LANDUSE CLASSIFICATION FOR STORMWATER MODELING USING BAYSIAN NETWORKS

M. Park* and M. K. Stenstrom**

*Department of Civil and Environmental Engineering, UCLA, 405 Hilgard Avenue, Los Angeles, CA, 90095, USA (*mhpark@seas.ucla.edu **stenstro@seas.ucla.edu)*

ABSTRACT
Reducing pollution from diffuse sources, such as stormwater runoff, is now the major emphasis of regulatory agencies in many urban areas. Monitoring and modeling of such large-scale problems is inherently difficult, and many approaches require land use information to estimate emissions. Acquiring land use information can be time consuming and expensive using conventional ground surveys. Remotely sensed land use is a possible alternative, and remote sensing is receiving increasing attention from environmental engineers for its utility to more economically and accurately monitor large-scale processes. Land use from digital images has been classified by using statistical methods and artificial intelligence applications. For our research, Bayesian networks were evaluated as an intelligent classification tool. The Landsat TM image covering parts of Los Angeles was used. We identified the input parameters of the image that directly affect the classification label of pixels, which was spectral signature of band 5. Incorporating ancillary input data such as geospatial information significantly improved classification accuracy. Finally, we propose a new land use classification system suitable for stormwater modeling since the USGS land use and land cover classification systems may not be the best approach for environmental applications.

Keywords **Bayesian networks; land use classification; remote sensing; storm water monitoring**

INTRODUCTION
Remote sensing deals with information about the earth resource without actual contact (Jensen 1996). Recently, remote sensing has become more important for monitoring the environment of the Earth. For example, land use classifications from remotely sensed images are used for modeling non-point source pollution in urban area. Non-point source pollution has many different sources based on land use activities. Therefore, urban land use information is essential to manage non-point source pollution problems caused by stormwater runoff quality and quantity. Acquiring land use information is time consuming and expensive using conventional ground surveys but becomes inexpensive and repetitive using remote sensing data.
For application of remote sensing data, digital image classification has been examined by using statistical methods (e.g. maximum likelihood) and artificial intelligence applications (e.g. expert systems and neural networks, Foody *et al*., 1992; Paola *et al*., 1995). Among these methods, the use of artificial intelligence (AI) techniques have been increasingly popular and reported to be successful (Stefanov *et al*., 2001; Paola *et al*., 1995). Bayesian networks are one of the AI learning methods and have yet received little attention in remote sensing image classification since it is a relatively young AI methodology. Recently Bayesian networks have been recognized as excellent classification tools in many areas such as medicine, vision, natural language processing and decision-making (Russel, 1995; Lucas *et al*., 1999).

Bayesian networks were proposed by Pearl by combining Bayes' theorem and graph theory (Pearl, 1988) so that the networks have both rich statistical expression and clear graphical representation that shows relationships among variables. A Bayesian network is graphically represented as a directed acyclic graph (DAG) that consists of a set of nodes in which the dependent variables are connected by arcs (Neapolitan, 1990; Duda *et al*., 2001). The structure of the network explicitly shows the dependency relationships among variables. This helps us to easily understand the cause and effect relationship in the given domain even under uncertainty. Given a network structure and data, it is possible to compute a posterior probabilistic estimate of a target variable. For example, the simplest Bayesian networks, sometimes called naïve Bayesian classifiers, are used to predict the value of one variable, called the class node, from a set of measured variables.
In a typical Bayesian network, we are given a class node, C and several measured variables, Li. Hence, for example, if there are three measured variables and we assume that L1, L2 and L3 are conditionally independent and probability distributions sum to 1, we can write Bayes' theorem in the following form:

$$P(C\,|\,L1,L2,L3) = \alpha P(L1\,|\,C)P(L2\,|\,C)P(L3\,|\,C)P(C) \tag{1}$$

where á is set to normalize the probabilities to 1. Since we can easily estimate the quantities P(L1|C), P(L2|C) and P(L3|C) from known data values, we can use this equation to calculate a posterior probability distribution over C for any values of L1, L2 and L3.

An important feature of a Bayesian network is its assumption of conditional independence. A node is conditionally independent of its sibling nodes given its parent nodes. In other words, each parent node protects its child nodes from the effect of the non-descendant nodes in the network (Neapolitan, 1990). This makes Bayesian networks more computationally tractable because it is not necessary to use a large joint probability table listing the probabilities of all combinations of variables. Also conditional independence reduces the complexity of learning (Lucas *et al*. 1999). However, this also represents a limitation of naïve Bayesian classifiers since in practice, it is unlikely that a set of variables will be totally independent given a single class node. Nevertheless, many researchers have found that naïve Bayesian classifiers perform as well as or better than other more complicated networks (Friedman *et al*., 1997; Webb and Pazzani, 1998). Another desirable feature of a Bayesian network is its flexibility when updating. We can include additional information at any time and this does not affect or contradict the existing network.

One of the commonly used Bayesian networks are Maximum Weight Spanning Trees (MWST), which use Chow and Liu's approach to build an optimal dependence tree from data (Chow and Liu, 1968). In order to construct a MWST, first, it is necessary to compute the joint probabilities of the variables and their mutual information. Mutual information provides a way of measuring dependency. It is zero for completely independent variables but it increases as the variables become more strongly dependent.

$$\mathrm{Dep}(A,\ B) = \sum_{A\times B} P(a_i, b_j)\log_2\left[\frac{P(a_i, b_j)}{P(a_i)P(b_j)}\right] \quad (2)$$

Once the mutual information has been calculated for every pair of variables, a tree is constructed by choosing arcs. An arc is added in order of the magnitude of the mutual information of the corresponding variable pair while avoiding loops. The process ends when there are no unconnected nodes. The resulting network is a tree without causal directions. The class node is chosen using expert knowledge, and arrows are directed away from it to create a final directed structure. It requires a good representative set of data involving all variables in the model and sometimes an expert knowledge.

As described above, Bayesian networks can be constructed automatically from data with the relationships among variables (Pearl, 1999). The output of the networks can be viewed as both the range of values and the probability assigned to each value of the variables. This provides a straightforward way to draw inference and select the best result. In Bayesian networks, we can obtain structure and conditional distribution either in a subjective or objective way, or both ways. In the subjective way, causal direction and conditional probability are elicited from experts. In the objective way they are obtained from data. Besides, Bayesian networks are flexible in that we can treat any subset of the variables as inputs; one or more of the other variables can be used as outputs. Like other AI techniques, Bayesian networks are also capable of managing missing information. Most of all, Bayesian networks are able to provide not only a classification tool but also a higher level of explanation. For example, identification of the inputs that are most influential in reaching the conclusion is revealed. However, Bayesian networks have been restricted to problems having variables with discrete values (Mitchell 1997). Although they can handle continuous values with normal distributions, the algorithm for dealing with continuous values has not been extensively developed (Russell, 1995).

For our research, we proposed Bayesian networks as an alternative AI classification tool. This study investigated the performance of Bayesian network application to the remotely sensed image classification of urban land use. We identified input parameters of the image that directly affect the classification label of pixels. Furthermore we compared naïve Bayesian classifiers and MWSTs to find the optimal structure of the network in the given domain. We hope to achieve a better understanding of classification in remote sensing and to propose an improved approach in stormwater monitoring.

METHODS

The area of interest in this research was Marina del Rey and its vicinity, which is located in Santa Monica Bay watershed. Santa Monica Bay is regarded as an important natural resource, and restoring water quality and natural habitat is high priority. We used preprocessed Landsat Thematic Mapper (TM) images (1990, USGS scene no. 5237717480), which have been extensively used for urban land use analysis. We used all seven TM spectral bands that had been resampled to 25 by 25 m. Although this relatively low resolution allows at most level II classification, it is consistent with our categories of classification. The categories we used were modified classification level II in urban area based on Anderson's U.S. Geological Survey (USGS) classification system (Anderson *et al*., 1976). Land use categories were residential, commercial, light industrial, transportation, and open area. As ancillary input data, the coordinate values of the centroid of each pixel were calculated geospatially with GIS based on NAD1983 UTM (zone 11N) projection. Table 1 shows statistical information of input data for classification. Each pixel of each band is associated with digital number (DN) from 0 to 255 and each coordinate value has unit of meters.

Table 1 Input data statistics

Input	Minimum	Maximum	Mode	Median	Mean	Standard deviation
band1	5	255	101	107	112.91	32.57
band2	0	255	44	49	51.63	19.13
band3	0	255	65	65	68.26	29.11
band4	0	255	64	65	66.13	22.95
band5	0	255	80	94	97.87	39.39
band6	134	193	164	163	163.20	6.31
band7	0	255	430	53	56.05	26.06
X	365,234	369,534	368,259	367,734	367,674	1,127.93
Y	3,756,968	3,762,618	3,761,993	3,759,818	3,759,822	1,666.61

Land use data was obtained from Southern California Association of Governments (SCAG, 1993). This data was used for computation of prior probability for Bayesian networks and accuracy assessment. The land use data had been resampled to

25 by 25 m to match the resolution of TM image. The data contained 26,614 homogeneous land use pixels, and the subset used for training is shown in Table 2. Within this area of interest, the ratios of residential and open are higher than other class labels.

Table 2 Summary of land use information of training data

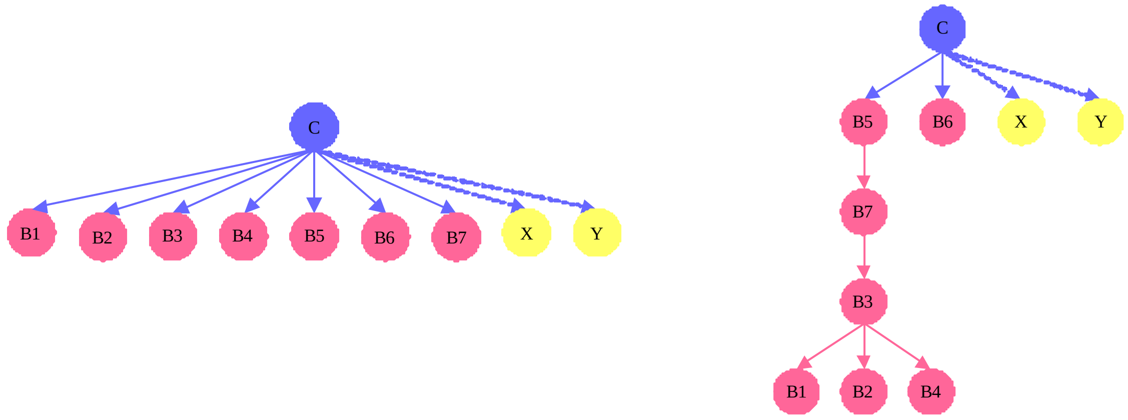
Land use class	USGS land use class	Number of pixels	Ratio
residential	residential (11)	1,553	39 %
commercial	commercial (12)	455	11 %
light industrial	industrial (13)	178	4 %
transportation	transportation (14)	426	11 %
open	agriculture (2), water (5)	1388	35 %
Total		4,000	100%

We generated two different data sets that used only spectral bands: DN values of band 1 to 7, and spectral bands with geospatial ancillary data (X-centroid and Y-centroid). The training data set for the networks were collected randomly from all classes in order to avoid undersampling the small classes (Jensen 1996). As mentioned above, all selected samples were homogeneous to properly represent each class. The total number of training data pixels was 4,000, which corresponds to 15% of total data. Since Bayesian networks mainly deal with discrete data, all data values were quantized. For example, all spectral band data were quantized to 10 values based on equal interval. The class node had 5 values corresponding to each land use category of interest.

With the given data, we conducted urban land use classification using naïve Bayesian classifiers and MWST. Firstly, in order to construct a naïve Bayesian classifier, we set land use category as a class node and all input data as child nodes of the class node. In this case, the networks were constructed with the spectral data only, and with spectral data and geospatial ancillary data respectively. In order to construct a MWST, we calculated mutual information to discover relationships between variables and chose the land use categories as the class node.

RESULTS AND DISCUSSIONS

The resulting structures of naïve Bayesian classifiers and MWSTs are given in Figure 1. The structures with solid lines represent the network with spectral data only, and the structures with dotted lines represent the network including geospatial ancillary data. In naïve Bayesian networks, every child node contributes to the value of the class node. Conversely, the structures of MWSTs show that the class node is determined mainly by bands 5 and 6 when considering only spectral data. Moreover, X- and Y-centroids are important when incorporating geospatial ancillary data. The class node values mainly depend on the Y-centroid. This indicates for this case that the geospatial information is more important than the spectral bands. Among spectral bands, band 5 has more effect on the class node than other spectral bands. This agrees with previous experts' belief that band 5 is important for land use classification. The visible bands show strong dependency, which agrees with the high correlation reported by other researchers. Infrared bands (5 and 7) are also strongly dependent, which are often found to be highly correlated. This result demonstrates that MWSTs reveal the dependency between variables in a reasonable way. However, band 6 dependencies with other nodes are quite low, and are at most 1/3 of the smallest value, in terms of mutual information.



(a) Naï ve Bayesian classifiers (b) Maximum weight spanning trees

Figure 1 *The structure of Bayesian networks for land use classification*

Table 3 shows the accuracy of classification as a confusion matrix. Residential, transportation and open are fairly well predicted whereas commercial and light industrial are not. For example, in the case of MWST with both spectral and geospatial input data, the user's accuracies of residential, transportation and open are 86%, 81% and 80%, respectively, and other user's accuracies are all below 30%. The producer's accuracies of residential, transportation and open are 71 %, 78% and 75% respectively but other class pixels are rather incorrectly assigned. This is due to the size of data set for the latter categories, which is rather small to properly train the network. Moreover, the range of the latter categories are quite overlapped to that of the former categories, so it is difficult to distinguish each class from the spectral signature. Table 4 shows the overall accuracy for each case.

Table 3 Confusion matrix of MWST with band1 to 7, X- and Y-centroids

	original						user's
	residential	commercial	industrial	transportation	open	total	accuracy
residential	130	4	1	0	16	151	**86%**
commercial	28	7	5	0	16	56	**13%**
industrial	8	0	3	0	3	14	**21%**
transportation	1	0	0	25	5	31	**81%**
open	15	2	5	7	119	148	**80%**
total	182	13	14	32	159	400	
producer's accuracy	**71%**	**54%**	**21%**	**78%**	**75%**		

Table 4 Overall accuracy of land use classification

Network	Input	Accuracy
Naïve Bayesian	band1 to band7	**51 %**
	band1 to band7, X and Y	**67 %**
MWST	band1 to band7	**52 %**
	band1 to band7, X and Y	**71 %**

As shown in Table 4, the accuracy of MWST with input including geospatial data shows the best prediction for each pixel. Although the difference in accuracy between naïve Bayesian classifiers and MWSTs is small, the difference arises from MWSTs' data contribution to the class node. Moreover, we can observe that the accuracy is considerably improved by incorporating geospatial data, which demonstrates that the spatial information is very important to determine the class value. However, the overall accuracy is still less than 85%, which is the recommended accuracy by Anderson at USGS (Anderson *et al.*, 1976). This might be due to errors in the SCAG data set or the difference in timing (SCAG and TM data were collected three years apart).

Other inaccuracies may be introduced because the ranges of each class are overlapped. For example, Figure 2 illustrates the distribution of pixels in two-dimensional spectral space with bands 4 and 5. In this figure, the range of each class is defined by its mean DN value and standard deviation. As shown in the figure, none of the classes is separated from other classes. The range of residential, in particular, resides within almost all other classes except transportation. For this reason, the overall accuracy when using spectral data is lower.

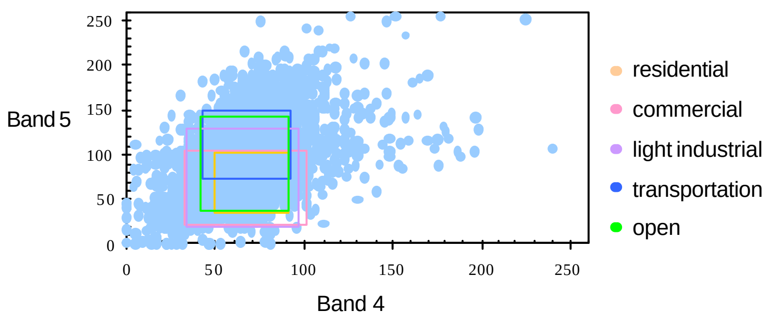
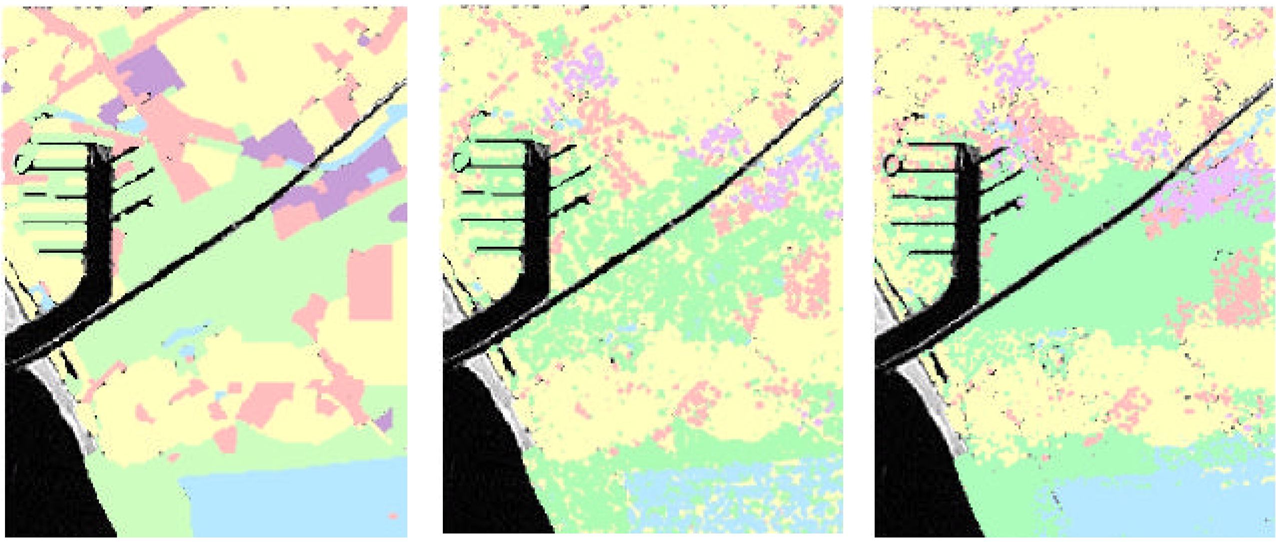


Figure 2 *Distribution of DN in two-dimensional spectral space with band 4 and 5*

The resulting thematic maps using MWSTs overlaid with TM image band 5 are shown in Figure 3. Compared with SCAG land use data, most thematic maps capture residential, transportation and open classes well. Also the maps clearly show that incorporating geospatial ancillary data improves the classification performance. This implies that incorporating other spatial ancillary data might improve the accuracy of classification. For example, digital elevation model (DEM) and contexture data can be incorporated for this purpose. We strongly believe that this can help the networks in learning more about the relationships between variables and accordingly generate more accurate thematic maps.

CONCLUSIONS

Our results indicate that Bayesian networks can perform urban land use classification well. Although the overall accuracy still needs improving, Bayesian networks are a useful, alternative tool for the classification of remotely sensed images. If we obtain higher resolution images such as the Landsat ETM$^+$ with panchromatic band or SPOT images and proper ground truth, Bayesian networks can be used as a classification system that further divides residential into single and multiple family residential. It may also be possible to isolate public land use, which is currently imbedded in commercial land use.



(a) SCAG land use data (b) with spectral data only (c) with spectral and geospatial data

residential commercial light industrial transportation open

Figure 3 *Land use classification thematic maps by using MWSTs*

Moreover, the ancillary data was useful for improving classification. In future research we will examine the effect of incorporating ancillary data such as DEM since Bayesian network structure can be easily updated by incorporating additional data without contradicting the existing network. Contexture will also be adopted to improve classification accuracy after the initial classification. Bayesian networks could be used to rank classes of mixed pixels by providing their individual probability of belonging to different classes. Also, the performance of Bayesian networks will be compared with existing supervised networks, including neural networks and maximum likelihood methods, to determine if they are a comparable or superior classification tool. An important advantage of the Bayesian networks is their ability to describe the relationship between variables, which is not provided by other supervised classification. Finally, we will explore different classification systems for environmental engineering purposes. The current classification systems have been extensively used but may not be best for stormwater modeling. Ultimately, we hope to propose a new classification system suitable for environmental engineering area and conduct urban land use classification in particular for stormwater monitoring.

ACKNOWLEDGEMENT

The author wishes to thank Professor Tom Gillespie in department of geography at UCLA for his advice and comment on this research.

REFERENCES

Anderson, J.R., Hardy, E., Roach, J., and Witmer, R. (1976). A land-use and land-cover classification system for use with remote sensor data, *U. S. Geological Survey Profession Paper* 964.

Chow, C.K., and Liu, C.N. (1968). Approximating Discrete Probability Distributions with Dependence Trees, *IEEE Transactions on Information theory*, IT-14 (3), 462-467.

Duda, R.O., Hart, P.E., Stork, D.G. (2001). *Pattern classification*, Wiley, 2nd ed.

Foody, G.M., Campbell, N.A., Trood, N.M., and Wood, T.F. (1992). Derivation and Applications of Probabilistic Measures of Class Membership from the maximum-likelihood Classification, *Photogram. Engng and Remote Sensing*,58(9),1335-1341.

Friedman, N., Geiger, D. and Goldszmidt, M. (1997). Bayesian Network Classifiers, *Machine Learning*, 29, 131-163.

Jensen, J.R. (1996). *Introductory digital image processing: A remote sensing perspective*, Prentice Hall.

Lucas, P. and Abu-Hanna, A. (1999). Prognostic methods in medicine, *Artificial Intelligence in Medicine*, 15, 105-119.

Mitchell, T.M. (1997). *Machine Learning*, McGraw Hill.

Neapolitan, R.E. (1990). *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*, Wiley.

Paola, J.D. and Schowengerdt, R.A. (1995). A detailed comparison of back propagation neural network and maximum-likelihood classification for urban land use classification, *IEEE Trans. Geoscience and Remote Sensing*, 33, 981-996.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann: San Mateo, CA

Pearl, J. (1999). Bayesian Networks, UCLA Cognitive Systems Laboratory, Technical Report (R-246), Revision I, July 1997 in *MIT Encyclopedia of the Cognitive Sciences*, Cambridge, MA.

Russell, S., and Norvig, P. (1995) *Artificial intelligence: A modern approach*, Prentice Hall.

Stefanov, W.L., Ramsey, M.S. and Christensen, P.R. (2001). Monitoring urban land cover change; An expert system approach to land cover classification of semiarid to arid urban centers, *Remote Sensing of Environment*, 77(2), 173-185.

Webb, G. and Pazzani, M. (1998), Adjusted Probability Naïve Bayesian Induction, Proc. *10th Australian Joint Conf. Artificial Intelligence*.