ELSEVIER

Journal of Environmental Management 86 (2008) 181–192

Journal of Environmental Management

www.elsevier.com/locate/jenvman

Classifying environmentally significant urban land uses with satellite imagery

Mi-Hyun Park*, Michael K. Stenstrom

Department of Civil and Environmental Engineering, University of California, Los Angeles, 405 Hilgard Avenue, Los Angeles, CA, 90095-1593, USA

Received 7 April 2006; received in revised form 27 November 2006; accepted 5 December 2006
Available online 8 February 2007

Abstract

We investigated Bayesian networks to classify urban land use from satellite imagery. Landsat Enhanced Thematic Mapper Plus (ETM^+) images were used for the classification in two study areas: (1) Marina del Rey and its vicinity in the Santa Monica Bay Watershed, CA and (2) drainage basins adjacent to the Sweetwater Reservoir in San Diego, CA. Bayesian networks provided 80–95% classification accuracy for urban land use using four different classification systems. The classifications were robust with small training data sets with normal and reduced radiometric resolution. The networks needed only 5% of the total data (i.e., 1500 pixels) for sample size and only 5- or 6-bit information for accurate classification. The network explicitly showed the relationship among variables from its structure and was also capable of utilizing information from non-spectral data. The classification can be used to provide timely and inexpensive land use information over large areas for environmental purposes such as estimating stormwater pollutant loads.

Keywords: Urban; Land use classification; Satellite imagery; Remote sensing; Bayesian networks

1. Introduction

Human activity has impacted the quality of receiving waters such as rivers, lakes, and oceans. Urbanization has deteriorated water quality by discharging more pollutants and higher volumes of runoff at accelerated rates. Recent environmental concerns are focusing on diffuse source pollution due to its significant impact on the receiving waters, and progress in managing point source pollution. However, diffuse sources of many pollutants, e.g., suspended solids, nutrients, heavy metals, oil and grease, and pesticides, are difficult to measure and are released over wide areas. As conventional monitoring is impractical, alternative management approaches are needed. Characterizing pollution emission rates using the relationship between land use and runoff water quality is one possible method (Stenstrom et al., 1984; Chiew and McMahon, 1999). These studies used land use information to estimate diffuse source pollution to receiving waters, which is particularly useful in unmonitored watersheds or drainage basins. For small drainage basins, ground surveys can be used to obtain land use information but it is too expensive for large drainage basins or watersheds. Moreover, land uses keep changing due to the rapid population growth and ongoing urbanization in many watersheds. In this case, land use information from public records, that may be updated only once per decade, cannot provide timely information. Therefore, using satellite imagery will be useful to classify land uses in the given watershed or drainage basins and to estimate diffuse source pollution.

For the last three decades, many approaches to land cover classification using satellite imagery have been widely explored. For example, Landsat series imagery has been extensively used since the first satellite was launched in 1972. Satellite imagery has been used to mainly classify the biophysical characteristics of land surface. Land use, on the other hand, which is human activity on the land, is difficult to determine directly from satellite imagery. Although land cover and land use have been used interchangeably in some cases (Anderson et al., 1976),

*Corresponding author. Tel.: +1 310 825 1408; fax: +1 310 206 5476.
E-mail addresses: mhpark@seas.ucla.edu (M.-H. Park), stenstro@seas.ucla.edu (M.K. Stenstrom).

0301-4797/$ - see front matter
doi:10.1016/j.jenvman.2006.12.010

they are not necessarily equivalent in many cases (Brown et al., 2000). One land use can consist of various land covers. Conversely, one type of land cover can be found in different land uses. This is especially true in urban areas because of their heterogeneous characteristics. For instance, residential land use can include different roof materials such as asphalt, clay tiles, wood shingles, or metal. A land use also contains a mixture of different urban features with different proportions. For instance, a pixel of residential land use could contain mixed signatures of buildings, pavement, driveways, and vegetation (i.e., grass yards and trees) (Clapham, 2003). The number of mixed pixels can be reduced using higher-resolution images, such as IKONOS and Quickbird, but these data are generally too expensive for watershed scale management. An example of the latter case is commercial and industrial land uses, which can exhibit similar spectral signatures if the roof or pavements use similar materials (Stefanov et al., 2001).

Land use information is important in many applications, i.e., tax assessment, urban planning, and environmental management. For diffuse source pollution management, successful urban land use classification is essential because the pollutant concentrations and runoff rates are correlated to land use and not land cover. Land cover for two different land uses may be impervious, but pollutant loads and environmental impacts can be quite different. For example, the pollutant loads from a local street are quite different from a school playground even though both could be composed of asphalt. In addition, a simple distinction between urban and non-urban land uses does not provide sufficient information for estimating diffuse source pollutant loading. The challenge of inferring different land use classifications from similar spectral signatures may depend upon using other information or ancillary data. Therefore, a method to infer land use directly from satellite imagery is needed for diffuse source pollution management in urban areas.

Bayesian networks, also called probabilistic networks or belief networks, have not been widely used for satellite image classification. However, they are known as excellent tools in many areas, such as medical informatics, natural language processing, computer vision, and decision making (Charniak, 1991; Russell and Norvig, 1995). Bayesian networks may be successful in satellite imagery application because they have been successfully used to classify images and signals (Malka and Lerner, 2004; Boutell and Luo, 2005; Gurwicz and Lerner, 2005). They were developed for reasoning and inference and are able to handle uncertainty with mixed signatures. Only recently have researchers attempted to adopt Bayesian networks for satellite image classification (Pal et al., 2001; Orun, 2004).

In this research, we used modified classification systems based on US Geological Survey (USGS) Levels I to III and evaluated the performance of Bayesian networks. We investigated the minimum sample size to train the network because there is no universal rule or theoretical background for selecting sample size. We examined the optimal number of variable states (i.e., radiometric resolution) since Bayesian networks are known to be more effective with quantized data. We also investigated the performance of Bayesian networks depending on different classification systems. Our hope is that these results will facilitate the use of Bayesian networks for classifying urban land use from satellite imagery.

2. Background of Bayesian networks

Bayesian networks are capable of automatically estimating the probability of target variable values by means of both probabilistic expression (Bayes' theorem) and graphical representation (graph theory). In the probability theory, $P(T)$, the probability of variable T, which lies in between 0 and 1, is equal to 1 if and only if T is certain (Cowell, 1998). If we have two variables, T and M, conditional probability of M given T is

$$P(M|T) = \frac{P(T,M)}{P(T)}. \tag{1}$$

The product rule of probability for two variables, T and M, is expressed as

$$P(T,M) = P(T|M)P(M) = P(M|T)P(T), \tag{2}$$

which defines the joint probabilities of T and M in terms of conditional and marginal probabilities of individual variables. By rearranging Eq. (2), the Bayesian theorem can be derived as

$$P(T|M) = \frac{P(M|T)P(T)}{P(M)}, \tag{3}$$

where $P(T|M)$ is the posterior probability of T given M, $P(T|M)$ is the likelihood of M given T, and $P(T)$ is the prior probability of T. Eq. (3) can be denoted as

$$\text{posterior probability} \propto \text{likelihood} \times \text{prior probability}. \tag{4}$$

The theorem makes it possible to compute the posterior probability of a target variable from the prior probability and likelihood under the conditional independence assumption. For example, if we have a target variable T and variable M_i, given $i = 3$, such that $P(M_i) \neq 0$ for all i and they are mutually exclusive and exhaustive, we can write their relationship in Bayes' theorem as follows:

$$P(T|M_1, M_2, M_3) = \frac{P(M_1, M_2, M_3|T)P(T)}{P(M_1, M_2, M_3)}. \tag{5}$$

If we can assume that M_1, M_2, and M_3 are conditionally independent given T, we can write:

$$P(T|M_1, M_2, M_3) = \frac{P(M_1|T)P(M_2|T)P(M_3|T)P(T)}{P(M_1)P(M_2)P(M_3)}. \tag{6}$$

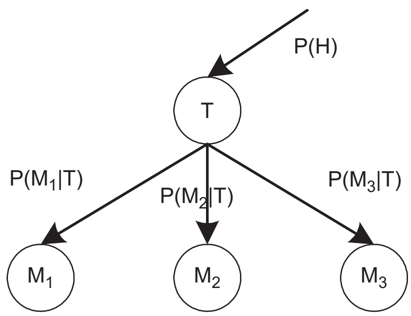

Fig. 1. An example of a Bayesian network.

Being that we are dealing with probability distributions that sum to 1 we can rewrite

$$P(T|M_1, M_2, M_3) = \alpha P(M_1|T)P(M_2|T)P(M_3|T)P(T), \tag{7}$$

where α is set to normalize the probabilities to 1. Because we can estimate the prior probability of $P(T)$ and the likelihoods of $P(M_1|T)$, $P(M_2|T)$, and $P(M_3|T)$ from the given data, we are able to calculate the posterior probability distribution over T for any values of M_1, M_2, and M_3. This equation is graphically represented in Fig. 1.

As shown in Fig. 1, a Bayesian network is a directed acyclic graph (DAG) that consists of nodes and arcs (Pearl, 1988; Neapolitan, 1990). A DAG is a directed graph that does not include any cycles. In a network, a node corresponds to a variable. An arc connects a parent node to a child node and the direction of the arc represents the direction of influence. Therefore, the structure of the network visually shows the relationships between connected nodes. The relationships are quantified by a conditional probability table, in which the probability distribution of a node is conditioned on its parent nodes.

In general, the joint probability distribution between any connected nodes in a network is written as

$$P(X) = \prod_{i=1}^{n} P(x_i|\mathrm{pa}(x_i)), \tag{8}$$

where $X = x_1, \ldots, x_n$ and $\mathrm{pa}(x_i)$ is a parent node of x_i. For example, T is a parent node of M_1, M_2, and M_3 in Fig. 1, In Bayesian networks, there is a conditional independence assumption where a node is conditionally independent of its non-descendant nodes given its parent node. In Fig. 1, M_1 is conditionally independent of M_2, and M_3 given T and vice versa.

$$P(M_1|T, M_2, M_3) = P(M_1|T). \tag{9}$$

3. Study area and data

3.1. Study area

Fig. 2 shows our study areas. The first study area was focused on Marina del Rey and its vicinity. Marina del Rey is located seaward of the Ballona Creek drainage basin, which is the southern part of the Santa Monica Bay watershed. There is growing concern in regards to Santa Monica Bay due to the importance of its water quality and natural resources. It receives the bulk of Los Angeles' treated wastewater and the stormwater runoff from the Los Angeles River via Ballona Creek, which has caused periodic beach closures (NCERQA, 2002). The highly urbanized area in the watershed has impacted the water quality of the Bay. Urban growth has also degraded the Ballona wetlands, which are included in this study area. Marina del Rey is one of the major sources of contamination into the Bay and Ballona wetlands. The size of the study area is approximately 25 km^2 and contains different urban land use features. The area is predominantly composed of residential and open land uses.

Another area was examined to verify the performance of Bayesian network classification. The second area includes six drainage basins adjacent to the Sweetwater reservoir as shown in Fig. 2. The Sweetwater River watershed, located in Southern California near San Diego, is a rapidly urbanizing watershed facing many challenges to water quality and the natural habitat. In order to protect the Sweetwater reservoir, a part of the Sweetwater River watershed, the Sweetwater Authority constructed an urban runoff diversion system immediately adjacent to the reservoir to divert urban runoff. The study area covers 37 km^2 and mainly consists of residential and open land uses.

3.2. Data sets

We used the Landsat Enhanced Thematic Mapper Plus (ETM$^+$) for both study areas. The Marina del Rey image was obtained on August 11, 2002 (path 41 row 36) and the Sweetwater reservoir image was obtained on December 18, 1999 (path 40 row 37). Subimages were extracted for both sites, including all bands except the panchromatic band. The thermal band, band 6, was resampled to match the resolution of the other bands based on the nearest neighbor method using RSI ENVI 4.0.

Three different levels of land use types were identified based on the USGS classification system (Anderson et al., 1976): (1) level I: urban, open, and water; (2) level II: residential, commercial, industrial, transportation, open, and water; and (3) level III: single family residential (SFR), multiple family residential (MFR), commercial, public, light industrial, transportation, open, and water. We also employed a classification system based on the imperviousness that is associated with land uses (Stenstrom and Strecker, 1993). The system is based upon a ground survey of different types of land uses and impervious areas, and was performed by the Los Angeles County Department of Public Works (LADPW) to document flood control benefits. Impervious area is important because it controls the amount of stormwater runoff and therefore the tendency of the land use to pollute receiving waters. Open

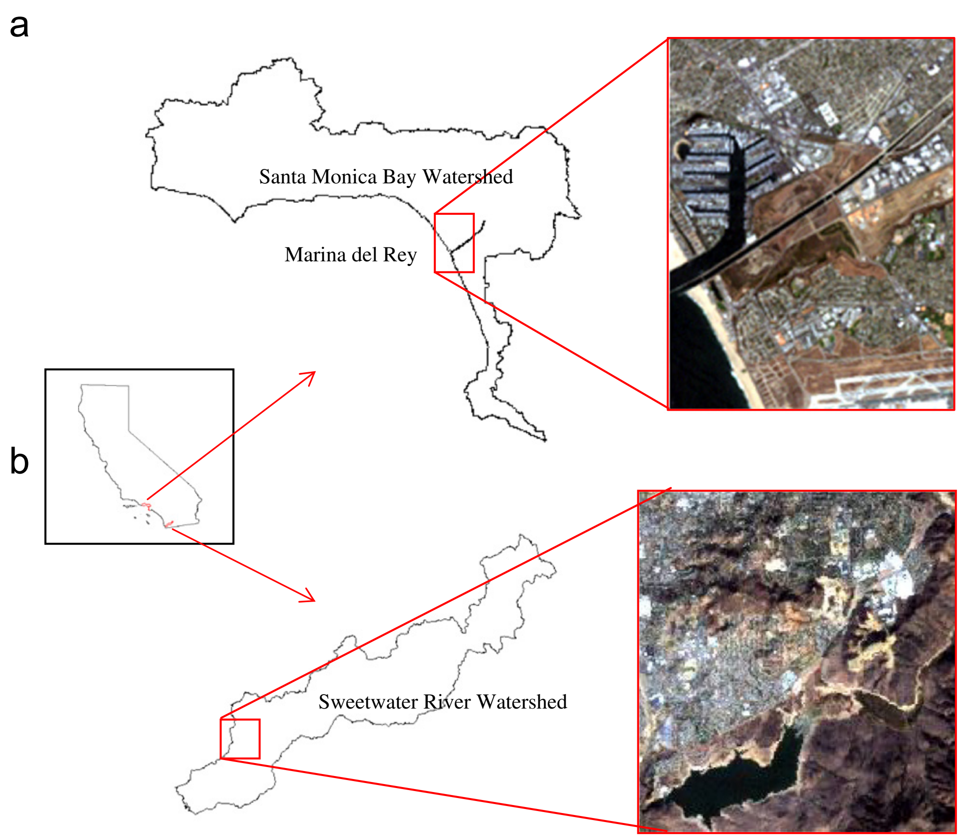

Fig. 2. Location of the study areas: (a) Marina del Rey and Santa Monica Bay watershed; (b) drainage basins into Sweetwater reservoir.

land use is pervious (0%) and produces little runoff. SFR land use is impervious (42%) while MFR land use has even greater roof and paved areas with imperviousness corresponding to 68%. Commercial, industrial, and transportation land uses have the greatest imperviousness (>80%).

Reference data for land use types were obtained from public records and aerial photography. Official land use data were obtained from the Southern California Association of Governments (SCAG, 2003) and San Diego's Regional Planning Agency (SANDAG, 2000). The SCAG land use data were updated from the existing data using digital orthophotography with 1 meter resolution. The SANDAG land use data were updated from the 1990 land use data using satellite images, digital orthophotography, the SanGIS landbase, the County Assessor Master Property Records file, and other ancillary information. However, the minimum land use polygon size from public records was larger than the Landsat pixels. For example, the minimum unit of SCAG data is 6070 m^2, which may contain mixed features. All samples in this research were checked against the aerial photography with 1 m resolution to assign sampling pixels into their appropriate categories.

The total number of pixels used for training and testing was 2110 for the Marina del Rey case, which corresponds to 7% of the total data, and 1120 for the Sweetwater Reservoir case, corresponding to 3% of the total data. The test data were excluded from the training data to avoid biased results. Table 2 shows statistical characteristics of the samples.

The coordinate values of each pixel (i.e., longitude and latitude), referred to as spatial ancillary data, were calculated from Landsat ETM$^+$ images using RSI ENVI 4.0 based on WGS 1984 UTM (zone 11N). These data were used to evaluate the benefit of incorporating spatial information of each pixel to classification accuracy.

4. Methods

4.1. Bayesian network structure

We used two structures of Bayesian networks: naïve Bayesian classifiers and Maximum Weight Spanning Trees (MWSTs). Naïve Bayesian classifiers are the simplest Bayesian networks and have only one class node. Every child node contributes to the value of the class node. Naïve Bayesian classifiers are easy to construct, although the relative contribution of each node is hard to identify.

MWSTs are alternative structures that are constructed from data using the optimal dependence tree developed by Chow and Liu (1968). Construction of the network does not require any information other than the mutual information between pairs of variables, which is calculated as follows:

$$\mathrm{MI}(X_i, X_j) = \sum P(x_i, x_j) \log\left(\frac{P(x_i, x_j)}{P(x_i)P(x_j)}\right), \tag{10}$$

where $\mathrm{MI}(X_i, X_j)$ is mutual information of random variables of X_i and X_j. $P(x_i, x_j)$ is a joint probability of x_i and x_j, and $P(x_i)$ and $P(x_j)$ are the probabilities of each random variable. The network was constructed by adding arcs between variable nodes with the greatest mutual information. Cycles or loops of arcs were avoided. The tree structure of MWSTs shows the relationship among variables.

We constructed both network structures by selecting land use as a class node. In both cases, two networks were constructed using spectral data only and using spectral data and spatial data. Prior to classification, the separability of the spectral signatures of the various land uses was evaluated using the Jeffries-Matusita distance (Richards and Jia, 1999), which produces separability indices ranging from 0 to 2 with higher numbers suggesting greater separability.

4.2. Sample size

We investigated the effect of different sample size on classification performance. The minimum training data size can be estimated from the statistical techniques. For example, Fitzpatrick-Lins (1981) suggested the following binomial formula:

$$N = \frac{4p(100 - p)}{\varepsilon^2}, \tag{11}$$

where N is a training data size, p the expected accuracy (%), and ε the allowable error. For instance, with 85% of expected accuracy and 2% of allowable error, the required training data size becomes 1275. Mather (1999) suggested an alternative method to estimate the minimum training data set size, as follows:

$$N = 30nc, \tag{12}$$

where n is the number of spectral bands and c the number of classes. This equation agrees with others (Swain, 1978; Van Genderen et al., 1978; Foody et al., 1995) who found that training data size for each class should be at least $30n$ to form representative training samples. In this case, with 7 spectral bands and 8 classes, at least 1680 training observations are needed. There are other methods of selecting data from each class (Jensen, 1996). An equal or proportional number of data can be selected from each class. The rule of thumb that Congalton (1991) proposed is to collect a minimum of 50 observations for each class.

In order to test the effect of the training data size, we used eight subsets of sample sizes: 500, 750, 1000, 1250, 1500, 1750, 2000, and 2110. The proportion of each class is consistent for all different training data sizes as shown in Table 1. For this analysis, only a naïve Bayesian classifier was used with 256 states (8 bits) for both spectral data and spatial data, using USGS classification system level III (8 land uses) (see also Table 2).

Table 1
Training data for USGS classification systems

Level	USGS land use class	Marina del Rey		Sweetwater reservoir	
		Samples	Ratio (%)	Samples	Ratio (%)
I	Urban	1310	62	525	51
	Open: wetland, barren land	550	26	300	29
	Water	250	12	200	20
	Total	2110	100	1025	100.0
II	Residential	780	37	300	30
	Commercial	230	11	120	11
	Industrial	120	6	55	5
	Transportation	180	9	50	5
	Open: wetland beach	550	26	300	29
	Water	250	12	200	20
	Total	2110	100	1025	100.0
III	SF residential	550	26	200	20
	MF residential	230	11	100	10
	Commercial	100	5	55	5
	Public	130	6	65	6
	Industrial	120	6	55	5
	Transportation	180	9	50	5
	Open: wetland, beach, parks	550	26	300	29
	Water	250	12	200	20
	Total	2110	100	1025	100

4.3. Quantization of variable states

Landsat ETM$^+$ imagery provides 8-bit information per pixel, which is by nature suitable for Bayesian network classification because Bayesian networks often require variables with discrete values. To evaluate the robustness to reduced information, a sensitivity analysis was performed using 3–8 bits. Equal width intervals were used for quantizing radiometric resolution. The equal width interval method breaks the range of the observed data values with k equally sized intervals (Liu et al., 2002; Yang and Webb, 2003). The number of intervals, k, can be selected subjectively. Each interval width is defined as follows:

$$\text{interval width} = \frac{x_{\max} - x_{\min}}{k}, \tag{13}$$

where $x_{\min}$ and $x_{\max}$ are the minimum and the maximum value of a variable x, respectively. This method was applied to all bands and the spatial ancillary data simultaneously. Therefore, spatial ancillary data had the same states as the spectral data when they were quantized. This method has been widely used in many machine learning algorithms because it is simple and effective (Hsu et al., 2000). Again, we used a naïve Bayesian classifier with sample size of 2110 pixels. The class node had eight land uses.

Table 2
Landsat ETM$^+$ image data statistics from samples

Band	Wave length (μm)	Minimum	Maximum	Range	Median	Mean	Standard deviation
(a) Marina del Rey							
1 (Blue)	0.45–0.52	80	228	148	107	113	23
2 (Green)	0.52–0.60	51	216	165	90	94	25
3 (Red)	0.63–0.69	34	253	219	98	100	33
4 (Near IR)	0.76–0.90	13	155	142	64	60	19
5 (IR)	1.55–1.75	7	255	248	96	95	38
6 (Thermal IR)	10.4–12.5	133	210	77	180	175	15
7 (IR)	2.08–2.35	3	255	252	73	74	32
(b) Sweetwater Reservoir Drainage							
1 (Blue)	0.45–0.52	46	127	81	60	64	14
2 (Green)	0.52–0.60	29	111	82	45	48	15
3 (Red)	0.63–0.69	21	127	106	50	51	20
4 (Near IR)	0.76–0.90	12	107	95	48	45	20
5 (IR)	1.55–1.75	9	124	115	63	58	29
6 (Thermal IR)	10.4–12.5	127	191	64	156	153	15
7 (IR)	2.08–2.35	8	123	115	51	46	23

4.4. Accuracy assessment

Overall accuracy and kappa (κ) coefficient (Congalton, 1991; Foody, 2002) were used to assess the Bayesian network performance as follows:

$$\text{Overall accuracy} = \frac{\sum_k n_{kk}}{N}, \tag{14}$$

$$\kappa = \frac{N\sum_k n_{kk} - \sum_k(\sum_j n_{kj}\sum_i n_{ik})}{N^2 - \sum_k(\sum_j n_{kj}\sum_i n_{ik})}, \tag{15}$$

where n_{ij} is the number of pixels with row i and column j and N is the total number of test pixels in the confusion matrix.

The accuracy was assessed using 10-fold cross-validation that is widely used in the machine learning community. In 10-fold cross-validation, the data were randomly divided into 10 subsets of equal size. The networks were trained 10 times with 9 subsets as training data and 1 subset as test data for each time. The average of overall accuracies and κ coefficients from each test were used to assess the accuracies.

Producer's and user's accuracy (Congalton, 1991) were used to assess per class accuracy using a confusion matrix:

$$\text{producer's accuracy} = \frac{n_{ij}}{\sum_i n_{ij}} 100, \tag{16}$$

$$\text{user's accuracy} = \frac{n_{ij}}{\sum_j n_{ij}} 100, \tag{17}$$

where n_{ij} is the number of pixels of each element in the matrix and i and j are the indices for row and column, respectively.

5. Results and discussion

5.1. Training data characteristics

Table 3 shows the Jeffries–Matusita distances between the pairs of land uses. Water and open land uses had excellent separability with all other land uses. SFR was the next most separable. Other land uses had lower separability. In Marina del Rey, the pairs of MFR, commercial, public, and industrial land uses especially had low separability, less than 1. The Sweetwater reservoir case had similar results and the pairs of residential, commercial, public and industrial land uses had low separability. Low separability less than 1 indicates the pair of land uses should be lumped because of the difficulty in separability. The difficulty may be due to similar spectral signatures of roofing materials (Stefanov et al., 2001; Herold et al., 2003).

5.2. Sample size

Fig. 3 shows the classification accuracies as a function of sample size. Accuracies increased monotonically with the sample size—up to 1500 pixels. After 1500 pixels, the accuracies reached a plateau and the increase in accuracy between 1500 and 2110 pixels was trivial (<1%). When incorporating spatial information, the overall accuracies and κ coefficients plateau at approximately 79% and 74%, respectively. Without incorporating spatial information, the classification accuracies decreased by 5%.

The results show that the Bayesian networks performed classification well with only a small sample size. A sample size of 1500 pixels, or 5% of the total data set, was sufficient for classification with acceptable accuracy. The required sample size is comparable to Mather's (1999) suggestion (Eq. (12)), which predicts a minimum training

Table 3
Spectral separability for level III classification using Jeffries–Matusita distance

	SFR	MFR	Commercial	Public	Industrial	Transportation	Open	Water
Marina del Rey								
SFR								
MFR	1.59							
Commercial	1.98	1.13						
Public	1.73	0.72	0.89					
Industrial	1.99	1.51	0.84	0.99				
Transportation	1.99	1.61	1.36	1.56	1.35			
Open	1.87	1.94	1.99	1.94	2.00	1.98		
Water	2.00	2.00	2.00	2.00	2.00	2.00	2.00	
Sweetwater								
SFR								
MFR	1.74							
Commercial	1.43	1.62						
Public	1.29	1.68	1.58					
Industrial	1.64	1.28	1.14	1.50				
Transportation	1.78	1.86	1.61	1.51	1.66			
Open	1.99	2.00	1.98	1.99	1.98	1.99		
Water	2.00	2.00	2.00	2.00	2.00	2.00	2.00	

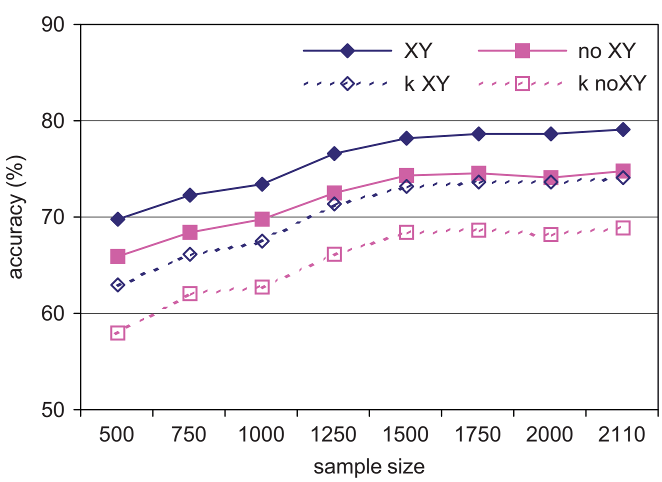

Fig. 3. Performance of Bayesian networks with different sample size. Note that solid lines with closed markers represent overall accuracy and those with open markers represent kappa coefficient.

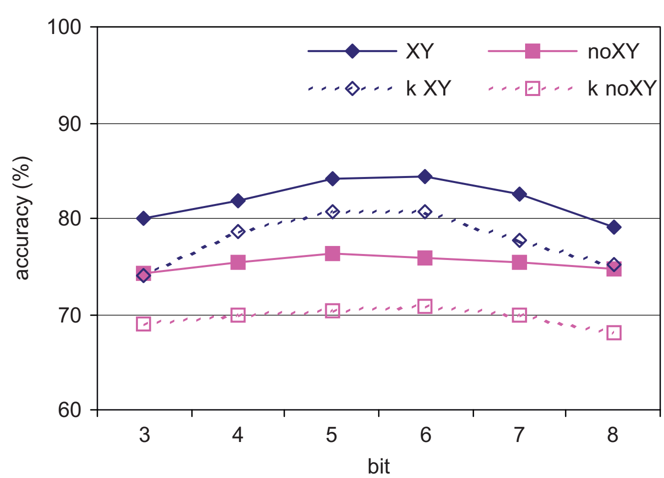

Fig. 4. Performance of Bayesian networks with quantized radiometric resolution. Note that solid lines with closed markers represent overall accuracy and those with open markers represent kappa coefficient.

data size of 1680 pixels (e.g., $30 \times 7 \times 8$). Our results confirm this minimum sample size for Bayesian network classification, although there is no theoretical research on the required size for Bayesian network training data sets. The required training data size may be dependent on the quantization level or the number of variable states. Landsat images contain 8-bit information (256 states) which may impact the required training data size.

5.3. Quantization of radiometric resolution

Fig. 4 illustrates the effect of quantization on classification accuracy. Overall accuracies and κ coefficients increased as the number of states increased, achieving a peak when using only 5 or 6 bits of the available 8 bits. At this point, the overall accuracy was 84% and κ coefficient was 81%. After 6 bits, the accuracies decreased as the number of variable states increased. Therefore, classification with 8-bit information was less accurate.

The results show that classification using Bayesian networks of Landsat imagery does not require 8-bit information. Our results show that classification with 3–6-bit information was just as accurate as using 8-bit information, where 5 or 6 bits were optimal. Generally, a large number of quantization levels can represent a continuous variable more accurately, but increases complexity and may introduce inaccuracies. Inaccuracies may be introduced because there may be insufficient training data to represent all pixels for all states. For example, 256 pixel values in the Landsat data with 8 land use states produces 2048 separate cases for the Bayesian network. For the examples in this paper, only 1899 training pixels out of 2110 samples are used, and some states will have no observations. Conversely, a small number of quantization levels may reduce the chance of states with no observations, but may lose information from the original continuous distribution. Malka and Lerner (2004) have shown that the optimum number of quantization levels can depend on the sample size.

The results show that using both extremes of quantization levels was less accurate. Using 5 or 6 bits resulted in more accurate representation of probabilistic tables for Bayesian learning. An advantage of using fewer bits is reduced complexity and computing resources, because the problem dimensionality is proportional to the number of quantization levels to the power of the number of the nodes (Malka and Lerner, 2004). This indicates that reducing the number of quantization levels may provide faster and more accurate learning (Dougherty et al., 1995; Liu et al., 2002; Yang and Webb, 2003). Reduced computing resources may be advantageous for very large problems.

5.4. Different classification systems

Fig. 5 shows the results of classification based on different classification systems. In the Marina del Rey case, naïve Bayesian classifiers provided overall accuracy of 93% and κ coefficient of 86% for level I classification. For level II classification, overall accuracies and κ coefficients were 82% and 76%, respectively. For level III classification, overall accuracies were slightly reduced to 79%. When spatial information was omitted, level II and III classification accuracy decreased by approximately 5%.

Imperviousness classifications were slightly more accurate than level II land use classification even though their class nodes had a similar number of states. In the Marina del Rey case, the overall accuracy and κ coefficient for imperviousness classification were 5% and 7% greater than the corresponding accuracy in level II land use classification. For the Sweetwater reservoir case, overall accuracies and κ coefficients for imperviousness were 10% and 15% greater, respectively, than the corresponding accuracy in

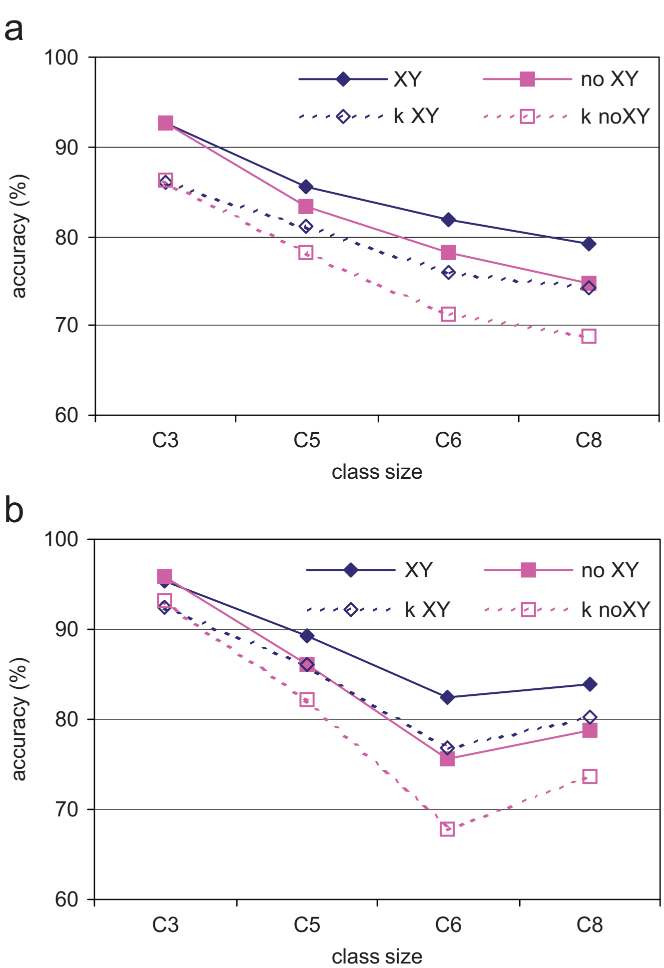

Fig. 5. Performance of Bayesian network classification for different classification systems: (a) Marina del Rey; (b) Sweetwater reservoir. Note that C3 corresponds to USGS land use/land cover level I classification system, C6 to level II system, C8 to level III system, and C5 to imperviousness classification associated with land use.

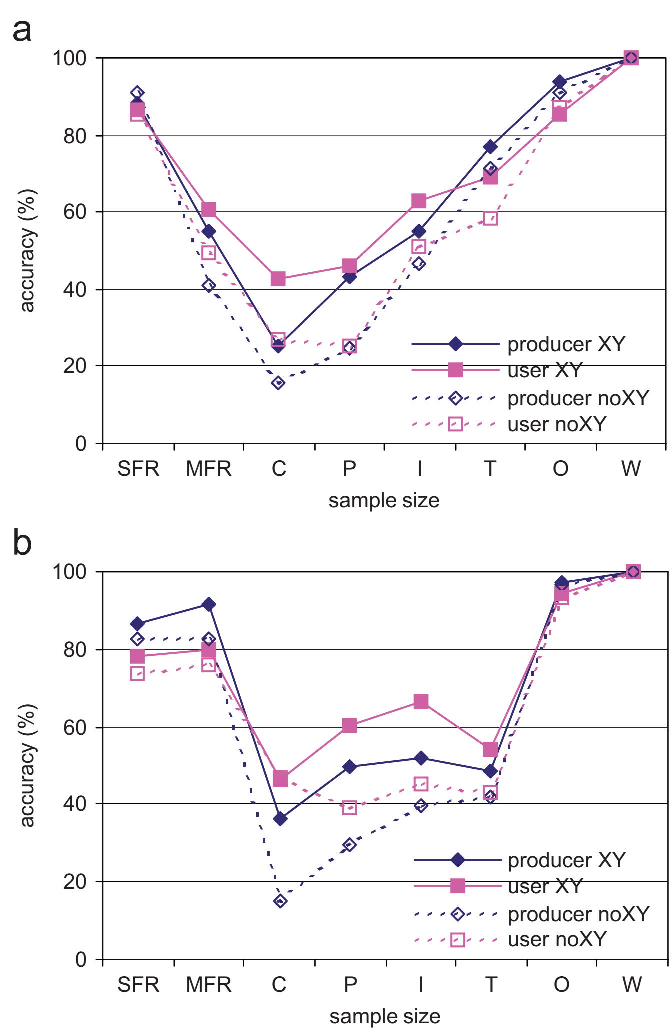

Fig. 6. Per-class accuracy assessment: (a) Marina del Rey; (b) Sweetwater reservoir.

level II land use classification. This is to be expected because the impervious surfaces have less varied spectral signatures.

Fig. 6 shows the producer's and user's accuracies for each land use based on confusion matrices with the level III classification system. In the Marina del Rey case, SFR, open and water land uses were successfully assigned, with accuracies greater than 85%, even without incorporating spatial information. By contrast, the producer's and user's classification accuracies of commercial and public land uses were approximately 50%. In the Sweetwater reservoir case, the accuracies of SFR, MFR, open, and water land uses were all above 85%. The producer's and user's accuracies of commercial land use were also approximately 50%. Omitting spatial information reduced the accuracies of commercial and public land uses by 9–21%. The separability results support this finding in that some urban land uses such as commercial, public, and industrial land uses are difficult to separate due to similar spectral signatures of roofing materials (Stefanov et al., 2001; Herold et al., 2003). However, misclassification among commercial, public and industrial land uses is not a fatal problem for stormwater management because these land uses usually generate similar stormwater pollutant loads and therefore can be combined together into a single class (Park and Stenstrom, 2006). When these three classes are combined into a single class, the overall classification accuracy improved to 86% and omission error of the combined land uses was reduced to 16% (Park, 2004).

In general, the Landsat ETM$^+$ imagery is considered to be suitable for level I or level II classification. In the Marina del Rey study, level III accuracy was only 2–3% lower than level II accuracy, and in the Sweetwater reservoir case, level III classification outperformed level II classification by 1 to 3%. This may be due in part to the differences in land use distribution (the Sweetwater example had less commercial, industrial, and public land use—see Fig. 9). Confusion exists among these land uses because of the inherent similarity of their spectral signatures. An advantage of the Bayesian network approach is its ability to learn zoning effects from spatial information. Therefore, spatial data will be helpful for a higher level of classification using Landsat ETM$^+$ imagery for areas with zoning rules or other rules regulating land use.

5.5. Bayesian network structures

Fig. 7 illustrates the two Bayesian network structures. Using naïve Bayesian classifiers, all input variables (bands 1–7) contribute to the value of the class node (land use or imperviousness class). By contrast, the structure of MWSTs shows that bands 2, 5, and 6 mainly contribute to the value of the class node. The importance of bands 2 and 5 is supported by other research. For example, Herold et al. (2003) measured spectral signature and also found

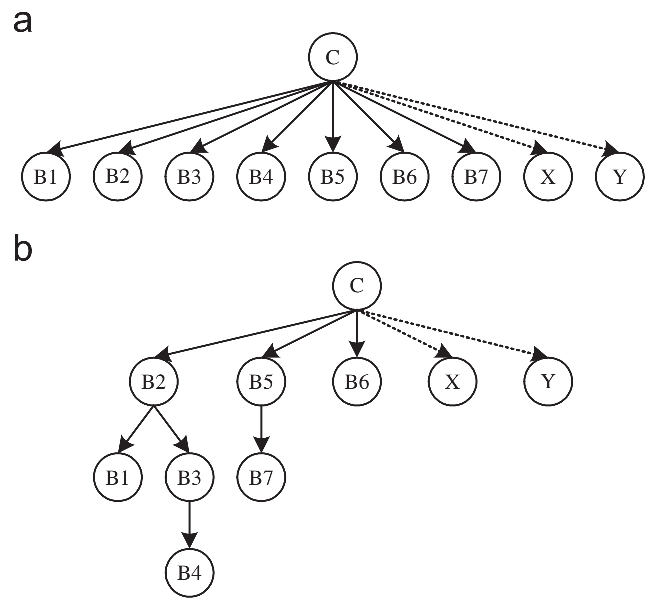

Fig. 7. Bayesian network structures for urban land use classification: (a) naïve Bayesian classifiers; (b) maximum weight spanning trees. The structures with solid lines represent the network with spectral data only, and the structures with dotted lines represent the network including spatial ancillary data.

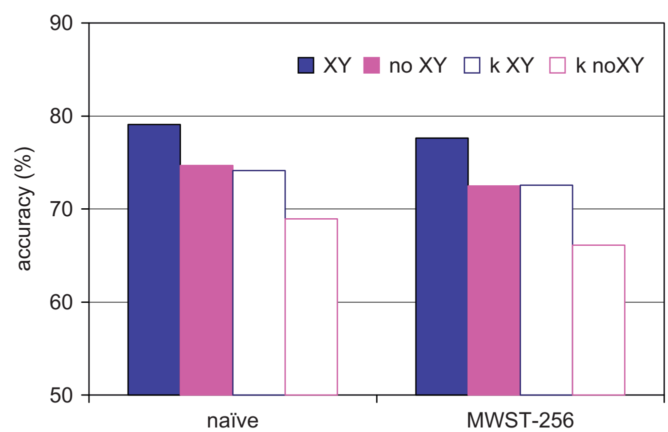

Fig. 8. Classification accuracies with different network structures.

that bands 2 and 5 provided suitable spectral information for separating urban features.

Our results showed high dependency between visible bands (1, 2, and 3) and middle infrared bands (5 and 7), which corresponds to similar findings of strong correlation between these bands (Jensen, 1996). This result implies that information among these bands may be redundant and implies that the network size can be reduced without losing information. Many researchers have not used band 6 for land cover classification, but in our case, including band 6 improved accuracy. Previous studies have reported that thermal information is related to urban surface characteristics (Voogt and Oke, 2003). These results illustrate how the structure of MWSTs can be helpful in understanding the relationships among variables for classification. Fig. 8

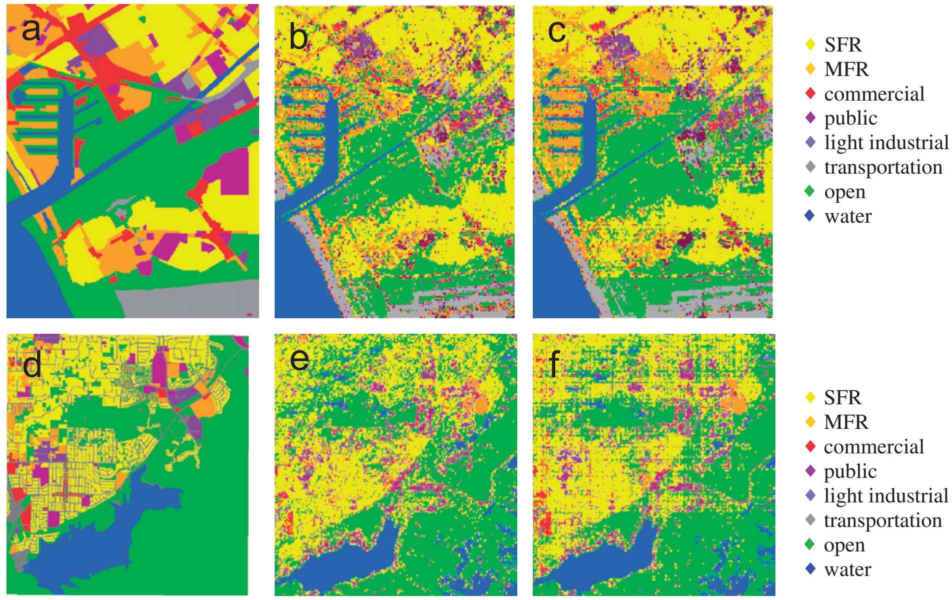

Fig. 9. Urban land use maps: (a) SCAG land use data; (b) classification with spectral data only in Marina del Rey; (c) classification with spectral and spatial data in Marina del Rey; (d) SANDAG land use data; (e) classification with spectral data only in Sweetwater reservoir; (f) classification with spectral and spatial data in Sweetwater reservoir.

shows that the performances of MWSTs and naïve Bayesian classifiers were similar.

5.6. Urban land use maps

Fig. 9 shows the resulting urban land use maps using naïve Bayesian classifiers in both the Marina del Rey and Sweetwater reservoir cases. Compared with existing land use information, i.e., SCAG and SANDAG land use data, most residential, transportation, open, and water pixels were well captured. The visual correspondence of the commercial, public, and industrial land use pixels is only approximate, which is due in part to the ambiguous definitions of land uses within these classes. The classification errors in the pixels belonging to the beach in Marina del Rey were assigned to transportation because beach pixels were undefined in the training data. In the Sweetwater reservoir case, the upper part of the reservoir shows an urban spectral signature because this part of the reservoir was dry at the time the image was taken. The pixels of the shadows in the open area were assigned to water land use because they had not been included in the training data.

Incorporating spatial information improved the visualization of all land uses even when the digital numbers of the spectral signatures were similar. The land use maps in Fig. 9 had less "salt and pepper" effect as the spatial information was added. This is a logical result because the same types of land uses tend to be located together. This occurs because of zoning laws and human nature, i.e., industrial sites are usually not located in SFR areas and vice versa.

The quantized spatial data provided to the Bayesian networks gives them the ability to make classification decisions based upon the locational or neighborhood information associated with each pixel. The number of contiguous pixels assigned to a particular land use having similar properties can be used by a Bayesian network to improve inferences. The roof on a large industrial complex or factory may be composed of the same material as an apartment building (e.g., asphalt), but a Bayesian network can make inferences based upon the number of contiguous pixels associated with the same land use. Although the examples provided in this paper have no way of making direct comparisons of contiguous pixels, the spatial information gives the network the ability to make these inferences. This implies that incorporating other ancillary data, for example, a digital elevation model, might also improve the accuracy of classification. Others have reported the benefits of incorporating ancillary data, such as population and road density information (Greenberg and Bradley, 1997), and contexture information (Stuckens et al., 2000) for land cover classification. We believe that incorporating ancillary data can provide more information to the networks in learning the relationships among variables and in generating more accurate land use maps. Bayesian networks have an important advantage in that they can show which variables contribute more to classification accuracy.

6. Conclusions

This study has shown that Bayesian networks can be used to accurately classify urban land use from Landsat ETM^+ imagery. Both naïve Bayesian classifiers and

MWSTs were evaluated in the Marina del Rey and Sweetwater reservoir drainage basins. The following specific conclusions are drawn:

(1) Bayesian networks were successful in urban land use classification from Landsat ETM$^+$ imagery. Classifying imperviousness, that was related to land use, showed better accuracy due to the low separability of urban features.
(2) Using spatial information, such as X and Y coordinate values of each pixel, improved classification accuracy. The greatest improvement was for commercial, public, and industrial land uses, which had lower per class accuracy. This shows that locational information can enhance highly impervious urban land use classification.
(3) Bayesian networks required only a small sample size for training, equal to approximately 5% of the total number of pixels.
(4) Quantization of variable states (i.e., radiometric resolution) was useful, showing that using all 8 bits were no more accurate than using only 3–6 bits. Using 5 or 6 bits provided the highest accuracy and reduced the required computing resources, which may be important for large problems.
(5) MWSTs have the advantage of showing which variables are most useful for classification, and overcome the ''black box'' criticism of neural networks, which generally do not provide indication of the cause and effect relationships.

This study can provide a guideline for urban land use classification using Bayesian networks. Bayesian networks handled the uncertainty associated with mixed signatures from urban land use. The success of this study should encourage others to use Bayesian networks for urban land use classification and its application to urban planning and environmental management. Well-identified land use information will improve estimates of environmental quality by identifying significant hot spots.

Acknowledgements

This work was supported in part by the US EPA grant R825831 and by the Sweetwater Authority. The authors thank Prof. Keith Clarke (Department of Geography, University of California, Santa Barbara), Prof. Thomas Gillespie (Department of Geography, University of California, Los Angeles), and Prof. Duncan F. Gillies (Department of Computing, Imperial College London) for their reviewing this manuscript.

References

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

Boutell, M., Luo, J.B., 2005. Beyond pixels: exploiting camera metadata for photo classification. Pattern Recognition 38 (6), 935–946.

Brown, D.G., Pijanowski, B.C., Duh, J.D., 2000. Modeling the relationships between land use and land cover on private lands in the Upper Midwest, USA. Journal of Environmental Management 59, 247–263.

Charniak, E., 1991. Bayesian network without tears. AI Magazine 12 (4), 50–63.

Chiew, F.H.S., McMahon, T.A., 1999. Modelling runoff and diffuse pollution loads in urban areas. Water Science and Technology 39, 241–248.

Chow, C.K., Liu, C.N., 1968. Approximating discrete probability distributions with dependence trees. IEEE Transactions on Information theory 14 (3), 462–467.

Clapham, W.B., 2003. Continuum-based classification of remotely sensed imagery to describe urban sprawl on a watershed scale. Remote Sensing of Environment 86, 322–340.

Congalton, R., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment 37, 35–46.

Cowell, R., 1998. Introduction to inference in Bayesian networks. In: Jordan, A. (Ed.), Learning in Graphical Models. The MIT Press, Cambridge, MA, pp. 9–26.

Dougherty, J., Kohavi, R., Sahami, M., 1995 Supervised and unsupervised discretization of continuous features. In: Proceedings of the 12th International Conference on Machine Learning, Los Altos, pp. 194–202.

Fitzpatrick-Lins, K., 1981. Comparison of sampling procedures and data analysis for a land-use and land-cover map. Photogrammetric Engineering and Remote Sensing 47 (3), 343–351.

Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote Sensing of Environment 80 (1), 185–201.

Foody, G.M., McCulloch, M.B., Yates, W.B., 1995. Classification of remotely sensed data by an artificial neural network: issues related to training data characteristics. Photogrammetric Engineering and Remote Sensing 61, 391–401.

Greenberg, J.D., Bradley, G.A., 1997. Analyzing the urban–wildland interface with GIS. Journal of Forestry 95, 18–22.

Gurwicz, Y., Lerner, B., 2005. Bayesian network classification using spline-approximated kernel. Pattern Recognition Letters 26 (11), 1761–1771.

Herold, M., Gardner, M.E., Roberts, D.A., 2003. Spectral resolution requirements for mapping urban areas. IEEE Transaction of Geoscience and Remote Sensing 41 (9), 1907–1919.

Hsu, C.N., Huang, H.J., Wong, T.T., 2000. Why discretization works for naive Bayesian classifiers. In: Proceedings of the Seventeenth International Conference on Machine Learning, pp. 309–406.

Jensen, J.R., 1996. Introductory Digital Image Processing: A Remote Sensing Perspective. Prentice-Hall, Upper Saddle River, NJ, pp. 248–249.

Liu, H., Hussain, F., Tan, C.L., Dash, M., 2002. Discretization: an enabling technique. Data Mining and Knowledge Discovery 6, 393–423.

Los Angeles County Department of Public Works (LADPW), Alhambra, CA.

Malka, R., Lerner, B., 2004. Classification of fluorescence in situ hybridization images using belief networks. Pattern Recognition Letters 25, 1777–1785.

Mather, P.M., 1999. Computer Processing of Remotely-Sensed Images: An Introduction. Wiley, Chichester, UK.

NCERQA, 2002. Integrated urban watershed analysis: the Los Angeles basin and coastal environment. Institute of the Environment Report #02-02.

Neapolitan, R.E., 1990. Probabilistic Reasoning in Expert Systems: Theory and Algorithms. Wiley, New York, pp. 120–123, 153–190.

Orun, A.B., 2004. Automated identification of man-made textural features on satellite imagery by Bayesian networks. Photogrammetric Engineering and Remote Sensing 70 (2), 211–216.

Pal, C., Swayne, D., Frey, B., 2001. The automated extraction of environmentally relevant features from digital imagery using Bayesian multi-resolution analysis. Advances in Environmental Research 5 (4), 435–444.

Park, M.-H., 2004. Bayesian network application to satellite image classification for stormwater management. Ph.D. Dissertation, University of California, Los Angeles/

Park, M-H., Stenstrom, M.K., 2006. Spatial estimates of stormwater-pollutant loading using Bayesian networks and geographic information systems. Water Environment Research 78 (4), 421–429.

Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo, CA, pp. 77–133.

Richards, J.A., Jia, X., 1999. Remote Sensing Digital Image Analysis: An Introduction, third ed. Springer, New York, pp. 240–247.

Russell, S., Norvig, P., 1995. Artificial Intelligence: A Modern Approach. Prentice Hall, Upper saddle river, NJ, pp. 436–470, 588–593.

SANDAG land cover and activity center, 2000. 〈http://www.sandag.cog.ca.us/〉.

Southern California Association of Governments, 2003. 〈http://wagsdata.scag.ca.gov〉.

Stefanov, W.L., Ramsey, M.S., 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.

Stenstrom, M.K., Silverman, G.S., Bursztynsky, T.A., 1984. Oil and grease in urban stormwaters. Journal of the Environmental Engineering ASCE 110, 58–72.

Stenstrom, M.K., Strecker, E., 1993. Assessment of Storm Drain Sources of Contaminants to Santa Monica Bay, Vol. I, Annual Pollutants Loadings to Santa Monica Bay from Stormwater Runoff, UCLA-ENG-93-62, Los Angeles, USA.

Stuckens, J., Coppin, P.R., Bauer, M.E., 2000. Integrating contextual information with per-pixel classification for improved land cover classification. Remote Sensing of Environment 71, 282–296.

Swain, P.H., 1978. Fundamentals of pattern recognition. In: Swain, P.H., Davis, S.M. (Eds.), Remote Sensing: The Quantitative Approach. McGraw-Hill, New York, pp. 136–187.

Sweetwater Authority 〈http://www.sweetwater.org/〉.

Van Genderen, J.L., Lock, B.F., Vass, P.A., 1978. Remote sensing: statistical testing of thematic map accuracy. Remote Sensing of Environment 7 (1), 3–14.

Voogt, J.A., Oke, T.R., 2003. Thermal remote sensing of urban climates. Remote Sensing of Environment 86, 370–384.

Yang, Y., Webb, G.I., 2003. Discretization for naïve-Bayes learning: managing discretization bias and variance. Technical Report 2003/131, School of Computer Science and Software Engineering, Monash University.