NRAC415: West Virginia Wetland Inventory and Modeling Project Phase III

Dr. Mike Strager and Aaron Maxwell

Statewide Forest Cover:

Forest cover was derived from 2011, 4 band, leaf-on, growing season, 1 meter pixel size imagery. The National Agricultural Imagery Program (NAIP) orthophotography were obtained from the West Virginia GIS Technical Center as 1831 quarter quads in uncompressed TIFF format covering the full extent of West Virginia. The West Virginia GIS Technical Center obtained the data from the USDA Farm Service Agency. The data represent growing season 2011 conditions with minimal cloud coverage.

We merged the quarter quads by contiguous areas averaging 16 quarter quads or 4 quads. This was done in ERDAS Imagine using the mosaic wizard. The imagery was converted to IMG format. The file size was commonly between 1 and 3 Gb per merged area, and this process resulted in 117 areas to classify.

The bulk of the processing time was spent training each of the 117 pieces of imagery. We trained for forested/woody, grasslands/herbaceous, and barren/non-vegetated cover. We found there was a need to collect a large number of training samples in order to accurately extract the LULC. An average of 1 to 3 hours was spent training each area. The following number of training data was collected statewide:

Forested/Woody: 21,904 polygons
Grasslands/Herbaceous: 31,322 polygons
Barren/Non-vegetated: 8,503 polygons

The classes of interest are defined as follows:

Forested/Woody: Any forest or woody area covered by tree canopy.

Grasslands/Herbaceous: Any vegetated area dominated by non- woody vegetation. This includes some agricultural fields, pastureland, grasslands, reclaimed mine lands, some developed areas, and natural grasslands such as glades.

Barren/Non-Vegetated: Any non-vegetated area. This includes impervious areas, parking lots, exposed rock, roadways, paved surfaces, and natural non-vegetated features.

Our goal was to capture the spectral, textural, and land use variability within our defined classes. The training data were collected as polygon shapefiles by manual photograph interpretation of the 2011 NAIP orthophotography. The training data are provided as a product, and it would be possible to use this training data on future extractions of NAIP imagery as opposed to creating a new training dataset.

Each piece of imagery was processed in the object-based classification tool Feature Analyst 5 by Overwatch, an extension within ArcMap 10. In order to collect textural information for the classification, we used a Manhattan pattern with a pattern width of 9. We did not resample the imagery, so a 1 m pixel size LULC was obtained. The imagery was analyzed with a standard deviation stretch and the analysis was masked to the extent of the imagery. We performed an aggregation to remove areas of less than 300 square meters in order to smooth the output and remove “salt and pepper” artifacts.

We merged all of the extractions, as categorical raster data in GRID format, in ArcMap 10 using the Mosaic to New Raster tool. The original tiles were processed at a 1 m cell size; however, we resampled the tiles to a 9 m cell size using Nearest Neighbor resampling in order to produce a statewide mosaic of a manageable file size. We also merged in the open water data form the 2003 SAMB water body polygons. Any marsh or island was excluded. The assumption was made that open water cover had not changed significantly since 2003. This was necessary as we found that water could not be accurately delineated from the NAIP orthophotography using object-based image analysis techniques. There was much spectral and textural variability in this class, and it was often confused with shadows. We removed this class from the analysis so as not to induce error, and we relied on an ancillary dataset for this class. The final dataset contained forested/woody, grasslands/herbaceous, barren/non-vegetated, and open water as classifications.

We assessed the three classes obtained from object-based image analysis using a manual photograph interpretation process. The accuracy was assessed using randomly selected point locations and also majority class within a buffered area with a radius of 15 meters. The point data were assessed by Aaron Maxwell and Glenn Jansen, and the buffered data were assessed by Glenn Jansen and Elise Austin. The accuracy of the data was assessed by comparing the object-based classification results to manual interpretation at random point locations and an assessment of the majority class in a buffered area.

We create a point layer of random points throughout West Virginia using the software tool Geospatial Modeling Environment (GME). Using this same tool, we determined the resulting thematic classification at that cell. We then selected a random stratified subsample of 200 examples of each class within each LULC category that was derived from object-based image analysis. The water class was excluded. The interpreter then manually interpreted the cover conditions at the point location by manual photograph interpretation of the NAIP orthophotography.

For the error assessment based on the majority class within a buffered area, we first buffered the 50,000 points by 15 meters. We calculated the distribution of the thematic classification throughout each circular area. We then determined which circular areas were classified as at least 80% one cover type so that the interpreter could more easily interpret the dominant class. We then selected a random stratified subsample of 200 examples of each class within each LULC category that was derived from object-based image analysis. The water class was excluded. The interpreter then manually interpreted the majority cover condition of the area by manual photograph interpretation of the NAIP orthophotography.

The resulting error matrixes are shown in Tables 1 through 4. As can be seen from the assessment data, forest cover was extracted with a high degree of accuracy, both in terms of user’s accuracy and producer’s accuracy. The most confusion existed between the grasslands/herbaceous and barren/non-vegetated class. Error was not high for any class. We argue that the error assessment performed supports the conclusion that the thematic data are accurate for delineating basic cover types at a high spatial resolution. The resulting cover is shown in Figure 1. This data is a 9 m, categorical representation of forest cover and basic land cover statewide and relative to growing season 2011 conditions. For the delivery, we are providing the training data utilized, the forest cover, and the error assessment.

Statewide Forest Fragmentation:

We are providing a statewide forest fragmentation derived from the forest cover. First, the 9 m cover was reclassified to forest and non-forest. In order to provide a more general representation of fragmentation we smoothed the resulting cover in Feature Analyst 5 using an aggregation of 15 pixels. Any contiguous area of less than 15 pixels were removed and absorbed by the surrounding class. We did this to remove small canopy gaps or forest interruptions from the dataset. Such small areas were deemed inadequate to segment or fragment the forest and not worth considering in the analysis.

In order to complete the analysis, we made us of morphological image analysis. Morphological image processing uses mathematical morphology to analyze shape and form of objects. Forest pixels are classified as patch, edge, perforated, or core using this model. This method has been shown to provide a more accurate representation of fragmentation as the single pixel or landscape level when compared to image convolution (Vogt, 2007).

We made use of a forest fragmentation model after Vogt (2007) that can be obtained at the following URL: http://clear.uconn.edu/tools/lft/lft2/index.htm. This tool was produced by the Center for Land Use Education and Research at the University of Connecticut. The Landscape Fragmentation Tool (LFT) Version 2.0 was utilized, which functions as a Toolbox within ArcMap. We found that this tool could not process the entire state at a 9 m cell size due to file size limitations. As a result, we segmented the state into 25 regions which we then buffered by 10 km. This was necessary so as not to induce artificial edges into the final model. Once all sections were completed, we produced a negative 9 km buffer from each original buffer so as to remove edges from the final product. This resulted in 25 sections that overlapped by 1 km with adjacent regions. We then produced a 9 km statewide dataset using the Mosaic to New Raster tool in ArcMap. The result was inspected to ensure that no edges were induced by the segmentation process.

The only user defined parameter required for this tool is a defined edge width. The scientific literature, according to the creators of the tool, indicates that the width of edge-effects vary by the species or issue being studied. An appropriate edge-width for one species may be 50 meters whereas for a different species and appropriate edge-width may be several hundred meters. A 100 meter edge-width is commonly used for general purposes. For this analysis, we used an edge width of 100 meters because we were attempting a general purpose fragmentation layer for the state.

The final forest fragmentation model is shown in Figure 2. The state is differentiated into patch, edge, perforated, core (< 250 acres), core (250-500 acres), and core (>500 acres). This is a 9 m pixel size representation of forest fragmentation. Non-forest areas are defined as NoData in this raster dataset.

Statewide Wetland Probability Surface:

A statewide wetland probability surface was produced using the land cover results described above and maximum entropy probabilistic modeling in MAXENT, which can be obtained at the following URL: http://www.cs.princeton.edu/~schapire/maxent/. MAXENT was utilized to obtain the topographic probability of wetland occurrence while object-based image analysis was used to determine cover conditions.

We developed presents only training data, as required for maximum entropy modeling, from the National Wetland Inventory (NWI) dataset for West Virginia. This dataset has recently been edited at NRAC relative to recent color infrared orthophotography. We were only attempting to model palustrine emergent wetlands (PEM) and palustrine forested (PFO) wetlands /palustrine shrub/scrub (PSS) wetlands. As a result, we extracted only PEM, PFO, or PSS wetlands from the dataset. The polygon data were converted to raster data at a 10 m cell resolution. Each cell was then converted to a point feature. This resulted in a regular spacing of point features throughout wetland extents.

We segmented our modeling in the state by ecological subregions that can be obtained from the United States Department of Agriculture Forest Service at the following URL: http://fsgeodata.fs.fed.us/other_resources/ecosubregions.php. A total of 14 subregions intersect the state. For each subregion, we extracted the NWI wetland points that were created. Using GME, we collected a random selection of these points of 2,000 points for each of the 14 sub regions both for PEM wetlands and PFO/PSS wetlands. We created two separate models using these random points: one for PEM wetlands and one for PFO/PSS wetlands. These models represent the topographic likelihood of wetland occurrence.

We created 5 grids as predictor variables. All of these grids were derived from the 3 m cell size 2003 SAMB DEM data. However, the elevation data were resampled to 9 m so that a statewide processing could be completed. We derived slope in degrees and surface curvature using the Spatial Analyst Extension in ArcMap. We created a cost path from 1:4,800 scale streams weighted by slope using the Cost Distance tool within the Spatial Analyst Extension. This grid is an estimate of the friction moving away from a stream channel. We created ecological land units from the DEM data that segment the DEM into the following landforms: cliff, steep slope, slope crest, upper slope, flat summit, sideslope, cove, dry flat, moist flat, and slope bottom. This was the only categorical predictor variable used in the model. We also produced a moisture index using the following equation:

$$\ln[(\text{flow accumulation} + 1) / (\text{slope} + 1)]$$

This equation provides a relative measure of the level of moisture at a cell. We found that it was necessary to smooth the moisture index and the curvature grid data because they were noisy. To do this, we used the Focal Statistics tool in ArcMap with a circular neighborhood of 9 cells. The mean value was used.

These 5 grids were utilized because they offered estimators of topographic conditions that may influence wetland occurrence and because they could be calculated from the available DEM data.

Before running MAXENT, it was necessary to determine the cell value for each of the input grids at each randomly selected point location. This was done using GME. The attributed tables were exported, edited in Microsoft Excel, and saved as CSV files to use in the modeling. The grids were converted to ASCII text files.

Within MAXENT, we used a cumulative output format and we withheld a 25% random test percentage. As a result, the modeling was conducted using 1,500 points and 500 were withheld for model validation. We obtained response curves and jackknife plots to measure variable importance. The grid output was obtained in ASCII text format.

Each ecological subregion was buffered so that it overlapped with adjacent regions. We ran a total of 28 models, or two for each subregion. Once the models were produced, we used the Mosaic to New Raster tool in ArcMap to create statewide layers. One layer was produced for the topographic probability of wetland occurrence for PEM wetlands and a second for the topographic probability of wetland occurrence for PFO/PSS wetlands. A 9 m statewide probabilistic grid surface was created for each model. All response curves and jackknife plots are provided for each individual model. Figure 3 show the statewide output for PEM occurrence.

The MAXENT output was combined with the cover data produced from the 2011 NAIP orthophotography. PEM probability was only considered in grasslands/herbaceous areas and PFO/PSS probability was only considered in forested/woody area. This method was utilized to associate the topographic model with the appropriate cover conditions. Open water and barren/non-vegetated areas were considered inadequate for wetland occurrence. An example is shown in Figure 4.

Table 1: Aaron Maxwell: Manual photo interpretation at randomly selected points (Equal number of points in each class).

		Manual Interpretation				User's Accuracy
		Forested	Grasslands	Barren	Total	
Object-Based Classification	Forested	191	7	2	200	96%
	Grasslands	14	172	14	200	86%
	Barren	27	29	144	200	72%
	Total	232	208	160	600	
	Producer's Accuracy	82%	83%	90%		Overall Accuracy: 85% K_{Hat}: 77%

Table 2: Glenn Jansen: Manual photo interpretation at randomly selected points (Equal number of points in each class).

		Manual Interpretation				User's Accuracy
		Forested	Grasslands	Barren	Total	
Object-Based Classification	Forested	186	12	2	200	93%
	Grasslands	8	188	4	200	94%
	Barren	23	65	112	200	56%
	Total	217	265	118	600	
	Producer's Accuracy	86%	71%	95%		Overall Accuracy: 81% K_{Hat}: 72%

Table 3: Elise Austin: Manual photo interpretation at randomly selected points as majority class (Equal number of points in each class and buffered to a radius of 15 meters).

		Manual Interpretation				User's Accuracy
		Forested	Grasslands	Barren	Total	
Object-Based Classification	Forested	194	5	0	199	97%
	Grasslands	2	195	3	200	98%
	Barren	2	24	174	200	87%
	Total	198	224	177	599	
	Producer's Accuracy	98%	87%	98%		Overall Accuracy: 94% K_{Hat}: 91%

Table 4: Glenn Jansen: Manual photo interpretation at randomly selected points as majority class (Equal number of points in Each Class and buffered to a radius of 15 meters).

		Manual Interpretation				User's Accuracy
		Forested	Grasslands	Barren	Total	
Object-Based Classification	Forested	196	3	1	200	98%
	Grasslands	1	196	3	200	98%
	Barren	2	21	177	200	89%
	Total	199	220	181	600	
	Producer's Accuracy	98%	89%	98%		Overall Accuracy: 95% K_{Hat}: 92%

Figure 1: Forest cover for West Virginia representing growing season 2011 conditions and derived from NAIP orthophotography.

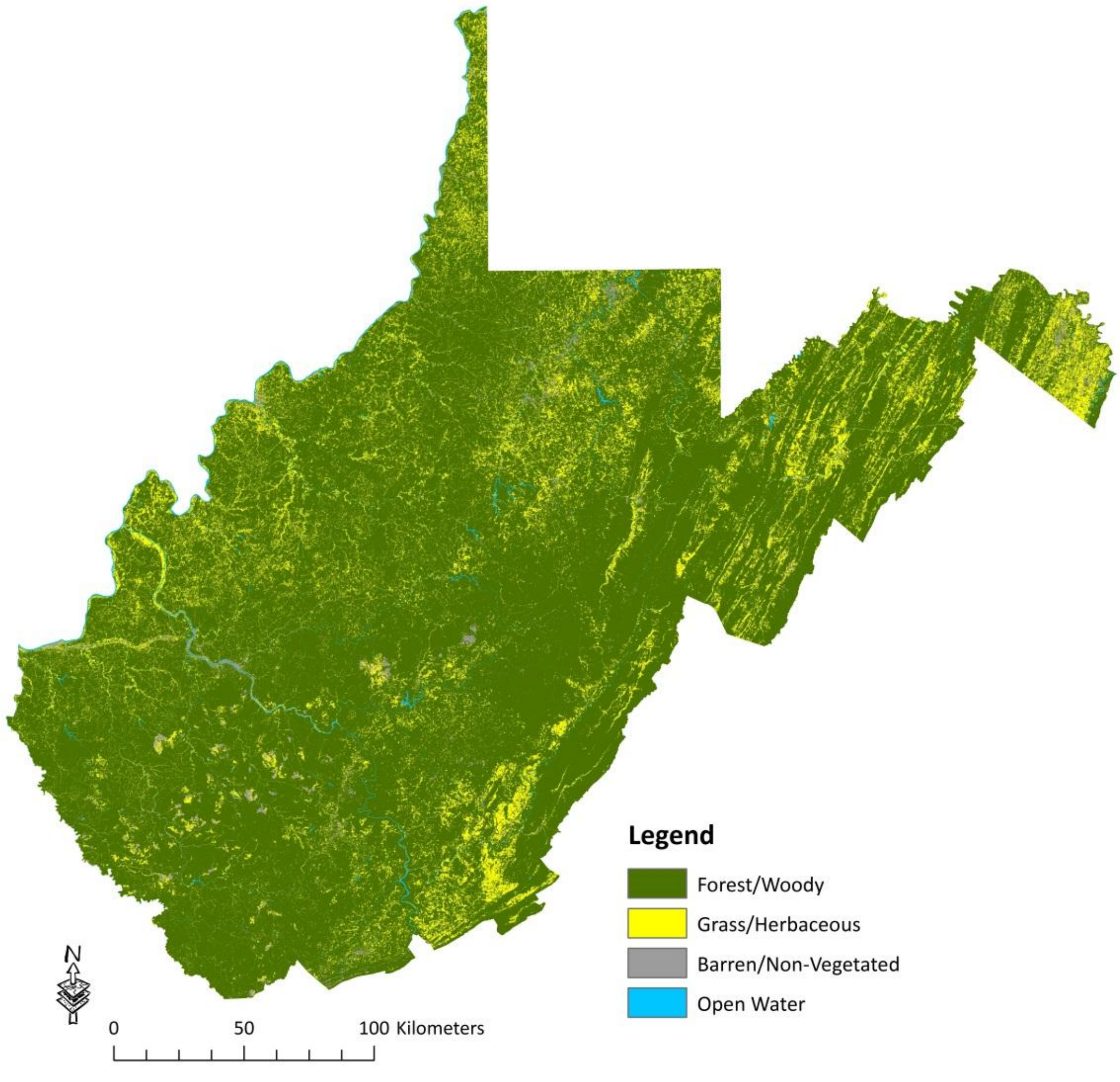

Figure 2: Forest fragmentation derived from morphological image analysis representing growing season 2011 conditions.

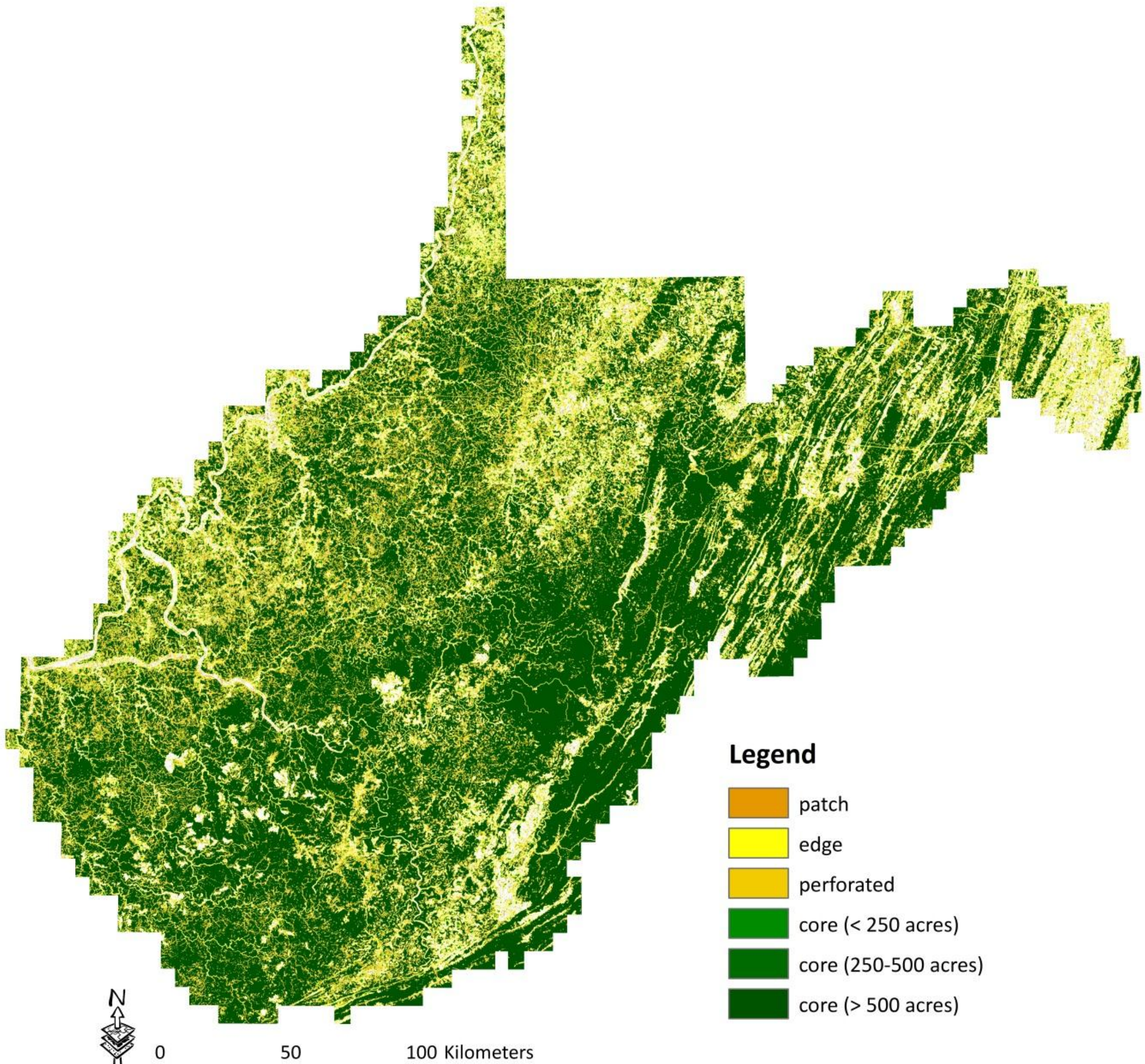

Figure 3: MAXENT Output: Statewide topographic probability of PEM occurrence.

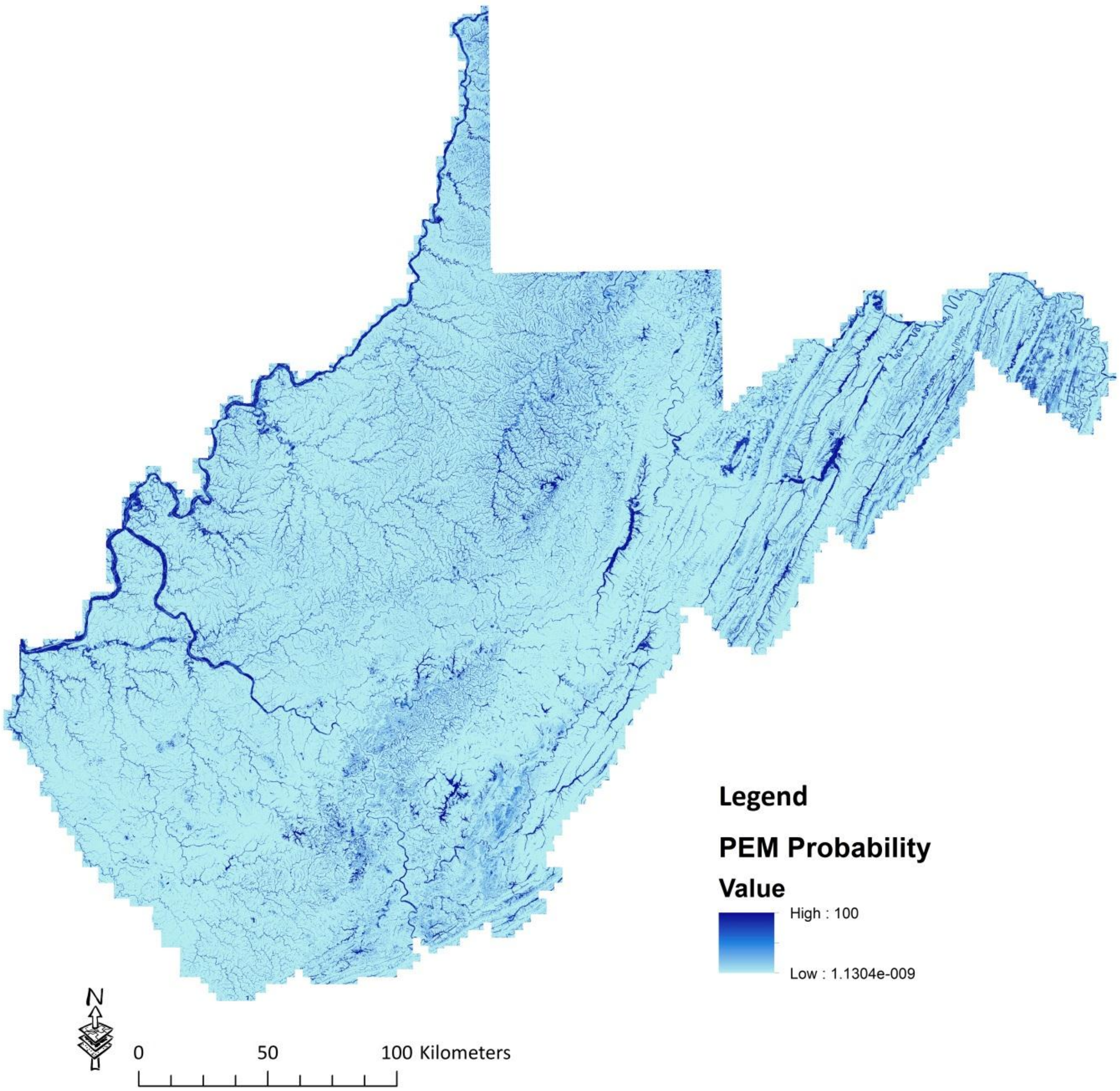

Figure 4: Probability of PEM wetlands in herbaceous areas.