ALI), and Hyperion (all at 30 m) are potential sources for local area coverage, and these may be augmented in many local applications with finer resolution sources such as SPOT (10/20 m). New commercially available "small sat" data may be employed in certain instances to enhance spatial precision, temporal frequency, or spectral definition.

Populated Places

VMAP-Level 0 contains three categories of human settlement features. Two of them are point features distinguished only as "named" or "unnamed" populated places; the other consists of polygon boundaries for larger urban areas. Attributes for named populated places and populated polygons provide the name but not the population count for each place. Populated polygons originally were digitized from small-scale maps, sometimes aeronautical charts dating from the 1970s.

We matched the populated polygons with nighttime lights (discussed in the following section) and assigned a greater probability weighting for LandScan cells containing both a populated polygon and nighttime lights than that for cells containing only nighttime lights.

Nighttime Lights

Several deficiencies of the previously discussed databases can be overcome with satellite data produced by the Defense Meteorological Satellite Program (DMSP) which measures nighttime light emanating from the Earth's surface at a 1 km resolution (Elvidge *et al.*, 1997; Sutton *et al.*, 1997; Sutton, 1997). LandScan employs the Nighttime Lights of the World light frequency data processed and provided by NOAA's National Geophysical Data Center (NGDC) (Plate 2). Frequency data cover the Northern Hemisphere and South America, but most areas south of the equator are limited to a binary value indicating lights present versus no lights present. A better source would be data from the DMSP Operational Line Scanner (OLS) data which measure light intensity but these data have not been processed globally and released to the public.

Investigating the efficacy of nighttime lights for estimating population in the United States, Sutton *et al.* (1997) found that saturated pixels (i.e., adjusted pixel value of 64) cover almost 8 percent of the territory of the contiguous 48 states and account for about 80 percent of the population in those states. Conversely, about 17 percent of the population, occupying about 90 percent of the land area, is dispersed too sparsely for detection (i.e., adjusted pixel value of one) by this particular sensor. Sutton (1997) further investigated the correlation of nighttime lights with population density, and his model accounted for 25 percent of the variation in population density. Thus, at the high end of the population/light spectrum, no further distinction of population densities is possible once light saturation occurs. At the low end of the spectrum, no further distinction is possible in pixels with undetected lights. Sutton *et al.* (1997) suggest that nighttime lights "might also be used as a primary informant to a 'smart' interpolation program for modeling human population distributions in areas where only large scale aggregate data are available." They recommend candidate variables to include city locations, coastlines, landforms, railroads, airports, harbors, and rivers.

Exclusion Areas

Areas with ambient populations of less than one person per LandScan cell are determined by identifying the Census Bureau's 20- by 30-minute cells with zero rural populations and no P-95 circles. These are then compared with populated places, roads, land cover, and nighttime lights. If none of these databases contradict the 20- by 30-minute cell data, zero population is assigned to all LandScan cells inside the 20- by 30-minute zero-population cells. The exclusion is then extended to adjacent LandScan cells if they show no indicators of population (roads, nighttime lights, etc.), even if the LandScan cells lie within 20- by 30-minute cells that contain population. Water cells and ice cells are assigned zero population.

Urban Density Factor

We matched the point locations and diameters of P-95 circles with nighttime lights, and increased the probability weighting for LandScan cells containing both features over cells containing only nighttime lights. The associated P-95 population values proportionally increased the probability weighting, but absolute P-95 values were not employed in the final calculation of LandScan cell populations.

Coastlines

Considerable effort is required to reconcile the positional accuracy of diverse global databases, and mismatches among databases are most conspicuous on coastlines. Globally, LandScan coastlines are based on NIMA's World Vector Shoreline (WVS) at 1:250,000 scale. Typically, this coastline differs somewhat from the related line representing the seaward boundary of administrative units, and both of these differ from the land/water boundary indicated on the GLCC gridded database. In the final

LandScan 1998 Global Population Database, the WVS took precedence, and no population was apportioned to cells extending more than one-half cell beyond the WVS coastline.

Population Model

Best available census counts (usually at province level) are allocated to 30- by 30-second cells through a "smart" interpolation based on the relative likelihood of population occurrence in cells due to road proximity, slope, land cover, and nighttime lights. Probability coefficients are assigned to each value of each input variable, and a composite probability coefficient is calculated for each LandScan cell. Coefficients for all regions are based on the following factors:

- Roads, weighted by distance from cells to roads;
- Slope, weighted by favorability of slope categories;
- Land Cover, weighted by type with exclusions for certain types; and
- Nighttime Lights of the World, weighted by frequency.

The resulting coefficients are weighted values, independent of census data, which can then be used to apportion shares of population counts within any area of interest. Coefficients vary considerably from country to country and even from province to province. The generic model remains the same for all regions, but the probability weights of individual variables must be customized for each area due to economic, physical, and cultural factors. For example, nighttime lights tend to be intense in energy-rich nations, like Kuwait, and less intense in energy-poor nations like North Korea. The main highway westward from Kuwait City is brightly lit with streetlights, giving a false impression of urban populations sprawled across what is actually uninhabited desert. Similarly, population densities of cultivated land in one region may differ greatly from cultivated land in another region. All weighting values for all areas are retained and archived as metadata for future reference.

Control totals can be for any administrative unit (nation, province, district, minor civil division) or arbitrary polygon for which census data are available. The resulting population distribution is normalized and compared with appropriate control totals to ensure that aggregate distributions are consistent with census control totals. Successful operation of the model has been demonstrated for various control totals, control areas, and weighting values.

Verification and Validation

Verification of any spatially explicit global population database is inherently limited by the difficulty of establishing a suitable reference database. The ideal would be actual counts for sample areas at the same resolution or finer resolution than the database being evaluated. Even in the United States, such data are available only for urbanized areas, and they do not exist for most of the world. Thus, verification of data and validation of underlying models necessarily depends on indirect measures, including

- *Verification Based on Best Available Census Counts at Finest Available Resolution*
  This check is conducted for all countries comprising the LandScan Global Population Database. However, the results do not constitute an accuracy assessment because the same data are employed in calculation of the LandScan data.
- *Surrogate Area Analysis*
  The results of V & V in areas of good reference data (e.g., United States) may be extrapolated to areas of poor reference data. Care should be taken to match areas that are not too distant and whose physical, cultural, and economic circumstances are similar.
- *Ancillary Data Analysis*
  High resolution population estimates may be compared to indicators of population (e.g., buildings, settlements, or pertinent land-cover classes such as high intensity developed, low intensity developed, and cultivated) derived from satellite imagery or aerial photographs. The imagery must not have been employed in the calculation of the LandScan database and should be at a finer spatial resolution than the input data. However, buildings, even those that appear to be residential, are only a subjective, non-quantative indicator of population.
- *Input Data Analysis*
  Verification, validation, and sensitivity analyses may be conducted for input data (in this case, land cover, elevation, roads, and nighttime lights).

In the following sections, we summarize a variety of V&V efforts for selected areas in the southwestern United States, Germany, and Israel. Additional maps and tables are available from the lead author.

Census Validation for the Southwestern United States

The United States provides a unique opportunity for V&V of the LandScan methodology due to the availability of population counts and census unit boundary geometries at fine spatial resolution. We focused on the southwestern region due to its arid climate and other physical similarities with the Middle East. State population counts for Arizona, California, Nevada, and Utah were distributed to 30- by 30-second cells based on the LandScan population model with coefficients modified to account for distinctive regional differences between the southwestern United States and the Middle East. The resulting cell values were then aggregated to counties and compared to actual census counts at county level.

For V&V purposes, the input census data were deliberately entered as aggregate state totals, artificially limiting the calculation to the type of census counts available for most of the world, presuming states to be equivalent to countries elsewhere and counties equivalent to provinces elsewhere. California at 411,000 sq km, for example, compares with Iraq at 438,000 sq km but Iraq contains many provinces, so the input data available there are far more detailed than that used for the southwestern United States V&V analysis. This effect is even greater for small countries like Qatar, Bahrain, and Oman.

The overall correspondence is such that 87.8 percent of the simulated LandScan population for the southwestern United States corresponds with the county totals of the official census (i.e., only 12.2 percent of the total population is placed in a county other than that indicated in the official census count). Respectively this correspondence is 90.4 percent in Arizona, 88.2 percent in California, 88.9 percent in Nevada, and 74.9 percent in Utah. The results indicate a difference of less than 20 percent ($\pm$) between the census count and the simulated coarse LandScan ambient population in 40.3 percent of the counties, and these counties contain the vast majority of the total population. Of course, small percentages of urban populations redistributed to rural cells cause large percentage differences in sparsely populated counties. Thus, most of the sizable percentage differences occur in Nevada, Utah, and a few sparsely populated counties in Arizona and California. Most substantial differences occur in counties whose populations are negligible at a regional scale.

Among urban counties, the most conspicuous "over" estimation of simulated ambient population is for Sacramento, California (49.0 percent), which fairs unusually well in LandScan variables. As the state capital, Sacramento has a disproportionately large number of administrative buildings compared with its resident population. As with other administrative and institutional centers, such as college towns, ambient populations may justify the higher LandScan values relative to official census counts.

Again, we remind the reader that none of the differences discussed above apply to the final LandScan database which is based on tract level census data. Indeed, some of the advantages

Table 1. LandScan Population and P-95 Circle/Rural-Cell Population Compared with Census Counts in the Southwestern United States, Germany, and Israel

Location	Census Aggregation	Database	Percentage of Areas with Less Than 10 % (±) Difference	Percentage of Areas with More Than 50 % (±) Difference	Standard Deviation of Absolute Difference Per Area	Percentage of Total Population Allocated Differently from Census Count
SW US	6811 Census Tracts	LandScan	100.0	0	0	0
		P-95/ Rural Cell	16.2	42.5	2,429	41.0
	119 Counties	LandScan	100.0	0	0	0
		Simulated Coarse LandScan	23.5	31.1	79,979	12.2
Germany	445 Counties	LandScan	100.0	0	0	0
		P-95\Rural Cell	55.3	0.7	46,194	11.3
Israel	14 Sub-Provinces	LandScan	100.0	0	0	0
		Simulated Coarse LandScan	50.0 (Excluding Jerusalem and Tel Aviv)	0	21,030	9.0 (Excluding Jerusalem and Tel Aviv)
		P-95/Rural Cells	42.9	0	52,705	13.2

claimed for LandScan, such as recognition of ambient populations in state capitals and national parks, actually may diminish with finer resolution census data input. Consider, for example, that our simulated statewide calculation for Arizona may account for Arizonans who depart from Phoenix to visit the Grand Canyon while our final database only accounts for movements within each census tract that touches the Grand Canyon. As a global policy, we have adopted the geometry associated with the "best available census count" as the areal unit for which such travel patterns will be reconciled. We interpret the simulated results in the southwestern United States to mean that the LandScan algorithm works as intended.

P-95 Circles/Rural Cells Compared to Tract-Level Census Data in the Southwestern United States

For comparison, consider the accuracy and precision inherent in other attempts to characterize local population distributions. We mapped P-95 circles and rural cells (employing minicells wherever available) and intersected them with census tracts in the southwestern United States, based on a uniform distribution within each circle or cell (Table 1). (See Plate 1 for an illustration of the relative sizes of LandScan cells, P-95 circles, rural cells, and mini-cells.) A histogram of percentage differences between these estimates and official census counts for census tracts depicts substantial differences. Of 6,811 census tracts, only 1,105 (16.2 percent) show differences of less than 10 percent (±), and another 1,077 (15.8 percent) show differences of 10 to 20 percent (±). Some 701 census tracts (10.3 percent) differ by 100 percent or more. Most of the large differences occur in census tracts with small populations, and they are due not to error *per se* but to the spatial resolution of the P-95 circles and rural cells.

For comparison, a histogram of percentage differences between final LandScan estimates and official census counts for census tracts in the southwestern United States would show a difference of 0 for every census tract. The correspondence is perfect because census tracts are employed as control totals in the LandScan calculation. Thus, it is the finer spatial resolution of LandScan, rather than any fundamental error in the P-95/rural-cell values, that results in this highly favorable comparison.

P-95 Circles/Rural Cells Compared to County Census Data in Germany

We mapped P-95 circles and rural cells (employing mini-cells that were available for most of Germany) and intersected them with counties in Germany, based on a uniform distribution within each circle or cell (Table 1). A histogram of percentage differences between these estimates and official census counts for counties depicts substantial differences. Of 445 counties, 246 (55.3 percent) show differences of less than 10 percent (±), and none differ by 100 percent or more. Most of the large differences occur in counties with small populations, and they are due not to error *per se* but to the spatial resolution of the P-95 circles and rural cells.

For comparison, a histogram of percentage differences between final LandScan estimates and official census counts for counties in Germany would show a difference of 0 for every county. The correspondence is perfect because counties are employed as control totals in the LandScan calculation. Thus, it is the finer spatial resolution of LandScan, rather than any fundamental error in the P-95/rural-cell values, that results in this highly favorable comparison.

P-95 Circles/Rural Cells Compared to Sub-Province Census Data in Israel

We mapped P-95 circles and rural cells (employing mini-cells

Table 2. P-95 Circle/Rural-Cell Population Compared with Sub-Province Census Counts in Israel

Sub-Province	Census Count Projected to 1998	P-95/Rural Cell	Sub-Province Difference	Percentage Difference
51-Tel Aviv	1,188,251	1,178,948.50	−9,302.50	−0.78
11-Jerusalem	706,049	513,767.19	−192,281.81	−27.23
31-Haifa	520,257	466,372.03	−53,884.97	−10.36
62-Be'er Sheva	465,208	368,146.81	−97,061.19	−20.86
42-Petah Teqwa	458,155	414,763.31	−43,391.69	−9.47
24-Akko	441,333	353,744.50	−87,588.50	−19.85
44-Rehovat	378,181	341,382.72	−36,798.28	−9.73
61-Ashqelon	367,517	345,396.72	−22,120.28	−6.02
23-Yizre'el	364,389	351,905.34	−12,483.66	−3.43
41-Sharon	296,227	235,373.11	−60,853.89	−20.54
32-Hadera	270,242	157,740.36	−112,501.64	−41.63
43-Ramla	178,507	155,275.33	−23,231.67	−13.01
22-Kinneret	90,080	99,471.34	9,391.34	10.43
21-Zefat	89,768	94,159.34	4,391.34	4.89
TOTAL	5,814,164	5,076,446.59	765,282.75[1]	13.16[2]

[1]Total sub-province difference disregarding sign.
[2]Total sub-province difference disregarding sign/Country total * 100.

wherever available) and intersected them with sub-provinces in Israel based on a uniform distribution within each circle or cell (Table 1). An analysis of percentage differences between these estimates and official census counts for sub-provinces depicts substantial differences (Table 2). Of 14 sub-provinces, only 6 (42.9 percent) show differences of less than 10 percent (±), and another 4 (28.6 percent) show differences of 10 to 20 percent (±). All other sub-provinces differ by less than 28 percent (±). These differences are due not to errors *per se* but to the spatial resolution of the P-95 circles and rural cells.

For comparison, a histogram of percentage differences between final LandScan estimates and official census counts for sub-provinces in Israel would show a difference of 0 for every sub-province. The correspondence is perfect because

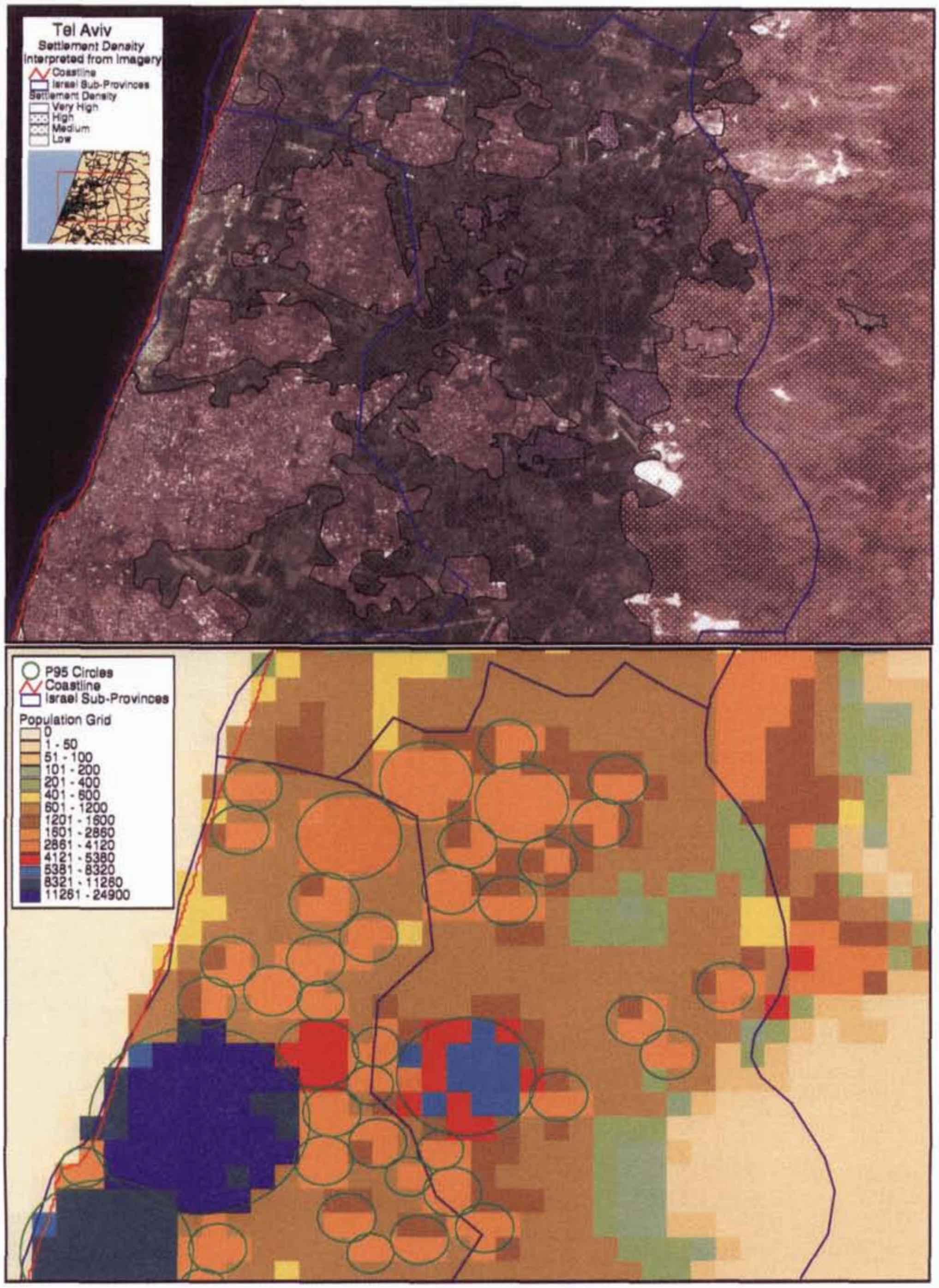

Plate 3. LandScan 1998 population distribution compared to high-resolution (10-m) panchromatic imagery for Tel Aviv, Israel, P-95 circles, and photo-interpreted building denstities.

Table 3. Simulated LandScan Population Estimates Compared with Sub-Province Census Counts in Israel. All LandScan Values in This Table Are Artificially Coarse for V&V Purposes Only.

Sub-province	Census Count Projected to 1998	Simulated Coarse LandScan	Sub-Province Difference	Percentage Difference	Final LandScan
51-Tel Aviv	1,188,251	1,188,251.00	0.00	0.00	1,188,251.00
11-Jerusalem	706,049	706,049.00	0.00	0.00	706,049.00
31-Haifa	520,257	523,334.13	−3,077.13	−0.59	520,257.00
62-Be'er Sheva	465,208	421,900.25	43,307.75	9.31	465,208.00
42-Petah Teqwa	458,155	482,038.31	−23,883.31	−5.21	458,155.00
24-Akko	441,333	513,124.78	−71,791.78	−16.27	441,333.00
44-Rehovat	378,181	346,177.44	32,003.56	8.46	378,181.00
61-Ashqelon	367,517	410,824.78	−43,307.78	−11.78	367,517.00
23-Yizre'el	364,389	359,246.41	5,142.59	1.41	364,389.00
41-Sharon	296,227	330,719.81	−34,492.81	−11.64	296,227.00
32-Hadera	270,242	267,137.66	3,104.34	1.15	270,242.00
43-Ramla	178,507	152,134.48	26,372.52	14.77	178,507.00
22-Kinneret	90,080	58,382.52	31,697.48	35.19	90,080.01
21-Zefat	89,768	54,843.50	34,924.50	38.91	89,768.00
TOTAL	5,814,164	5,814,164.06	353,105.56[1]	6.07[2]	5,814,164.01
TOTAL Excluding Jerusalem and Tel Aviv	3,919,864	3,919,864.06	353,105.56[1]	9.01[2]	3,919,864.01

[1]Total sub-province difference disregarding sign.
[2]Total sub-province difference disregarding sign/Country total * 100.

sub-provinces are employed as control totals in the LandScan calculation. Thus, it is the finer spatial resolution of LandScan, rather than any fundamental error in the P-95/rural-cell values, that results in this highly favorable comparison.

Census Validation for Israel

Census validation can be conducted for certain foreign areas that have recent, high-quality, fine-resolution census counts. For Israel, we (a) employed province level census totals as input data, (b) simulated ambient population for 1-km cells, (c) aggregated cell values for sub-provinces, and (d) compared to official census data for sub-provinces. The results (Table 3) indicate good correspondence between census data and LandScan data, except for the same trend observed in the southwestern United States, i.e., areas with small populations were "over" estimated. The two sub-provinces with the smallest official census counts showed differences of 35 and 39 percent, respectively. Even so, the overall correspondence is such that 91 percent of the simulated LandScan population for Israel corresponds with the sub-province totals of the official census (i.e., only 9 percent of the total population is placed in a sub-province other than that indicated in the official census count). Conversely, among sub-provinces with census populations of 100,000 or more, differences range from 1.1 percent to 16.3 percent (±). For completeness, Jerusalem and Tel Aviv are shown in the table, but they were omitted from the V&V analysis because each contains only one sub-province.

Again, this V&V analysis is a conservative assessment based on artificially coarse results obtained by keeping the input data at unnecessarily high levels of aggregation. The final LandScan results are based on sub-province input data and will correspond precisely with official census totals for sub-provinces.

Comparison with Tel Aviv Imagery

An ancillary data analysis for Tel Aviv, Israel (Plate 3) reveals excellent correspondence between LandScan gridded population densities and developed land cover identifiable on high resolution panchromatic imagery. In total, the image contains about 21 settlements identifiable through visual interpretation. Of these, 17 appear as elevated population values in the LandScan database. The image contains 42 settlements designated as P-95 circles; all of these also appear as elevated population values in LandScan. This is not surprising because the locations of P-95 circles are used in the LandScan calculation. Conversely, however, in the easternmost sub-province of the image one substantial settlement, not identified as a P-95 circle, was identified by elevated population values in LandScan and does appear on the imagery.

Throughout the image, abrupt LandScan gradients correspond with abrupt shifts from developed land to sparsely settled arid land. This correspondence is conspicuous, for instance, in the southwestern quadrant of the image. In one case, however, the LandScan gradient appears artificially abrupt on a province boundary.

Plume Intersections with Population Databases

Our main thrust is to estimate populations at risk, and that often means intersecting contaminant plumes with LandScan cells. We tested the LandScan database in comparison to P-95 circles/rural-cell distributions and official census counts for notional plumes in Germany and the southwestern United States. The results (Table 4) indicate that LandScan produces more precise and accurate results by a considerable margin for small plumes and by a non-negligible margin for large plumes. Again, however, V&V is limited to residential counts, because ambient counts are not available.

Conclusions

LandScan provides global coverage of population at 30 by 30-second resolution, the finest spatial resolution yet developed, employing a "smart" interpolation procedure based on variables similar to those recommended by Tobler *et al.* (1995), Sutton (1997), and Sutton *et al.* (1997). V&V conducted in the southwestern U. S., Israel, and Germany indicate that greater spatial precision can been be achieved with no sacrifice in aggregate accuracy compared to previous global population databases. Indeed, LandScan's inherent correspondence with best available census counts for finest available census units actually represents an improvement in accuracy over previous global population databases. Indeed, for Israel even the simulated coarse LandScan results matched official census counts better than did P-95 circles/rural cells (9.0 percent versus 13.2 percent, respectively). In addition, high resolution imagery for Tel Aviv, Israel shows excellent correspondence between Land-Scan cell values and settlements identifiable on the imagery. Census validation efforts in the United States and Israel

TABLE 4. LANDSCAN POPULATION AND P-95 CIRCLE/RURAL-CELL POPULATION COMPARED WITH CENSUS COUNTS FOR PLUME INTERSECTIONS IN THE SOUTHWESTERN UNITED STATES AND GERMANY

Location	Area (sq km)	Census Count	LandScan Number	LandScan Percentage Difference	P-95 Circle/ Rural Cells Number	P-95 Circle/ Rural Cells Percentage Difference
Southwestern United States		By Block				
Polygon 1	3	52,870	57,219	+8.2	26,160	−50.5
Polygon 2	24	20,117	17,192	−14.5	10,868	−46.0
Polygon 3	111	202,038	192,497	−4.7	226,156	+11.9
Polygon 4	2,000	192,580	193,920	+0.7	211,807	+10.0
Germany		By County				
Polygon 1	947	2,254,863	2,254,863	0	1,969,491	−12.7
Polygon 2	22,143	7,237,523	7,237,523	0	7,780,000	+7.5

indicate that an overwhelming majority of the total population is properly apportioned to census areas, even when LandScan is artificially constrained to unnecessarily coarse aggregations of census input data. Most of the significant differences in the southwestern United States and in Israel occur in sparsely populated areas.

LandScan 1998 appears to be the most suitable, currently available global database for estimating populations at risk. It compares favorably with census counts and P-95 circles/rural cells in terms of accuracy and precision, covers the entire world for a consistent date, and offers a single data format for global applications. Ongoing research is addressing improvements such as distinction between daytime and nighttime populations, age and gender pyramids by country, and higher resolution distributions for urban areas.

Acknowledgments

This research was supported by the United States Department of Defense. We express our gratitude to William G. McArthur for cartographic and image processing assistance and to Margaret E. Setser for assistance in data collection and manuscript preparation. We are especially thankful to all those cited in our references who contributed to the ancillary databases on whose quality and availability LandScan relies.

References

Dobson, J.E., E.A. Bright, R.L. Ferguson, D.W. Field, L.L. Wood, K.D. Haddad, H. Iredale, III, V.V. Klemas, R.J. Orth, and J.P. Thomas, 1995. *NOAA Coastal Change Analysis Program; Guidance for Regional Implementation, Version 1.0*, NOAA Technical Report NMFS 123, Scientific Publications Office, National Marine Fisheries Service, NOAA, Seattle, Washington, 92 p.

Elvidge, C.D., K.E. Baugh, E.A. Kihn, H.W. Kroehl, and E.R. Davis, 1997. Mapping city lights with nighttime data from the DMSP Operational Linescan System, *Photogrammetric Engineering & Remote Sensing*, 63(6):727–734.

Loveland, T., J. Merchant, D. Ohlen, and J. Brown, 1991. Development of a land-cover characteristics database for the conterminous U.S., *Photogrammetric Engineering & Remote Sensing*, 57(11): 1453–1463.

Leddy, R., 1994. Small area populations for the United States. Presented at the Annual Meeting of the Association of American Geographers, San Francisco, California **(http://infoserver.ciesen.org/datasets/cir/jdysedit.html)**.

Sutton, P., 1997. Modeling population density with night-time satellite imagery and GIS, *Computers, Environment, and Urban Systems*, 21(3/4):227–244.

Sutton, P., D. Roberts, C. Elvidge, and H. Meij, 1997. A comparison of nighttime satellite imagery and population density for the continental United States, *Photogrammetric Engineering & Remote Sensing*, 63(11):1303–1313.

Tobler, W.R., U. Deichmann, J. Gottsegen, and K. Maloy, 1995. *The Global Demography Project*, Technical Report No. 95-6, National Center for Geographic Information and Analysis, University of California at Santa Barbara, Santa Barbara, California, 75 p.

Received 11 February 1999; revised and accepted 02 August 1999)