Counting the trees in – and outside of – the forest: A best estimate of forest cover in the Delaware River Basin

Forests play an important role in the health of watersheds by regulating hydrology (through canopy interception, infiltration, transpiration), soil quality, temperature, and pollutant uptake. Clearing or degradation of forests can be directly linked to water quality changes. But tracking loss of forest is not necessarily straightforward, and there is no single, authoritative data source for forest cover.

Typically, forest area is estimated based on some type of **land use** or **land cover** data, which is generally derived from processing of satellite or other aerial imagery. These datasets represent continuous coverages of the land (and water surface) and are broken down into pixels, most commonly at 30-meter resolution. *Land use* datasets assign a value to each pixel representing the land use type (such as forest, developed areas, or agriculture).[1] *Land cover* datasets instead assign values that describe literally what is covering the land surface such as tree canopy, low vegetation, pavement, buildings, etc. For example, certain land cover datasets assign a 'percent forested' value to each pixel to represent how much of each pixel an algorithm determines is under tree canopy. Both types of data have their drawbacks when estimating forest cover.

- **Land use** datasets assign a single value to each pixel such that 100% of the area within each pixel is either classified as forest or "not forest" (essentially any other land use type). Due to the pixel resolution, and this binary choice, many forested pixels may not in fact be fully covered by tree canopy, and many non-forest pixels can contain some tree canopy.

- Percent forested **land cover** datasets classify the percentage of total area that is covered by tree canopy in each pixel on a scale from 1 to 100, but it is difficult to tell which pixels in fact represent forests and which are picking up tree cover in other land uses.

Accordingly, we refer to forest pixels in land use datasets as forested land use, or area classified as forest. When we refer to forest cover, tree cover, or tree canopy coverage, we are talking about actual land cover (any area, regardless of land use that is under natural tree cover).

A complete estimate of forest cover in the Delaware River Basin (DRB) would ideally include both land use and land cover. That is, it would identify the amount (i.e. total area) of tree cover found in areas classified as forest land use, as well as tree cover in areas classified as other land use types.

In this analysis, we complete three tasks to determine how to best estimate forest cover in the DRB:

A. Case-study comparison of forest cover area by land use or land cover dataset
B. Comparing forest change over time by dataset
C. Change analysis - New land cover breakdown in 'lost' forest areas

Datasets

One key to this analysis is the availability of high-resolution land cover (HRLC) data being made available for the Delaware River Basin. This HRLC data is created by processing very detailed laser elevation (LIDAR) and imagery data and is available at 1-meter resolution for Pennsylvania portions of

[1] Note that 'natural' land use types such as wetlands, forests, etc., are also land cover types, so sometimes land use datasets are also referred to as land use/land cover (LULC) datasets.

December, 2016
Contact: Lars S. Hanson
hansonl@cna.org
3003 Washington Boulevard, Arlington VA 22201 703-824-2000 www.cna.org

the DRB representing 2013 conditions. The dataset classifies land cover into types including tree canopy, shrub/scrub, low vegetation, emergent wetlands, roads, other impervious area, and buildings, among others. This dataset, made available by the University of Vermont Spatial Analysis Laboratory and Shippensburg University, should represent the most detailed picture of land cover, and specifically tree cover, available. In this analysis, we compare this HRLC data to other available land use/land cover (LULC) datasets shown in Table 1.

Table 1. Land use and land cover datasets used in this analysis.

Type	Abbrev.	Data Years	Resolution (m)	Description	Source (link to general download site or most recent data year)
High-Resolution	HRLC	2013	1	High-Resolution Land Cover dataset for Pennsylvania (PA) (Chesapeake and Delaware River Basin)	http://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=3193
Land Use	NLCD	2001, 2006, 2011	30	National Land Cover Dataset - Land Cover	http://www.mrlc.gov/nlcd11_data.php
Land Use	CCAP	2001, 2006, 2010	30	NOAA Coastal Change Analysis Program Regional Land Cover Data	https://coast.noaa.gov/dataregistry/search/dataset/info/ccapregional
Land Use	CDL	2002, 2010-2015	30	United States Department of Agriculture _ Cropland Data Layer	https://nassgeodata.gmu.edu/CropScape/
Percent Forest	GFC	2000-2014	30	Global Forest Change dataset by (Hansen et al., 2013)	https://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.2.html
Percent Forest	NLCD_PF	2011	30	National Land Cover Dataset - 2011 Tree Canopy Cartographic	http://www.mrlc.gov/nlcd11_data.php
Percent Forest	HRLC_PF	2013	30	HRLC 1-m forest canopy data aggregated to 30m cell size, expressed as a percent.	See HRLC above. (The original data was reprocessed by CNA to develop this data coverage.)

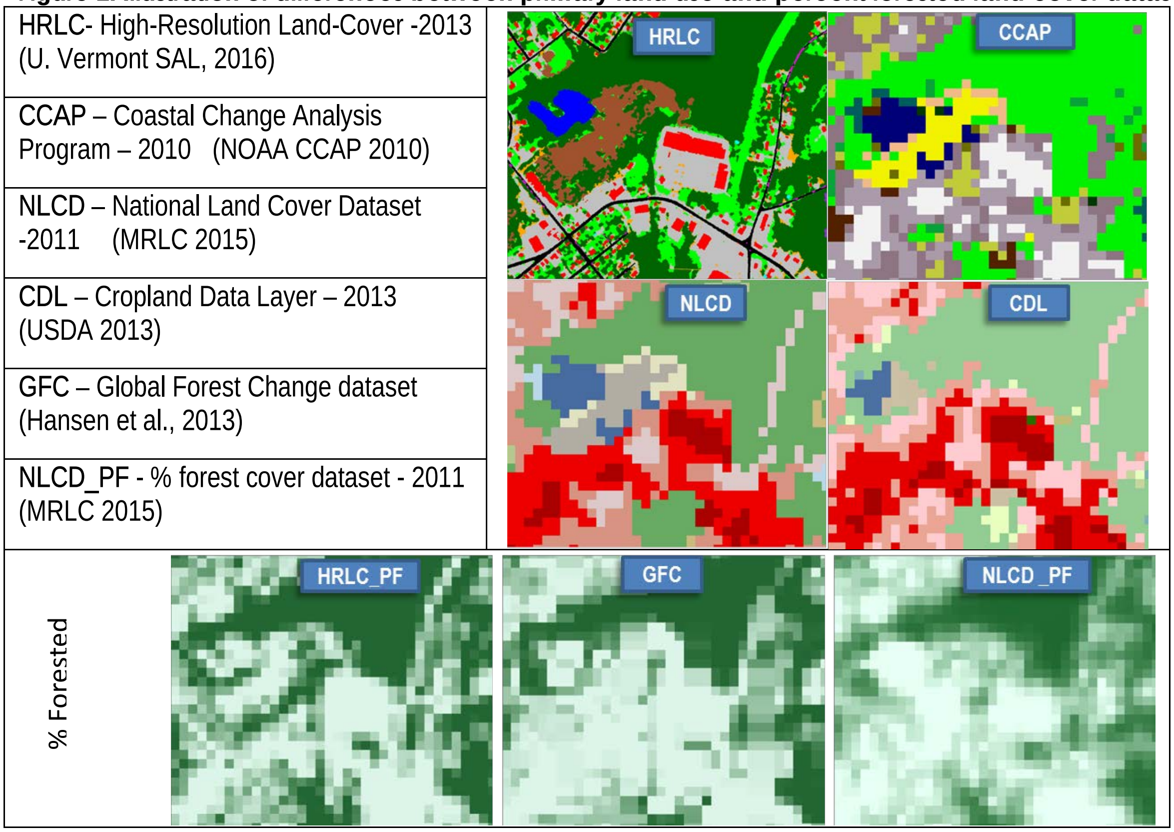
A. Case study – Forest Cover Area by Land Use/ Land Cover dataset

Since the HRLC data are not yet available over the entire Delaware River Basin, the HRLC data can be compared to other land use/land cover datasets over the domain of the available data: the Pennsylvania portion of the DRB. This area is established as the case study area.

The datasets are compared to the HRLC data on the basis of total area of forest cover, and percent forested cover in each of land use or land cover classifications. Figure 1 illustrates the differences between the HRLC and other datasets (Land use datasets shown on top, percent forested in the bottom row). The HRLC data at 1-meter resolution offers a far clearer picture of the state of forest cover in the Delaware River Basin. In order to compare this data with existing land use datasets, it was most efficient to aggregate the forest canopy coverage to 30-meter cells aligned with the existing datasets.[2] Figure 1 shows how the aggregated HRLC percent forested data (bottom left) compares to the other percent forested land cover datasets available.

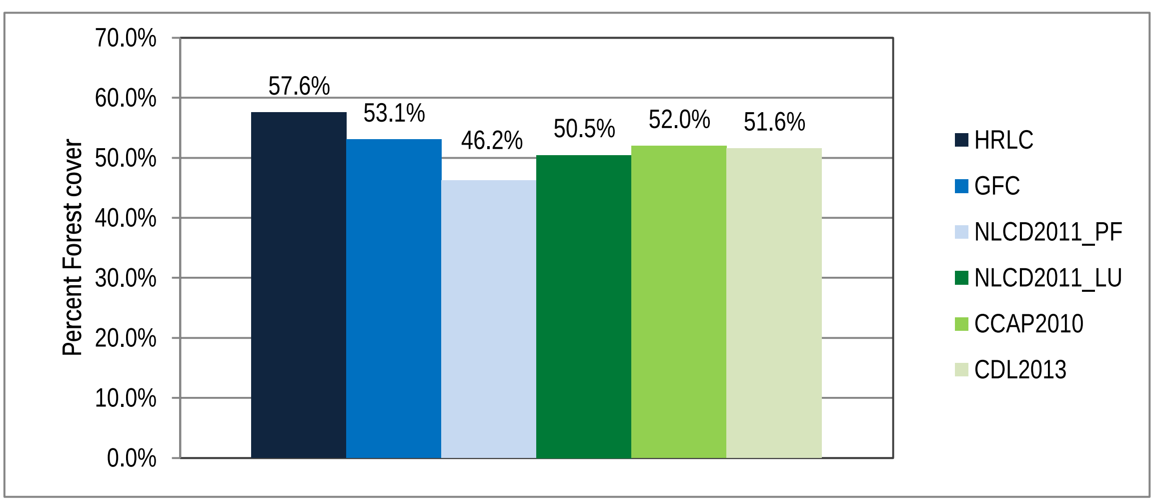
[2] To compute the percent forest coverage in each cell, we totaled the number of 1-m pixels denoted as "tree canopy" and divided by 900 (the number of 1 m pixels in a 30x30 meter pixel). This does **not** include classifications in the HRLC that are shown as tree canopy over roads, tree canopy over buildings or tree canopy over other impervious.

Figure 1. Illustration of differences between primary land use and percent forested land cover datasets.



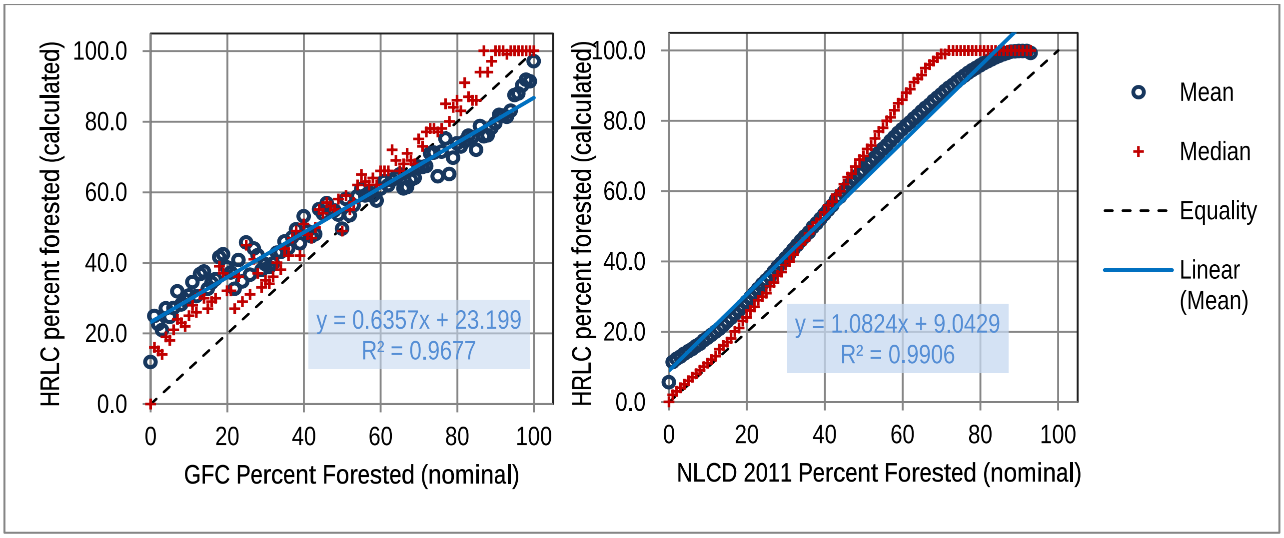
We computed the estimated total forested area within the study area for each dataset. Figure 2 shows the forest cover as a percentage of the study area (PA portion of the DRB). There is a large spread among the datasets, equivalent to over 11% of the total study area! Clearly, the HRLC data shows the highest amount of forest cover as a percentage of the area.

Figure 2. Percent of area as forested cover by dataset within case study area (PA portion of DRB).



Further investigation comparing the HRLC to individual datasets on a pixel by pixel basis can illuminate why there is such a difference. First, the percent forested land cover datasets (GFC[3] and NLCD_PF) are compared to the HRLC_PF dataset to determine if the nominal and actual percent forested values are aligned. Figure 3 shows the nominal percentage forested versus the calculated percentage forested (within each nominal 1% gradation) from the HRLC data. The GFC and HRLC data are correlated, but the correlation is noisy, and the slope of the line is far below unity. In general, the GFC overpredicts when percent forest cover is low and underpredicts when forest coverage is high. By contrast, the pattern of agreement between the HRLC and NLCD_PF data is well-ordered, but biased. The NLCD_PF nominal percent forest values are uniformly lower than the calculated values from the HRLC_PF data, though with a similar overall trend. It is worth noting that the NLCD_PF dataset has a maximum percent forested value of 93 percent in the DRB. Compared to the GFC, though, the NLCD_PF data are more closely correlated with the HRLC data, and its regression equation is better suited for estimating actual forest cover.

Figure 3. Comparison of nominal percent forested and calculated percent forested based on HRLC data.



The HRLC_PF data can also be used to evaluate the forest cover by land use classification within the land use datasets. For simplicity, we aggregated the (numerous) land use classifications of each dataset to four major land cover types including forest, other natural land cover, developed area, and agriculture. Table 2 shows the percentage that is tree canopy based on the HRLC_PF data. In general, areas classified as having forest land use are roughly 90% tree canopy on average. Other natural areas (grassland, shrub, barren, emergent wetland) are generally around 30-40% tree canopy. Developed areas are between 25 and 35% tree canopy, with the percentage declining as development intensity increases. Finally, agricultural areas are roughly 15% tree canopy (higher for pasture areas, lower for cultivated crops).

From the same data, it also possible to compute the land use type breakdown of total forest cover area (rightmost three columns of Table 2). It seems in the PA portion of the DRB that about 80% of tree cover is located in lands designated as forest, while about 8-13% of tree cover is located in land classified as a developed (especially Developed – Open Space). Very little of the total forested cover is found in other natural areas, while about 5 percent of tree cover can be found in agricultural areas.

[3] Note that the GFC dataset was adjusted to reflect 2013 conditions. The GFC data downloads include both a percent forested coverage reflecting conditions in 2000, and a separate file with forest loss areas for each year through 2014. We set all pixels representing forest loss areas from 2001-2013 to zero in the percent forest raster.

These proportions are averages for the case study area; in heavily developed or heavily agricultural areas, a much higher percentage of the total tree cover may be found in developed or agricultural areas.

Table 2. Percent tree canopy and percent of total forest cover area in major land use classifications.

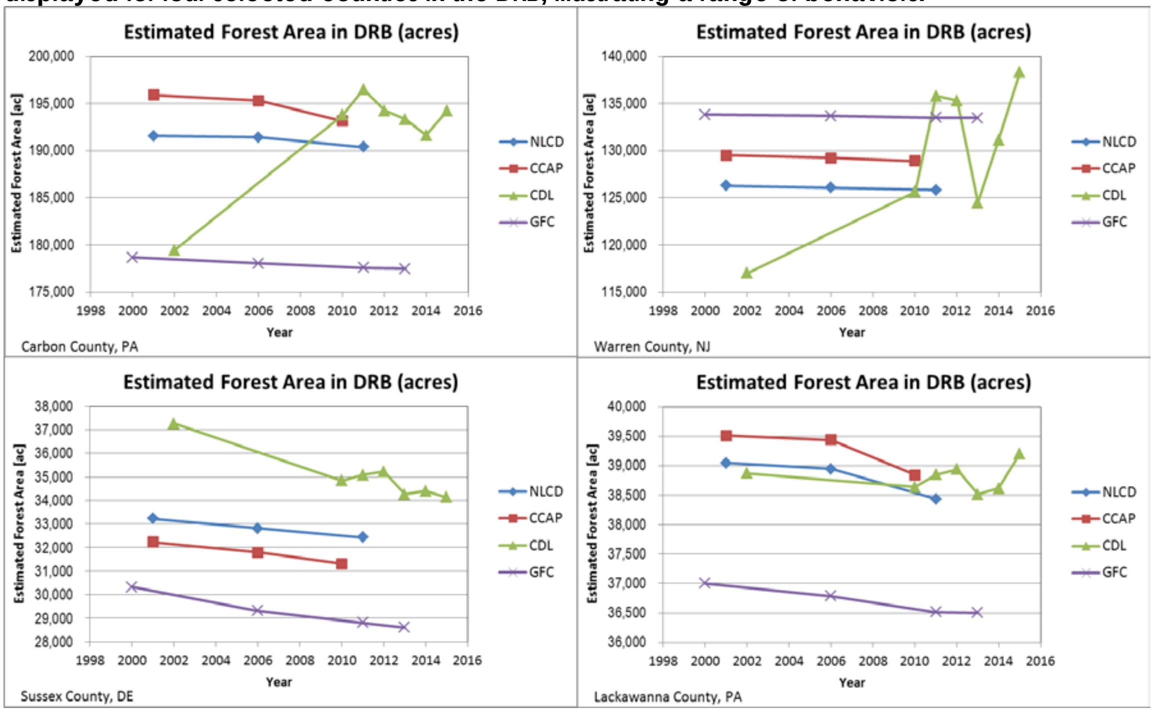
	Mean % tree canopy			% of total forested area		
Land Use Type	NLCD 2011	CDL 2013	CCAP 2010	NLCD 2011	CDL 2013	CCAP 2010
All Forest (forest & forested wetland)	88.7	88.8	89.6	78.2	81.0	79.5
All Other Natural (shrub, grassland, etc.)	31.9	27.5	39.2	3.0	5.0	1.8
All Developed	31.8	33.0	24.6	13.7	8.5	13.6
All Agriculture	15.0	14.0	15.3	5.1	5.5	5.1

B. Comparing Forest Change over time by Dataset

One benefit of established land cover datasets that are released regularly is that they can be used to assess land cover change over time, so long as the methodology does not change too much between successive iterations. The NLCD and CCAP datasets are both released at roughly 5-year intervals, and have been methodologically comparable since about 2001. The CDL dataset is now released annual, but has changed over time with releases having different resolutions for certain years. The 2002 release and all years 2010 and later have a 30-m pixel size. The GFC tree cover dataset has a companion product ("Loss Year") showing "lost" forest area by year from 2000 through 2014. We used this data to generate estimates for total GFC tree coverage for the years 2001, 2006, 2011, and 2013 to match other datasets.

We assessed the forest coverage versus time for all counties in the DRB for the NLCD, CCAP, CDL, and GFC datasets. A sampling of results in Figure 4 shows the range of behavior of the datasets over time for four selected counties.

Figure 4. Comparison of forest area trends for four datasets over the 2000 – 2015 time period. Data are displayed for four selected counties in the DRB, illustrating a range of behaviors.



Several key findings are apparent from the figure:

- CDL data should not be used to estimate changes in forest cover over time. It has very unpredictable year-to-year changes, of varying magnitude, depending on the county.
- The CCAP and NLCD data are generally in fairly close agreement with respect to general amount of forest cover, as well as trends. The CCAP estimates are usually, but not always, higher than NLCD estimates.
- The GFC dataset can be highly variable with respect to the total amount of forest cover. Sometimes it is line with the CCAP or NLCD estimates, and sometimes it is quite different. The general trend in forest loss often is close to NLCD or CCAP data, though the GFC tends to be slightly shallower in slope.
- The difference between datasets often exceeds the total rate of loss within each dataset.

Table 3 shows the mean annual rate of change aggregated to states, and the basin. Note that the range of years is determined by available data years for each dataset.

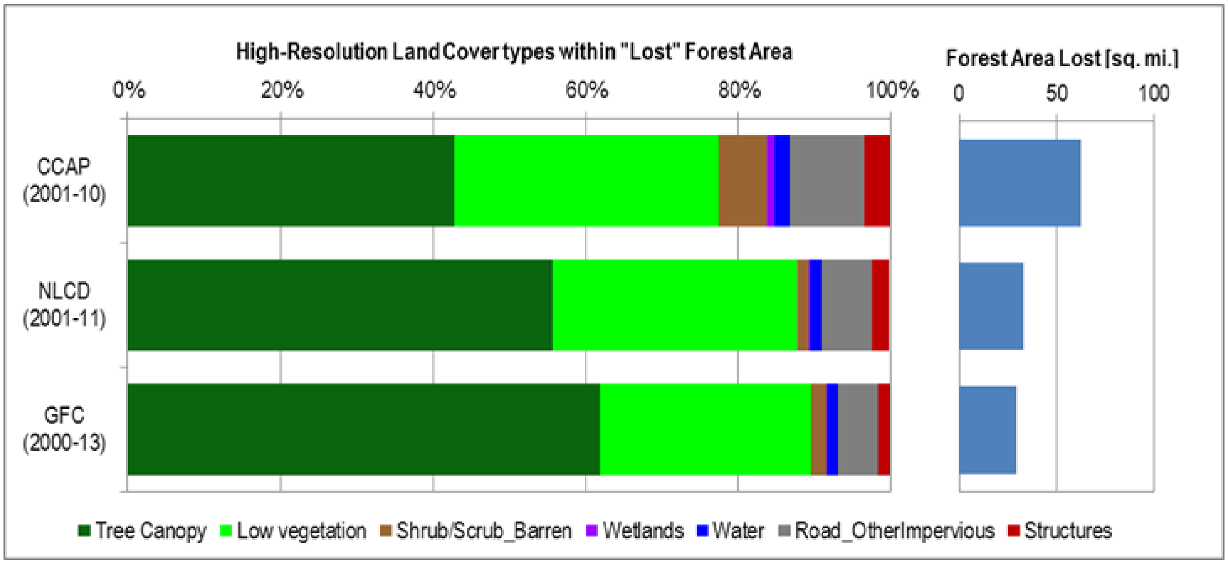
Table 3. Mean annual rate of forest area change by dataset within the Delaware River Basin. [sq. mi./yr]

State	CDL (2002-2015)	CCAP (2001-2010)	NLCD (2001-2011)	GFC (2000-2013)
Delaware	-0.69	-0.55	-0.45	-0.35
New Jersey	3.04	-1.98	-1.63	-1.13
New York	11.93	-1.17	-0.88	-0.58
Pennsylvania	20.88	-6.19	-4.00	-2.40
Basin (DRB)	35.16	-9.89	-6.95	-4.47

C. Land cover breakdown within lost forest area

We identified forest area lost over roughly the 2000-2013 period for the NLCD, CCAP, and GFC datasets on a pixel basis by comparing successive versions of the datasets. Due to its variability, we did not elect to use the CDL data. Using the HRLC data, we investigated for each dataset the current (i.e. 2013 conditions in the HRLC) land cover breakdown within the forest area that was 'lost'. Figure 5 displays the results.

Figure 5. Land cover breakdown from the HRLC dataset (2013) for 'lost' forest areas.



The results are surprising. We observe that for the GFC dataset, over 60 percent of the forest cover that was supposedly "lost" actually remained as tree canopy! This finding was confirmed by detailed visual inspection in ArcGIS. The GFC dataset identified numerous areas of forest loss in still-forested areas, and these areas were not even near other areas flagged as forest loss in other datasets. The NLCD data perform slightly better in identifying non-forested area within the "lost" forest area. The CCAP dataset included the least forest cover (still more than 40% tree canopy) in the areas where forest was lost. Interestingly, the CCAP data also had the greatest total forest loss over the shortest analysis period. So the CCAP dataset appears to more reliably indicate land cover change than the other datasets, even while identifying a greater amount of forest loss.

Recommended Method for determining total forest cover change

Based on the findings in parts A, B, and C, we have identified a recommended methodology for assessing forest cover and the rate of loss in the DRB. In short, we use the estimates of percent tree canopy coverage by land use classification from part A to estimate total forest cover in each land use classification of a selected land use dataset. These percentages are relatively stable, and can be calculated based on a study area and then applied across the whole domain of the selected land use dataset, and forward or backwards in time (i.e., to other data years available).

The result of this method is the ability to not only estimate the total forest cover, but also to estimate where it is located. For instance, this method allows differentiating forest in area classified as forest from forest in developed area, or forest in agricultural area. In some areas, there may be very little area actually classified as forest, but a substantial area of forest cover found in other land use types that would otherwise be completely ignored.

Steps 1 to 5 show how to estimate total tree canopy coverage using a selected land use dataset by computing the percent forest cover within land use classifications. (Steps 1 through 4 should only have to be performed once for each land use dataset, since it is likely only one year of high-resolution land cover data will be available.) Steps 6 to 9 provide instructions for how to estimate forest change between successive iterations of land use dataset. The method accounts for "over-clearing" during the process of land conversion, as it assumes the land use classes that increase in area start out less-forested on average than the mean value for that land use class. A lower estimate of change can be computed by using simply the mean percent forested cover value in Step 8b.

1. Acquire high-resolution land cover data for the study area, and make note of the year the dataset represents.
2. Select a land use dataset (such as the NLCD) and select the data year that most closely approximates the year of the high resolution dataset.
3. Reprocess the high-resolution dataset to create a percent-forest canopy dataset at the same resolution as the land use dataset, with pixels aligned. (In ArcGIS, this involves a 'Conditional' raster process to identify forest versus not forest in the high-resolution dataset. Then, an 'Aggregate' process to reprocess the cell size to match the land use dataset, with the raster snapped to the same grid. Then, apply additional mathematic transformations as needed to end up with a percent forested raster.)
4. Determine the mean percentage of each land use code that is forested by comparing the land use and forest cover dataset (built from the high-resolution land cover dataset) over the spatial domain of the forest cover dataset. (In ArcGIS, use the Zonal Statistics as Table tool. At minimum, select both mean and standard deviation as statistics.) If desired, compute separate statistics for smaller scale zones such as counties or watersheds.
5. Estimate the total forest area in each land use class by multiplying the total area for the class by the mean percent forest cover computed in step 4.
6. Acquire a land use dataset from the same source representing a later year. For example, NLCD 2016 data when it becomes available.
7. Determine the net change in area for each land use class between the original and later year for the land use dataset.
8. Compute the change in forest area for each of the land use classes:
    a. For land use classes that declined in area, multiply the net area change (loss) by the mean forested percentage for that class determined in Step 4.
    b. For land use classes that increase in area, multiply the net area change (gain) by mean forested percentage minus the standard deviation (both determined in Step 4). You may which to apply a different adjustment to meet your needs, such as simply reducing the mean forest cover by 50%.
9. Determine the sum of changes for all land use classes to estimate the total change in forest area. You may wish to break down the study area into logical zones (e.g. counties, watersheds), each with its own mean percentage forested by land use class.

NOTE: This method will work best working forward in time. Suppose the high-resolution land cover data represent 2013 conditions, and the base-year land use data matched to it represent 2011 conditions. The method will work better if the next year of land use data represents 2015 or 2016 conditions, than it will for determining the change from an earlier to the base year. Of course, this backward looking analysis has been tested for NLCD data representing 2006 and 2011, and the results seem reasonable. The main issue is assuming that the mean forest cover percentages applied to the 'lost' areas are valid for the earlier (non-base) year.

Best Estimate of Forest Cover, Forest Loss in the Delaware River Basin

Using the recommended methodology above, we generated an estimate of total DRB forest cover and forest change using the NLCD 2006 and 2011 land use datasets. The CDL data from 2015 represent the most recent land use data that is available for the entire river basin, but unfortunately, the analysis discovered numerous issues with using CDL data for estimating forest change. Instead, the most recent available data is the NLCD 2011 land use data. We recommend repeating this analysis when the 2016 NLCD or CCAP data are released.

Table 4 shows the best estimate of forest cover generated from the method above, and the average annual rate of change from 2006-2011. These estimates were created using the NLCD 2006 and NLCD 2011 land use data. The low estimate reflects a constant assumption for percent forested by land use class. The high estimate assumes the land uses that gain in area start with a percent forested that is equal to the mean percent forested minus one standard deviation. (Means and standard deviation for percent forested by land use were calculated in Part A.) These rates can be compared to the change by dataset in Table 3.

Table 4. Best estimate of total forest cover and range for rate of annual change.

State	Forest Cover (2011)	Forest Cover Change (2006-2011)	Forest Cover Change (2006-2011)
	sq.mi	Low [sq.mi/year]	High [sq.mi/year]
Delaware	347	-0.06	-0.44
New Jersey	1,520	-0.91	-1.69
New York	1,867	-0.65	-1.08
Pennsylvania	3,726	-2.37	-4.58
Delaware River Basin	7,460	-3.99	-7.79

Summary of Data Products provided with this Memo

The table below includes a summary of the data products provided with this analysis. These files include a wealth of useful information, methodology, and background about the data sources, and how the analysis was completed. They also provide useful information in spreadsheet form such as percent forested values by land use for each land use dataset (in this case, see A_Casestudy_LULCDatasets_vsHighResLC.xslx).

File	Description
A_Casestudy_LULCDatasets_vs_HighResLC.xlsx	From part A - Excel file comparing land use and percent forested datasets with high-resolution land cover (percent of area as tree canopy), and estimates of total forest cover by dataset (c. 2010-2013) for PA portion of DRB.
B_ForestLossTrends_byCounty_Dataset.xlsx	From part B – Excel file comparing trends in forest land use from roughly 2000 to 2015 for the CCAP, NLCD, CDL, and GFC datasets. Results are presented by county.
B_ForestLossTrends_byHUC10_Dataset.xlsx	Related to part B – Excel file comparing trends in forest land use from roughly 2000 to 2015 for the CCAP, NLCD, and CDL datasets. Results are presented by HUC-10 watershed (NOTE: watershed results were not described in this memo).
C_NewLandCoverBreakdowninLossArea.xlsx	From part C – Land cover breakdown from the HRLC within the areas of forest land use 'lost' from roughly 2000 – 2013 for the CCAP, NLCD, and GFC datasets.
BestEstimate_County.xlsx	Excel file with best estimate of forest cover by county. Based on NLCD 2011, with variations for generalized percent forested values (by land use) and specific values referenced from individual PA counties.
BestEstimate_HUC10.xlsx	Excel file with best estimate of forest cover by HUC-10. Separate estimates are provided based on NLCD 2011 and CCAP 2010 land use data.
BestEstimate_ForestCover_Map.pdf	Set of maps of forest cover by county using recommended forest cover method. Shows total forest acreage, percent of county area as forest cover, breakdown of forest cover among land uses, and comparison with NLCD area classified as forest.
NLCD_2006-2011_LossCalc_Example.xlsx	Excel file showing an example using three methods for calculating forest loss rates with successive iterations of a land use dataset.
ForestAnalysis_GISProducts.gdb	ArcGIS file geodatabase including best estimates of forest cover by county and by HUC-10 within the DRB.

Recommendations for future forest cover analysis

Based on the findings of this analysis, we have the following recommendations for future assessment of forest cover in the DRB.

- **Estimate forest cover based on land use and high-resolution land cover** – Use the high-resolution land cover data, where available, to estimate the actual forest cover by land use. Use land cover and land use datasets with data years as closely matched as possible. This was the methodology used in this analysis. It should be repeated when more high-resolution land cover become available in other states. Delaware data are already available, and New Jersey and New York will be released soon. Once they are released, estimates of forest cover by land use can be improved with locally derived percent cover value instead of those from Pennsylvania. Be aware that the base-year for analysis may differ by state.

- **Use NLCD and CCAP land use data for forest change analysis**– These datasets appear well-suited to change analysis (post 2001) and are relatively stable and easy to work with (few land use classifications). Forthcoming updates to these datasets for the 2016 data year will be an opportunity to test the methods described. The CCAP data generally has higher baseline forest cover, and higher rate of forest loss as compared to the NLCD. The CCAP data may be more useful in low-lying coastal areas, and the NLCD more appropriate for upland analysis.

- **Do *not* use CDL or GFC data for forest change analysis** – For multiple reasons, the CDL is not useful for forest change analysis. It is methodologically inconsistent, highly variable year to year, and has a huge number of land use classifications. The GFC data are intriguing, but on closer analysis, are not particularly useful, do not identify lost forest cover consistently, and do not correlate well with the high-resolution land cover data. They also appear to underestimate forest loss rates. If updated GFC "percent tree cover" data are released for a more recent year (last release for year 2000), it may be worth re-investigating.

- **Perform finer-scale analysis within land use types using the HRLC data** – One possible extension is to use the HRLC to analyze percent tree canopy (or impervious cover, etc.) by land use at a finer scale such as counties or watersheds. For instance, we note that the average tree canopy coverage *within* land use areas classified as forest appears lower in counties with more development than in sparsely developed areas. That is, the average percent tree cover within the each pixel classified as forest may be lower in a suburban county like Delaware, PA than in heavily-forested Pike County, PA. (This may simply be a result of edge-effects uncertainty between land use types with a 30-m resolution, but is worth further investigation).

References

Han, W., Yang, Z., Di, L., Yagci, A., Han, S., 2014. Making Cropland Data Layer data accessible and actionable in GIS education. Journal of Geography, 113(3), 129-138. Data download available online from: https://nassgeodata.gmu.edu/CropScape/

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. 2013. "High-Resolution Global Maps of 21st-Century Forest Cover Change." Science 342 (15 November): 850–53. Data available on-line from: https://earthenginepartners.appspot.com/science-2013-global-forest/download_v1.2.html

MRLC. 2015. Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K., 2015, Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing*, v. 81, no. 5, p. 345-354. Available online via the Multi-resolution Land Characteristics consortium (MRLC): http://www.mrlc.gov/nlcd2011.php

NOAA, 2016. "Coastal Change Analysis Program (C-CAP) Regional Land Cover Data." Digital Coast. Office of Coastal Management, National Oceanographic and Atmospheric Administration. Data available online from: https://coast.noaa.gov/dataregistry/search/collection/info/ccapregional

University of Vermont Spatial Analysis Laboratory. 2016. "High-Resolution Land Cover, Commonwealth of Pennsylvania, Chesapeake Bay Watershed and Delaware River Basin, 2013." Geographic Information Systems imagery file. Data available on-line from: http://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=3193