Genetic algorithm-based method for forest type classification using multi-temporal NDVI from Landsat TM imagery

Hong Tao\textsuperscript{a,b,c}, Manqi Li\textsuperscript{d}, Ming Wang\textsuperscript{a,b,c} and Guonian Li\textsuperscript{a,b,c}

\textsuperscript{a}Ministry of Education, Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Nanjing, China; \textsuperscript{b}State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing Normal University, Nanjing, China; \textsuperscript{c}Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing Normal University, Nanjing, China; \textsuperscript{d}Department of Geography and Geoinformation Science, George Mason University, Fairfax, VA, USA

1. Introduction

Forest type composition provides valuable information for forest inventory and management, thus it is increasingly desirable to obtain the spatially explicit forest type distribution for a range of applications such as forest monitoring (Lehmann et al. 2015; Liu, Im, and Quackenbush 2015; Sankey et al. 2017), natural resources (Corona 2016; Li et al. 2014), biodiversity and ecosystems (Barlow et al. 2016; Laurin et al. 2014; Van der Plas et al. 2018) and climate change (Charney et al. 2016; Asner et al. 2017). Traditional forest mapping methods such as field-based investigation are usually limited by the great demands for time and workload, as well as the difficulty to conduct over large geographic areas and precipitous topography. Advanced remote sensing technology offers a unique capability of large-scale forest mapping in a cost-effective manner (Gudex-Cross, Pontius, and Adams 2017; Khatami, Mountrakis, and Stehman 2016; Zhen, Quackenbush, and Zhang 2016).

The usefulness of multi-temporal Landsat imagery for forest species/type classification has been extensively explored, and favorable performances have been achieved at varied geographical scales (Gudex-Cross, Pontius, and Adams 2017; Li, Im, and Beier 2013; Zhu and Woodcock 2014). In the Landsat family, the Landsat Thematic Mapper (TM) data have been proved ideal for mapping forest type composition at local and regional scales, considering its moderate spatial resolution (30 m) and temporal frequency (16 days) (NASA 2018), as well as the low cost and wide availability (Mancino et al. 2014; Foody and Hill 1996; Li et al. 2018).

Existing methods for forest types mapping using Landsat TM data range from statistical-based classifiers such as clustering (Lo and Choi 2004) and maximum likelihood (Hagner and Reese 2007), to machine learning algorithms such as neural network (Kempeneers et al. 2011), decision tree (Sesnie et al. 2008), random forest (Zhu and Woodcock 2014) and support vector machine (Li, Im, and Beier 2013). Comprehensive comparisons of algorithm performances have been made, but so far no agreement has been reached regarding the most appropriate classifiers for any given task (Phiri and Morgenroth 2017).

Nowadays there is growing interest in machine learning approaches fusion, aiming at improving classification accuracy. Progress has been made in various studies; for example Gong, Im, and Mountrakis (2011) coupled genetic algorithm and artificial immune networks to improve land cover classification, and Chatterjee et al. (2016) employed genetic algorithm to optimize the input weight vector of a neural network classifier. However, enhancements of
these hybrid machine learning approaches are somehow limited, and at the cost of algorithmic complexity (Chatterjee et al. 2016). In addition, an evident limitation of the above-reviewed classifiers lies in the high dimension of input features, which could result in decreasing classification accuracy due to the Hughes effect (Zhu and Liu 2014). A common solution is to employ feature selection that keeps the most informative features (Zhu and Liu 2014; Pal and Foody 2010).

Genetic algorithms (GAs) are a family of computational models motivated by the process of natural selection in evolutionary biological systems (Mitchell 1998). As one of the most popular heuristic algorithms for optimization, GA shows its advantage of avoiding the convergence to a locally optimal solution, which is easily incurred in other algorithms such as hill climbing (Yuret and Maza 1993). Its ability to handle the search space in many directions simultaneously has greatly increased the probability of finding a global optimum (Gen and Cheng 2000). Thus it has been reported as a robust and efficient optimization technique in a wide range of applications, and has been commonly utilized in parameter optimization, optimum threshold identification and feature selection in classification tasks (Chatterjee et al. 2016; Gong, Im, and Moutrakis 2011; Van Coillie, Verbeke, and De Wulf 2007). In this study, we are interested in examining its utility in detecting optimum thresholds as well as useful features for forest type classification. We expect a simple and low-demanding method that minimizes the data inputs and algorithmic complexity, while still produces satisfactory results.

Normalized Difference Vegetation Index (NDVI) was found highly correlated with vegetative characteristics such as vegetation cover (Huang et al. 2017), leaf area index (Xavier and Vettorazzi 2004), biomass and carbon stock (Liu, Su et al., 2015) and crown density (Zaitunah, Samsuri, and Safitri 2018). Its weak sensitivity to atmospheric and topographic effects enabled consistent performance in multi-temporal analysis and scene-to-scene comparisons (Krishnaswamy, Kiran, and Ganeshiaah 2004). In the past few decades, much emphasis has been placed on using multi-temporal NDVI as the major indicator for vegetation mapping because of its seasonal distinction (Wang and Tenhunen 2004). Krishnaswamy, Kiran, and Ganeshiaah (2004) accomplished vegetation cover mapping using multi-season NDVI data in diverse tropical deciduous ecosystems. Mancino et al. (2014) successfully detected the natural forest expansion area via Landsat TM imagery and NDVI differencing techniques. Smoothed NDVI data were evaluated regarding its potential of distinguishing various land cover types in a vegetated area in Shao et al. (2016). Therefore considering the strong predicting ability in vegetation mapping of NDVI, multi-temporal NDVI bands are chosen as the predicting variables for forest type classification in this study.

The objectives of this research include (1) separating local forest communities at the Huntington Wildlife Forest (HWF) into deciduous, mixed/coniferous and non-forests types using multi-temporal Landsat-derived NDVI, (2) assessing the effectiveness of the developed GA-based method for forest type classification when using NDVI as the sole predictor and (3) providing interpretations of forest cover conditions over HWF for long-term continuous inventory and management.

2. Study area and data preparation

2.1. Study area and data

Located in the central Adirondack Park in New York State of the United States, the Huntington Wildlife Forest is a 6,000-ha research forest with a Continuous Forest Inventory (CFI) system, which has been updated with new measurements every 10 years since 1950. According to the manual of 1991 Continuous Forest Inventory Measurement Procedures for Huntington Wildlife Forest, there are 285 CFI plots in this study site, and each plot has a radius of about 17 m, suggesting the area of a plot (908 m²) approximates to that of one pixel in Landsat TM images (900 m²). Therefore, it is reasonable to represent a plot with a pixel in the Landsat imagery. HWF is an ideal site for remote-sensing studies because of the availability of forest inventory and environmental monitoring data on-site. Figure 1 demonstrates the location and extent of the study area along with the 285 CFI plots.

Two types of datasets are collected for this study: Landsat TM images (Path 14, Row 29) covering the HWF area and CFI data of each study plot located in this site. Six Landsat images of leaf-on and leaf-off seasons (i.e. images acquired on 20th May, 8th August and 9th September for leaf-on season, and on 28th January, 17th March and 30th December for leaf-off season) in 1991 are downloaded from the website of United States Geological Survey (USGS, https://glovis.usgs.gov/app). And the red and near-infrared bands of these images are of particular interest. The CFI data on HWF include a manual instructing the CFI measurement procedures, a GIS shapefile containing the geographical location of the study plots, and a spreadsheet recording the in situ measurements obtained in 1991.

In the in situ measurement records, tree species of each plot is determined by the surveyors through visual inspection of its dominant and codominant canopy, and is used as the reference data. Summarizing the tree species, three major forest types are identified: northern hardwood, hardwood/conifer and conifer. Major hardwood include sugar maple, red maple and yellow birch; common conifers are
red spruce, balsam fir, eastern hemlock and white pine (Demers and Breitmeyer 2008). For practical concern, which is to balance the sample size of the forest cover types, in this study tree species are categorized into deciduous and mixed/coniferous, and the nonforested plots are assigned nonforests (Table 1).

### 2.2. Data preprocessing

Preliminary processing steps were performed upon the acquired data. First using the extent of the GIS shapefile of CFI plots, the original Landsat TM images were masked for the exact study area. Then the spectral reflectance values of visible red (R) and near-infrared (NIR) bands were converted from the digital numbers of the original TM images. NDVI, the normalized difference of near-infrared and visible red reflectance values were then calculated as follows:

\[
\text{NDVI} = \frac{\text{NIR} - \text{R}}{\text{NIR} + \text{R}}
\]

The above equation indicates that NDVI for a given pixel always ranges from minus one (−1) to plus one (+1); and nonvegetated pixel produces a value close to zero. Negative values are indicative of clouds, snow, water and other nonvegetated, nonreflective surfaces, while positive values imply vegetated or reflective surfaces (Burgan and Hartford 1993).

Layers containing the NDVI value for every single pixel were generated for the 6 months chosen, and then by referring to the locations of the study plots, pixel values on these NDVI layers were extracted. These NDVI values at each plot were treated as the predicting variables for forest type classification in this study.

The proportions of deciduous, mixed/coniferous and nonforests are 63.16%, 35.09% and 1.75%, respectively. Considering such uneven distribution among different classes, equal percentage allocation on these forest types were applied for dividing subdatasets for calibration and validation. In this way, the original data were split into two datasets using stratified random sampling, keeping the ratio of sample counts for calibration and validation 2:1. The allocation of the two subdatasets is summarized in Table 2.
Table 2. Plot number allocation for the calibration and validation datasets.

| Dataset | Deciduous | Mixed/coniferous | Nonforests | Total |
|---------|-----------|-----------------|------------|-------|
| Calibration | 120       | 65              | 3          | 188   |
| Validation  | 60        | 35              | 2          | 97    |
| Total      | 180       | 100             | 5          | 285   |

(63.16%) (35.09%) (1.75%) (100%)

2.3. Feature selection

Feature selection aims at choosing a small subset of features that ‘ideally is necessary and sufficient to describe the concept, to speed up learning, and to improve concept quality’ (Kira and Rendell 1992). Thus, for optimized performance, best combination of NDVI values from the 6 months as well as the sequence of these variables when importing to the model were sought by exploiting statistical approaches. Boxplots, Analysis of Variance (ANOVA) and Canonical Discriminant Analysis (CDA) were utilized in this process. Boxplots visually depict the variables through their quartiles; ANOVA analyzes the differences among variable means and CDA augments the understanding of the contributions from each variable to the canonical components.

3. GA modeling

3.1. Conceptual model description

Normally a GA model consists of four stages: Initialization, Selection, Crossover and Mutation. The core of GA is to search for an enhanced fitness of a population of attempted solutions toward a convergence at the global optimum (Tseng et al. 2008).

In the Initialization stage, a population of attempted solutions, called chromosomes, which are randomly generated in the solution space, is selected as the starting point of the search. In this study, one chromosome is represented by the array of NDVI values of the month(s) that are used for searching. After Initialization, each chromosome is evaluated by a user-defined fitness function. The role of the fitness function is to assess the performance of the chromosome as a solution for the targeted problem (McCall 2005).

In the Selection phase, a common rule of selecting parents for reproduction is that only the best members will be retained and will pass their characteristics from generation to generation. This selective process encourages the best-performing chromosomes to contribute the most to the next generation over time. Then Crossover and Mutation steps are applied to create new generations. The Crossover operation produces a new offspring from two randomly selected ‘good parents’, while Mutation randomly chooses the chromosomes in the population, and arbitrarily changes a subset of the genes on the chromosomes.

Briefly, in a GA, a population of potential solutions evolves over successive generations using a set of genetic operators. Figure 2 demonstrates the structure of a basic GA model.

3.2. Genetic algorithms model

In this study, A GA model is trained to classify the vegetated area into deciduous and mixed/coniferous. In this model, various combinations of NDVI values in different months are tested to find the best fitness. Figure 3 displays a flow diagram of the key phases in the proposed GA model. In this model, genes on the chromosomes refer to the NDVI variables. The fitness function for determining the ability of a particular chromosome to survive is represented by the classification accuracy in the calibration process. The number of generations that the chromosomes undergo, and number of population of the chromosomes can be arbitrarily predefined or based on a threshold of fitness. For this study, these two parameters are set as static numbers of 50 as maximum, for the sake of simplicity and efficiency.

In Selection phase, fitness for the formed chromosomes is firstly calculated. The chromosomes with the highest five fitness values are excluded from the processes of regeneration and directly proceed to the next generation, while others will be regenerated aiming for better replacements. Crossover and Mutation are controlled by random processes that depend on crossover and mutation rate. Crossover simulates the sexual reproduction while Mutation imitates asexual reproduction. Considering their practical meanings, we chose a crossover rate of 0.8 and mutation rate of 0.2, consistent with tggghose in Im, Lu, and Jensen (2011). When a chromosome is selected for Crossover, it switches certain genes with others on chosen chromosomes by predefined rules, which depends on the season of NDVI. If a chromosome is picked for Mutation, certain genes are replaced by either random values or a percentage change of their original values. In the Mutation stage chromosome with the highest fitness is exempt from this operation, ensuring that the best set of genes is kept until it is surpassed. This design enables the GA model to produce thresholds of NDVI bands corresponding to the highest classification accuracy.

4. Results and discussion

Results of this study consist of four components. First, thresholds for separating the forested area and nonforests are found by statistical analysis. Second, thresholds for separating deciduous and mixed/coniferous are generated from the developed GA model. Third, overall performance is evaluated by combining the results of
the above two-step separations, which is referred to as the GA-based classification method. Further, the entire study area is classified into deciduous, mixed/coniferous and nonforests using the best fitted thresholds, which produces a classification map showing the geographical distribution of forest types of HWF in 1991.

### 4.1. Statistical analysis

For uncovering the distributions of the NDVI values of each month, boxplots are drawn using the 120 deciduous plots, 65 mixed/coniferous plots and three plots of nonforests in the calibration dataset. These boxplots are illustrated in Figure 4, from which general interpretations arise on the range of NDVI values across the 6 months, and on the distinction of NDVI values among the three types within each month. These understandings help to find the thresholds for extracting the nonforests category, and to decide the combinations of the six variables as the input data in the following searching step in the GA model.

In the graphs of May and September, the occurrence of clear distinctions between nonforests and the other two types (deciduous and mixed/coniferous) suggests the usefulness of NDVI in these 2 months for extracting the nonforests type from the whole dataset. However, a few outliers exist in the deciduous category that might influence the accuracy. The thresholds derived from NDVI bands of May and September are shown in Table 3. Another important observation from the graphs is that the NDVI of deciduous are lower than that of mixed/coniferous in leaf-off seasons, while higher in leaf-on seasons. Therefore, in the GA model, only two kinds of crossover scenarios are considered since the season a month belongs to is known, which greatly simplifies the model (see the Crossover stage in Figure 3).

Results from ANOVA suggest a rejection of the hypothesis that the mean NDVI of the 6 months are equal. This rejection means that the tested data are of little correlation with each other, and further supports that all the six NDVI sets are suitable for being the predicting variables in the subsequent classification. In this study, CDA is used for discovering the importance
of the variables. For this purpose, canonical coefficients for the first and second canonical components are inspected: the larger the absolute value of the coefficient, the more contributions it makes to the corresponding canonical component; in other words, the better the variable is to distinguish different classes. Thus, the canonical coefficients lead to a ranking list of the months showing their importance from high to low: March, December, August, September, January and May. This list assists to determine the combinations of months as inputs to the GA model.

4.2. Genetic algorithms model

Without using any prior knowledge, first for each month, its NDVI is imported to the GA model individually, and the threshold along with its corresponding accuracy is gained from the model. Since this model executes heuristic search, five trials are performed for each month, and the mean value of the thresholds corresponding to the highest accuracy is considered as the threshold for that month. Results of using single months as inputs are shown in Table 4.

The rank of the months offering accuracy from high to low is: March, December, August, January, September and May. This list is almost the same as that from CDA, except for the switched places of January and September. Thus in the next step, March is used as the first input, and then the rest of the months are added to the model one by one according to the ranking lists given by CDA and GA modeling with single months, to search for the highest accuracy with a combination of the variables. In this process, we also run the model for five times, and then find the best threshold of each month in the combinations

*CR: Crossover Rate
*MR: Mutation Rate
that produce the highest accuracy. Table 5 shows the results of this process. It suggests that the highest accuracy (89.73%) occurs when the combination of NDVI values in March and August are used, and the thresholds are 0.1451 and 0.7349, respectively. Therefore, for classifying the forested area into deciduous and mixed/coniferous, NDVI values in March and August are chosen by the GA model.

Summarizing the results from statistical analysis and GA modeling, thresholds of NDVI for classifying the three forest types are found, providing an optimized solution for the classification (Table 6). For separating forested area (i.e. deciduous and mixed/coniferous) and nonforests, NDVI values of May and September are used. Specifically, plots with NDVI values of May larger than 0.45 and NDVI of September larger than 0.35 are classified as forested area.

Table 3. Thresholds for separating nonforests and forested area.

| Forest type    | Nonforests | Forested area |
|----------------|------------|---------------|
| NDVI of May    | NDVI < 0.45| NDVI ≥ 0.45   |
| NDVI of September | NDVI < 0.35| NDVI ≥ 0.35   |

Figure 4. Boxplots of NDVI values of the selected months in 1991. (a)–(c) for leaf-off season, and (d)–(f) for leaf-on season.

Table 4. Thresholds and highest accuracy using single months.

| Generation | Population | CR* | MR** | 1 | 2 | 3 | 4 | 5 | Mean*** | Rank |
|------------|------------|-----|------|---|---|---|---|---|---------|------|
| 50         | 50         | 0.8 | 0.2  |   |   |   |   |   |         |      |
| January    | Threshold  | 0.1173 | 0.1188 | 0.1196 | 0.1164 | 0.1184 | 0.1181 | 4 |
|            | accuracy (%) | 78.38 | 78.38 | 78.38 | 78.38 | 78.38 | 78.38 |   |
| March      | Threshold  | 0.1445 | 0.1452 | 0.1473 | 0.1439 | 0.1435 | 0.1449 | 1 |
|            | accuracy (%) | 88.65 | 88.65 | 88.65 | 88.65 | 88.65 | 88.65 |   |
| May        | Threshold  | 0.6212 | 0.6139 | 0.6087 | 0.6204 | 0.6201 | 0.6202 | 6 |
|            | accuracy (%) | 76.22 | 75.68 | 76.22 | 76.76 | 76.76 | 76.76 |   |
| August     | Threshold  | 0.7073 | 0.7089 | 0.7096 | 0.7094 | 0.7077 | 0.7086 | 3 |
|            | accuracy (%) | 83.24 | 83.24 | 83.24 | 83.24 | 83.24 | 83.24 |   |
| September  | Threshold  | 0.5025 | 0.5052 | 0.5020 | 0.5050 | 0.5039 | 0.5037 | 5 |
|            | accuracy (%) | 77.84 | 77.84 | 77.84 | 77.84 | 77.84 | 77.84 |   |
| December   | Threshold  | 0.1407 | 0.1397 | 0.1403 | 0.1410 | 0.1415 | 0.1407 | 2 |
|            | accuracy (%) | 84.86 | 84.86 | 84.86 | 84.86 | 84.86 | 84.86 |   |

*CR: Crossover rate
**MR: Mutation rate
***Mean value of the thresholds corresponding to the highest accuracy (in bold) is chosen as the threshold for that month.
categorized as forested area, and the rest of the plots are regarded as nonforests. For distinguishing deciduous and mixed/coniferous, NDVI values of March and August are employed. Thresholds for these 2 months are 0.1451 and 0.7349, respectively.

It is suggested that for identifying nonforests from the forested area, NDVI in leaf-on season is more useful than that of leaf-off season. Because in leaf-off season, all trees grow leaves that causes the NDVI values generally high in the vegetated area, while the NDVI values in the nonforested area stay low. This enhanced discrepancy of NDVI promotes the capability to separate the forested and nonforested area by simple thresholding. As for discriminating deciduous and mixed/coniferous, combined NDVI in both leaf-on and leaf-off seasons are necessary. This is supported by the NDVI distributions in Figure 4, where in leaf-off season, NDVI of deciduous is generally lower than that of mixed/coniferous, and in leaf-on season the situation is exactly the opposite.

### 4.3. GA-based method evaluation

The proposed GA-based method contains two steps, one for extracting nonforests type from the whole data using statistical approaches, and one for separating deciduous and mixed/coniferous using a GA model. This classification method is evaluated using the prepared validation dataset. Applying the thresholds of NDVI in March, August, May, and September in Table 6 to the validation dataset, the 97 CFI plots are classified into deciduous, mixed/coniferous and nonforests.

Comparing the results with reference data, a matrix showing classification results and accuracy statistics is created (Table 7), providing producer’s accuracy and user’s accuracy of each category, as well as overall accuracy and kappa coefficient. The overall accuracy indicates that 80.41% plots are correctly classified. The kappa coefficient (0.56) suggests a moderate agreement between classification results and reference information. The accuracy of the deciduous type is the highest: 58 out of 60 deciduous plots are correctly classified. As for mixed/coniferous, only 54.29% of the plots are classified correctly, and 16 plots are misclassified into the deciduous class. For the two validation plots of nonforests, the 50% accuracy means one plot is classified correctly and the other one is placed to another class (deciduous). The low accuracy of nonforests is greatly attributed to its small sample size. To offset this deficiency, we may manually add additional plots that locate in the open field or lakes, where visual interpretation of the Landsat images is sufficient for this action.

### 4.4. Forest type classification

A classified map containing deciduous, mixed/coniferous and nonforests types over the study area is produced by utilizing the proposed GA-based classification method, which is shown in Figure 5. In the figure, detailed distribution of the three types throughout the entire area is attained. Forested area covers more than 90% of the map, and deciduous and mixed/coniferous spread through the entire site, mostly aggregated by type. Large area

| Classification data | Deciduous | Mixed/coniferous | Nonforests | Row total | User’s accuracy |
|---------------------|-----------|------------------|------------|-----------|----------------|
| Deciduous           | 58        | 16               | 1          | 75        | 77.33%         |
| Mixed/coniferous    | 2         | 19               | 0          | 21        | 90.48%         |
| Nonforests          | 0         | 0                | 1          | 1         | 100.00%        |
| Column total        | 60        | 35               | 2          | 97        | 100.00%        |
| Producer’s accuracy | 96.67%    | 54.29%           | 50.00%     |           |                |
| Overall accuracy    | (58 + 19 + 1)/97 = 80.41% |                |            |           |                |
| Kappa coefficient*  | 0.56      |                  |            |           |                |

*Degree of agreement (Landis and Koch 1977): 0.81–1.00 as ‘almost perfect’, 0.61–0.80 as ‘substantial’, 0.41–0.60 as ‘moderate’, 0.21–0.40 as ‘fair’, 0.01–0.20 as ‘slight’ and <0.00 as ‘poor/none’.

| Table 5. Thresholds and accuracy with different combinations of variables. |
|-----------------------------|-----------------------------|-----------------------------|
| Threshold                  | March | December | August | September | January | May | Highest accuracy (%) |
| 0.1449                     | 0.1465 | 0.0219   | 0.1451 | 0.0616    | 0.0783 | 0.1354 | 0.1102 | 0.1393 | 88.65 |
| 0.1465                     | 0.0783 | 0.1354   | 0.107 | 0.0954    | 0.0620 | 0.0835 | 0.0313 | 0.6961 | 89.73 |
| 0.1451                     | 0.1354 | 0.0835   | 0.1716 | 0.7283    | 0.0210 | 0.7778 | 0.6961 | 0.6777 | 89.73 |
| 0.0783                     | 0.0313 | 0.6961   | 0.6777 | 0.0035    | 0.6499 | 77.30 |

| Table 6. Thresholds of NDVI for the highest classification accuracy. |
|-----------------------------|-----------------------------|-----------------------------|
| Forest type                 | March | August | May | September |
| Deciduous                   | <0.1451 | >0.7349 | ≥0.45 | ≥0.35 |
| Mixed/coniferous            | ≥0.1451 | ≤0.7349 |        |
| Nonforests                  |        | <0.45  | <0.35 |        |

| Table 7. Confusion matrix and accuracy measures of the GA-based classification method. |
|---------------------------------|-----------------------------|-----------------------------|
| Classification data             | Deciduous | Mixed/coniferous | Nonforests | Row total | User’s accuracy |
| Deciduous                       | 58        | 16               | 1          | 75        | 77.33%         |
| Mixed/coniferous                | 2         | 19               | 0          | 21        | 90.48%         |
| Nonforests                      | 0         | 0                | 1          | 1         | 100.00%        |
| Column total                    | 60        | 35               | 2          | 97        | 100.00%        |
| Producer’s accuracy             | 96.67%    | 54.29%           | 50.00%     |           |                |
| Overall accuracy                | (58 + 19 + 1)/97 = 80.41% |                |            |           |                |
| Kappa coefficient*             | 0.56      |                  |            |           |                |
belonging to nonforests are lakes and open fields. By checking the Digital Elevation Model (DEM) covering this area, we observed that mixed/coniferous mostly grew in higher elevations, and low-lying area is generally dominated by deciduous type. Area and proportion of each forest type are calculated and exhibited in Table 8. These results demonstrate that the mixed/coniferous is the dominant category over the study area with a proportion of 64.72%, and there are 28.09% area covered by the deciduous. In addition, 7.19% of the area belong to the nonforests type. These interpretations have provided valuable information of the forest cover conditions over HWF in the year 1991, and will contribute to long-term continuous inventory and forest management in this area.

5. Conclusion

In this study, a genetic algorithm-based method was proposed for classifying the Huntington Wildlife Forest into deciduous, mixed/coniferous, and nonforests types, using multi-temporal NDVI derived from the Landsat TM imagery. Statistical approaches such as ANOVA and canonical discriminant analysis were utilized for priori knowledge of the NDVI variables, facilitating subsequent decisions on selecting different groups of predicting variables for model calibration. Results show that the proposed method successfully achieved a favorable overall classification accuracy of 80.41%, and the kappa coefficient of 0.56 indicates moderate agreement between the classification results and reference. As the sole type of input variable, multi-temporal NDVI proved its capability of separating the forest types as well as extracting nonforested area effectively. In addition, a map displaying the geographical distributions of the three forest types of the entire study area was produced, to promote our understanding of the studied forest and to assist its future monitoring and management. Future work include applying the developed method to the same study area at 10 years’ interval, taking advantage of the decennial forest inventory system of HWF. Also, there are potential improvements for the GA-based method in its flexibility and robustness, for example testing other forms of variables.

Table 8. Area and proportion of each forest type of the study area.

| Forest type      | Deciduous | Mixed/coniferous | Nonforests | Total     |
|------------------|-----------|------------------|------------|-----------|
| Area (m²)        | 39,883,500 | 91,890,900       | 10,215,000 | 141,989,400 |
| Proportion       | 28.09%    | 64.72%           | 7.19%      | 100.00%    |

Figure 5. A map of the study area composed of deciduous, mixed/coniferous, and nonforests types.

Acknowledgments

The authors would like to thank the State University of New York, College of Environmental Science and Forestry for providing in situ measurements of the study site. Work done
in this paper was supported by the Priority Academic Program Development of Jiangsu Higher Education Institutions [grant number 164320H116].

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by [Priority Academic Program Development of Jiangsu Higher Education Institutions]; [164320H116].

ORCID

Manqi Li http://orcid.org/0000-0002-5206-0369

References

Asner, G. P., R. E. Martin, D. E. Knapp, R. Tupayachi, C. B. Anderson, F. Sinca, N. R. Vaughn, and W. Llactayo. 2017. “Airborne Laser-Guided Imaging Spectroscopy to Map Forest Trait Diversity and Guide Conservation.” Science 355 (6323): 385–389. doi:10.1126/science.aaj1987.

Barlow, J., G. D. Lennox, J. Ferreira, E. Berenguer, A. C. Lees, R. M. Nally, J. R. Thomson, S. F. de Barros Ferraz, J. Louzada, V. H. F. Oliveira, and L. Parry. 2016. “Anthropogenic Disturbance in Tropical Forests Can Diversify Biodiversity Loss from Deforestation.” Nature 535 (7610): 144–147. doi:10.1038/nature18326.

Burgan, R. E., and R. A. Hartford. 1993. “Monitoring Vegetation Greenness with Satellite Data.” Gen. Tech. Rep. INT-GTR-297. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 13: 297. doi:10.2737/INT-GTR-297.

Charney, N. D., F. Babst, B. Poulter, Valerie M. Trouet, D. Frank, B. J. Enquist, and M. E. K. Evans. 2016. “Observed Forest Sensitivity to Climate Implies Large Changes in 21st Century North American Forest Growth.” Ecology Letters 19 (9): 1119–1128. doi:10.1111/ele.12650.

Chatterjee, S., S. Ghosh, S. Dawn, S. Hore, and N. Dey. 2016. “Forest Type Classification: A Hybrid NN-GA Model Based Approach.” In Information Systems Design and Intelligent Applications, edited by S. C. Satapathy, J. K. Mandal, S. K. Udgata, and V. Bhatje. 227–236. India: Springer. Advances in Intelligent Systems and Computing.

Corona, P. 2016. “Consolidating New Paradigms in Large-Scale Monitoring and Assessment of Forest Ecosystems.” Environmental Research, the Provision of Ecosystem Services in Response to Global Change 144 (Jan.): 8–14. doi:10.1016/j. envres.2015.10.017.

Demers, C., and B. Breitmeyer. 2008. Huntington Wildlife Forest Natural Resources Management Plan. New York: State University of New York College of Environmental Science and Forestry Newcomb. https://www.esf.edu/aec/documents/HWF-NaturalResourcesPlan.pdf.

Foody, G. M., and R. A. Hill. 1996. “Classification of Tropical Forest Classes from Landsat TM Data.” International Journal of Remote Sensing 17 (12): 2353–2367. doi:10.1080/01431169608948777.

Gen, M., and R. Cheng. 2000. Genetic Algorithms and Engineering Optimization. New York: John Wiley & Sons.

Gong, B., J. Im, and G. Mountrakis. 2011. “An Artificial Immune Network Approach to Multi-Sensor Land Use/Land Cover Classification.” Remote Sensing of Environment 115 (2): 600–614. doi:10.1016/j.rse.2010.10.005.

Gudex-Cross, D., J. Pontius, and A. Adams. 2017. “Enhanced Forest Cover Mapping Using Spectral Unmixing and Object-Based Classification of Multi-Temporal Landsat Imagery.” Remote Sensing of Environment 196 (Jul.): 193–204. doi:10.1016/j.rse.2017.05.006.

Hagner, O., and H. Reese. 2007. “A Method for Calibrated Maximum Likelihood Classification of Forest Types.” Remote Sensing of Environment, ForestSAT Special Issue 110 (4): 438–444. doi:10.1016/j.rse.2006.08.017.

Huang, S., C. Ramirez, K. Kennedy, J. Mallory, J. Wang, and C. Chu. 2017. “Updating Land Cover Automatically Based on Change Detection Using Satellite Images: Case Study of National Forests in Southern California.” GIScience & Remote Sensing 54 (4): 495–514. doi:10.1080/15481603.2017.1286727.

Im, J., L. Zhenyu, and J. R. Jensen. 2011. “A Genetic Algorithm Approach to Moving Threshold Optimization for Binary Change Detection.” Text. February. doi:10.14358/PERS.77.2.167.

Kempeneers, P., F. Sedano, L. Seebach, P. Strobl, and J. San-Miguel-Ayanz. 2011. “Data Fusion of Different Spatial Resolution Remote Sensing Images Applied to Forest-Type Mapping.” IEEE Transactions on Geoscience and Remote Sensing 49 (12): 4977–4986. doi:10.1109/TGRS.2011.2158548.

Khatri, R., G. Mountrakis, and S. V. Stehman. 2016. “A Meta-Analysis of Remote Sensing Research on Supervised Pixel-Based Land-Cover Image Classification Processes: General Guidelines for Practitioners and Future Research.” Remote Sensing of Environment 177 (May): 89–100. doi:10.1016/j.rse.2016.02.028.

Kira, K., and L. A. Rendell. 1992. “The Feature Selection Problem: Traditional Methods and a New Algorithm.” Aaai 2: 129–134.

Krishnaswamy, J., M. C. Kiran, and K. N. Ganeshiaha. 2004. “Tree Model Based Eco-Climatic Vegetation Classification and Fuzzy Mapping in Diverse Tropical Deciduous Ecosystems Using Multi-Season NDVI.” International Journal of Remote Sensing 25 (6): 1185–1205. doi:10.1080/0143116031000149989.

Landis, J. R., and G. G. Koch. 1977. “The Measurement of Observer Agreement for Categorical Data.” Biometrics 33 (1): 159–174. doi:10.2307/2529310.

Laurin, G. V., J. C.-W. Chan, Q. Chen, J. A. Lindsell, D. A. Coomes, L. Guerriero, F. D. Frate, F. Miglietta, and R. Valentini. 2014. “Biodiversity Mapping in a Tropical West African Forest with Airborne Hyperspectral Data.” PLoS one 9 (6): e97910. doi:10.1371/journal.pone.0097910.

Lehmann, E. A., P. Caccetta, K. Lowell, A. Mitchell, Z.-S. Zhou, A. Held, T. Milne, and I. Tapley. 2015. “SAR and Optical Remote Sensing: Assessment of Complementarity and Interoperability in the Context of a Large-Scale Operational Forest Monitoring System.” Remote Sensing of Environment 156 (Jan.): 335–348. doi:10.1016/j.rse.2014.09.034.

Li, G., L. Dengsheng, E. Moran, M. F. Calvi, L. V. Dutra, and M. Batistella. 2018. “Examining Deforestation and...
