

Identifying grain-size dependent errors on global forest area estimates and carbon studies

Daolan Zheng,¹ Linda S. Heath,² and Mark J. Ducey¹

Received 19 August 2008; accepted 8 October 2008; published 14 November 2008.

[1] Satellite-derived coarse-resolution data are typically used for conducting global analyses. But the forest areas estimated from coarse-resolution maps (e.g., 1 km) inevitably differ from a corresponding fine-resolution map (such as a 30-m map) that would be closer to ground truth. A better understanding of changes in grain size on area estimation will improve our ability to quantify bias and uncertainty, and provide more accurate estimates of forest area and associated carbon stocks and fluxes. We simulated that global forest area estimated from a 1-km land-cover map (the most practical and finest resolution currently used for global applications) was 947,573 km² less than that of its corresponding 30-m map (excluding Antarctic and Greenland). This amount of forest could produce 0.57 ± 0.01 petagrams of carbon per year (PgC/yr) as NPP or 0.11 ± 0.002 PgC/yr as NEP equivalent to 4% of the 2.9 PgC/yr missing carbon sink in the 1990s. The most significant underestimation of forest area was in the temperate zone (the more uncertain region regarding the “missing carbon sink”) while a relatively small overestimation occurred in the tropic zone as the grain size of satellite-derived land-cover maps increases from 30 m to 1 km. **Citation:** Zheng, D., L. S. Heath, and M. J. Ducey (2008), Identifying grain-size dependent errors on global forest area estimates and carbon studies, *Geophys. Res. Lett.*, 35, L21403, doi:10.1029/2008GL035746.

1. Introduction

[2] Accurately identifying forest area is necessary for improving our understanding of the global carbon cycle [Townshend *et al.*, 1991]. Variation in grain size is one of many factors that can alter the outputs through a scaling process, thus, plays a substantial role in estimation and interpretation of landscape attributes and patterns [Woodcock and Strahler, 1987; Turner *et al.*, 2000]. Effects of land-cover maps derived from different sensor resolutions (e.g., 30 m and 1 km) on forest area estimates have been reported from regional to national scales [Liknes *et al.*, 2004; Zheng *et al.*, 2008] but there has been no such a quantification at the global scale.

2. Methods and Materials

[3] We present a ‘conceptual’ model for quantifying grain-size dependent errors on forest area estimates, and estimate

how much this may affect global forest carbon fluxes. The model was developed using the 1992 National Land Cover Dataset (NLCD) derived from Landsat 5 Thematic Mapper (TM, 30-m resolution) that covers the conterminous USA [Vogelmann *et al.*, 2001]. Although the scaling effect on area estimates likely increases as the difference between two grain sizes increases [Turner *et al.*, 1989; Zheng *et al.*, 2008], this study focuses on examining the difference in forest area estimates based on 30-m and 1-km resolutions because the latter is commonly used for global applications [Hansen *et al.*, 2000; Friedl *et al.*, 2002].

[4] The model was developed from the 3,111 county polygons with sizes ranging from over 100 km² to 52,200 km², with a mean size of 2,500 km², in the conterminous USA (Figure 1). All land cover classes in the NLCD map were aggregated to 2 general types: forest (classes 41, 42, and 43) and non-forest (all other classes). The 30-m NLCD map was aggregated to 1-km resolution following the majority scheme (<http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=BlockMajority>). The scheme, compared to other scaling schemes such as random or nearest-neighbor, is more often used in ecological studies for scaling up distinct variables derived from remote sensing [Stuckens *et al.*, 2000; Ahl *et al.*, 2005]. The profiles of forest cover percentages were well-represented in either 30-m (ranging from 0% to 99%) or 1-km (ranging from 0% to 100%) maps used for model development. The scaling effect, that is, the difference in forest-cover percentages between 30-m and 1-km resolution maps for each county, was calculated as $(\text{Forest}\%_{1\text{km}} - \text{Forest}\%_{30\text{m}})$. We used regression analysis to fit a cubic model estimating the grain-size effect as a function of the forest cover percentages at 1-km map for all counties. Counties with forest cover <0.5% or ≥99.5% in the 1-km resolution land-cover map were classified as zero percent or 100% forest cover, respectively; all other counties were binned to the nearest whole percentage point.

[5] We applied our conterminous U.S. derived conceptual model to identify grain-size effect on global forest area estimates using 1-km 1992 University of Maryland (UMD) [Hansen *et al.*, 2000] and 2001 MODIS land-cover maps to represent the average land-cover condition of the 1990s. Averaged forest cover percentage of the 1990s in the conterminous U.S. was about 28% and the corresponding percentage in the world was a similar 29%. Scale investigation requires that the scale of analysis be commensurate with the intrinsic scale of the phenomenon under study [Legendre *et al.*, 2002; Zheng *et al.*, 2008]. The mean county size of the conterminous U.S., (2,500 km²) where the model was developed is approximately that of the cell size of half by half degree across the globe. Thus, we simulated grain-size effect on area estimation at half by half degree resolution. We aggregated the cell-level calculations

¹Department of Natural Resources and the Environment, University of New Hampshire, Durham, New Hampshire, USA.

²Northern Research Station, USDA Forest Service, Durham, New Hampshire, USA.

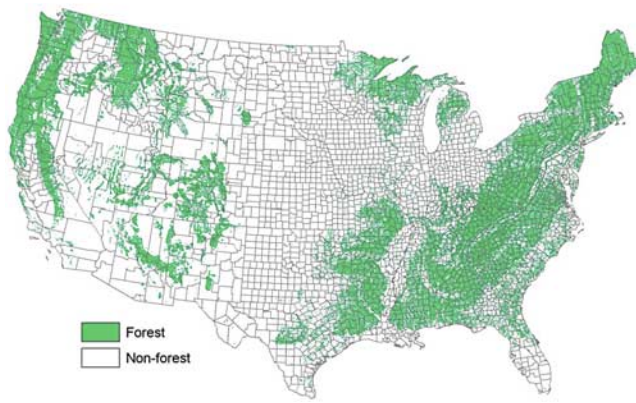


Figure 1. NLCD 1992 county-level observations at 30-m resolution for the conterminous USA from which the 1-km observations were obtained and the ‘conceptual’ model was developed to examine grain-size effect on forest area estimates.

to the country level, and then tallied for geographic zones and the globe.

[6] To reduce uncertainty in examining grain-size effects on global forest area estimation, we made two adjustments on the model application for the globe: First, we included bare ground to calculate forest cover percentages (as proportion of the total land area) but we excluded bare ground when we estimated the scaling effect. This adjustment was appropriate because bare ground in the conterminous U.S. accounted for only 1.6% of its total land area. In other words, including or excluding bare ground would have little impact on error identification. However, bare ground accounted for 18% of the total land area globally in large blocks, such as the Sahara Desert, where forest area would have been affected very little by the grain-size effect. Second, we calculated grain-size effect on forest areas for the 148 remaining countries worldwide that could be identified at half by half degree resolution and whose forest cover percentages fell into the 95% percentile ranging from 2.5% to 97.5%. The 148 countries were grouped into 3 geographic zones: 1) tropical zone – including the countries with at least 50% of their land mass between the tropics of

Cancer and Capricorn according to the list from the World Conservation Union (http://www.nhm.ac.uk/hosted_sites/bbstbg/tropctry.htm); 2) boreal zone – including Russia, Canada, the four Nordic countries, three Baltic countries of the former USSR, Belarus, Poland, UK, Ireland, and Iceland; and 3) temperate zone – all other remaining countries. Forest area changes caused by grain-size effect were summed by zone based on the calculations at country level.

[7] Prediction errors (PE) of area estimates caused by grain-size effect for different countries (Table 1), and for geographic zones and the globe (Table 2) were initially calculated assuming that prediction errors of half by half degree grains were independent,

$$PE_{zone} = \left(\sum_{j=1}^N \sum_{i=1}^K A_{ij}^2 SD_{ij}^2 \right)^{0.5} \quad (1)$$

where $j = 1 \dots N$ indexes the countries in a given zone ($N = 1$ for a single country), $i = 1 \dots K$ indexes the grain in the j th country, A_{ij} is the area of a half by half degree grain, and SD_{ij} is the standard deviation of the residuals from the grain-size effect prediction, based on the 1% forest cover bin calculated from the observed 1-km map (Figure 2). Exploratory analysis indicated positive autocorrelation among the counties in the U.S. used to develop the scaling relationship. Prediction errors were adjusted upward following the methods outlined by Griffith [2005]. We fit a spatial autoregressive model to the residuals ($\rho = 0.4193$), yielding a variance inflation factor of 2.6584 which was applied to all PEs (Table 2). To illustrate the importance of the error identification associated with changes in grain size on forest area estimation in terms of carbon flux estimation, we converted the forest areas to net primary production (NPP, PgC/yr) using the unit NPP rates in boreal, temperate, and tropic zones of the world (Table 2).

[8] We further converted the global NPP to net ecosystem production (NEP). Net ecosystem production was calculated as the residual of the NPP after deducting heterotrophic respiration (R_H) because by definition $NEP = NPP - R_H$ [Euskirchen et al., 2002; Hibbard et al., 2005]. Raich et al. [2002] demonstrated that mean annual global forest soil respiration (including woodlands) was about 46 PgC/yr over a 15-year period (1980–1994) while Hanson et al. [2000]

Table 1. Estimated Top 20 Countries Worldwide Affected Most in Terms of Forest Area Difference in the 1990s Caused by the Grain-Size Effect Between a Hypothetic Global 30-m Land-Cover Map and a Global 1-km Map Based on the Averaged Forest Cover Percentages of 1992 UMD and 2001 MODIS Maps at the Country Level Using the ‘Conceptual’ Model^a

Order	Countries Gaining Area ($\pm$ PE), Forest Cover %	Countries Losing Area ($\pm$ PE), Forest Cover %
1	Zaire (136 ± 2), 83	Australia (262 ± 4), 11
2	Indonesia (91 ± 2), 77	China (218 ± 4), 18
3	Brazil (65 ± 5), 59	United States (164 ± 5), 29
4	Central African Rep. (34 ± 1), 72	India (97 ± 3), 17
5	Papua New Guinea (34 ± 1), 89	Argentina (82 ± 3), 21
6	Bolivia (31 ± 2), 61	Russia (67 ± 5), 37
7	Burma (30 ± 1), 72	South Africa (55 ± 2), 11
8	Malaysia (20 ± 0.9), 82	Sudan (50 ± 1), 9
9	Sweden (20 ± 0.7), 73	Mongolia (47 ± 1), 6
10	Cameroon (18 ± 1), 70	Nigeria (43 ± 2), 21

^a $\pm$ PE, prediction error. ‘Conceptual’ model found in Figure 2. Both area and PE are in unit of 1,000 km².

Table 2. Area Differences Caused by Scaling (ADCS) from 30-m to 1-km Resolution for Each Individual Country Were Simulated at Half by Half Degree Resolution and Talled to Three Geographic Zones and the Globe^a

Zone	Forest Area	NPP (PgC/yr)	Unit NPP Rate (PgC/yr/km ²)	ADCS	NPP (Flux ± PE)
Boreal	13 700	2.9	2.12E-07	−98	−0.02 ± 0.001
Temperate	10 400	7.3	7.02E-07	−991	−0.70 ± 0.01
Tropical	17 550	17.8	10.14E-07	141	0.14 ± 0.01
Globe	41 650	28		−948	−0.57 ± 0.01

^aAnnual NPP flux and prediction errors (PE) caused by scaling were calculated using the corresponding NPP rates obtained from mean forest area and NPP (the 2nd and 3rd columns) values of two studies reported by *Intergovernmental Panel on Climate Change* [2001]. Both flux and PE are in unit of PgC/yr. Forest area and ADCS are in unit of 1000 km².

found an overall mean of 50% in world forest (varying from 46% at annual step to 56% at daily step) could be attributed to heterotrophic respiration. We used monthly heterotrophic rate of 50% because that matches the temporal scales used in both studies. As a result, world forest soil heterotrophic respiration (R_H) is about 82% of world forest NPP on average (23 PgC/yr divided by 28 PgC/yr in Table 2). We used this proportion to estimate the global scaling-affected NEP (0.11 PgC/yr) from NPP flux in Table 2.

3. Results and Discussion

[9] The conceptual model allows us to quickly examine the difference in forest area estimation caused by the scaling-up process from 30-m to 1-km resolution if its forest cover percentage at 1-km resolution is known. Our model suggested that the forest cover would be underestimated on average by up to 7.9% with a standard deviation (SD) of 4.7% when forest was 18% of the total land area based on the 1-km cover map, whereas it was overestimated by up to 8.6% with a SD of 2.4% when forest occupied 84% of the total land area (Figure 2). A general phenomenon revealed

by this study is that given 2 land-cover types, the one with non-prevailing cover percentages (<50%) at finer resolution will more likely lose cover percentages during the scaling-up process while the other type with coverages >50% will more likely gain cover percentages. SDs tend to get smaller when forest cover percentages at the 1-km map approach the median value (50%), and the variations within the range less than 50% (3.6% on average) are larger than those in the range greater than 50% (2.4% on average) (Figure 2). The degrees of forest cover change and variation caused by the scaling-up process are primarily controlled by actual forest and non-forest proportions (as demonstrated in this study), as well as forest and non-forest patch mosaics and configurations across landscapes. Our scaling-effect model developed from a larger sample size allows us to quantify the percent difference for each of the whole percent bins (Figure 2). The variation within each bin is a reflection of how landscape configurations affect forest area estimates during the scaling-up process.

[10] Nine of the top ten countries that gained apparent forest area by scaling from 30m to 1-km data were in the tropics (Table 1). The average forest cover of the 10 gaining

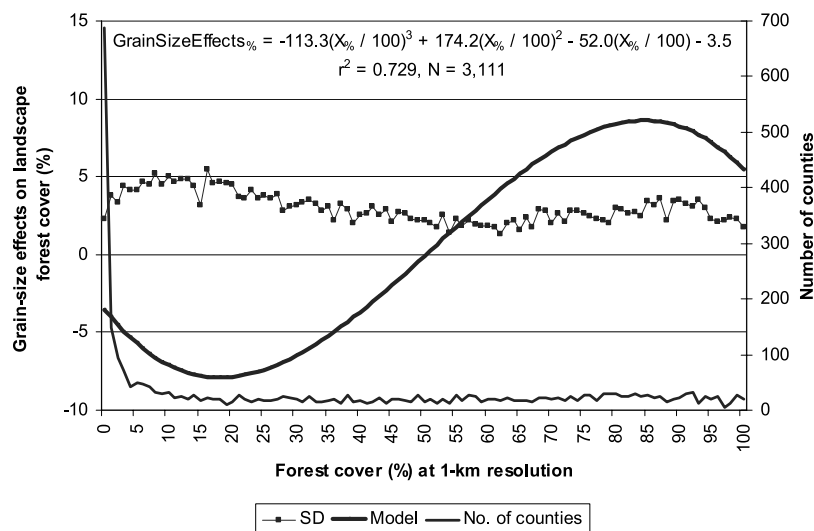


Figure 2. This ‘conceptual’ model was produced using 3,111 county polygon observations obtained from the NLCD 30-m and 1-km land-cover maps for the conterminous USA, which allows us to calibrate forest area estimation from coarse-resolution (1 km) map. Standard deviation (SD) averaged across bins was 3.0% ranging from 1.3 to 5.4%, while mean number of counties in a percent bin was 31, ranging from 5 to 688. Standard error of the model estimated grain-size effect for all the observations was 3.4%.

countries was 74%, ranging from 59% in Brazil to 89% in Papua New Guinea. Conversely, seven of the top ten countries that lost apparent forest area when scaling from 30 m to 1-km data were from temperate or boreal zones. The average forest cover of the 10 losing countries was 18%, ranging from 6% in Mongolia to 37% in Russia (Table 1). The scaling process tends to underestimate forest area in temperate and boreal latitudes where the forest cover percentages in major countries (in size) are <50%. At the same time, it tends to overestimate forest areas in the tropics, where large countries (in size) have forest cover percentages > 50%. The differences in forest area estimates are large enough to affect evaluations of regional carbon balance. A 1-km land-cover map could overestimate forest area by 141,352 km² in the tropical zone, on the one hand, in comparison to what would be estimated using 30-m data. This translated to an overestimate of forest NPP flux by 0.14 PgC/yr (Table 2). The results agree well with the previous findings that estimates of deforested areas in Brazilian Amazon were generally larger based on ground surveys than estimates from satellite data [Andersen *et al.*, 2002; Ometto *et al.*, 2005]. On the other hand, forest area in boreal and temperate zones was underestimated by 1,088,925 km², leading to an underestimate of forest NPP flux in middle-high latitudes by 0.72 PgC/yr (Table 2). As a result, the averaged global 1-km resolution land-cover map of the 1990s likely underestimated forest area by 947,573 km², compared to what it would be observed for a comparable hypothetical 30-m land-cover map as estimated by our model. Although there is no probability-based assessment that includes uncertainty bounds for global forest area estimation, Mayaux and Lambin [1997] indicated such a calibration could improve the accuracy of 1-km based forest area estimation by up to 35% in tropical zone, which is non-trivial progress.

[11] Our identified global forest area caused by changes in grain size from 30 m to 1 km could produce changes in estimates of 0.57 ± 0.01 PgC/yr carbon NPP flux (Table 2) or 0.11 ± 0.002 PgC/yr NEP. This amount of NEP is equivalent to approximate 4% of the 2.9 PgC/yr “missing” carbon based on 1990s’ estimates [Plattner *et al.*, 2002; Houghton, 2003]. Our results reinforce the statements that temperate/boreal forests are considered to be net carbon sinks at present [Heath and Birdsey, 1993; Nabuurs *et al.*, 1997; Liski *et al.*, 2003] while tropical forests function as a net carbon source in global carbon cycle [Grace *et al.*, 1995; Phillips *et al.*, 1998; Houghton *et al.*, 2000; Ometto *et al.*, 2005].

4. Conclusions

[12] The sink power of global forest ecosystems may be underestimated if coarse-resolution land-cover data are used without calibration. Current global forest cover is between 30%–40% (less than 50%) so the global forest area observed at 1-km resolution as a whole tends to be less than actual forest area observed at fine resolution. Spatially, increase in grain size from 30 m to 1 km results in a significant underestimation of forest area in temperate and boreal zones (the more uncertain regions regarding the “missing carbon sink”), and relatively small overestimation

of forest area in the tropics. Our results provide an insightful perspective to improve our understanding of reducing uncertainties in regional and global forest area estimation and associated carbon studies. Our calibrated country-level results can also provide useful information to improve estimation of carbon credits for international trading (http://en.wikipedia.org/wiki/Carbon_credit, 2008). Although the magnitude, location, and causes of error identifications associated with global forest area estimation and carbon fluxes are complex and can vary greatly by regions and years due to consistently changing in landscape configurations and patch size distributions, this study clearly suggests that incorporating grain-size effect can significantly reduce uncertainties in forest area estimation using satellite derived land-cover maps at 1-km resolution.

[13] **Acknowledgments.** This study is supported by the USDA Forest Service, Northern Research Station through grant 05-DG-11242343-074. We thank M. Friedl and D. J. Sulla-Menashe at Boston University for providing us an assembled global MODIS 2001 land-cover map.

References

- Ahl, D. E., *et al.* (2005), The effects of aggregated land cover data on estimating NPP in northern Wisconsin, *Remote Sens. Environ.*, 97, 1–14, doi:10.1016/j.rse.2005.02.016.
- Andersen, L. E., C. W. J. Granger, E. J. Reis, D. Weinhold, and S. Wunder (2002), *The Dynamics of Deforestation and Economic Growth in the Brazilian Amazon*, Cambridge Univ. Press, Cambridge, U. K.
- Euskirchen, E. S., J. Chen, H. Li, E. J. Gustafson, and T. R. Crow (2002), Modeling landscape net ecosystem productivity (LandNEP) under alternative management regimes, *Ecol. Modell.*, 154, 75–91, doi:10.1016/S0304-3800(02)00052-2.
- Friedl, M. A., *et al.* (2002), Global land cover from MODIS: Algorithms and early results, *Remote Sens. Environ.*, 83, 135–148, doi:10.1016/S0304-4257(02)00078-0.
- Grace, J., J. Lloyd, J. McIntyre, A. C. Miranda, P. Meir, H. S. Miranda, C. Nobre, J. Moncrieff, J. Masseder, and Y. Malhi (1995), Carbon dioxide uptake by an undisturbed tropical rain forest in southwest Amazonia, 1992 to 1993, *Science*, 270, 778–780, doi:10.1126/science.270.5237.778.
- Griffith, D. A. (2005), Effective geographic sample size in the presence of spatial autocorrelation, *Ann. Assoc. Am. Geogr.*, 95, 740–760, doi:10.1111/j.1467-8306.2005.00484.x.
- Hansen, M. C., R. S. DeFries, J. R. Townshend, and R. Sohlberg (2000), Global land cover classification at 1 km spatial resolution using a classification tree approach, *Int. J. Remote Sens.*, 21, 1331–1364, doi:10.1080/014311600210209.
- Hanson, P. J., N. T. Edwards, C. T. Garten, and J. A. Andrews (2000), Separating root and soil microbial contributions to soil respiration: A review of methods and observations, *Biogeochemistry*, 48, 115–146, doi:10.1023/A:1006244819642.
- Heath, L. S., and R. A. Birdsey (1993), Carbon trends of productive temperate forests of the conterminous United States, *Water Air Soil Pollut.*, 70, 279–293, doi:10.1007/BF01105002.
- Hibbard, K. A., B. E. Law, M. Reichstein, and J. Sulzman (2005), An analysis of soil respiration across Northern Hemisphere temperate ecosystems, *Biogeochemistry*, 73, 29–70, doi:10.1007/s10533-004-2946-0.
- Houghton, R. A. (2003), Revised estimates of the annual net flux of carbon to the atmosphere from changes in land use and land management 1850–2000, *Tellus, Ser. B*, 55, 378–390.
- Houghton, R. A., D. L. Skole, C. A. Nobre, J. L. Hackler, K. T. Lawrence, and W. H. Chomentowski (2000), Annual fluxes of carbon from deforestation and regrowth in the Brazilian Amazon, *Nature*, 403, 301–304.
- Intergovernmental Panel on Climate Change (2001), *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Report on Climate Change*, edited by J. T. Houghton *et al.*, Cambridge Univ. Press, Cambridge, U. K.
- Legendre, P., M. R. T. Dale, M.-J. Fortin, J. Gurevitch, M. Hohn, and D. Myers (2002), The consequences of spatial structure for the design and analysis of ecological field surveys, *Ecography*, 25, 601–615, doi:10.1034/j.1600-0587.2002.250508.x.
- Liknes, G. C., M. D. Nelson, and R. E. McRoberts (2004), Evaluating classified MODIS satellite imagery as a stratification tool, paper presented at the Joint Meeting of The Fifteenth Annual Conference of The

- International Environmetrics Society and The Sixth Annual Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Portland, Maine.
- Liski, J., A. V. Korotkov, C. F. L. Prins, T. Karjalainen, D. G. Victor, and P. E. Kauppi (2003), Increased carbon sink in temperate and boreal forests, *Clim. Change*, *61*, 89–99, doi:10.1023/A:1026365005696.
- Mayaux, P., and E. F. Lambin (1997), Tropical forest area measured from global land-cover classifications: Inverse calibration models based on spatial textures, *Remote Sens. Environ.*, *59*, 29–43, doi:10.1016/S0034-4257(96)00077-6.
- Nabuurs, G. J., R. Päivinen, R. Sikkema, and G. M. Mohren (1997), The role of European forests in the global carbon cycle: A review, *Biomass Bioenergy*, *13*, 345–458, doi:10.1016/S0961-9534(97)00036-6.
- Ometto, J. P. H. B., A. D. Nobre, H. R. Rocha, P. Artaxo, and L. A. Martinelli (2005), Amazonia and the modern carbon cycle: Lessons learned, *Oecologia*, *143*, 483–500, doi:10.1007/s00442-005-0034-3.
- Phillips, O. L., et al. (1998), Changes in the carbon balance of tropical forests: Evidence from long-term plots, *Science*, *282*, 439–442, doi:10.1126/science.282.5388.439.
- Plattner, G., F. Joos, and T. F. Stocker (2002), Revision of the global carbon budget due to changing air-sea oxygen fluxes, *Global Biogeochem. Cycles*, *16*(4), 1096, doi:10.1029/2001GB001746.
- Raich, J. W., C. S. Potter, and D. Bhagawati (2002), Interannual variability in global soil respiration, 1980–94, *Global Change Biol.*, *8*, 800–812, doi:10.1046/j.1365-2486.2002.00511.x.
- Stuckens, J., P. R. Coppin, and M. E. Bauer (2000), Integrating contextual information with per-pixel classification for improved land cover classification, *Remote Sens. Environ.*, *71*, 282–296, doi:10.1016/S0034-4257(99)00083-8.
- Townshend, J., C. O. Justice, W. Li, C. Gurney, and J. McManus (1991), Global landcover classification by remote sensing: Present capabilities and future possibilities, *Remote Sens. Environ.*, *35*, 243–255, doi:10.1016/0034-4257(91)90016-Y.
- Turner, D. P., W. B. Cohen, and R. E. Kennedy (2000), Alternative spatial resolutions and estimates of carbon flux over a managed forest landscape in western Oregon, *Landscape Ecol.*, *15*, 441–452, doi:10.1023/A:1008116300063.
- Turner, M. G., R. V. O'Neill, R. H. Gardner, and B. T. Milne (1989), Effects of changing spatial scale on the analysis of landscape pattern, *Landscape Ecol.*, *3*, 153–162, doi:10.1007/BF00131534.
- Vogelmann, J. E., S. M. Howard, L. Yang, C. R. Larson, B. K. Wylie, and N. Van Driel (2001), Completion of the 1990s National Land Cover Dataset for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources, *Photogramm. Eng. Remote Sens.*, *67*, 650–652.
- Woodcock, C. E., and A. H. Strahler (1987), The factor of scale in remote sensing, *Remote Sens. Environ.*, *21*, 311–332, doi:10.1016/0034-4257(87)90015-0.
- Zheng, D., L. S. Heath, M. J. Ducey, and J. E. Smith (2008), Quantifying scaling effects on satellite derived forest area estimates for the conterminous U.S., *Int. J. Remote Sens.*, in press.

M. J. Ducey and D. Zheng, Department of Natural Resources and the Environment, 215 James Hall, University of New Hampshire, Durham, NH 03824, USA. (daolan.zheng@unh.edu)

L. S. Heath, Northern Research Station, USDA Forest Service, 271 Mast Road, Durham, NH 03824, USA.