Remote Sens. **2014**, *6*, 4907-4926; doi:10.3390/rs6064907

OPEN ACCESS

remote sensing

ISSN 2072-4292

www.mdpi.com/journal/remotesensing

Article

Automated Detection of Cloud and Cloud Shadow in Single-Date Landsat Imagery Using Neural Networks and Spatial Post-Processing

M. Joseph Hughes [1,*] and Daniel J. Hayes [1,2]

[1] Department of Ecology and Evolutionary Biology, University of Tennessee Knoxville, Knoxville, TN 37996, USA

[2] Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA; E-Mail: hayesdj@ornl.gov

* Author to whom correspondence should be addressed; E-Mail: jhughes@utk.edu; Tel.: +1-865-974-9760.

Received: 22 February 2014; in revised form: 15 May 2014 / Accepted: 19 May 2014 / Published: 28 May 2014

Abstract: The use of Landsat data to answer ecological questions is greatly increased by the effective removal of cloud and cloud shadow from satellite images. We develop a novel algorithm to identify and classify clouds and cloud shadow, SPARCS: Spatial Procedures for Automated Removal of Cloud and Shadow. The method uses a neural network approach to determine cloud, cloud shadow, water, snow/ice and clear sky classification memberships of each pixel in a Landsat scene. It then applies a series of spatial procedures to resolve pixels with ambiguous membership by using information, such as the membership values of neighboring pixels and an estimate of cloud shadow locations from cloud and solar geometry. In a comparison with FMask, a high-quality cloud and cloud shadow classification algorithm currently available, SPARCS performs favorably, with substantially lower omission errors for cloud shadow (8.0% and 3.2%), only slightly higher omission errors for clouds (0.9% and 1.3%, respectively) and fewer errors of commission (2.6% and 0.3%). Additionally, SPARCS provides a measure of uncertainty in its classification that can be exploited by other algorithms that require clear sky pixels. To illustrate this, we present an application that constructs obstruction-free composites of images acquired on different dates in support of a method for vegetation change detection.

Keywords: cloud detection; Landsat; image analysis; neural networks

1. Introduction

The Landsat archive provides an unprecedented opportunity to discover how our landscape has changed over the last 30 years. Much of the imagery, however, is contaminated with clouds and their associated shadows, particularly in the tropics and other forested areas with high transpiration [1]. Therefore, the usefulness of this imagery for landscape change studies depends on reliably separating clear sky regions from those obstructed by clouds and cloud shadow. Because of the large number of scenes over multiple dates needed for such studies, accurate and reliable automated methods are essential for this task.

Significant work has been devoted to cloud and cloud shadow identification. Many algorithms for cloud and cloud shadow masking have been developed for other sensors, particularly for AVHRR [2–5] and MODIS [6–8]. Some of these algorithms have then been adapted for use on Landsat data (see, for example, Oreopoulos *et al.*, 2011 [9]). Since clouds are bright and cold and cloud shadows are darker than the surrounding landscape, a common approach is to apply a threshold to the spectral values [10] or some simple function of two or more spectral values [9,11]. An early example is the automatic cloud cover assessment (ACCA) [12], which uses a series of successive thresholds over bands and band combinations to define a hierarchical set of rules for clouds. ACCA was not designed for precise spatial detection of clouds, but rather, to estimate the percentage of cloud cover in a given Landsat scene. These methods operate over imagery acquired on a single date; multi-temporal methods can leverage additional data to detect clouds and their shadows [13,14]. Since clouds are typically bright objects in a scene and shadows necessary darken an area, the obstructions can be identified by looking for outliers from a reference scene. In an automated approach, however, where the reference scene must be selected algorithmically, this is only effective when a good cloud detection method for single-date imagery is already in use or when most images are cloud-free and therefore clouds and cloud shadows are outliers from the mean, an assumption that may not hold in areas with frequent cloud cover.

Cloud shadows are more difficult to identify than clouds, because the spectral information does not discriminate between shadows caused by clouds and shadows arising from other causes, such as terrain. Additionally, other dark land covers, such as dark vegetation or water bodies, have similar spectral signatures to shadows. To address this confusion, the cloud mask itself has been used to distinguish cloud shadows, using the known sensor and solar geometry to estimate where cloud shadows should occur given the location of clouds [5,10,11,15–18]. Such approaches are sensitive to the height of clouds above the land surface, as this height is proportional to the two-dimensional distance of a cloud from its shadow, as seen in imagery. Cloud height can be estimated using the thermal band [17,18], and combining this with a digital elevation model, as in the algorithm from Huang *et al.* [17], can further reduce error.

In this paper, we develop a novel method for identifying clouds from Landsat TM and ETM+ imagery: Spatial Procedures for Automated Removal of Cloud and Shadow (SPARCS). This development was motivated by our need for efficient and reliable cloud and cloud shadow masking in a forest change detection application over highly heterogeneous land cover in the eastern U.S.; existing methods were either too computationally intensive or missed many clouds or cloud shadows, which were detected as change. Four design objectives directed our development. First, the method should only use bands

present on all sensors and avoid ancillary data sources in order to ensure that the method is applicable over the entire archive. Second, the method should be completely automated and free from operator input. Third, the method should be sufficiently computationally efficient to be applied over thousands of scenes. Finally, the method should provide a spatially-explicit measure of classifier certainty that can be propagated to the products relying on the resultant cloud and cloud shadow masks, a feature we are unaware of in other cloud detection algorithms.

To meet these goals, we use neural network classifiers [19] to explore different methods of using spatial information contained in a single-date Landsat scene to address the cloud and cloud shadow detection problem. These classifiers are trained using scenes with clouds and cloud shadow labeled by human operators at USGS [20] and evaluated using additional manually labeled data. Using this evaluation, we choose a high-quality classifier to become the basis of SPARCS and apply a series of spatial post-processing procedures to resolve ambiguous pixels in the classifier outputs. We then compare SPARCS to a high-quality, commonly used method, FMask [18]. Like FMask, we also include a class for water and snow/ice, for completeness. Finally, we demonstrate the utility of our method using an example that exploits the classifier uncertainty provided by SPARCS to combine a multi-temporal image stack into an obstruction-free composite.

2. Background

Neural networks are non-linear supervised learning algorithms that can be trained to partition an input space into a set of classes. Neural networks work by learning h linear combinations of the input data, where h is determined by the operator, and passing each of these through a given non-linear thresholding function. These results are temporarily stored as hidden values. Then, the network repeats the process by taking c linear combinations of those hidden values, where c is the number of desired classes, and again passing them through a given non-linear thresholding function. These results are then interpreted as the input observation's membership in each output class and are wholly dependent on the weights of the linear combinations at both stages. These memberships are continuous values between zero and one; if desired, a crisp classification can be performed by setting the highest membership value to one and all others to zero. The weights themselves are learned using an optimization procedure over a training dataset that contains observations labeled with their correct class. The correctness of the labels in this training dataset directly controls the quality of the resulting classifier. Additionally, a higher number of hidden values (h) allows more complex patterns in the input data to be learned, though it also increases the likelihood of learning spurious correlations in the training data and thereby reducing the generality of the classifier [19].

Importantly, the input data must include non-ambiguous information about the desired output classes to be able to discriminate between observations. This, however, is not the case when attempting to identify cloud and cloud shadow from aspatial Landsat pixel data over a wide range of land cover types, as the exact same spectral data can be associated with pixels of clouds, snow or some other bright and cold feature, or associated with cloud shadow, terrain shadow, water bodies or some other dark terrain feature. In short, pixel data by itself is ambiguous. As such, an additional source of information is needed to resolve these cases, such as elevation data, which is useful for distinguishing cloud shadow from

terrain shadow, observations from multiple time periods to filter ephemeral values or spatial relationships between pixels. We are most interested in harnessing spatial information, because it is already present in the Landsat scene, and because human classifiers can almost always visually identify clouds and cloud shadows within a scene, spatial information should be sufficient for discrimination. In addition to applying spatial adjustments to the classifier output in a post-processing step, we examine two simple methods for incorporating space into neural network inputs. The intuition behind both methods is that, by providing an estimate of "average value" in a region, ambiguity caused by variation within objects could be reduced, creating a simpler problem for the neural network to learn. The first method is to simply calculate the mean spectral value within a neighborhood around each pixel. The second method uses the pixel values from the image after denoising using total variation regularization (TVR) [21]. TVR removes noise from an image, f, by finding a new image, u, that minimizes the functional:

$$\min_u \|u - f\|^2 + 2\alpha |\nabla u| \quad (1)$$

The first term is the sum of squared-errors between the new image and the original image; minimizing prevents the new image from diverging too far from the original. The second term measures the magnitude of differences between adjacent pixels; minimizing favors outputs with similar adjacent values. Taken together, the method balances smoothing parts of the image with keeping original details. How this balance is struck is determined by α, which should be positive, with smaller values favoring more detail and larger values favoring more smoothing. Due to the nature of the gradient term [21,22], the smoothing manifests as regions of the image with constant values and sudden discontinuous jumps at the edges. This constant value can be thought of as an average of the spectral values within the region.

3. Methods

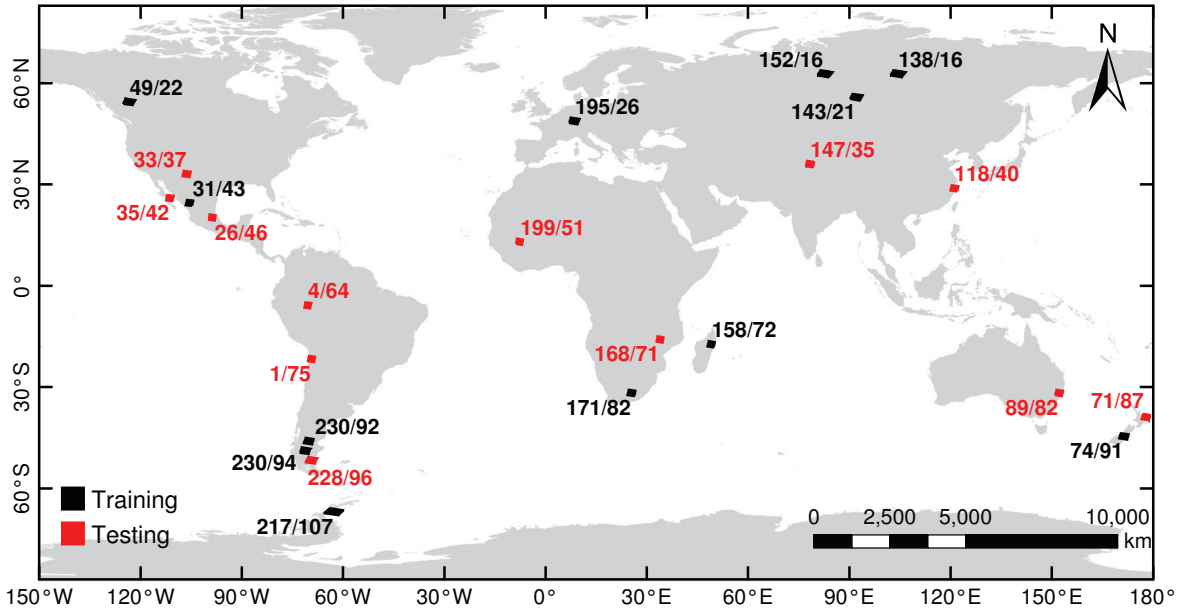
3.1. Data Sets

Manually-generated cloud masks from the USGS LDCM Cloud Cover Assessment Data [20] were subset, refined and used as a training dataset for several different neural network classifiers. The core-cloud and thin-cloud classes, which were separate in the USGS data, were combined into a single cloud class. Of the 157 scenes in the dataset, only the 18 scenes with more than 1% of pixels in both the cloud and cloud shadow classes were considered. From these, twelve representative scenes (Figure 1) from different hemispheres and latitudes were selected for the training dataset; the other six scenes had inadequately accurate cloud and cloud shadow masks for training. To reduce the amount of data used, while retaining land cover variability within scenes, four 1000 $\times$ 1000 pixel (30 $\times$ 30 km) regions, each separated by at least 2000 (60 km) pixels, were subset from each of these twelve scenes. These 48 sub-scenes were used for training only.

The 48 sub-scene masks were visually assessed for classification label accuracy. The USGS masks provided classes for clear sky, cloud, and cloud shadow. Additionally, water and snow/ice classes were added to the masks. To do so, the Normalized Difference Snow Index (NDSI) [23] and tassel-cap brightness [24] were used to locate potential water and snow/ice regions in the imagery. These proposed masks were then hand-edited to include regions missed by the thresholding and to remove regions incorrectly included. The water and snow/ice masks were then combined with the USGS masks using

the follow precedence rules. Pixels labeled as cloud in the USGS mask were unchanged. Pixels labeled clear sky in the USGS masks and flagged as either water or snow/ice were changed to water or snow/ice, as appropriate. Pixels labeled cloud shadow in the USGS masks and flagged as water were hand set appropriately. Cloud shadows over water were labeled as cloud shadow where a distinction could be made.

Figure 1. WRS2 path/row locations of imagery used in training (black) and testing (red).



Twelve additional scenes were selected for testing, and one 1000 × 1000 pixel (30 × 30 km) sub-scene was extracted from each. These 12 sub-scenes were then clustered over their spectral data using k-means clustering [25] with 100 cluster seeds, which sorts pixels into 100 groups with similar spectral values. Each resulting cluster was assigned to one of the five classes by an operator. Then, mislabeled image regions were hand-corrected using image editing software. These 12 sub-scenes were used only in classifier assessment, and not during training, in order to reduce the risk of neural network over-fitting during comparisons. All analyses comparing different networks and methods used only these testing scenes.

Landsat 7 ETM+ Level 1T imagery and metadata for all scenes used in training and testing were acquired from the USGS archive. All scenes were from 2001, before the scan line corrector failure. The Landsat ETM+ data in Bands 1–5 and 7 in each sub-scene were corrected to top-of-atmosphere reflectance [26,27] and then further corrected using dark-object subtraction [28]. The low-gain scaling of the thermal band (B_6) was converted to brightness temperature and then arbitrarily rescaled to values near the other bands to facilitate neural network learning:

$$\hat{B}_6 = B_6/100 - 2 \tag{2}$$

These values were used in all analyses.

3.2. Neural Network Classification

A total of 15 neural network configurations were used to explore the role of classifier complexity and the inclusion of spatial information on classification accuracy (Table 1). Networks with 10, 20 and 30 hidden nodes were constructed using five different types of spatial inputs. In addition, all configurations included the aspatial spectral information (ETM+ Bands 1–5, 7) and rescaled brightness-temperature (ETM+ Band 6) for each individual pixel. The first spatial input type added no spatial information and was used for baseline comparison. The second through fifth spatial input types added spatial information summarized from the first three components of the tassel-cap transformation [24]. Spatial input types two and three added the local average in a region around each pixel in the three tassel-cap bands, using a 5×5 and 9×9 pixel neighborhood, respectively. Spatial input type four and five used tassel-cap pixel values after removing spatial noise using TVR [21,22] with $\alpha = 0.05$ and $\alpha = 0.10$, respectively, which remove amounts of detail similar to the local averaging neighborhoods. The five types therefore represent a no-space baseline, plus two spatial-averaging methods, each using two intensities.

Table 1. All network configurations included aspatial ETM+ bands, but varied in network size (h) and the type of spatial inputs. TVR, total variation regularization.

#	Network Size (h)	Spatial Averaging Method (Calculated over Tassel-Cap)	Intensity of Spatial Averaging
1	10	No Space	-
2	10	Local Average	5×5 Window
3	10	Local Average	9×9 Window
4	10	TVR	$\alpha = 0.05$
5	10	TVR	$\alpha = 0.10$
6	20	No Space	-
7	20	Local Average	5×5 Window
8	20	Local Average	9×9 Window
9	20	TVR	$\alpha = 0.05$
10	20	TVR	$\alpha = 0.10$
11	30	No Space	-
12	30	Local Average	5×5 Window
13	30	Local Average	9×9 Window
14	30	TVR	$\alpha = 0.05$
15	30	TVR	$\alpha = 0.10$

Training data for the neural networks was randomly sampled from the 48 training sub-scenes after stratifying each sub-scene by class (cloud shadow, cloud, water, snow/ice, clear sky). Where possible, 1500 pixels from each class were selected from each sub-scene. If a sub-scene did not have 1500 pixels of a class, all pixels of that class were selected. The aspatial spectral and brightness-temperature data, the 5×5 and 9×9 pixel averages in the tassel-cap indices and the two TVR-denoised values of the tassel-cap indices were extracted for each pixel. This process was performed three times, with different stratified random samples selected each time, to generate three sets of training samples. A total of

166,639 samples were used for each network. Each network configuration was trained using each training set, generating a total of 45 networks consisting of the 15 configurations replicated three times.

Networks were trained using scaled conjugate gradient backpropagation by the MATLAB Neural Network Toolbox [29] *patternnet*() function.

3.3. Spatial Post-Processing

Clouds and cloud shadows are spatially-coherent objects in satellite imagery. SPARCS uses information on surrounding pixels to exploit this spatial coherency for the purposes of reducing classification error. In exploratory classifications using several different neural networks, error maps were generated to visually assess the spatial patterns of errors. From these observations, a series of six rules were developed to address spatially-definable error. Each of these rules operates over the continuous-valued membership images for each class. First, a 3×3 median filter is applied to the cloud and cloud shadow membership images to reduce noise.

The second rule addresses confusion at water-land boundaries. Shallow water is often confused for clouds or snow/ice, and the wet soil in the transition zone between water and land is often confused with cloud shadow. To correct this, the cloud, cloud shadow and snow-ice membership of pixels within three pixels of large bodies of water are decreased.

The third rule uses sun and sensor geometry to identify areas of potential cloud shadow using the cloud membership to reduce the significant ambiguity between hill shade, wet ground and cloud shadow. Our approach closely resembles the method of Luo *et al.* [8] in that it defines a broad area of potential cloud shadow that accounts for a range of potential cloud heights and then combines it with an estimate based on spectral values. Our method uses the cloud shadow membership values for the initial estimate. The direction of the sun is first determined from scene metadata. Then, a copy of the cloud membership image is transposed away from the sun a distance determined by the sun elevation and a cloud height of 2250 m and then expanded (dilated) to include potential cloud heights from 1800 m to 2700 m above the ground, a height range chosen to capture the dark shadows created by the optically thick cumulus clouds. This potential cloud shadow location is then further expanded and blurred using a 15 × 15 pixel filter to create a feasible zone of cloud shadow. Finally, this estimate is multiplied with the cloud shadow membership to increase cloud shadow membership within the feasible zone and to reduce the cloud shadow membership in areas outside of the feasible zone that likely represent terrain shadow erroneously classified as cloud shadow.

The fourth rule addresses confusion between water and deep shadow, which can have equivalent spectral signatures. Pixels that have similar memberships in both cloud shadow and water, meaning that the neural network classifier has identified them as ambiguous, are selected. Those that are also surrounded by pixels of high cloud shadow membership then have their own cloud shadow membership increased and their water membership decreased.

The fifth rule performs a similar function between clouds and snow/ice, which can have similar ambiguity, biasing membership toward clouds and away from snow/ice in pixels surrounded by clouds.

The final rule identifies pixels of high overall uncertainty and uses the membership of nearby pixels to predict the correct membership of the uncertain pixels. First, uncertainty is calculated as the variance

between memberships, rescaled to be between 0 and 1. Then, a weighted average of nearby pixels is calculated for each membership class, with weights calculated as the product of each pixel's certainty and a Gaussian decay function over distance with $\sigma = 2$ pixels. Finally, new memberships are calculated as a linear combination of the original value and the spatial average, weighted using the pixel's uncertainty, such that more certain pixels retain their original value and uncertain pixels become more like the average value of pixels around them. This rule has the effect of homogenizing areas and removing noise.

3.4. Classifier Assessment

Each of the 45 neural networks was scored on each of the 12 evaluation sub-scenes based on the total classifier accuracy. For each pixel, the assigned class from the network was taken as the class with the maximum membership value over all classes. These classes were compared to the evaluation masks described in Section 3.1. Because clouds and cloud shadows are not discrete objects, there are many semi-obstructed pixels that form a transition zone between cloud or cloud shadow and clear sky. Since we are more concerned with identifying potential clouds and cloud shadows than with precisely defining their extent, a three-pixel buffer around the areas labeled as cloud and cloud shadow in the evaluation masks was constructed. Pixels within this buffer were scored as correct if they were labeled as either cloud or cloud shadow, as appropriate, or the label in the evaluation mask. This reduced commission errors and increased the overall accuracy for all methods, including FMask, by approximately 2%.

A multiway ANOVA [30] was performed on the 540 accuracy scores to assess the contributions of classifier complexity and the type of spectral information to accuracy, while accounting for the variations between sub-scenes. Networks with more inputs or more hidden nodes were not penalized for additional model complexity, since the purpose was to find the most effective classifier and not necessarily the most efficient.

After selecting a network to serve as the basis of our method, we then compared our method to FMask using the same classifier accuracy statistics derived from the testing data with buffered masks. We further compared the methods by calculating omission errors as the percentage of cloud or cloud shadow pixels that were mislabeled as clear sky and commission errors as the percentage of clear sky pixels, outside of the buffered area, mislabeled as clouds or cloud shadow. This was done, because FMask has significant confusion between clouds and cloud shadow, and we do not feel that including that confusion is relevant to the core question of whether the classifiers can separate clear sky pixels from obstructions.

3.5. Application: Obstruction-Free Summertime Composites

Though SPARCS can provide a crisp classification, wherein each pixel is labeled as exactly one class, by using the raw membership values, neural networks provide a measure of how certain the classifier is in assigning class membership. In this method, we use these original membership values to create clear sky composite images from a multi-temporal stack of Landsat TM scenes acquired within the same year. For the purposes of illustration, we selected four scenes acquired during late summer of 1990 from the same location in eastern Tennessee that had moderate cloudiness on visual inspection. Each of these scenes were classified using SPARCS to generate memberships in cloud, cloud shadow, water, snow/ice and clear sky classes.

For each scene, the pixel memberships in the water and clear sky classes were then combined to generate a clarity index (Q) for each pixel. Additionally, because including a marginally contaminated pixel is more harmful than excluding a marginally clear pixel, this index was squared.

$$Q = (m_W + m_L)^2 \tag{3}$$

where m_W is the water membership and m_L is the clear sky membership. For the purposes of this example, snow and ice are considered as obstructions, as snow is seasonal in the area of interest. In high altitude or latitude areas where snow and ice are persistent features, the class should be included.

In order to reduce the influence of phenology, scenes are also weighted by a Gaussian decay function of day-of-year, such that scenes further away from a given date are weighted less than scenes near that date. For this example, we chose a late summer day (day of year 225) and used a standard deviation of 30 days:

$$w_j = \exp\left(-\left(\frac{d_j - 255}{60}\right)^2\right) \tag{4}$$

where d_j is the day of year that the j-th scene was acquired. Other days could be chosen, and a comparison between composites weighted to different days could be fruitful to examine phenological effects, as long as care is taken that the decay function itself does not span significant phenological change.

Yearly summertime composites are then generated as a weighted average of all the summertime scenes each year, using the clarity index (Q) and the Gaussian-transformed distance from a target date (w_j) as weights. Each of the seven Landsat bands are computed independently:

$$A_b = \frac{\sum_{(j \in S)} B_{b,j} Q_j w_j}{\sum_j Q_j w_j} \tag{5}$$

where A_b is the composite image of band b, S is the set of selected scenes to be combined, $B_{b,j}$ is the image data of band b for scene j and Q_j and w_j are the weights for scene j described above.

4. Results and Discussion

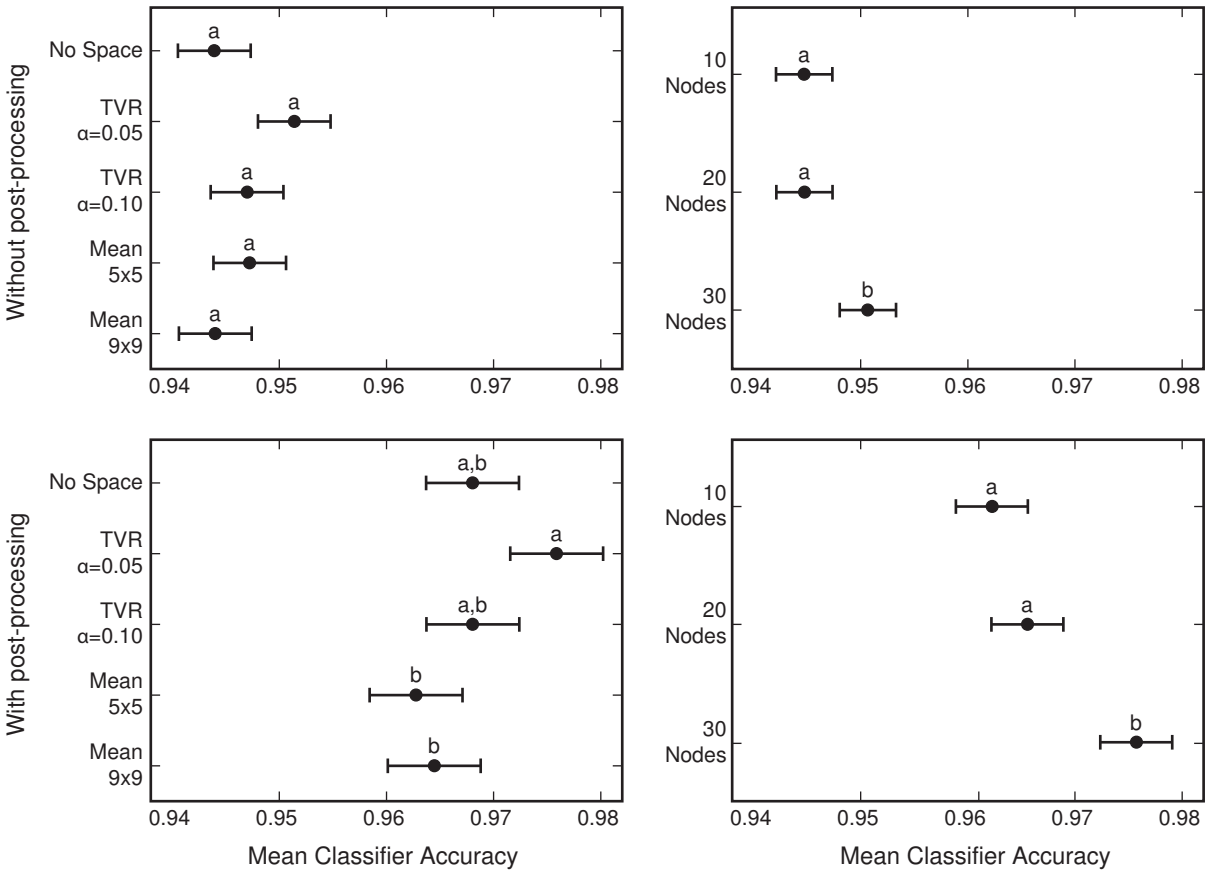
4.1. Network Selection

We trained 45 neural networks over 48 training sub-scenes and explored the effects of network size and the type of spatial inputs on classification accuracy over 12 evaluation sub-scenes. The inclusion of spatial inputs into the neural network had no statistically significant impact on classifier accuracy over the evaluation dataset. (Table 2). Network size was significant at the 0.05 confidence level; in a *post hoc* test using Tukey's honestly significant difference (HSD) criterion [31], statistically significant increases in total accuracy were seen in networks with 30 hidden nodes over those with 10 and 20 hidden nodes, which were similar (Figure 2). Absolute gains, though, were small; networks with 30 hidden nodes increased total accuracy by approximately 0.5% or about a 15% decrease in error. The interaction term between network size and sub-scene was also significant, suggesting that the increase in accuracy is due to the ability of more complex networks to learn additional land cover features and that more complex networks are not simply overfitting training data, but become more general classifiers.

Table 2. Multiway ANOVA of accuracy over the 12 evaluation sub-scenes for all neural networks, without applying post-processing spatial procedures.

Source	Sum Sq.	df	Mean Sq.	F	*p*
Type of Space	0.004	4	0.001	1.51	0.198
Network Size (h)	0.004	2	0.002	3.18	0.043
Sub-scene	1.004	11	0.091	137.26	<0.001
Space×Size	0.011	8	0.001	1.99	0.045
Space×Sub-scene	0.030	44	0.001	1.02	0.439
Size×Sub-scene	0.058	22	0.003	3.98	<0.001
Error	0.298	448	0.001		
Total	1.408	539			

Figure 2. Mean ranks of different methods for including spatial information in the network (**left**) and of different numbers of hidden nodes (**right**), before (**top**) and after (**bottom**) applying post-processing spatial procedures. Methods with the same letter are not significantly different by Tukey's honestly significant difference criterion.



After post-processing the neural network output with spatial procedures, classification accuracy increased overall, from approximately 94.5% to approximately 97%. The post-processing procedures also exaggerated differences between the methods of including spatial inputs to the neural network,

which became statistically significant by the ANOVA F-test (Table 3). Tukey's HSD separated the TVR method with $\alpha = 0.05$ from the methods that used a mean over a local neighborhood, with the TVR method with $\alpha = 0.10$ and the method with no spatial inputs being intermediate between the groups (Figure 2). However, the difference between the means of methods to incorporate space within the neural network are within 1% of total classifier accuracy, much less than the increase gained by the inclusion of spatial post-processing procedures. The spatial procedures preserved the patterns in accuracy between networks with different numbers of hidden nodes. Importantly, the application of spatial procedures greatly enhances the effectiveness of methods that have no spatial inputs to the neural network, suggesting that at least part of the classification rules learned by the networks that incorporated space are replicated by the post-processing spatial procedures. Given that calculating the spatial inputs, particularly TVR, is computationally expensive, there is no clear choice for an operational method. We selected a network with 30 hidden nodes and no spatial inputs to the neural network for further evaluation and to use in our cloud and cloud shadow detection package, SPARCS: Spatial Procedures for the Automated Removal of Cloud and Shadow.

Table 3. Multiway ANOVA of accuracy over the 12 evaluation sub-scenes for all neural networks after applying spatial post-processing procedures.

Source	Sum Sq.	df	Mean Sq.	F	*p*
Type of Space	0.011	4	0.003	2.53	0.040
Network Size (h)	0.018	2	0.009	8.17	<0.001
Sub-scene	1.271	11	0.116	106.37	<0.001
Space×Size	0.017	8	0.002	2.00	0.045
Space×Sub-scene	0.087	44	0.002	1.81	0.002
Size×Sub-scene	0.113	22	0.005	4.22	<0.001
Error	0.487	448	0.001		
Total	2.004	539			

4.2. Comparison to FMask

SPARCS compares favorably with FMask over the 12 sub-scene evaluation dataset (Table 4). The largest improvement is in correct identification of cloud shadow: SPARCS mislabels 3.2% of cloud shadow pixels as clear sky compared to FMask's 8.0%. Both methods perform well at identifying clouds, with FMask performing somewhat better by mislabeling 0.9% of cloud pixels as clear sky compared to with SPARCS mislabeling 1.3%. Considering errors of commission, SPARCS performs substantially better by mislabeling 0.5% of clear sky pixels as cloud shadow and 0.2% as clouds, compared to FMask mislabeling 2.4% of clear sky pixels as cloud shadow and 2.8% as clouds.

The spatial patterns of error are examined in Figures 3 and 4. A false color image of the scene mapping Bands 5, 4 and 2 to red, green and blue, respectively, is provided for reference (top rows). Classification output for SPARCS (left) and FMask (right) show agreement with the evaluation masks,

with commission errors (purple) and omission errors (red) highlighted for clouds (light colors) and cloud shadows (dark colors).

Table 4. Agreement over all 12 test sub-scenes for SPARCS and FMask compared to the evaluation masks.

Labeled as	**Shadow**	**Cloud**	**Water**	**Snow/Ice**	**Clear**
SPARCS					
Classed as Shadow	94.7%	1.0%	2.3%	1.7%	0.5%
Cloud	0.7%	97.2%	0.1%	0.9%	0.2%
Water	0.5%	0.0%	96.6%	1.0%	0.1%
Snow/Ice	0.9%	0.4%	0.0%	90.2%	0.0%
Clear	3.2%	1.3%	1.0%	6.2%	99.2%
FMask					
Classed as Shadow	69.9%	0.5%	0.6%	7.6%	2.4%
Cloud	20.9%	98.6%	0.3%	10.7%	2.8%
Water	1.0%	0.0%	96.6%	0.0%	0.0%
Snow/Ice	0.3%	0.0%	0.0%	72.4%	0.1%
Clear	8.0%	0.9%	2.4%	9.3%	94.7%

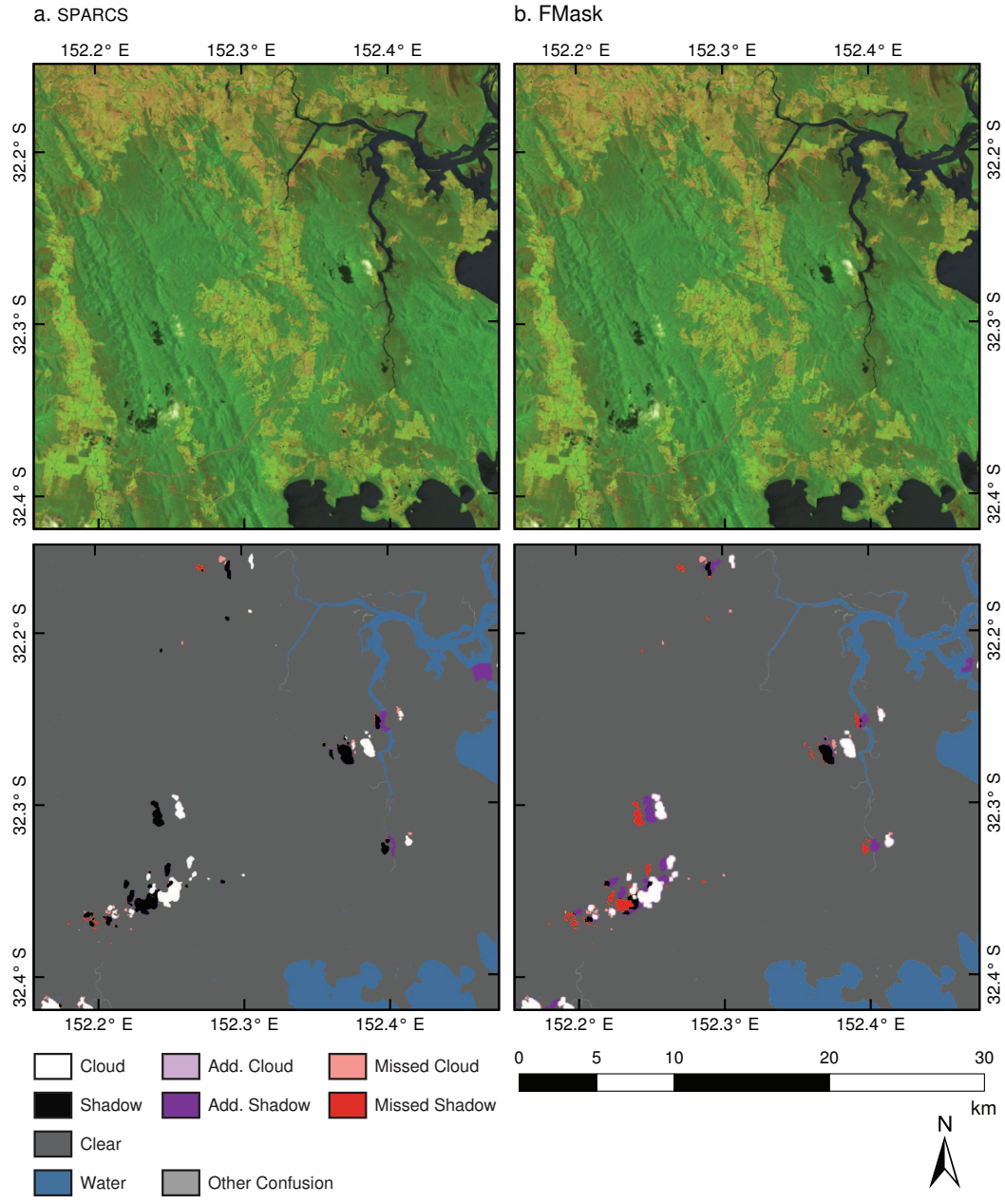
Figure 3 is of a scene with sparse mid-altitude clouds in the coastal region of New South Wales, Australia (WRS2 path/row 89/82). Both methods have strong cloud detection, though both miss some small, thin clouds in the northern portion of the image and thin clouds in the southwestern portion, as well as their respective shadows. Because the spectral signal of areas contaminated with thin clouds and their shadows is a mixture of cloud/cloud shadow and the underlying landscape, they are especially difficult to detect, as the resulting signal is ambiguous. Successful methods typically use multi-temporal image stacks and then detect deviations from an average or consensus signal [13].

FMask predicts cloud shadow by projecting the cloud mask onto the land-surface as a function of sun angle and topography without considering spectral information about shaded pixels. When this projection fails, this creates a pattern in the mask where the cloud shadow mask is offset from the actual location of the cloud shadow, as can be seen in several clouds in this image. SPARCS combines information about dark pixels from the neural network output with a similar, but much less precise, cloud projection approach to achieve more overall precision in cloud shadow locations. However, because of shadow/water ambiguity of dark pixels, this approach can causes over-shadowing in dark water near detected clouds, such as that in the eastern portion of the sub-scene.

A scene with dense, discrete clouds and cloud shadows in Hidalgo, Mexico (WRS path/row 26/46), is presented in Figure 4. For both SPARCS and FMask, clouds are detected very well, and most errors occur around the edges of cloud and cloud shadow objects. In SPARCS, some bright land cover in the eastern portion of the image is misidentified as clouds with some spurious associated cloud shadow. Additionally, some dark water is labeled as cloud shadow. FMask exhibits some bright land cover/cloud confusion, as well, though less than SPARCS. Again, the cloud shadow mask consistently fails to extend

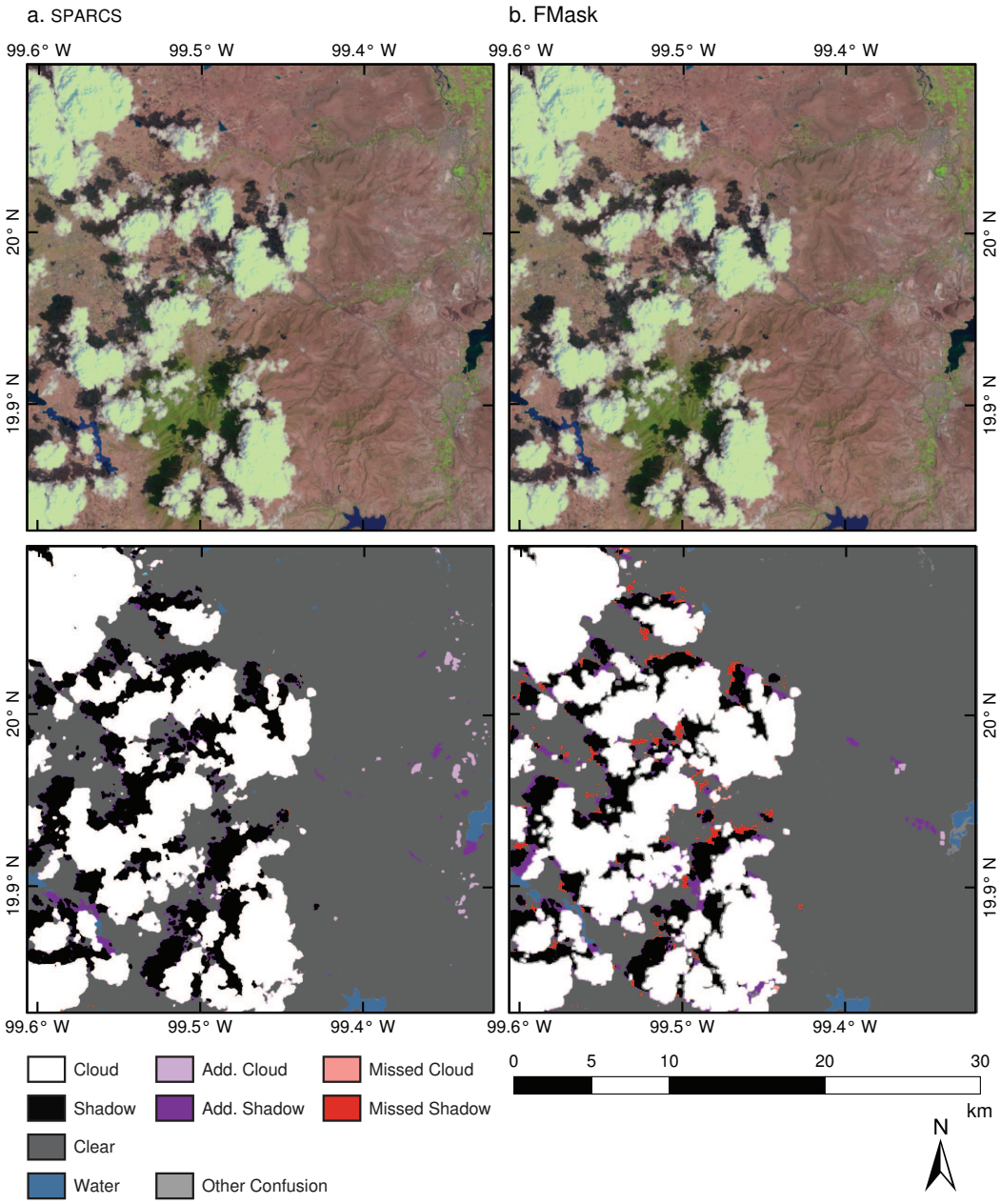
to the edges of cloud shadow objects, due to misprojection, resulting in substantial missed cloud shadow while simultaneously labeling unshadowed areas as shadowed.

Figure 3. Image classification of a sub-scene from New South Wales, Australia (WRS2 path/row 89/82), acquired on 21 April 2001, using SPARCS (**left**) and FMask (**right**) with confusion between the classifications and evaluation masks highlighted.



Both methods have a halo of commission error around cloud and cloud shadow objects that results from design decisions to reduce contaminated pixels by expanding those masks slightly. This expansion approach creates a trade-off between commission and omission errors, with larger expansions capturing more obstructed pixels by sacrificing nearby clear sky pixels. Much of this halo is ignored due to the three-pixel buffer, described above, but some extends past that buffer and can be seen in the images. The larger halos in FMask represent the method's more aggressive efforts toward this goal.

Figure 4. Image classification of a sub-scene from Hidalgo, Mexico (WRS2 path/row 26/46) acquired on 1 February 2001 using SPARCS (Spatial Procedures for Automated Removal of Cloud and Shadow) (**left**) and FMask (**right**) with confusion between the classifications and evaluation masks highlighted



Over all 12 evaluation sub-scenes, SPARCS performs consistently better than FMask (Table 5). In only one sub-scene, which is predominately cloud cover over the Amazon rainforest, is the overall accuracy for SPARCS less than that for Fmask, and then, only by 0.3%. In that scene, SPARCS misses some thin clouds on the edge of the bulk of the cloud mass. FMask and SPARCS perform relatively well or poorly on the same sub-scenes, that is, when SPARCS performs well, so does FMask, and *vice versa*, suggesting that in some sub-scenes, separating clouds and cloud shadow from clear sky is simply a more difficult problem than in others.

Table 5. Agreement over all 12 test sub-scenes for SPARCS and FMask compared to the evaluation masks.

	Missed Shadow	Missed Cloud	Over Shadow	Over Cloud	Overall
Jammu and Kashmir, India: pr 147/35 (36.4°N, 78.8°E). 20 February 2001.					
SPARCS	19.2%	10.1%	1.0%	0.6%	97.2%
FMask	6.9%	2.5%	7.2%	7.2%	86.8%
New Mexico, USA: pr 33/37 (33.5°N, 105.9°W). 11 February 2001.					
SPARCS	2.9%	0.2%	0.1%	0.0%	99.6%
FMask	3.4%	0.0%	4.7%	4.4%	92.4%
Zhejiang, China: pr 118/40 (28.6°N, 120.4°E). 11 March 2001.					
SPARCS	0.4%	0.1%	0.7%	0.2%	99.4%
FMask	5.0%	0.3%	3.3%	4.2%	94.9%
Baja California Sur, Mexico: pr 35/42 (25.5°N, 111.1°W). 22 March 2001.					
SPARCS	2.5%	0.0%	0.2%	0.0%	99.8%
FMask	17.1%	0.4%	0.2%	0.2%	99.2%
Hidalgo, Mexico: pr 26/46 (20.0°N, 99.5°W). 1 February 2001.					
SPARCS	0.2%	0.0%	1.0%	0.7%	98.9%
FMask	9.1%	0.5%	2.1%	0.5%	97.4%
Koulikoro, Mali: pr 199/51 (13.3°N, 7.3°W). 30 January 2001.					
SPARCS	0.0%	0.0%	0.0%	0.0%	100.0%
FMask	0.0%	0.0%	0.6%	0.3%	99.1%
Amazon, Brazil: pr 4/64 (5.2°S, 70.9°W). 13 March 2001.					
SPARCS	9.5%	1.9%	2.2%	0.8%	97.6%
FMask	9.4%	1.3%	8.4%	1.0%	97.9%
Tete, Mozambique: pr 168/71 (16.2°S, 33.3°E). 10 April 2001.					
SPARCS	0.0%	0.0%	0.0%	0.0%	100.0%
FMask	0.0%	0.0%	0.0%	0.0%	100.0%
Antofagasta, Chile: pr 1/75 (21.4°S, 68.9°W). 20 January 2001.					
SPARCS	17.7%	18.0%	2.5%	0.0%	96.4%
FMask	24.8%	7.4%	7.2%	13.5%	79.8%
New South Wales, Australia: pr 89/82 (32.3°S, 152.3°E). 21 April 2001.					
SPARCS	6.9%	11.2%	0.2%	0.0%	99.6%
FMask	59.3%	9.7%	0.5%	0.1%	98.9%
Hawke's Bay, New Zealand: pr 71/87 (39.2°S, 177.1°E). 12 April 2001.					
SPARCS	9.1%	0.8%	1.1%	0.2%	98.6%
FMask	12.7%	2.1%	0.1%	0.1%	99.8%
Santa Cruz, Argentina: pr 228/96 (51.9°S, 70.5°W). 11 January 2001.					
SPARCS	3.9%	1.4%	1.0%	2.5%	97.8%
FMask	6.8%	0.4%	5.0%	2.5%	97.1%

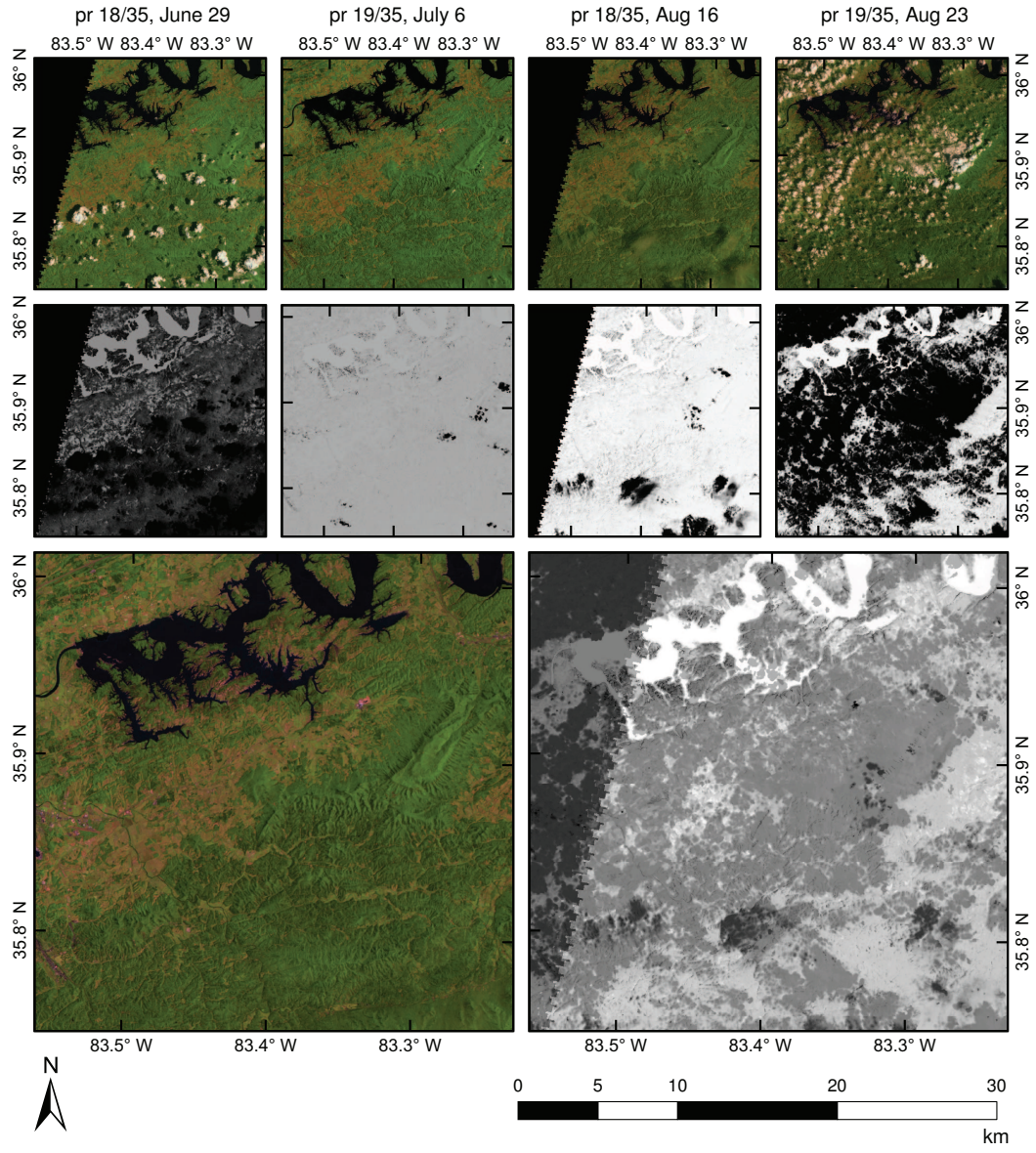
In forest disturbance identification, mislabeling cloud and cloud shadow as clear sky (omission) is generally a larger problem than mislabeling clear areas as clouds or cloud shadow (commission), because dark or bright spots can be mistaken for ephemeral disturbance [17]. Commission errors become important, however, in areas with frequent cloud cover, such as the tropics or mountainous regions, where actual clear sky views are rare. In these cases, proper identification of clear pixels is necessary to increase the number of observations and maximize the likelihood of comparing images with similar days of the year and phenology. We believe SPARCS provides a balance between these competing objectives. Although SPARCS consistently misses more cloud cover than FMask, it does so at the gain of substantially reducing commission error in otherwise obstructed scenes and, so, increases the likelihood of observing rare clear sky pixels in cloudy areas.

4.3. Creating a Multi-Temporal Composite

To illustrate the usefulness of our method, we generated a composite image from Landsat 5 TM images acquired over eastern Tennessee on four dates in the summer of 1990 (Figure 5). Though SPARCS was developed using data from Landsat 7 ETM+, since it does not use the panchromatic band, it can be easily applied to the TM archive. Two of the images in this example have significant cloud cover. One scene, from July 6, is mostly clear sky, though a few cloud and cloud shadow pairs are detected and removed from the composite. A scene from August 16, contains several thin clouds. The scene weights shown are the sum of the clear sky and water memberships from the SPARCS output, multiplied by a deceasing function of distance from August 1, which is used to account for shifts in phenology, as images are acquired further away from the target day. Additionally, the algorithm provides the total of the weights used (Figure 5 bottom right), which can be useful in calculations and analyses further down a processing pipeline to determine, for example, that insufficient data was available for an area during a certain year or that a seemingly anomalous area actually has significant support.

In the scene from 16 August, large portions of the thin cloud are assigned intermediate weights due to classifier uncertainty. By squaring these values, our algorithm trusts these areas substantially less than the same areas in clear images. By providing a continuous measure of uncertainty, though, different algorithms and operators can choose their own thresholds for how conservative they wish to be with data inclusion. The classifier is also consistently uncertain about several areas around Douglas Lake, the body of water in the northern part of the images. However, because the algorithm takes a weighted mean of each image on a pixel-by-pixel basis, only the relative weight through time is important, allowing them to be combined successfully. They are, however, somewhat more susceptible to contamination by clouds or cloud shadow, as the relative difference between the weights of contaminated and uncontaminated pixels is less.

Figure 5. A 30×30 km region in eastern Tennessee from four Landsat 5 TM scenes acquired during the summer of 1990 (**top row**) and their respective weights from zero to one (black to white) from SPARCS (**second row**), where weights of zero signify contaminated or unusable pixels. The results from compositing (**bottom left**) and the sum of the weights used to determine each pixel (**bottom right**) are included.



5. Conclusions

We presented a neural network approach to detect cloud and cloud shadow, as well as water and snow/ice, in Landsat TM and ETM+ imagery: SPARCS (Spatial Procedures for the Automated Removal of Clouds and Shadow). SPARCS uses only single date imagery, does not rely on ancillary datasets and outperforms another high quality method that operates on similar constraints with an overall accuracy of 98.8% compared to 95.3%. Additionally, it is completely automated and does not require specifying new parameters for different scenes, and the classification of a Landsat scene completes in under 5 min on a

desktop computer using an AMD Athalon II 3.1 GHz dual-core processor with 8 Gb of RAM, meeting our design goals.

Unlike other cloud and cloud shadow detection algorithms, SPARCS is a fuzzy classifier; crisp classification can be achieved by labeling each pixel as the highest-valued membership class and then using the variance among class memberships as an accompanying measure of uncertainty. Knowing uncertainty allows spatial analyses that utilize SPARCS-generated cloud masks to create more accurate spatial products that have more robust estimates of error.

We explored the inclusion of spatial information as an input to the neural network classifier and found limited support for their inclusion. No method summarizing spatial information increased overall accuracy by more than 0.5%. However, a post-processing stage using expert-defined rules increased accuracy by 3.5%. Of particular usefulness is the rule to differentiate between cloud shadows and terrain shadow that combines predicted cloud shadow locations from solar geometry and cloud locations with neural network output in the cloud shadow and water classes. Inclusions of data from larger spatial areas summarized using total variation regularized denoising (with $\alpha > 0.1$) or using a log-polar representation of the local neighborhood [32], which has promise in the field of robotic vision, as neural network inputs may be useful avenues for future research. However, these methods are computationally intensive. We believe that a multi-stage method that first classifies several single-date scenes of the same location using a method, such as the one described here, and then using those classifications in a second-stage multi-date classifier to resolve cloud and cloud shadow within each single-date scene is the most promising way forward.

Acknowledgments

This research was supported through Grant NO. NNX10AT66G of the NASA New Investigator program and the Scalable Computing and Leading Edge Innovative Technologies fellowship, part of the NSF's Integrative Graduate Education and Research Traineeship program. We thank Pat Scaramuzza for directing us to the USGS cloud masking dataset, Zhe Zhu for providing assistance with FMask, and Lou Gross, Doug Kaylor and the anonymous reviewers who provided valuable comments on earlier drafts of this manuscript.

Author Contributions

Joseph Hughes developed methods, created the evaluation dataset, and analyzed results. Daniel Hayes supervised research. Joseph Hughes and Daniel Hayes wrote the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. 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.

2. Saunders, R.; Kriebel, K. An improved method for detecting clear sky and cloudy radiances from AVHRR data. *Int. J. Remote Sens.* **1988**, *9*, 123–150.
3. Derrien, M.; Farki, B.; Harang, L. Automatic cloud detection applied to NOAA-11/AVHRR imagery. *Remote Sens. Environ.* **1993**, *46*, 246–267.
4. Cihlar, J.; Howarth, J. Detection and removal of cloud contamination from AVHRR images. *IEEE Trans. Geosci. Remote Sens.* **1994**, *32*, 583–589 .
5. Simpson, J.; Stitt, J. A procedure for the detection and removal of cloud shadow from AVHRR data over land. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 880–890.
6. Ackerman, S.A.; Strabala, K.I.; Menzel, W.P.; Frey, R.A.; Moeller, C.C.; Gumley, L.E. Discriminating clear sky from clouds with MODIS. *J. Geophys. Res.* **1998**, *103*, 32141–32157.
7. Gao, B.; Kaufman, Y. Selection of the 1.375-μm MODIS channel for remote sensing of cirrus clouds and stratospheric aerosols from space. *J. Atmos. Sci.* **1995**, *52*, 4231–4237.
8. Luo, Y.; Trishchenko, A.; Khlopenkov, K. Developing clear-sky, cloud and cloud shadow mask for producing clear-sky composites at 250-meter spatial resolution for the seven MODIS land bands over Canada and North America. *Remote Sens. Environ.* **2008**, *112*, 4167–4185.
9. Oreopoulos, L.; Wilson, M.J.; Várnai, T. Implementation on Landsat data of a simple cloud-mask algorithm developed for MODIS land bands. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 597–601.
10. Martinuzzi, S.; Gould, W.; González, O. *Creating Cloud-Free Landsat ETM+ Data Sets in Tropical Landscapes: Cloud and Cloud-Shadow Removal*; General Technical Report IITF-GTR-32; International Institute of Tropical Forestry: Ro Piedras, Puerto Rico, 2007; pp. 1–18.
11. Choi, H.; Bindschadler, R. Cloud detection in Landsat imagery of ice sheets using shadow matching technique and automatic normalized difference snow index threshold value decision. *Remote Sens. Environ.* **2004**, *91*, 237–242.
12. Hollingsworth, B; Chen L.; Reichenbach S.E.; Irish R.R. Automated cloud cover assessment for Landsat TM images. *Proc. SPIE* **1996**, *2819*, 170–179.
13. Goodwin, N.R.; Collett, L.J.; Denham, R.J.; Flood, N.; Tindall, D. Cloud and cloud shadow screening across Queensland, Australia: An automated method for Landsat TM/ETM+ time series. *Remote Sens. Environ.* **2013**, *134*, 50–65.
14. Kennedy, R.E.; Schroeder, T.A.; Cohen, W.B. Trajectory-based change detection for automated characterization of forest disturbance dynamics. *Remote Sens. Environ.* **2007**, *110*, 370–386.
15. Berendes, T.; Sengupta, S.; Welch, R.; Wielicki, B.; Navar, M. Cumulus cloud base height estimation from high spatial resolution Landsat data: A Hough transform approach. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 430–443.
16. Hagolle, O.; Huc, M.; Pascual, D.; Dedieu, G. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENS, LANDSAT and SENTINEL-2 images. *Remote Sens. Environ.* **2010**, *114*, 1747–1755.
17. Huang, C.; Thomas, N. Automated masking of cloud and cloud shadow for forest change analysis using Landsat images. *Int. J. Remote Sens.* **2010**, *31*, 37–41.
18. Zhu, Z.; Woodcock, C.E. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* **2012**, *118*, 83–94.

19. Haykin, S. *Neural Networks and Learning Machines 3rd Ed.*; Pearson Education, Inc.: Upper Saddle River, NJ, USA, 2008.
20. Scaramuzza, P.L.; Bouchard, M.A.; Dwyer, J.L. Development of the Landsat data continuity mission cloud-cover assessment algorithms. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1140–1154.
21. Rudin, L.; Osher, S.; Fatemi, E. Nonlinear total variation based noise removal algorithms. *Phys. D Nonlinear Phenom.* **1992**, *60*, 259–268.
22. Goldstein, T.; Osher, S. The split bregman method for L1-regularized problems. *SIAM J. Imaging Sci.* **2009**, *2*, 323–343.
23. Hall, D.K.; Foster, J.L.; Verbyla, D.L.; Klein, A.G.; Benson, C.S. Assessment of snow-cover mapping accuracy in a variety of vegetation-cover densities in central Alaska. *Remote Sens. Environ.* **1998**, *137*, 129–137.
24. Crist, E.P. A TM tasseled cap equivalent transformation for reflectance factor data. *Remote Sens. Environ.* **1985**, *306*, 301–306.
25. MacQueen, J. Some Methods for Classification and Analysis of Multivariate Observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, California, USA, 21 June–18 July 1967; University of California Press: Berkeley, CA, USA, 1967; Volume 1, pp. 281–297.
26. Moran, M.; Jackson, R.; Slater, P.; Teillet, P. Evaluation of simplified procedures for retrieval of land surface reflectance factors from satellite sensor output. *Remote Sens. Environ.* **1992**, *41*, 169–184.
27. Chander, G.; Markham, B.L.; Helder, D.L. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sens. Environ.* **2009**, *113*, 893–903.
28. Chavez, P.S. Image-based atmospheric corrections—Revisited and improved. *Photogramm. Eng. Remote Sens.* **1996**. *62*, 1025–1036.
29. Beale, M.H.; Hagan, M.T.; Demuth, H.B. *Neural Network Toolbox User's Guide*, r2013b ed.; The MathWorks, Inc: Natick, MA, USA, 2013.
30. Zar, J.H. *Biostatistical Analysis*, 5th ed.; Prentice-Hall: Upper Saddle River, NJ, USA, 2010.
31. Kramer, C.Y. Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics* **1956**, *12*, 307–310.
32. Traver, V.J.; Bernardino, A. A review of log-polar imaging for visual perception in robotics. *Robot. Auton. Syst.* **2010**, *58*, 378–398.