

# Assessment of Landsat 7 ETM+ SLC-off Gap-filled Data for Impervious Surface and Canopy Cover Estimation

Prepared by Jon Dewitz, SAIC, April 2004  
USGS EROS Data Center, Sioux Falls, SD

## Introduction

The National Landcover Database 2001 (NLCD) has three primary products, which consist of an Anderson level 2 Landcover, a percent impervious surface estimation and a percent canopy estimation. Each of these products is generated by using three different dates of imagery, all with distinct phenological characteristics corresponding to a generalized spring, leaf-on, and leaf-off definition. Impervious surface and percent canopy are extrapolated using a rule based regression tree. Training data for this extrapolation are derived from DOQQ imagery, and scaled up to Landsat resolution. In this test, three Landsat scenes with the SLC anomaly artificially created were used, and the gaps were filled with histogram-matched data from phenologically similar dates. These scenes were used to create canopy and impervious estimates, and compared with canopy and impervious estimates created from the original non SLC-off Landsat scenes.

## Methodology

NLCD 2001 utilizes multiple Landsat scenes mosaicked to eliminate clouds and provide a clean image for extrapolation. NLCD fills any remaining cloudy areas using regression tree analysis. The first noticeable problem with the SLC-off histogram filled data was that significant areas in two of the three cloud-free original images had been filled with scenes that contained clouds (see figure 1). This rendered approximately one fourth of the gap-filled imagery useless for comparison because of clouds.

Using training data that included only the cloud-free areas common to all three scenes in both the gap-filled and original imagery, an estimation of percent impervious surface and canopy cover were created. Subsets of these extrapolations from the center (which should have little or no gap filled areas) and from the edges (which would have mostly gap filled areas) were compared to the impervious and canopy estimations made from the original SLC-on imagery. The subsets chosen were in areas of high canopy cover and impervious surface area. These subsets were used as the basis for a statistical comparison using the mean and median of these subsets.

## Results

The impervious surface comparison was a very close match. The mean of the impervious estimate was within 0.1% for the comparison areas, and the median was identical. A visual inspection of impervious areas showed no lines from the gap filling process, and all impervious areas were almost identical in estimation.

The percent canopy cover comparison was much more variable. Mean of all areas differed by two to five percent, and median differed in all areas from five to nine percent. A visual inspection did not show any lines from the gap-filling process, but showed large areas of differing estimation (see figure 2) throughout the canopy filled portion of the scenes.

## Conclusions

One initial conclusion is that filling of gap areas with cloudy imagery is not acceptable. The NLCD project goes to great lengths to have imagery it will be using cloud free to provide consistent, comparable data. Imagery that is filled with clouds will not work.

## Imperviousness

The impervious surface prediction fared reasonably well with the gap filled product, as most impervious areas are pseudo-invariant, therefore *impervious cover estimation should be relatively consistent even if filled with poorly matching dates*. The exception to this could be areas of low imperviousness that had canopy and herbaceous cover, but this could not be evaluated with the products chosen for this test.

## Canopy Cover

The results of percent canopy cover estimates were more variable. Upon visual review of the scenes, the areas without clouds that had been gap filled were seamless, and it could not be differentiated where they had been filled. The NLCD project strives to minimize phenology differences when mosaicking as even small differences can produce an evident “seam” in the canopy estimation. It was encouraging that there were no evident “seam” lines from the gap filling process. However, the modeling phase did produce dramatic differences, which most likely relates back to how the canopy estimate is extrapolated. The DOQQ training data included 3 training scenes, of which two were located in heavily gap filled areas. As the canopy model trains in these areas, it looks at the TM imagery in all bands in this same spatial extent and extrapolates these features to rules, which produce the canopy estimation over the much larger area of the entire TM scene. The gap filling process evidently produced enough variability in the localized training area and subsequent rules that different results when extrapolating to the larger scale were seen. This could be minimized if training data from non-gap filled areas was used, but this is not a practical option. *We conclude that consistent canopy estimation is unlikely with gap filled data.*

Figure 1.  
Gap filled product on the left, original Landsat scene on the right.

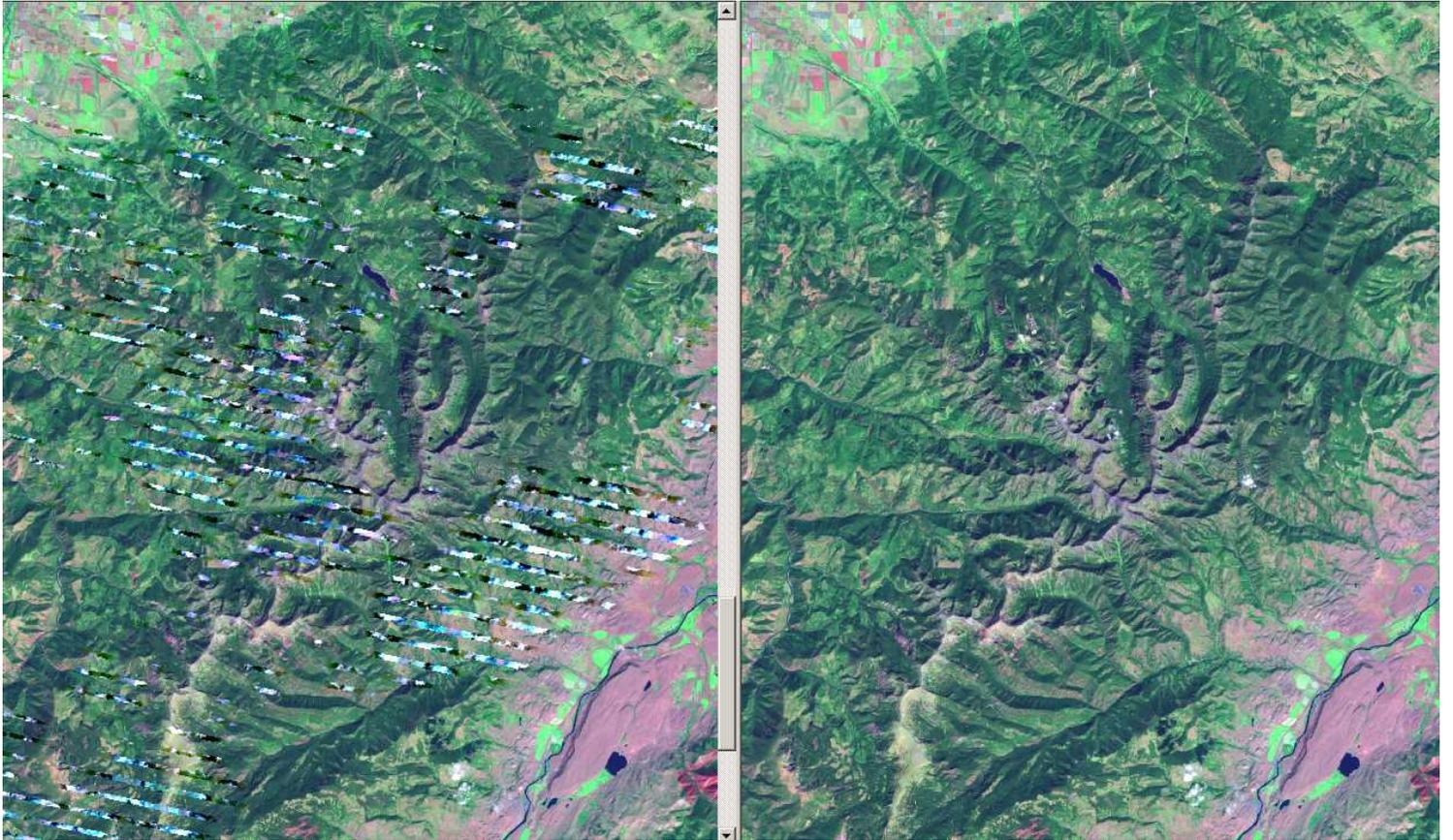
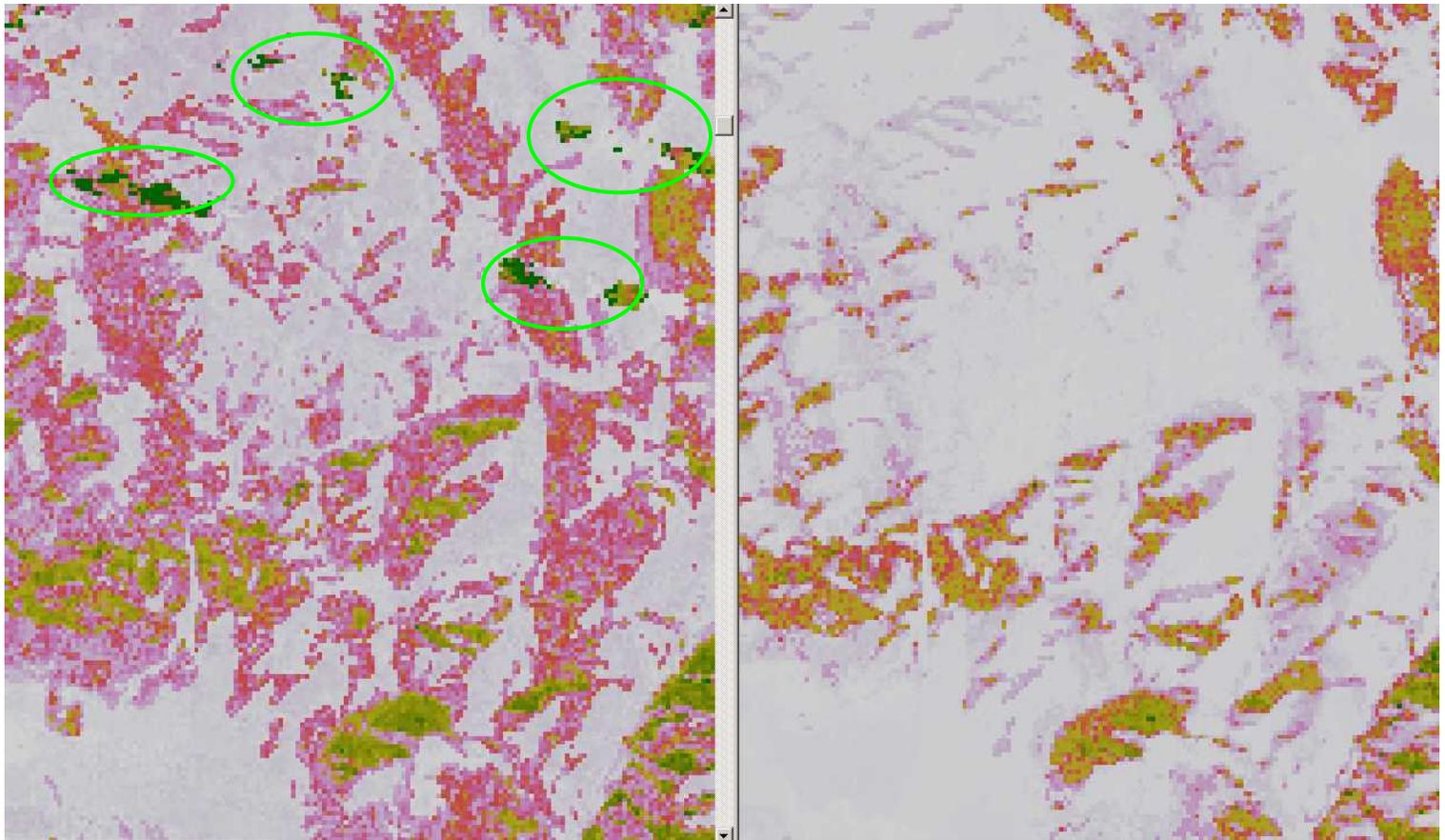
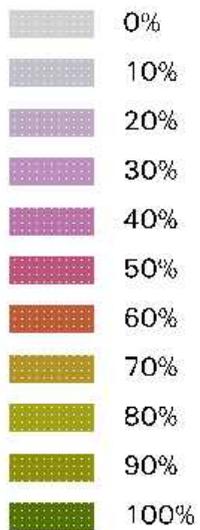


Figure 2



Percent Canopy Cover



Identical subset of canopy estimation with the gap filled product on the left, and original Landsat canopy estimation on the right. Note the large differences in estimation, as well as the SLC-off line circled that was filled with clouds.