Using Support Vector Machines to Automatically Extract Open Water Signatures from POLDER Multi-Angle Data Over Boreal Regions

J. Pierce[1], M. Diaz-Barrios[1], J. Pinzon[2], S.L. Ustin[1] , P. Shih[3], S. Tournois[1],
P.J. Zarco-Tejada[1], V.C. Vanderbilt[4] and G.L. Perry[4]
[1] Department of Land, Air and Water Resources, University of California, Davis CA 95616, USA
[2] NASA Goddard Space Flight Center, Greenbelt MD
[3] National Chiao Tung University, Taiwan
[4] NASA Ames Research Center, Moffett Field, CA 93401, USA

***Abstract* - This study used Support Vector Machines to classify multiangle POLDER data.**

I. Introduction

Boreal wetland ecosystems cover an estimated 90 x 10^6 ha, about 36% of global wetlands, and are a major source of trace gases emissions to the atmosphere [1]. Four to 20 percent of the global emission of methane to the atmosphere comes from wetlands north of 40°N latitude [2]. Large uncertainties in emissions exist because of large spatial and temporal variation in the production and consumption of methane. Accurate knowledge of the areal extent of open water and inundated vegetation is critical to estimating magnitudes of trace gas emissions. Improvements in land cover mapping have been sought using physical-modeling approaches [3],[4], neural networks [5],[6], and active-microwave [7], examples that demonstrate the difficulties of separating open water, inundated vegetation and dry upland vegetation. Here we examine the feasibility of using a support vector machine to classify POLDER data representing open water, inundated vegetation and dry upland vegetation.

II. Methods

Airborne POLDER sensor [8] data were collected in the red (665 nm) and infrared (865 nm) spectral regions on 21 July 1994 over the Southern Study Area (SSA) of the *BOReal Ecosystem-Atmosphere Study* (BOREAS) centered at 54ºN, 105ºW in central Saskatchewan, Canada. After resampling and registration, each POLDER pixel represented a ground area of 150m x 150m that was observed in each of 16 different view directions spaced at 5° intervals in the principle plane ±50º about nadir.

The forest cover map used as ground truth, derived from forest cover data provided by the *Saskatchewan Environment and Resource Management, Forestry Branch-Inventory Unit*, indicated the land cover consisted of fens (6.9%), black spruce (27.4%), jack pine (13.6%), and aspen (23.5%), grasses, mixed land use, and open water.

The Support Vector Machine (SVM) classifier [9] was designed to meet two objectives, i) that it is accurate when assessed against a test set and ii) that we can deduce from its design a level of confidence that the trained machine will accurately classify comparable data outside of the training and testing sets. Our analysis was restricted to four candidate polynomial support vector machines: linear, quadratic, quartic (4^{th} degree) and septic (7^{th} degree). Machines were trained to discriminate open water from all other classes "without penalty", that is, to seek to minimize the empirical risk function allowing no error in classification. All machines were trained on 25% of the classified data from each of the two wavelength bands, red and near infrared, of the POLDER data.

III. Results and Discussion

Table I provides estimates of the Vapnik-Chervonenkis (VC) dimension and the mean probability of misclassification when applied to subsequent data as well as a conservative estimate for the upper bound on the 95% confidence interval. While higher order machines provide better training results, they generally prove less robust when applied to subsequent data. The loss of robustness is seen by the growth of the VC dimension; the loss of confidence is manifested in the growth of the upper bound of the confidence interval. The balance of empirical risk versus generality favors the linear machine, with an estimated mean probability of misclassification of subsequent data of a remarkable 2%, but with a 95% confidence interval for the conservative upper bound of 68.23%. Both red and near infrared spectral bands are nearly equal in predictive ability.

By establishing a classifier solely based on the presence or absence of open water, the machines were less able to discriminate inundated vegetation from dry vegetation for either red or infrared bands. Accurate discrimination of inundated vegetation was achieved only by overfitting the training data, thereby sacrificing robustness of the support vector machine. After open water has been filtered from the image, a support vector regression machine, rather than a classifier, might provide a basis for discriminating the remaining community types. Alternatively, machines that use multispectral multiple view direction bands may provide sufficiently more information about the surface cover that accuracy is improved. The results also demonstrate that the septic 7th degree machine was the most robust of the three polynomial machines for each of the three binary classifiers of interest.

IV. Conclusions

This study reports results of applying Support Vector Machines to classify land cover types using multi-angle POLDER data. These results suggest that SVM can be applied to extend our analysis to classify additional types of inundated vegetation and non-inundated vegetation by masking open water after it is classified and/or adding additional POLDER spectral bands to the multiple view angle data set.

References

[1] A. Joabsson T.R. Christensen, B. Wallen, "Vascular plant controls on methane emissions from northern peatforming wetlands," *Trends Ecol. Evol.* vol. 14 (10) pp. 385-388, 1999.

[2] M.K. Cao, S. Marshall, K. Gregson, "Global carbon exchange and methane emissions from natural wetlands: application of a process-based model," *J. Geophys. Res.* vol. 101 (D9) pp. 14399-14414, 1996.

[3] F.G. Hall, D. E. Knapp, and F. Huemmrich, "Physically based classification and satellite mapping of biophysical characteristics in the southern boreal forest," *J. Geophys.Res.*, vol. 102, (D24), pp. 29567-29580, 1997.

[4] D.R. Peddle, F. G. Hall, E. F. LeDrew, and D. E. Knapp, "Classification of forest land cover in BOREAS. II: comparison of results from a sub-pixel scale physical modeling approach and a training based method," *Can. J. Remote Sens.*, vol. 23, pp. 131-142, 1997.

[5] J.A. Benediktson, P.H. Swain, and O.K. Ersoy, "Neural network approaches versus statistical methods in classification of multi source remote sensing data," *IEEE Trans. Geosci. and Remote Sens.*, vol. 28, pp. 540- 552, 1990.

[6] C.R. Duguay and D. R. Peddle, "Comparison of evidential reasoning and neural network approaches in a multi-source classification of alpine tundra vegetation," *Can. J. Remote Sens.*, vol. 22, pp. 433-440, 1996.

[7] K.J. Ranson, G. Sun, R. H. Lang, N. S. Chauhan, R. J. Cacciola, and O. Kilic, "Mapping of boreal forest biomass from spaceborne synthetic aperture radar," *J. Geophys. Res.*, vol. 102 (D24), pp. 29599-29610, 1997.

[8] F.-M. Bréon, V. Vanderbilt, M. Leroy, P. Bicheron, C.L. Walthall, and J.E. Kalshoven, "Evidence of hot spot directional signature from airborne POLDER measurements," *IEEE Trans. Geosci. and Remote Sens.* vol. 35, pp. 479-484, 1997.

[9] V. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. New York, NY: Springer, 2000.

TABLE I.

Support Vector Machine (SVM) classification results

	Support Vector Machine							
	linear		quadratic		quartic		septic	
	Spectral Band							
	red	nir	red	nir	red	nir	red	nir
support vectors	4	3	5	6	5	5	3	5
Margin	1.661	1.648	3.857	2.170	16.71	3.461	112.5	5.490
Bounding Sphere	8.59	13.0	73.9	196.	5467	3.856e+4	4,316,800	2.343e+8
VC Dimension	15	17	21	137	80	3877	1365	170,545
Mean Probability of Misclassification (%)	2.67	2.00	3.33	4.00	3.33	3.33	2.00	3.33
95% Confidence Interval Upper Bound (%)	61.52	68.23	73.41	100	100	100	undefined	100.