

## A NEW CLASSIFICATION SYSTEM FOR URBAN STORMWATER POLLUTANT LOADING: A CASE STUDY IN THE SANTA MONICA BAY AREA

M. Park and M. K. Stenstrom

*Department of Civil and Environmental Engineering, UCLA,  
Los Angeles, CA, 90095, USA*

### ABSTRACT

Urban stormwater runoff has been identified as the major source of many pollutants. Estimating stormwater mass emissions is inherently difficult due to the large area, many emission points, and the difficulty associated with sampling episodic storm events. Modeling urban stormwater runoff often requires land use information to estimate emissions.

Conventional collection of land use information, *i.e.* ground surveys, is time-consuming and expensive. Alternatively, land use data can be estimated from satellite imagery, and pollutants can be correlated using previously developed relationships. An alternate but less often used approach is to estimate stormwater pollutant loadings directly from satellite imagery, which is the objective of this research.

We conducted Bayesian network classification with Landsat TM image of Marina del Rey area in Santa Monica Bay Watershed. The results suggest an improved classification system for stormwater modeling, using open land use as low pollutant loading areas and transportation land use as high pollutant loading areas. Commercial and industrial land uses were medium and high loading areas, depending on the pollutant type. This indicates that management strategies should generally address transportation areas first. Classification systems should be developed for each water quality parameter. These results are useful in developing management practices for stormwater runoff.

### KEYWORDS

Bayesian networks, satellite imagery, stormwater pollutant loading

### INTRODUCTION

Proper management of stormwater runoff is required for continued improvement of receiving waters throughout the United States, and Santa Monica Bay is one example in the Los Angeles area. Urban stormwater runoff has been identified as a major source of many pollutants, which results from highly developed, impervious land use. However, estimating stormwater pollutant mass emissions is inherently difficult due to the lack of stormwater quality and rainfall data (Vaze and Chiew, 2003). For monitoring and modeling of urban runoff processes, many approaches have been developed using land use and/or impervious surface information to estimate the emissions. Conventional collection of land use or imperviousness information, such

as ground surveys, is time-consuming and prohibitively expensive. Land use data based on ground surveys might not be available for some areas of concern. Moreover, ground surveyed land use data were not developed for stormwater management purposes, which causes errors when used for estimating stormwater pollution.

Recent studies have developed land use estimated from remote sensing data such as satellite imagery (Ridd and Liu, 1998; Stefanov *et al.*, 2001; Pal and Mather, 2003; Park and Stenstrom, 2003). These approaches provide less expensive information in terms of time and expenses, and almost entire coverage of the earth. Satellite image classification can provide more suitable interpretation from spectral signatures of the given area for environmental purpose. However, most land use classification from satellite imagery has been conducted based on USGS land cover/land use classification system (Anderson *et al.*, 1976), and Levels I and II are not optimized for environmental purposes. Levels III and IV are too detailed and land uses must be regrouped into environmentally similar uses. For example, recreation and resorts are categorized as commercial and services, which show a different spectral signature from other commercial or official buildings, which are also in the same land use category. Communications and utilities belong to the same land use category with transportation, and have different stormwater pollution impacts. Parking lots in commercial and industrial areas have different impacts from commercial complexes and industrial plants. Therefore, a classification system suitable for stormwater management is needed.

Land use classification has long been used to quantify stormwater emissions. The Nationwide Urban Runoff Program (EPA, 1983) did not find significant differences among land uses, but studies accounting for regional differences did observe significant differences. For example, Stenstrom *et al.* (1984) found that residential property had mean oil and grease concentrations of 3.9 mg/L, while commercial property and parking lots had 13 mg/L concentrations. The Los Angeles County Department of Public Works (2000) monitored eight different land uses and found significant differences in 21 different pollutants. Others (Wong *et al.*, 1997; Burian *et al.*, 2002; Ackerman and Schiff, 2003; Goonetilleke, *et al.*, 2005) also found significant differences. The relationship between land use and pollutant emission rates is the basis for the load estimates developed in this research.

In this study, we propose a new, simplified classification system optimized for stormwater pollutant emissions in the Santa Monica Bay Watershed. This new classification system will be compared to the existing USGS classification system. In order to estimate stormwater pollutant emissions, we propose an alternate approach that directly estimates the emissions from satellite imagery using Bayesian networks. The results are displayed as thematic maps of stormwater pollutant loadings. These maps can be used to develop best management practices for stormwater pollution by identifying the areas that discharge high pollutant loading into receiving waters.

## METHODS

Our area of interest was Marina del Rey and its vicinity in the Santa Monica Bay Watershed. Santa Monica Bay is one of the most popular recreational resources in the United States as well as important ecological resources of natural habitat (Dojiri *et al.*, 2003). It receives contaminants including wastewater and stormwater runoff from the City of Los Angeles that impair water quality. Improved wastewater treatment plants have reduced the pollutant emissions from wastewater discharges. Therefore, emissions from stormwater runoff are receiving more concern since stormwater runoff has become the dominant source of many contaminants (Bay *et al.*, 2003).

A Landsat Thematic Mapper (TM) image (obtained on 09/03/1990, path 41 row 36) was used as our satellite image. We used all seven TM spectral bands and geospatial ancillary data, the coordinate values of each pixel, as inputs since geospatial ancillary data improves accuracy (Park and Stenstrom, 2003).

We classified stormwater pollutant loadings using Bayesian networks. Bayesian networks (Pearl, 1988) 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). The structure of the network explicitly shows the dependency relationships among variables.

Maximum weight spanning trees (MWSTs) were built from the given data as the network structures since MWSTs were found to outperform naïve Bayesian classifiers (Park and Stenstrom, 2003). The networks 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.

$$\text{Dep}(A, B) = \sum_{A \times B} P(a_i, b_j) \log_2 \left[ \frac{P(a_i, b_j)}{P(a_i)P(b_j)} \right] \quad (1)$$

A class node is selected. Mutual information is calculated and arcs are added in order of the magnitude of the mutual information of the corresponding variable pair. No loops are allowed. Arrows are directed away from the class node to create a final directed structure. In this case, the stormwater pollutant emissions for each water quality parameter were selected to be the class node. The network structure for estimating stormwater pollutant emissions is shown in Figure 1.

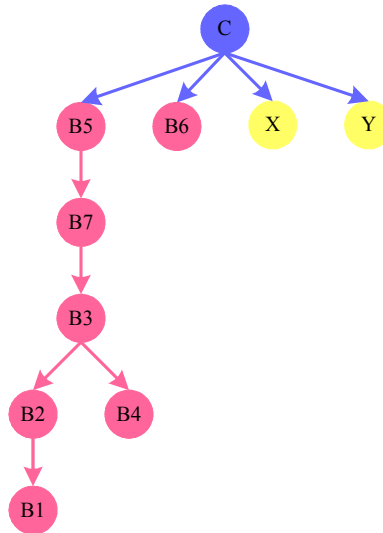


Figure 1. The structure of Bayesian networks

Note that  $B_n$  denotes the band, X and Y denotes the coordinate value of each pixel, and C denotes the class that is pollutant loading.

For a known a network structure and data, it is possible to compute a posterior probabilistic estimate of a target variable. In this case, the class node value was determined using equation (2) from dependency in the network.

$$P(C|B1,B2,B3,B4,B5,B6,B7,X,Y) = \frac{P(B1|B2) P(B2|B3) P(B4|B3) P(B3|B7) P(B7|B5) P(B5|C) P(B6|C) P(X|C)P(Y|C)P(C)}{P(C)} \quad (2)$$

The training data for the networks were collected using a random subset from all classes. The total number of training data pixels was 4,000, which corresponds to 15% of total data. All input data were quantized to 15 values based on equal width interval for optimal Bayesian network performance. The class node had three values: low, medium, and high pollutant emissions.

Six water quality parameters were examined to estimate pollutant loading: chemical oxygen demand (COD), biochemical oxygen demand, (BOD<sub>5</sub>), total Kjeldahl nitrogen (TKN), nitrite and nitrate (NO<sub>2&3</sub>), total phosphorus (TP) and soluble phosphorus (SP). For each parameter, unit pollutant emissions were calculated from runoff coefficients (RCs) and event mean concentrations (EMCs) based on existing empirical relationship developed for the Ballona Creek watershed as shown in Table 1 (Stenstrom *et al.*, 1984; Stenstrom *et al.*, 1993; Wong *et al.*, 1997; Ackerman and Schiff, 2003). The rainfall and number of storms per year can be assumed equal for all pixels due to the small size of the study area, and all stormwater runoff discharges to Santa Monica Bay. The following equation represents unit pollutant loadings.

$$P_i = \alpha \times RC \times EMC_i \quad (3)$$

where  $P_i$  is pollutant loads per unit pixel and unit rainfall for each water quality parameter  $i$  and  $\alpha$  is a normalization factor that depends on units and conversion factors.

Table 1 Runoff coefficient and EMCs for land uses in Ballona Creek Watershed

|     | RC   | COD | BOD <sub>5</sub> | TKN | NO <sub>2&amp;3</sub> | TP   | SP   |
|-----|------|-----|------------------|-----|-----------------------|------|------|
| SFR | 0.39 | 140 | 17               | 4.3 | 1.85                  | 0.85 | 0.25 |
| MFR | 0.58 | 130 | 15               | 2.4 | 1                     | 0.62 | 0.13 |
| C   | 0.74 | 90  | 14               | 2.0 | 1.2                   | 0.43 | 0.18 |
| P   | 0.66 | 90  | 14               | 2.0 | 1.2                   | 0.43 | 0.18 |
| I   | 0.74 | 130 | 14               | 2.0 | 1.2                   | 0.43 | 0.18 |
| T   | 0.66 | 130 | 15               | 2.4 | 1                     | 0.62 | 0.13 |
| O   | 0.10 | 95  | 2                | 2.8 | 1.45                  | 0.52 | 0.14 |

## RESULTS AND DISCUSSION

The stormwater pollutant emissions were classified into three classes for all water quality parameters. Figure 2 shows the thematic maps of pollutant emission estimates compared with official land used data obtained from SCAG land use data.

The pollutant loading maps of all water quality parameters show that the low pollutant emitting areas correspond to open land use due to its low RC associated with low imperviousness. Many pollutant loading maps show transportation land use such as Los Angeles International Airport as the high pollutant emitting areas. The pollutant loading map of BOD<sub>5</sub> shows commercial, public, industrial land uses as well as transportation land use contributes to high pollution emission. The

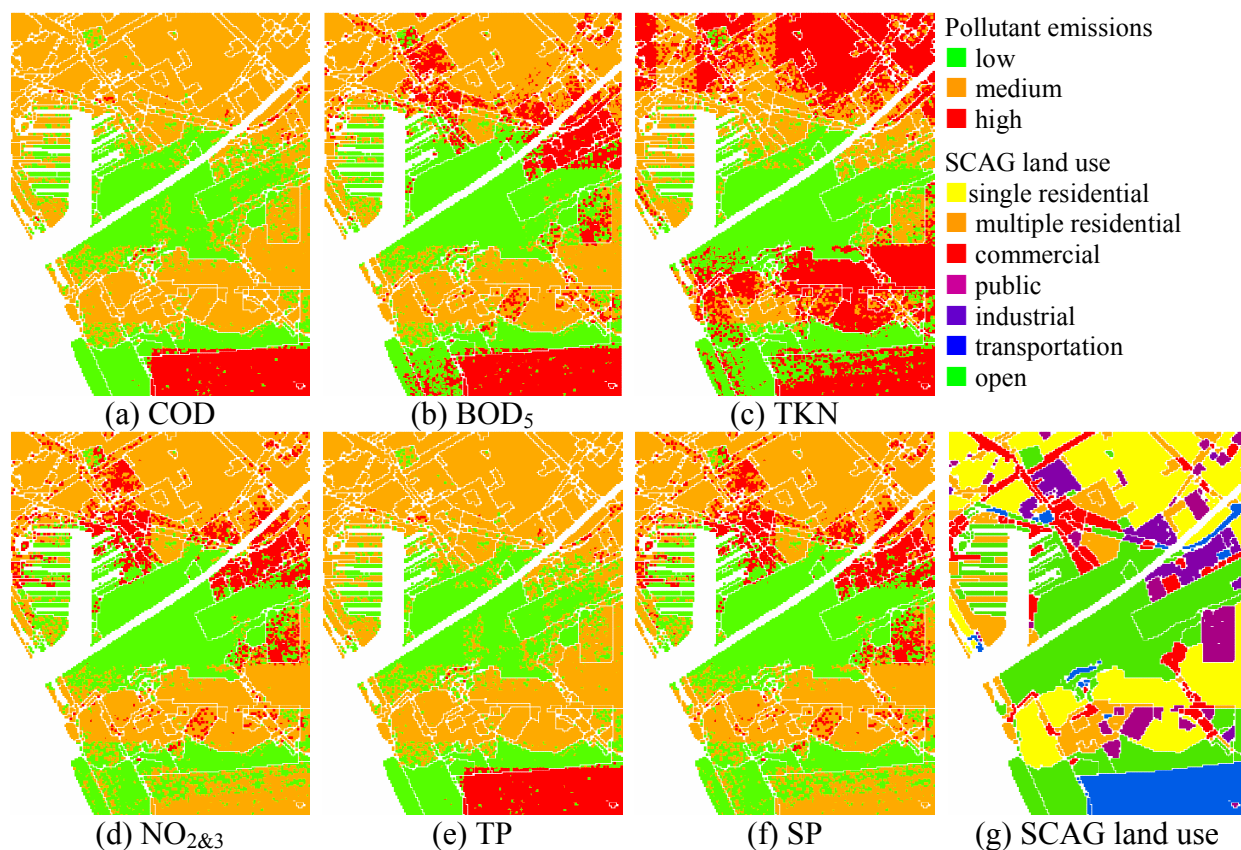


Figure 2. Maps of stormwater pollutant mass emissions

pollutant loading map of TKN shows single-family residential land use as well as transportation land use is high pollutant emitters. However, for NO<sub>2&3</sub> and SP loading, commercial, public, industry land uses are high pollutant loading areas, and transportation land use is a medium loading area.

These results imply that classification system for stormwater pollutant loading should be developed differently for individual water quality parameter. The only common class for all water quality parameters is low pollutant loading area. For the other classes, each water quality parameter exhibited different results. As shown above, COD and TP have only transportation land use as high pollutant loading areas. But TKN adds single-family residential and BOD<sub>5</sub> adds commercial, public, industrial as high pollutant loading areas. Conversely NO<sub>2&3</sub> and SP do not have transportation land use as high pollutant loading area. From these results, the new classification system for managing stormwater pollutant loads was developed as shown in Table 2.

Table 2 New classification system for stormwater pollutant mass emissions

|        | COD       | BOD <sub>5</sub> | TKN     | NO <sub>2&amp;3</sub> | TP        | SP    |
|--------|-----------|------------------|---------|-----------------------|-----------|-------|
| Low    | O         | O                | O       | O                     | O         | O     |
| Medium | S,M,C,P,I | S,M              | M,C,P,I | S,M,T                 | S,M,C,P,I | S,M,T |
| High   | T         | C,P,I,T          | S,T     | C,P,I                 | T         | C,P,I |

Note that S is single-family residential, M is multiple-family residential, C is commercial, P is public, I is industrial, T is transportation and O is open land use.

Satellite images were useful to estimate stormwater pollutant mass emissions, since conventional land use data such as Southern California Association of Governments (SCAG, 1993) data were not developed for environmental purposes. For example, some areas of the buffer zone inside the Los Angeles International Airport were identified as low pollutant loading areas in some cases. They are actually open land use, which were classified as transportation land use in the SCAG data and USGS classification system. Vegetated areas inside the institutional areas were also identified as low pollutant loading areas, which were classified as public land use in the SCAG data and USGS classification system. Recreational facilities including parks were also classified as low pollutant loading areas, which were often categorized into public land use in USGS classification system.

Figure 3 shows the overall accuracies of Bayesian network performance to predict stormwater pollutant mass emissions. Most estimates of the pollutant loading were improved compared to land use classification (72%, Park and Stenstrom, 2003) except for TKN. For COD and TP loading estimate, our methodology appeared to be useful because their overall accuracies were 84%. For other parameters, the inaccuracy might result from the fact that the spectral signatures of the different classes are not easy to distinguish. Even the spectral signatures in the same class could be more varied compared with those in different classes. For the example of TKN, pixels of medium pollutant loading areas are not separable from those of high pollutant loading areas. As mentioned earlier, SCAG data lumped nearby pixels to represent each land use, which might have different spectral signatures from satellite image. Therefore, the actual accuracies could be better than shown here since the accuracy was compared to SCAG land use data. Generally, Bayesian networks gave a reasonable estimate for stormwater pollutant mass emissions despite the potential problems of SCAG data.

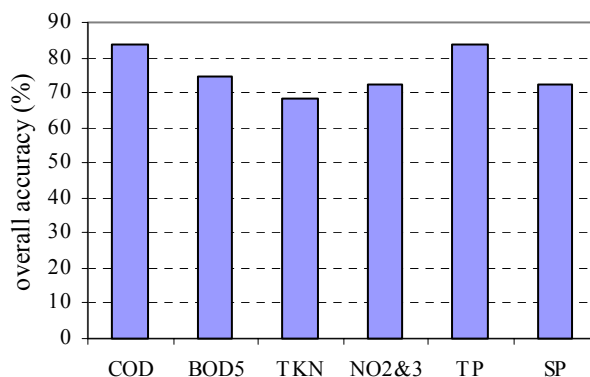


Figure 3. Overall accuracy of Bayesian networks

## CONCLUSIONS

Stormwater pollutant mass emissions were estimated from satellite imagery using Bayesian networks. This methodology facilitates pollutant loading estimates with more coverage compared to the conventional land use based model. This methodology provides practical information with reasonable accuracy, which can be used for environmental planning and management.

Open land use corresponds to low pollutant loads for all water quality parameters used in this study. Therefore, it is a better strategy to distinguish open and non-open areas for strict



management of stormwater pollution. In addition, transportation land use often contributes to high pollutant loads, which suggests that it should be an early target for stormwater management. Classification systems should be developed for each water quality parameter. The class of high pollution emitting areas can be expanded depending on the water quality parameter of interest.

## 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.
- Ackerman, D., and Schiff, K. (2003). Modeling storm water mass emissions to the Southern California Bight, *J. Environ. Eng. - ASCE*, Vol.129, No.4, 308-317.
- Bay, S., Jones, B.H., Schiff, K., and Washburn, L. (2003). Water quality impacts of stormwater discharges to Santa Monica Bay, *Mar. Environ. Res.*, Vol.56, 205–223.
- Burian, S.J., Brown M.J., and McPherson T.N. (2002). Evaluation of land use/land cover datasets for urban watershed modeling, *Water Sci. Tech.*, Vol. 45, No. 9, 269-276.
- Chow, C.K., and Liu, C.N. (1968). Approximating Discrete Probability Distributions with Dependence Trees, *IEEE T. Inform. Theory*, Vol.14, No.3, 462-467.
- Dojiri, M., Yamaguchi, M., Weisberg, S. B., and Lee H. J. (2003). Changing anthropogenic influence on the Santa Monica Bay watershed, *Mar. Environ. Res.*, Vol. 56, 1–14.
- Environmental Protection Agency (1983). *Results of the Nationwide Urban Runoff Program*, Vol. 1 EPA report No. 832R83112.
- Goonetillekea, A., Thomasb, E., Ginnc, S., and Gilbertc, D. (2005). Understanding the role of land use in urban stormwater quality management, *J. Environ. Manage.*, Vol. 74, No. 1, 31-42.
- Los Angeles County Department of Public Works (2000). *Los Angeles County 1994-2000 Integrated Receiving Water Impacts Report*.
- Neapolitan, R.E. (1990). *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*, Wiley.
- Park, M., and Stenstrom, M. K (2003). Landuse classification for stormwater modeling using Bayesian networks, In *Proc.7th IWA Conference, Diffuse Pollution and Basin Management*, Dublin, Ireland.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann: San Mateo, CA
- Ridd, M. K., and Liu, J. (1998). A comparison of four algorithms for change detection in an urban environment, *Remote Sens. Environ.*, Vol.63, 95–100.
- Southern California Association of Governments (1993). <http://www.scag.ca.gov>
- 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 Sens. Environ.*, Vol.77, No.2, 173-185.
- Stenstrom, M. K., Silverman, G. S. and Bursztynsky, T. A. (1984). Oil and Grease in Urban Stormwaters, *J. Environ. Eng. - ASCE*, Vol.110, No.1, 58-72.
- Stenstrom, M. K. and 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, I, 1-248.
- Vaze, J. and Chiew, F. H. S. (2003). Comparative evaluation of urban storm water quality models, *Water Resour. Res.*, Vol.39, No.10, 1280-1289.
- Wong, K., Strecker, E. W., and Stenstrom, M. K. (1997). A geographic information system to estimate stormwater pollutant mass loadings, *J. Environ. Eng. - ASCE*, Vol.123, 737-745.