


Article

# Time-Series Model-Adjusted Percentile Features: Improved Percentile Features for Land-Cover Classification Based on Landsat Data

Shuai Xie <sup>1,2</sup>, Liangyun Liu <sup>1,\*</sup>  and Jiangning Yang <sup>1</sup>

<sup>1</sup> Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; xiesshuai@radi.ac.cn (S.X.); yangjiangning@aircas.ac.cn (J.Y.)

<sup>2</sup> College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

\* Correspondence: liuly@radi.ac.cn; Tel.: +86-10-8217-8163

Received: 21 August 2020; Accepted: 18 September 2020; Published: 21 September 2020



**Abstract:** Percentile features derived from Landsat time-series data are widely adopted in land-cover classification. However, the temporal distribution of Landsat valid observations is highly uneven across different pixels due to the gaps resulting from clouds, cloud shadows, snow, and the scan line corrector (SLC)-off problem. In addition, when applying percentile features, land-cover change in time-series data is usually not considered. In this paper, an improved percentile called the time-series model (TSM)-adjusted percentile is proposed for land-cover classification based on Landsat data. The Landsat data were first modeled using three different time-series models, and the land-cover changes were continuously monitored using the continuous change detection (CCD) algorithm. The TSM-adjusted percentiles for stable pixels were then derived from the synthetic time-series data without gaps. Finally, the TSM-adjusted percentiles were used for generating supervised random forest classifications. The proposed methods were implemented on Landsat time-series data of three study areas. The classification results were compared with those obtained using the original percentiles derived from the original time-series data with gaps. The results show that the land-cover classifications obtained using the proposed TSM-adjusted percentiles have significantly higher overall accuracies than those obtained using the original percentiles. The proposed method was more effective for forest types with obvious phenological characteristics and with fewer valid observations. In addition, it was also robust to the training data sampling strategy. Overall, the methods proposed in this work can provide accurate characterization of land cover and improve the overall classification accuracy based on such metrics. The findings are promising for percentile-based land cover classification using Landsat time series data, especially in the areas with frequent cloud coverage.

**Keywords:** percentile feature; time-series model; land-cover classification; Landsat; continuous change detection

## 1. Introduction

Earth observation data acquired by satellites are commonly utilized to map and monitor land covers [1], which is essential for research on biological diversities [2], wetland ecosystems management [3], and forest disturbances and recoveries [4,5]. Due to the long-term data record and moderate spatial resolutions which can capture spatial pattern of land-covers at a detailed level [6], images acquired by Landsat satellites are important dataset used to map land cover and monitor change [7]. The statistical metrics that are extracted using Landsat time-series data over a single year or consecutive years provide a novel spectro-temporal feature space for Landsat-based land-cover classification. These spectro-temporal statistic measurements have been demonstrated as feasible tools

for distinguishing land cover classes [8–10]. Their use in characterizing land-cover types began with coarse spatial resolution imagery [8,11].

Recently, statistical metrics have been widely used as input features to classify land cover over large areas based on Landsat data [12–15]. Statistical metrics, e.g., the percentile values for the specific time span, are insensitive to the geographical location-induced phenological differences in the same land-cover class because these metrics capture the magnitude of the changing reflectance instead of the timing [14]. For example, the 10th, 25th, 50th, 75th, and 90th percentile composites were generated using Landsat observations across four consecutive years to characterize land cover in Zambia [12]. The 20th, 50th, and 80th percentiles composites derived from monthly global Web-enabled Landsat Data (WELD) were used for land-cover mapping of North America [14]. The 0th, 25th, 50th, 75th, and 100th percentiles derived from the WELD monthly composites were included in the spectral inputs that were used to generate the continuous field for land-cover classes of the United States [13].

Unlike coarse spatial resolution data, the characteristics of Landsat data are that the observation counts are highly unequal due to the different Landsat acquisition strategies [16]. Furthermore, the occurrence of cloud, cloud shadow, snow, and also the missing data resulting from the Landsat 7 ETM+ scan line corrector (SLC) failure to reduce the amount of data available in the single Landsat scene. Consequently, the number and acquisition dates of available observations made over time varied unpredictably by pixels [12]. The percentiles derived from such data may not help to normalize the feature space, especially in the case of areas that are frequently cloudy. In addition, using consecutive multiple years of Landsat data for generating percentile syntheses is likely to be helpful in gaining spatial coverages; however, the inter- and intra-annual land cover changes (i.e., from one cover type to another) have not been considered and this method also introduces some contamination into the percentile composites derived from such data.

To accurately characterize land cover using percentile features, these issues need to be considered more carefully. First, the missing observations caused by the SLC-off problem and cloud cover need to be accurately estimated to increase the number of available observations from which the percentiles features are derived, especially in the case of frequently cloud-covered areas. Secondly, changes in land cover over time need to be detected before the Landsat time-series data are used to generate percentile features. The land-cover changes mentioned here refer to abrupt changes—i.e., from one cover type to another—rather than to seasonal land-surface changes that are due to, for example, vegetation phenology. In the case of Landsat data, the temporal scales of change-detection algorithm shift from the decennial [4] and annual scales [17,18] to sub-annual scales [7,19,20]. Sub-annual scales can reveal more detailed changes that were missed at the decadal and annual scales. In contrast with other sub-annual scale algorithms commonly employed for detecting forest changes, the continuous change detection (CCD) algorithm [7,21] is able to continuously detect various land-cover changes using all available Landsat images, and also has the ability to estimate any missing observation for any given date.

The objective of this study was to test the use of our new percentile feature approach for improving land cover classification and examine their performance over three study areas and under different levels of observation frequency. An improved percentile called the time-series model (TSM)-adjusted percentile was proposed. For the given Landsat time-series data, the land-cover changes were first detected using the CCD algorithm; the time-series observations between any two break points were then estimated using three different time-series models to fill the gaps caused by SLC-off and cloud cover. The TSM-adjusted percentiles that were derived from the synthetic time-series data without gaps were then used for the land-cover classification. The classification results based on TSM-adjusted percentile features were then compared with those based on the original percentile features that were derived from the original time-series data with gaps. The difference in classification accuracy between the two sets of results was comprehensively analyzed, and the impact of phenological characteristics and frequency of valid observations on the percentile features before and after the adjustment was discussed.

## 2. Data and Study Area

### 2.1. Landsat Data

The Landsat 5 and 7 satellites have a revisit period of 16 days but this revisit cycle can be reduced to 8 days via the complementarity of the two satellites [22,23]. The TM and ETM+ sensors, carried by Landsat 5 and 7, respectively, have similar spectral band configurations, and their data are collected using the Worldwide Reference System-2 (WRS-2) and defined in the Universal Transverse Mercator (UTM) projection. In this paper, the images acquired by the two sensors were used together to achieve higher temporal frequencies of Landsat observation.

All available Landsat TM/ETM+ surface reflectance (SR) images (with cloud cover less than 80%) acquired from 2000 to 2011 for selected study areas were exported from Google Earth Engine (GEE). This platform provides massive satellite imagery combined with cloud computing service, which makes satellite data access and processing fast and easy [24]. Landsat TM/ETM+ SR images are atmospherically corrected with the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) [25], and include the clouds, cloud shadows, and snow mask band generated from CFMASK [26,27]. Additionally, these images are geometrically aligned over time, meaning that it is straightforward to use them in time-series applications. In this paper, all of our analysis later were performed with the exported Landsat imagery using MATLAB codes in the personal computer (PC), not within the GEE platform, because at time of review CCD is available within GEE, but it was not at time of analysis.

In this study, all the exported SR images containing spectral bands stacked in the order Bands 1, 2, 3, 4, 5, 7, 6, and CFMASK were used as input data for a continuous change detection (CCD) algorithm, which was used to develop a time-series model and continuously detect abrupt changes. In addition, the Bands 1, 2, 3, 4, 5, 7, and Normalized Difference Vegetation Index (NDVI) were used for testing the proposed land-cover classification method.

### 2.2. Reference Data

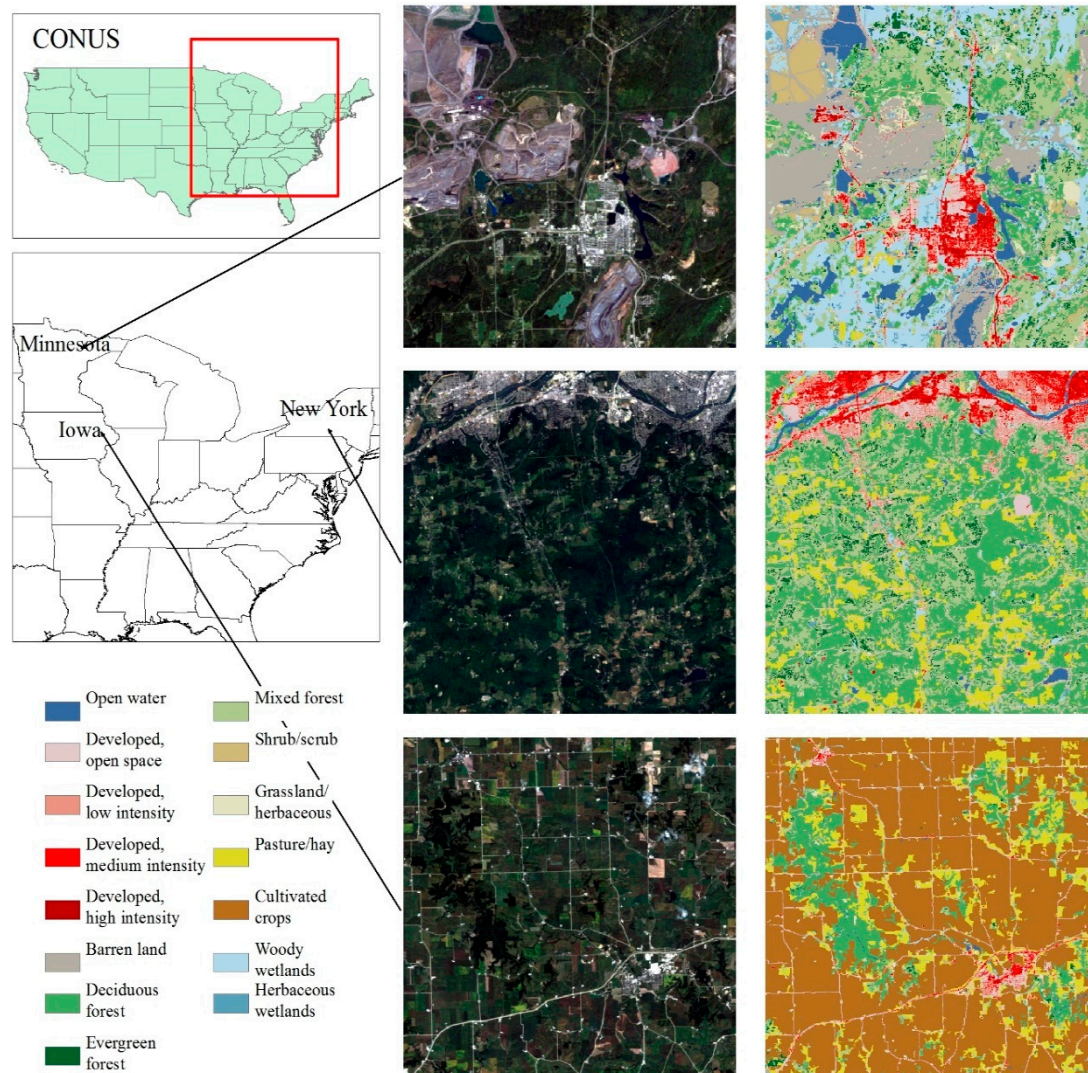
The National Land Cover Database (NLCD) is the well-established and commonly utilized sources of information related to land covers [28]. The NLCD 2011 product provides a land-cover dataset generated using Landsat data for three eras (2001, 2006, and 2011) at 30-m scale [29]. These products were also exported from the GEE platform and utilized as the source of training and test data for supervised classifications in this study. They were used because these products consist of various land-cover types, have a high classification accuracy, and provide national wall-to-wall coverage of the United States [30]. The NLCD 2011 product includes 16 land-cover classes together with ancillary datasets such as the Natural Resources Conservation Service Soil Survey Geographic database “Hydric Soils”, the National Agricultural Statistics Service Cropland Data Layer and the National Wetlands Inventory [29]; these were used to assist in post-classification refinement for specific land cover types. The overall accuracies of NLCD products are 82%, 83%, and 83% at Level II of the classification hierarchy and 88%, 89%, and 89% at Level I, for 2011, 2006, and 2001, respectively [31].

In this study, the NLCD data were reprojected from the Albers Equal Area projection to a UTM projection, and also resampled to have the same dimensions and same upper-left corner as Landsat TM images, so as to make them geographically compatible and to facilitate class label subsampling and spectral feature extraction.

### 2.3. Study Areas

The three study areas (see Figure 1) used in this research were covered by Landsat footprint WRS-2 path/row 027/027, 025/031 and 015/031, and located in Minnesota, Iowa, and New York in the United States. We took a subset of the full Landsat scenes from each study area. These areas were selected because (1) they have various land-cover types, including typical vegetation and non-vegetation types, which was of benefit to testing the effectiveness of the proposed method for various cover

types; (2) the land-cover changes in these areas affected a relatively small proportion of the study areas; and (3) a considerable fraction of the Landsat time-series data that cover these areas were missing, which helped in testing the robustness of the proposed method to changes in the number of valid observations.



**Figure 1.** Illustration of the 2011 Landsat TM (middle column) and National Land Cover Database (NLCD) (right column) data of the three study areas. A subset of the full Landsat scenes was taken to form each study area. The Landsat images are displayed as 3-2-1 TM band combinations. Landsat data of Minnesota, Iowa, and New York study area were acquired on 11 September, 13 September, and 21 July, respectively. Maps in the upper left corner show the location of the three study areas in the United States.

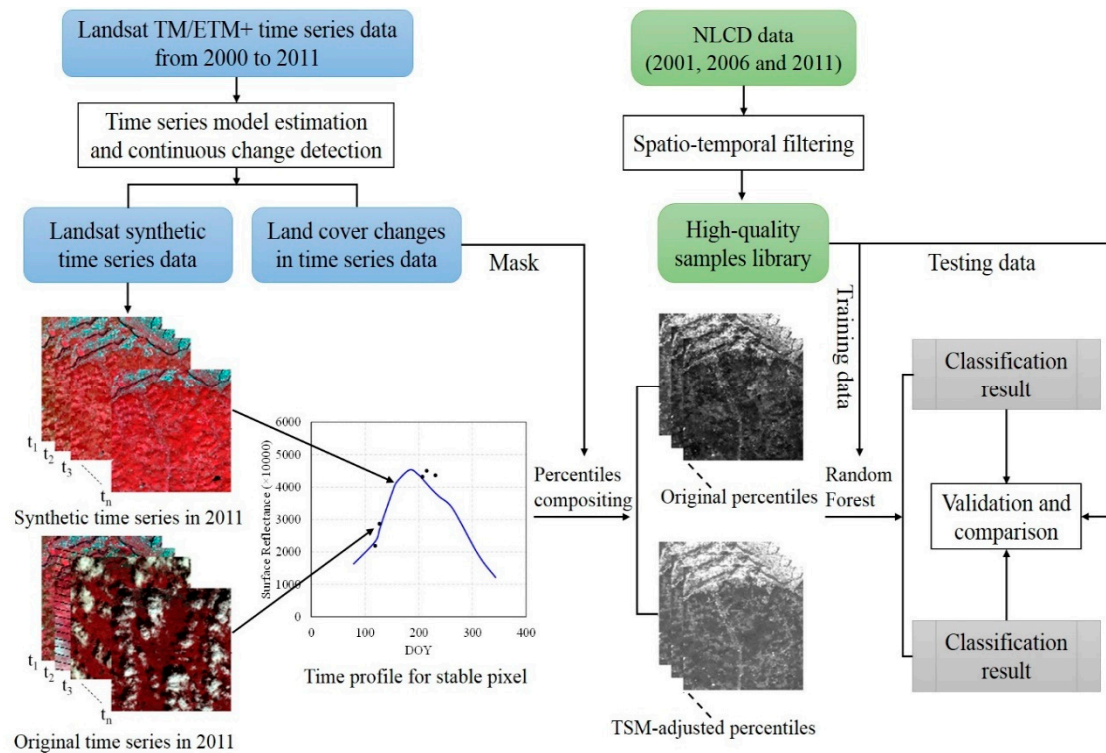
Any individual study area we used contained approximately 240,000 to 310,000 30-m pixels. We did not use the larger areas because the computational loads required for the time series models estimation and continuous change detection increased greatly with the image's spatial sizes. Table 1 summarizes the number of images from 2000 to 2011 used for developing the time-series models and continuous change detection, the number of images from the target year 2011 used in the classification experiments, and the geographic characteristics of the three study areas. Figure 1 illustrates the 2011 Landsat TM and corresponding NLCD data for the three study areas. The NLCD data are displayed with the standard color legend that is available from the Multi-Resolution Land Characteristics Consortium (MRLC) (<https://www.mrlc.gov/data/legends/national-land-cover-database-2016-nlcd2016-legend>).

**Table 1.** Summaries of the number of images from 2000 to 2011 used for continuous change detection, the number of images from the target year 2011 used for classification, and the geographic characteristics of the three study areas.

Study Area	Number of Images from 2000 to 2011 Used for Continuous Change Detection	Number of Images from Target Year 2011 Used for Classification	Spatial Size (Pixels)	Area (km <sup>2</sup> )	Longitudinal Extent	Latitudinal Extent
Minnesota	306	12	489 × 505	222.25	92.6687°W to 92.4689°W	47.4744°N to 47.6056°N
Iowa	318	14	545 × 562	275.66	91.1540°W to 90.9542°W	42.2589°N to 42.4026°N
New York	299	13	541 × 558	271.69	76.0972°W to 75.8974°W	41.9614°N to 42.1057°N

### 3. Methodology

The proposed TSM-adjusted percentile method consisted of two major steps (Figure 2): (1) development of the time-series models and continuous change detection, and (2) percentile feature generation. In this study, the TSM-adjusted percentile features were then utilized to classify land cover.



**Figure 2.** Flowchart of the proposed method.

#### 3.1. Development of the Time-Series Model and Continuous Change Detection (CCD)

Clouds, cloud shadows, and ephemeral snow limit the availability of Earth observations acquired by the Landsat series of satellites. Therefore, the number of clear observations (not contaminated by clouds, cloud shadows, and ephemeral snow) across a certain time span varies from pixel to pixel. For every pixel, Fmask [27,32] and Tmask algorithms [33] were first used to mask out the clouds, cloud shadows, and ephemeral snow to obtain the time-series of clear observations. The surface reflectance of different spectral bands were then estimated using three different time-series models—simple, advanced, and full (Equations (1)–(3))—that included harmonic and trend components for the observations [21]. Which model was used depended upon the number of clear observations: 12 to 18 clear observations were required for the simple model; the advanced model needed 18 to 24 clear observations; the full model would be selected if there were more than 24 clear observations. Our ability to model the intra-year variation of Landsat time-series observations was improved with a more complex model. The least absolute shrinkage and selection operator (LASSO) regression technique was used for estimating the coefficients of time-series models [34,35]. The LASSO technique can minimize the sum of the squares of the residuals and has a constraint on the sum of the absolute values of the coefficients [36]. This allowed a time-series model that did not have the problem of overfitting to be developed [21].

$$\hat{\rho}(i, x)_{simple} = a_{0,i} + a_{1,i} \cos\left(\frac{2\pi}{T}x\right) + b_{1,i} \sin\left(\frac{2\pi}{T}x\right) + c_{1,i}x \quad (1)$$

where

$x$ —Julian day

$i$ —Landsat band  $i$  ( $i = 1, 2, 3, 4, 5,$  and  $7$ )

$T$ —number of days of the year ( $T = 365$ )

$a_{0,i}$ —constant term that represents the mean for Landsat band  $i$

$a_{1,i}, b_{1,i}$ —coefficients of intra-year variation components for Landsat band  $i$

$c_{1,i}$ —coefficient of inter-year variation component (slope) for Landsat band  $i$

$\hat{\rho}(i, x)_{simple}$ —surface reflectance of Landsat band  $i$  on Julian day  $x$  obtained using the simple model.

$$\hat{\rho}(i, x)_{advanced} = \hat{\rho}(i, x)_{simple} + a_{2,i}\cos\left(\frac{4\pi}{T}x\right) + b_{2,i}\sin\left(\frac{4\pi}{T}x\right) \quad (2)$$

where

$a_{2,i}, b_{2,i}$ —coefficients of the intra-year bimodal variation components for Landsat band  $i$

$\hat{\rho}(i, x)_{advanced}$ —surface reflectance of Landsat band  $i$  on Julian day  $x$  obtained using the advanced model.

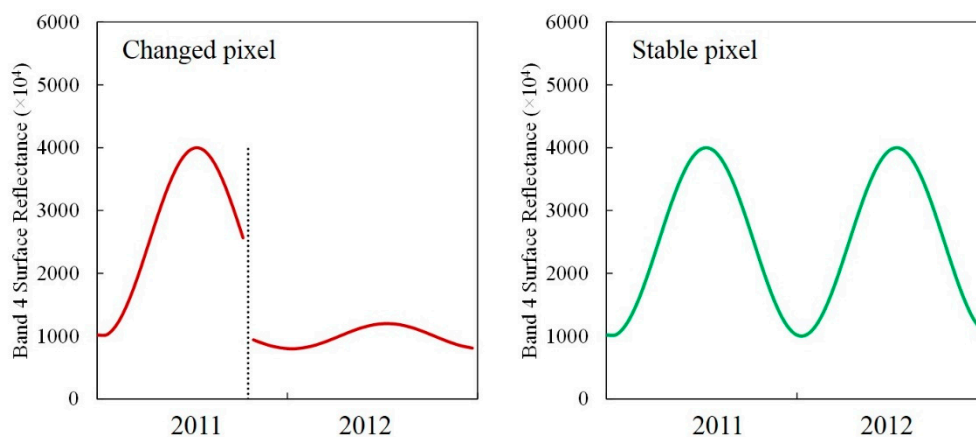
$$\hat{\rho}(i, x)_{full} = \hat{\rho}(i, x)_{advanced} + a_{3,i}\cos\left(\frac{6\pi}{T}x\right) + b_{3,i}\sin\left(\frac{6\pi}{T}x\right) \quad (3)$$

where

$a_{3,i}, b_{3,i}$ —coefficients of intra-year trimodal variation components for Landsat band  $i$

$\hat{\rho}(i, x)_{full}$ —surface reflectance of Landsat band  $i$  on Julian day  $x$  obtained using the full model.

Abrupt surface changes were detected based on comparisons of model-predicted values with real observation data from Landsat. If the difference was larger than a given threshold on six consecutive occasions, the pixel was identified as a changed pixel. To detect various surface change accurately, a change was defined using all the spectral bands except for blue and TIR bands, because these two spectral bands are quite sensitive to atmospheric effect and less sensitive to most of the surface changes; and the change threshold was determined using root mean square error (RMSE) from the time series model fit for each spectral band. If an abrupt surface change had occurred, a break would occur in the time-series model and newly collected clear observations were added to fit to a new time-series model. Figure 3 illustrates changed and stable pixels detected by the CCD. In this paper, changed pixels were removed from further classification analyses, and labeled as “disturbed” in the final classification maps. The time-series of stable pixels were used to generate the percentile features.



**Figure 3.** Illustration of changed and stable pixels detected by CCD. The time-series models shown are based on Landsat band 4 data. The left-hand panel illustrates the signal for a changed pixel, and the vertical black dotted line represents the change that occurred in 2011. This pixel was labeled as “disturbed” in the land-cover map for 2011. The right-hand panel illustrates the signal of a stable pixel; the percentile features for this land-cover type were derived from time-series data such as these.

### 3.2. TSM-Adjusted Percentile Features

#### 3.2.1. Method Used for Calculating Percentiles

A temporal metric is the feasible conversion of time-series data over a given interval. Metrics can summarize the multi-temporal feature space, which captures the prominent phenological features regardless of the specified period of year [37]. Percentiles have been commonly used as the temporal metric in land-cover classification. Method for calculating percentiles used in this study can be formulated as follows (Equation (4)).

$$P_{i,k} = \begin{cases} (\rho_{i,ascending}(R) + \rho_{i,ascending}(R+1))/2, & R \in N_+ \\ \rho_{i,ascending}(\lceil R \rceil), & R \notin N_+ \end{cases} \quad (4)$$

where  $i$  denotes the  $i$ th Landsat band ( $i = 1, 2, 3, 4, 5$ , and  $7$ );  $k$  is any number between zero and one hundred;  $P_{i,k}$  denotes the  $k$ th percentile for the surface reflectance of the  $i$ th Landsat band over a given temporal interval;  $\rho_{i,ascending}$  denotes the surface reflectance data in ascending order for the  $i$ th Landsat band over the given temporal interval; and  $R$  is the rank of the  $k$ th percentile, which is computed as

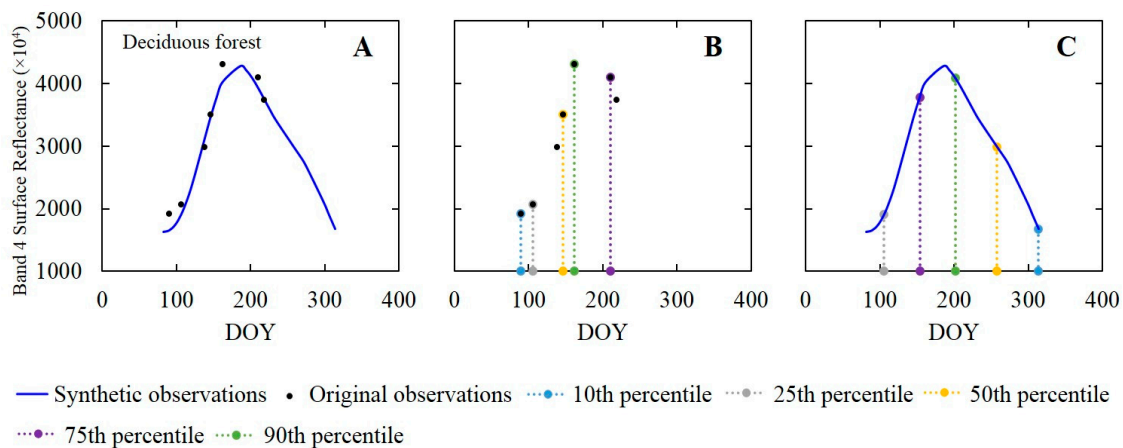
$$R = \frac{k}{100} \times N \quad (5)$$

where  $N$  indicates the total number of clear observations for a given temporal interval. Specifically, if the rank obtained using Equation (5) is a whole number, the  $k$ th percentile is the average of the  $R$ th and  $(R+1)$ th values in the surface reflectance data in ascending order; if the rank obtained using Equation (5) is not a whole number, it is rounded up to the nearest whole number and the  $k$ th percentile is then the  $\lceil R \rceil$ th value in the surface reflectance data in ascending order.

#### 3.2.2. Generation of TSM-Adjusted Percentile Features

Generally, percentile features are generated based on clear observations acquired during a given temporal interval [13]. However, the gaps caused by clouds, cloud shadows, snow, and the SLC-off issue lead to changes in the frequency of Landsat observations with time. According to Equations (4) and (5), the values of percentiles are affected by the total number of clear observations and the surface reflectance values over the given time interval. Therefore, the original percentiles features may vary for pixels belonging to the same land-cover type but for which there are different numbers of clear observations within the time interval. In this study, we proposed the TSM-adjusted percentile features with the aim of characterizing land cover accurately and improving the classification accuracy substantially. The time-series of surface reflectances were first estimated using the models based on clear observations (detailed in Section 3.1). Next, percentile features based on synthetic time-series observations were generated. For illustration purposes, we generated the 10th, 25th, 50th, 75th, and 90th percentile features of a deciduous forest pixel over the period of a year based on both the original and synthetic time-series of surface reflectances. The original percentiles were generated based on time-series of clear observations that were temporally discrete due to the gaps resulting from the SLC-off issue and cloud cover (see Figure 4B). The TSM-adjusted percentiles were generated based on time-series of synthetic observations that were temporally continuous because the gaps had been estimated by the time-series models (see Figure 4C). Therefore, the proposed TSM-adjusted percentiles completed the total number of clear observations over a given time interval for the entire study area.





**Figure 4.** Illustration of the generation of time-series model (TSM)-adjusted percentile features for deciduous forest: (A) time-series model based on clear Landsat band 4 observations over a given year, (B) the original percentile features generated using time-series of clear observations, and (C) the TSM-adjusted percentile features generated using synthetic time-series observations.

### 3.3. Classification Experiment Methodology

The pixels that were stable over the chosen temporal interval were classified with supervised random forest (RF) classifier and the changed pixels detected by CCD were labelled as “disturbed” in land-cover maps. The percentile features, including the 10th, 25th, 50th, 75th, and 90th percentiles for Landsat bands 1-5 and band 7, and the NDVI (i.e., 35 features in total), were used as input data for the RF classifier. The RF classifier was an ensemble machine-learning algorithm which operates by constructing sets of decision trees for classification during the training [38]. All the generated decision trees were used to classify the newly unlabeled data, and the category receiving the largest number of votes will be given to this data. The forest trees were generated by setting two parameters: in this study, we set the number of trees to 500; in addition, the number of split variables was set to the defaults, i.e., square root of the total number of input features. We chose RF classifier for use due to its superiority in handling high-dimensional input features without dimension reduction, its robustness against outliers, as well as the high classification accuracies achieved by the use of ensemble techniques [39].

The training and testing data for the classification were collected using 2011 NLCD land cover maps of the three study areas. The four land-cover types related to impervious surfaces were spatially merged into one type named “developed”. Spatio-temporal filtering methods were used to assist in the selection of highly accurate class labels. The pixels in the 2011 NLCD data were selected only if the following filtering criteria were met. First, the NLCD pixel values for 2001, 2006, and 2011 had to be identical. The use of this temporal rule helped select the NLCD pixels having the identical land-cover type between 2001 and 2011. Second, the 2011 NLCD pixel had to have the same value as the eight pixels surrounding it. This spatial rule was used to help reduce 30-m pixel edge effects which may produce apparent mix in land cover. The generated class labels used for training and testing were illustrated in Supplementary Materials Figure S1.

The sampling technique adopted in this study is stratified random sampling; that is, each land cover type is sampled independently and randomly. The number of samples collected for each type is proportional to the area occupied by each type. Additionally, the effect of training-data balance was also considered because an imbalanced sample size among classes would substantially decrease the accuracies of rare class because of the very few extracted training samples [40]. A total of 3437, 1869, and 2261 training pixels (see Table 2 for the number of training pixels of each land cover type) were selected for the Minnesota, Iowa, and New York study areas, respectively; the remaining candidate NLCD pixels are treated as testing data.

**Table 2.** The number of training pixels for each land cover type in the three study areas.

Land Covers	Minnesota	Iowa	New York
Open water (OW)	420	210	210
Developed (D)	245	140	280
Barren land (BL)	700	-	-
Deciduous forest (DF)	350	350	630
Evergreen forest (EF)	182	49	140
Mixed forest (MF)	490	140	420
Shrub/scrub (S)	280	-	-
Grassland/herbaceous (G)	140	70	-
Pasture/hay (P)	-	350	420
Cultivated crops (CC)	-	560	-
Woody wetlands (WW)	560	-	161
Herbaceous wetlands (HW)	70	-	-

The NLCD class labels and the original and proposed TSM-adjusted percentile features were extracted for each training sample. The original percentiles were also extracted in order to compare the classification results obtained using the proposed TSM-adjusted percentiles with those obtained using the original percentiles. The RF classification results were generated using the training data; these results were then quantitatively assessed using the common classification accuracy metrics of overall accuracy (OA), per-class producer's accuracy (PA) and user's accuracy (UA), which were derived from the confusion matrices using the corresponding test data [41]. To apply the statistical significance testing, the above classification experiment was conducted repeatedly 10 times. Paired *t*-tests were used to determine if the differences in accuracy of the two sets of classifications were significant at the 5% level. In addition, a final land cover classification was generated for the spatial visualization using the most frequent category from the 10 individual ones.

In order to investigate the effect of valid observation frequency and training data sampling strategy on classification results, some other experiments were performed, and the results were presented in Sections 5.2 and 5.3. In Section 5.2, the classification accuracies of land cover types with different numbers of valid observations were calculated. More specifically, for each independent classification experiment, the test data for each land-cover type were stratified according to valid observation counts, and then the accuracy for each layer was calculated by comparing the classified data against the corresponding test data. More specifically in Section 5.3, in addition to the sampling strategy used in previous experiments, i.e., random sampling across valid observation frequency strata for each class, another two sampling strategies were used, i.e., random sampling from the pixels with low (high) observation frequencies for each class. The experiments were designed to test the robustness of the proposed TSM-adjusted percentiles to the sampling strategies of training data.

## 4. Results

### 4.1. Classification of Percentiles Derived from Multispectral Reflectance and NDVI Time Series

The original and TSM-adjusted percentiles that were respectively derived from the original and synthetic Landsat 6-bands reflectances and NDVI time-series for April to October of the climatological year 2011 were classified. A comparison of the classification results obtained using the original percentiles with those obtained using the proposed TSM-adjusted percentiles provided insights into whether using the TSM-adjusted percentiles led to improved classification results. The overall accuracies of original and TSM-adjusted percentiles-based classification results are summarized in Table 3. The overall accuracies derived from the TSM-adjusted percentiles are significantly higher than those derived from the original percentiles in any single selected test area. This indicates that compared to original percentiles, the proposed TSM-adjusted percentiles have the improved overall classification performance. In addition, the standard deviations in the OA for the TSM-adjusted percentiles-based

classification are smaller than those for the original percentiles-based classification for all three study areas. This result suggests that the former yielded more stable results than the latter. The original percentiles for training and test pixels randomly selected each time might exhibit some degree of variability in the repeatedly performed experiments, because the uneven temporal distribution of valid observations exist in the original time series from which the original percentiles were derived. In contrast, the TSM-adjusted percentiles are produced from the synthetic time-series observations without gaps resulting from clouds, cloud shadows, snow, and missing data. Thus, the uncertainty in the temporal distribution of observations for training and test data can be alleviated.

**Table 3.** Overall classification accuracies achieved by the use of original percentiles and TSM-adjusted percentiles for the three study areas. The average values and standard deviations (in parentheses) of the overall accuracies were derived from ten independent classification experiments. Note that the upward arrows indicate that the overall accuracies obtained using the TSM-adjusted percentiles are significantly (at the 5% level) higher than that obtained using the original percentiles.

Study Area	Original Percentiles-Based Classification	TSM-Adjusted Percentiles-Based Classification
Minnesota	81.98% (0.35%)	82.82%↑(0.33%)
Iowa	93.53% (0.63%)	94.31%↑(0.38%)
New York	85.99% (0.41%)	90.05%↑(0.39%)

Table 4 summarizes the user's and producer's accuracies of each land cover class for original percentiles-based classification, and the counterpart for TSM-adjusted percentiles-based classification are given in Table 5. User's accuracy of one land cover class is the ratio of the number of pixels correctly classified as that class to the total number of pixels classified as that class; the relative higher user's accuracy indicates fewer commission errors. Producer's accuracy of one land cover class is the ratio of the number of pixels correctly classified as that class to the total number of pixels specified as that class in the reference data; the relative higher producer's accuracy indicates fewer omission errors. The significant difference of UA and PA between the two sets of classification was reported in Table 5. This result indicated that the improvements obtained using the proposed method were different between specific land cover classes, and also between producer's and user's accuracies for the same class. Further, these results also varied across the three study areas. In order to discuss the difference in improvements obtained from the proposed method between specific land cover classes and across the three study areas, we chose five different land cover types (i.e., open water, developed, deciduous forest, evergreen forest, and mixed forest) as examples for analysis in Sections 4.1, 5.1 and 5.2 because they can be found in all three study areas.

**Table 4.** Average producer and user accuracies of the classifications derived from original percentiles (see Table 3 for the associated overall accuracies). Ten independent classification experiments in total were carried out.

		OW	D	BL	DF	EF	MF	S	G	WW	HW
Minnesota	UA	95.13%	32.51%	96.78%	63.89%	17.72%	78.26%	83.71%	23.95%	81.86%	51.84%
	PA	95.86%	77.42%	93.03%	72.18%	75.32%	78.74%	89.31%	83.28%	67.53%	62.89%
Iowa				OW	D	DF	EF	MF	G	P	CC
	UA			64.09%	9.90%	96.08%	84.43%	5.05%	4.56%	55.10%	99.33%
	PA			90.58%	71.41%	90.75%	90.00%	73.33%	37.66%	86.02%	94.35%
				OW	D	DF	EF	MF	P	WW	
New York	UA				99.24%	67.36%	94.46%	19.29%	70.29%	84.43%	20.43%
	PA				99.53%	74.27%	87.60%	61.28%	75.30%	92.46%	57.56%

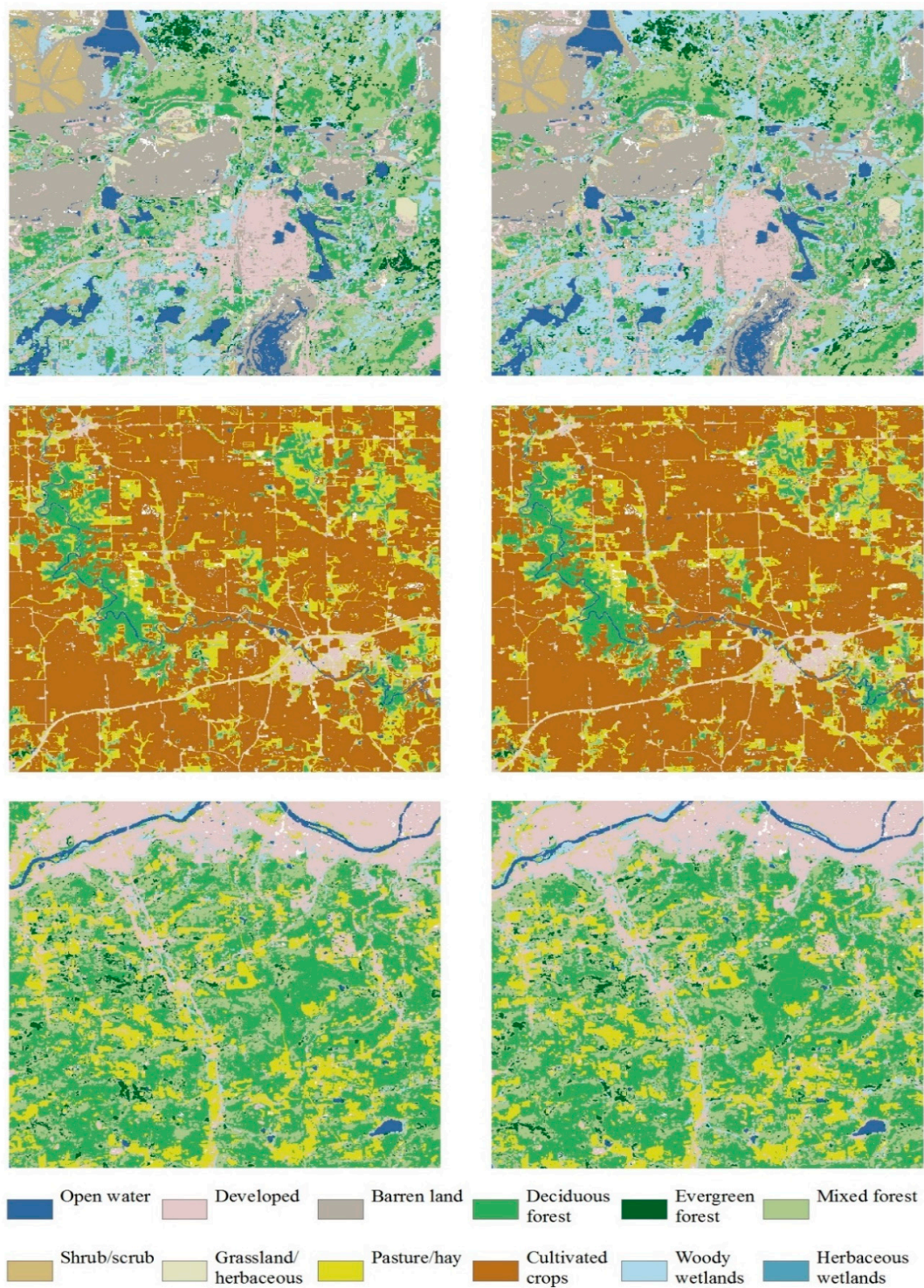
**Table 5.** Average producer and user accuracies of the classifications derived from TSM-adjusted percentiles (see Table 3 for the associated overall accuracies). Ten independent classification experiments in total were carried out. Note that the upward (downward) arrows indicate that the accuracies obtained using the TSM-adjusted percentiles are significantly (at the 5% level) higher (lower) than that obtained using the original percentiles. Figure without arrow behind it represents that there is no significant difference between the two sets of classification.

		OW	D	BL	DF	EF	MF	S	G	WW	HW
Minnesota	UA	94.96%	31.57%	96.12%↓	65.84%↑	21.14%↑	82.25%↑	74.90%↓	23.20%	83.02%↑	32.17%↓
	PA	95.73%	79.80%↑	91.84%↓	75.60%↑	82.71%↑	76.16%↓	84.36%↓	66.94%↓	73.67%↑	52.47%↓
Iowa				OW	D	DF	EF	MF	G	P	CC
	UA			60.70%↓	9.57%	96.54%	83.84%	5.37%	3.24%↓	60.79%↑	99.29%
	PA			90.53%	64.93%↓	91.37%↑	91.33%	80.00%↑	43.13%↑	84.53%↓	95.31%↑
				OW	D	DF	EF	MF	P	WW	
New York	UA				99.17%	64.87%↓	97.07%↑	26.58%↑	83.64%↑	88.05%↑	17.88%↓
	PA				98.97%↓	81.90%↑	91.77%↑	79.01%↑	83.12%↑	91.73%↓	75.75%↑

More specifically, the accuracies of open water had no significant difference, but it had lower UA in Iowa and lower PA in New York. The improvement of developed was seen in terms of PA in Minnesota and New York, but its PA was lower in Iowa and its UA was lower in New York. The accuracies of deciduous forest were significantly improved across all three study areas except the UA in Iowa. Evergreen forest had higher UA and PA in Minnesota and New York but had no significant difference in classification accuracy in Iowa. The PA of Mixed forest was improved in Iowa and New York but was decreased in Minnesota; the UA of it was improved in Minnesota and New York, but have no significant difference in Iowa.

#### 4.2. Spatially Explicit Classification Results

This section provides a spatial visualization of the classification results obtained using the original and TSM-adjusted percentiles (see Figure 5). An individual final classification result of any single test area is produced instead of examining the 10 random independent classification results (see Tables 3–5 for a summary of the associated classification accuracy metrics). The final classification results were obtained by assigning the most frequent category in the 10 random forest classifications to each pixel. The land-cover changes detected by the CCD algorithm for 2011 were labeled as “disturbed” in the final classification results.

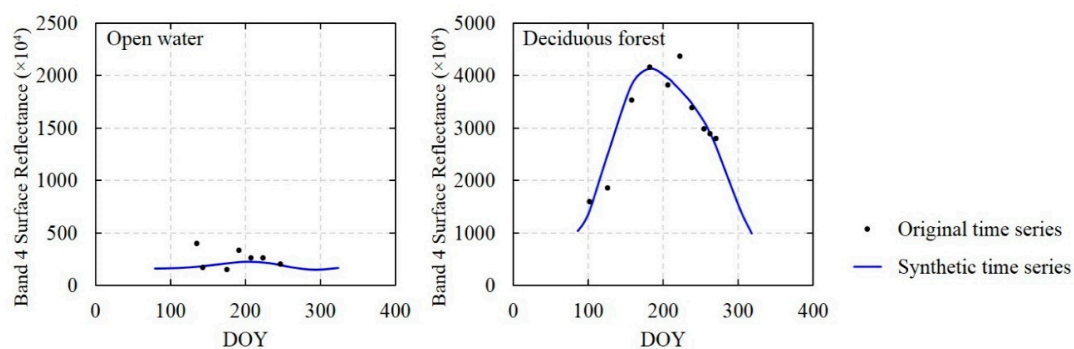


**Figure 5.** Spatially explicit classifications of the three study areas for the target year 2011. Left-hand column: the final classifications generated using 10 independent random forest (RF) supervised classification results of original percentiles. Right-hand column: the final classifications generated using 10 independent RF supervised classification results of proposed TSM-adjusted percentiles. Top: Minnesota classification results, middle: Iowa classification results, and bottom: New York classification results. Colors correspond to those used in Figure 1 except for the “Developed” class, which is rendered in light pink. White shows the pixel locations where land-cover changes occurred: these pixels were labeled as belonging to the “disturbed” class.

## 5. Discussion

### 5.1. Effect of Phenological Characteristics

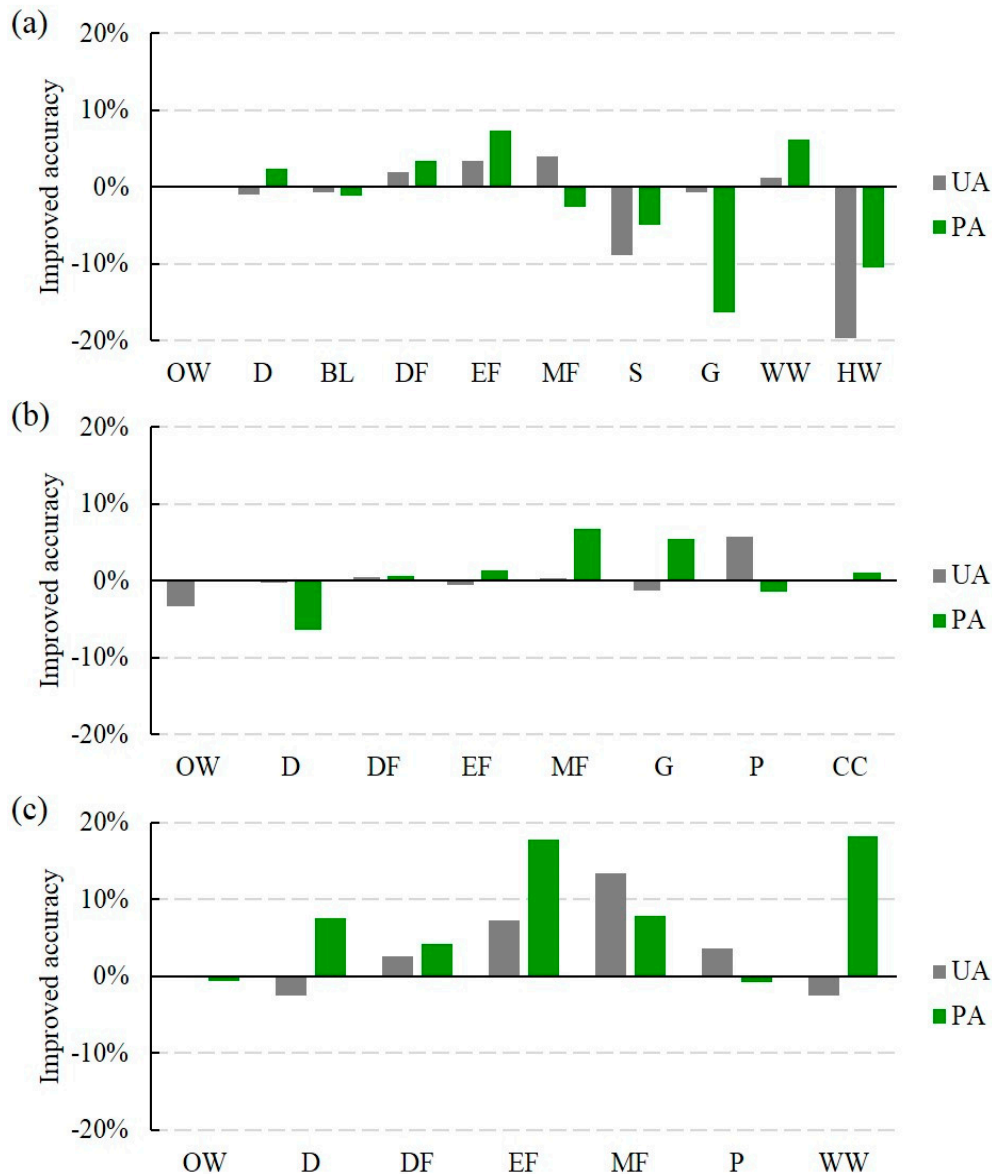
The mechanisms producing the changes in surface reflectance over time varied according to the land-cover type. For example, the reflectance of deciduous forest changed over time due to the phenological characteristics. For cover types that have no seasonal features, such as open water, changes in illumination geometry led to the changes in surface reflectance (due to, for example, the bidirectional reflectance distribution function (BRDF) effect). Nevertheless, the magnitude of the changes caused by the illumination geometry was much smaller than that caused by phenological characteristics. Figure 6 illustrates the changes in Landsat band-4 reflectance in 2011 for the deciduous forest and open water cover types. It is evident that the band-4 reflectance reached its peak during the growing season for deciduous forest and that for open water it did not vary significantly during the whole year.



**Figure 6.** Illustration of Landsat band-4 time-series for open water and deciduous forest in 2011.

Figure 7 illustrates the differences in PA and UA between the TSM-adjusted and original percentiles-based classifications for the three study areas to show how the improvement in classification accuracies using the proposed TSM-adjusted percentiles varied by cover type and across different study areas. As expected, the improvement both in UA and PA for deciduous forest, evergreen forest, and mixed forest were consistently observed across all the three study areas (except for the slightly lower PA of mixed forest in Minnesota); the improvement was most significant in New York study area, followed by Minnesota and least significant in Iowa. On the other hand, higher PA for forest types with less improvement in UA would indicate lower rates of omission (more of what is forest in the reference data is captured as such by the adjusted percentile feature classification), but similar rates of commission (the adjusted percentile feature classification still misclassifies other classes as forest at the same levels). There is no noteworthy improvement in classification accuracy for open water across the three study areas using the proposed method (except for the slightly lower UA of open water in Iowa). These results can be interpreted by considering that the percentiles capture the magnitude of time-series reflectances. The gaps in time-series observations that are used to generate percentile features have little impact on open water, where there is little change in reflectance over time. However, these gaps have a great impact in the case of deciduous, evergreen and mixed forest where there are great changes in reflectance over time due to phenological characteristics; this is especially true in frequently cloud-covered areas—i.e., the areas with fewer valid observations. Thus, the original percentiles derived from pixels with fewer valid observations might be not as much reliable to produce accurate classification results as that derived from pixels with a good number of valid observations. In contrast, the proposed TSM-adjusted percentiles could normalize the number of valid observations for pixels using time series models. As a result, compared to using original percentiles, the use of TSM-adjusted percentiles can significantly improve the classification accuracy for forest with obvious phenological characteristics. The developed class exhibited a certain degree of uncertainty in accuracy

variation across the three study areas, due to the complex spectral heterogeneity of the developed type. For example, the developed type in NLCD consists of four sub-types with different levels of mixture of constructed materials and vegetation. Therefore, the seasonal and non-seasonal spectral variation both exist in developed type.

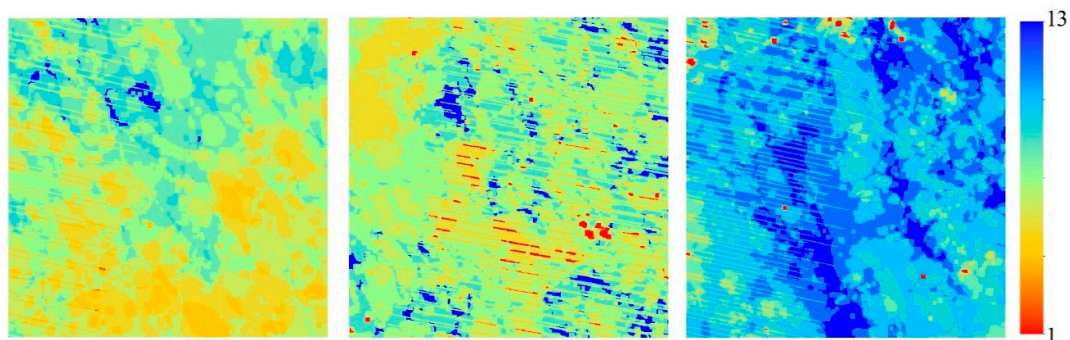


**Figure 7.** The differences in PA and UA between the TSM-adjusted and original percentiles-based classification for individual land-cover types in Minnesota (a), Iowa (b), and New York (c).

### 5.2. Effect of the Frequency of Valid Observations

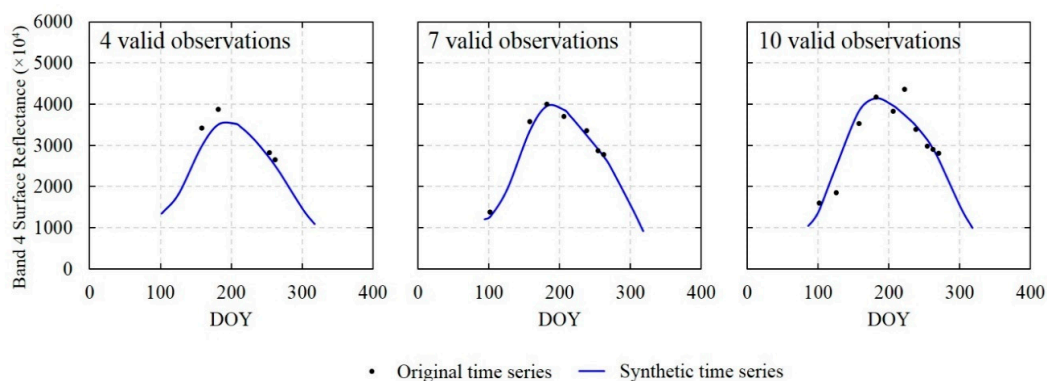
Missing data resulting from clouds, cloud shadows, snow, and the SLC-off problem is normal in Landsat observation time-series. Although the implementation of percentile compositing can alleviate the impact of missing data, the more missing data there are, the less reliable the resulting percentile bands. Generally the distribution of missing satellite observations is not even over time [42–44] and, consequently, the number and acquisition date of valid observations varied by pixels. Figure 8 illustrates the spatial patterns in the frequency of valid observations for the three study areas in 2011. The corresponding histograms of the valid observations frequency for each study area were shown in

Figure S2. In addition, the level of time series models (simple, advanced, and full) used for surface reflectance estimation in 2011 for each pixel of each study area were provided in Figure S3.



**Figure 8.** Spatial patterns in the number of valid observations in 2011 for the three study areas. (left): Minnesota; (middle): Iowa; and (right): New York.

As can clearly be seen from Figure 9, the profiles of the original reflectance time-series of deciduous forest samples with different numbers of valid observations are different; however, the profiles of the synthetic reflectance time-series estimated by the time-series models are similar for all of the deciduous forest samples regardless of valid observation counts. In fact, the percentile values were affected by the number and magnitude of the time-series data, according to the formulae given in Section 3.2.1. Therefore, it was expected that the proposed TSM-adjusted percentiles derived from the synthetic time-series were less influenced by valid observation frequency, and that they would show less variation and be more accurate than those obtained using original time-series in characterizing land-cover.

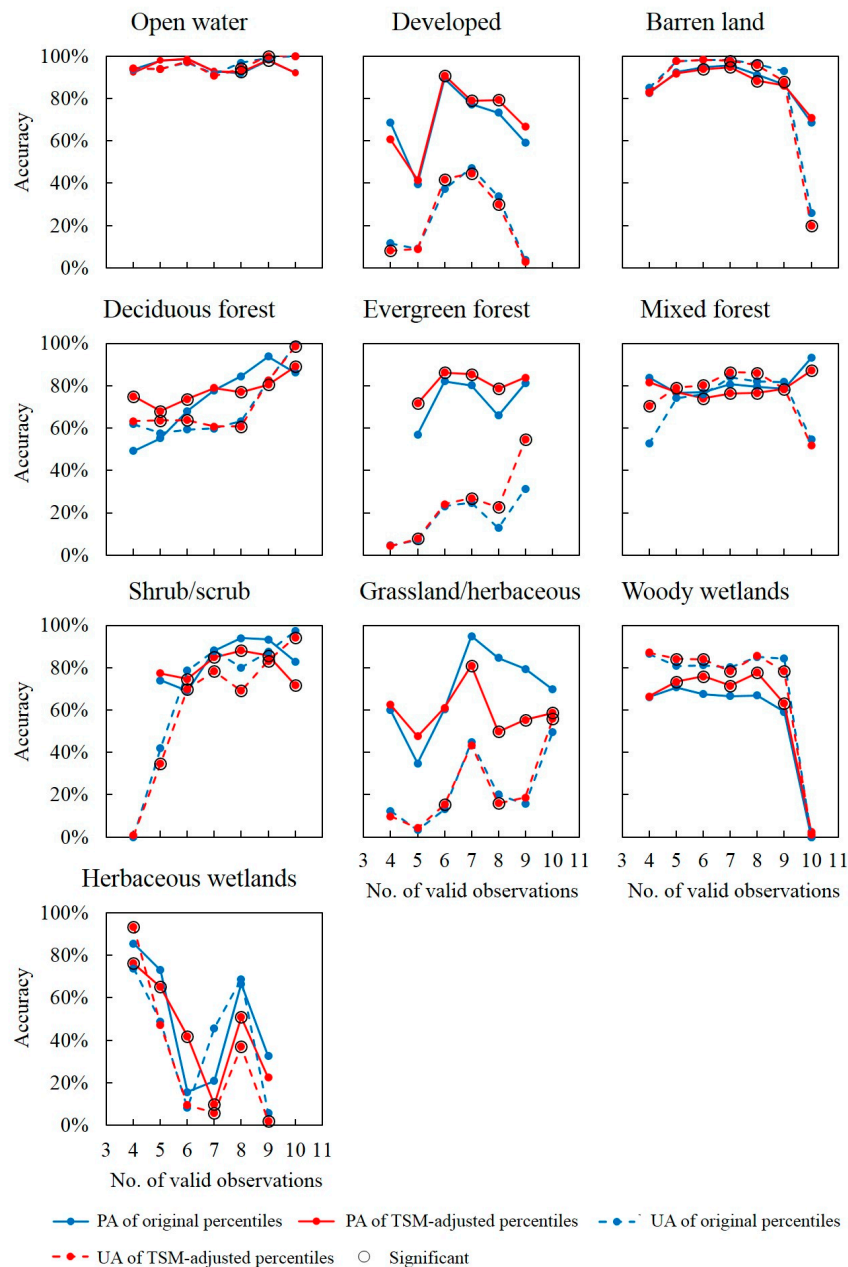


**Figure 9.** Illustration of Landsat band-4 time-series of deciduous forest samples with different numbers of valid observations during the year 2011. (The black dots indicate original time-series observations and blue curves indicate synthetic time-series).

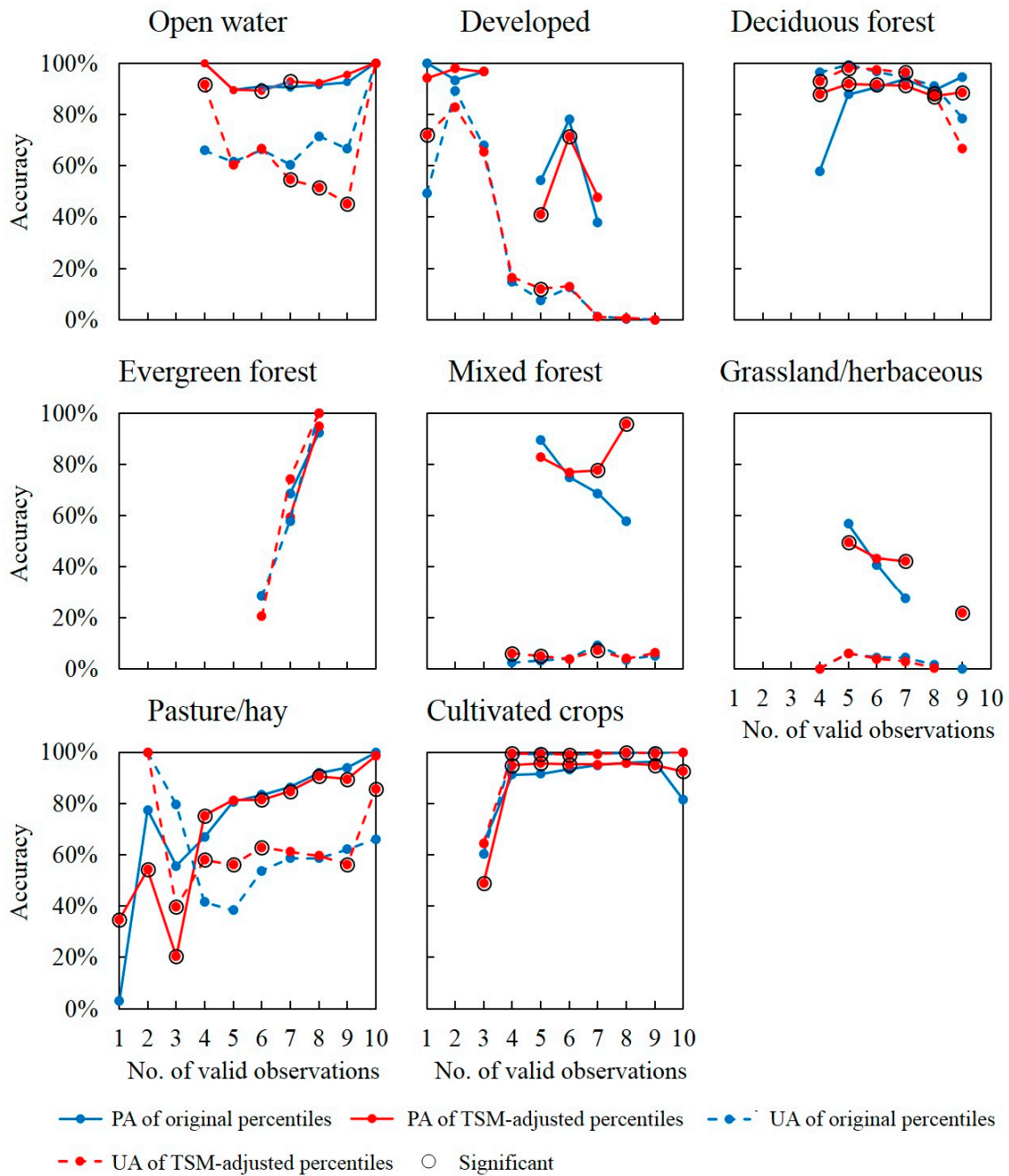
The effect of valid observation frequency (i.e., the number of valid observations) was examined by calculating the classification accuracies of land cover types with different numbers of valid observations. Figures 10–12 illustrate the responses of the classification accuracies for each land cover class to valid observation counts in the three study areas. As evident, for open water and developed types, there is no obvious difference between the accuracies obtained using the TSM-adjusted percentiles and those obtained using the original percentiles regardless of the number of valid observations, except for UA curve of open water in Iowa and PA curve of developed in New York. By contrast, for deciduous forest in all three study areas and evergreen forest in Minnesota and New York study areas, the PA obtained by using the TSM-adjusted percentiles are much better than those obtained using the original percentiles especially when the number of valid observations is lower; the PA difference between the two sets of results becomes smaller as the number of valid observations increases. Interestingly, except



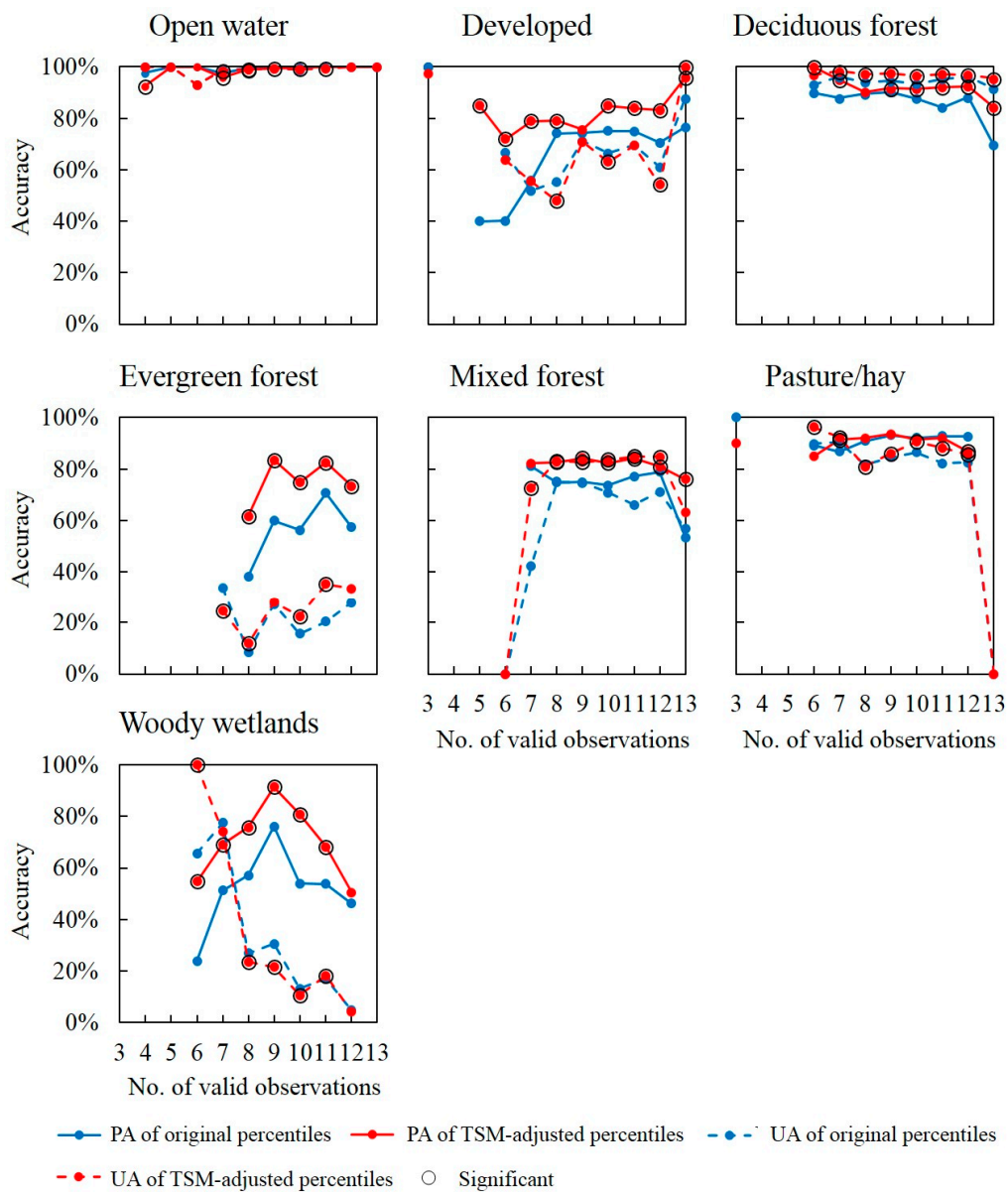
for evergreen forest in Iowa and mixed forest in Minnesota and New York study areas, UA curves for forest types showed no obvious difference in performance between original percentiles-based and TSM-adjusted percentiles-based classification across the different number of valid observations in the three study areas. It was worth noting that, compared to using original percentiles, the use of the TSM-adjusted percentiles can greatly improve the PA for deciduous forest types, especially when the number of valid observations was fewer. Additionally, the accuracy differences between the TSM-adjusted and original percentiles are smaller for open water and developed types regardless of the frequency of valid observations.



**Figure 10.** The average PA and UA of each land cover type against the number of valid observations in Minnesota study area. Note that the black circles indicate that the accuracies obtained using the TSM-adjusted percentiles are significantly (at the 5% level) different from the accuracies obtained using the original percentiles.



**Figure 11.** The average PA and UA of each land cover type against the number of valid observations in Iowa study area. Note that the black circles indicate that the accuracies obtained using the TSM-adjusted percentiles are significantly (at the 5% level) different from the accuracies obtained using the original percentiles.



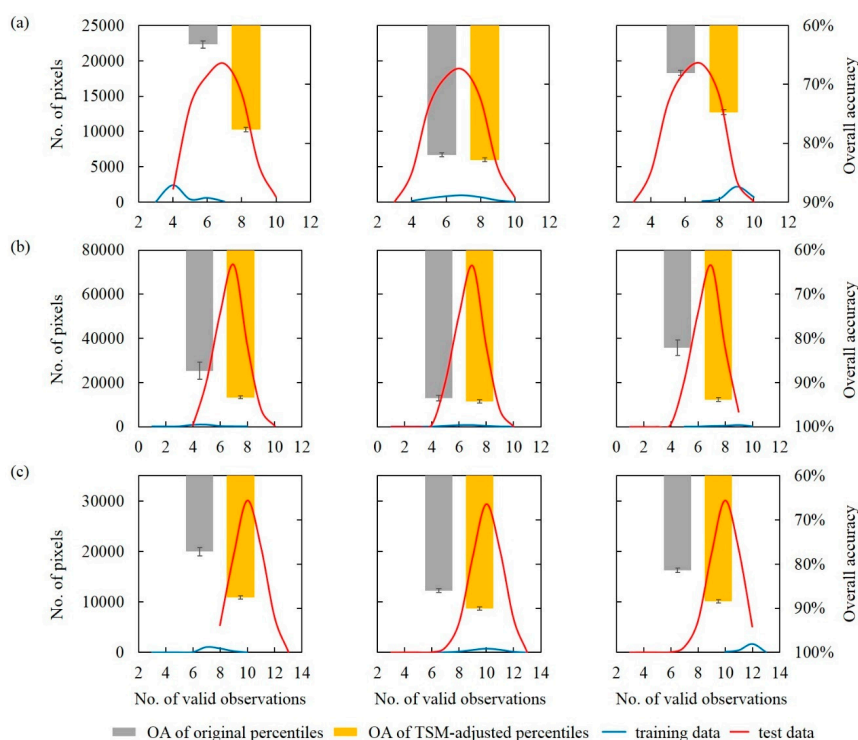
**Figure 12.** The average PA and UA of each land cover type against the number of valid observations in New York study area. Note that the black circles indicate that the accuracies obtained using the TSM-adjusted percentiles are significantly (at the 5% level) different from the accuracies obtained using the original percentiles.

### 5.3. Effect of Training Data Sampling

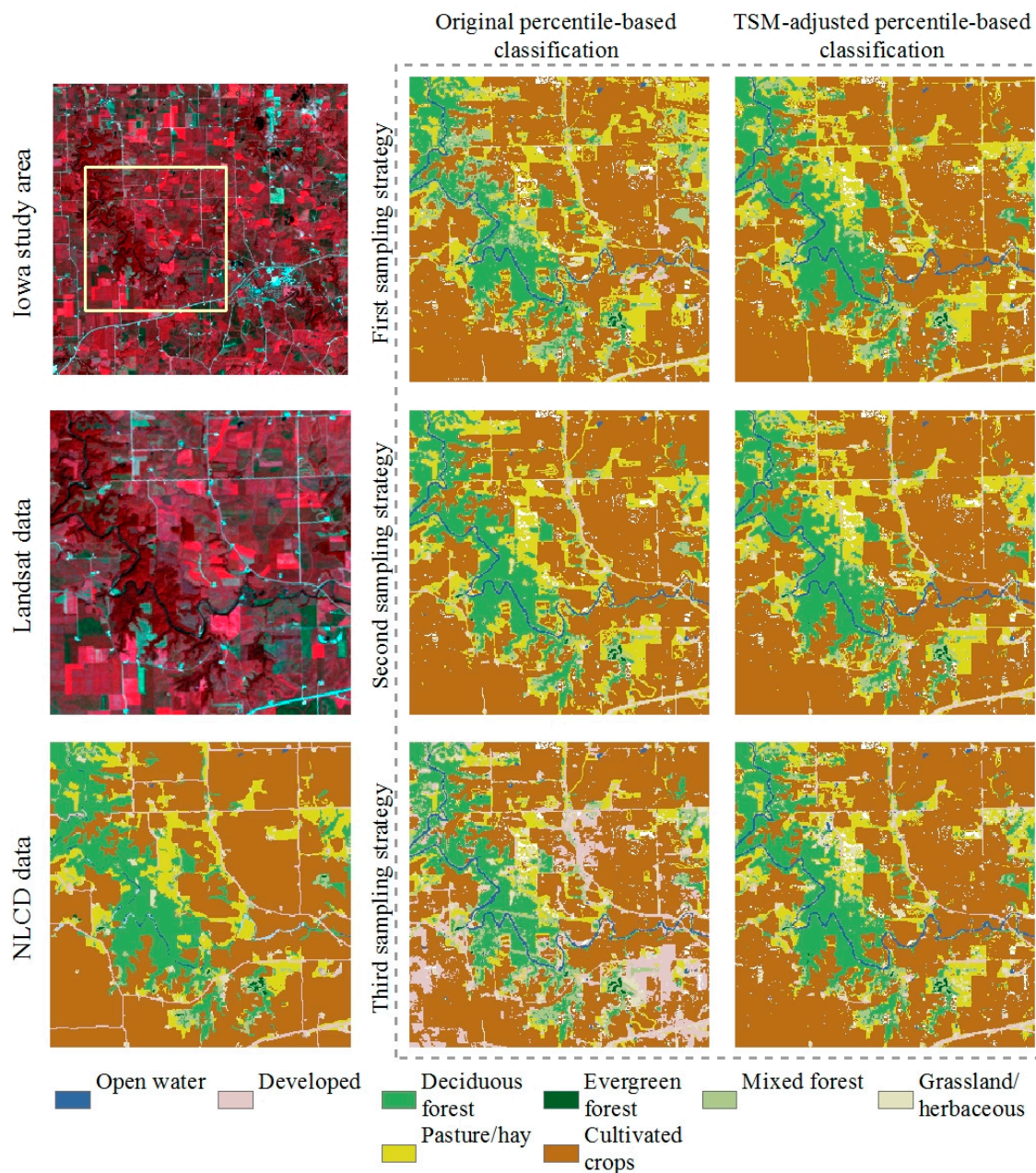
In this section, the effect of training data sampling on overall classification accuracy obtained using percentile features was investigated. In addition to the sampling strategy of training data (i.e., random sampling across valid observation frequency stratum for each class) used in the previous experiment, another two sampling strategies were adopted: one is random sampling of training pixels limited to locations with high observation frequencies, another is random sampling of training pixels limited to locations with low observation frequencies. Due to the spatially random intersection of land cover and valid observation frequency stratum, the superior percentile features should be robust to sampling strategies of training data.

Figure 13 illustrated the overall classification accuracy provided by original and TSM-adjusted percentiles using the three different sampling strategies in the three study areas. All the overall

accuracies obtained using TSM-adjusted percentiles were consistently significantly higher (at the 5% level) than that obtained using original percentiles. The great variation of OA arising from different sampling strategies for original percentiles demonstrated that the pixel-wise uncertainty of valid observation frequency leads to a certain level of bias in original percentile features. In contrast, the OA variation induced by sampling strategy for TSM-adjusted percentiles was less than that for original percentiles; this result suggested that the proposed TSM-adjusted percentiles were robust to different sampling strategies. It is mainly because the valid observation frequencies, from which the TSM-adjusted percentiles were derived, for the entire classified study area were harmonized using time series model estimation. The greatest improvement with the proposed method was seen in the scenarios of the first and third sampling strategies. This also could be supported by some close-ups illustrated in Figure 14. As shown in Figure 14, the spatial patterns of classification results derived from TSM-adjusted percentiles were consistent regardless of the sampling method used. In contrast, the original percentiles-based classifications obtained using different sampling strategies have obvious difference. Based on the visual consistency with the reference data of NLCD and Landsat, the largest improvement using the proposed TSM-adjusted percentiles in Iowa study area could be observed from the third scenario, followed by the first scenario, and the second scenario was least. This result suggests that the large range of variation in valid observation frequency for the training and test data provides more opportunities for the robustness to missing observations provided by the TSM-adjusted percentiles to become evident. In addition, it was worth noting that the OA difference between the two sets of classification was substantially reduced using the second strategy. It is demonstrated that the random sampling across valid observation frequency stratum for each land cover class could slightly compensate for the bias of original percentile features caused by pixel-wise uncertainty of valid observation frequency.



**Figure 13.** Sensitivity of the overall classification accuracy to different sampling strategies of training data in Minnesota (a), Iowa (b), and New York (c) study area. Left column: random sampling of training pixels limited to locations with low observation frequencies. Middle column: random sampling across valid observation frequency stratum for each class (used in this paper). Right column: random sampling of training pixels limited to locations with high observation frequencies.



**Figure 14.** The close-ups of exemplified Iowa study area for a spatial comparison between original percentile-based and TSM-adjusted percentile-based classification using the three sampling strategies.

## 6. Conclusions

Due to gaps resulting from clouds, cloud shadows, snow, and the SLC-off problem in Landsat time-series, the temporal distribution of valid observations is highly uneven across different pixels. Therefore, the percentile features derived from such time-series data may not provide the accurate characterization of land covers. The intent of this study was to propose the TSM-adjusted percentile to enhance the performance of original percentiles in characterizing land covers, and then improve the accuracy of land cover classification based on these metrics.

The proposed method was implemented in time-series of Landsat data covering three study areas wherein time-series had a considerable amount of missing data. The classification results were compared with those obtained using the original percentiles. According to the experimental results, the main findings of the work were highlighted:

- (i) The land-cover classifications obtained using the proposed TSM-adjusted percentiles had significantly higher overall accuracies than those obtained using the original percentiles.
- (ii) The TSM-adjusted percentile features were more effective for forest types with obvious phenological characteristics and with less valid observations.
- (iii) The performance of TSM-adjusted percentiles was robust to the training data sampling strategy. The performance difference between the two sets of results was alleviated when using the random sampling across valid observation frequency strata for each land cover class.

The proposed method also has limitations. First, the continuous change detection algorithm based on all available Landsat data was computationally complicated, the out of memory or computation time out problems may occur for the large area application although CCD algorithms are available within GEE platform. Second, future work could focus on using the cross-platform images (for example, Landsat and Sentinel-2 together). An increase number of valid observations benefit from the combined using of multi-platform data would minimize the issue of the original percentiles, although the Sentinel-2 data are only available after the year 2015.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2072-4292/12/18/3091/s1>, Figure S1: Maps with only pixels can be selected for training/testing. Black indicated the pixels which were removed from NLCD data using spatio-temporal filtering methods. Left: Minnesota study area; middle: Iowa study area; and right: New York study area, Figure S2: Histograms showing observations of each study area. Left: Minnesota study area; middle: Iowa study area; and right: New York study area, Figure S3: Maps showing the time series models used for surface reflectance estimation in 2011. White shows the pixel locations where land-cover changes occurred. Left: Minnesota study area; middle: Iowa study area; and right: New York study area.

**Author Contributions:** Conceptualization, L.L.; methodology, S.X. and J.Y.; software, S.X.; validation, S.X.; investigation, S.X. and J.Y.; resources, J.Y.; writing—original draft preparation, S.X.; and writing—review and editing, L.L. and S.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Strategic Priority Research Program of the Chinese Academy of Sciences, grant number XDA19090200, the Key Research Program of the Chinese Academy of Sciences (ZDRW-ZS-2019-1), and the National Natural Science Foundation of China, grant number 41825002.

**Acknowledgments:** The MATLAB versions of CCDC codes developed by Global Environmental Remote Sensing (GERS) Laboratory at University of Connecticut, were downloaded from the GitHub page (<https://github.com/GERSL/CCDC>). The authors thank the editor and anonymous reviewers for their valuable comments and suggestions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Rodriguez-Galiano, V.F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; Rigol-Sanchez, J.P. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.* **2012**, *67*, 93–104.
2. Pereira, H.M.; Ferrier, S.; Walters, M.; Geller, G.N.; Jongman, R.; Scholes, R.J.; Bruford, M.W.; Brummitt, N.; Butchart, S.; Cardoso, A. Essential biodiversity variables. *Science* **2013**, *339*, 277–278.
3. Baker, C.; Lawrence, R.L.; Montagne, C.; Patten, D. Change detection of wetland ecosystems using Landsat imagery and change vector analysis. *Wetlands* **2007**, *27*, 610–619.
4. Masek, J.G.; Huang, C.; Wolfe, R.; Cohen, W.; Hall, F.; Kutler, J.; Nelson, P. North American forest disturbance mapped from a decadal Landsat record. *Remote Sens. Environ.* **2008**, *112*, 2914–2926.
5. Pickell, P.D.; Andison, D.W.; Coops, N.C.; Gergel, S.E.; Marshall, P.L. The spatial patterns of anthropogenic disturbance in the western Canadian boreal forest following oil and gas development. *Can. J. For. Res.* **2015**, *45*, 732–743.
6. Pflugmacher, D.; Cohen, W.B.; Kennedy, R.E. Using Landsat-derived disturbance history (1972–2010) to predict current forest structure. *Remote Sens. Environ.* **2012**, *122*, 146–165.
7. Zhu, Z.; Woodcock, C.E. Continuous change detection and classification of land cover using all available Landsat data. *Remote Sens. Environ.* **2014**, *144*, 152–171.

8. DeFries, R.; Hansen, M.; Townshend, J. Global discrimination of land cover types from metrics derived from AVHRR pathfinder data. *Remote Sens. Environ.* **1995**, *54*, 209–222.
9. Gebhardt, S.; Wehrmann, T.; Ruiz, M.A.M.; Maeda, P.; Bishop, J.; Schramm, M.; Kopeinig, R.; Cartus, O.; Kellndorfer, J.; Ressler, R. MAD-MEX: Automatic wall-to-wall land cover monitoring for the Mexican REDD-MRV program using all Landsat data. *Remote Sens.* **2014**, *6*, 3923–3943.
10. Petitjean, F.; Inglada, J.; Gançarski, P. Satellite image time series analysis under time warping. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3081–3095.
11. Reed, B.C.; Brown, J.F.; VanderZee, D.; Loveland, T.R.; Merchant, J.W.; Ohlen, D.O. Measuring phenological variability from satellite imagery. *J. Veg. Sci.* **1994**, *5*, 703–714.
12. Azzari, G.; Lobell, D. Landsat-based classification in the cloud: An opportunity for a paradigm shift in land cover monitoring. *Remote Sens. Environ.* **2017**, *202*, 64–74.
13. Hansen, M.C.; Egorov, A.; Roy, D.P.; Potapov, P.; Ju, J.; Turubanova, S.; Kommareddy, I.; Loveland, T.R. Continuous fields of land cover for the conterminous United States using Landsat data: First results from the Web-Enabled Landsat Data (WELD) project. *Remote Sens. Lett.* **2011**, *2*, 279–288.
14. Zhang, H.K.; Roy, D.P. Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification. *Remote Sens. Environ.* **2017**, *197*, 15–34.
15. Wessels, K.J.; Van den Bergh, F.; Roy, D.P.; Salmon, B.P.; Steenkamp, K.C.; MacAlister, B.; Swanepoel, D.; Jewitt, D. Rapid land cover map updates using change detection and robust random forest classifiers. *Remote Sens.* **2016**, *8*, 888.
16. Kovalskyy, V.; Roy, D.P. The global availability of Landsat 5 TM and Landsat 7 ETM+ land surface observations and implications for global 30 m Landsat data product generation. *Remote Sens. Environ.* **2013**, *130*, 280–293.
17. Hermosilla, T.; Wulder, M.A.; White, J.C.; Coops, N.C.; Hobart, G.W. Regional detection, characterization, and attribution of annual forest change from 1984 to 2012 using Landsat-derived time-series metrics. *Remote Sens. Environ.* **2015**, *170*, 121–132.
18. Kennedy, R.E.; Yang, Z.; Cohen, W.B. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—Temporal segmentation algorithms. *Remote Sens. Environ.* **2010**, *114*, 2897–2910.
19. Hamunyela, E.; Verbesselt, J.; Herold, M. Using spatial context to improve early detection of deforestation from Landsat time series. *Remote Sens. Environ.* **2016**, *172*, 126–138.
20. Zhu, Z.; Woodcock, C.E.; Olofsson, P. Continuous monitoring of forest disturbance using all available Landsat imagery. *Remote Sens. Environ.* **2012**, *122*, 75–91.
21. Zhu, Z.; Woodcock, C.E.; Holden, C.; Yang, Z. Generating synthetic Landsat images based on all available Landsat data: Predicting Landsat surface reflectance at any given time. *Remote Sens. Environ.* **2015**, *162*, 67–83.
22. Loveland, T.R.; Dwyer, J.L. Landsat: Building a strong future. *Remote Sens. Environ.* **2012**, *122*, 22–29.
23. Teillet, P.; Barker, J.; Markham, B.; Irish, R.; Fedosejevs, G.; Storey, J. Radiometric cross-calibration of the Landsat-7 ETM+ and Landsat-5 TM sensors based on tandem data sets. *Remote Sens. Environ.* **2001**, *78*, 39–54.
24. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27.
25. Masek, J.G.; Vermote, E.F.; Saleous, N.E.; Wolfe, R.; Hall, F.G.; Huemmrich, K.F.; Gao, F.; Kutler, J.; Lim, T.-K. A Landsat surface reflectance dataset for North America, 1990–2000. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 68–72.
26. Foga, S.; Scaramuzza, P.L.; Guo, S.; Zhu, Z.; Dille, Jr, R.D.; Beckmann, T.; Schmidt, G.L.; Dwyer, J.L.; Hughes, M.J.; Laue, B. Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sens. Environ.* **2017**, *194*, 379–390.
27. Zhu, Z.; Woodcock, C.E. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* **2012**, *118*, 83–94.
28. Wickham, J.; Homer, C.; Vogelmann, J.; McKerrow, A.; Mueller, R.; Herold, N.; Coulston, J. The multi-resolution land characteristics (MRLC) consortium—20 years of development and integration of USA national land cover data. *Remote Sens.* **2014**, *6*, 7424–7441.

29. Homer, C.; Dewitz, J.; Yang, L.; Jin, S.; Danielson, P.; Xian, G.; Coulston, J.; Herold, N.; Wickham, J.; Megown, K. Completion of the 2011 National Land Cover Database for the conterminous United States—representing a decade of land cover change information. *Photogramm. Eng. Remote Sens.* **2015**, *81*, 345–354.
30. Zhou, Q.; Tollerud, H.J.; Barber, C.P.; Smith, K.; Zelenak, D. Training Data Selection for Annual Land Cover Classification for the Land Change Monitoring, Assessment, and Projection (LCMAP) Initiative. *Remote Sens.* **2020**, *12*, 699.
31. Wickham, J.; Stehman, S.V.; Gass, L.; Dewitz, J.A.; Sorenson, D.G.; Granneman, B.J.; Poss, R.V.; Baer, L.A. Thematic accuracy assessment of the 2011 national land cover database (NLCD). *Remote Sens. Environ.* **2017**, *191*, 328–341.
32. Zhu, Z.; Wang, S.; Woodcock, C.E. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sens. Environ.* **2015**, *159*, 269–277.
33. Zhu, Z.; Woodcock, C.E. Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change. *Remote Sens. Environ.* **2014**, *152*, 217–234.
34. Friedman, J.; Hastie, T.; Höfling, H.; Tibshirani, R. Pathwise coordinate optimization. *Ann. Appl. Stat.* **2007**, *1*, 302–332.
35. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **2010**, *33*, 1–22.
36. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* **1996**, *58*, 267–288.
37. Hansen, M.; Egorov, A.; Potapov, P.; Stehman, S.; Tyukavina, A.; Turubanova, S.; Roy, D.P.; Goetz, S.; Loveland, T.R.; Ju, J. Monitoring conterminous United States (CONUS) land cover change with web-enabled Landsat data (WELD). *Remote Sens. Environ.* **2014**, *140*, 466–484.
38. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.
39. Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31.
40. Zhu, Z.; Gallant, A.L.; Woodcock, C.E.; Pengra, B.; Olofsson, P.; Loveland, T.R.; Jin, S.; Dahal, D.; Yang, L.; Auch, R.F. Optimizing selection of training and auxiliary data for operational land cover classification for the LCMAP initiative. *ISPRS J. Photogramm. Remote Sens.* **2016**, *122*, 206–221.
41. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46.
42. Brooks, E.B.; Thomas, V.A.; Wynne, R.H.; Coulston, J.W. Fitting the multitemporal curve: A Fourier series approach to the missing data problem in remote sensing analysis. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3340–3353.
43. Ju, J.; Roy, D.P. The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally. *Remote Sens. Environ.* **2008**, *112*, 1196–1211.
44. Kovalsky, V.; Roy, D.P.; Zhang, X.Y.; Ju, J. The suitability of multi-temporal web-enabled Landsat data NDVI for phenological monitoring—a comparison with flux tower and MODIS NDVI. *Remote Sens. Lett.* **2012**, *3*, 325–334.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).