This is a post-print version of an article published in IEEE Journal of selected Topics in Applied Earth observations and remote sensing in November 2018. The original article can be found here: 10.1109/JSTARS.2018.2870978

Image Classification by Integrating Reject Option and Prior Information

Gregory N. Taff, Yang Shao, Jie Ren, and Ruoyu Zhang

Abstract—**Accuracy in land cover classification using remotely sensed imagery can be increased using Bayesian methods that incorporate prior probabilities of classes. However, estimating these prior probabilities can be expensive and data intensive. We propose methods to improve classification accuracy using Bayesian methods to classify ambiguous (or low-confidence) pixels, using only the remotely-sensed imagery or existing land cover maps to estimate prior probabilities. We propose a spatial method that predicts prior probabilities from the original image, and a temporal method that incorporates land cover maps from previous years. We illustrate our methods with a Neural Network (NN) classifier on the US state of Iowa to classify crops into corn/soybean/other using MODIS data. USDA Cropland Data Layers (CDLs) were aggregated to the 250m resolution of MODIS and used as ground truth, based on a cropland mask from the National Land Cover Database (NLCD). Results show that the spatial prior adjustment method, which predicts prior probabilities for low-confidence pixels based on class percentages of initial NN classification, increased overall accuracy of low-confidence pixels between 2 – 3.3 % over the standard NN classification. The temporal prior adjustment method, which uses crop classes from the previous 6 years to estimate prior probabilities for the current year, shows significantly greater accuracy improvement for low-confidence pixels (almost 7%) over the standard NN classification. Increased benefit of the temporal prior adjustment method relative to the spatial prior adjustment method is likely due to increased information from more ground truth data (from previous years) than the spatial method.**

Index Terms—**Neural Network applications, Bayes procedures, remote sensing, algorithms, agriculture**

I. Introduction

IMAGE classification is one of the most common and important tasks in satellite remote sensing applications. Inherent in image classification is error. Researchers are continually trying to develop methodology to achieve lower classification error rates with minimal added data and labor costs. In the past several decades, a variety of supervised classification algorithms have been developed. Early works mainly focus on parametric (e.g., parallelepiped, maximum likelihood) approaches, while non-parametric algorithms such as neural networks, random forest, and support vector machines receive increased attention in recent years [1], [2], [3]. Higher classification accuracies are often achieved with non-parametric methods compared to those from traditional parametric algorithms, particularly when large and complex remote sensing datasets are employed in land cover classification [4], [5]. In addition to remote sensing spectral bands/features, spatial textural information and spectral indices (e.g. NDVI) might be incorporated to improve the land cover classification performances [6]. Ancillary data such as census data, topographic, hydrographic, or road data in a GIS are also found to be useful for improving land cover classification accuracies [7], [8], [9], [10].

Independent of land cover classification algorithms and input data used, there is always a problem related to the overlap of land cover classes in feature space [11]. The spectral confusions are well documented for land class pairs such as primary forest and secondary forest, impervious cover and bare soil, and water and shadow in numerous remote sensing applications [12], [13], [14]. In addition to the spectral confusion between classes, spectral mixture of land cover classes at the sub-pixel level might also introduce additional confusions and cause difficulties in land cover classification [15], [16]. The spectral confusions between classes and/or spectral mixtures often lead to suspicious classification results because pixels are ambiguous to the classifiers. For instance, a maximum likelihood algorithm may have suspect results in cases where classes have substantial overlap areas in feature space. Moody et al. (1996) [16] found that when classes are similar, a neural network classification approach may produce low network output values that require caution in the pixel label process. Another common problem that may lead to a large number of

Manuscript was received on October 12, 2017; revised on October 31, 2017; revised again on January 20, 2018; revised again on March 29, 2018; revised again on July 18, 2018; and accepted on September 2, 2018. The authors would like to thank the Norwegian Research Council for supporting our research. Funding was partially provided by the Norwegian Research Council grant number 194051.

G. N. Taff is with the Norwegian Institute for Bioeconomy Research, Postbox 115, 1431 Ås, Norway (e-mail: gregory.taff@nibio.no).

Y. Shao is with Department of Geography, Virginia Tech, Blacksburg, VA 24061 USA (e-mail: yshao@vt.edu).

J. Ren is with the School of Geographic and Oceanographic Sciences, Nanjing University, China (e-mail: jren4@vt.edu).

R. Zhang is with Department of Geography, Virginia Tech, Blacksburg, VA 24061 USA (e-mail: ruoyu1@vt.edu).

This is a post-print version of an article published in IEEE Journal of selected Topics in Applied Earth observations and remote sensing in November 2018. The original article can be found here: 10.1109/JSTARS.2018.2870978

ambiguous image pixels, and therefore a high classification error rate, is when limited or unrepresentative training pixels are used in image classification practices, especially for classes with high spectral variation [17]. In addition, the users may arbitrarily choose any number of training samples (e.g., 200 pixels for one class and 1000 pixels for another class) and the distribution of the number of selected training samples can be very different from true class prior probabilities, therefore seeding the classification with a false set of prior probabilities [18].

These sources of classification errors can be reduced using a so-called “reject” option in the classification algorithm, which leaves ambiguous pixels unclassified [19]. Specifically, a thresholding technique can be used to reject ambiguous pixels from an initial classification. Chow’s rule (1970) is one of the commonly used methods, in which a threshold is set by the user for the class posterior probability: when the posterior probability for a pixel is less than this threshold for all classes, the pixel is rejected. More recently, Fumera et al. (2000) [20] recommended a reject option with multiple thresholds (one for each class) to determine which pixels to reject from the classification. When this is done, classification accuracy for the remaining (classified) pixels is typically substantially higher, but then the problem remains: what is to be done with the ‘rejected’ (unclassified) pixels? With the aid of ancillary data and manual interpretation (e.g., using high resolution imagery or air photos as reference), these pixels can be classified with an increase in classification accuracy. However, little work has been done to suggest how to deal with these rejected pixels in a more efficient manner. In fact, little research has employed or evaluated the rejection option in remote sensing classification due to the data and personnel costs associated with acquiring ancillary data. Therefore, there is a need to develop methods that consistently increase classification accuracy through automated classification of rejected pixels with the minimum use of ancillary data.

Here we develop and demonstrate the benefit of automated methods that use neural networks to classify, reject, and then re-label ambiguous pixels. Ambiguous and unambiguous pixel groups are identified by thresholding of neural network output signals. In one proposed “spatial” method, the class percentages found in the initial NN classification are used to estimate image prior probabilities. Such image spatial prior information is then used to adjust neural network signals to re-classify initially rejected or ambiguous pixels. Within this prior adjustment framework, we also examine a “temporal” method of integrating temporal information to estimate prior probabilities to improve classification accuracy of initially rejected pixels. We demonstrate how these methods perform using a crop-specific mapping application. Specifically, multi-temporal Moderate Resolution Imaging Spectroradiometer (MODIS) Normalized Difference Vegetation Index (NDVI) data are used as input to classify corn, soybean, and other crops. Multiple years (2009-2015) of crop-specific mapping are tested to examine the robustness of proposed methods. In the remote sensing community, crop-specific mapping remains a challenging image classification problem, especially for large-scale operational applications [21], [22], [23]. The main reason to choose corn and soy mapping for our study is because they are two dominant crops in the Midwestern US, and they are also difficult to separate because both are summer crops (with a similar crop calendar) [24], [25]. Furthermore, this output is valuable because a range of agricultural remote sensing applications such as crop rotation change and crop phenological analysis require annual corn/soybean map products as a key input, and such a map is only available from NASS approximately a year after the crops are harvested [26], [27].

The rest of the paper is organized as follows: in the Method section we describe the data used in this study, illustrate how we apply neural network classification and pixel rejection, and proposed methods of integrating spatial and temporal information to improve image classification. Detailed accuracy assessments and comparisons of the various tested methods are presented in the Results and Discussion sections. Finally, we summarize key findings in the Conclusion section.

II. METHOD

A. Data and Data Pre-processing

We used multi-temporal MODIS NDVI classification to illustrate neural network-based pixel rejection and pixel re-label methods. Multi-temporal MODIS time-series data are now increasingly used in land cover classification applications [28], [29]. For this study, we examined a crop-specific mapping task focusing on classification of corn, soybean, and other crops using time-series MODIS NDVI data as input for the US state of Iowa (Figure 1). The 250m resolution 16-day composite MODIS NDVI data from 2009-2015 were downloaded from the NASA Reverb (http://reverb.echo.nasa.gov/reverb/) website. MODIS NDVI imagery were projected to an Albers Equal Area Conic projection and then clipped using the Iowa State boundary. The dimension of the MODIS NDVI image was 1403 by 2137. The Whittaker Smoother (WS) was applied to the original MODIS time-series NDVI data to reduce image noise of pseudo-hikes and pseudo-lows [30], [31]. The WS algorithm shows good balance in fidelity to the original data and smoothness of the fitted curve and has high potential for crop-specific mapping tasks [32]. Figure 2 shows an example of smoothed NDVI signals for corn and soybean pixels for year 2009.

We downloaded the 2011 National Land Cover Database (NLCD, 30m resolution) from the Multi-Resolution Land Characteristics Consortium (http://www.mrlc.gov/). The Cultivated Crops class (code=82) was extracted from the NLCD and rescaled to the 250m MODIS grid by computing sub-pixel land cover proportions. MODIS grids with greater than 50% of Cultivated Crops were used as an image mask and our crop-specific mapping was limited within this cropland mask. After applying the cultivated crops mask, there were 1,419,318 pixels (~ 47% of the total area) remaining for our detailed crop mapping experiments. The USDA’s Cropland

This is a post-print version of an article published in IEEE Journal of selected Topics in Applied Earth observations and remote sensing in November 2018. The original article can be found here: 10.1109/JSTARS.2018.2870978

Data Layers (CDLs) from 2009 to 2015 were downloaded from the CropScape website. These CDLs have 30-56m spatial resolution, depending on the input data and data analysis protocols [33]. Classification accuracies for corn and soybean are reported to be higher than 90% for Iowa for all study years from 2009 to 2015. Corn and soybean pixels were extracted from CDLs and class proportions were then calculated for each 250m MODIS grid. MODIS grids with greater than 50% of corn (or soybean) proportions were labelled as a corn (or soybean) class. All the remaining MODIS grids within the cropland mask were considered as other crops (e.g., winter wheat, grassland, etc.). Such CDL data processing procedures were repeated for each study year from 2009 to 2015, generating a set of high quality reference datasets to test MODIS-NDVI classification performance. We note that all MODIS, NLCD and CDLs have been re-projected to an Albers Equal Area Conic (AEA) projection to ensure consistency in spatial projection. MODIS images have been geolocated to sub-pixel (250m) accuracies. Although it is not possible to have a perfect geometrical match between MODIS and the CDLs, it was generally acceptable to use CDLs as 'ground truth', especially for our study area of Midwestern US, where agricultural field patches are quite large.

B. Neural Network Classification and Pixel Rejection

We used a three-layer Multi-layer perceptron (MLP) neural network (NN) to test our pixel rejection and re-labelling methods. MLP NN has been widely used for general land cover and crop-specific mapping applications [34], [35]. Given a moderate-large number of training data points, MLP NN often generates good image classification results [36]. In addition, MLP NN can be trained to approximate the posterior probabilities [16], [18]. The Bayesian posterior probability can be written as:

$$P(c_i \mid x) = \frac{p(x \mid c_i) \times P(c_i)}{P(x)} \tag{1}$$

In remote sensing applications, x can be seen as the digital number (DN) value from one spectral band or a DN vector from multispectral space. $P(c_i|x)$ is the posterior probability, indicating the probability that a pixel belongs to class c_i given the pixel's DN value(s), x. $p(x|c_i)$ is the likelihood – the probability that a pixel has DN value/vector x, given it is in class c_i. $P(c_i)$ is the class prior probability. It is the probability that a randomly chosen pixel in the image is of class c_i, or simply, the fraction of pixels in the image belonging to class c_i. $P(x)$ is a normalization term. Chow's Rule thus is readily applicable to identify unambiguous or ambiguous pixel groups by applying a user-defined threshold value (e.g., T=0.75) on the posterior probability estimation. Specifically, for a given pixel, if the largest NN output signal (or posterior probability) is less than the threshold value, the classification result for the pixel can be rejected or considered as low confidence that requires further analysis.

The MLP NN used in this study includes 13 input nodes to represent multi-temporal MODIS NDVI features from DOY 97 to DOY 289. Previous studies showed that these early-spring to late-fall NDVI work best in crop-specific mapping, especially for corn/soybean [37]. The NN output layer includes three nodes representing corn, soybean, and other crops. We used a 1-of-M target coding system (e.g. 1,0,0) for three classes. A sigmoid transfer function was used as the NN transfer function and the NN was trained to minimize the mean squared error. The most suitable number of nodes in the hidden layer needs to be tested with a trial-and-error method, which was automated through a cross-validation approach. Training data points were randomly selected using CDL-derived reference images. Initially, only 1000 pixels per class were selected as a starting point to evaluate performance of various classifications. More training data points (e.g., 1,000-8,000 per class) were then examined to compare classification performance based on varying training data sample sizes. For each sample size, we repeated training data selection 10 times to incorporate variation of training data characteristics.

The trained NNs were used to classify the full MODIS image covering the state of Iowa. For most NN classification applications, a given pixel is labelled based on the largest output signal (i.e., winner-takes-all) of three output nodes (corn, soybean, and other crops). We compared the largest output signal for each pixel with a user-defined threshold (e.g., T=0.75) to identify high confidence and low confidence pixel groups. Pixels in the low confidence group were still labelled using a winner-takes-all approach based on initial network output signals, however, additional analyses were conducted (see sections below) for these pixels to improve their accuracy. A general difficulty with the utilization of a thresholding rule is the identification of the appropriate threshold value. For this study, a threshold value of 0.75 was used as a starting point. Different threshold values (0.5 and 0.9) were further examined to assess how the threshold selection affects classification performance.

C. Integrating Spatial Prior Information

To improve accuracy in classifying the rejected pixels, the results of the initial NN classification were used to estimate the class prior probabilities in the image. The prior probabilities were simply estimated as the proportions of the classified data points that belong to each of the three classes. The estimated prior probabilities were considered a better approximation to the real prior probabilities than the equal prior assumption which we used in the initial NN training. A simple method to compensate for these different proportions can be employed, based on the Bayesian equation [18]. Specifically, the network outputs were multiplied by the prior probabilities estimated from the full image and then normalized (equation (1)). After adjusting the network outputs and generating new posterior probabilities for each MODIS pixel, a new class label was assigned based on the maximum posterior probability among the three target classes.

D. Integrating Temporal Prior Information

For this crop-specific mapping task, we also designed a

This is a post-print version of an article published in IEEE Journal of selected Topics in Applied Earth observations and remote sensing in November 2018. The original article can be found here: 10.1109/JSTARS.2018.2870978

method to incorporate temporal prior information to potentially improve classification performance. To support 2015 MODIS image classification, we analysed the cropping frequency for corn and soybean, pixel-by-pixel using CDL as reference data for the period 2009-2014. The temporal prior was calculated pixel-by-pixel as a proportion of previous years in each crop type. For a given MODIS pixel, if the cropping sequence for the period 2009-2014 was corn-corn-soybean-corn-soybean-corn, the temporal prior for corn and soybean can be estimated as 0.67 and 0.33, respectively. Each pixel thus would have its own temporal prior information with respect to the three different land cover classes. Such temporal prior information was integrated to adjust neural network output signals using the same approach described as with the spatial prior adjustment procedure. The temporal prior adjustment was implemented for all pixels, including both the low confidence and high confidence groups, because such temporal prior information provides useful additional knowledge (i.e., independent from spatial information) and could potentially improve classification performance for all pixels. We note that the integration of temporal prior was only conducted for the 2015 MODIS image classification, because temporal information needs to be derived from previous years' CDLs.

E. Accuracy Assessment

Reference data for the accuracy assessment were derived from the high accuracy CDLs from 2009-2015. MODIS image classification results were compared to these CDL references at the 250m spatial scale to generate a confusion matrix, overall accuracy, kappa coefficients, and user's and producer's accuracies [38]. The main reason to include multiple years of MODIS classification and accuracy assessments was to evaluate the robustness of the proposed prior-adjusting methods. Many agricultural landscapes are constantly changing and have sharp year-to-year differences; multiple years of crop-specific mapping and accuracy assessment thus provide an ideal scenario to test new classification algorithms.

III. Results and Discussion

A. Neural Network Classification and Rejection Option

NN training was begun with 1000 pixels per class as training data points. For the 2015 MODIS classification, about 66% of pixels were labelled as high confidence based on thresholding ($T \geq 0.75$) of the largest network output signals. The remaining 34% of pixels ($T<0.75$) were labelled as low confidence. Table 1 compares classification accuracies of these two groups using a commonly used winner-takes-all pixel labelling method. The overall accuracy for the high confidence group was 71.7% (kappa = 0.562), as compared with 52.5% (kappa = 0.196) for the low confidence group. The user's accuracies of the corn, soybean, and other crops were 85.3%, 76.8% and 45.3%, respectively, for the high confidence group, as compared with just 65.7%, 44.2%, and 27.5%, respectively, for the low confidence group. Similarly, the producer's accuracies for the high confidence group were substantially higher than for the low confidence group. We visually compared the locations of low confident pixels with the CDL. Figure 3 shows that most of these low confident pixels were located at edges of agricultural patches or on smaller fields.

Figure 4 depicts large differences of overall accuracies and kappa coefficients for high confidence and low confidence groups for all MODIS classification years from 2009 to 2015. Overall accuracies for high confidence groups ranged from 68.5-72.7%, as compared with 49.4-54.1% for low confidence groups. When these two groups were combined, MODIS classifications for 2009 - 2015 resulted in overall accuracies ranging from 62.5-65.7% (kappa 0.425-0.468). Such moderate levels of overall accuracy were expected, because corn and soybean are both summer crops and it is generally difficult to separate these two classes at a 250m spatial resolution, even with multi-temporal classification approaches [23], [24].

We examined impacts of training sample sizes on pixel rejection (or confidence labelling) and associated classification accuracy statistics. Figure 5 shows that the percent of rejected (or low confidence) pixels decreases when training sample size is increased from 1000 to 8000 pixels per class. For example, rejection rates for the 2015 MODIS classification were 33% and 29% for training sample size of 1000 and 8000, respectively. For the same 2015 MODIS classification, the overall accuracies for the low confidence groups ranged from 51.4% to 52.5% using varying training sample sizes. For all pixels including low and high confidence groups, the overall accuracies ranged from 64.5% to 65.9% using varying training sample sizes from 1000 to 8000 pixels per class. Such results suggest that 1000 pixel per class might be sufficient because further increase of training data points only slightly improved classification accuracy. We thus focus our analyses using 1000 pixel per class to reduce redundancy.

B. Integrating Spatial Prior

The initial NN classification results were used to compute estimated class prior probabilities to potentially improve classification accuracy. For the 2015 MODIS classification, the estimated class prior probabilities from the initial NN classification were 0.38, 0.37, and 0.25 for corn, soybean, and other crops, respectively. Initial NN outputs thus were slightly adjusted to favour corn (and soybean to a lesser extent), as compared to the other crop class. Table 2 compares confusion matrices of the initial NN classification and the spatial prior-adjusted results for the low confidence pixel group. More pixels (20,161 vs 18,401) were labelled as corn after integrating spatial prior. As a result, the user's accuracy for corn was slightly reduced from 65.7% to 65.0%. However, the producer's accuracy for corn improved from 57.7% to 62.5%. Integration of spatial prior produced better users' accuracy for the other crops class, which has the lowest prior probabilities of the three classes. On the other hand, the producer's accuracy decreased compared to those from the initial NN classification. This brings out the potential weakness (for certain applications) of the spatial prior adjusting method. A reduction in producer's accuracy of the less dominant class is not desirable in situations

This is a post-print version of an article published in IEEE Journal of selected Topics in Applied Earth observations and remote sensing in November 2018. The original article can be found here: 10.1109/JSTARS.2018.2870978

when the primary goal of classification is to minimize the chance of misclassifying a less dominant class (e.g., when the analyst's primary goal is to locate all instances of a relatively rare land cover type). In such cases, spatial prior adjusting is not ideal. For the high confidence group pixels, the application of spatial prior did not improve classification accuracy - the difference in overall accuracy was less 0.1% when using vs. not using spatial prior information. This is because high confidence pixel groups were relatively easy to classify and incorporating spatial prior information rarely affects their classification.

Figure 6 compares overall accuracies (kappa coefficients) from the initial NN classification and the spatial prior-adjusted approach for all MODIS classifications from 2009 to 2015. Overall accuracy improved 2-3.3%, depending on the year (2009-2015). Kappa coefficients also slightly improved for almost all mapping years, suggesting that the spatial prior-adjusting approach generated relatively balanced commission errors and omission errors as compared with those from the initial NN classifier.

C. *Integrating Temporal Prior*

Even with spatial prior adjustment, there were high levels of confusion among the three classes of corn, soybean and other crops. We examined a new method of temporal prior integration to improve classification accuracy. Temporal prior information was derived from multiple years of CDLs by calculating cropping frequency, pixel-by-pixel. Figure 7 shows cropping frequency (2009-2014) for corn and soybean. The color red indicates pixels have been cropped more frequently as corn (Fig. 7a) or soybean (Fig. 7b) during these 6 years. For a given pixel, we expected that temporal prior probability (or cropping frequency) derived from previous years would contribute to farmers' likelihood of crop choice for the current year (2015). Such information could potentially improve classification performance of the 2015 MODIS classification.

The classification accuracy statistics for the standard NN and temporal prior-adjusting algorithm are shown in Table 3. The overall accuracy for the low confidence group increased from 52.5% (kappa = 0.196) for the standard NN to 59.2% (kappa = 0.270) for the temporal prior-adjusting algorithm, an almost 7% of increase. The user's accuracies increased to 66.7%, 47.2%, and 51.2% for corn, soybean, and other crops, as compared to 65.7%, 44.2%, and 27.5% from the initial NN classification. The producer's accuracies for corn and other crops also improved from 57.7% and 29.3% to 70.5% and 39.6%, respectively. The only weakness of temporal prior adjustment is that we observed a reduction in producer's accuracy for soybean (from 52.9% to 46.8%). The integration of temporal prior increased the likelihood of low confidence pixels to be classified as the most dominant class (corn), which had the consequence of increased misclassification of soybean as corn. However, the increase in all other user's and producer's accuracies (and in the overall accuracy and Kappa) clearly show that for these data, the benefits outweighed the weaknesses of this method. For the high confidence group, the integration of temporal prior slightly improved the overall accuracy to 72.9% (compared to 71.7% in the initial NN classification). Table 4 shows accuracies for the individual classes. Because more MODIS pixels were classified as corn when incorporating temporal prior, there was a decrease in user's accuracy for corn. Accordingly, users' accuracies slightly increase for the other two classes.

Incorporating the spatial prior moderately increased classification accuracy, while the temporal prior appeared to be significantly more effective in improving classification performance. One of the possible reasons is that the temporal prior is able to make use of information from much more ground truth data (CDLs from multiple years) compared to the spatial prior, which only uses CDLs from one year combined with estimates from the initial classification. True prior information for a new input image is always unknown and it is documented to be a challenging task to approximate true prior probabilities based on initial image classification alone [39], [40].

Regarding the temporal prior method, confusion between the corn and soybean classes might be reduced if a longer time-series of CDLs is processed to provide better cropping frequency estimates. We conducted a sensitivity analysis to evaluate how classification accuracy varies when we use two to six years of CDLs for temporal prior calculation. Table 5 compares overall accuracies and kappa coefficients using different numbers of years for temporal prior estimation. For the high confidence pixel group, the overall accuracies (and kappa values) were similar when using three to six years for temporal prior estimation – all were only slightly higher than those from the initial NN classification. The use of two years of CDLs actually led to slightly worse results compared to not integrating temporal prior information. In this study area, farmers may follow various crop rotation practices (e.g., corn-soybean, corn-corn-hay). The short-term (e.g., 2 years) cropping frequency at the pixel scale does not reveal the full picture of cropping patterns. This shows that a reduction of overall accuracy is possible if unrepresentative temporal prior information is used for the image classification. For the low confidence group, the overall accuracy ranged from 56.4% to 59.2% using two to six years for temporal prior estimation. Using four or six years of CDLs led to the highest overall accuracies.

D. *Impact of Threshold Selection*

For both spatial and temporal prior-adjusting approaches, additional threshold values were examined (0.5 and 0.9) to identify ambiguous pixels to be labelled as being in the low confidence group from the initial classification. For example, for the 2015 MODIS classification, a low threshold value of 0.5 led to a very small percent (3.0%) of MODIS pixels being labelled as low confidence. Applying of spatial- or temporal-prior adjustment for such a small percentage of pixels is thus not particularly meaningful in improving overall classification performance.

A relatively high threshold value (0.9) led to about 55% of total pixels labelled as low confidence. Following the

This is a post-print version of an article published in IEEE Journal of selected Topics in Applied Earth observations and remote sensing in November 2018. The original article can be found here: 10.1109/JSTARS.2018.2870978

procedures described above, the low confidence pixels from the initial NN classification were re-labeled using the estimated spatial prior probabilities. The integration of spatial prior improved overall accuracy by 1-2%, relative to the initial NN classification, for the groups labelled as low confidence, depending on the mapping years from 2009 to 2015. Such results were quite similar in levels of increased accuracy as the results from the use of the 0.75 threshold value. In practice, there is no easy way to determine which specific threshold value to use to label pixels as falling in the high or low confidence groups. A user may start from an arbitrary threshold value (e.g., 0.75) to examine whether the estimated spatial prior probability matches reasonably well with historical classification maps or other statistics [41]. Another method of threshold selection is to use a cross-validation approach. The validation data set can be used to determine the best threshold value among a range of threshold values. Finally, the users may arbitrarily choose a percentage of total pixels that needs to be rejected or labelled as low confidence. Integration of spatial or temporal prior could improve classification accuracy for these low confidence pixel groups. We also note that there are many other classification rejection options that could be utilized instead of Chow's rule [42]. They can be further evaluated in the context of prior integration for potential improvement of classification performance.

E. Impact of Training Sample Distribution and the Use of Smaller Training Sample Sizes

To evaluate impacts of training sample distribution on classification results, we conducted a sensitivity analysis by varying training sample sizes. Specifically, we arbitrarily selected more training samples (5 times the original sample size) for one class while keeping the same training sample size for the other two classes. Because the training sample distribution changed from an equal number per class, the network outputs were first divided by the percentage that each class is represented in the training set, and multiplied by the spatial or temporal prior probabilities estimated and then normalized. The impacts of training sample sizes on classification performance are shown in the following table 6. For ease of illustration, the impacts were analyzed for temporal prior integration only.

Compared to results from equal sample sizes (i.e., 1000 pixels per class), the increase in training sample size for soybean and other crops classes led to a reduction in overall accuracy. The initial NN misclassified a significant portion of corn pixels as soybean or other crops classes, because the training samples were imbalanced (favoring soybean or the other crops class) and do not represent class proportions in the full image. However, incorporating temporal prior largely corrected such error, especially for pixels labeled as low confidence in the initial NN classification, raising overall accuracies for low confidence pixels as much as 20%.

For our study, we mainly examined our image classification methods using moderate-sized training samples (e.g., >= 1000 pixels per class) to represent the spectral variability of individual crop types. Additional analyses were conducted to evaluate NN classification performance using smaller training sample sizes. Without integrating prior information, we found that the use of 50-200 training samples per class resulted in low overall accuracies (57%-61%). Although incorporating temporal prior still increased overall accuracy by around 3%, the overall performance was still worse than those obtained from 1000 samples per class. Therefore, it is important to use a relatively large training sample size to better represent the spectral variability of individual cover types and achieve higher classification accuracy. In addition, with increased data volume, a parallel and distributed implementation needs to be considered to improve efficiency [43], [44].

F. Logistic Regression Classification

To examine the robustness of the methods, we also applied the methods using another commonly used classifier, a logistic regression classifier. The same spatial and temporal prior adjustment methods were applied. Using 1000 pixels per class for training, the initial logistic regression generated overall accuracies of 48.1% to 53.8% for the low confidence group pixels, depending on the year (2009-2015). By integrating spatial prior, the overall accuracy increased to 50.4% to 56.9%. For the temporal prior, we focused on the year 2015 classification, the overall accuracy improved from 52.6% to 59.0% by applying temporal prior adjustment. Such level of improvement was very similar to those obtained using the NN-based images classification, suggesting high potential of our methods in terms of generalizability. Further studies are recommended to examine our prior adjustment methods to other classification algorithms (e.g. Support Vector Machine, Random Forest, etc.) and a larger study area (e.g., all Midwestern US states).

IV. CONCLUSION

This study examined NN-based image classification by integrating a rejection option and implementing prior probability adjustment methods to improve classification accuracy. By applying a threshold value to NN output signals, we divided input pixels to high confidence and low confidence groups. Pixels defined as low confidence were re-labelled based on image spatial prior probabilities, as estimated from the initial NN image classification, or temporal prior probabilities derived from ancillary data. These methods were tested for crop mapping tasks using multi-temporal MODIS images as inputs with CDLs as ground truth.

For a user-defined threshold value of 0.75, accuracy assessments for MODIS 2009-2015 classifications showed a large difference between overall accuracies for the high confidence pixel group (overall accuracy ~ 70%) and the low confidence group (overall accuracy ~ 50%). The spatial prior-adjusted NN method improved overall accuracy by 2.0-3.3% for the low confidence pixel group, depending on the mapping year. The main advantage of the spatial prior adjustment method is that it increases mapping accuracy without the need

This is a post-print version of an article published in IEEE Journal of selected Topics in Applied Earth observations and remote sensing in November 2018. The original article can be found here: 10.1109/JSTARS.2018.2870978

for ancillary data. With historical CDL data, we further examined pixel-by-pixel cropping frequency for corn and soybean from 2009 to 2014, and this temporal prior information was used to adjust NN output signals for 2015. This temporal prior-adjusting method resulted in 7% of improvement in overall accuracy, much higher than those obtained from the spatial-prior adjustment method, though it requires ancillary data to support the classification tasks.

REFERENCES

[1] M. Pal, "Random forest classifier for remote sensing classification", International Journal of Remote Sensing, vol. 26, no. 1, pp. 217-222, 2005.

[2] G. Mountrakis, J. Im and C. Ogole, "Support vector machines in remote sensing: A review", ISPRS Journal of Photogrammetry and Remote Sensing, vol. 66, no. 3, pp. 247-259, 2011.

[3] Y. Shao and R. Lunetta, "Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points", ISPRS Journal of Photogrammetry and Remote Sensing, vol. 70, pp. 78-87, 2012.

[4] J. Paola and R. Schowengerdt, "A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification", IEEE Transactions on Geoscience and Remote Sensing, vol. 33, no. 4, pp. 981-996, 1995.

[5] M. Friedl, D. McIver, J. Hodges, X. Zhang, D. Muchoney, A. Strahler, C. Woodcock, S. Gopal, A. Schneider, A. Cooper, A. Baccini, F. Gao and C. Schaaf, "Global land cover mapping from MODIS: algorithms and early results", Remote Sensing of Environment, vol. 83, no. 1-2, pp. 287-302, 2002.

[6] A. Puissant, J. Hirsch and C. Weber, "The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery", International Journal of Remote Sensing, vol. 26, no. 4, pp. 733-745, 2005.

[7] V. Mesev, " The use of census data in urban image classification", Photogrammetric Engineering and Remote Sensing, vol. 64, no. 5, pp. 431-436, 1998.

[8] D. McIver and M. Friedl, "Using prior probabilities in decision-tree classification of remotely sensed data", Remote Sensing of Environment, vol. 81, no. 2-3, pp. 253-261, 2002.

[9] R. Schowengerdt, Remote sensing: models and methods for image processing, Academic press, 2006.

[10] C. Homer, J. Dewitz, L. Yang, S. Jin, P. Danielson, G. Xian, J. Coulston, N. Herold, J. Wickham and K. Megown, " Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information", Photogrammetric Engineering and Remote Sensing, vol. 81, no. 5, pp. 345-354, 2015.

[11] S. Walsh, A. McCleary, C. Mena, Y. Shao, J. Tuttle, A. González and R. Atkinson, "QuickBird and Hyperion data analysis of an invasive plant species in the Galapagos Islands of Ecuador: Implications for control and land use management", Remote Sensing of Environment, vol. 112, no. 5, pp. 1927-1941, 2008.

[12] K. Sawaya, "Extending satellite remote sensing to local scales: land and water resource monitoring using high-resolution imagery", Remote Sensing of Environment, 2003.

[13] C. Wu and A. Murray, "Estimating impervious surface distribution by spectral mixture analysis", Remote Sensing of Environment, vol. 84, no. 4, pp. 493-505, 2003.

[14] H. Wang, Y. Shao and L. Kennedy, "Temporal generalization of sub-pixel vegetation mapping with multiple machine learning and atmospheric correction algorithms", International Journal of Remote Sensing, vol. 35, no. 20, pp. 7118-7135, 2014.

[15] Y. Shao, G. Taff and S. Walsh, "Comparison of Early Stopping Criteria for Neural-Network-Based Subpixel Classification", IEEE Geoscience and Remote Sensing Letters, vol. 8, no. 1, pp. 113-117, 2011.

[16] A. Moody, S. Gopal and A. Strahler, "Artificial neural network response to mixed pixels in coarse-resolution satellite data", Remote Sensing of Environment, vol. 58, no. 3, pp. 329-343, 1996.

[17] R. Duda, P. Hart and D. Stork, Pattern classification, 2nd ed. New York: John Wiley & Sons, 2000.

[18] M. Richard and R. Lippmann, "Neural Network Classifiers Estimate Bayesian a posteriori Probabilities", Neural Computation, vol. 3, no. 4, pp. 461-483, 1991.

[19] C. Chow, "On optimum recognition error and reject tradeoff", IEEE Transactions on Information Theory, vol. 16, no. 1, pp. 41-46, 1970.

[20] G. Fumera, F. Roli, and G. Giacinto, " Multiple reject thresholds for improving classification reliability", In Proceedings of Advances in Pattern Recognition: Joint IAPR International Workshops, pp. 863-871, 2000.

[21] N. Kussul, G. Lemoine, F. Gallego, S. Skakun, M. Lavreniuk and A. Shelestov, "Parcel-Based Crop Classification in Ukraine Using Landsat-8 Data and Sentinel-1A Data", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 9, no. 6, pp. 2500-2508, 2016.

[22] S. Skakun, N. Kussul, A. Shelestov, M. Lavreniuk and O. Kussul, "Efficiency Assessment of Multitemporal C-Band Radarsat-2 Intensity and Landsat-8 Surface Reflectance Satellite Imagery for Crop Classification in Ukraine", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 9, no. 8, pp. 3712-3719, 2016.

[23] F. Waldner, D. De Abelleyra, S. Verón, M. Zhang, B. Wu, D. Plotnikov, S. Bartalev, M. Lavreniuk, S. Skakun, N. Kussul, G. Le Maire, S. Dupuy, I. Jarvis and P. Defourny, "Towards a set of agrosystem-specific cropland mapping methods to address the global cropland diversity", International Journal of Remote Sensing, vol. 37, no. 14, pp. 3196-3231, 2016.

[24] B. Wardlow and S. Egbert, "Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. Central Great Plains", Remote Sensing of Environment, vol. 112, no. 3, pp. 1096-1116, 2008.

[25] Y. Shao, R. Lunetta, J. Ediriwickrema and J. Iiames, "Mapping Cropland and Major Crop Types across the Great Lakes Basin using MODIS-NDVI Data", Photogrammetric Engineering & Remote Sensing, vol. 76, no. 1, pp. 73-84, 2010.

[26] J. Ren, J. Campbell and Y. Shao, "Spatial and temporal dimensions of agricultural land use changes, 2001–2012, East-Central Iowa", Agricultural Systems, vol. 148, pp. 149-158, 2016.

[27] J. Ren, J. Campbell and Y. Shao, "Estimation of SOS and EOS for Midwestern US Corn and Soybean Crops", Remote Sensing, 9(7), 722; doi:10.3390/rs9070722, 2017.

[28] J. Knight, R. Lunetta, J. Ediriwickrema and S. Khorram, "Regional Scale Land Cover Characterization Using MODIS-NDVI 250 m Multi-Temporal Imagery: A Phenology-Based Approach", GIScience & Remote Sensing, vol. 43, no. 1, pp. 1-23, 2006.

[29] C. Alcantara, T. Kuemmerle, A. Prishchepov and V. Radeloff, "Mapping abandoned agriculture with multi-temporal MODIS satellite data", Remote Sensing of Environment, vol. 124, pp. 334-347, 2012.

[30] P. Eilers, "A Perfect Smoother", Analytical Chemistry, vol. 75, no. 14, pp. 3631-3636, 2003.

[31] C. Atzberger and P. Eilers, "A time series for monitoring vegetation activity and phenology at 10-daily time steps covering large parts of South America", International Journal of Digital Earth, vol. 4, no. 5, pp. 365-386, 2011.

[32] Y. Shao, R. Lunetta, B. Wheeler, J. Iiames and J. Campbell, "An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data", Remote Sensing of Environment, vol. 174, pp. 258-265, 2016.

This is a post-print version of an article published in IEEE Journal of selected Topics in Applied Earth observations and remote sensing in November 2018. The original article can be found here: 10.1109/JSTARS.2018.2870978

[33] C. Boryan, Z. Yang, R. Mueller and M. Craig, "Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program", Geocarto International, vol. 26, no. 5, pp. 341-358, 2011.

[34] P. Atkinson, M. Cutler and H. Lewis, "Mapping sub-pixel proportional land cover with AVHRR imagery", International Journal of Remote Sensing, vol. 18, no. 4, pp. 917-935, 1997.

[35] L. Bruzzone, D. Prieto and S. Serpico, "A neural-statistical approach to multitemporal and multisource remote-sensing image classification", IEEE Transactions on Geoscience and Remote Sensing, vol. 37, no. 3, pp. 1350-1359, 1999.

[36] C. Bishop, Neural networks for pattern recognition, Oxford university press, 1995.

[37] R. Lunetta, Y. Shao, J. Ediriwickrema and J. Lyon, "Monitoring agricultural cropping patterns across the Laurentian Great Lakes Basin using MODIS-NDVI data", International Journal of Applied Earth Observation and Geoinformation, vol. 12, no. 2, pp. 81-88, 2010.

[38] R. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data", Remote Sensing of Environment, vol. 37, no. 1, pp. 35-46, 1991.

[39] B. Gorte and A. Stein, "Bayesian classification and class area estimation of satellite images using stratification", IEEE Transactions on Geoscience and Remote Sensing, vol. 36, no. 3, pp. 803-812, 1998.

[40] J. Ju, E. Kolaczyk and S. Gopal, "Gaussian mixture discriminant analysis and sub-pixel land cover characterization in remote sensing", Remote Sensing of Environment, vol. 84, no. 4, pp. 550-560, 2003.

[41] L. Bruzzone and S. Serpico, "An iterative technique for the detection of land-cover transitions in multitemporal remote-sensing images", IEEE Transactions on Geoscience and Remote Sensing, vol. 35, no. 4, pp. 858-867, 1997.

[42] C. De Stefano, C. Sansone and M. Vento, "To reject or not to reject: that is the question-an answer in case of neural classifiers", IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), vol. 30, no. 1, pp. 84-94, 2000.

[43] Z. Wu, Y. Li, A. Plaza, J. Li, F. Xiao and Z. Wei, "Parallel and distributed dimensionality reduction of hyperspectral data on cloud computing architectures", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9, no. 6: 2270-2278, 2016.

[44] Z. Wu, L. Shi, J. Li, Q. Wang, L. Sun, Z. Wei, J. Plaza, and A. Plaza, "GPU parallel implementation of spatially adaptive hyperspectral image classification", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017.