Environmental Pollution xxx (2017) 1–8

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol

# Influence of traffic on build-up of polycyclic aromatic hydrocarbons on urban road surfaces: A Bayesian network modelling approach☆

Yingxia Li [a, *], Ziliang Jia [a], Buddhi Wijesiri [b, **], Ningning Song [a], Ashantha Goonetilleke [b]

[a] State Key Laboratory of Water Environment Simulation, School of Environment, Beijing Normal University, Beijing 100875, China
[b] Science and Engineering Faculty, Queensland University of Technology (QUT), GPO Box 2434, Brisbane, 4001, Queensland, Australia

ARTICLE INFO

*Article history:*
Received 7 June 2017
Received in revised form
29 October 2017
Accepted 30 October 2017
Available online xxx

*Keywords:*
Bayesian networks
Pollutant build-up
Polycyclic aromatic hydrocarbons
Stormwater quality
Stormwater pollutant processes

ABSTRACT

Due to their carcinogenic effects, Polycyclic Aromatic Hydrocarbons (PAHs) deposited on urban surfaces are a major concern in the context of stormwater pollution. However, the design of effective pollution mitigation strategies is challenging due to the lack of reliability in stormwater quality modelling outcomes. Current modelling approaches do not adequately replicate the interdependencies between pollutant processes and their influential factors. Using Bayesian Network modelling, this research study characterised the influence of vehicular traffic on the build-up of the sixteen US EPA classified priority PAHs. The predictive analysis was conditional on the structure of the proposed BN, which can be further improved by including more variables. This novel modelling approach facilitated the characterisation of the influence of traffic as a source of origin and also as a key factor that influences the re-distribution of PAHs, with positive or negative relationship between traffic volume and PAH build-up. It was evident that the re-distribution of particle-bound PAHs is determined by the particle size rather than the chemical characteristics such as volatility. Moreover, compared to commercial and residential land uses, mostly industrial land use contributes to the PAHs load released to the environment. Carcinogenic PAHs in industrial areas are likely to be associated with finer particles, while PAHs, which are not classified as human carcinogens, are likely to be found in the coarser particle fraction.



## 1. Introduction

Polycyclic Aromatic Hydrocarbons (PAHs) are a subset of the large group of hydrocarbons generated mainly through automobile use activities and other sources specific to urban land uses such as industrial activities. Exhaust emissions from incomplete combustion and leakages of fuel and lubricants release significant amounts of these toxicants into the urban environment (Li et al., 2017; Mummullage et al., 2016a, 2016b; Pitt and Voorhees, 2004). PAHs build up on urban surfaces such as roads during dry weather periods, and are generally found in association with particulate solids. During rainfall events, the runoff transports these hydrocarbons into receiving waters, deteriorating urban water quality. Based on the carcinogenic effects of PAHs, US EPA has identified sixteen species of PAHs as priority pollutants which require specific mitigation in the context of urban stormwater pollution for safeguarding human and ecosystem health (Brown and Peake, 2006; Gobel et al., 2007; Herngren et al., 2005; Ma et al., 2017; Manoli and Samara, 1999).

The design of stormwater pollution mitigation strategies needs informed decision making that relies on stormwater quality models. However, the deterministic modelling approach that is commonly adopted in stormwater quality modelling does not adequately assist the decision makers to evaluate the influence of anthropogenic and environmental factors on the degradation of stormwater quality. This is due to the poor replication of the interdependencies between pollutant processes and their influential factors (Beck et al., 2017; Wijesiri et al., 2016; Zoppou, 2001). This highlights the need for modelling approaches that can be easily implemented, while generating reliable and adequate information for stakeholders and decision makers to understand and evaluate the problem of interest.

Bayesian Networks (BNs) is an emerging approach for modelling

☆ This paper has been recommended for acceptance by Joerg Rinklebe.

* Corresponding author.
** Corresponding author.
*E-mail addresses:* yingxia@bnu.edu.cn (Y. Li), 201621180080@mail.bnu.edu.cn (Z. Jia), b.mahappukankanamalage@qut.edu.au (B. Wijesiri), 201331180007@mail.bnu.edu.cn (N. Song), a.goonetilleke@qut.edu.au (A. Goonetilleke).

https://doi.org/10.1016/j.envpol.2017.10.125
0269-7491/

complex environmental systems and processes (Korb and Nicholson, 2010). This modelling approach enables the incorporation of both, quantitative and qualitative data into a model and the quantitative evaluation of the interdependencies among model variables. BNs also enable the visual interpretation of a given problem. As such, BNs have been employed in a diversity of past studies in order to seek solutions to environmental problems such as adaptive management of bird habitats (Howes et al., 2010), catchment-based integrated water resources management (Chan et al., 2010), and assessing the influence of nutrients and climatic conditions on the occurrence of cyanobacterial blooms in water bodies (Rigosi et al., 2015). However, the wide ranging capabilities of BNs have not been exploited to their full potential in the case of stormwater pollution mitigation (Aguilera et al., 2011).

The research study discussed in this paper aimed to characterise the influence of vehicular traffic on the build-up of different PAH species, and to characterise this influence in relation to the type of land use. Accordingly, a BN modelling approach was employed to quantify the interdependencies between vehicular traffic and the build-up of PAHs that can exert significant impacts on human health. The use of BNs enabled the investigation of the influence of traffic on PAH build-up far beyond the typical perspective of traffic as a source of PAHs. The PAH build-up in this paper refers to the concentrations of PAH adsorbed by particles. The outcomes of this investigation are expected to contribute to enhancing urban stormwater quality modelling practices and thereby, to strengthen the design of pollution mitigation strategies.

## 2. Materials and methods

### 2.1. Study area

Three cities located in Northern China, namely, Daqing (DQ), Harbin (HEB) and Jilin (JL) that have different urban characteristics were selected as the study areas. Daqing, where the largest oil field in China is located, produces approximately 40 million tons of oil every year. Jilin has a long history of chemical industries since the 1950s, while Harbin, which is China's third biggest heavy industrial city, hosts food, equipment manufacturing, petro-chemical, metallurgical and electric power industries. The sampling sites for the study were selected such that they are evenly located within the central area of each city and encompass different land uses. Fig. 1 and Figs. S1 and S2 in the Supplementary Information show the locations of the sampling sites. Further details regarding the study areas can be found in Song et al. (2015).

### 2.2. Experimental data

The dust samples were collected within 0.5 m from the kerb or street edges at the selected sites using a clean brush. At least three cycles of back-and-forth sweeping were performed along 2–30 m length of the street to collect one sample at each site within 0.5–1 h time period. Accordingly, 23 samples from DQ site and 21 samples from each of HEB and JL sites were collected.

The dust samples were air-dried for at least 15 days, and then sieved through a screen with 500 μm opening to remove large particulates such as stones and plant debris. This approach has been commonly used in previous studies (eg. Lorenzi et al., 2011; Nguyen et al., 2014; Peng et al., 2011; Wang et al., 2015). The concentrations of the sixteen PAHs identified by the US EPA as priority pollutants (Schoeny and Poirier, 1993; USEPA, 1984) were determined for each dust sample from each site. Detailed information about sampling and laboratory analysis of PAHs can be found in Song et al. (2015).

To determine the daily traffic volume, the traffic at each study site was video-recorded between 7:30am–8:30am and 10am–11am. This was followed by counting the traffic volume during these two periods using the recorded videos. On the same day, the traffic for 24 h was also video-recorded at selected site and the daily traffic volume was estimated using the data obtained from the manual counting of the traffic for the two hourly study periods noted above.

### 2.3. Bayesian network modelling

BNs are a graphical modelling approach based on Bayesian statistical methods. The *directed acyclic graphical (DAG)* structure (i.e. a directed graph without any loops) of BNs integrates a given set of random variables (discrete and/or continuous), which describes the system or process being modelled using probabilistic conditional dependencies (Fig. 2). This is achieved by factorising the global probability distribution of the set of random variables into local probability distributions of individual variables. The factorisation is based on the *Markov Property* of Bayesian Networks, such that each random variable depends only on its immediate parent variables (Scutari, 2009).

BN modelling is a two-step approach. It first learns the structure of the BN using *Structure Learning Algorithms* and then estimates the parameters of the variables commonly based on *Maximum Likelihood Estimates* given the data and the model structure (Ben-Gal, 2007; Scutari, 2009; Uusitalo, 2007). Moreover, the parameters of the discrete and continuous random variables are estimated in the form of conditional probabilities and conditional regression coefficients, respectively. Further, in the case where a continuous variable has discrete parent variables which form different configurations, regression coefficients for the continuous variable are estimated for each configuration.

In this study, a BN was proposed to investigate the interdependencies between vehicular traffic in different types of urban land use and build-up of particle-bound PAHs. Accordingly, the three land use types, namely, residential, industrial and commercial, and traffic volume corresponding to each land use type were identified as the factors that influence PAHs build-up. This was based on the outcomes of the previous investigation by Song et al. (2015) on PAHs adsorbed to street dust in the same study areas as shown in Figs. S3 and S4 in the Supplementary Information.

Accordingly, the input data of the BN model included observed concentrations of particle-bound PAHs, observed volumes of vehicular traffic and types of land uses. The PAH concentrations and traffic volumes were fed into the model as quantitative data, while land use type was fed as qualitative data. This means that the land use data were provided in terms of 'Yes' and 'No' scenarios. For example, for residential sites, the data for variable 'Residential' would be fed into the model as 'Yes', while variables 'Industrial' and 'Commercial' would be fed as 'No'.

The proposed BN model was fitted with observed data using the *bnlearn* R statistical computing package to conduct the predictive analysis. This means that the probability density functions corresponding to a specific BN (Fig. 2) are fitted with observed data, such that the difference between observed and predicted values is minimised (Scutari, 2009, 2016; Team, 2014). Accordingly, parameters for discrete and continuous variables were estimated, and then the estimated BN model was utilised to predict the concentrations of PAHs at each study site.

The BN model facilitated the quantitative evaluation of the influence of traffic on the build-up of the sixteen US EPA priority PAHs and total PAHs. This was done by assessing the type and relative strength of relationships between traffic and PAH build-up as informed by the estimated conditional regression coefficients. Positive regression coefficients would imply that the concentration