# Detecting and Downscaling Wet Areas on Boreal Landscapes

Yasir H. Kaheil and Irena F. Creed

***Abstract*—This letter presents an approach to classify wet areas from European Remote Sensing 2 (ERS-2) synthetic aperture radar (SAR)-, Landsat Thematic Mapper (TM)-, and Light Detection and Ranging (LiDAR)-derived terrain data and downscale the result from the coarse resolution of satellite images to finer resolutions needed for land managers. Using discrete wavelet transform (DWT) and support vector machines (SVM), the algorithm finds multiple relationships between the radar, optical, and terrain data and wet areas at different spatial scales. Decomposing and reconstructing processes are performed using a 2-D DWT (2D-DWT) and inverse 2D-DWT respectively. The underlying relationships between radar, optical, and terrain data and wet areas are learned by training an SVM at the coarse resolution of the wet-area map. The SVM is then applied on the predictors at a finer resolution to produce wet-area detailing images, which are needed to reconstruct a finer resolution wet-area map. The algorithm is applied to a boreal landscape in northern Alberta, Canada, characterized by many wet-area features including ephemeral and permanent streams and wetlands.**

***Index Terms*—Data fusion, downscaling, hydrologically sensitive areas, hydrology, learning machines, remote sensing, synthetic aperture imaging, wavelet transforms, wet areas, wetland.**

## I. Introduction

WET AREAS include landscape areas that are permanently or transiently wet [1]. These hydrologic features play a fundamental role in the water cycle and are important in regulating the movement of nutrients and biota within the landscape [2]. Detecting the occurrence of these areas at a fine spatial scale is critically important for land managers and decision makers when implementing land management practices [3].

Distributed hydrological models have been used to detect the formation of wet areas (e.g., regional hydro-ecologic simulation system (RHESSys) [4]). Creed *et al.* [5] applied RHESSys to a boreal landscape in northern Alberta, Canada. They found that while they were able to model realistic water budgets and the timing of peak and base flows, they were not able to model realistic wet areas distributed throughout the boreal watershed. That is, they were able to develop a model that could accurately predict runoff but not the wet areas that contribute to runoff due largely to the lack of data in this remote boreal watershed.

Manuscript received May 30, 2008; revised August 27, 2008 and September 29, 2008. First published January 6, 2009; current version published April 17, 2009. This work was supported by the Network of Centres of Excellence on Sustainable Forest Management.

Y. H. Kaheil is with the International Research Institute for Climate and Society, Lamont–Doherty Earth Observatory, Columbia University, Palisades, NY 10964 USA (e-mail: ykaheil@iri.columbia.edu).

I. F. Creed is with the Department of Biology, The University of Western Ontario, London, ON N6A 5B7, Canada (e-mail: icreed@uwo.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2008.2010001

In contrast, increasing availability of remote sensing products offers a new opportunity for wet-area mapping in remote landscapes. There are numerous multivariate statistical clustering and classification methods that can be used with remote sensing imagery to predict the occurrence of wet areas [6]. Clustering methods (i.e., unsupervised classification) do not require a map of wet features as they cluster the predictors (e.g., maps of reflectance, topography, and soils that regulate the occurrence of wet features) into a set of classes that have no *a priori* meaning. Classification methods (i.e., supervised classification) require a map of wet features to rotate the predictor space in a way to match it. Such methods include support vector machines (SVM). Maps of wet areas exist for many boreal landscapes although at coarse spatial resolutions ($>= 25$ m).

The selection of potential predictors of wet areas is nontrivial. Several remote sensing products have been correlated with wet soils. For example, early remote sensing studies used optical reflectance to predict wet soils, relying on the fact that soils get darker when they are wetter. These analyses were often hindered by interference of the optical signature by clouds, atmosphere, and other factors inherent in optical remote sensing [7]. More recent studies have used C-band radar data with synthetic aperture processing. Synthetic aperture radar (SAR) images contain backscatter information, where higher backscatter coefficients correspond to wetter soils [8]. However, the SAR backscatter coefficient is highly influenced by surface roughness, vegetation structure, and water content [9], [10].

The use of SAR images as a predictor of wet areas can be enhanced using information on forest cover and topography [11]. Multispectral images can be used to calculate vegetation indices and thereby provide a good source of information on vegetation [12], albeit highly influenced by the water content of vegetation. Digital elevation models obtained by Light Detection and Ranging (LiDAR) technology at a fine spatial resolution provide information on topographic features (e.g., flats and/or depressions) that have been correlated with the position of wet areas [13] or features that have been associated with hydrologic processes leading to the formation of wet areas (e.g., topographic index and distance to stream) [14].

Fusion of radar, optical, and topographic data represents a promising approach for improving mapping of wet areas. For example, Li and Chen [15] proposed a rule-based methodology to detect and classify wet areas across Canada, employing a suite of data sets similar to what this letter proposes. Due to their model structure with a separate classification component for each source of input data, the combined decision is bound to be affected by errors from each classification component, deeming this methodology vulnerable to errors associated with the input data. Töyrä and Pietroniro [16], on the other hand, proposed a more comprehensive geomatics-based methodology

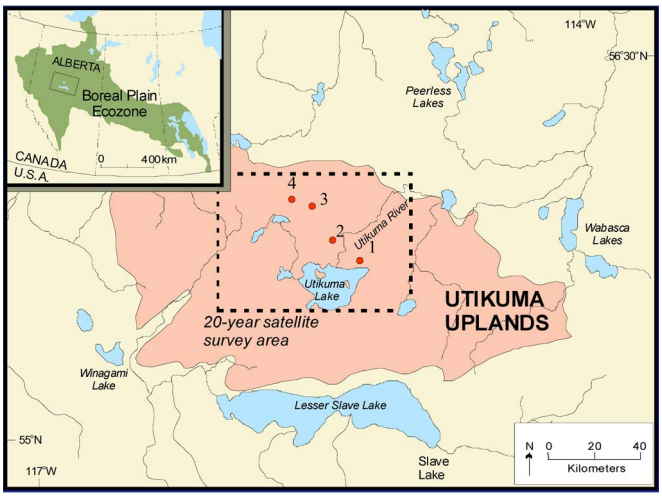


Fig. 1. Map of the Utikuma Lake drainage basin. The modeling training site (1) and the three additional sites (2, 3, and 4) where the synthetic model was applied are depicted as small circles on this map.

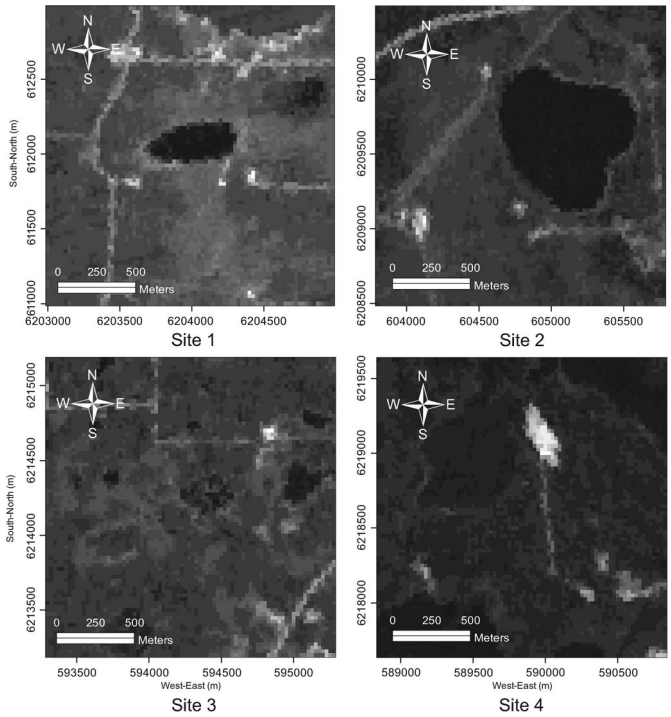


Fig. 2. Landsat TM red-band images of the four selected sites. Coordinates appear on the axes in meters; projection: UTM Zone 12N.

that was focused on spatiotemporal monitoring of wet areas in northern Canada. Similar to this study, the methodology accommodated inputs from multiple data sources; however, it did not address issues of scale discrepancy. Also, the methodology cannot be automated, and therefore, it is difficult to be viewed as applied research that can ultimately be used by resource managers. The radar, optical, and topographic data that may be used as predictors of wet areas are often obtained at different spatial resolutions. Consequently, a multiresolutional framework is needed to carefully map these predictors onto the wet-area map.

In a recent study [17], an algorithm was developed to address the problem of predictors with different spatial resolutions, given a predictant at coarse spatial resolution. This algorithm embeds 2-D discrete wavelet transform (2D-DWT) operations to facilitate multiresolutional analysis with multiple SVMs to map predictors to the wet-area map. SVMs are designed to accommodate a relatively high dimensional predictor space [18]. Furthermore, the data used as predictors contain substantial errors. This fact provides further support for the use of SVMs, given their capability to avoid overfitting and thereby circumvent the errors embedded in the predictors [18].

In this letter, we modify the algorithm described in [17] to handle classification problems. We combine European Remote Sensing 2 (ERS-2) SAR images (12.5 m), Landsat Thematic Mapper (TM) multispectral reflectance images (25 m), and LiDAR data (6.25 m) as predictors to map wet areas at a fine spatial resolution (6.25 m) on a boreal landscape. A map of wet areas (25 m) generated by the Alberta Provincial Government using aerial photography is used as a wet-area map training and testing the synthetic model.

## II. Study Area

The boreal landscape located in Alberta, Canada, is characterized by simple canopy structure on gentle topography, with hydrologic features that range from small isolated wet areas to large complex wetland systems [19]. A study area was selected from within the Utikuma River drainage basin which lies in the western portion of the Boreal Plain ecozone on an upland plateau (Fig. 1). The climate is continental with cold long winters and short cool summers. Average annual temperature and precipitation are 1.7 °C and 503 mm, respectively. The surficial geology consists of deep (40–200-m) glacial deposits, including moraines, glaciofluvial outwash, and glaciolacustrine clay plains. The topography is flat to gently undulating, where flat areas are covered by extensive wetlands. Peat-forming wetlands (i.e., fens and bogs) are dominated by black spruce and tamarack, whereas nonpeat-forming wetlands (i.e., swamps and marshes) are dominated by willow, birch, alder, sedges, and grasses. Uplands are dominated by trembling aspen with occasional stands of white spruce and pine. Within the study area, the synthetic model was trained and tested, given separate subsets from four different sites that collectively represent the range of surficial geological deposits within the drainage basin (Fig. 1). Composite Landsat TM images of these sites projected on UTM 12N are shown in Fig. 2. The reader is referred to [19] for further information on the study area.

## III. Theory

A flowchart of the algorithm used in this letter is shown in Fig. 3. The reader is hence advised to follow Fig. 3 closely throughout this section. The algorithm uses a 2D-DWT to decompose and an inverse 2D-DWT to reconstruct wet areas across scales and a classification-based SVM to build relationships between predictors and wet areas at a coarse scale that can be applied at finer scales.

The algorithm is a modification of the one presented by Kaheil *et al.* [17] to handle classification rather than regression problems. In the modified algorithm, to downscale a wet-area map (for example, between 25 and 12.5 m), a 2D-DWT uses a 12.5-m predictor image to generate a 25-m datum image and three 25-m detailing images that are then used to generate a 12.5-m wet-area map (step 1, Fig. 3). A datum image (low–low, LL) is the result of applying a low-pass filter on both rows and

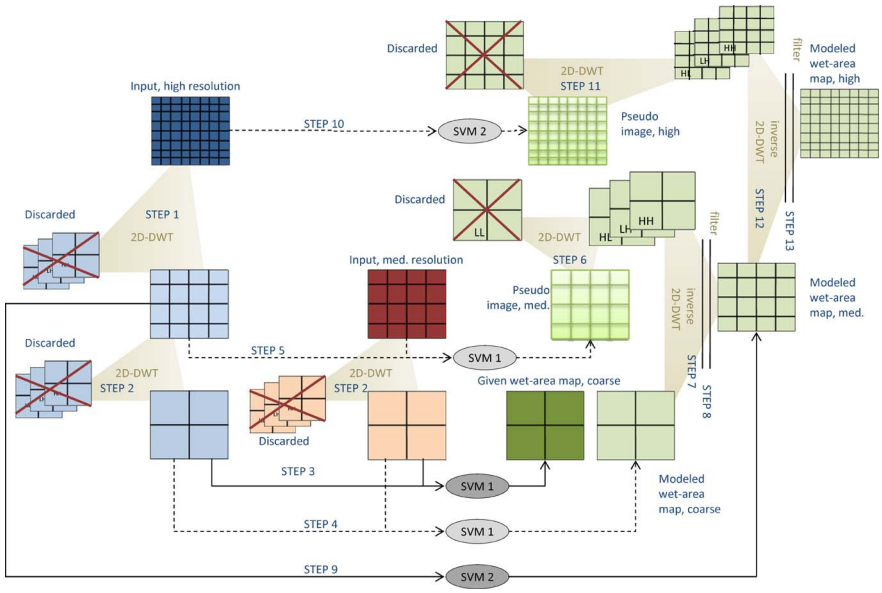


Fig. 3. Flowchart of the algorithm used in this letter. Dark colors represent "observed" variables used in the algorithm; different colors represent different variables (shades of blue: input 1; shades of red: input 2; shades of green: response or output); labels demonstrate steps/processes used to arrive at the results at different scales.

columns of the original image. A detailing image is any one of three combinations where a high-pass filter was applied on rows (high–low, HL), columns (low–high, LH), or both (high–high, HH) on an original image. Similarly, 2D-DWT is applied on other inputs (step 2, Fig. 3).

An SVM relationship (SVM 1, Fig. 3) is built at the resolution of the given wet-area map by training on a subset of pixels and then applied to obtain a modeled wet-area map (steps 3–4, Fig. 3).

In order to downscale the wet-area map, an inverse 2D-DWT is implemented. The inverse 2D-DWT takes an existing 25-m datum image and three 25-m detailing images to construct a 12.5-m wet-area map. The three detailing images are obtained by a two-step procedure. In the first step, an SVM built at a scale of 25 m using the available predictors at 25 m and the wet-area map available at 25 m (SVM 1, Fig. 3) is applied at 12.5 m using the same predictors, to create a pseudoimage of wet areas (step 5, Fig. 3) to obtain detailing images required to reconstruct a wet-area map from 25 to 12.5 m (step 6, Fig. 3). In this step, one assumes that the SVM relationship at the 25-m scale is applicable at the 12.5-m scale so that the pseudoimage of wet areas can be modeled at 12.5 m. In the second step, the pseudoimage of wet areas at 12.5 m is decomposed, resulting in a datum image of wet areas at 25 m which is discarded, and three detailing images at 25 m which are used (step 6, Fig. 3). With the modeled wet-area map at the coarser resolution and the modeled detailing images, the wet-area map can be constructed at the next finer resolution using an inverse 2D-DWT (step 7, Fig. 3). A filtering step follows each reconstruction step in order to coerce pixel values into categorical values and hence produce classes with distinct edges (step 8, Fig. 3). It is assumed that a pixel with multiple-class membership can be coerced to its closest class.

Once the wet-area map at a finer resolution is generated, the algorithm further downscales the wet-area map using those predictors that still contain information at even finer spatial resolutions (steps 9–13, Fig. 3).

TABLE I
LIST OF EACH PREDICTOR, ITS RESOLUTION, AND ITS SOURCE

| Predictor | Source | Resolution (m) |
|---|---|---|
| DEM | LiDAR | 10 (resampled to 6.25) |
| Topographic index | DEM | 10 (resampled to 6.25) |
| Distance to stream | DEM | 10 (resampled to 6.25) |
| TM bands 1, 2, 3, 4, 5, 7 | LANDSAT TM | 25 |
| SAR | ERS-2 | 25 |

## IV. APPLICATION

Our synthetic model was applied using the predictors that appear in Table I to predict wet areas based on a map generated by the Alberta Provincial Government at a spatial resolution of 25 m and then downscaled to 6.25 m. A majority of the data were acquired during the summer period in the mid-1990s (SAR, TM, and Alberta vegetation index (AVI) data). The LiDAR data were later acquired in 2002. The LiDAR data were acquired during a dry period to ensure that we covered as much of the landscape as possible (i.e., to avoid problems of new land areas being revealed with no LiDAR coverage as wet areas contracted). These data were also acquired in a landscape that is relatively undisturbed. Therefore, the LiDAR data are considered "stable" and representative of the mid-1990s. The wet-area map was extracted from the AVI data. The focus of the study was on four sites within the Utikuma drainage basin: Site 1 was used to train the SVM model, and sites 2, 3 and 4 were used to assess the performance of the SVM across a range of hydrogeological conditions. Each site covered an area of 40 000 $m^2$ or 80 $\times$ 80 pixels at 25-m resolution (Fig. 1).

### *A. Generation of Wet Area Maps at 25-, 12.5-, and 6.25-m Resolutions*

An SVM was trained using predictor data and the wet-area map at 25-m resolution (Table I). The basis matrix used for the 2D-DWT upscaling was arithmetic averaging (i.e., a Haar two-coefficient diagonal matrix with a value of 0.5) in so-doing,

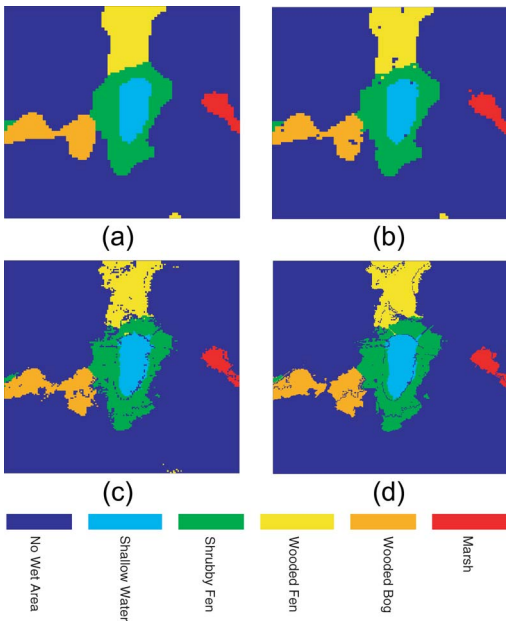


Fig. 4. Results of modified algorithm used to downscale wet areas at site 1 (Fig. 1). (a) True image at 25 m (given the provincial map of wet areas). (b) Modeled image at 25 m. (c) Downscaled modeled image at 12.5 m. (d) Downscaled modeled image at 6.25 m.

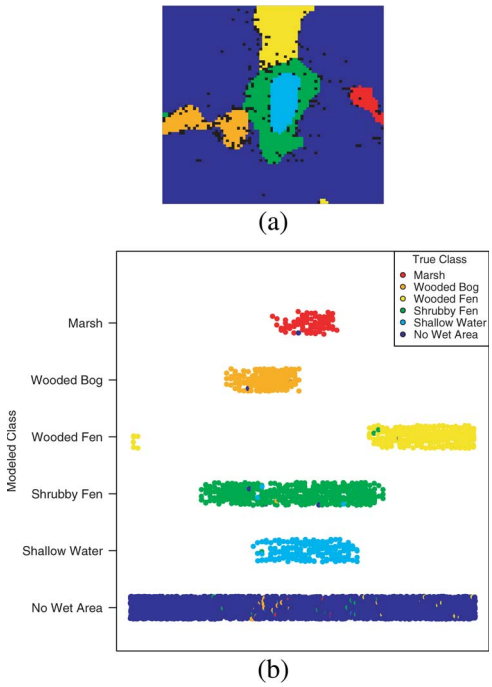


Fig. 5. Accuracy assessment for site 1 at 25-m resolution. (a) Map of misclassified pixels (black) overlaying the given provincial map. (b) Jitter plot of different wet-area classes. The point color indicates the true class (which the point should be classified as), while the $y$-axis shows the modeled class. Jittering allows better visualization and does not imply any lack of performance. The overall misclassification error was 0.97%.

it was assumed that the arithmetic averages produced images with the same units.

A second SVM was trained using LiDAR-derived topographic data only at finer resolutions ($<$ 25 m) using training sets of their upscaled datum images at 25 m versus corresponding training set of wet-area map at 25 m. This SVM was then applied on the topographic data at 12.5-m resolution to obtain a pseudoversion of the wet-area map at 12.5 m, $\mathrm{PO}_{12.5}$. $\mathrm{PO}_{12.5}$ was then decomposed, and its datum component was overwritten with $\mathrm{O}_{25}$ to collectively reconstruct the wet-area map at 12.5 m, or $\mathrm{O}_{12.5}$.

Assuming that $\mathrm{O}_{12.5}$ was the true image, $\mathrm{O}_{6.25}$ was reconstructed in a similar process, using only the predictors available at $<$ 12.5 m. The results show that as more information was captured from the topographic thematic layers, the class edges became smoother and better defined in terms of location and extent (Fig. 4).

### B. Test of SVM to Predict Wet Areas

The performance of the SVM (prediction alone) was assessed by the proportion of misclassified pixels (in percent) of the wet-area map at 25 m (the scale at which the existing wet-area map was created). Accuracy assessment was based on pixels that were not used in model development.

For site 1, Fig. 5 shows a spatial display of the misclassified pixels overlaying the original map [Fig. 5(a)] and as a jitter plot [Fig. 5(b)]. The jitter plot indicates the misclassification overlap among classes. Pixels that are located near or at class borders have higher chances to be misclassified [Fig. 5(a)]. Given that "no wet area" has the highest number of near-border pixels, we expect most misclassified pixels to originate from it. Table II shows the classification matrix for site 1. The misclassification error was very low ($< 2\%$) for all six classes.

TABLE II
DISTRIBUTION OF CORRECTLY VERSUS INCORRECTLY CLASSIFIED PIXELS AMONG THE DIFFERENT CATEGORIES OF WET AREAS AT SITE 1

| | | Classified | | | | | |
|---|---|---|---|---|---|---|---|
| | | No wet area | Shallow Water | Shrubby Fen | Wooded Fen | Wooded Bog | Marsh |
| Observed | No wet area | 4898 | 4 | 16 | 8 | 12 | 3 |
| | Shallow Water | 0 | 208 | 2 | 0 | 0 | 0 |
| | Shrubby Fen | 5 | 2 | 520 | 1 | 0 | 0 |
| | Wooded Fen | 1 | 0 | 2 | 359 | 0 | 0 |
| | Wooded Bog | 4 | 0 | 1 | 0 | 263 | 0 |
| | Marsh | 1 | 0 | 0 | 0 | 0 | 90 |
| | *Misclassified (%)* | *0.870* | *0.952* | *1.515* | *0.829* | *1.866* | *1.099* |

Misclassification among classes could be due to the following several reasons, sorted from most to least plausible: 1) uniform sampling that reduces the chances to sample points from classes represented by a small number of pixels; 2) the quality of the wet-area map; 3) grid misalignment due to poor georeferencing of the SAR, TM, LiDAR, and wet-area maps; and/or 4) the learning machine's incapability of modeling the wet-area map given the selected predictors.

The performance of the synthetic model (prediction and downscaling) was evaluated by first resampling the 25-m provincial map at 100 m and then downscaling it to 25 m using the model, while concealing the original 25-m image for corroboration. Fig. 6(a) shows the resampled image at 100 m used as a starting point for the algorithm. Fig. 6(b) shows the model results at 25 m, and Fig. 6(c) shows the locations of the points that were misclassified. The model successfully downscaled the 100-m image to 25 m and matched the concealed true image at 25 m with 4% misclassification error.

The performance of the synthetic model was also assessed on three other sites representing different surficial geological

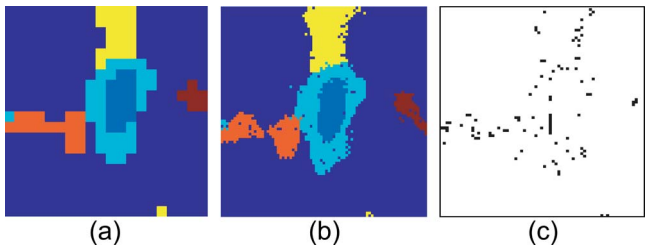


Fig. 6. Testing of synthetic model at site 1. (a) Resampled true image at 100 m. (b) Downscaled modeled image from 100 to 25 m—compare with Fig. 4(a) for cross-validation. (c) Misclassified points (modeled versus true images at 25-m resolution).

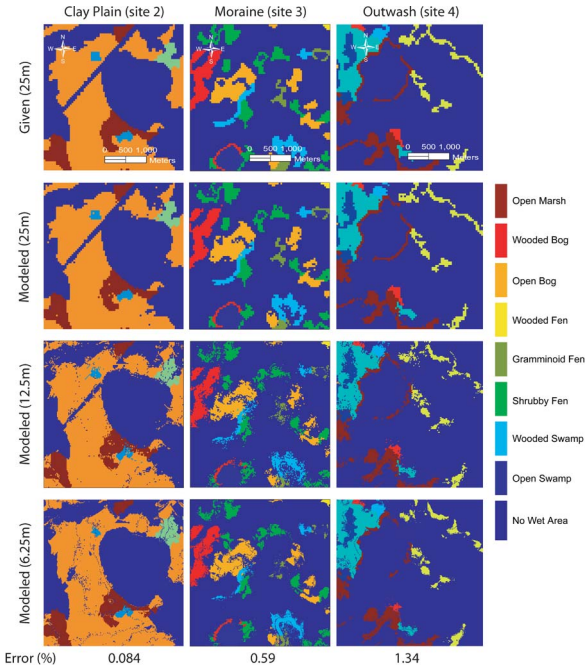


Fig. 7. Application of synthetic model to test sites with different surficial geological deposits. Error (in percent) refers to the misclassification error (in percent) at 25-m resolution (i.e., between the first and second rows).

deposits. In all cases, the model reproduced the true image at 25 m with a small misclassification error, and the model captured the hydrologic features at a finer resolution (Fig. 7).

## V. Conclusion

This letter aimed at predicting wet areas and then mapping them at finer resolutions required for land managers for a surprise-free experience when implementing land management practices. A synthetic model that combined 2D-DWT and SVM was applied on a boreal landscape characterized by simple canopy structure on gentle topography, with wet areas that range from small isolated wetlands to large complex wetland systems. The synthetic model used different predictors, including SAR and TM images and LiDAR digital terrain data, to reproduce a provincial government map of wet areas (25 m) and then downscale it from a spatial resolution of 25–6.25 m. The synthetic model produced promising results, with misclassification errors in all cases being less than 2%. Future work will evaluate the robustness of this model on other landscapes characterized by more complex canopies on variable topography.

## Acknowledgment

The authors would like to thank the six reviewers who provided comments that substantially improved this letter.

## References

[1] R. B. Clark, G. Z. Sass, and I. F. Creed, "Mapping hydrologically sensitive areas on the Boreal Plain: A multi-temporal analysis of ERS synthetic aperture radar data," *Int. J. Remote Sens.*, 2008. To be published.

[2] K. J. Devito, I. F. Creed, R. J. Rothwell, and E. E. Prepas, "Landscape controls on the loading of phosphorous to boreal lakes following timber harvest: A physical basis for adaptive riparian forest buffer strip strategies," *Can. J. Fish. Aquat. Sci.*, vol. 57, pp. 1977–1984, 2000.

[3] I. F. Creed, G. Z. Sass, M. B. Wolniewicz, and K. J. Devito, "Incorporating hydrologic dynamics into buffer strip design on the sub-humid Boreal Plain of Alberta," *For. Ecol. Manag.*, vol. 256, no. 11, pp. 1984–1994, Nov. 2008.

[4] C. L. Tague and L. E. Band, "RHESSys: Regional hydro-ecologic simulation system—An object-oriented approach to spatially distributed modeling of carbon, water, and nutrient cycling," *Earth Interact.*, vol. 8, no. 19, pp. 1–42, Dec. 2004.

[5] I. F. Creed, C. L. Tague, R. B. Clark, M. B. Wolniewicz, G. Zsigovics, and C. C. Krezek, "Modeling hydrologic processes in Boreal watersheds: The proof is in the pattern," *EOS Trans. AGU*, vol. 83, no. 47, 2002. Fall Meet. Suppl., Abstract H61B-0775.

[6] J. Lattin, J. D. Carroll, and P. E. Green, *Analyzing Multivariate Data*. Pacific Grove, CA: Duxbury, 2003.

[7] Y. H. Kerr, "Soil moisture from space: Where are we?" *Hydrogeol. J.*, vol. 15, no. 1, pp. 117–120, Feb. 2007.

[8] F. T. Ulaby, P. P. Batlivala, and M. C. Dobson, "Microwave backscatter dependence on surface roughness, soil moisture, and soil texture: Part I—Bare soil," *IEEE Trans. Geosci. Electron.*, vol. GE-16, no. 4, pp. 286–295, Oct. 1978.

[9] P. C. Dubois, J. J. van Zyl, and E. T. Engman, "Measuring soil moisture with imaging radars," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 4, pp. 915–926, Jul. 1995.

[10] J. C. Shi, J. Wang, A. Y. Hsu, P. E. O'Neill, and E. T. Engman, "Estimation of bare surface soil moisture and surface roughness parameter using L-band SAR image data," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 5, pp. 1254–1266, Sep. 1997.

[11] R. M. Lucas, A. C. Lee, and M. L. Williams, "Enhanced simulation of radar backscatter from forests using LiDAR and optical data," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2736–2754, Oct. 2006.

[12] S. A. Sader, R. B. Waide, W. T. Lawrence, and A. T. Joyce, "Tropical forest biomass and successional age class relationships to a vegetation index derived from Landsat TM data," *Remote Sens. Environ.*, vol. 28, pp. 143–156, 1989.

[13] J. B. Lindsay, I. F. Creed, and F. D. Beall, "Drainage basin morphometrics for depressional landscapes," *Water Resour. Res.*, vol. 40, no. 9, p. W09 307, 2004.

[14] S. W. Lyon, M. T. Walter, P. Gérard-Marchant, and T. S. Steenhuis, "Using a topographic index to distribute variable source area runoff predicted with the SCS curve-number equation," *Hydrol. Process.*, vol. 18, no. 15, pp. 2757–2771, Oct. 2004.

[15] J. Li and W. Chen, "A rule-based method for mapping Canada's wetlands using optical, radar and DEM data," *Int. J. Remote Sens.*, vol. 26, no. 22, pp. 5051–5069, Jan. 2005.

[16] J. Töyrä and A. Pietroniro, "Towards operational monitoring of a northern wetland using geomatics-based techniques," *Remote Sens. Environ.*, vol. 97, no. 2, pp. 174–191, Jul. 2005.

[17] Y. H. Kaheil, E. Rosero, M. K. Gill, M. McKee, and L. Bastidas, "Downscaling and forecasting of evapotranspiration using a synthetic model of wavelets and support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 9, pp. 2692–2707, Sep. 2008.

[18] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1998.

[19] G. Z. Sass and I. F. Creed, "Characterizing hydrodynamics on boreal landscapes using archived synthetic aperture radar imagery," *Hydrol. Process.*, vol. 22, no. 11, pp. 1687–1699, May 2008.