remote sensing

Article

Evaluation of ALOS PALSAR Data for High-Resolution Mapping of Vegetated Wetlands in Alaska

Daniel Clewley [1],*, **Jane Whitcomb** [1], **Mahta Moghaddam** [1], **Kyle McDonald** [2], **Bruce Chapman** [3] **and Peter Bunting** [4]

[1] Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089, USA;
   E-Mails: jbwhitco@usc.edu (J.W.); mahta@usc.edu (M.M.)

[2] The City College of New York, The City University of New York, NY 10031, USA;
   E-Mail: kmcdonald2@ccny.cuny.edu

[3] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA;
   E-Mail: bruce.d.chapman@jpl.nasa.gov

[4] Department of Geography and Earth Sciences, Aberystwyth University, Aberystwyth, Ceredigion,
   Wales SY23 3DB, UK; E-Mail: pfb@aber.ac.uk

* Author to whom correspondence should be addressed; E-Mail: daniel.clewley@gmail.com;
   Tel.: +44-1752-633-100.

Abstract: As the largest natural source of methane, wetlands play an important role in the carbon cycle. High-resolution maps of wetland type and extent are required to quantify wetland responses to climate change. Mapping northern wetlands is particularly important because of a disproportionate increase in temperatures at higher latitudes. Synthetic aperture radar data from a spaceborne platform can be used to map wetland types and dynamics over large areas. Following from earlier work by Whitcomb *et al.* (2009) using Japanese Earth Resources Satellite (JERS-1) data, we applied the "random forests" classification algorithm to variables from L-band ALOS PALSAR data for 2007, topographic data (e.g., slope, elevation) and locational information (latitude, longitude) to derive a map of vegetated wetlands in Alaska, with a spatial resolution of 50 m. We used the National Wetlands Inventory and National Land Cover Database (for upland areas) to select training and validation data and further validated classification results with an independent dataset that we created. A number of improvements were made to the method of Whitcomb *et al.* (2009): (1) more consistent training data in upland areas; (2) better distribution of training data across

all classes by taking a stratified random sample of all available training pixels; and (3) a more efficient implementation, which allowed classification of the entire state as a single entity (rather than in separate tiles), which eliminated discontinuities at tile boundaries. The overall accuracy for discriminating wetland from upland was 95%, and the accuracy at the level of wetland classes was 85%. The total area of wetlands mapped was 0.59 million km^2, or 36% of the total land area of the state of Alaska. The map will be made available to download from NASA's wetland monitoring website.

Keywords: wetlands; SAR; Alaska; random forests; PALSAR

1. Introduction

Wetlands play an important role in global climate models as both carbon stores and methane (CH_4) emitters. They represent the largest single natural source of CH_4 [1], accounting for 20%–45% of total emissions [2]. As CH_4 has 25-times the global warming potential of carbon dioxide (CO_2) [1] and current models predict increased CH_4 emissions from wetland areas in response to rising CO_2 emissions [2,3], a better understanding of current emissions and changes in response to climate change is particularly important. In addition to their role in the global greenhouse gas cycle, wetlands are home to a number of plant and animal species and are important to the hydrological cycle [4].

Although temperatures are increasing globally, in the Arctic, they have been increasing at a rate almost double the average, due to what is known as "Arctic amplification" [5]. In the state of Alaska, accelerated warming has been observed between 1970 and 2000 [6]. In subsequent years, temperatures have continued to rise in northern Alaska, although a decrease in temperature has been observed in the south [7]. This disproportionate increase in temperatures, at high latitudes in general and in Alaska in particular, makes Alaskan wetlands especially vulnerable to climate change.

It has been estimated that wetlands cover 43% of the surface area of Alaska [4], representing a greater total area of wetlands than in all of the conterminous United States (U.S.) [8]. High-resolution maps are critical to understanding the response of wetlands to climate change [9]. Maps produced decadally, in sufficient detail with respect to the Cowardin *et al.* [10] wetland classification scheme, have been recognized as being beneficial for CO_2 budgets [8]. To generate maps meeting these requirements, it is necessary to develop a methodology that can be applied to produce maps for multiple years, as required for a monitoring system [11]. As different types of wetlands sequester and emit at different rates [8], it is important to discriminate between wetland types in mapping efforts. This is particularly important for monitoring changes in wetland type, with previous studies noting increases in shrub abundance [12], a reduction in size or loss of water bodies [13,14] and drying of wetland areas [15] within Alaska.

The use of spaceborne synthetic aperture radar (SAR) provides a number of advantages for mapping and monitoring the extent and type of wetlands over large areas and at a relatively high spatial resolution (20–100 m), including the ability to acquire data regardless of illumination conditions or cloud cover and a sensitivity to vegetation structure and moisture content, particularly at L-band wavelengths [16].

Although the use of remote sensing has been established in mapping wetlands in the United States, through the National Wetlands Inventory (NWI) [17,18], the methodological approach has primarily been through manual interpretation of aerial photographs. Photointerpretation is a time-consuming process, making it costly to generate maps over large areas, particularly if they need to be updated regularly. Therefore, more automated techniques are required. There are a number of algorithms available that can be applied to generate thematic maps from remotely-sensed data. These include unsupervised (e.g., k-means and ISODATA), supervised (e.g., maximum likelihood), rule or knowledge-based [19,20] and machine learning approaches (e.g., support vector machines). Random forests [21] is a machine learning approach capable of handling discrete (thematic) and continuous input data. Fernandez-Delgado *et al.* [22] compared a large number of classification algorithms and implementations and found random forests to produce the best accuracy across a number of datasets. Random forests has previously been used for classifying remotely-sensed data in a number of studies (e.g., [23,24]) and has demonstrated the capability to generate classifications with a high accuracy. A particular advantage of random forests over other machine learning algorithms, such as support vector machines (SVM), is that it only requires a small number of tuning parameters [25] and is computationally efficient. Random forests was applied by Whitcomb *et al.* [26] to data from the Japanese Earth Resources Satellite (JERS-1) and ancillary layers to derive a map of wetlands in Alaska and was found to produce maps with a higher accuracy than applying unsupervised (ISODATA) and supervised (maximum likelihood) algorithms. Following from the successful application of random forests to map wetlands from JERS-1 data and ancillary data [26], subsequent studies focused on applying the same method to data from the Phased-Array L-band SAR (PALSAR) carried onboard the Advanced Land Orbiting Satellite (ALOS) [27] and expanding the technique to also include wetlands in Canada [28].

Although the original [26] classification in Alaska and subsequent work demonstrated that the technique was capable of providing highly accurate (∼90%) maps of wetland type and represented a significant improvement over existing mapping in the area, the approach had a number of limitations that needed to be addressed. One major problem was that the method necessitated breaking the mosaic into sixteen tiles. For each tile, a separate "random forest" was generated from training data within that tile and applied. This meant that only those classes for which training data fell within the tile were considered; any other classes were omitted. Given that a separate classification, with a different subset of classes, was applied for each tile, discontinuities became apparent when the classified tiles were combined to create the mosaicked map. These problems persisted in subsequent attempts to classify small regions of PALSAR imagery using the methodology presented in Whitcomb *et al.* [26], such that production of a complete map for Alaska with the code of [26] would not have been possible. Whitcomb *et al.* [27] thus presented PALSAR classification results only for key areas. In Clewley *et al.* [11], the code used for pre-processing and classification was re-written to address these and other limitations, resulting in a greatly improved software suite, particularly as regards the manipulation of large datasets. The updated software was used to produce an improved map derived from JERS-1 data and ancillary data and an initial map from ALOS PALSAR data and ancillary data, both at 100-m spatial resolution.

Building on previous work [11,26], this study has aimed to produce an enhanced classification of vegetated wetlands in Alaska based on ALOS PALSAR data from 2007 and ancillary data at a higher

spatial resolution than existing maps. The method presents a number of improvements to that of Whitcomb *et al.* [26], both to the input and training data (described in Section 3) and to the method of applying random forests to the entire state at once by using stratified sampling (described in Section 4). This classification was then used to provide an estimate of the extent of vegetated wetlands in Alaska, as of 2007, and the proportion within each wetland class.

2. Study Area

Alaska, which spans latitudes from 51° to 72°N, is the largest state in the United States with a total surface area of over 1.6 million km^2 [4] and more than 53,000 km of shoreline [29]. The range of latitudes within Alaska leads to variations in climate spanning arctic to subarctic conditions. Permafrost is an important feature of Alaska, with continuous permafrost occurring in the north and discontinuous or sporadic permafrost further south [29].

Alaska can be divided into seven broad physiographic units: (1) northwest Alaska, with moist and wet tundra types (*Eriophorum* spp.), ericaceous shrub polygons and saline meadows; (2) arctic Alaska, which includes extensive wet tundra and wet sedge meadows (*Eriophorum angustifolium* and *Carex aquatilis*); (3) south-central Alaska, which ranges from the peaks of the Alaska Range to coastal marshes and includes forest cover with extensive areas of black spruce muskeg; (4) southwest Alaska, which includes wet sedge meadows, halophytic wet meadows and wet shrub tundra; (5) southeast Alaska, where forest cover includes extensive regions of black spruce muskeg with halophytic and freshwater sedge and wet meadows dominated by *C. lyngbyei* on coastal deltas; (6) interior Alaska, with extensive black spruce muskeg forest cover, subarctic lowland sedge and sedge-moss bog meadows; and (7) the Aleutian Islands, where the most widespread community is *Empetrum* heath [30].

3. Data

3.1. ALOS PALSAR Data

Data from ALOS PALSAR, the successor to JERS-1, formed the base layer of the classification. PALSAR was an L-band (24-cm wavelength) SAR sensor operational during 2006–2011. It had similar characteristics to JERS-1 [31], with a number of enhancements, including the ability to acquire fully polarimetric data (horizontal and vertical transmit and receive), as well as enhancements to geometric and radiometric accuracy [32]. Under the framework of the Kyoto & Carbon Initiative (K & C; [33]) a mosaic was generated for Alaska, primarily using data from the summer of 2007, but with scenes from 2008, 2009 and 2010 used to fill in missing data. Only summer data were available, unlike Whitcomb *et al.* [26], for which a summer and winter JERS-1 mosaic were available through the Global Boreal Forest Mapping Campaign [34].

Data were acquired in fine beam dual (FBD) mode at HH- and HV-polarization with a look angle of 34.3° and were supplied as Level 1.1 format (single look complex). These data were multi-looked, registered to a DEM and topographically corrected to a spatial resolution of 1 arc s (∼30 m) using GAMMA Remote Sensing's GAMMA software [35,36]. The scenes were mosaicked to create a single image, with the acquisition date of each pixel retained as a separate layer. To compensate for a

drop off in power towards the edge of each swath, the maximum pixel value in the overlap between swaths was used, implemented as part of the mosaic algorithm in the Remote Sensing and GIS Library (RSGISLib; [37]). This approach reduced artifacts between strips, without the need for color balancing (as shown in Figure 1), although variations of as much as several decibels remained due to differences in environmental conditions at the time of acquisition (e.g., soil and vegetation moisture). The mosaic was created using scenes from 2007 (86.7% of all pixels), with scenes from 2008 (7.5%), 2009 (4.3%) and 2010 (1.4%) used to fill data gaps. Scenes were mostly acquired during the summer months, with the earliest scene from June and the latest from September; the majority of the data were collected from late July through the end of August. The mosaic was reprojected to Albers Conical Equal Area projection and resampled to a pixel size of 50 × 50 m using the average pixel value. The same grid as the map produced by Whitcomb *et al.* [26] and Clewley *et al.* [11] was used to make comparisons easier between the two maps.
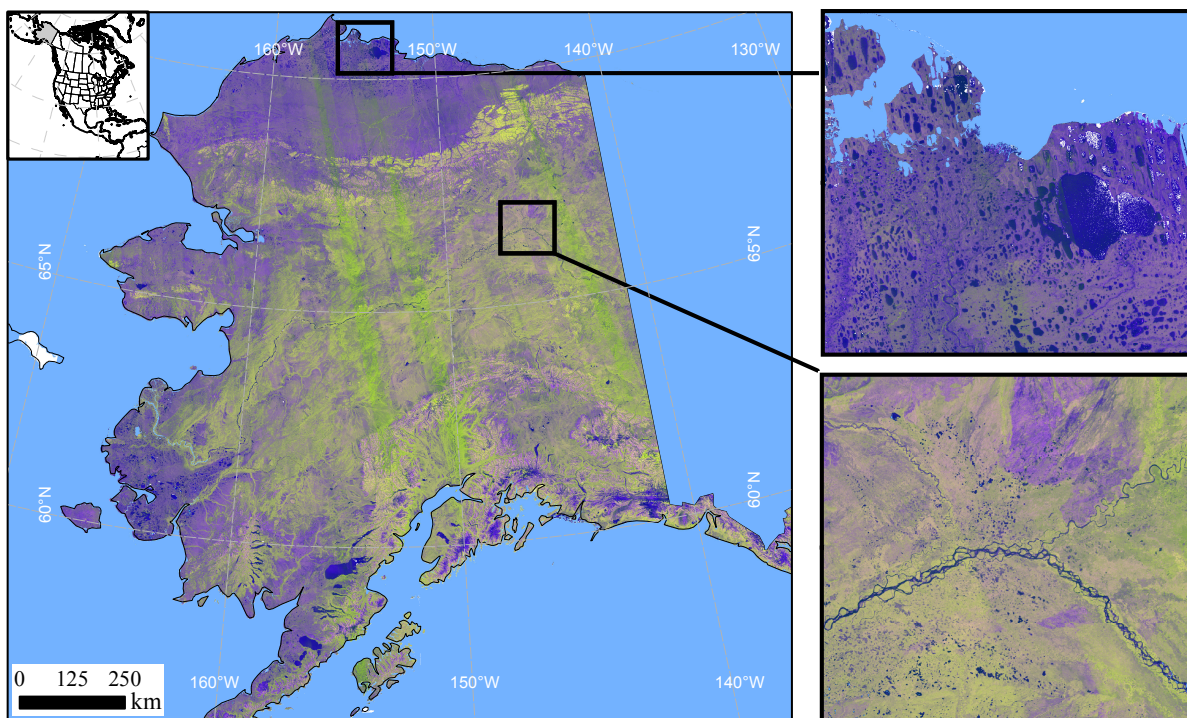

Figure 1. ALOS PALSAR mosaic, primarily using data from 2007, shown as a false-color composite of HH (red), HV (green) and HH/HV (blue). Green areas indicate denser vegetation. Some vertical striping between strips is visible, likely due to changes in ground conditions between overpasses.

The role of the HH and HV PALSAR images in the classification were to capture dynamic variations in wetlands water extent, vegetation structure, vegetation water content and soil moisture. Previous analysis in Clewley *et al.* [11] indicated that the normalized backscattering coefficient (σ^0) at both polarizations increased from emergent to shrub/scrub to forested wetland classes, that is with increasing biomass. Some vegetation categories also exhibited different levels of σ^0 in different types of wetland systems (e.g., emergent vegetation exhibited a lower σ^0 at HH-polarization in the lacustrine class than in the palustrine class). However, the differences between wetland systems were not sufficient to allow separation using σ^0 alone.

3.2. Ancillary Data

Due to the difficulty in accurately separating all wetland classes using L-band SAR data alone, in Whitcomb *et al.* [26], a number of ancillary layers were included as part of the classification to aid discrimination, including SAR texture, SAR acquisition date, slope, elevation, proximity to water, latitude and longitude. We chose to incorporate these layers, as well, because they provided information relevant to the Cowardin *et al.* [10] classification scheme or relevant to the characteristics of the SAR data (*e.g.*, SAR acquisition date).

3.2.1. SAR Texture and Acquisition Date

The coefficient of variation (standard deviation divided by the mean) measure of texture was computed over a 3×3-pixel window from the PALSAR data to provide a measure of local SAR brightness variability. Such texture layers reveal landscape structural patterns that are characteristic of different wetland types. We also incorporated a layer representing the acquisition date of each scene to help compensate for temporal differences that affect backscatter responses and SAR texture, due to seasonality or changes in environmental conditions (*i.e.*, moisture) between different dates.

3.2.2. Slope and Elevation

Elevation is an important data source within the classification, as high-elevation wetlands often support significantly different mixes of species than low-elevation wetlands, even within the same geographic region [38]. The National Elevation Dataset (NED; [39,40]) digital elevation model (DEM) was used for this information. Data were provided for Alaska at 2-arc s (~60 m) resolution in 1×1 degree tiles through the U.S. Geological Survey (USGS) bulk data distribution program; these were mosaicked to create a single file.

A slope layer was calculated from the DEM based on the change in elevation over a 3×3-pixel window. The slope calculation was performed prior to reprojection so that it would not be corrupted by residual geometric errors resulting from the reprojection process. Horizontal distances needed in the slope computation (expressed in m) were converted from the decimal degrees of the original geographic coordinate system on a pixel-by-pixel basis.

The slope layer was used as an input into the classification and to differentiate wetland from non-wetland areas. Only areas with a slope less than 3.8° were considered possible wetlands, with this value chosen as it represented the 75th percentile of all of the slope pixels extracted from all NWI wetland polygons.

3.2.3. Water Mask and Proximity to Water

A water mask was included to eliminate areas of open water from the classification. The water mask was generated by applying a rule-based classification to the PALSAR data and slope layer, unlike in Whitcomb *et al.* [26], where the water mask was generated from only the JERS-1 imagery using a supervised maximum likelihood estimator (MLE) classification approach [41]. Water was assumed to have a slope $<3°$ and HH-backscattering coefficient < -14 dB. Because the water mask and its derivative

proximity to water layer are used in the main wetlands classification, this rule-based classification had to be performed prior to the main classification. A proximity to water layer, providing the distance of each pixel to the nearest pixel of water (expressed in m), was then generated from the water mask. It served as an additional layer in the wetlands classification that helped to distinguish wetland systems (estuarine, riverine, lacustrine, *etc.*) from each other and from uplands.

Although other water masks were available for Alaska, such as the MODIS-derived MOD44W product [42] and the National Hydrography Dataset (NHD; [43]), we created one based on the PALSAR data, so it would be at the same resolution as the classification (unlike the MOD44W product, which is at a coarser resolution of 250 m) and would represent the state of water bodies for the same period as the rest of the mapping. The PALSAR-derived water mask was compared with each of these pre-existing water masks on a pixel-by-pixel basis to generate a land/water confusion matrix. Good agreement was observed between the PALSAR-derived water mask and both the MOD44W and NHD products, with accuracies of 98% and 97%, respectively. One of the main causes of discrepancy was the different scales of the three maps.

3.2.4. Latitude and Longitude

Because geographic location has a strong impact on vegetated wetlands and the state of Alaska covers a large number of biogeographical zones with differing climates and species compositions, we included positional information as part of the classification. One way to incorporate positional information is to divide the study area into separate biogeographical zones, as in Lucas *et al.* [20]. Rather than applying different classifications to different areas, we included layers providing the latitude and longitude of each pixel as input to random forests. We preferred latitude and longitude, because they are continuous variables and therefore less likely to create hard boundaries within the classification, as can happen with a thematic regional ecosystems layer (e.g., [44]).

3.3. Training Data

The primary source of data for training and validating the classification was the NWI dataset produced by the U.S. Fish and Wildlife Service. The NWI currently is the only available source of high-resolution geospatial data on the types and extents of wetlands in Alaska that includes at least some data for all Level I terrestrial ecoregions in a consistent set of wetland classes [45]. The NWI categorizes wetlands according to the Cowardin classification system [10] and has been derived for the United States from photointerpretation of aerial photography.

The Cowardin system separates wetlands into classes based on major wetland systems and vegetation type. These basic classes are augmented by modifiers that identify supplementary wetland traits, such as water regime. The four major systems it defines are: estuarine (tidal, semi-closed by land with partial access to ocean), riverine (non-tidal, contained within a channel of moving fresh water), lacustrine (non-tidal, situated in topographic depressions, including wetlands bordering freshwater lakes) and palustrine (non-tidal, dominated by vegetation with shallow water depth). Within these broad systems, the following vegetation classes are defined for Alaska: moss-lichen, emergent (herbaceous plants protruding from saturated soil or the water surface, typically dominated by *Carex* sedges and cotton-grass

(*Eriophorum*)), scrub-shrub (including shrubs, such as dwarf alder (*Alnus*) and willow (*Salix*)) and forested (most frequently containing white spruce and/or black spruce (*Picea glauca, P. mariana*)). As only a subset of these combinations of wetland system and vegetation cover a significant area in the Alaska NWI dataset, our classification was based on nine classes (Table 1). Each of these nine classes had a water regime modifier (permanently flooded, seasonally flooded, saturated, *etc.*), creating a total of 23 classes. The water regime modifiers initially were used as part of the classification, but due to the difficulty in accurately discriminating them, they were aggregated post-classification.

Table 1. Area of wetland classes available in the National Wetlands Inventory (NWI) data used to train the classification. Only a subset of these data was actually used as input to the classification, due to the stratified random sampling approach used.

Class	Area (km^2)	Percent by Area of NWI Data
Estuarine emergent (E2EM)	1608	0.84
Estuarine scrub-shrub (E2SS)	65	0.03
Estuarine forested (E2FO)	5	0.00
Riverine emergent (REM)	4	0.00
Lacustrine emergent (L2EM)	265	0.14
Palustrine moss-lichen (PML)	138	0.07
Palustrine emergent (PEM)	108,881	56.58
Palustrine scrub-scrub (PSS)	58,980	30.65
Palustrine forested (PFO)	22,478	11.68
Total	192,424	

The accuracy of the NWI has been evaluated at two levels, wetland identification (correct discrimination of wetlands and non-wetlands) and wetland attribution (correct identification of wetland type). Previous studies of NWI accuracy based on targeted field surveys in Virginia and around the Great Lakes [17,18] found the accuracy of wetland identification to be greater than 90%, with lower accuracy associated with forested wetlands, which are difficult to discriminate using photointerpretation [46]. Within Alaska, the NWI dataset covers 12% of the entire state (192,400 km^2; Figure 2) and was produced mainly from color-infrared (CIR) and some true color aerial photographs acquired between 1974 and 2007, with the majority from the late 1970s and early 1980s. A single polygon layer was supplied for the entire state. We removed wetland polygons less than 0.25 ha (1 pixel) and assigned a unique integer code to each class of the remaining polygons. We generated a blank raster with a 50-m resolution and assigned those pixels where the center fell within an NWI polygon to the integer code for the polygon's wetland class.

To avoid confusion with non-wetland classes, we included training data over non-wetland areas. For this, we used the National Land Cover Database (NLCD; [47]), produced from Landsat 5 TM and Landsat 7 ETM+ data from around 2001. The NLCD was chosen over the Alaska Geospatial Data Clearinghouse data used in Whitcomb *et al.* [26], as it is synoptic across Alaska and is provided at a comparable resolution (30 m) to the PALSAR data. The overall accuracy of the NLCD dataset in Alaska was found to be 76.2%, based on an aerial survey [48]. From the NLCD classification, only the barren land (BL, comprising areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial

debris, sand dunes, strip mines, gravel pits and other accumulations of earthen material, with a vegetation cover generally less than 15%), deciduous forest (DF) and coniferous forest (CF) classes were chosen. Other classes were excluded to avoid confusion with similar wetland classes. The user's accuracy of these classes in Alaska was found by Selkowitz and Stehman [48] to be 85% (BL), 60% (DF) and 84% (CF). The NLCD was combined with the NWI data to produce a single training data layer; where there were discrepancies between the NWI and NLCD label for a given pixel, the NWI label was chosen. The combined training data provided 26 land cover classes (Figure 2), with the wetland classes displayed aggregated to wetland system-vegetation classes to form the 12 classes used in the final classification.
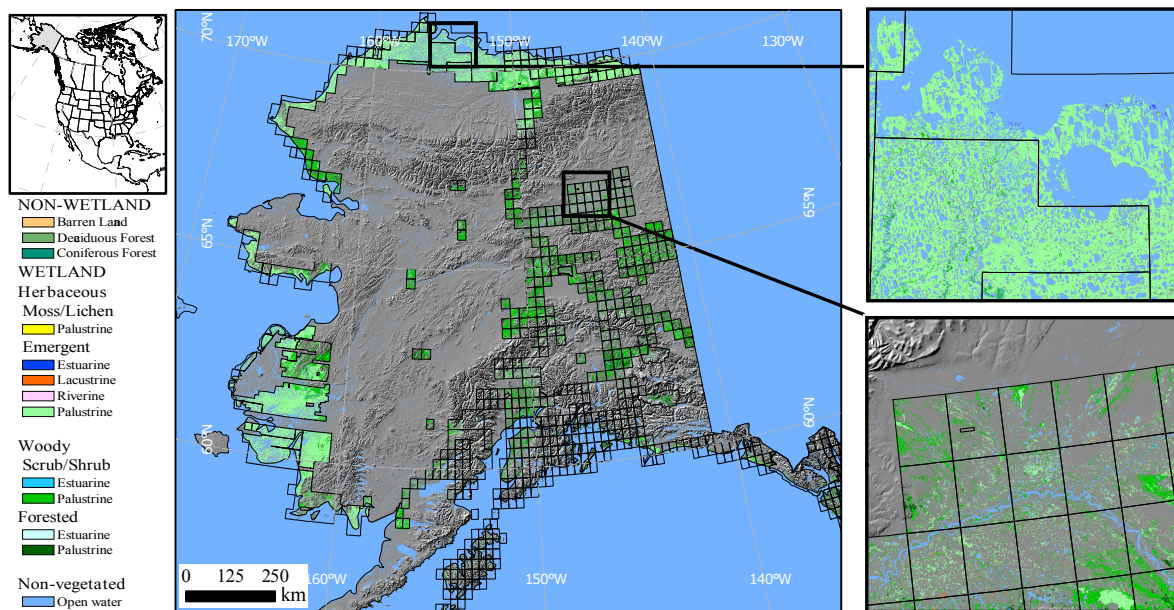

Figure 2. The complete set of training data available, compiled mainly from the National Wetlands Inventory (NWI) data, but with data from the National Land Cover Database (NLCD) used for non-wetland areas. Polygons show map quadrangles for the NWI data.

4. Methods

4.1. Classification

The data ingestion and classification process is shown in Figure 3. All data layers were reprojected to Albers Conical Equal Area projection and resampled to a spatial resolution of 50 m using the average of overlapping pixels. As part of resampling and reprojection, all data layers were converted to the KEAfile format [49] to make use of lossless compression.

As in Whitcomb *et al.* [26], the classification was produced with random forests [21]. The algorithm is an extension of the classification and regression trees (CART; [50]) algorithm utilizing multiple decision trees (a forest) generated from random samples within a set of training data. Each decision tree is constructed by randomly selecting N samples (where N is the size of the pool of samples submitted to random forests) with replacement (bootstrap aggregation, bagging), so as not to alter the characteristics of the pool as selections are made [51]. The $\sim 2N/3$ samples that get chosen at least once are used to build the tree, while the remaining samples (out-of-bag samples) are used to validate the tree.

The combination of out-of-bag errors across all trees is used to evaluate overall classification accuracy. As part of the random forests algorithm, the importance of each variable is calculated and summarized over all trees as the mean decrease in the Gini impurity index. Optionally, random forests also computes the importance of each variable based on the mean decrease in accuracy observed when values of the variable are randomly permuted; the latter measure of importance is calculated overall and also on a class-by-class basis.
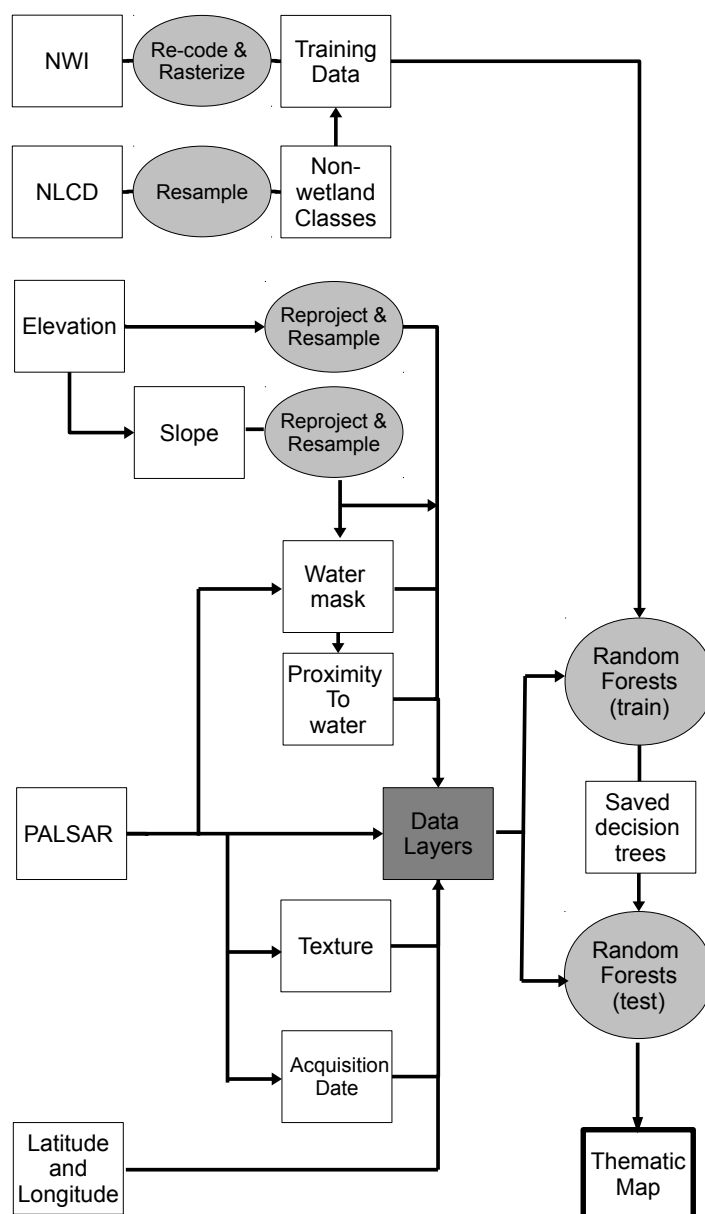

Figure 3. The pre-processing and classification processes used to produce a thematic map from PALSAR and ancillary data. Multiple layers used as input to the classification were derived from the PALSAR data. The NWI and NLCD data were used to train the random forests classifier. Decision trees generated by the random forests classifier were then applied to generate the classification.

When all available training pixels were used, there was a large disparity in the amount of samples available for each class, with 87% of wetland training data falling within two classes. To produce

a more even distribution of samples across all classes, a stratified random sample of training data was taken, with the maximum number of samples in each class set to 100,000, as in [11]. This had two purposes: (1) to reduce the amount of data used as input into random forests, as the full dataset required large amounts of memory and computational time to form each tree; and (2) to adjust for the skewed distribution of training samples.

The "randomForest" package in R [52,53] was used with rGDAL [54] to read and write the data. The number of data layers (m) used for each tree was set to four based on a sensitivity analysis around the recommended starting parameter of $m = \sqrt{M}$ (where M is the total number of layers) to minimize the out-of-bag error without causing artifacts in the classification. The number of trees was set to 300 because, although a large number of trees are preferred, experiments showed that there was little increase in the out-of-bag accuracy after 200 trees.

Following generation of the "random forest" from the available training data, we applied the forest to the stack of data layers to produce an output classification. For each pixel, a class was generated from each decision tree by executing the sequence of comparisons of data layer values to thresholds that constitute that decision tree. Once the pixel had been classified by all trees, the final class was assigned based on the class with the greatest number of votes across all decision trees.

4.2. Accuracy Assessment

The accuracy of the classification was evaluated with confusion matrices. Two methods were used to generate validation data. The first method used out-of-bag samples from random forests, and the second method used a set of points randomly selected from the available training data and not used as input to random forests. For the second case, a maximum of 100 points from each class of the training data were randomly selected. For each point, a vegetation class was assigned based on manual interpretation of high-resolution optical data available through GoogleMaps; the assignment was carried out without reference to the original classification. Points on the boundary between classes or where it was difficult to assign a class based on the available imagery were dropped. The class manually assigned to each point was compared with the assignment from the NWI/NLCD classification, retaining only those points for which the two classes matched. Only the vegetation class was considered in the comparison; major wetland systems (estuarine, palustrine *etc.*) were assumed to be correct and not to have changed since the original mapping, which used mostly data from the late 1970s and late 1980s. Through this procedure, 888 points were produced for validation. The wetland class in the PALSAR-based map was extracted from each point and used to create a confusion matrix.

4.3. Area Calculation

Whereas a simple count of pixels within each wetland class would provide an estimate of total wetland area and the proportion within each wetland class, this approach is biased due to errors in the classification [55,56]. To provide an unbiased estimate of area, it is necessary to adjust the pixel counts based on the accuracy of each class, as provided by the confusion matrix. As in

Carreiras *et al.* [57], the inverse method was used [55,56]. The adjusted area of each class (A_i') is given as:

$$A_i' = \sum_{j=1}^{k} A_j \left(n_{ij}/n_j \right) \tag{1}$$

where i is the observed class (NWI data, row in confusion matrix) and j the predicted class (column in confusion matrix).

4.4. Variable Sensitivity Analysis

To provide an additional evaluation of the importance of different variables within the classification, multiple runs were performed with a subset of all available data layers. In the first, we used the NHD water mask and generated a proximity-to-water layer from this instead of from our PALSAR-derived water mask. In the second, we excluded the proximity-to-water layer from the classification. A third classification run was performed using only the PALSAR-derived data layers. In a fourth classification run, the latitude and longitude layers were replaced by a thematic layer generated using the ecoregions map of Gallant [44], with an integer code assigned to each ecoregion. In the fifth run, no positional information was included, and in the sixth run, no topographic information was included. As with the main classification, the accuracy for these reduced/alternative dataset classification runs was evaluated using out-of-bag samples and the separate points verified against high-resolution optical data.

5. Results

5.1. Classification

The map of wetlands in Alaska, based on PALSAR data from 2007 and ancillary data on topography and geographic location, is presented in Figure 4. Palustrine scrub-shrub wetlands and palustrine forested wetlands are dominant in the interior, with palustrine emergent wetlands dominant along the coast. Lining coastal outlets are estuarine emergent wetlands with patches of estuarine scrub-shrub wetlands. Lacustrine emergent wetlands are visible around the edges of lakes.

5.2. Accuracy Assessment

The confusion matrix generated as part of the random forests algorithm using out-of-bag samples and aggregated across water regimes is shown in Table 2. The overall accuracy was 84.5%. In comparing this overall accuracy to the 89.5% accuracy achieved by Whitcomb *et al.* [26], it should be noted that the aggregate classes presented here are more detailed than those used by Whitcomb *et al.* [26], in which individual Cowardin wetlands classes were aggregated into four broad categories (herbaceous, scrub-shrub, forested and barren) prior to the computation of accuracy statistics. That previous classification also did not include a coniferous upland class. Discrimination between wetlands and non-wetlands yielded an overall accuracy of 94.7%.

Table 2. Aggregated confusion matrix by wetland class, generated from out-of-bag samples for the classification resulting from all input variables. See Table 1 for the key to wetland codes. For uplands: BL = barren land; DF = deciduous forest; CF = coniferous forest.

| Observed	Predicted												Producer's Accuracy (%)
	E2EM	E2SS	E2FO	REM	L2EM	PML	PEM	PSS	PFO	BL	DF	CF	
E2EM	160,086	876	61	0	77	12	1301	236	369	675	71	1980	96.59
E2SS	439	11,506	3	0	0	0	18	4	11	5	0	78	95.37
E2FO	59	6	981	0	0	0	1	0	5	0	0	25	91.09
REM	0	0	0	667	27	0	248	6	0	0	0	0	70.36
L2EM	69	0	0	0	76,604	1	3557	365	25	73	274	415	94.13
PML	10	0	0	0	14	40,420	145	488	677	15	43	232	96.14
PEM	5133	322	10	56	9501	613	337,854	29,591	5900	3774	3283	4204	84.41
PSS	1397	211	18	14	923	2241	21,000	324,097	24,981	9304	7016	8798	81.02
PFO	303	30	10	0	27	438	1814	11,577	241,121	1582	2353	7537	90.38
BL	2115	25	1	1	550	31	3048	6191	2550	80,037	2426	3025	80.04
DF	464	1	1	0	291	108	4774	5806	7011	2253	52,626	6512	65.91
CF	692	91	7	0	380	362	3923	6749	8674	2387	7923	68,812	68.81
User's accuracy (%)	93.75	88.05	89.84	90.38	86.66	91.39	89.45	84.16	82.77	79.95	69.23	67.72	**84.53**
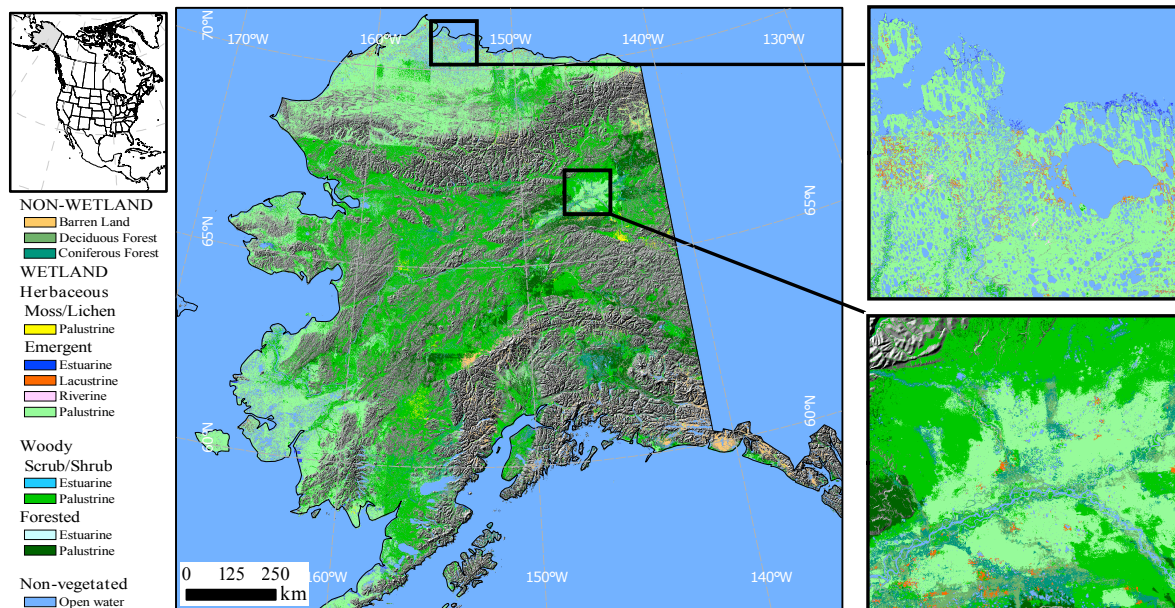
Figure 4. Map of wetlands in Alaska generated using ALOS PALSAR data from 2007 and ancillary information on elevation, slope and location. Insets show sections of Teshekpuk Lake near Barrow (upper) and Yukon Flats (lower). The classification is displayed over a hillshade image (grey) derived from the National Elevation Dataset (NED) DEM to show the topography of non-wetland areas.

Classification accuracy for the forested wetland classes was greatly improved in this study relative to that achieved in Whitcomb *et al.* [26]. Specifically, the accuracy for estuarine forested wetlands was about 90% (both producer's and user's), and that for palustrine forested wetlands was ∼90% for producer's, ∼83% for user's. This compares to the ∼81% producer's accuracy and ∼74% user's accuracy of [26] for the aggregate combination of forested wetland classes. Performance for these classes was likely enhanced by the availability of HV polarized data, which is known to be especially sensitive to volume scattering, such as occurs in a forest canopy. The accuracy obtained for the estuarine forested wetland class, in particular, was unattainable before the classification code was updated to incorporate stratified sampling, since this class had previously been almost entirely suppressed by random forest's heavy emphasis on prevalent classes.

In addition to the accuracy assessment using out-of-bag samples, the overall accuracy was evaluated with points that were not used within random forests and checked against high-resolution optical data; it was found to be 94% (Table 3). For comparison, the same points were also used to re-evaluate the accuracy of the map generated in Whitcomb *et al.* [26] and the accuracy was found to be 48.6%. It should be noted that geometric errors in the original JERS-1 mosaic, which were not considered as part of this analysis, are likely to have impacted the classification accuracy we found; as in Whitcomb *et al.* [26], training and validation data were aligned to match the original JERS-1 mosaic.

Table 3. Confusion matrix for classification of wetland class, generated with verified sampling points not provided to random forests to derive the classification. See Table 1 for the key to wetland codes. Note, PML was excluded because of insufficient samples.

Observed	Predicted											Producer's Accuracy (%)
	E2EM	E2SS	E2FO	REM	L2EM	PEM	PSS	PFO	BL	DF	CF	
E2EM	83	0	0	0	0	1	0	0	0	0	0	98.81
E2SS	0	18	0	0	0	0	0	0	0	0	0	100
E2FO	0	0	4	0	0	0	0	0	0	0	0	100
REM	0	0	0	63	0	2	0	0	0	0	0	96.92
L2EM	0	0	0	0	79	3	0	0	0	0	0	96.34
PEM	2	0	0	0	5	225	10	3	1	1	0	91.09
PSS	0	0	0	0	0	0	20	0	1	0	0	95.24
PFO	0	0	0	0	0	0	0	78	0	0	0	100
BL	0	0	0	0	1	0	3	1	18	1	0	75
DF	0	0	0	0	0	0	0	0	0	12	1	92.31
CF	1	0	0	0	0	0	0	1	0	1	11	78.57
User's Accuracy (%)	96.51	100	100	100	92.94	97.4	60.61	93.98	90	80	91.67	**94**

5.3. Importance of Variables

The importance of each data layer used within the random forests classification, as evaluated with the mean decrease in the Gini impurity over all trees and for all classes, is presented in Table 4. The most important layer was latitude, followed by longitude, then elevation. The importance of positional information was likely due to the wide spatial distribution of the training samples and the differences related to this, with different wetland systems and water regimes occurring at different locations. Since wetland classes are influenced by hydrology, layers that provided information related to this (e.g., DEM-derived layers and the proximity to water layer) were of high importance. Of the SAR data layers, HV-polarization was found to be more important than HH-polarization, likely due to the role of volume scattering in differentiating different vegetation classes. Both were considered more important than the derived texture layers.

Table 4. The importance of each data layer used within the random forests classification, measured by the mean decrease in the Gini impurity index.

Layer	Importance
Latitude	368,677
Longitude	349,531
Elevation	217,587
HV	127,177
Proximity to water	112,341
HH	102,832
Slope	84,869
HH Texture	78,372
HV Texture	73,669
Date	42,499

Class-specific variable importance was found to vary significantly (taking the ratio of maximum to minimum importance for each class, the variation was found to be over 4:1 across classes). As was the case for the overall variable importance, the most important data layer for most classes was latitude, but the most important layer for the palustrine emergent wetland, palustrine scrub-shrub wetland and barren classes was longitude; and the most important layer for the estuarine emergent wetland class was elevation.

To determine which classes the different input data layers were most useful for distinguishing, we normalized the importance for each class, then identified the class with the largest importance for each data layer. The wetland class for which each data layer was most useful in distinguishing is listed in Table 5. Aggregating the SAR-derived layers, topographic layers (slope, elevation and proximity to water) and positional layers, the SAR-derived layers were most important for distinguishing the palustrine emergent wetland class; the topographic layers were most useful for distinguishing the estuarine emergent wetland class; and the positional layers were best for distinguishing the palustrine scrub-shrub wetland class.

Table 5. Wetland class each data layer was most important for distinguishing.

Layer	Class
HH	Coniferous forest
HH Texture	Coniferous forest
HV	Palustrine emergent
HV Texture	Palustrine emergent
Date	Estuarine emergent
Elevation	Estuarine Emergent
Slope	Riverine emergent
Proximity to water	Lacustrine emergent
Latitude	Palustrine scrub-shrub
Longitude	Palustrine scrub-shrub
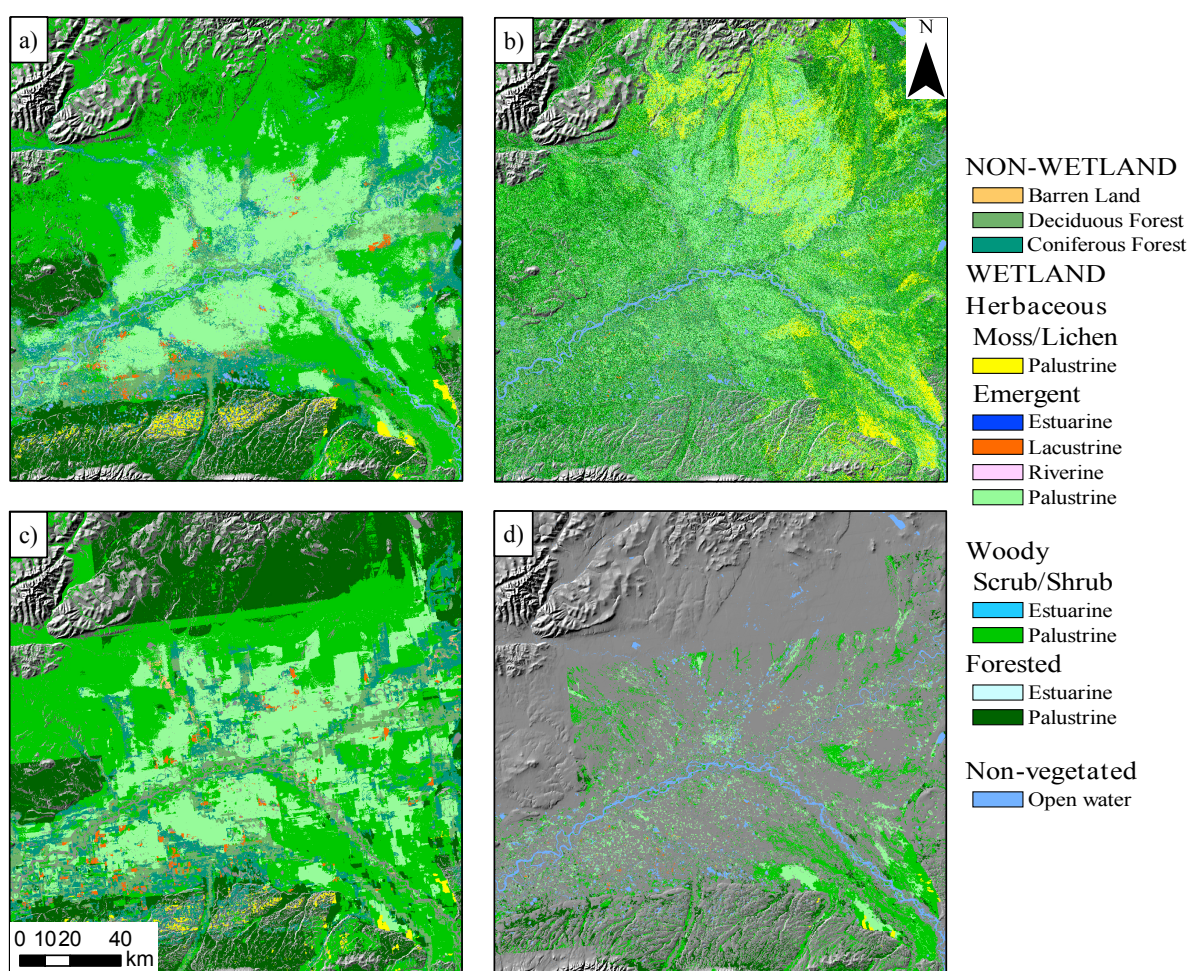
Figure 5. Subset around the Yukon Flats showing the classification for a number of layer combinations: (**a**) all available layers; (**b**) only PALSAR layers; (**c**) no PALSAR layers and no water; (**d**) ground reference layer. For the legend, see Figure 4.

The accuracies for classification runs using a subset of all available layers are shown in Table 6. Similar accuracies were observed when all layers were included and when the PALSAR layers were excluded. However, for the scenarios where PALSAR data were not included, spatial artifacts (e.g., hard boundaries) were visible in the classification results (Figure 5). This can be seen by comparing Figure 5a, for which PALSAR imagery was included in the classification, and Figure 5c, for which PALSAR

imagery was not included. The latter shows harsh blocky patterns in many locations, with particularly unrealistic boundaries between the palustrine scrub-shrub and palustrine forested wetland classes in the northern central and western parts of the image (it is noted that there is no significant agriculture or other human development in the region).

Table 6. Accuracy for classification using a subset of all available layers.

Scenario	Out of Bag Accuracy	Points Accuracy
All layers	84.5	94.0
No PALSAR (NHD Water)	88.3	95.5
No PALSAR (no Water)	87.6	94.2
Only PALSAR	46.7	73.7
Ecoregions instead of latitude and longitude	74.2	89.5
No latitude and longitude	68.2	86.5
No topography	82.0	93.2

5.4. Area of Wetlands

Using a count of pixels within each class, adjusted for classification error [55], the total area of vegetated wetlands in Alaska was 585,400 km^2 (the unadjusted area was 599,900 km^2), which is 35.9% of the total surface area of Alaska. In Table 7, the proportion of each wetland class is provided, with palustrine scrub-shrub (47%; 275,100 km^2) and palustrine emergent (39.3%; 230,200 km^2) wetlands covering the largest areas.

Table 7. Total area of each wetland class, calculated using the classification derived from all input variables and adjusted for classification accuracy. See Table 1 for the key to the codes.

Class	Total Area (km^2)	Percent of Wetlands	Percent of Alaska
E2EM	5877	1.00	0.36
E2SS	235	0.04	0.01
E2FO	19	0.00	<0.01
REM	182	0.03	0.01
L2EM	4509	0.77	0.28
PML	3093	0.53	0.19
PEM	230,165	39.32	14.10
PSS	275,095	47.00	16.86
PFO	66,171	11.30	4.05
Total	585,346		35.87

6. Discussion

6.1. Significant Improvements to Existing Mapping

Although the idea of applying random forests to a combination of L-band SAR data and ancillary layers was initially proposed in a Whitcomb *et al.* [26] study investigating the application of JERS-1 data and ancillary information for mapping wetlands in Alaska, the current study has broken new ground

through the development of an entirely new software suite that addressed a host of unresolved issues that had plagued the earlier study and thereby allowed the production of a more accurate map of vegetated wetlands in Alaska.

Firstly, the current study employed higher quality input data than that used in the 2009 study. Its ALOS PALSAR data exhibited much better geometric and radiometric accuracy than did the JERS-1 data used in [26]. The PALSAR data were also available at a higher spatial resolution than had been used in the JERS-1 mosaics of the previous study, thereby enabling us to produce a wetlands map with 2:1 better spatial resolution than our previous wetlands map. Improvements to the quality of ancillary data layers were also utilized, namely the NED DEM [39,40], which was a vast improvement over the previous elevation data available for Alaska. Improvements were also made to the training data through the inclusion of newly available quadrangles of NWI data and the use of the NLCD data for non-wetland areas, which was standardized across the state, rather than region-specific data from the Alaska Geospatial Data Clearinghouse.

The study also benefited from an overhaul of the classification method, as first outlined in Clewley *et al.* [11]. A major advance was the choice to generate and apply a single random forest model for the entire state using a stratified sample of training pixels rather than applying the classification to sixteen separate tiles using all available training pixels for each tile. This improvement compensated for a large disparity in the number of samples available for each class, which had caused sparse classes to be underrepresented in the [26] classification, while also eliminating what had been quite prominent tile boundary discontinuities in the classification. Additional enhancements over previous work included improved PALSAR swath processing that nearly eliminated what had been prominent swath edge anomalies, correction of substantial errors in the slope data layer, planar-regression filtering to reduce the effects of DEM terracing on the slope data layer, addition of a longitude data layer and addition of a coniferous uplands class. As a consequence of all of these improvements, an assessment with points verified using high-resolution optical data demonstrated a greatly improved accuracy of 94%, compared to 48.6% for the map of Whitcomb *et al.* [26].

The software developed for the current study, which exploited a newly-developed compressed file format, was many times more efficient in the handling of large datasets than that used in the [26] study. This, in combination with more direct and flexible interface processing, establishes a solid baseline from which future high-resolution wetlands maps can be quickly developed.

6.2. Wetlands Area

The total wetlands area of 585,400 km^2 (35.9%) is lower than the estimated 688,800 km^2 (42.2%) reported for vegetated wetlands by Hall *et al.* [4], but is much larger than the 410,000 km^2 (26.3%) provided by Whitcomb *et al.* [26]. However, the quality of the DEM and the commercial routine used for the slope calculation by Whitcomb *et al.* [26] in the earlier classification led to substantial areas of wetlands being masked out [58]. The revised map of vegetated wetlands derived from JERS-1 data and ancillary data [11] used an improved DEM and slope calculation method and mapped a total of 613,800 km^2 of wetlands, following correction for bias due to classification errors [55]. This revised JERS-1-derived estimate of wetlands area represents a more accurate estimate of the extent of vegetated wetlands in Alaska in the 1990s based on comparison with other studies (e.g., [8]).

Despite some correction for areal bias using the confusion matrix, there likely are errors in both areal estimates and the effects of different spatial resolutions used for the JERS-1 classification (100 m) and the PALSAR classification presented here (50 m), which need to be quantified before developing a baseline estimate of changes in wetland extent between the 1990s and 2000s. The development of techniques to better understand these errors will be the focus of future work and will enable changes in wetland type and extent within Alaska to be better understood.

6.3. Comparison with 1990s Map

The map was visually compared with both the original map of Whitcomb *et al.* [26] and the updated version of Clewley *et al.* [11], with reference to historical optical data available through GoogleEarth, primarily from the Landsat program. A number of ponds classified in the JERS-1-derived classification [11,26] had either shrunk or disappeared altogether. The shrinking of ponds has also been noted in other studies (e.g., [13,14]), with degradation of the permafrost layer leading to lake drainage identified as a possible cause.

The percentage of wetlands in each class was compared with the JERS-1-derived classification [11], with the largest changes being a 1.9% increase in the proportion of palustrine forested wetlands and a 1.8% decrease in the proportion of palustrine emergent wetlands between the 1998 (JERS-1) and 2007 (PALSAR) products. An increase in shrub abundance, as well as an increase in the extent and density of spruce forest, has been previously noted in Alaska [12].

When comparing changes between the 1990s map and the current product, it is important to separate differences in classes due to inter-annual variability and long-term trends. Given that both maps present only a snapshot at a particular time with a gap of nearly ten years, additional data are required for change analysis. For example, incorporating higher temporal resolution data from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument [59] with our wetland mapping procedures might be an option.

6.4. Importance of SAR Data

When considering the overall importance, across all classes, the positional and elevation layers were found to be more important than the PALSAR-derived layers in terms of the decrease in the Gini impurity index. By separating the importance by class, it was found the PALSAR-derived layers were most important for distinguishing the palustrine emergent class, which made up nearly 40% of the area mapped.

Despite the PALSAR-derived data layers being assigned a lower importance within the random forests classification than the topographic and position layers, the latter "static" layers provide only an approximation of where wetlands are likely to occur. It is possible to derive a map using only these static layers, but the map will contain no dynamic component and, thus, provide no way to monitor long-term wetland dynamics. The SAR data used in our classification provided information from a given time period that refined the classification. As only points where there had been no change in wetland type were used for evaluating accuracy, there was little difference in overall accuracy when all data layers were included (94%), compared to when only non-PALSAR layers were used (95.5%). If more dynamic

areas were considered as part of the accuracy assessment, we would expect to see lower accuracy for classification runs where PALSAR data were excluded.

Although we encountered problems inherent in the use of SAR data, namely variations in backscatter between strips due to variations in environmental conditions, SAR data offered a number of benefits over optical Earth observation data. For example, over 85% of the PALSAR data used for the classification was acquired during a single year (2007), most of which were from a single season (summer). Data over a much longer time period would be required to produce a cloud-free mosaic of optical data at an equivalent resolution for the entire state of Alaska.

6.5. Limitations of Approach

Both the NWI and NLCD datasets used to train the classification had errors associated with them that likely influenced the accuracy of the classification. These errors include inaccuracies in the original mapping and changes that have occurred in the 30–40 years since the NWI maps were produced. Given the large areas covered by both datasets and ongoing changes in the Alaskan landscape, identifying which areas of training data were correct or incorrect is difficult. One approach would be to select a small number of training samples and to compare them manually with aerial photography or high-resolution satellite data from around the period for which the classification is being produced (2007 in our case) to confirm that they are an accurate representation of classes for the study period. This is labor intensive, and for the current study, we could check only a small number of training points using this method. An alternative is to use a large number of training samples, of which a proportion will be incorrect, and reserve points that have been verified for validation. The impact of the incorrect training samples will depend on the classification method used. Random forests is relatively robust to outliers and noise in the training data [21], provided sufficient samples are available. Therefore, the approach adopted here was to use a large amount of training data that undoubtedly contained some errors, rather than a very small amount of well-validated training data.

A large factor determining the amount of wetlands mapped with the method proposed was the slope threshold used to mask out non-wetland areas. We followed the method used by Whitcomb *et al.* [26]; however, as was noted in Hall *et al.* [4], wetlands are also likely to occur on slopes, particularly north-facing slopes, due to the presence of permafrost. Therefore, refinement of the method used for the initial wetland/upland split is required.

Although the classification method can incorporate SAR data from multiple seasons (as in [26], where summer and winter JERS-1 data were used), only summer data were used here, as no PALSAR data from other seasons were obtainable. If PALSAR data from other seasons, or multi-season data from another sensor, became available, the new data could easily be incorporated into the random forests classification algorithm and would be expected to increase the accuracy of the classification.

6.6. Future Work

Although an initial comparison has been made between the current map produced from ALOS PALSAR and ancillary data and the 1990s map produced using JERS-1 data and ancillary data [11,26], work is ongoing to further quantify the uncertainty associated with both datasets and the implications for

detecting change. We continue to improve the quality of our wetland mapping, especially in relation to improving discrimination of wetland/non-wetland vegetation types.

The use of geographic object-based image analysis (GEOBIA) for classification has been increasing in popularity in recent years [20,60,61]. Applying the random forests-based classification described here at the object level is one area for evaluation. The use of GEOBIA is expected to become more relevant as higher resolution data become available. One particular advantage of GEOBIA is that the polygon classification output is closer to the classification produced from air photo interpretation (e.g., the NWI dataset).

In addition to developing ways to improve the accuracy of the existing classification, future work will continue the time series using the next generation of Earth observation data. The successor to ALOS, ALOS-2, was launched in May 2014 and will provide a continued time series of L-band SAR data, allowing continued monitoring of wetland areas in Alaska.

We also hope to assess the improvement in classification accuracy achievable through the incorporation of C-band SAR data, which could be expected to enhance performance for the herbaceous vegetation classes.

7. Summary

Given the importance of wetlands in Alaska and the elevated rates of climate change, there is a need for up-to-date mapping. Building on previous work with JERS-1 data by Whitcomb *et al.* [26], we evaluated the use of ALOS PALSAR L-band SAR imagery from 2007 and ancillary topographic and locational information to generate a map of vegetated wetlands in Alaska at 50-m spatial resolution. We found that topographic information (slope and elevation) and positional (latitude, longitude) information could be used to determine general locations where wetlands were likely to occur, but that refinement using Earth observation data was needed to capture wetland dynamics. The accuracy of our classification was over 94.7% when discriminating wetlands from uplands and 84.5% when discriminating among nine wetland and three upland classes. Our map showed a total of 0.59 million km^2 of vegetated wetlands in Alaska, a figure lower than previous estimates [4,8]. Work is ongoing to evaluate decadal change between the new map and the previous one based on JERS-1 imagery from the 1990s [11,26].

Acknowledgments

Author Contributions

Jane Whitcomb, Mahta Moghaddam, Kyle McDonald and Bruce Chapman developed the original classification method. Dan Clewley added a number of enhancements, applied the classification to PALSAR data and performed the analysis with input from the other authors. Bruce Chapman processed

the PALSAR data. Dan Clewley, Pete Bunting and Jane Whitcomb developed the software used for pre-processing and the classification. Dan Clewley, Jane Whitcomb and Mahta Moghaddam prepared the manuscript with contributions from the other authors.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Bridgham, S.D.; Cadillo-Quiroz, H.; Keller, J.K.; Zhuang, Q. Methane emissions from wetlands: Biogeochemical, microbial, and modeling perspectives from local to global scales. *Glob. Chang. Biol.* **2013**, *19*, 1325–1346.

2. Shindell, D.T.; Bernadette, W.P.; Faluvegi, G. Impacts of climate change on methane emissions from wetlands. *Geophys. Res. Lett.* **2004**, *31*, L21202, doi:10.1029/2004GL021009.

3. Melton, J.R.; Wania, R.; Hodson, E.L.; Poulter, B.; Ringeval, B.; Spahni, R.; Bohn, T.; Avis, C.A.; Beerling, D.J.; Chen, G.; *et al.* Present state of global wetland extent and wetland methane modelling: Conclusions from a model inter-comparison project (WETCHIMP). *Biogeosciences* **2013**, *10*, 753–788.

4. Hall, J.V.; Frayer, W.E.; Willen, B.O. *Status of Alaska Wetlands*; Technical Report; U.S. Fish and Wildlife Service: Anchorage, AK, USA, 1994.

5. Serreze, M.C.; Barry, R.G. Processes and impacts of Arctic amplification: A research synthesis. *Glob. Planet. Chang.* **2011**, *77*, 85–96.

6. Serreze, M.; Walsh, J.; Chapin, F.S.I.; Osterkamp, T.; Dyurgerov, M.; Romanovsky, V.; Oechel, W.; Morison, J.; Zhang, T.; Barry, R.; *et al.* Observational evidence of recent change in the northern high-latitude environment. *Clim. Chang.* **2000**, *46*, 159–207.

7. Wendler, G.; Chen, L.; Moore, B. The first decade of the new century: A cooling trend for most of Alaska. *Open Atmos. Sci. J.* **2012**, *6*, 111–116.

8. Bridgham, S.D.; Megonigal, J.P.; Keller, J.K.; Bliss, N.B.; Trettin, C. The carbon balance of North American wetlands. *Wetlands* **2006**, *26*, 889–916.

9. Miller, C.E.; Dinardo, S.J. CARVE: The Carbon in arctic reservoirs vulnerability experiment. In Proceedings of the 2012 IEEE Aerospace Conference, Big Sky, MT, USA, 3–10 March 2012; pp. 1–17.

10. Cowardin, L.M.; Carter, V.; Golet, F.C.; LaRoe, E.T. *Classification of Wetlands and Deepwater Habitats of the United States*; U.S. Fish and Wildlife Service, Office of Biological Services: Washington, DC, USA, 1979.

11. Clewley, D.; Whitcomb, J.; Moghaddam, M.; McDonald, K. Mapping the state and dynamics of boreal wetlands using synthetic aperture radar. In *Advances in Wetlands Mapping*; Taylor Francis: Boca Raton, FL, USA, 2015; Chapter 17.

12. Sturm, M.; Racine, C.; Tape, K. Climate change: Increasing shrub abundance in the Arctic. *Nature* **2001**, *411*, 546–547.

13. Riordan, B.; Verbyla, D.; McGuire, A.D. Shrinking ponds in subarctic Alaska based on 1950–2002 remotely sensed images. *J. Geophys. Res.* **2006**, *111*, G04002.

14. Yoshikawa, K.; Hinzman, L.D. Shrinking thermokarst ponds and groundwater dynamics in discontinuous permafrost near council, Alaska. *Permafr. Periglac. Process.* **2003**, *14*, 151–160.

15. Klein, E.; Berg, E.E.; Dial, R. Wetland drying and succession across the Kenai Peninsula Lowlands, south-central Alaska. *Can. J. For. Res.* **2005**, *35*, 1931–1941.

16. Rosenqvist, A.; Finlayson, C.M.; Lowry, J.; Taylor, D. The potential of long-wavelength satellite-borne radar to support implementation of the Ramsar Wetlands Convention. *Aquat. Conserv. Mar. Freshw. Ecosyst.* **2007**, *17*, 229–244.

17. Stolt, M.; Baker, J. Evaluation of national wetland inventory maps to inventory wetlands in the southern Blue Ridge of Virginia. *Wetlands* **1995**, *15*, 346–353.

18. Kudray, G.M.; Gale, M.R. Evaluation of National Wetland Inventory maps in a heavily forested region in the upper Great Lakes. *Wetlands* **2000**, *20*, 581–587.

19. Pierce, L.; Ulaby, F.; Sarabandi, L.; Dobson, M. Knowledge-Based classification of polarimetric SAR images. *IEEE Trans. Geosci. Remote Sens.* **1994**, *32*, 1081–1086.

20. Lucas, R.M.; Medcalf, K.; Brown, A.; Bunting, P.J.; Breyer, J.; Clewley, D.; Keyworth, S.; Blackmore, P. Updating the Phase 1 habitat map of Wales, UK, using satellite sensor data. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 81–102.

21. Leo, B. Random forests. *Mach. Learn.* **2001**, *45*, 5–32.

22. Fernandez-Delgado, M.; Cernadas, E.; Barro, S.; Amorim, D. Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.* **2014**, *15*, 3133–3181.

23. Mellor, A.; Haywood, A.; Stone, C.; Jones, S. The performance of random forests in an operational setting for large area sclerophyll forest classification. *Remote Sens.* **2013**, *5*, 2838–2856.

24. Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. Random forests for land cover classification. *Pattern Recognit. Lett.* **2006**, *27*, 294–300.

25. Pal, M. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* **2005**, *26*, 217–222.

26. Whitcomb, J.; Moghaddam, M.; McDonald, K.; Kellndorfer, J.; Podest, E. Mapping vegetated wetlands of Alaska using L-band radar satellite imagery. *Can. J. Remote Sens.* **2009**, *35*, 54–72.

27. Whitcomb, J.; Moghaddam, M.; McDonald, K.; Podest, E.; Chapman, B. Decadal change in northern wetlands based on differential analysis of JERS and PALSAR data. In Proceedings of the 2009 IEEE International on Geoscience and Remote Sensing Symposium (IGARSS 2009), Cape Town, South Africa, 12–17 July 2009; Volume 3, pp. 951–954.

28. Whitcomb, J.; Moghaddam, M.; McDonald, K.; Podest, E. Mapping Canadian wetlands using L-band radar satellite imagery. In Proceedings of the 2009 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2009, Cape Town, South Africa, 12–17 July 2009; Volume 2, pp. 1032–1035.

29. Shulski, M.; Wendler, G. *The Climate of Alaska*; University of Alaska Press: Fairbanks, AK, USA, 2007.

30. Viereck, L.; Dyrness, C.; Batten, A.; Wenzlick, K. *The Alaska Vegetation Classification, General Technical Report*; Technical Report PNW-GTR-286; U.S. Department of Agriculture: Portland, OR, USA, 1992.

31. Rosenqvist, A.; Shimada, M.; Chapman, B.; McDonald, K.; de Grandi, G.; Jonsson, H.; Williams, C.; Rauste, Y.; Nilsson, M.; Sango, D.; *et al.* An overview of the JERS-1 SAR Global Boreal Forest Mapping (GBFM) project. In Proceedings of the 2004 IEEE International Geoscience and Remote Sensing Symposium (IGARSS '04), Anchorage, AK, USA, 20–24 September 2004; Volume 2, pp. 1033–1036.

32. Shimada, M.; Tadono, T.; Rosenqvist, A. Advanced Land Observing Satellite (ALOS) and monitoring global environmental change. *Proc. IEEE* **2010**, *98*, 780–799.

33. Rosenqvist, A.; Shimada, M.; Lucas, R.M.; Chapman, B.; Paillou, P.; Hess, L.; Lowry, J. The Kyoto & carbon initiative—A brief summary. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2010**, *3*, 551–553.

34. Rosenqvist, A.; Shimada, M.; Chapman, B.; McDonald, K.; de Grandi, G.; Jonsson, H.; Williams, C.; Rauste, Y.; Nilsson, M.; Sango, D.; *et al.* An overview of the JERS-1 SAR Global Boreal Forest Mapping (GBFM) project. In Proceedings of the 2004 IEEE International Geoscience and Remote Sensing Symposium (IGARSS'04), Anchorage, AK, USA, 20–24 September 2004.

35. Wegmüller, U.; Werner, C.; Strozzi, T. SAR interferometric and differential interferometric processing chain. In Proceedings of the 1998 IEEE International, Geoscience and Remote Sensing Symposium Proceedings (IGARSS'98), Seattle, WA, USA, 6–10 July 1998; Volume 2, pp. 1106–1108.

36. Wegmüller, U. Automated terrain corrected SAR geocoding. In Proceedings of the IEEE 1999 International Geoscience and Remote Sensing Symposium (IGARSS'99), Hamburg, Germany, 28 June–2 July 1999; Volume 3, pp. 1712–1714.

37. Bunting, P.; Clewley, D.; Lucas, R.M.; Gillingham, S. The Remote Sensing and GIS Software Library (RSGISLib). *Comput. Geosci.* **2014**, *62*, 216–226.

38. Wettstein, W.; Schmid, B. Conservation of arthropod diversity in montane wetlands: Effect of altitude, habitat quality and habitat fragmentation on butterflies and grasshoppers. *J. Appl. Ecol.* **1999**, *36*, 363–373.

39. Gesch, D.; Oimoen, M.; Greenlee, S.; Nelson, C.; Steuck, M.; Tyler, D. The mational elevation dataset. *Photogramm. Eng. Remote Sens.* **2002**, *68*, 5–11.

40. Gesch, D. The national elevation dataset. In *Digital Elevation Model Technologies and Applications: The DEM Users Manual*, 2nd ed.; American Society for Photogrammetry and Remote Sensing: Bethesda, MD, USA, 2007; pp. 99–118.

41. Podest, E.; McDonald, K.; Bohn, T.; Lettenmaier, D.P. Mapping boreal wetlands using spaceborne syntheric aperture radar. In Proceedings of the Proceedings of the American Geophysical Union Fall Meeting, San Francisco, CA, USA, 11–15 December 2006.

42. Carroll, M.; Townshend, J.; DiMiceli, C.; Noojipady, P.; Sohlberg, R. *MOD44W MODIS Water Mask*; Technical Report; University of Maryland: College Park, MD, USA, 2009.

43. U.S. Geological Survey. *National Hydrography Dataset*; Technical Report; U.S. Geological Survey: Juneau, AK, USA, 2015.

44. Gallant, A. *Ecoregions of Alaska*; Technical Report; USGS: Denver, CO, USA, 1996.

45. Commission for Environmental Cooperation (CEC). *Ecological Regions of North America*; Technical Report; Commission for Environmental Cooperation: Montreal, QC, Canada, 2009.

46. Tiner, R.W. *Wetland Indicators: A Guide to Wetland Identification, delineation, classification, and mapping*; Lewis Publishers: Boca Raton, FL, USA, 1999.

47. Homer, C.; Dewitz, J.; Fry, J.; Coan, M.; Hossain, N.; Larson, C.; Herold, N.; McKerrow, A.; van Driel, J.N.; Wickham, J.; *et al.* Completion of the 2001 National Land Cover Database for the conterminous United States. *Photogramm. Eng. Remote Sens.* **2007**, *73*, 337–341.

48. Selkowitz, D.J.; Stehman, S.V. Thematic accuracy of the National Land Cover Database (NLCD) 2001 land cover for Alaska. *Remote Sens. Environ.* **2011**, *115*, 1401–1407.

49. Bunting, P.; Gillingham, S. The KEA image file format. *Comput. Geosci.* **2013**, *57*, 54–58.

50. Breiman, L.; Friedman, J.; Stone, C.J.; Olshen, R.A. *Classification and Regression Trees*; Chapman and Hall: New York, NY, USA, 1993.

51. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140.

52. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.

53. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2014.

54. Bivand, R.; Keitt, T.; Rowlingson, B. *rgdal: Bindings for the Geospatial Data Abstraction Library*; R Package Version 0.8-11; 2013.

55. Tenenbein, A. A double sampling scheme for estimating from misclassified multinomial data with applications to sampling inspection. *Technometrics* **1972**, *14*, 187–202.

56. Walsh, T.A.; Burk, T.E. Calibration of satellite classifications of land area. *Remote Sens. Environ.* **1993**, *46*, 281–290.

57. Carreiras, J.; Pereira, J.; Campagnolo, M.L.; Shimabukuro, Y. Assessing the extent of agriculture/pasture and secondary succession forest in the Brazilian Legal Amazon using SPOT VEGETATION data. *Remote Sens. Environ.* **2006**, *101*, 283–298.

58. Whitcomb, J.; Moghaddam, M.; McDonald, K.; Podest, E.; Chapman, B. Mapping and change detection for boreal wetlands of North America based on JERS and PALSAR data. In Proceedings of the 2010 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Honolulu, HI, USA, 25–30 July 2010; pp. 1371–1373.

59. Potter, C.; Li, S.; Crabtree, R. Changes in Alaskan tundra ecosystems estimated from MODIS greenness trends, 2000 to 2010. *J. Geophys. Remote Sens.* **2013**, doi: 10.4172/2169-0049.1000107.

60. Benz, U.; Hofmann, P.; Willhauck, G.; Lingenfelder, I. Multi-Resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS J. Photogramm. Remote Sens.* **2004**, *58*, 239–258.

61. Clewley, D.; Bunting, P.J.; Shepherd, J.; Gillingham, S.; Flood, N.; Dymond, J.; Lucas, R.M.; Armston, J.; Moghaddam, M. A Python-based open source system for Geographic Object-Based Image Analysis (GEOBIA) utilizing raster attribute tables. *Remote Sens.* **2014**, *6*, 6111–6135.