Accuracy of Road Density Estimates Derived from USGS DLG Data for Use in Environmental Applications

Timothy G. Wade, James D. Wickham, and David F. Bradford

Abstract

Collection and analysis of information on accuracy of spatial data is a growing and active field of research, but has tended to emphasize land cover. Roads are another spatial data set that is widely used in environmental studies, but information on its accuracy is generally lacking. There are several sources of road data, but all are based on U.S. Geological Survey (USGS) Digital Line Graph (DLG) data. A three-part weight-of-evidence approach is used to evaluate accuracy of road density estimates based on 1:100,000-scale USGS DLG data at different spatial scales. The first two components use population to estimate accuracy of road density, because roads are built to improve human access. The third part compares USGS DLG roads with roads derived from digital orthophotoquads. The results suggest that estimates of road density using USGS DLG road data are of adequate accuracy for use in habitat fragmentation or other studies requiring road density data. Commercially available road data would not likely provide significantly improved estimates of density.

Introduction

Accuracy assessment of spatial data is a growing field of research because of its broad utility in environmental and other studies (Walsh *et al.*, 1987; Lunetta *et al.*, 1991; Hess, 1994). However, most studies have focused on land-cover misclassification (Fenstermaker, 1994; Janssen and van der Wel, 1994) and, to a lesser extent, errors in Digital Elevation Models (DEM) (Walsh *et al.*, 1987; Bolstad and Stowe, 1994). There has been little investigation of error in other spatial data. One example is roads, which are important factors in habitat and water resource studies.

Several studies have shown that roads adversely affect larger animals by fragmenting habitat. Elk (*Cervus elaphus*) habitat was reduced by 75 percent at a road density of only 2 km/km^2 (Lyon, 1979; Lyon, 1983). Gray wolf (*Canis lupus*) populations in Wisconsin failed to survive at road densities greater than 0.93 mi/mi^2 (Thiel, 1985), and red fox (*Canus vulpes*) avoids roads (Storm *et al.*, 1967). These findings illustrate that, for a given species, habitats (e.g., forests) which seem suitable in terms of size and connectivity may not be suitable because of the existing road network. Measurements commonly derived from land-cover, such as habitat size and connectedness, might not detect this problem, given that roads are not typically included in land-cover data unless they are of sufficient width (e.g., interstate highway).

Roads are also one of two principal aspects of impervious surface estimates (Schueler, 1994). Quality of water resources tends to decline as the amount of impervious surface increases (Schueler, 1994). Road density can be used as an approximation of impervious surface to rank watersheds according to potential impacts on water resources (Jones *et al.*, 1997; Wickham *et al.*, in press).

Road data are available from the federal government in the form of U.S. Geological Survey (USGS) Digital Line Graphs (DLG) and Bureau of the Census TIGER files. Commercially produced data are available from Etak Inc. (Menlo Park, California), Wessex (Winnetka, Illinois), and DeLorme Inc. (Freeport, Maine), among others (Environmental Systems Research Institute, 1995). Of these data, USGS DLGs are the most accessible (available free of charge via the internet). Other road data listed above are based to some degree on USGS DLGs. Commercial data are generally enhanced to include address geocoding information, or are improved through the use of data obtained from local sources. Commercial data may also have the advantage of being more recent than USGS DLGs (DeLorme, pers. comm.; Etak, pers. comm.; USGS, Reston, Virginia, pers. comm.). Nevertheless, no two road data sources represent independent samples that can be compared to each other. Therefore, assessing the data quality of roads data requires the use of other types of data.

A three-part, weight-of-evidence approach was used to evaluate the accuracy of road densities of DLG road data. The three phases are (from smallest to largest spatial scale) (1) mapping spatial distribution of road density and visually comparing distribution of road density with population; (2) correlation of road density and population density; and (3) matched-pair comparison of density calculated from DLGs and digital orthophotoquads (DOQs). The first two components of the approach are based on the fact that roads are built to improve access for human activity (Rice, 1989; Dale *et al.*, 1993). Therefore, there should be a strong positive relationship between road density and population density. The third phase derived an independent roads data set from high resolution imagery, allowing for a quantitative measurement of accuracy of DLG road density estimates.

Methods

DLGs are available at 1:24,000-, 1:100,000-, and 1:2,000,000-scales. The 1:100, 000-scale data are the largest scale data set available nationwide. These data are created by photomosaick-

T.G. Wade and D.F. Bradford are with the U.S. EPA National Exposure Research Laboratory, P.O. Box 93478, Las Vegas, NV 89193 (wade.timothy@epa.gov).

J.D. Wickham is with the U.S. EPA (MD-56) National Exposure Research Laboratory, Research Triangle Park, NC 27711.

Photogrammetric Engineering & Remote Sensing
Vol. 65, No. 12, December 1999, pp. 1419–1425.

0099-1112/99/6512–1419$3.00/0

ing and then photoreducing the 1:24,000-scale USGS topographic maps. The photoreductions are manually scribed and then scanned to create the digital 1:100,000-scale data (USGS, 1989; USGS, Reston, Virginia, pers. comm.). Loss of information from 1:24,000- to 1:100,000-scale is minimal and typically only occurs in urban areas (USGS, Reston, Virginia, pers. comm.). USGS 1:100,000-scale data for Oregon and Pennsylvania were chosen for evaluation.

In the first part of the analysis, road densities were calculated for 1:100, 000-scale USGS quads (one-half degree of latitude by one degree of longitude) by dividing the total road length in the quad by the total land area in the quad. Large water bodies (e.g., oceans and Lake Erie) were not included in the calculation of quad area. For quads extending beyond state borders, the entire road network and land area were used to calculate densities. The road densities were grouped into ten equal-frequency categories and ranked from one (highest density) to ten (lowest density). These ranks were mapped to determine whether quads with the highest road densities were also the most urbanized.

For the second phase, graphs of road density versus population density were generated and Pearson correlation coefficients were calculated. Because road development is closely tied to human activity (Rice, 1989; Dale *et al.*, 1993), there should be a strong linear correlation between road density and population density. Population density was calculated as persons per square kilometer, and road density was calculated as in part one above. Population estimates were taken from the Department of Commerce, Bureau of Census 1990 Decennial Census data (Bureau of the Census, 1990a). Comparisons of road and population density were made by county in Pennsylvania, and by census tract in Oregon. Census tracts were used in Oregon because the large size of many counties in the western United States is much less a reflection of population than in the eastern United States. Census tracts are subdivisions within counties having between 2,500 to 8,000 people (Bureau of the Census, 1990b). National Forest lands were excluded from calculation of road and population density, because these lands are not representative of conditions in the rest of the state (population is generally very low, while road density associated with logging activity may be high).

For the third component, roads were manually digitized on-screen using USGS DOQs as background. DOQs are digital images created from the National Aerial Photography Program (NAPP). The digitization process results in a 1-meter ground resolution when using 1:40,000-scale photography. Digital orthophotos comply with National Map Accuracy Standards, i.e., 90 percent of the points selected for testing must fall within 40 feet of correct ground location at 1:24,000 scale. The data are stored as 3.75-minute quadrangles (one-quarter of a 1:24,000-scale (7.5-minute) topographic map) (USGS, 1998). A total of 42 DOQs were available, 18 for Pennsylvania and 24 for Oregon. Nine DOQs, representative of urban, suburban, and rural areas, were chosen from each state for analysis. Road density was calculated from the screen digitized data, and compared to density values for DLGs calculated for the same area.

To maintain objectivity, screen digitization of DOQs was performed by persons without prior knowledge of DLG data. To most closely replicate the USGS methodology, digitizers were provided only the survey rules used by USGS in creating their maps (USGS, 1980; USGS, 1985). These survey rules were used to determine which features on the DOQs should be kept or rejected as roads (the rules mainly affected Class 4 (unimproved) roads). A synthesis of the rules for inclusion of Class 4 roads is provided in Table 1.

Results

The distributions of 1:100,000-scale quads as a function of road density in Pennsylvania and Oregon are shown in Figures 1a and 1b, respectively. The rank order map for Pennsylvania (Figure 1a) shows the highest road densities in the southeastern and northwestern corners of the state, in the vicinities of Philadelphia and Erie, respectively. The Erie quad had the highest road density because the small amount of land area was highly urbanized. The Pittsburgh quad would replace Erie in the highest rank order group if Erie was removed. The lowest rank orders occur in the rural portions of the state where one would expect lower road densities.

The spatial pattern of road densities in Oregon also shows the influence of urban centers (Figure 1b). The highest road densities are in the vicinity of the cities of Portland, Salem, Corvallis, and Boise. However, there are exceptions. High road densities are also found in the Blue Mountains in the northeastern part of the state (the quad south of La Grande) and near Bend. These areas have high road densities associated with the timber industry. The Newport and Astoria quads also show high densities because the small land area is highly urbanized, similar to the Erie quad in Pennsylvania. The lowest densities are located in the rural southeastern corner of the state.

For both Pennsylvania and Oregon, population density versus road density initially showed a non-linear relationship. In large urban centers (e.g., Philadelphia, Pittsburgh, Portland), population density increases much faster than road density. Removing counties and census tract that include large urban centers produced a linear relationship between population density and road density for both Pennsylvania (Figure 2a) and Oregon (Figure 2b). Pearson correlation coefficients were 0.951 and 0.903, respectively ($Prob > |R| = 0.0001$). Removal of

TABLE 1. SYNTHESIS OF USGS RULES FOR INCLUSION OF UNIMPROVED ROADS (USGS, 1980)

The surface construction of Class 4 (unimproved) roads is typically stabilized soil, sand-clay, or disintegrated rock with a poor or no foundation. Continual maintenance is required to keep them passable. Class 1, 2, and 3 roads have either improved, loose-surface or hard surface construction; they are always mapped.

(1) Whether or not an unimproved road is mapped depends on its local importance as a travel route.
- (a) Only unimproved roads leading to a mapped feature or providing access to a particular area are shown. When two such roads lead to the same feature in approximately the same direction, only the better of the two roads is shown.
- (b) In an area with a dense network of unimproved roads (e.g., logging areas), only the more permanent and direct roads are shown and the others are omitted.
- (c) Field and mine roads are included only when they lead to some mapped feature.
- (d) Logging roads are mapped only if they are useful permanent routes. These include connecting roads and roads leading to camps or sawmill sites. Feeder roads are not mapped.
- (e) Service roads along pipe- or powerlines are mapped only if they provide access to isolated areas.

(2) All driveways are considered Class 4 (unimproved) roads, regardless of construction.
- (a) Driveways are mapped if they are at least 500 feet long.
- (b) Access roads (roads into cemeteries, parks, or other special-purpose areas) are mapped if they are at least 500 feet long. Within cemeteries only the main inter-connecting roads are shown.

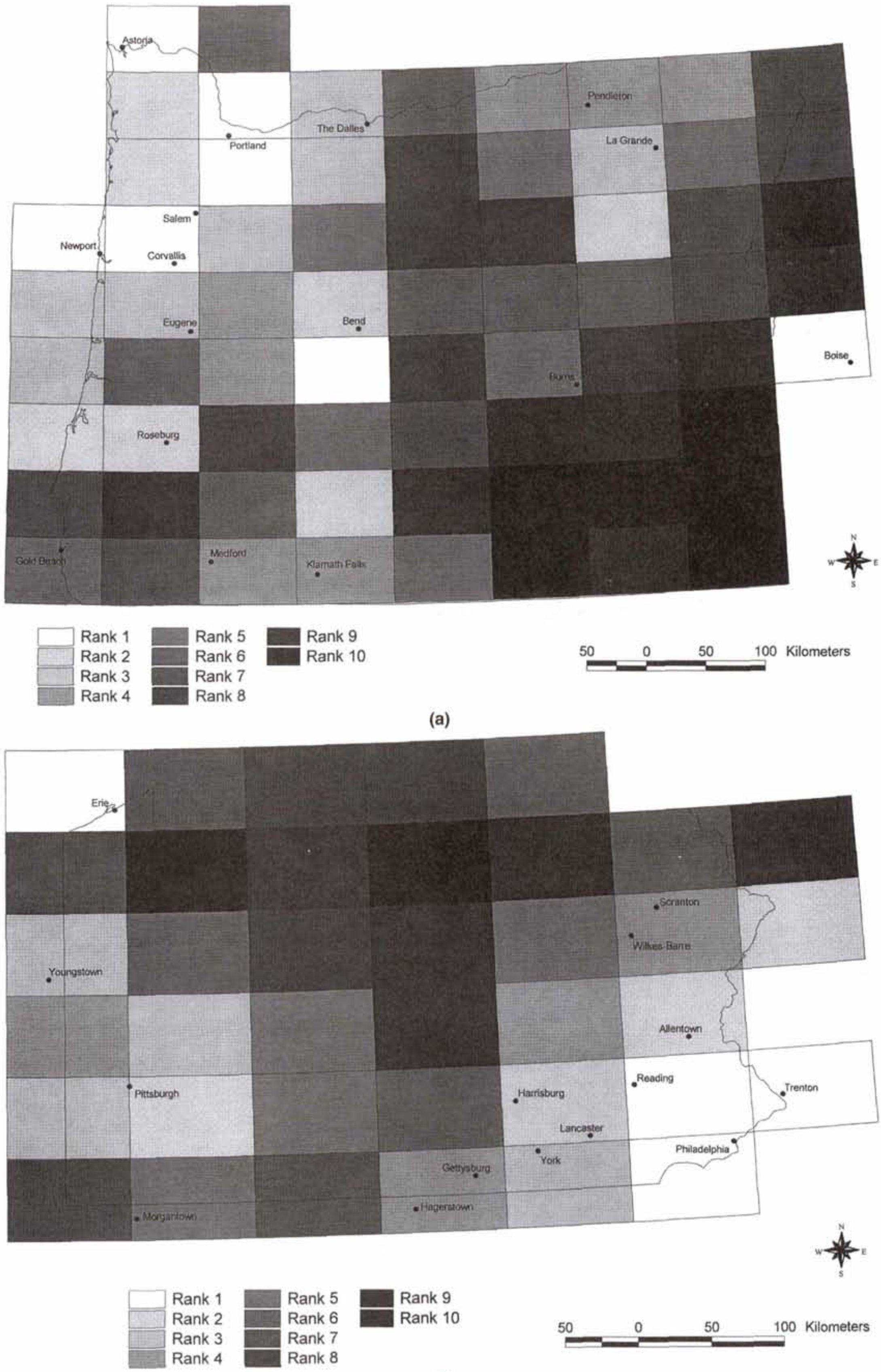
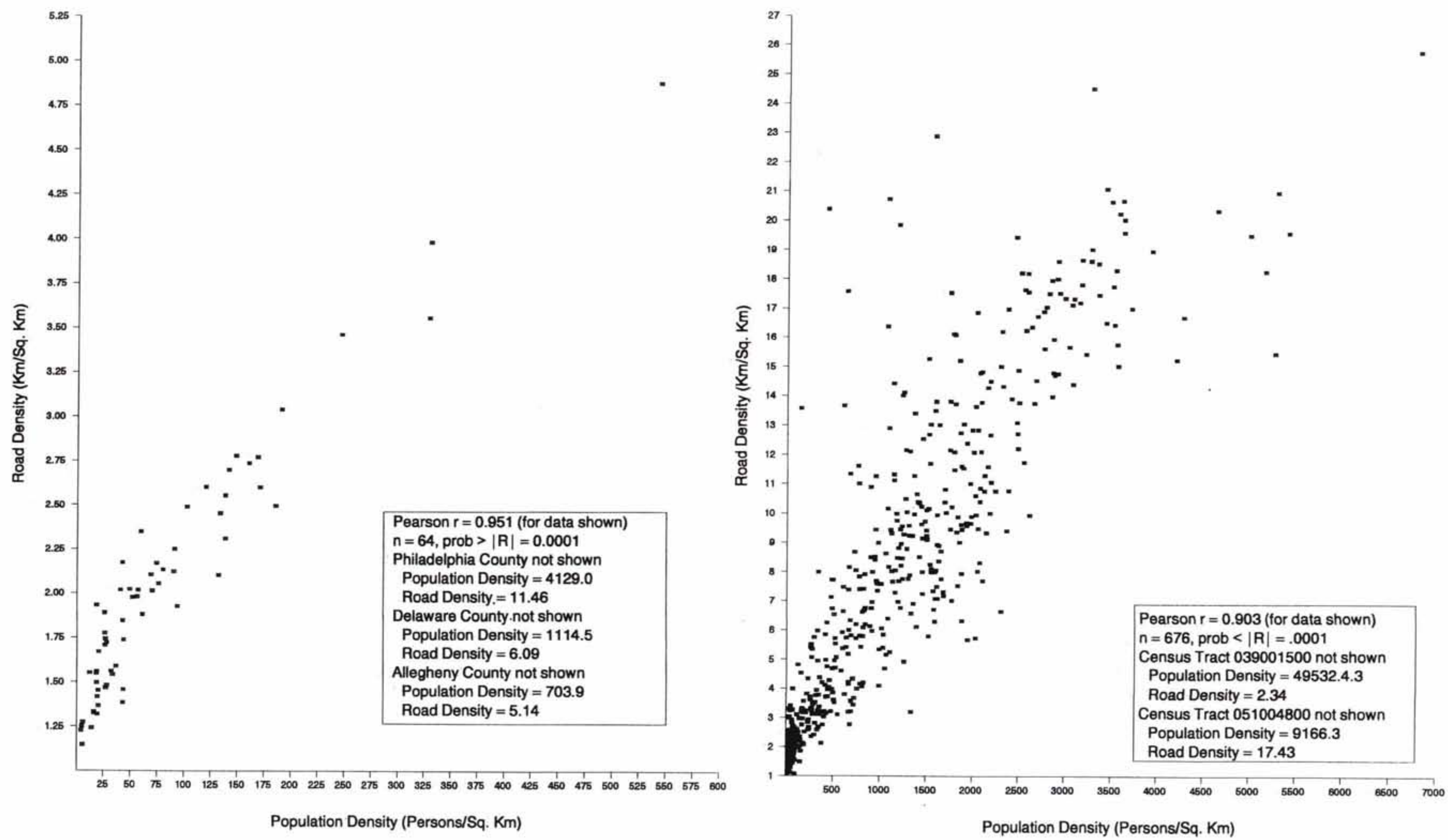


Figure 1. (a) Rank order of road densities by 1:100,000-scale quad in Pennsylvania. (b) Rank order of road densities by 1:100,000-scale quad in Oregon. The highest road densities are in white and the lowest are in black.



(a) (b)

Figure 2. (a) Road density versus 1990 population density by county in Pennsylvania. The counties shown in the box produced a power curve functional form and were not included in the correlation analysis. (b) Road density versus 1990 population density by census tract in Oregon. The census tracts shown in the box produced a power curve functional form and were not included in the correlation analysis.

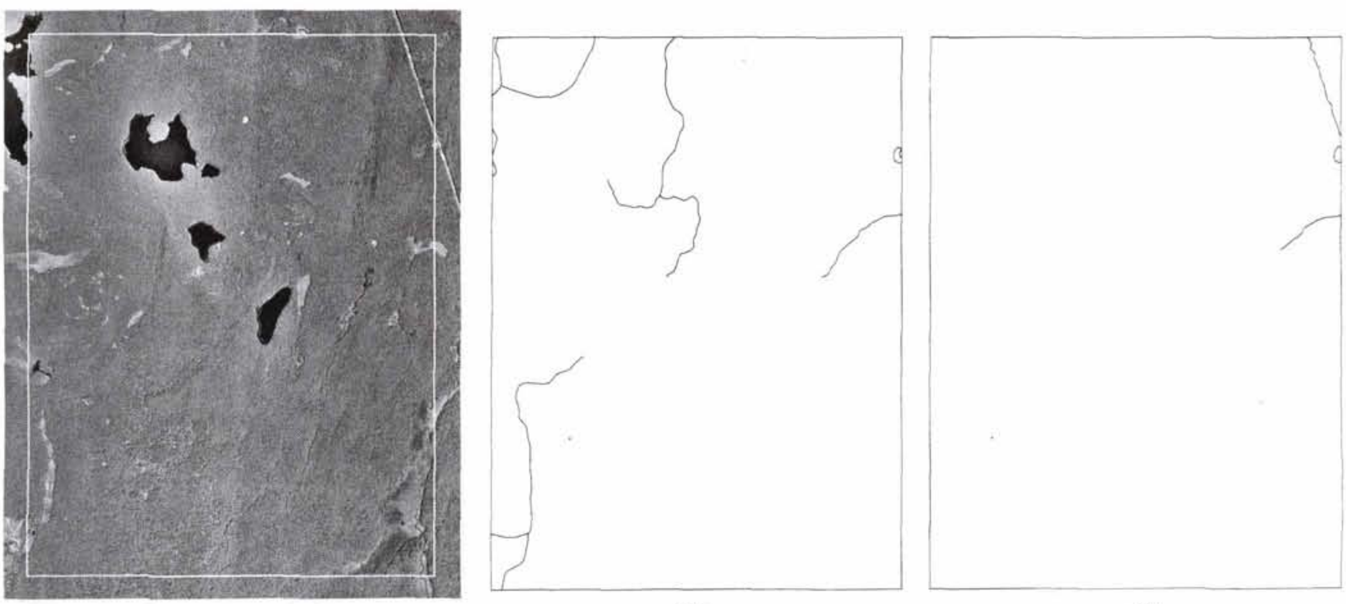
TABLE 2. MATCH-PAIR COMPARISON OF ROAD DENSITIES FROM DOQS AND DLGS.

Pennsylvania DOQ Quarter-Quadrangle	Density (km/km²)				|DOQ-DLG|
	DOQ	DLG	Difference	DOQ/DLG	DLG
Trenton West - SW	9.187	9.903	−0.716	0.928	0.072
Smithsburg - NE	2.486	2.878	−0.392	0.864	0.136
Promised Land - SE	0.072	0.458	−0.386	0.157	0.843
Emmitsburg - NW	2.279	2.472	−0.193	0.922	0.078
Blue Ridge Summit - NE	2.127	2.178	−0.050	0.977	0.023
Blue Ridge Summit -	2.919	2.958	−0.039	0.987	0.013
Honeybrook - NW	2.382	2.197	0.185	1.084	0.084
Taneytown - NW	1.929	1.743	0.186	1.107	0.107
Emmitsburg - NE	1.852	1.638	0.214	1.131	0.131
Oregon DOQ Quarter-Quadrangle	Density (km/km²)				|DOQ-DLG|
	DOQ	DLG	Difference	DOQ/DLG	DLG
Portland - SW	7.040	7.220	−0.180	0.975	0.025
Chapman - NE	1.960	2.000	−0.040	0.980	0.020
Dixie Mountain - SW	1.238	1.197	0.041	1.034	0.034
Hillsboro - SE	1.820	1.620	0.133	1.124	0.123
Dixie Mountain - SE	1.654	1.403	0.251	1.179	0.179
Dixie Mountain - NE	2.277	1.627	0.650	1.400	0.400
Rainier - SW	1.630	0.935	0.695	1.743	0.743
Hillsboro - SW	5.671	4.321	1.350	1.312	0.312
Camas - SW	7.590	5.640	1.940	1.346	0.346

urban counties and census tracts resulted in the loss of three out of 67 observations for Pennsylvania, and two out of 678 observations in Oregon.

Matched-pair comparison of roads digitized from DOQs with those from DLGs (third phase of the approach) generally showed good agreement for both Pennsylvania and Oregon (Table 2). A perfect match would result in a difference (DOQ-DLG) of zero, a ratio (DOQ/DLG) of one, and a fractional deviation from DLG (absolute value of the difference divided by DLG) of zero in the three right columns of Table 2. In Pennsylvania, the mean ratio for all areas was 0.906, but excluding Promised Land-SE, which had the poorest agreement between DOQ and DLG, the ratio becomes 1.00. The mean fractional deviation from DLG was 0.165 (0.081 after removing Promised Land-SE). For all areas in Oregon, the mean ratio was 1.233, omitting Rainier-SW (the area with the worst agreement) improves the ratio to 1.169. The mean fractional deviation was 0.243 (0.180 excluding Rainier-SW).

In Pennsylvania, there was a tendency to find fewer roads on DOQs than those mapped in DLGs. The Promised Land-SE



(a) (b) (c)

Figure 3. (a) DOQ of Promised Land, Pennsylvania (SE quadrant). The white rectangle delineates the area used for DOQ/DLG comparison. DLG roads (b) and roads digitized from DOQs (c) for Promised Land, Pennsylvania (SE quadrant). Note the decrease in roads due to closed canopy forest conditions on the DOQ.

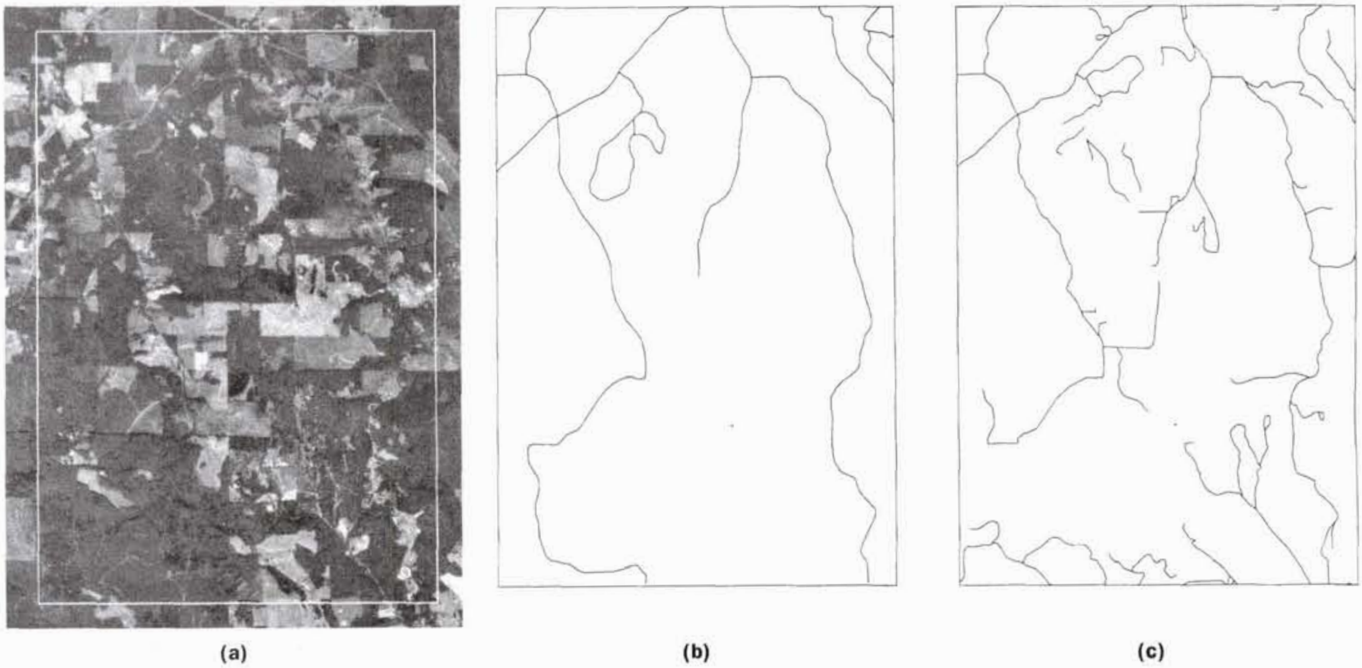


Figure 4. (a) DOQ of Rainier, Oregon (SW quadrant). The white rectangle delineates the area used for DOQ/DLG comparison. DLG roads (b) and roads digitized from DOQs (c) for Rainier, Oregon (SW quadrant). Note the increase in roads due to recent logging activity on the DOQ.

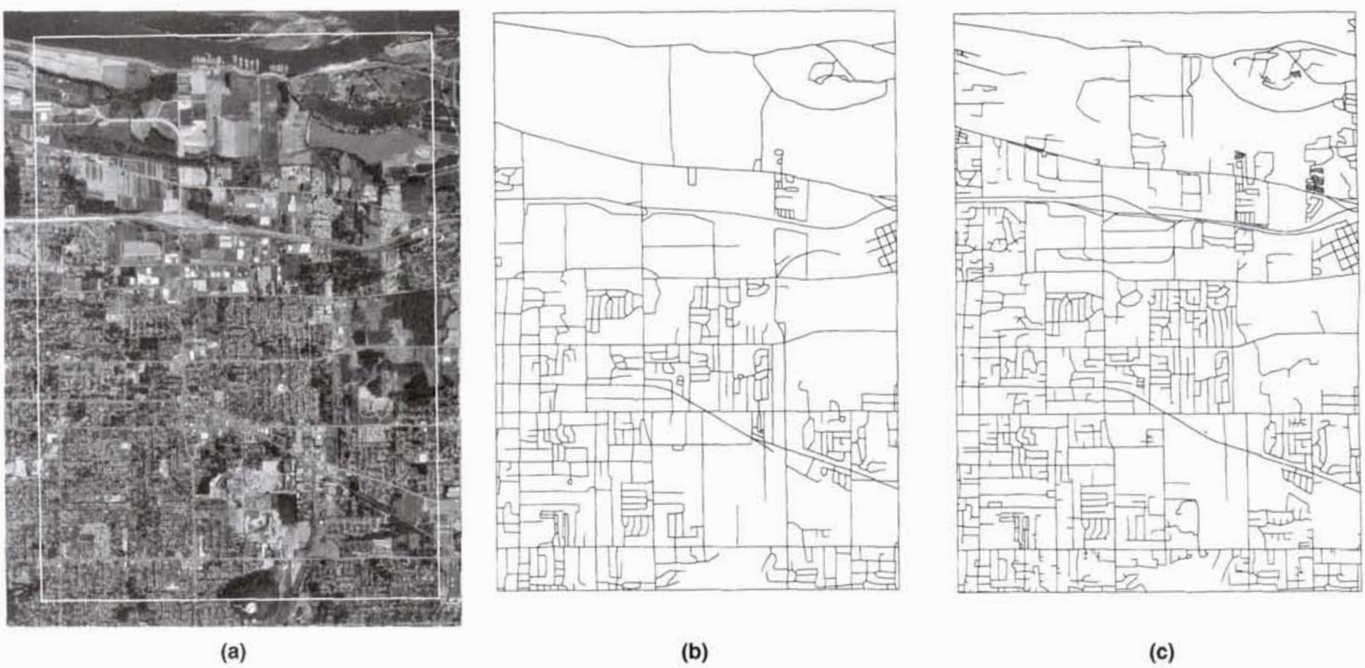


Figure 5. (a) DOQ of Camas, Oregon (SW quadrant). The white rectangle delineates the area used for DOQ/DLG comparison. DLG roads (b) and roads digitized from DOQs (c) for Camas, Oregon (SW quadrant). Note the increase in roads due to recent development on the DOQ.

(southeast quarter of the 7.5-minute Promised Land topographic map) DOQ is the most extreme example (Figures 3a, 3b, and 3c). The USGS road data were compiled from ground surveys and updated using aerial photography in the 1970s. Lack of ground survey data prevented the identification of some roads under closed-canopy forest conditions and mature residential landscaping in several of the Pennsylvania DOQs.

In Oregon, the opposite tendency was evident; there was a tendency toward higher density estimates from DOQs than from DLGs. Like Pennsylvania, many of the DOQs covered rural, forested areas. The higher density of roads in rural DOQs in Oregon relative to Pennsylvania is a reflection of timber harvesting. Rainier-SW (Figures 4a, 4b, and 4c) shows evidence of timber access roads that were built since the DLGs were updated. The large difference in road density estimates for Camas-SW (Figures 5a, 5b, and 5c) and Hillsboro-SW reflects housing development since the DLGs were updated.

The other urban areas covered by DOQs (Portland-SW (Oregon) and Trenton West-SW (Pennsylvania)) were apparently already well established at the time DLGs were updated in the 1970s. The densities of roads from DOQs for Portland-SW and Trenton West-SW actually show slightly lower densities than their DLG counterparts, indicating that some roads were overlooked during screen-digitizing.

Summary and Conclusion

A three-part, weight-of-evidence approach was used to assess the quality of USGS DLG road data. Part one showed that the spatial pattern of road density followed the spatial pattern of population at small (1:100,000) scales. Part two confirmed this pattern, showing that road density was strongly correlated with population density at larger (county or census tract) scales. Part three provided independent confirmation of parts one and two, showing that the density of roads calculated from DLGs generally showed good agreement with estimates derived from high resolution digital photography.

The primary shortcoming of USGS DLG road data appears to be age. However, with the exception of logging related activity, the majority of new roads are associated with urban and suburban areas, where habitat fragmentation is already high. In addition, there is now a coordinated effort to provide more current data (Forrest, 1994).

Road density has been shown to provide useful information on suitability of habitat for larger animals (Lyon, 1979; Lyon, 1983; Thiel, 1985).The results of this study indicate that USGS DLG road data provide accurate road density estimates. Commercially available road data would not likely provide significantly improved estimates of density.

Acknowledgments

We thank Kurt Riitters for his advice on comparing DLG and DOQ density estimates. The information in this paper has been funded in part by the United States Environmental Protection Agency, under Cooperative Agreement CR-819549-0105 to the Desert Research Institute. This paper has not been reviewed by the Environmental Protection Agency, and no official endorsement should be inferred. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

References

Bolstad, P.V., and T. Stowe, 1994. An evaluation of DEM accuracy: Elevation, slope, and aspect, *Photogrammetric Engineering & Remote Sensing*, 60(11):1327–1332.

Dale, V.H., R.V. O'Neill, M. Pedlowski, and F. Southworth, 1993. Causes and effects of land-use change in central Rondonia, Brazil, *Photogrammetric Engineering & Remote Sensing*, 59(6):997–1005.

Environmental Systems Research Institute (ESRI), 1996. *Arc Data Catalog, Fifth Edition*, Environmental Systems Research Institute, Redlands, CA California.

Fenstermaker, L.K., (editor), 1994. *Remote Sensing Thematic Mapper Accuracy Assessment: A Compendium*, American Society of Photogrammetry and Remote Sensing, (ISBN 1-57083-002-9), Bethesda, Maryland.

Forrest, D., 1994. Efforts under way to create a national transportation database, *GIS World*, 7(9):28.

Hess, G., 1994. Pattern and error in landscape ecology: A commentary, *Landscape Ecology*, 9(1):3–5.

Janssen, L.L.F., and F.J.M. van der Wel, 1994. Accuracy assessment of satellite derived land-cover data: A review, *Photogrammetric Engineering & Remote Sensing*, 60(4):419–426.

Jones, K.B., K.H. Riitters, J.D. Wickham, R.D. Tankersley, R.V. O'Neill, D.J. Chaloud, E.R. Smith, and A.C. Neale, 1997. *An Ecological Assessment of the United States Mid-Atlantic Region*, EPA/600/R-97/130, Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C.

Lunetta, R.S., R.G. Congalton, L.K. Fenstermaker, J.R. Jensen, K.C. McGwire, and L.R. Tinney, 1991. Remote sensing and geographic information system data integration: Error sources and research issues, *Photogrammetric Engineering & Remote Sensing*, 57(6):677–687.

Lyon, L.J., 1979. Habitat effectiveness for elk as influenced by roads and cover, *Journal of Forestry*, 77:658–660.

———, 1983. Road density models describing habitat effectiveness for elk, *Journal of Forestry*, 81:592–595.

Rice, R.E., 1989. *The uncounted costs of logging, Volume 5, National Forests, Policies for the Future*, The Wilderness Society, Washington, D.C.

Schueler, T., 1994. The importance of imperviousness, *Watershed Protection Techniques*, 1(3):100–111.

Storm, G.L., R.L. Andrews, R.L. Phillips, R.A. Bishop, D.B. Sniff, and J.R. Tester, 1967. Morphology, reproduction, dispersal, and mortality of midwestern red fox populations, *Wildlife Monographs*, 49:5–82.

Thiel, R.P., 1985. Relationship between road densities and wolf habitat suitability in Wisconsin, *American Midland Naturalist*, 113(2):404–407.

Bureau of the Census, 1990a. *Census of Population and Housing, 1990*: Summary tape file 1B extract on CD-ROM (Oregon) [STF1BXLR.DBF]/ prepared by the Bureau of the Census, Washington: The Bureau [Bureau of the Census, Data User Services Division], 1991.

———, 1990b. *Census of Population and Housing, 1990*: Summary Tape File 1B Extract on CD-ROM Technical Documentation/prepared by the Bureau of Census. Washington: The Bureau, 1991.

———, 1998. *USGS Standards for Digital Orthophotos*, 01 December 1998 (available at **http://mmcweb.cr.usgs.gov/public/nmpstds/doqstds.html**).

USGS, 1980. *Topographic Instructions*, U.S. Geological Survey Washington, D.C.

———, 1985. *Standards for 1:100,000-Scale Quadrangle Maps, Part 3: Feature Specifications and Compilation*, U.S. Geological Survey, Washington, D.C.

———, 1989. *Digital Line Graphs from 1:100,000-Scale Maps, National Mapping Program Technical Instructions, Data Users Guide 2*, U.S. Geological Survey, Washington, D.C.

Walsh, S.J., D.R. Lightfoot, and D.R. Butler, 1987. Recognition and assessment of error in geographic information systems, *Photogrammetric Engineering & Remote Sensing*, 53(10):1423–1430.

Wickham, J.D., K.B. Jones, K.H. Riitters, R.V. O'Neill, R.D. Tankersley, E.R. Smith, A.C. Neale, and D.J. Chaloud, in press. An integrated environmental assessment of the U.S. mid-Atlantic region, *Environmental Management*.

(Received 12 September 1997; revised and accepted 09 February 1999)