Final Report

# Vermont High-Resolution Land Cover

*Updated: 16Aug2019*

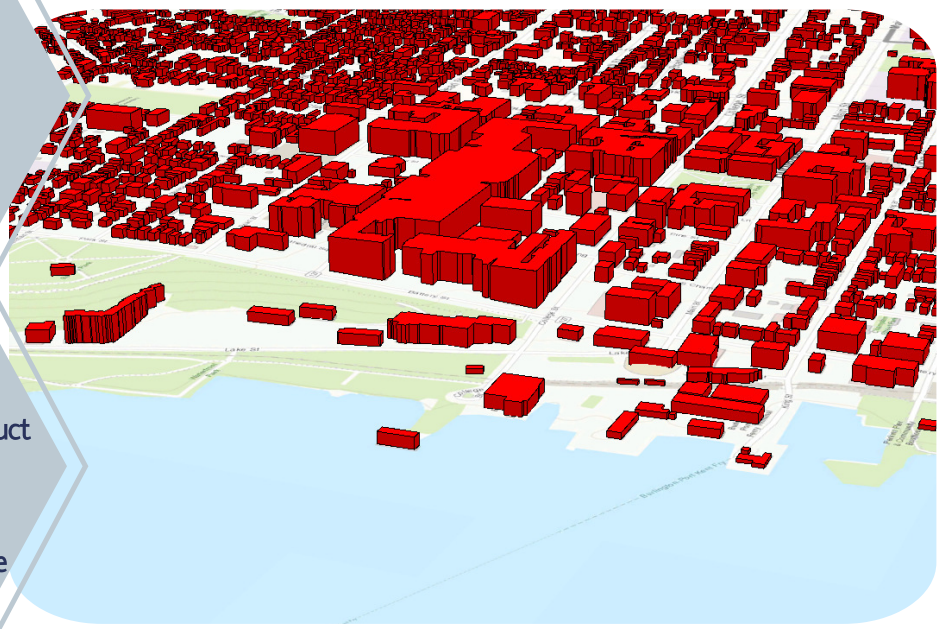
# Project Summary

## Products

- High-resolution land cover
- High-resolution impervious surfaces
- Forest patch
- 3D buildings
- Shurbs, wetlands, and agriculture

## Benefits

- The state's most accurate land cover product
- 900-times more detailed than existing statewide land cover products.
- Impervious surface information that can be summarized at the property parcel level.



Accurate, high-resolution land cover maps are essential to a wide range of landscape-analysis and monitoring efforts, including tree canopy change detection, nonpoint source pollution modeling, carbon stock estimation, spatial epidemiology, and urban planning. These maps provide quantitative evidence that decision-makers and researchers need to understand and materially affect positive solutions to climate change, environmental degradation, and socioeconomic inequality. When produced with high cartographic realism, they are also useful tools for highlighting and simplifying complex inter-relationships among ecological and social phenomena.

Land cover maps do exist for Vermont, but until now they were derived from relatively coarse satellite imagery; underestimating narrowly configured features such as impervious surfaces and failing to detect key components of the landscape that are crucial to water quality such as riparian buffers. This project yielded the most detailed and accurate land cover dataset ever produced for the State of Vermont. Federal, state, and regional organizations have recently made substantial investments in high-resolution imagery and LiDAR for Vermont, and this project harnessed these data to map land cover that is 900-times more detailed than any existing land cover dataset. Leveraged were the data, methods, and efforts of past land cover mapping initiatives in the state, sponsored by the Lake Champlain Basin Program and AmericaView. It succeeded in generating a seamless, high-resolution land cover dataset of Vermont reflecting 2016 conditions.

**Jarlath O'Neil-Dunne**
Director, UVM Spatial Analysis Lab

# Product Descriptions

This project yielded several distinct yet related products. These products were designed to meet the needs of a broad range of stakeholders and can be used separately or integrated, depending on analytical requirements. There are two reasons for having distinct products. First, feature types may overlap. For example, a forested wetland is a collection of other land cover features (tree canopy, water, shrub) and thus including forested wetlands in the base land cover dataset distorts the representation and accounting of features such as tree canopy. The second, but related reason, is that different mapping standards are required for certain land cover features. For example, individual shrubs are difficult to map, but shrubby areas can be mapped with reasonable accuracy.

## Base Land Cover

Raster land cover product with a resolution of 0.5-meters. This is the foundational, "top-down" land cover product from which the supplemental land cover products are derived. Includes vegetated, impervious, soil, and water features.

## Supplemental Land Cover

Vector land cover product derived from the base land cover dataset. This supplemental product includes shrub, agricultural, and wetland land cover types that are delineations of base land cover features.

## Tree Canopy

Vector product that refines the base land cover tree canopy class into individual tree canopy polygons and categorizes them as either deciduous or coniferous.

## Impervious Surfaces

Vector land cover product that takes a bottom-up approach to mapping impervious surfaces by including features obscured by tree canopy. Includes buildings, roads, other paved, bare soil, and railroad classes.

## Forest Patch

Vector land product that separates the tree canopy into forest patch classes and based on size, shape, and morphology.

## 3D Buildings

Buildings from the impervious surface dataset attributed with height information obtained from LiDAR data.

# Product Specifications

## Base Land Cover

- Tree Canopy
- Grass/Shrub
- Bare Soil
- Water
- Buildings
- Roads
- Railroads
- Other Impervious

## Forest Patch

- Small
- Medium
- Large

## Supplemental Land Cover

### Wetlands

- Forested Wetlands
- Shrubby Wetlands
- Emergent Wetlands

### Agriculture

- Hay
- Pasture
- Cultivated Crops

### Shrub

- Shubland

## Tree Canopy

- Deciduous
- Coniferous

## Impervious Surfaces

- Buildings
- Roads
- Railroads
- Other Impervious
- Bare Soil

## 3D Buildings

- Buildings attributed with height

## Deliverables

Final deliverables consisted of separate geospatial datasets in either raster or vector format. The format was selected based on the type of feature, number of features, and extent of coverage. All geospatial products were documented with metadata using the North American ISO metadata profile. A det

# Methods & Standards

The source data and standards for the products are presented below. Note that the minimum mapping unit and accuracy information apply only to the base land cover and impervious surface products.

## Source data

- Leaf-on imagery: 2016 NAIP
- Leaf-off imagery: latest VT orthoimagery
- LiDAR: latest LiDAR
- Road center-lines
- Property parcels
- VT Hydrograpy Dataset

## Mapping Approach

- Automated feature extraction using expert systems within an object-based image analysis framework
- Manual review and editing of all output
- Minimum mapping unit: 20 square meters

## Quality & Accuracy

- Stratified sampling. 5000 points with a minimum of 100 points per class.
- Overall accuracy >90%
- Per-class accuracy >80%
- Topologically correct (vector only)
- Cartographically realistic

# Feature Extraction

Land cover features were extracted using a combination of automated feature extraction techniques and manual editing. An object-based image analysis (OBIA) system was designed, developed, and deployed to handle the automated feature extraction. This system integrated the relevant source datasets into a data stack, then used a combination of segmentation, classification, and morphology algorithms to extract features. OBIA approaches are extraordinarily useful for data fusion projects because they are capable of harnassing the information contained within vector, raster, and point cloud datasets without the need for data transformation. Furthermore, OBIA systems can incorporate spectral, height, and spatial information that these datasets possess, thereby replicating the human cognitive process.

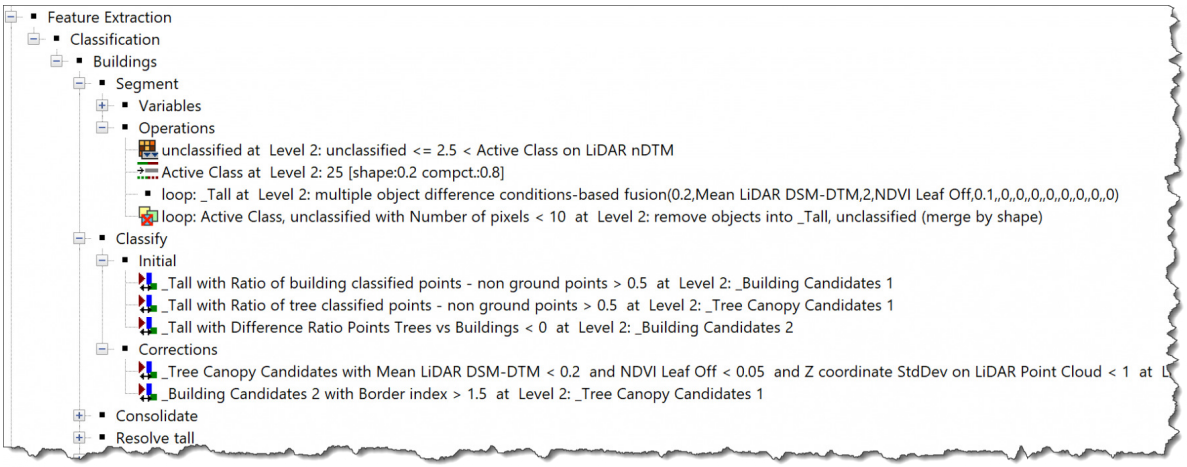



*Figure 1. Example of a portion of the rule-based expert system used for building extraction.*

The general workflow for feature extraction focused on mapping individual land cover classes and then reintegrating these features back into the feature extraction workflow. Over the course of the project, this resulted in increasingly greater amounts of contextual information available for the automated feature extraction process. For example, In the case of impervious surface mapping, buildings were first automatically mapped and edited. Next, the buildings were brought back into the feature extraction workflow to map roads as roads and buildings are mutually exclusive classes. Both these datasets were then used to extract the remaining impervious surface classes as these features are spatially associated with buildings and roads.

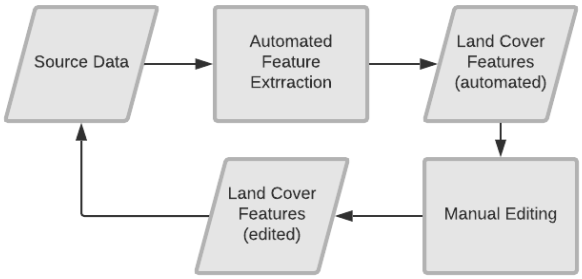



*Figure 2. The general feature extraction workflow. Features were first automatically mapped from the vector, LiDAR, and image source datasets. They were then manually edited and finally re-incorporated into the source data to extract the next set of features.*

# Buildings

This project mapped 398,851 buildings polygons. Buildings were extracted using a combination of E911 building points, Microsoft building polygons, LiDAR point clouds, LiDAR surface models, leaf-off Vermont orthophotos, and NAIP imagery. The E911 buildings gave a good indication of where structures do exist, but points were not present for all buildings, and there were some positional accuracy issues. The Microsoft building polygons are an excellent dataset. Derived from photogrammetric data, they have the limitation in that they tend to underrepresent buildings with overhanging tree canopy. The LiDAR point clouds were run through a customized algorithm that normalized the point Z-values relative to the ground and then classified the points into ground, tree, and building classes. This classification was generally effective, but there was confusion between the tree and buildings classes in many areas. The LiDAR surface models produced included a Digital Surface Model (DSM) consisting of the first returns, a Digital Terrain Model (DTM) consisting of the last returns, and a Digital Elevation Model (DEM) consisting of the ground classified points. The general logic for extracting buildings was first to use the Microsoft buildings, carrying out refinements as required using the LiDAR. For buildings not captured by Microsoft the logic was: 1) near the E911 building points, 2) a high percentage of building classified points, 3) tall in the nDSM (DSM-DEM), 4) little difference between the DSM and DTM, and 5) low NDVI values in the imagery. Each building had its maximum height assigned based on the building classified LiDAR point Z-values. It should be noted that complex roof structures were not extracted. End users will note that some buildings have jagged edges due to side effects of the orthogonalization routines employed.

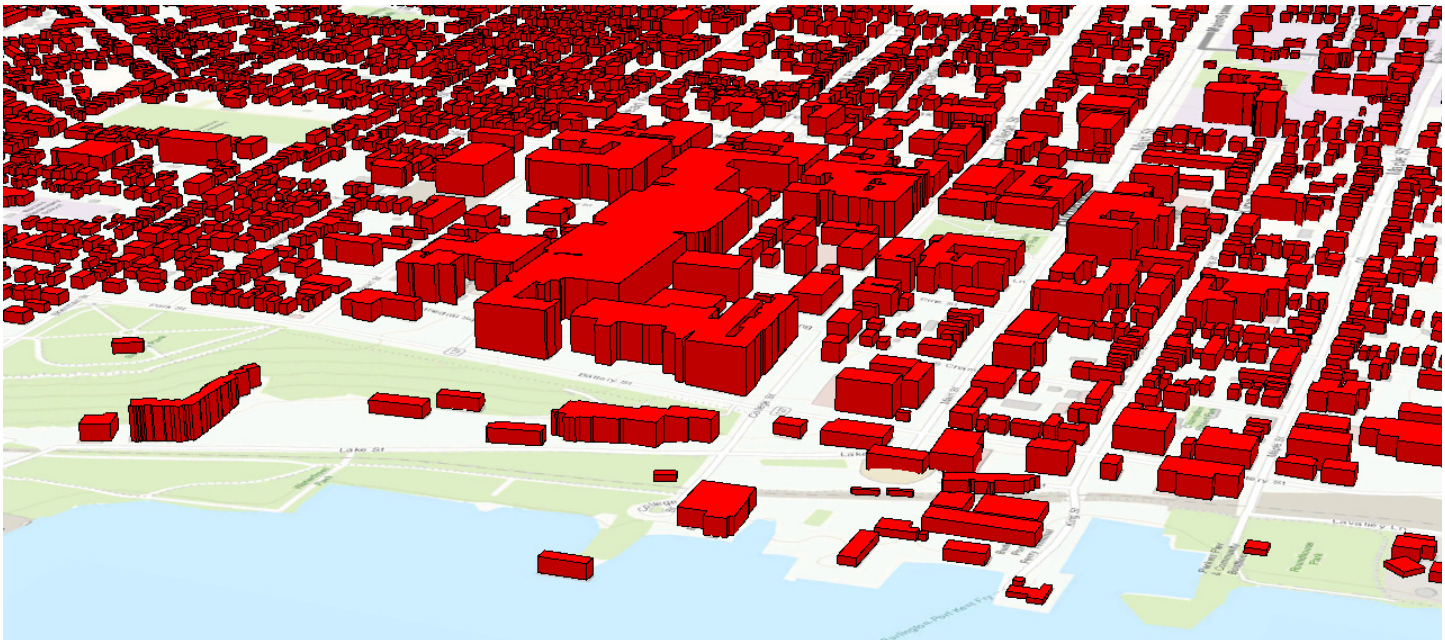


*Figure 3. Example of the 3D buildings final product.*

# Impervious Surfaces

The impervious surface feature extraction data stack consisted of all the dataset used to extract buildings plus the building polygons, road centerlines, and E911 driveways. 1,018,792 impervious features were mapped. An iterative growing routine was used to expand the road and railroad centerlines into road and railroad polygons, respectively. The method evaluated the spectral characteristics of the imagery immediately around the centerline then incrementally grew the polygons as long as the spectral and textural characteristics remained similar. Other paved surfaces (driveways and parking lots) were mapped based on their spectral properties in the imagery (principally low NDVI and brightness), spatial configuration, and proximity to roads and buildings. Bare soil features were initially mapped to the other paved class then separated primarily based on the fact that they reflected more red light than hard impervious surfaces. Manual editing focused heavily on the other paved class, particularly in areas where the surfaces were obscured by tree canopy. These obscured impervious features proved to be the most challenging to map. In many instances, they were difficult to discern using manual interpretation. Dense tree branches, inaccurate or missing driveway centerlines, and low-quality LiDAR intensity data all contributed to the difficulty of mapping other paved features, particularly in rural, forested areas.

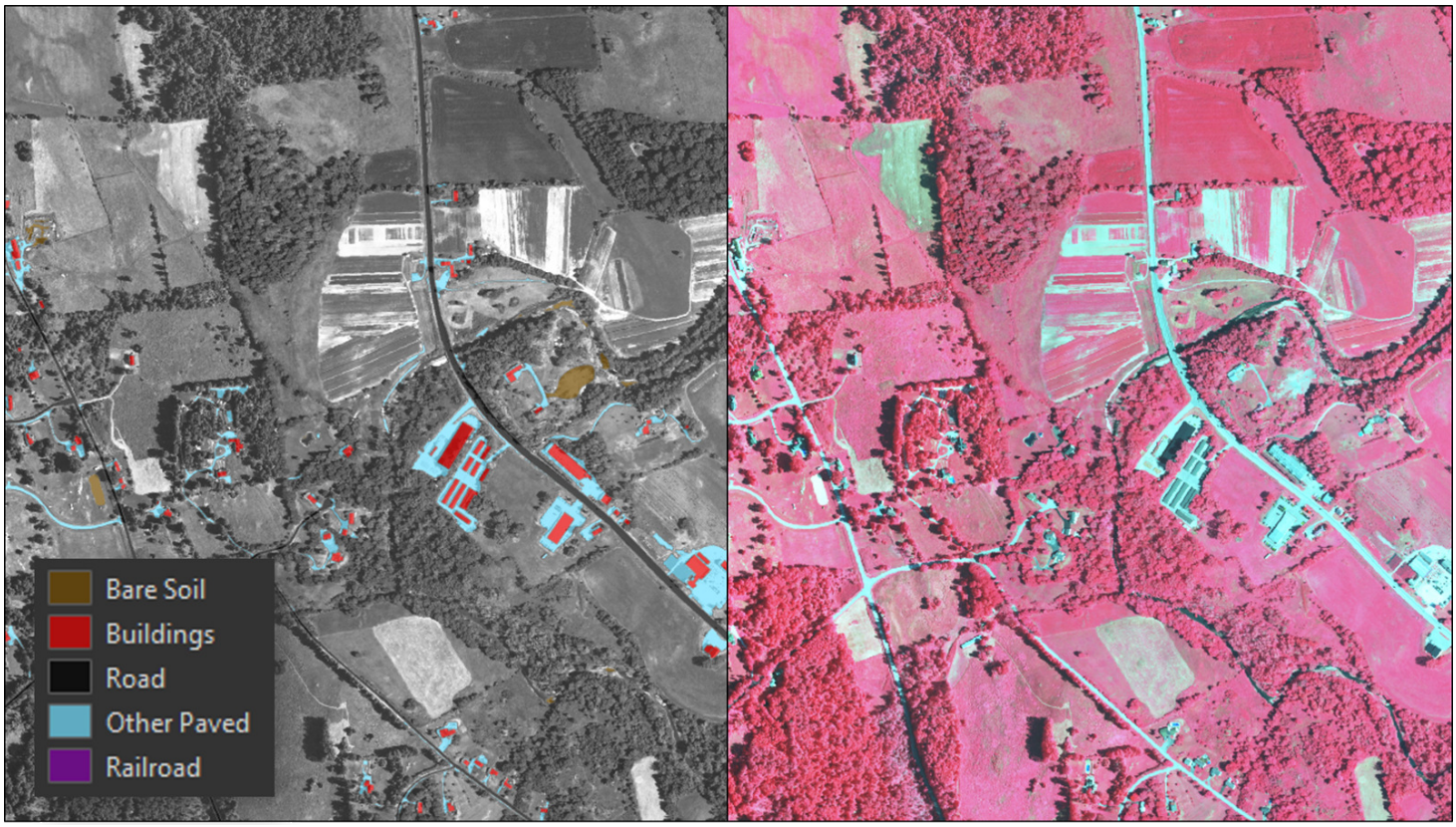



*Figure 4. Example of the impervious surface dataset.*

# Tree Canopy

The same data stack used to map buildings was adopted to map tree canopy, with the addition of the building polygons. Tree canopy feature extraction involved two principal workflows. The first focused on mapping tree canopy; the second focused on categorizing tree canopy as either deciduous or coniferous. Manual editing was only performed on the aggregate tree canopy; the deciduous/coniferous distinction was entirely an automated process. The expert system developed for extracting tree canopy was virtually identical to buildings, but with the logic reversed. Trees were tall in the nDSM, had a high percentage of tree classified points in the LiDAR point cloud, a more significant difference when the DTM was subtracted from the DSM, and high NDVI values. Once manual edits on the aggregate tree canopy dataset were complete, it was fed back into the OBIA system where the deciduous/coniferous distinction was performed by looking at the differences between NDVI in the NAIP (leaf-on) and NDVI in the VT orthophotos (leaf-off). Coniferous species had high NDVI values in both datasets, whereas deciduous had significantly lower NDVI values in the leaf-off orthophotos. The greatest challenging in extracting tree canopy occurred in areas of dense deciduous forest. All of the LiDAR datasets were acquired under leaf-off conditions, but the point density within the deciduous forested areas varied greatly among LiDAR collects. In many cases, trees that could be seen in the NAIP had no values that met the height threshold for trees in the LiDAR nDSM.

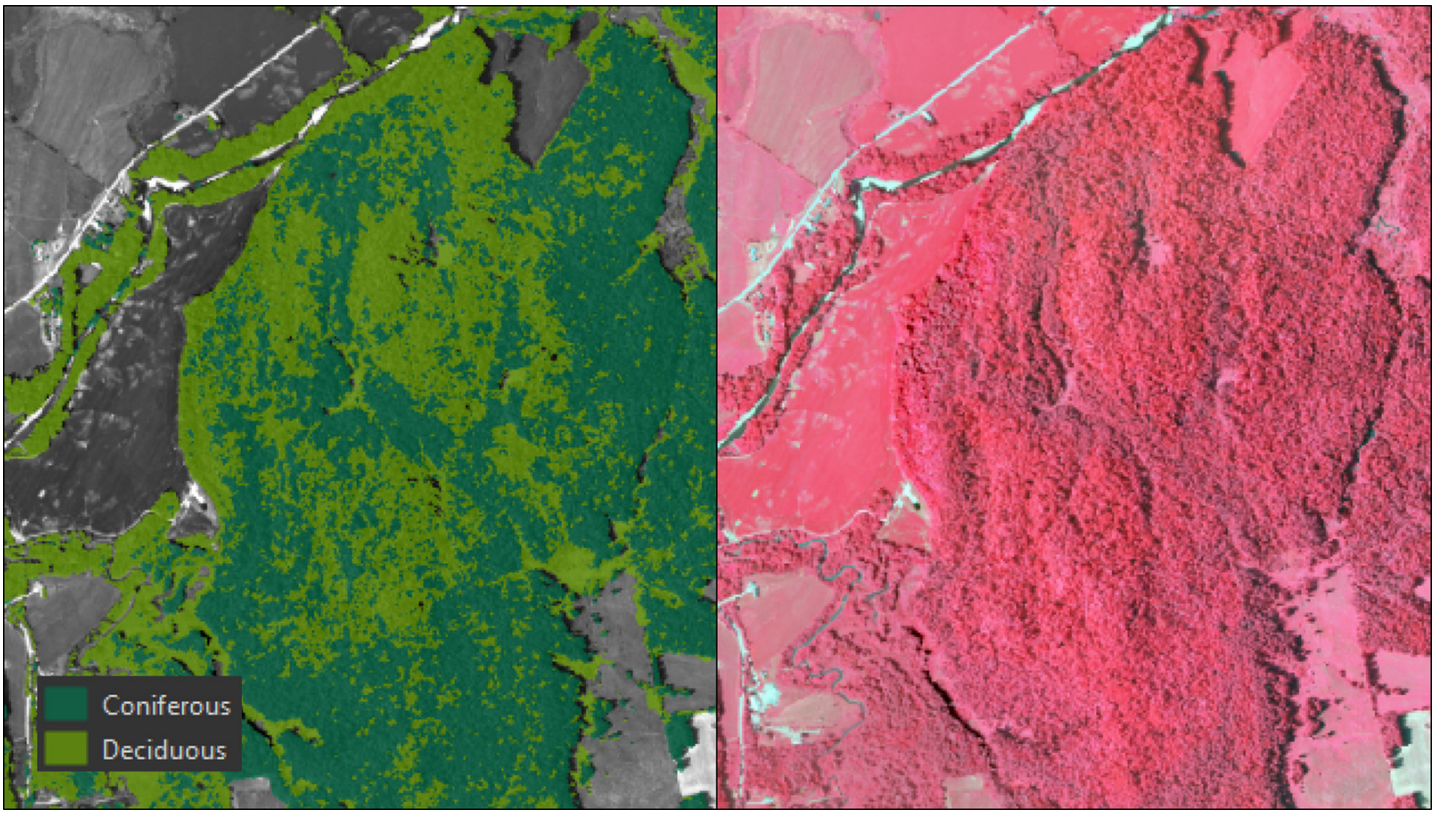



*Figure 5. Example of the tree canopy decidous/coniferous final product.*

# 8-Class Land Cover

The 8-class land cover dataset was a top-down synthesis of the impervious and tree canopy datasets plus two additional classes, water and grass/shrub. The data stack employed for feature extraction matched that of the impervious dataset plus the Vermont Hydrography Dataset (VHD) water polygons and lines. Tree canopy was first incorporated, followed by the impervious classes - buildings, roads, railroads, other paved, and bare soil. These classes were added in by exception, meaning that only those portions of these features that did not overlap tree canopy were included. Added in were the water and grass/shrub classes. The VHD served as the basis for extracting water polygons, but as it was both dated and contained errors, it was modified using an expert system that relied heavily on the spectral information present in the VT orthophotos and NAIP imagery to improve the delineation of water features. These two image datasets, acquired at different times, did not agree, most noticeably along rivers and streams, many of which were at or near peak spring flow when the VT orthophotos were collected. To address the resulting issues from this disagreement, manual edits for the 8-class land cover primarily centered on the water class, with all remaining areas of the landscape assigned to the grass/shrub class. The 8-class land cover served as the foundational dataset for deriving the supplemental shrub and wetland datasets.

*Figure 5. Example of the 8-class land cover final product for an agricultural landscape.*

# 8-Class Land Cover

*Figure 6. Example of the 8-class land cover final product for an urban landscape.*

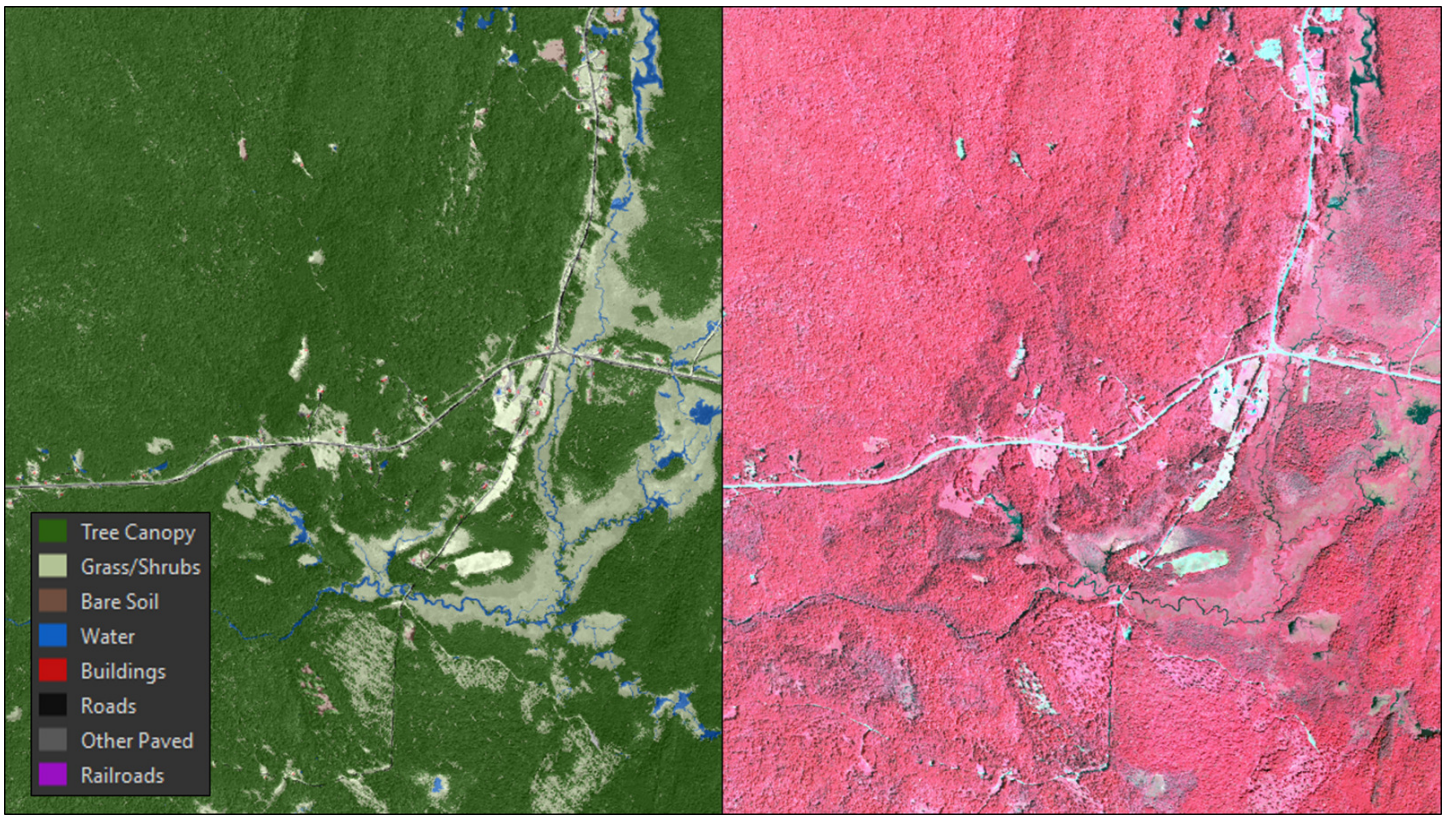



*Figure 7. Example of the 8-class land cover final product for a forested landscape.*

# Agriculture

National Land Cover Database (NLCD) and the 2016 Cropscape dataset. A segmentation algorithm created objects that generally reflected field boundaries using the grass/shrub class was used to seed the segmentation process. Morphological routines were employed to create more homogeneous objects to eliminate the salt and pepper appearance associated with having isolated and small rows of trees in the fields. Sub-objects within the fields were then mapped to indicate as to whether the field was crop, hay, or pasture. The agriculture cover type was challenging to determine from imagery alone as these cover types look similar, exist within rotation schemes, and the image acquisition time may not reflect the actual cover type in the case of crop or hay harvests. To help to improve the distinction of the classes the NLCD and Cropscape datasets were used for attribution, but as these two datasets often had conflicting classifications, class assignment for some fields was pushed down to the best guess of the technician conducting the manual review. As evident by the low accuracies for the agricultural classes mapped by both NLCD and Cropscape, agricultural mapping from remotely sensed data is challenging. It is recommended that end-users of this product conduct their own evaluation to determine its utility, keeping in mind that agricultural fields can change their crop type and alter between crops and hay on an annual basis.

*Figure 8. Example of the agricultural final product.*

# Shrubs

Grass/shrub features in the 8-class land cover that did not fall into the agricultural class were considered for assignment to the shrub class. The focus on shrub mapping was delineating patches of shrubs as opposed to individual shrubs. To be considered a shrub patch at least 10% of the polygon had to consist of shrubby vegetation. These are areas that are typically in a successional phase, transitioning to forest from agriculture or as the result of openings in a forest due to harvesting or natural events. The data stack for shrub mapping consisted of the VT orthophotos, the NAIP, and the LiDAR surface models. The expert system classified objects as shrubby based on the appearance of features that fell within the shrub height tolerance or the presence of small shadows throw from shrubs. The latter was more often used to classify shrubs, which are most often deciduous and did not usually appear in the LiDAR datasets. As shrubs are a transitional class end users of the product should keep in mind that shrub areas could have changed to forested tree canopy or have reverted to active agricultural use since the 2016 time period. The manual editing of the shrub dataset was not as extensive as it was for the base datasets. Shrubs are mutually exclusive to the agricultural classes, although minor overlaps do occur due to the morphological routines employed. The shrub features do overlap with the wetland dataset discussed on the next page as an area can be considered to be both shrubby and scrub/shrub wetland.

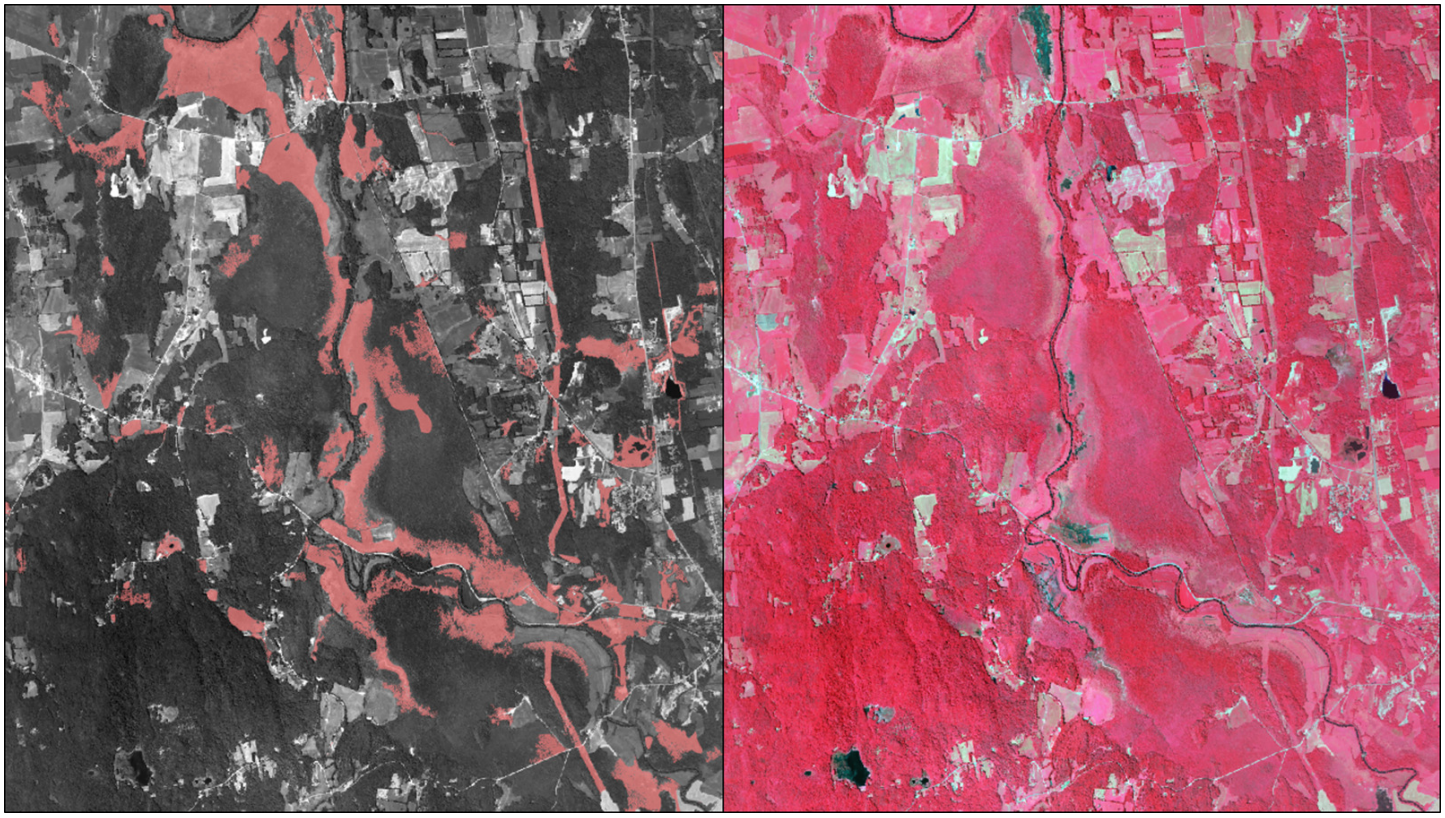


*Figure 8. Example of the shrub final product.*

# Wetlands

Unlike the shrub and agricultural classes, the initial wetlands mapping was initially done independently of the 8-class land cover, with the 8-class land cover used to attribute the wetland polygons after they were delineated. Wetlands delineation relied heavily on a LiDAR-derived Compound Topographic Index (CTI). CTI is based on a LiDAR DEM and combines slope and flow potential, identifying topographies where water could theoretically collect; these sites are very highly textured in CTI layers. Leaf-off VT orthophotos assisted with the classification of sites with high CTI textures, especially the near-infrared band as water absorbs almost all near-infrared energy. After potential wetlands were identified, they were further classified into three primary types (Emergent, Scrub\Shrub, Forested) using the 8-class land cover. Vegetation height is not a perfect indicator of wetland type, overlooking species-specific variation in composition and morphology, but it is a reasonable proxy for generalized wetland features. Wetland mapping from remotely sensed data has been a long-standing challenge. While this mapping represents a definite improvement over past efforts in the state, it no doubt contains flaws. End-users of the product are encouraged to evaluate it carefully. This wetlands dataset should not be used for regulatory enforcement, nor is it a replacement for detailed field mapping.

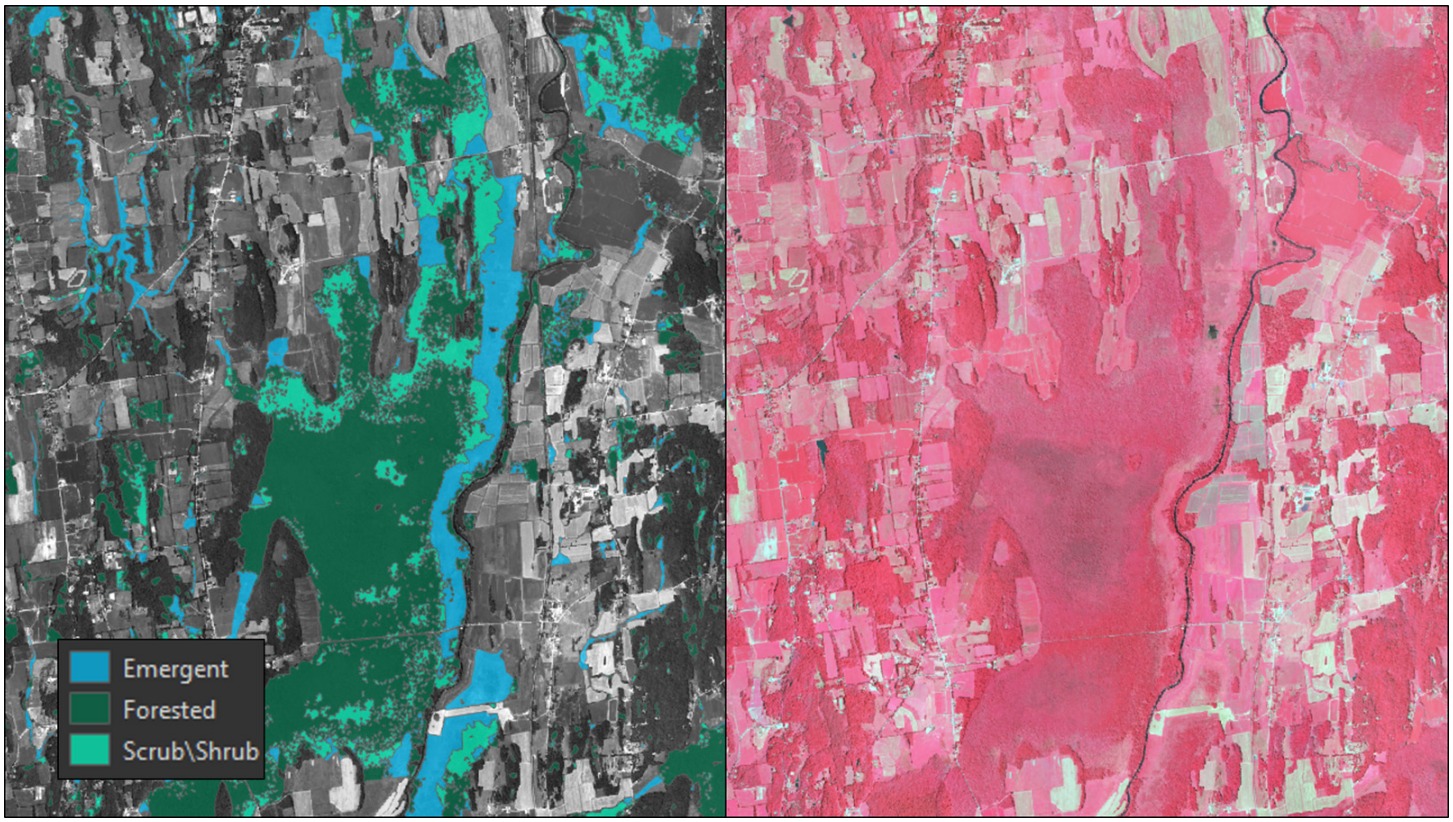



*Figure 8. Example of the wetland final product.*

# Forest Patch

The tree canopy dataset was used as the foundation for the forest patch mapping. The deciduous and coniferous classes were consolidated into a single tree canopy class. The OBIA system divided the tree canopy into three patch classes: 1) small, 2) medium, and 3) large. Small patches consist of individual trees, or small clumps of 2-3, medium patches comprised groups of trees that lack a definitive core, and large patches are consistent with what many would describe as core forest. Forest patch mapping is a largely cartographic exercise as there is no universally accepted definition for delineating forest patches. This mapping did not evaluate the understory, which would determine whether the forested area is mowed or is in a natural state. The approach to separating out the forest patches relied on a series of rules and morphological routines. The rules evaluated the patches based on their size, perimeter to edge ratio, and border index. The morphology routines used grow/shrink approaches and sanding algorithms to split off portions of larger patches into smaller ones. A downside of the forest patch delineation was that the routines were very CPU and RAM intensive, necessitating that the data be tiled. As a result, edge effects are present in the data. The forest patches, particularly the large ones, do allow for small gaps within the canopy, but any large breaks, whether due to anthropogenic or natural features, results in a distinct patch.

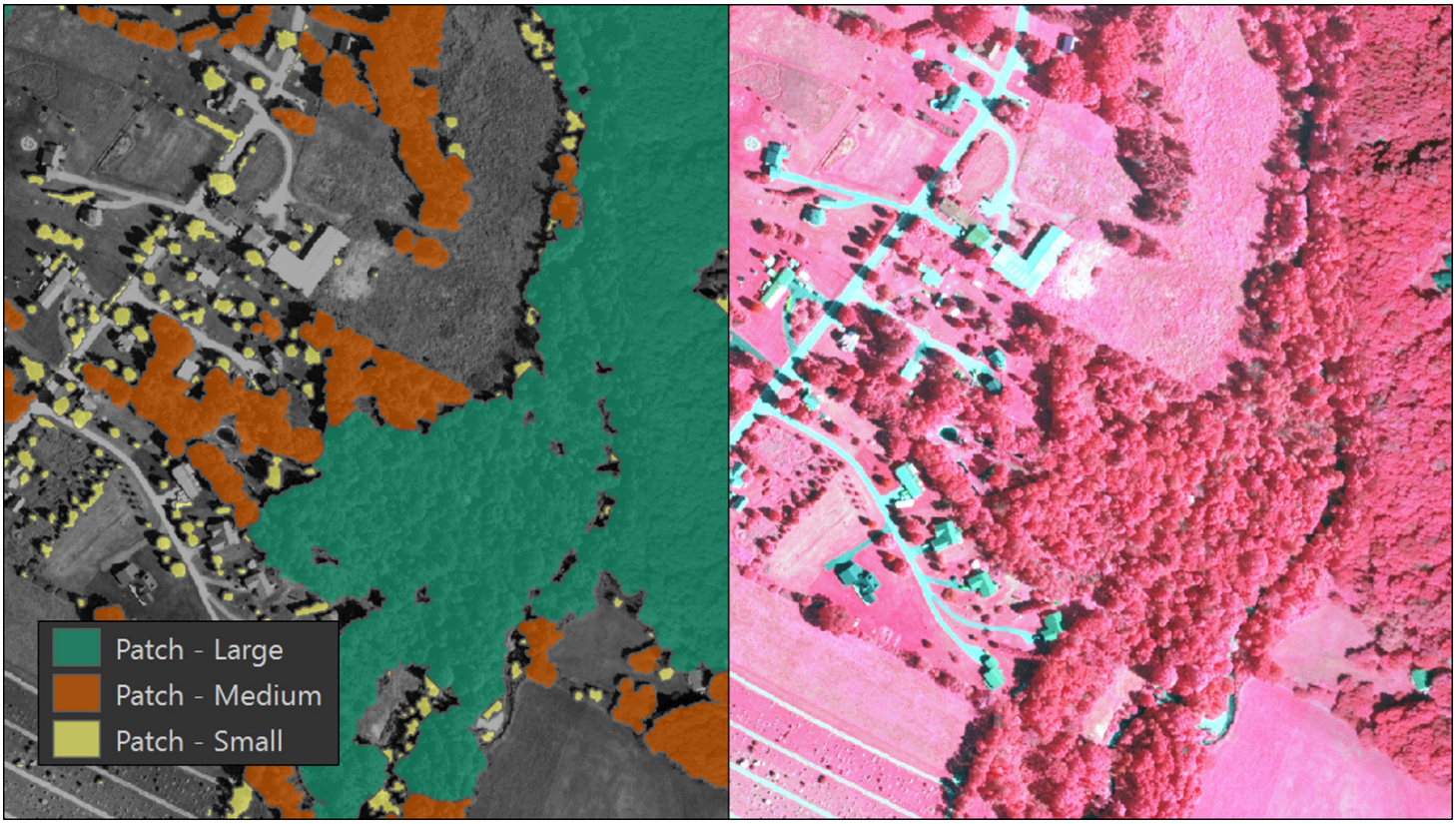



*Figure 9. Example of the forest patch final product.*

# Accuracy Assessment

An accuracy assessment of the base 8-class land cover product was carried out through an independent assessment of 5000 points. Establishing the 5000 points posed a challenge. The points needed to be randomly distributed to ensure there was no bias in point placement but the high percentage of tree canopy within Vermont meant that a strictly random placing of 5000 points resulted in the non-tree canopy classes having too few points to draw meaningful conclusions. To accommodate both of these needs, a stratified sampling first placed 100 points randomly within each class and then an additional random sampling paced 4200 points throughout the state. Each point was independently reviewed and assigned a reference class based on the 2016 NAIP imagery. The land cover was not used for reference class assignment. Once the reference class assignment was complete, the points were overlaid on the 8-class land cover to obtain the mapped class. This dataset was then used to calculate the producer's, user's, and overall accuracy. The user's accuracy, which is an estimate of the likelihood of any pixel in the land cover actually belonging to that class is the most pertinent statistic for end-users of the product.

The mapping surpassed the goal of 92% overall accuracy, but the user's accuracy of some classes are cause for concern. The bare soil user's accuracy was 78%, and the other paved user's accuracy was 84%. The values reflect the challenges in mapping features that are narrowly configured, exist in heterogeneous landscapes, and have similar characteristics in the imagery and LiDAR to other land cover features. Water, which is spectrally distinct in imagery, had the highest user's accuracy at 97%. Tree canopy and buildings, both of which benefit from having distinct characteristics in both LiDAR and imagers, had user's accuracies of 97% and 95%, respectively. A qualitative evaluation of the points in which the reference class and mapped class disagreed found that many errors occurred along the edges of features. These issues could often be traced to misalignment between source datasets and the side effects of morphology algorithms that were employed to improve the cartographic appearance of the data.

| | | Reference Data | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Tree Canopy | Grass/Shrub | Bare Soil | Water | Buildings | Roads | Other Paved | Railroads | Total | User's Accuracy |
| Mapped Data | Tree Canopy | 1723 | 33 | 8 | 12 | 27 | 7 | 4 | 1 | 1815 | 95% |
| | Grass/Shrub | 63 | 995 | 14 | 2 | 5 | 8 | 26 | 2 | 1115 | 89% |
| | Bare Soil | | 16 | 112 | | | 5 | 8 | 2 | 143 | 78% |
| | Water | 4 | 9 | | 844 | | | 9 | | 866 | 97% |
| | Buildings | 9 | 3 | | | 206 | 6 | 3 | | 227 | 91% |
| | Roads | 2 | 7 | | | | 301 | 12 | | 322 | 93% |
| | Other Paved | 26 | 33 | 6 | | | | 333 | | 398 | 84% |
| | Railroads | | 1 | | | | | 6 | 107 | 114 | 94% |
| | Total | 1827 | 1097 | 140 | 858 | 238 | 327 | 401 | 112 | 5000 | |
| | Producer's Accuracy | 94% | 91% | 80% | 98% | 87% | 92% | 83% | 96% | | 92% |

*Table 1. Accuracy asssessment error matrix.*

# Conclusions

This project succeeded in meeting its objectives, developing a high-resolution land cover dataset for Vermont whose overall accuracy exceeds 90%.

Vermont is now one of the few states in the nation with wall-to-wall high-resolution land cover data and the only state to currently have such a broad suite of base and supplementary land cover products.

The combination of top-down and bottom-up mapping, along with separating the base and supplemental land cover products allowed for the mapping of features that have different standards, providing the highest amount of flexibility to end-users.

High-resolution extraordinarily challenging. The high volume of data, the number of distinct data collects, temporal disagreements between data, data quality, and heterogeneous landscape all played a role. Carrying out even basic display and geoprocessing tasks can be slow or impossible due to the size of the data products.

# Recommendatons

The accuracy of the other paved class needs to be improved. This issue is especially important given the interest in the state in an accurate accounting of impervious surfaces. The accuracy issues are most acute in forested areas where impervious surfaces are obscured by tree canopy. An improved version of the E911 driveways dataset would help to resolve this issue.

Vermont should establish a plan to regularly update the high-resolution land cover data products as new source data become available. Updates to the existing products can be made cost-effective by performing hot spot analyses to determine areas of change and thereby reducing the mapping effort.