Environmental Pollution 234 (2018) 751–761

Contents lists available at ScienceDirect

Environmental Pollution

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

OrganoRelease — A framework for modeling the release of organic chemicals from the use and post-use of consumer products☆

Mengya Tao [1], Dingsheng Li [1], Runsheng Song, Sangwon Suh, Arturo A. Keller*

Bren School of Environmental Science and Management, University of California Santa Barbara, Santa Barbara, CA, 93106, United States

ARTICLE INFO

Article history:
Received 10 May 2017
Received in revised form
15 November 2017
Accepted 16 November 2017

Keywords:
Chemical release
Functional use
Consumer products
Use phase
Post-use phase

ABSTRACT

Chemicals in consumer products have become the focus of recent regulatory developments including California's Safer Consumer Products Act. However, quantifying the amount of chemicals released during the use and post-use phases of consumer products is challenging, limiting the ability to understand their impacts. Here we present a comprehensive framework, OrganoRelease, for estimating the release of organic chemicals from the use and post-use of consumer products given limited information. First, a novel Chemical Functional Use Classifier estimates functional uses based on chemical structure. Second, the quantity of chemicals entering different product streams is estimated based on market share data of the chemical functional uses. Third, chemical releases are estimated based on either chemical product categories or functional uses by using the Specific Environmental Release Categories and EU Technological Guidance Documents. OrganoRelease connects 19 unique functional uses and 14 product categories across 4 data sources and provides multiple pathways for chemical release estimation. Available user information can be incorporated in the framework at various stages. The Chemical Functional Use Classifier achieved an average accuracy above 84% for nine functional uses, which enables the OrganoRelease to provide release estimates for the chemical, mostly using only the molecular structure. The results can be can be used as input for methods estimating environmental fate and exposure.

1. Introduction

Concern over health risks of chemicals in consumer products has been increasing due to human exposure to chemicals released both indoors and outdoors (Trudel et al., 2008, 2011; Zota et al., 2014). Release of chemicals to the general environment from consumer products is also a concern for the health of the ecosystem especially for chemicals that are persistent and induce ecotoxicity, such as perfluorinated compounds, flame retardants, and certain antibiotics (Clarke and Smith, 2011; Ezechias et al., 2014; Janecko et al., 2016; Kunhikrishnan et al., 2015; Ortiz de Garcia et al., 2014; Rosal et al., 2010). As a result, chemicals contained in consumer products have become the focus of recent regulatory developments including the Safer Consumer Products Act of California, the Safe Chemical Act of the U.S., and the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) program of the European Union.

A major challenge to minimizing the human and ecological health risk of chemicals in consumer products is the lack of available release information on the vast majority of chemicals and the economical infeasibility of obtaining relevant information for all chemicals through traditional experimental testing. To determine release estimates of chemicals from consumer products, one must first determine the plausible uses of a given chemical. Recently, Phillips et al. (2017) proposed a methodology to screen out candidate chemical alternatives based on functional use similarities and hazard information by combining quantitative structure-use relationship models and high-throughput toxicity screening. However, to assess the human and ecological health risk chemicals may pose, exposure as well as the fate and transport of the chemicals must be characterized, which requires estimates of the amount of chemical released from different products and applications. Traditional exposure assessments have been done for chemicals in consumer products (Aronson et al., 2007; Dann and Hontela, 2011; Goebel et al., 2012; Kienhuis et al., 2015). However, these methods rely

☆ This paper has been recommended for acceptance by Eddy Y. Zeng.

* Corresponding author.
E-mail addresses: mengya@umail.ucsb.edu (M. Tao), dli@ucsb.edu (D. Li), runsheng@umail.ucsb.edu (R. Song), suh@bren.ucsb.edu (S. Suh), keller@bren.ucsb.edu (A.A. Keller).
[1] Authors with equal contribution.

https://doi.org/10.1016/j.envpol.2017.11.058
0269-7491/

on measuring the concentrations of chemicals in different environmental media, which would be cost prohibitive for evaluating thousands of chemicals in consumer products. Fate and transport models for both indoor and outdoor environments can be used to estimate the concentrations of the chemicals of interest (Garner et al., 2017; Hollander et al., 2016; Liagkouridis et al., 2015; Rosenbaum et al., 2011). However, a vital input for these models is the release information of the chemical(s) to different environmental compartments during and after the use of products that contain the chemical(s).

Holmgren et al. (2012) presented a generic emission model for organic chemicals embedded in solid materials but this model is rather mathematically complex with dozens of parameters for each chemical, the shape of the consumer product, and the room characteristics. A less complicated and parsimonious model was recently developed for the release of volatile organic compounds encapsulated in products (Huang and Jolliet, 2016) but is limited to volatile organic compounds. Furthermore, both models (Holmgren et al., 2012; Huang and Jolliet, 2016) provide release estimates only for indoor air. Various models to estimate the release of different types of nanoparticles to multiple environmental compartments during the life cycle of various products have been published (Gottschalk et al., 2010; Gottschalk and Nowack, 2011; Keller et al., 2013, 2014; Keller and Lazareva, 2013; Sun et al., 2016). However, such studies do not address the release of organic chemicals, which make up the majority of both existing and new chemicals.

Given the current status of research, a systematic methodology for quantitatively estimating the release of organic chemicals contained in a wide range of consumer products is needed as input to models used to characterize potential human and ecosystem exposure to chemicals. Here we present the OrganoRelease framework, a methodology to estimate the distribution (as a fraction of the total amount used) of the release of a chemical to different environmental compartments during the use and post-use phases of consumer products, when product-specific chemical release measurements are lacking. Post-use phase is when chemicals are being transported to wastewater or waste treatment plants after the direct use of consumer products. OrganoRelease was designed to accommodate different levels of data availability for any given organic chemical, with the minimum required input being the chemical structure alone. To achieve this goal, in OrganoRelease we connect the chemical's structure (e.g., topological, physicochemical properties, etc.), possible functional uses which are grouped based on the chemical's primary function (e.g., surfactant, solvent, etc.), potential product categories (e.g., cosmetics, paints, etc.), and release factors (as fractions of the chemical of interest released to indoor or outdoor air, wastewater, soil, and waste during use and post-use phases). OrganoRelease can provide rapid screening-level estimates of the release of chemicals contained in a range of consumer products during the use and post-use phases.

2. Methods

OrganoRelease estimates the distribution (in percent) of the release of a chemical directly released to indoor air, outdoor air, wastewater, soil, and waste, without considering any post-release environmental fate and transport processes. OrganoRelease consists of three components that connect the functional use (defined as the chemical categories grouped based on their primary function), product category, and release factors together: 1) a Chemical Functional Use Classifier that estimates the chemical's functional use if not known; 2) market share data that links the functional uses with different product categories to quantify the mass fraction of chemical(s) entering corresponding product streams; and 3) release factors, based on functional uses and product categories, to estimate the initial release of the chemical(s) in products to the environment. The conceptual framework of OrganoRelease is shown in Fig. 1.

2.1. Chemical Functional Use Classifier

An artificial neural network (ANN) was employed to develop a Chemical Functional Use Classifier ("Classifier" for short) that estimates chemical functional uses such as solvents and surfactants based on the molecular structure of the chemical. The molecular structure information is represented by the molecular descriptors, including constitutional, topological, chemical properties, and many other descriptor blocks (Todeschini and Consonni, 2009). ANNs serve as a nonlinear, universal approximation model to extract the intrinsic knowledge within a large amount of data (Hornik et al., 1989).

Pairs of chemical-functional use data points were collected from Chemical Book (ChemicalBook, 2016) based on chemical functional uses with available market share data. This included nine functional uses: aerosol propellants, antibacterial agents, flame retardants, flavors and fragrances, solvents, surfactants, fungicides, herbicides, and insecticides, which are the major functional uses in consumer products. This is to ensure a seamless connection between estimating product use categories and functional uses. Since the molecular descriptors can only be computed for organic compounds, we removed mixtures, inorganics, salts, and organometallics. After the data curation process, we had a total of 2900 pairs of chemical-functional use to build the Classifier.

Dragon 7 (Dragon 7.0, 2016) was used to generate over 4000 molecular descriptors for each chemical, including constitutional, topological, chemical properties, and many other descriptor blocks. In order to reduce the number of molecular descriptors to the recommended ratio of number of entries to number of variables for quantitative structure-activity relationship (QSAR) models (Dearden et al., 2009), a filter-based feature selection algorithm was performed to remove the molecular descriptors that have a variance lower than 15 or a correlation coefficient higher than 0.65 with other descriptors (Dutta et al., 2007; Gramatica, 2007).

The collected dataset was randomly split into training, validation and test datasets. For each functional use category, 20 chemicals were first randomly selected as test dataset. For the rest of the data points within each functional use, 85% were randomly chosen as training set and 15% as validation set. The ANN models were built based on the training set, and the validation set was used to select the best fit model. The test set was applied to evaluate the performance of the final model. The Classifier was built in Python, using the Scikit-learn package (Pedregosa et al., 2011) for dataset random selection and final model evaluation. TensorFlow (Abadi et al., 2016) was implemented to build neural network models. A combination of grid search and random search methods were performed for hyperparameter optimization to optimize the set of parameters in the learning algorithm as well as avoid overfitting (Bergstra et al., 2013). Fig. 2 illustrates the conceptual diagram of the Classifier building process.

The performance of the Classifier is measured by its precision, recall, and F1 score, which are defined in Equations (1)–(3). Precision indicates the ability of the Classifier to not label a negative sample as positive, while recall implies the ability of the Classifier to find all positive samples (Powers, 2011). F1 score represents a weighted harmonic mean of precision and recall. All three metrics range from 0 to 1; values closer to 1 indicate better model performance.