Prediction of the removal efficiency of emerging organic contaminants in constructed wetlands based on their physicochemical properties

Huma Ilyas, Ilyas Masih, Eric D van Hullebusch

Prediction of the removal efficiency of emerging organic contaminants in constructed wetlands based on their physicochemical properties

Huma Ilyas[1,3*], Ilyas Masih[2,3], Eric D. van Hullebusch[1]

[1] Université de Paris, Institut de physique du globe de Paris, CNRS, F-75005 Paris, France

[2] IHE Delft, Institute for Water Education, 2611 AX Delft, The Netherlands

[3] Water Treatment and Management Consultancy, B.V., 2289 ED Rijswijk, The Netherlands

* Corresponding author email: hi.wtmconsult@gmail.com

Abstract

This study investigates the prediction of the removal efficiency of emerging organic contaminants (EOCs) (pharmaceuticals-PhCs, personal care products-PCPs, and steroidal hormones-SHs) in constructed wetlands based on their physicochemical properties (e.g., molecular weight-MW, octanol-water partition coefficient-Log Kow, soil organic carbon sorption coefficient-Log Koc, octanol-water distribution coefficient-Log Dow, and dissociation constant-pKa). The predictive models are formed based on statistical analysis underpinned by principle component, correlation, and regression analyses of a global data set compiled from peer-reviewed publications. The results show that the physicochemical properties of EOCs emerged as good predictors of their removal efficiency. Log Koc, Log Dow, and Log Kow are the most significant predictors, and combination with MW and/or pKa often improved the reliability of the predictions. The best performing model for PhCs was composed of MW, Log Dow, and Log Koc (coefficient of determination-R^2: 0.601; probability value-$p < 0.05$; root mean square error-RMSE: training set: 11%; test set: 27%). Log Kow and Log Koc for PCPs (R^2: 0.644; $p < 0.1$; RMSE: training set: 14%; test set: 14%), and a combination of MW, Log Kow, and pKa for SHs (R^2: 0.941; $p < 0.1$; RMSE: training

set: 3%; test set: 15%) formed the plausible models for predicting the removal efficiency. Similarly, reasonably good combined models could be formed in the case of PhCs and SHs or PCPs and SHs, although their individual models were comparatively better. A novel decision support tool, named as REOCW-PCP, was developed to readily estimate the removal efficiency of EOCs, and facilitate the decision-making process.

Key words: Constructed wetlands; Personal care products; Pharmaceuticals; Physicochemical properties; Predictive models; Steroidal hormones.

1. Introduction

Pharmaceuticals (PhCs), personal care products (PCPs), and steroidal hormones (SHs) are among the emerging organic contaminants (EOCs) that are discharged to water resources and environment through various sources such as domestic wastewater, industrial wastewater, landfill leachate, runoff from concentrated animal feeding operations and aquaculture, and effluent discharge from wastewater treatment plants (e.g., Luo et al., 2014; Barbosa et al., 2016; Yi et al., 2017; Yin et al., 2019). Although EOCs are often present in relatively small concentrations (i.e., ng L^{-1} to µg L^{-1}) in water resources, their occurrence as individual compounds, transformation products, and multitude of compounds can still cause negative impacts on human health as well as aquatic and terrestrial life (e.g., Vymazal et al., 2015; Vystavna et al., 2017; Tran et al., 2018). It is well established that the higher concentration of EOCs compared with their potential no effect concentration could pose severe risk to human health, since many of these EOCs are considered as prospective endocrine disruptors (Caliman and Gavrilescu, 2009; Töre et al., 2012; Gogoi et al., 2018).

Constructed wetlands (CWs) are environmentally friendly, cost-effective, and nature-based treatment technologies that have been extensively investigated and practically used for the removal of EOCs from the wastewater (e.g., Zhang et al., 2014; Verlicchi et al., 2015; Gorito

et al., 2017; Vo et al., 2018). To date, a large number of individual case studies have been published in peer-reviewed journals related to the removal of EOCs by CWs. A review study by Ilyas et al. (2020) indicated that investigation of the impact of physicochemical properties on the removal efficiency has been an important research topic among other thematic areas such as evaluation of the performance of different types of CWs, role of design and operational factors on the treatment process, and environmental risk assessment. There is an increasing evidence that physicochemical properties of EOCs (molecular weight-MW, water solubility-WS, octanol-water partition coefficient-Log Kow, soil organic carbon sorption coefficient-Log Koc, octanol-water distribution coefficient-Log Dow, Henry's law constant-HC, dissociation constant-pKa, and charge at pH 7) play a very important role during the treatment process (Supplementary materials 1: Tables S1), and considerably influencing the possible removal mechanisms such as biodegradation, adsorption/sorption, and plant uptake (e.g., Zhang et al., 2012a, 2012b, 2018; Verlicchi et al., 2013; Vystavna et al., 2017; Ilyas et al., 2020; Ilyas and van Hullebusch, 2019, 2020a, 2020b, 2020c, 2020d).

Few experimental studies have indicated an inverse relationship between Log Dow and the removal efficiency of PhCs (e.g., Lee et al., 2011; Zhang et al., 2012b, 2018). However, the resulting correlations were not very strong (e.g., coefficient of determination-R^2: ~ 0.3) by the study of Zhang et al. (2012b) on horizontal subsurface flow constructed wetland (HFCW); but the relationship was statistically significant (e.g., probability value-p < 0.1). Zhang et al. (2018) reported a strong inverse relationship between the removal efficiency and Log Dow (R^2: 0.82-0.86; p < 0.05) in two FWSCWs. These studies indicated some potential of using Log Dow as a predictor of the removal efficiency of PhCs, and highlighted the important role of hydrophobicity in the removal mechanisms of PhCs in CWs. However, the results should be interpreted with caution because these are based on only few individual studies involving a limited number of PhCs (e.g., < 10). Moreover, the proposed regression models are only

74 based on the training data sets. A validation step on an independent test set was not carried
75 out, which limits the reliability and wider application of the proposed models.

76 On the other hand, some studies suggested a non-significant correlation between Log Kow
77 and the removal efficiency of PhCs (Park et al., 2009; Lee et al., 2011; Breitholtz et al., 2012;
78 Zhang et al., 2012a, 2012b). For example, Breitholtz et al. (2012) reported a non-significant
79 correlation ($p > 0.35$) between the removal efficiency of PhCs and Log Kow in FWSCW.
80 Verlicchi et al. (2013) stated the similar results for the relationship between the removal
81 efficiency of PhCs and Log Kow in HFCW. These studies indicated a non-substantial role of
82 hydrophobicity in the removal processes of PhCs in CWs. However, it was not fully clear
83 from the available studies why Log Kow does not yield a significant correlation with the
84 removal efficiencies like Log Dow, despite the fact that both are closely related and Log Dow
85 could be estimated using Log Kow (e.g., Carballa et al., 2008). Although Zhang et al. (2012b)
86 suggested that the removal efficiency of the selected PhCs can be predicted adequately in
87 HFCW based on their Log Dow, Verlicchi et al. (2013) suggested that Log Kow alone is not a
88 useful predictor for the behavior of PhCs in HFCW. These results indicate the potential role
89 of multiple physicochemical properties in their removal processes. However, the previous
90 research did not investigate the possibility of predicting the removal efficiency by combining
91 two or more physicochemical properties, i.e., multiple linear regression analysis to develop
92 predictive models for the removal efficiency with two or more physicochemical properties
93 was not conducted by the above-mentioned studies.

94 While most of the studies attempted to predict the removal efficiency of PhCs based on only
95 one physicochemical property (Log Kow and/or Log Dow), a study by Matamoros et al.
96 (2017) performed a correlation analysis between the removal efficiency of 16 EOCs including
97 PhCs, PCPs, and industrial chemicals and their corresponding values of Log Kow, HC, WS,
98 and MW. This study revealed a significant positive correlation of the removal efficiency with

Log Kow and MW. Nevertheless, the authors suggested a cautious interpretation of these results, and stated that differences in their findings compared to earlier studies could be due to the variability in the studied EOCs, weather conditions, sample size, and CW systems. Vystavna et al. (2017) conducted a principle component analysis (PCA) to investigate the role of physicochemical properties (WS, MW, and Log Kow) in the removal efficiency of 12 EOCs including PhCs, PCPs, and SHs in a hybrid constructed wetland (HCW). The results of PCA suggested a non-significant relationship between the physicochemical properties of the studied EOCs and their removal efficiency. The authors recommended a further analysis for the confirmation of their findings.

The CW models can provide a very useful information for scientific understanding of the processes and aiding in the design of the CWs (Meyer et al., 2015). However, existing models of CWs, both process-based or black-box types, only focus on the removal of conventional parameters such as nutrients, biochemical oxygen demand, and chemical oxygen demand (e.g., Kumar and Zhao, 2011; Meyer et al., 2015; Yuan et al., 2020). In general, there is a need to include EOCs in the further development of CW models (Yuan et al., 2020). Furthermore, there is a need to carry out a comprehensive statistical analysis to develop predictive models for the removal efficiency of EOCs (PhCs, PCPs, and SHs) based on their physicochemical properties. Further research is required on investigating the combined impact of several physicochemical properties (MW, WS, Log Kow, Log Koc, Log Dow, HC, and pKa) to predict the removal prospects of EOCs in CWs, while giving attention to the prediction of PCPs and SHs along with PhCs. The progress on predictive models will also advance scientific understanding and contribute to improve the decision-making process. For example, engineers and policy makers could readily use these models to obtain a first-hand information on the expected removal efficiency of a certain PhC, PCP or SH in CWs. Such information can be very valuable in data limited situations when no or a very limited

information on the removal efficiency of certain EOCs is available based on the field research, which could be quite costly. Moreover, the new models can help in the further development of modelling and decision support tools (DSTs) on EOCs.

Therefore, the main objective of this study is to comprehensively analyze the possibility of developing reliable predictive models for the removal efficiency of EOCs in CWs based on their physicochemical properties. The specific objectives are: (1) to explore the possibility of developing statistical models for the prediction of removal efficiencies of individual group of EOCs (PhCs, PCPs, and SHs) in CWs based on their physicochemical properties; (2) to explore the possibility of formulating a generic predictive model by combining the removal efficiencies of PhCs, PCPs, and SHs in CWs based on their physicochemical properties; (3) to examine the impact of physicochemical properties of EOCs on their removal mechanisms in CWs; (4) to study the inherent uncertainties involved in the development of the above-mentioned predictive models; and (5) to explore the possibilities of application of the plausible predictive models.

2. Methodology

2.1. Data

The flowchart of the research methodology is presented in Figure 1 and discussed below. Predictive models were formulated using data of widely studied EOCs (PhCs: 33; PCPs: 15; and SHs: 11) in CWs. The data was compiled from the peer-reviewed published sources in our previous work (Ilyas et al., 2020; Ilyas and van Hullebusch, 2020c, 2020d), and formed the basis of this modelling study. The removal efficiency is used as a dependent variable to be predicted. The independent variables were physicochemical properties: HC, WS, MW, Log Kow, Log Dow, Log Koc, and pKa. The data of these variables was collected from various sources: quantitative structure activity relationship-QSAR Toolbox (version 4.3.1), journal

papers, reports, and websites. Despite some differences in the data values, these sources were considered very useful and reliable. For example, the QSAR Toolbox contains data from several reliable sources, and was very helpful in collecting the data of most of the variables. However, data of Log Dow was not available for all the EOCs, and the missing values were filled by developing regression equations of Log Dow with Log Koc (for PhCs) and Log Kow (for PCPs). In the case of SHs, Log Kow values were used in place of Log Dow because the studied compounds were neutral (e.g., Carballa et al., 2008), and regression analysis was not possible due to a very limited number of data points for Log Dow. The final data sets used in this study are given in Supplementary materials 2: Tables S2-S4.

2.2. Statistical methods

First, PCA was performed in R software (version 3.6.2). The PCA was mainly aimed at understanding the role of different physicochemical properties in the removal processes. Moreover, the PCA helped to understand the key variables explaining most of the variance in the data set, and thus could be used in the model building process (Supplementary materials 3 and 4: Figures S1-S5 and Tables S5-S12). Second, a correlation of the removal efficiency of EOCs and their physicochemical properties was estimated to select the most promising variables to be used in the predictive models. Microsoft Excel software (Analysis ToolPak) was used to estimate the Pearson product moment correlation (Supplementary materials 5: Tables S13-S20).

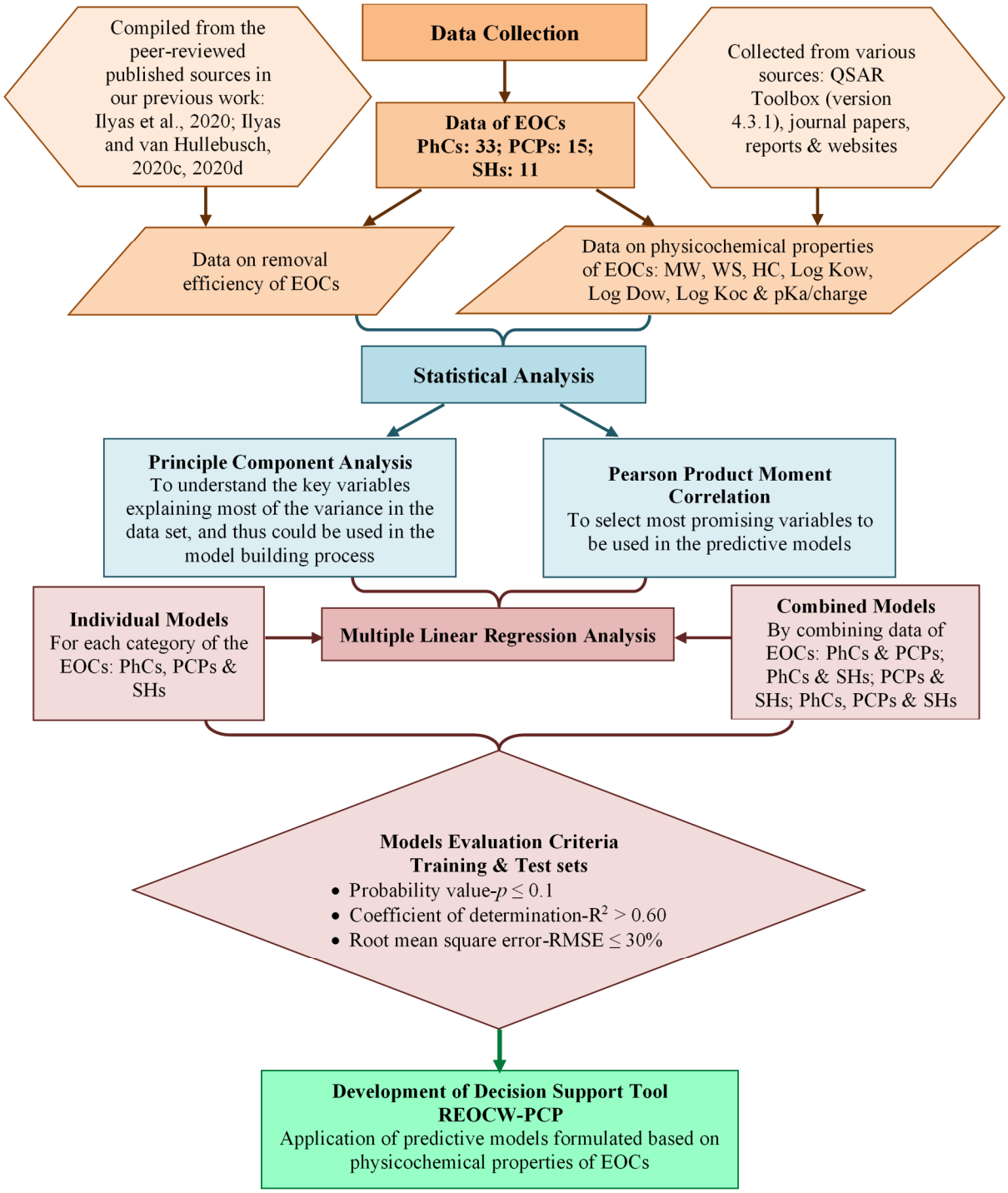


Figure 1. Flowchart of the research methodology.

Note: Emerging organic contaminants (EOCs); Pharmaceuticals (PhCs); Personal care products (PCPs); Steroidal hormones (SHs); Henry's Law Constant (HC); Water solubility (WS); Molecular weight (MW); Octanol-water partition coefficient (Log Kow); Soil organic carbon sorption coefficient (Log Koc); Octanol-water distribution coefficient (Log Dow); Dissociation constant (pKa).

Then, multiple linear regression analysis was carried out using Microsoft Excel software (Analysis ToolPak) (Supplementary materials 6: Tables S21-S27). The available EOCs' data sets were divided into the training and test sets using a purposive sampling approach based on the number of data points in the case of removal efficiency of EOCs (Tables S2-S4). For example, the data of PhCs with > 15 data points were used as one set in the model formulation processes. Other sets included PhCs with > 20 and ≥ 25 data points, respectively. It was assumed that more data points may reduce the uncertainty in the observed removal efficiency and, therefore, may results in the development of better models (Supplementary materials 7: Tables S28-S32). In the calibration process, various combinations of independent variables were used to formulate multiple linear regressions. For example, the removal efficiency was estimated using Log Koc and Log Dow in one model. Then, other variables (e.g., MW and pKa) were added into the model. The regressions models with $p \leq 0.1$ and R^2: > 0.60 were considered as satisfactory, and were selected for validation on an independent data (test set) (Tables S2-S4). The difference between the observed and predicted removal efficiency was used as an indicator to assess the performance of the regression models. Since these differences were both positive and negative, it was not suitable to estimate the overall mean difference because of the cancelling effect. Therefore, model performance was also assessed using root mean square error (RMSE), which was estimated from the observed and predicted removal efficiency. First, the regression models were developed separately for each category of the EOCs. For example, the removal efficiency of PhCs was predicted using the physicochemical properties of the examined PhCs. Similar approach was followed for PCPs and SHs. Then, an attempt was made to form the models by combining data of different EOCs. For example, the data of PhCs and PCPs was combined, and used to develop a predictive model that could represent both categories of these EOCs. In a similar way, combined models were developed for PhCs and SHs, PCPs and SHs, and PhCs, PCPs, and

SHs. The multiple linear regression approach was used for developing the individual as well as generic models.

A novel DST, named as REOCW-PCP was developed for the potential application of the predictive models formulated for predicting the removal efficiency of EOCs based on their physicochemical properties. REOCW-PCP was developed using Microsoft Excel 2016, and is provided as a supplementary material (Excel file: REOCW-PCP-secure, along with the user manual).

3. Results and Discussion

3.1. Predictive models for PhCs

The first three principle components (PCs) with eigenvalue > 1.0 explained 71% of the variance in the data of 33 widely studied PhCs (Figure 2 and Table 1). The first two PCs could explain 56% of the variance in the data (PC1: 39%; PC2: 17%). In PC1, the high positive loadings were observed for Log Koc, Log Kow, and Log Dow, and these variables cluster together (Figure 3). In PC2, pKa and MW indicated high positive loadings and cluster together (Figure 3 and Table 1). The removal efficiency indicated negative loadings in all three PCs. Although the variance explained by the first three PCs is not very high, the results show reasonably good prospects of establishing a relationship between the removal efficiency and physicochemical properties of PhCs. To further test the reliability of the PCA results, we also conducted PCA based on various combinations of the studied PhCs, i.e., samples with number of data points > 15, > 20, and ≥ 25. In general, the PCA results gave similar insights as obtained using 33 PhCs (Figures S1-S3 and Tables S5-S8). This reveals that processes behind the resulting models will follow a coherent pattern in all the combinations of selected PhCs. For example, Log Koc is likely to emerge as an important predictor in all the cases, and

is likely to show a significant negative correlation with the removal efficiency of PhCs (Figure 4 and Tables S13-S16).

Table 1. Principle component analysis (PCA) results based on widely studied 33 PhCs, 15 PCPs, and 11 SHs.

	PhCs			PCPs		SHs	
Description	**PC1**	**PC2**	**PC3**	**PC1**	**PC2**	**PC1**	**PC2**
Eigenvalue	1.758	1.162	1.120	2.109	1.236	1.971	1.574
Proportion of Variance	0.387	0.169	0.157	0.556	0.191	0.486	0.310
Cumulative Proportion	0.387	0.555	0.712	0.556	0.747	0.486	0.795
Variables and loadings							
HC (atm m^3 mol^{-1})	0.151	-0.639	-0.036	0.196	-0.558	0.139	0.415
WS (mg L^{-1})	-0.373	-0.062	0.355	-0.344	-0.066	-0.351	-0.395
MW (g mol^{-1})	0.042	0.469	-0.660	0.246	0.486	-0.364	0.247
Log Kow	0.498	-0.094	-0.259	0.466	-0.076	0.470	-0.071
Log Koc	0.491	-0.154	0.240	0.458	0.002	0.467	0.216
Log Dow	0.479	0.099	0.136	0.461	-0.093	0.473	-0.104
pKa	0.071	0.499	0.491	0.184	0.623	-0.014	0.574
RE (%)	-0.336	-0.277	-0.230	0.330	-0.212	-0.249	0.468

Note: Pharmaceuticals (PhCs); Personal care products (PCPs); Steroidal hormones (SHs); Principle component (PC); Henry's Law Constant (HC); Water solubility (WS); Molecular weight (MW); Octanol-water partition coefficient (Log Kow); Soil organic carbon sorption coefficient (Log Koc); Octanol-water distribution coefficient (Log Dow); Dissociation constant (pKa); Removal efficiency (RE).

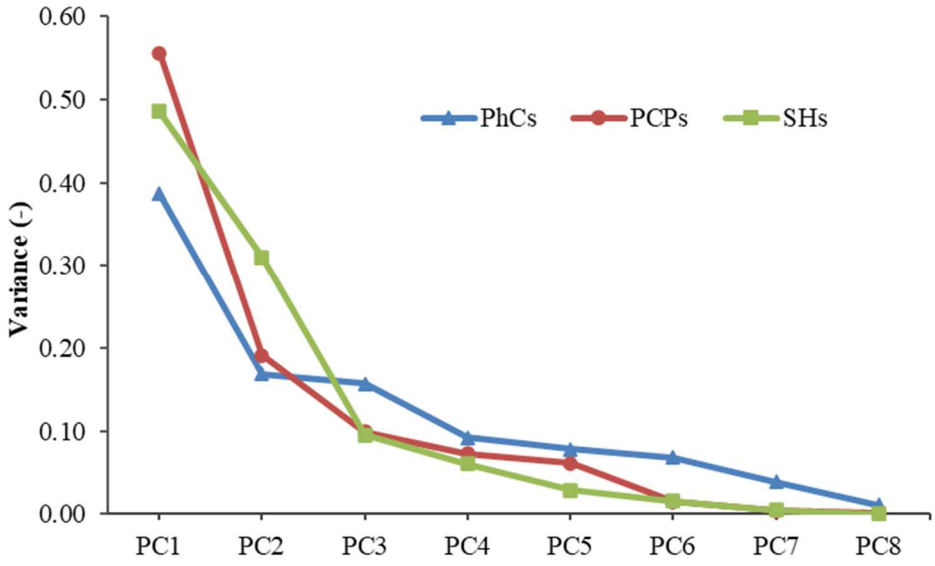


Figure 2. Scree plot indicated the variance explained by each principle component (PC) for the data of widely studied 33 PhCs; 15 PCPs; and 11 SHs.

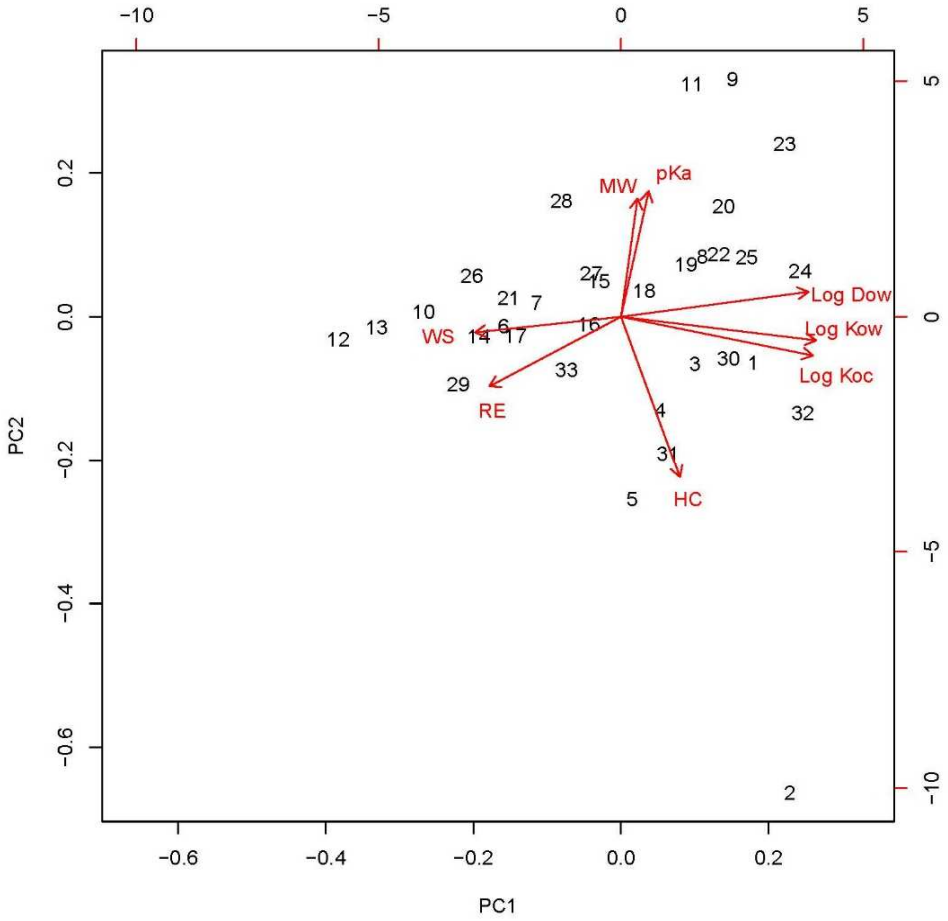


Figure 3. Byplot of PC1 and PC2 based on principle component analysis (PCA) of 33 widely studied PhCs.

Note: Removal efficiency (RE); Diclofenac (1); Ibuprofen (2); Ketoprofen (3); Naproxen (4); Salicylic acid (5); Acetaminophen (6); Codeine (7); Tramadol (8); Clarithromycin (9); Doxycycline (10); Erythromycin (11); Ofloxacin (12); Oxytetracycline (13); Sulfadiazine (14); Sulfamethazine (15); Sulfamethoxazole (16); Sulfapyridine (17); Trimethoprim (18); Monensin (19); Fexofenadine (20); Caffeine (21); Diltiazem (22); Carbamazepine (23); Mirtazapin (24); Venlafaxine (25); Atenolol (26); Metoprolol (27); Sotalol (28); Ranitidine (29); Bezafibrate (30); Clofibric acid (31); Gemfibrozil (32); and Furosemide (33).

The correlation and regression outcomes were quite consistent with the PCA results (Tables S5-S8, S13-S16, and S28). The best models were formed by combining Log Koc, Log Dow, Log Kow, MW, and pKa, mostly with three or four variables together in one model. The models with one or two variables were not acceptable (R^2: < 0.50; p > 0.1). This finding is consistent with Verlicchi et al. (2013) who also suggested that Log Kow alone is not a useful

predictor for the behavior of PhCs in CWs. On the other hand, these results are in contrast with the study of Zhang et al. (2018) which indicated the possibility of good prediction of the removal efficiency only with Log Dow based on high negative correlation between the two variables. Nevertheless, our results suggested that Log Dow is an important predictor when used along with other physicochemical properties.

Furthermore, a negative correlation of MW, Log Dow, Low Kow, and Log Koc with the removal efficiencies of PhCs suggested that the PhCs with high values of any of these physicochemical properties might lead to a low removal efficiency in CWs (Tables S13-S16). For example, this might be the reason of the low removal efficiency in the case of diclofenac, erythromycin, clarithromycin, fexofenadine, mirtazapin, carbamazepine, bezafibrate, and gemfibrozil in CWs (Table S2). In general, the PhCs which are readily biodegradable shows much better removal efficiencies in CWs (e.g., Hijosa-Valsero et al., 2010; Reyes-Contreras et al. 2012; Chen et al., 2016). In most of the cases, highly biodegradable PhCs have low values of Log Koc, Log Dow, Log Kow, and MW, which explains good prospects of their removal by CWs as substantiated by the statistical analysis conducted in this study.

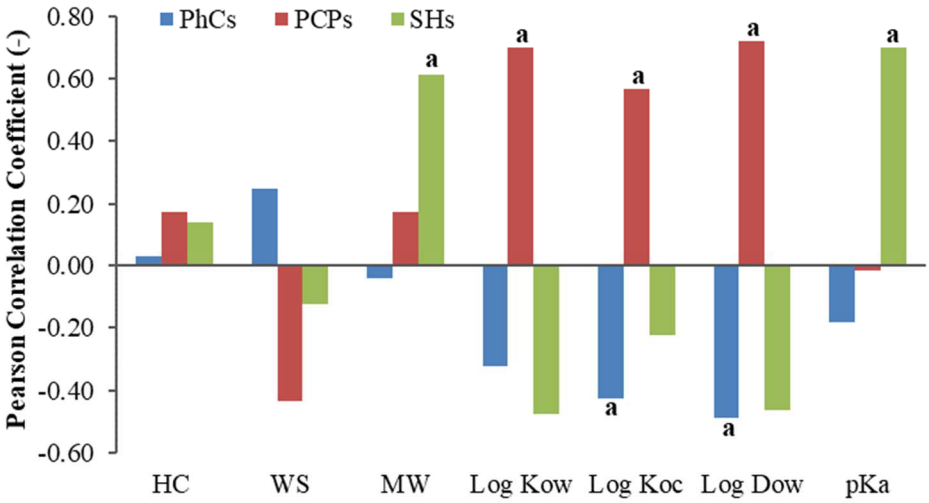


Figure 4. Pearson correlation statistics among the studied physicochemical properties and the removal efficiency of widely studied 33 PhCs, 15 PCPs, and 11 SHs.

Note: **'a'** shows a significant correlation between the parameters at 95% confidence level. The number of data points per pair was 33, 15, and 11 in the case of PhCs, PCPs, and SHs, respectively.

For acceptable models with R^2: > 0.60 and p < 0.05, the RMSEs in training set were < 15% in all the cases (Table S28). In general, the difference in the observed and predicted removal efficiency in the test set was quite good for most of the PhCs. For example, the difference in the observed and predicted removal efficiency was < 20% in the case of 13 out of 18 PhCs (72% of the cases) for the model formed based on MW, Log Dow, and Log Koc (Table S21). The differences were even < 10% for half of the cases indicating very high level of accuracy in predication in the case of venlafaxine, ofloxacin, codeine, mirtazapin, bezafibrate, oxytetracycline, tramadol, and atenolol (Figure 5). However, three PhCs (monensin, sotalol, and diltiazem) were not well predicted, as the difference in the removal efficiency was quite high (ranging from 43% to 76%).

The performance of the plausible models is quite similar, as the RMSE values were around 30% with small differences among different models (e.g., < 5%) (Table S28). This implies that any of these models could be used for predictions, depending upon the availability of data on physicochemical properties. However, acceptable models were further screened to facilitate the selection of the best model for application. The model given in Equation 1 is selected as the first choice, which is based on MW, Log Dow, and Log Koc (Tables S21 and S28). This model was formulated using a training set of 15 PhCs with ≥ 25 data points, and it was validated on an independent test set of 18 PhCs (Figure 5 and Table S2). The predicted removal efficiency was quite good for most of the PhCs, and the overall RMSE of 27% was the lowest compared with other models (Table S28).

$$RE = 108.667 - 0.084\,MW - 4.126\,Log\ Dow - 18.081\,Log\ Koc \quad (1)$$

Where, RE is removal efficiency in %; *MW* is molecular weight in g mol^{-1}; *Log Dow* is octanol- water distribution coefficient; and *Log Koc* is soil organic carbon sorption coefficient.

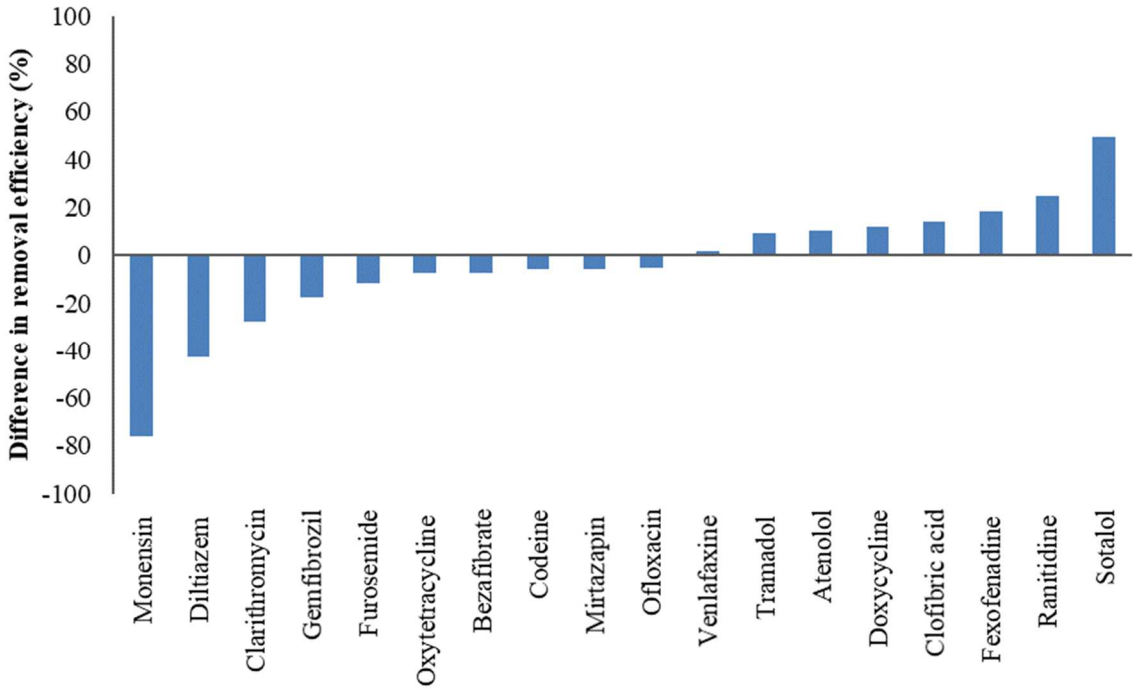


Figure 5. Difference between the observed and predicted removal efficiency based on the best model for PhCs with MW, Log Dow, and Log Koc.

3.2. Predictive models for PCPs

The first two PCs had eigenvalue > 1.0 and together could explain 75% of the variance in PCPs data (Figure 2 and Table 1). More details on PCA results with other combination of the studied PCPs are given in supplementary materials (Figure S4 and Tables S9 & S10). The first two PCs explained more variance in the case of PCPs compared with PhCs (75% VS 56%) (Table 1), which illustrates better prospects of predictions in the case of PCPs compared with PhCs. The PC1 alone explained 56% of the variance in the PCPs data, with the highest positive loadings of Log Kow, Log Dow, and Log Koc, whereas, pKa and MW had highest positive loadings in PC2 (Figure 6 and Table 1). These loading results were similar to those

obtained for PhCs. However, the removal efficiency showed positive loading in the case of PCPs compared with negative loadings for PhCs (Table 1). This could be due to the differences between PhCs and PCPs in dominant removal mechanisms and the contribution of respective physicochemical properties to the removal processes. For instance, adsorption to the substrate and/or sorption onto organic surfaces is the dominant removal mechanism in most of the widely studied PCPs (e.g., Vymazal et al., 2017; Wang et al., 2019; Ilyas and van Hullebusch, 2020c). This explains a significant positive correlation of Log Kow, Log Dow, and Log Koc with the removal efficiency of PCPs (Figure 4 and Tables S17 & S18).

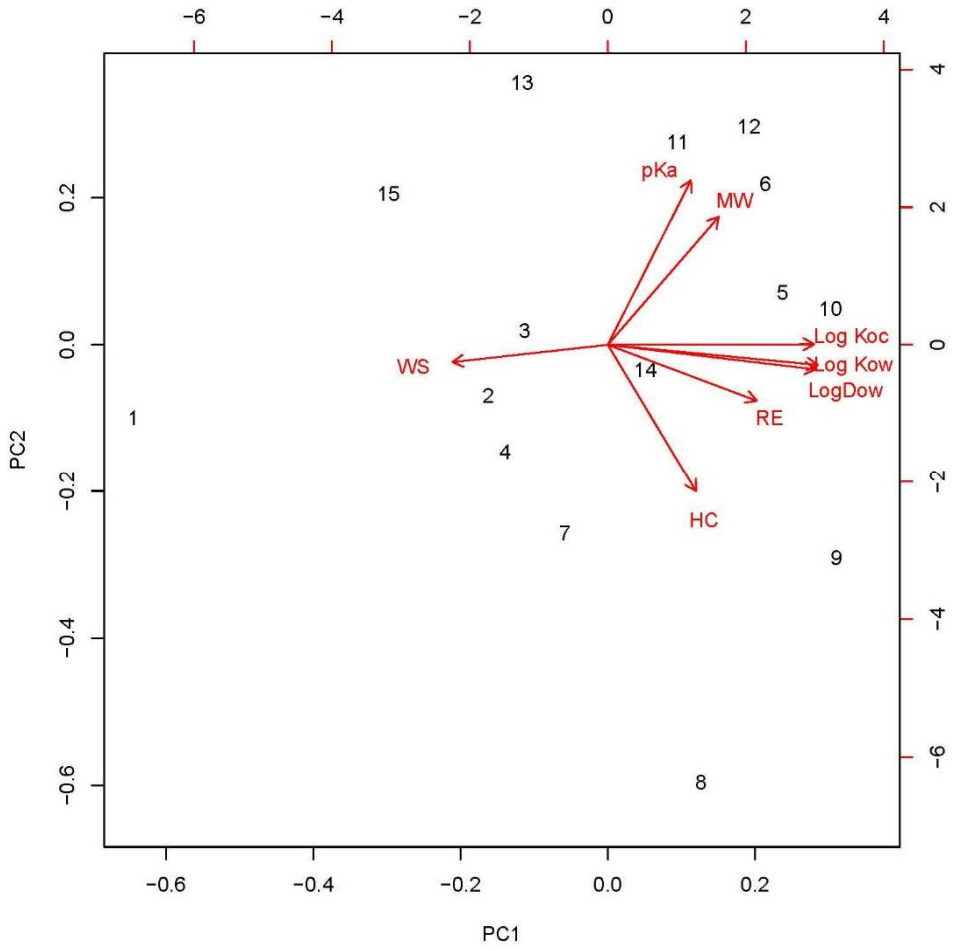


Figure 6. Byplot of PC1 and PC2 based on PCA of 15 widely studied PCPs.

Note: Acesulfame (1); Methylparaben (2); Propylparaben (3); N,N-diethyl-meta-toluamide (4); Triclosan (5); Triclocarban (6); Methyl dihydro-jasmonate (7); Cashmeran (8); Galaxolide (9); Tonalide (10); Tributyl

phosphate (11); Triphenyl phosphate (12); Tris (2-chloroethyl) phosphate (13); Oxybenzone (14); and Sulisobenzone (15).

The removal efficiency indicated a significant positive correlation with Log Dow, Log Koc, and Log Kow, which was consistent with the PCA results (Figure 4 and Table 1). The multiple linear regression models were formed using data of the training set of eight PCPs with > 5 data points and was tested to predict the removal efficiency of seven PCPs (Table S3). The RMSE values of the two promising models ($R^2 > 0.60$; $p \leq 0.1$) for the test set were 14% and 21% (Table S29). The model given in Equation 2 is selected as the best performing model, which was based on two variables: Log Koc and Log Kow (Table S22), and resulted in 14% RMSE in the test set.

$$RE = 66.292 + 21.880\ Log\ Kow - 30.505\ Log\ Koc \quad (2)$$

Where, *RE* is removal efficiency in %; *Log Kow* is octanol-water partition coefficient; and *Log Koc* is soil organic carbon sorption coefficient.

The addition of MW along with Log Kow and pKa resulted in the second-best model (Table S23) with RMSE of 21% in the case of test set. The model is presented in Equation 3.

$$RE = 49.970 - 0.056\ MW + 8.710\ Log\ Kow - 1.605\ pKa \quad (3)$$

Where, *RE* is removal efficiency in %; *MW* is molecular weight in g mol^{-1}; *Log Kow* is octanol-water partition coefficient; and *pKa* is dissociation constant.

In most of the cases, the difference in the observed and predicted removal efficiency was small in both models, as illustrated for the best model by Figure 7. The difference in the observed and predicted removal efficiency was < 10% in the case of five out of seven tested PCPs (cashmeran, triclocarban, triphenyl phosphate, N,N-diethyl-meta-toluamide, and tributyl phosphate). Furthermore, a significant positive correlation of Log Kow, Log Koc, and Log

Dow with the removal efficiencies of PCPs suggested that the PCPs with high values of any of these physicochemical properties might lead to a high removal efficiency in CWs (Tables S17 and S18). This might be the reason of the moderate to high removal efficiency of most of the studied PCPs with the exception of tributyl phosphate and propylparaben in CWs (Table S3).

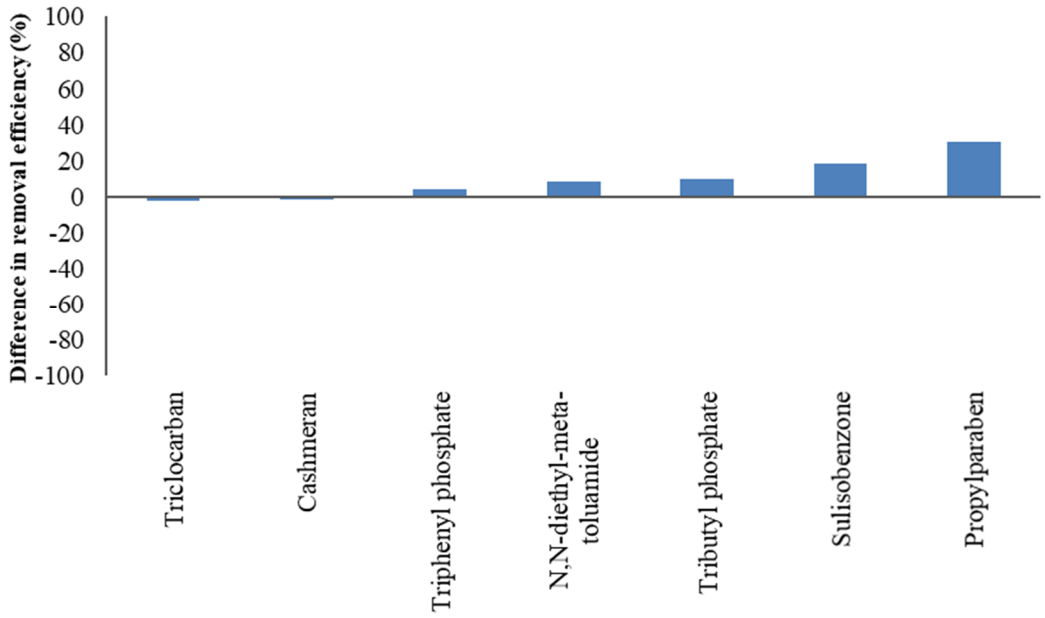


Figure 7. Difference between the observed and predicted removal efficiency based on the best model for PCPs with Log Kow and Log Koc.

3.3. Predictive models for SHs

The first two PCs had eigenvalue > 1.0 and could explain 80% of the variance in the SHs data (PC1: 49%; PC2: 31%) (Figure 2 and Table 1). Like PhCs and PCPs, Log Dow, Log Kow, and Log Koc showed high positive loadings in PC1. For PC2, high positive loadings were shown by pKa, RE, and HC (Figure 8 and Table 1). More detail on PCA analysis is given in supplementary materials (Figure S5 and Tables S11 & S12). The correlation and regression results were consistent with the PCA results (Figure 4 and Tables S19 & S20), indicating the important role of these physicochemical properties in estimating the removal efficiencies of the SHs in CWs.

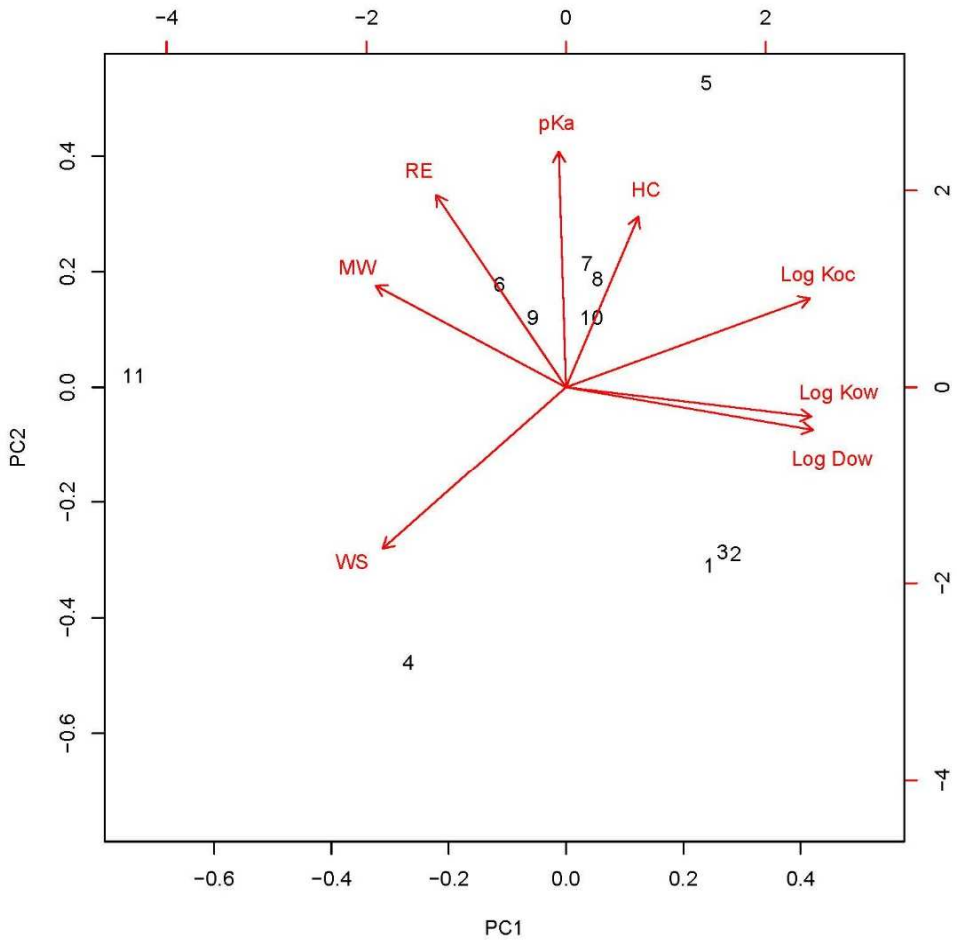


Figure 8. Byplot of PC1 and PC2 based on PCA of 11 widely studied SHs.

Note: 17α-ethinylestradiol (1); 17ß-estradiol (2); Estrone (3); Estriol (4); Progesterone (5); Norethisterone (6); Levonorgestrel (7); Androstenedione (8); Boldenone (9); Testosterone (10); and Prednisone (11).

The multiple linear regression models were formed using data of the training set of six SHs with > 10 data points and was tested to predict the removal efficiency of five SHs (Table S4). The resulting predictive models were very good (R^2: ≥ 0.90; $p \leq 0.1$), and are given in Table S30. The best model was formed by combining MW, Log Kow, and pKa (Table S24), and yielded quite good performance for the training set (RMSE: 3%) and the test set (RMSE: 15%) (Equation 4). Three out of five tested SHs (boldenone, norethisterone, and progesterone) had very less difference in the observed and predicted removal efficiency (< 10%) (Figure 9). The difference in the case of other two SHs (prednisone and androstenedione) was17% and 27%, respectively.

$$RE = 58.032 - 0.035\,MW - 7.244\,Log\,Kow + 3.863\,pKa \quad (4)$$

Where, *RE* is removal efficiency in %; *MW* is molecular weight in g mol^{-1}; *Log Kow* is octanol-water partition coefficient; and *pKa* is dissociation constant.

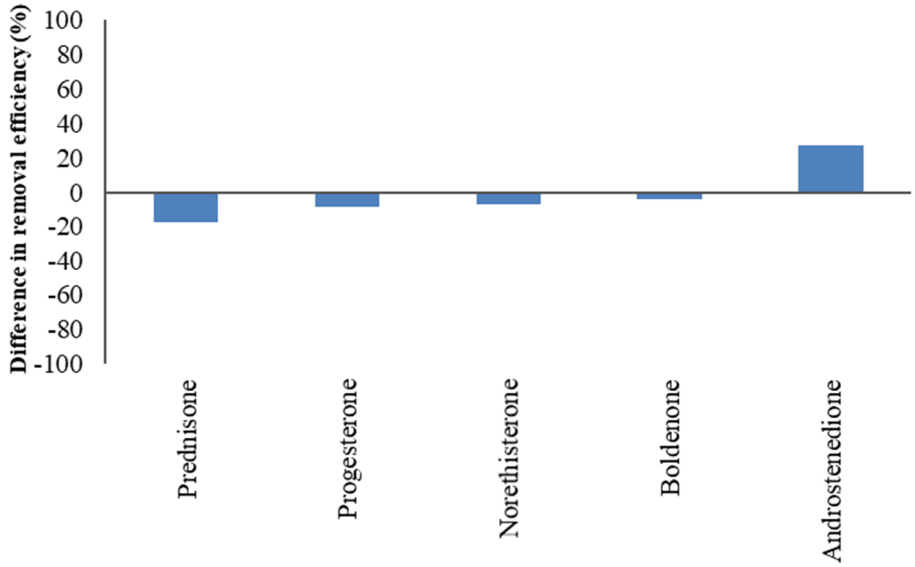


Figure 9. Difference between the observed and predicted removal efficiency based on the best model for SHs with MW, Log Kow, and pKa.

3.4. Predictive models for EOCs combining PhCs, PCPs, and SHs

3.4.1. Predictive models combining PhCs and PCPs

PCA results showed that removal efficiency has negative loadings for PhCs and does not cluster with Log Kow, Log Koc, and Log Dow, which have high positive loadings in PC1 (Figure 3 and Table 1). In contrast, the removal efficiency cluster with Log Kow, Log Koc, and Log Dow in the case of PCPs (Figure 6 and Table 1), and these variables indicated positive loadings in PC1. Due to these differences, the correlation and multiple linear regression models were not acceptable for the training set. For example, R^2 was very low (< 0.25), and the regression models were not statistically significant (p > 0.1). Thus, these models were not suitable for the validation step.

383 This lack of adequate predictive power of the combined model of PhCs and PCPs could be
384 attributed to the distinct differences in the physicochemical properties of these compounds as
385 well as differences in the dominant removal mechanisms in the CWs. For example,
386 biodegradation (aerobic and/or anaerobic) is the most dominant removal mechanism for PhCs
387 removal (e.g., Choi et al., 2016; Nivala et al., 2019; Ilyas et al., 2020). On the other hand,
388 PCPs have adsorption to the substrate and sorption onto organic surfaces as their major
389 removal mechanisms (e.g., Carranza-Diaz et al., 2014; Matamoros et al., 2016; Ilyas and van
390 Hullebusch, 2020c). Most of the PCPs have higher Log Kow, Log Dow, and Log Koc values
391 and low water solubility compared with PhCs (Tables S2 and S3). On the other hand,
392 Matamoros et al. (2017) reported a significant positive correlation ($p < 0.05$) of the removal
393 efficiency of the combined PhCs and PCPs with Log Kow. However, the studied PhCs
394 (diclofenac, naproxen, ibuprofen, ketoprofen, and carbamazepine) has moderate to high Log
395 Kow which is comparable with the Log Kow of the studied PCPs (tributyl phosphate,
396 triphenyl phosphate, tris (2-chloroethyl) phosphate, galaxolide, tonalide, oxybenzone, methyl
397 dihydro-jasmonate, triclosan, and methyl paraben) (Tables S2 and S3).

398 **3.4.2. Predictive models combining PhCs and SHs**

399 PCA results showed that removal efficiency has negative loadings for both PhCs and SHs in
400 PC1 and does not cluster with Log Kow, Log Koc, and Log Dow which have high positive
401 loadings (Figures 3 & 8 and Table 1). The consistency in PCA and correlation results (Figure
402 4 and Table 1) contributed in forming acceptable predictive models with the combined data of
403 PhCs and SHs (Tables S25 and S31). This adequate predictive power of the combined models
404 might be due to the similarities in the physicochemical properties of these compounds (Tables
405 S2 and S4), as well as correspondence in their dominant removal mechanisms in the CWs. For
406 example, the biodegradation (aerobic and/or anaerobic) is the most dominant removal

407 mechanism for PhCs and SHs removal (e.g., Herrera-Melián et al., 2018; Chen et al., 2019;
408 Ilyas et al., 2020; Ilyas and van Hullebusch, 2020d).

409 The models were formed by combining Log Koc, Log Kow, Log Dow, MW, and pKa, mostly
410 with three or four variables together in one model. Although R^2 was not high (only about 0.4),
411 the regression models were statistically significant at 95% or 90% confidence interval ($p \leq$
412 0.1) (Table S31). Therefore, the resulting regression statistics were acceptable for the training
413 set and were applied on the test set (Table S25). For acceptable models, the RMSEs in the
414 training set were 15% in all the cases.

415 The best model was formed by combining MW, Log Kow, Log Koc, and pKa (Equation 5),
416 and yielded satisfactory results for the test set (RMSE: 27%) (Table S31). The difference in
417 the observed and predicted removal efficiency of 17 out of 23 tested PhCs and SHs was ≤
418 25%. The difference in the observed and predicted removal efficiency was quite small (<
419 10%) in the case of codeine, atenolol, ranitidine, bezafibrate, androstenedione, and
420 gemfibrozil (Figure 10). However, six out of 23 PhCs and SHs had a quite high difference in
421 the observed and predicted removal efficiency (ranging from 30% to 60%).

422 $$RE = 75.093 - 0.094\,MW + 4.249\,Log\,Kow - 11.878\,Log\,Koc + 2.541\,pKa \quad (5)$$

423 Where, *RE* is removal efficiency in %; *MW* is molecular weight g mol^{-1}; *Log Kow* is octanol-
424 water partition coefficient; *Log Koc* is soil organic carbon sorption coefficient; and *pKa* is
425 dissociation constant.

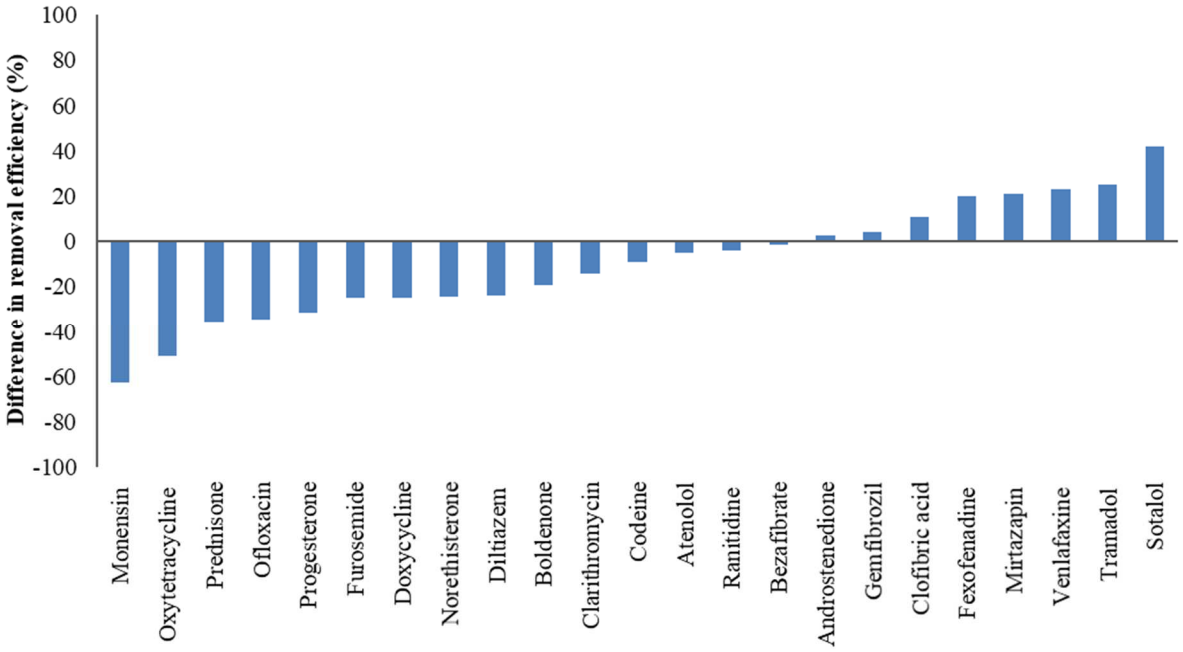


Figure 10. Difference between the observed and predicted removal efficiency based on the best model for combined PhCs and SHs with MW, Log Kow, Log Koc, and pKa.

3.4.3. Predictive models combining PCPs and SHs

In the case of PCPs and SHs, the models were formed by combining Log Koc, Log Kow, Log Dow, and MW, mostly with two or more variables together in one model. Although R^2 was not very high (about 0.5), the regression models were statistically significant at 95% or 90% confidence interval in most of the cases ($p < 0.1$) (Table S32). Therefore, the resulting regression statistics were acceptable (Table S26). For acceptable models, the RMSEs in the training set were 15% or 16%. The difference in the observed and predicted removal efficiency was not very large for most of the models, as illustrated by the results of the best model (Figure 11). The best performing model was formed by combining MW, Log Dow, and Log Koc (Equation 6), which yielded quite good results for both the training set (RMSE: 15%) and the test set (RMSE: 23%) (Table S32). The difference in the observed and predicted removal efficiency of seven out of 14 PCPs and SHs was < 20%. Two compounds (propylparaben and boldenone) showed difference in the observed and predicted removal efficiency up to 29%. The difference in the observed and predicted removal efficiency was

443 small (< 15%) in the case of cashmeran, triphenyl phosphate, triclocarban, N,N-diethyl-meta-
444 toluamide, tributyl phosphate, and androstenedione (Figure 11). This indicate that the removal
445 efficiency of these compounds could be adequately predicted by the proposed models.
446 However, in the case of three compounds (progesterone, norethisterone, and prednisone) the
447 difference in the observed and predicted removal efficiency was high (> 30%).

448 $$RE = 48.691 + 0.049\,MW + 15.545\,Log\ Dow - 18.451\,Log\ Koc \quad (6)$$

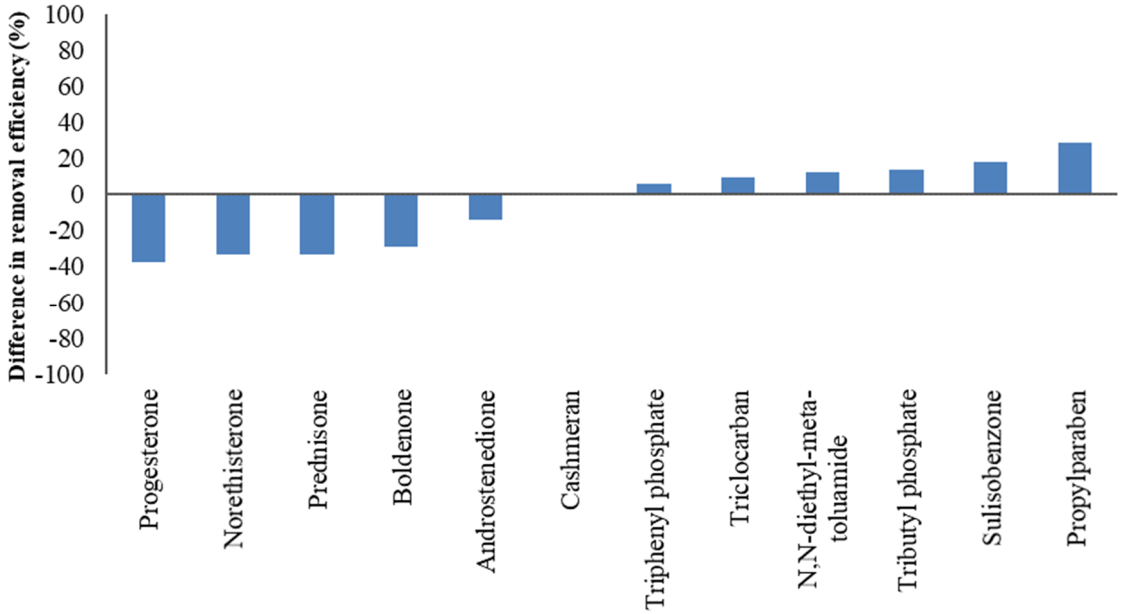
449 Where, *RE* is removal efficiency in %; *MW* is molecular weight g mol^{-1}; *Log Kow* is octanol-
450 water distribution coefficient; and *Log Koc* is soil organic carbon sorption coefficient.

451 The second-best model presented in Equation 7 is also possible with MW, Log Kow, and Log
452 Koc (Table S27), which yielded almost similar results for the training and test sets (RMSEs:
453 16% and 24%, respectively) (Table S32). However, this model was not able to explain more
454 variance compared with the best model. Nevertheless, the differences in observed and pre-
455 dicted removal efficiencies were somewhat similar to the best model.

456 $$RE = 39.301 + 0.098\,MW + 15.363\,Log\ Kow - 20.475 Log\ Koc \quad (7)$$

457 Where, *RE* is removal efficiency in %; *MW* is molecular weight g mol^{-1}; *Log Kow* is octanol-
458 water distribution coefficient; and *Log Koc* is soil organic carbon sorption coefficient.

459 The adequate predictive power of the combined model of PCPs and SHs might be due to the
460 similarities in some of the physicochemical properties of these compounds in most of the
461 cases (e.g., high Log Koc > 2.50 and low WS < 500) (Tables S3 and S4).



462

463 Figure 11. Difference between the observed and predicted removal efficiency based on the best model

464 for combined PCPs and SHs with MW, Log Dow, and Log Koc.

465 **3.4.4. Predictive models combining PhCs, PCPs, and SHs**

466 The performance of combined models with PhCs, PCPs and SHs was not satisfactory, as

467 demonstrated by weak correlations and multiple linear regression results. For example, the

468 resulting regression statistics were not acceptable for the training set because R^2 was very low

469 (< 0.3) and the regression models were not statistically significant (p > 0.1). Thus, these

470 models were not tested to predict the removal efficiency of the combined data of PhCs, PCPs,

471 and SHs.

472 This lack of adequate predictive power of the combined model could be attributed to the

473 distinct differences in the physicochemical properties of these compounds, in particular PhCs

474 and PCPs, as well as the differences in their dominant removal mechanisms in CWs. For

475 example, the biodegradation (aerobic and/or anaerobic) is the most dominant removal

476 mechanism for PhCs and SHs in CWs (e.g., Rühmland et al., 2015; Dai et al., 2017; Ilyas et

477 al., 2020; Ilyas and van Hullebusch, 2020d). The satisfactory predictive power of the

478 combined model of PhCs and SHs might be due to the similarities in some of the

physicochemical properties of these compounds (e.g., Log Kow, Log Dow, and Log Koc) in most of the cases (Tables S2 and S4). On the other hand, PCPs have adsorption to the substrate and/or sorption onto organic surfaces as their major removal mechanisms (e.g., Matamoros et al., 2016; Vystavna et al., 2017; Ilyas and van Hullebusch, 2020c). Most of the PCPs have much higher Log Kow, Log Dow, and Log Koc values and low WS compared with PhCs (Tables S2 and S3). The adequate predictive power of the combined model of PCPs and SHs might be due to the similarities in some of the physicochemical properties of these compounds (e.g., high Log Koc > 2.50 and low WS < 500) in most of the cases (Tables S3 and S4). These distinct features in their physicochemical properties lead to the differences in their dominant removal mechanisms and hinder the possibility to formulate a generic predictive model based on a combined data set of PhCs, PCPs, and SHs. These findings are consistent with the study by Vystavna et al. (2017). The authors reported no correlation among the physicochemical properties (WS, MW, and Log Kow) of PhCs, PCPs, and SHs with their removal efficiencies in CWs based on the PCA results.

3.5. Uncertainty Analysis

The removal efficiencies of PhCs, PCPs, and SHs could be predicted at variable degree of accuracy by using the proposed models. Although few models with combined data sets (e.g., PhCs and SHs; PCPs and SHs) indicated satisfactory performance, the predictions with individual models for PhCs, PCPs, and SHs were more reliable. However, there is some degree of uncertainty in the outcomes of all the predictive models, which could be considered in the analysis and the decision-making process. One way to include the uncertainty in the results is to make the predictions using more than one acceptable model, and then including the uncertainty range in the outcome. Using this approach, mean and standard deviation of the predicted removal efficiency of each PhCs, PCPs, and SHs was estimated using all the acceptable models. Figure 12, illustrates this point with an example of predicted mean and

504 standard deviation of PhCs based on the five acceptable models. The prediction uncertainty
505 among different models was not very high because the standard deviations were quite small in
506 most of the cases (< 5%). The RMSE values estimated based on the average predicted
507 removal efficiency of all the models was 28%, which was not much different from that of the
508 individual models (Table S28).

509 Similar procedure was followed for PCPs and SHs, and combined models (PhCs and SHs;
510 PCPs and SHs), and the resulting uncertainty range of predictions in their removal efficiencies
511 are shown in Figures 13-16. The results follow similar pattern, with relatively small
512 uncertainty ranges for most of the EOCs. The RMSE values for average of more than one
513 model were also comparable to the best model. Therefore, it seems reasonably good to use
514 either the best model to have one deterministic prediction of the removal efficiency or
515 estimate mean and standard deviation based on all the promising models. The application of
516 the latter option could be preferred, as this approach is a good reminder that the predicted
517 values could not be taken as exact estimates and are always have some degree of uncertainty
518 attached to them.

519 Another source of uncertainty is related to the data of physicochemical properties and removal
520 efficiencies. For example, there could be some degree of uncertainty attached to the mean of
521 the observed removal efficiency (Tables S2-S4), and it is not possible to take these
522 observations as fully accurate measure of the central tendency. Furthermore, values of some
523 physicochemical properties like MW and WS have high degree of accuracy (less uncertainty).
524 Whereas, values of some other physicochemical properties like Log Dow could have a bit
525 higher degree of uncertainty because these were not easily available from the experimental
526 studies, and were estimated from other models.

Thus, the predicated removal efficiencies should be interpreted with caution. These predictions could not be taken as absolute numbers; neither these should be considered as a substitute for rigorous experimental studies.

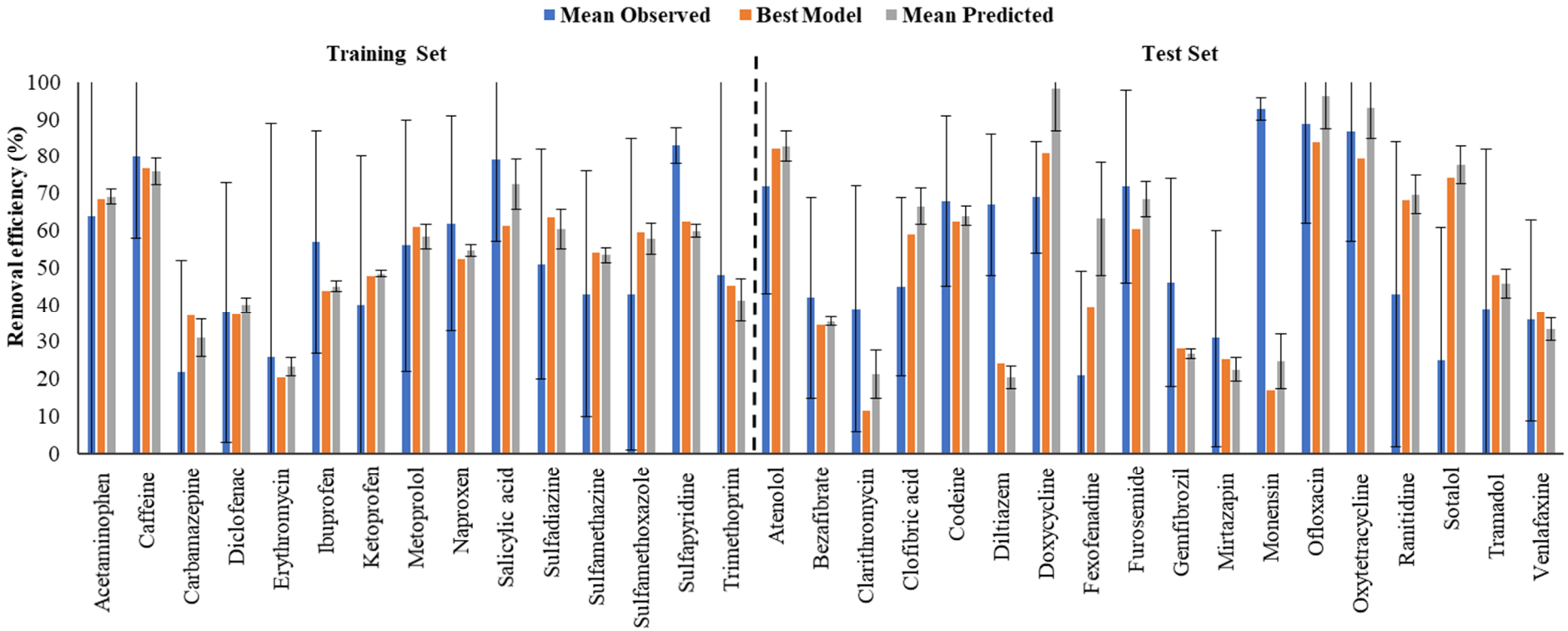
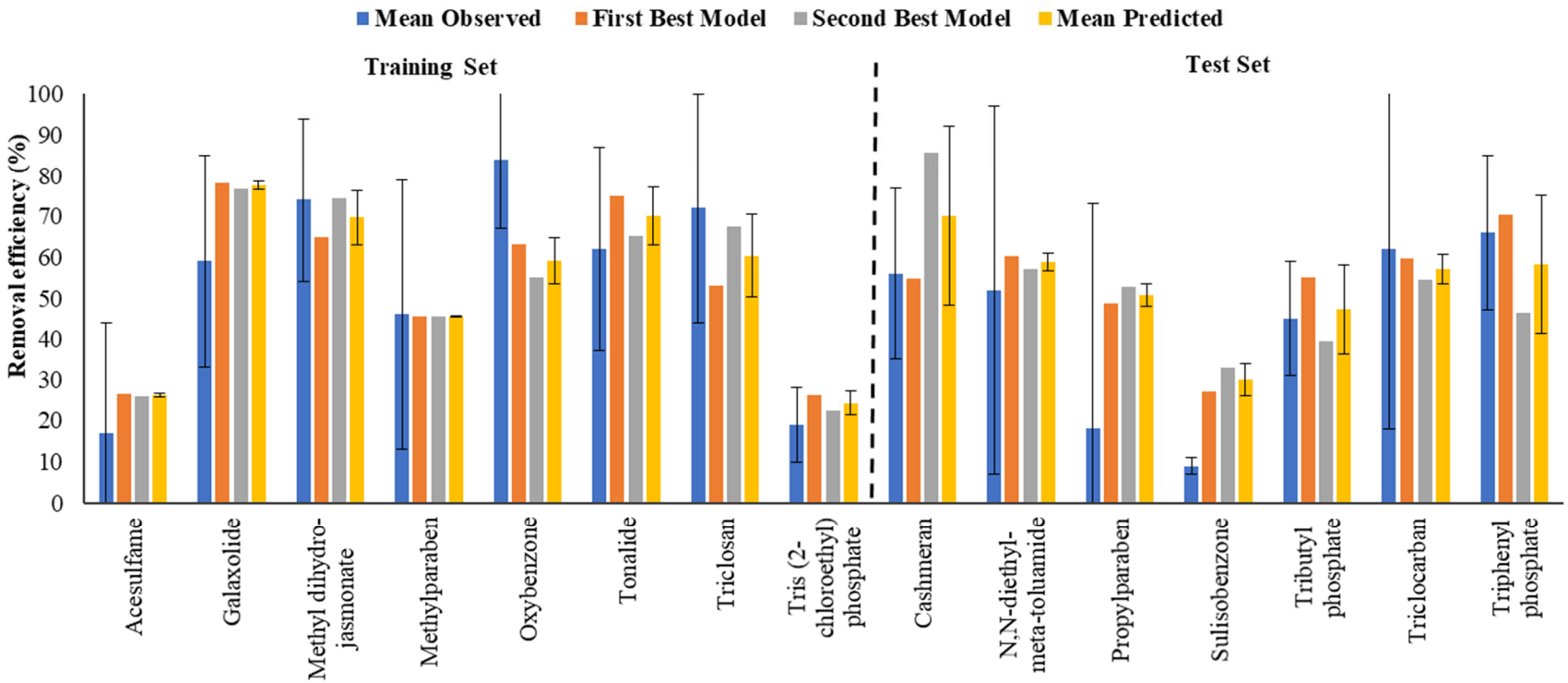


Figure 12. Difference between the observed and predicted removal efficiency based on five acceptable models formed using data of 15 PhCs with ≥ 25 data points.

Note: Standard deviation values were capped at 100 to improve the readability of the graph. Actual values can be found in the developed decision support tool (DST), REOCW-PCP provided as a supplementary material (Excel file: REOCW-PCP-secure, along with the user manual).

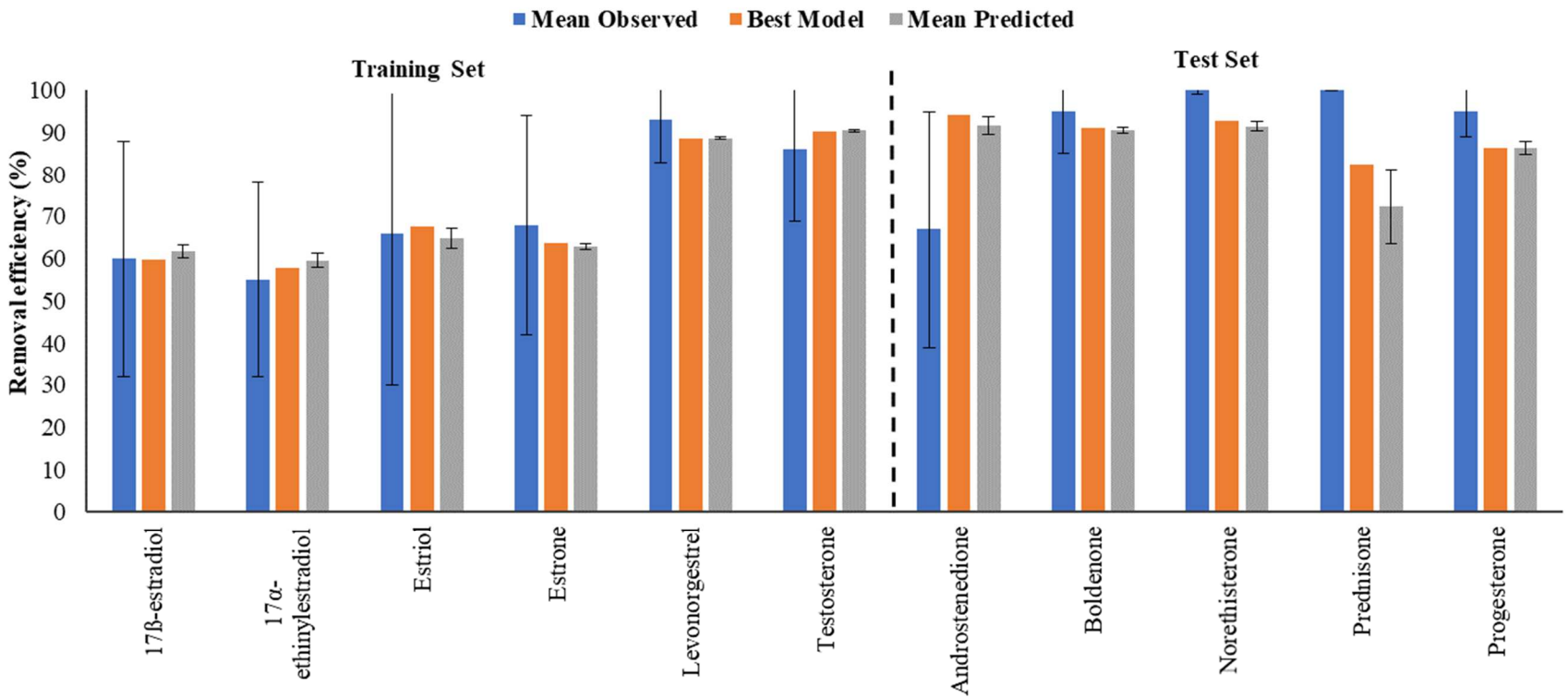


535

536 Figure 13. Difference between the observed and predicted removal efficiency based on two acceptable models formed using data of eight PCPs with > 5 data

537 points.

538 Note: Standard deviation values were capped at 100 to improve the readability of the graph. Actual values can be found in the developed DST, REOCW-PCP.

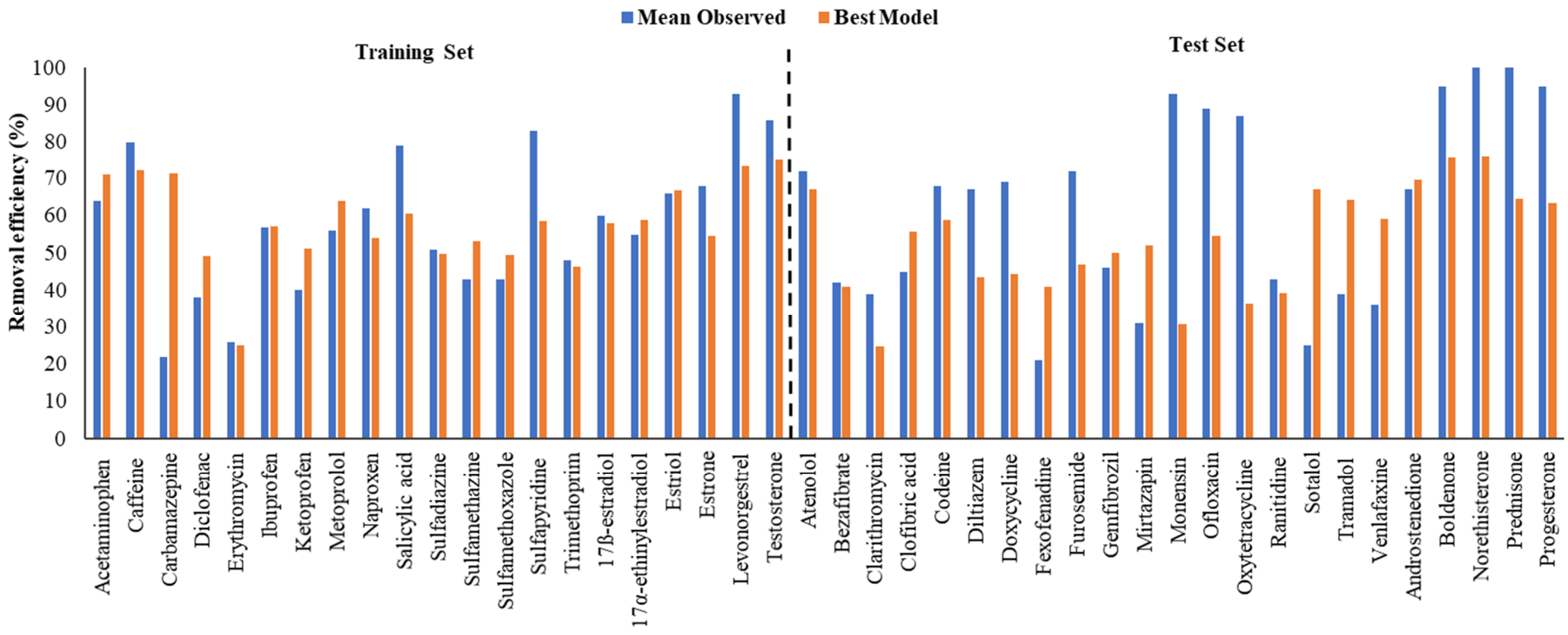


539

540 Figure 14. Difference between the observed and predicted removal efficiency based on three acceptable models formed using data of six SHs with > 10 data

541 points.

542 Note: Standard deviation values were capped at 100 to improve the readability of the graph. Actual values can be found in the developed DST, REOCW-PCP.

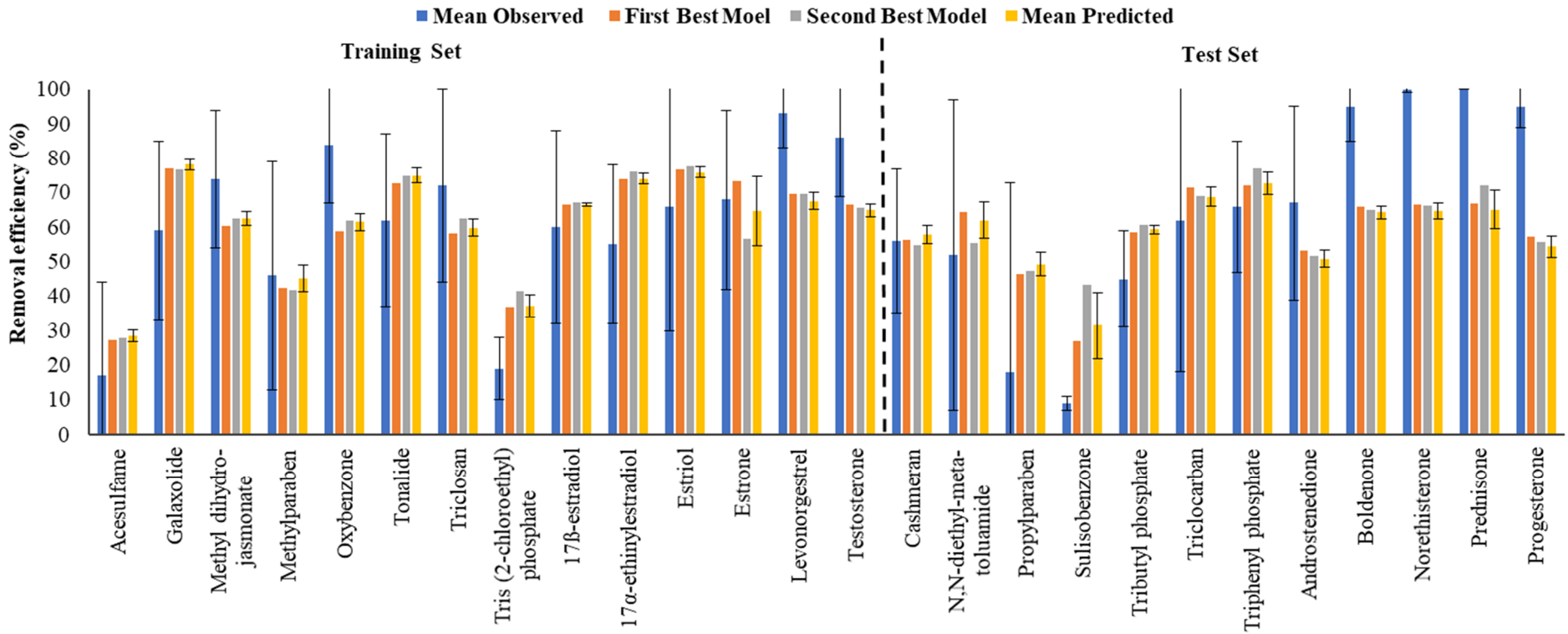


543

544 Figure 15. Difference between the observed and predicted removal efficiency based on one acceptable model formed using data of 15 PhCs and six SHs with

545 data points ≥ 25 and > 10, respectively.

546 Note: Standard deviation of the removal efficiency in the case of combined model of PhCs and SHs was not available because predictions are made based on one model.



547

548 Figure 16. Difference between the observed and predicted removal efficiency based on four acceptable models formed using data of eight PCPs and six SHs

549 with data points > 5 and > 10, respectively.

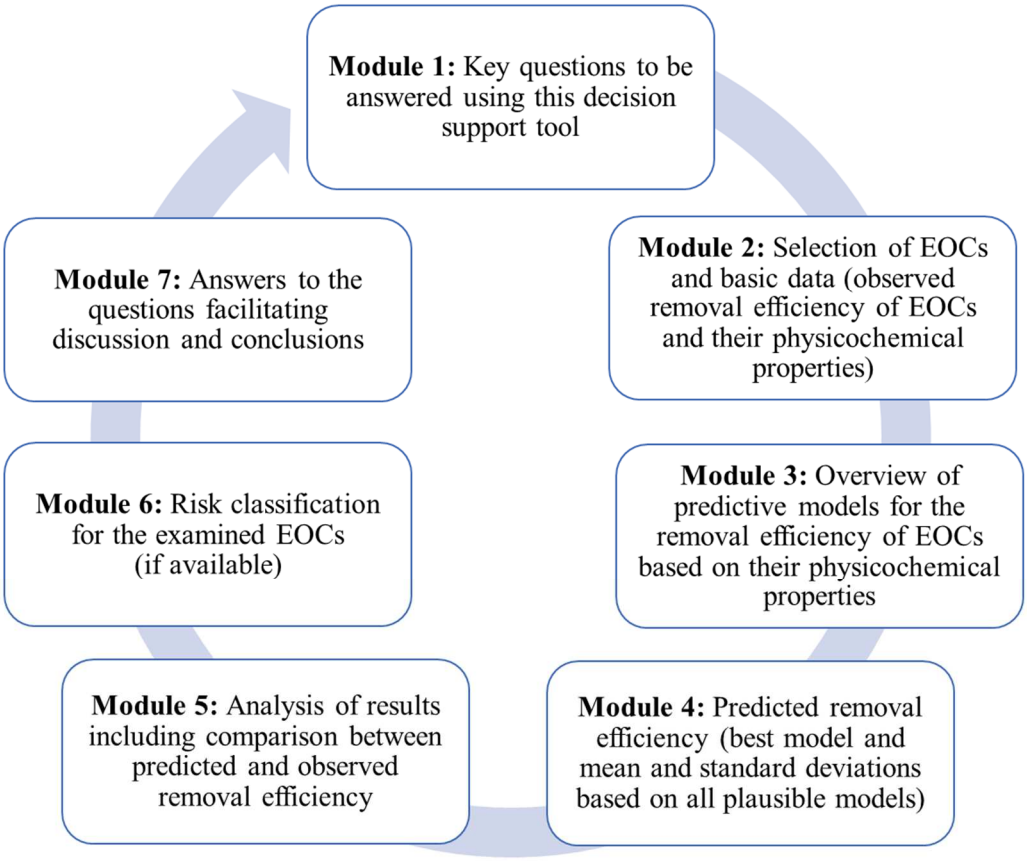
550 Note: Standard deviation values were capped at 100 to improve the readability of the graph. Actual values can be found in the developed DST, REOCW-PCP.

4. Further development and applications

The novel predictive models developed in this study could be useful DSTs. For example, the plausible models could be used as screening tools to obtain first-hand information on the removal efficiency of a certain PhC, PCP or SH when the wastewater containing them is treated using CWs. The predicted removal efficiency can give an overview of mean as well as range (e.g., standard deviation), providing insights into possible removal efficiency. This information may guide in decision making on research directions as well as policy actions, and practice (e.g., design and application of CWs in the field conditions).

Another area of application and further development could be to include these predictive models into already existing DSTs. For example, the addition of the proposed models to QSAR toolbox could extend its screening services to the domain of wastewater treatment using CWs. The QSAR toolbox contains a wealth of information on a large number of chemicals including PhCs, PCPs, and SHs, which may offer wider scale application of the novel models developed in this study. The QSAR tool is widely used for various purposes, most notably for the screening of new drugs based on existing QSAR in order to limit experimental and clinical trials to save time, costs, and animal lives. In a similar way, screening of a large number of PhCs, PCPs, and SHs could be possible when the models developed in this study will become part of the QSAR toolbox.

Moreover, the predictive models could be used in developing new DSTs for researchers, practitioners, policy makers, and citizens. For example, the authors have developed a novel DST based on this research, which is named as REOCW-PCP. The key features of the DST are schematized in Figure 17 and outlined below. REOCW-PCP was developed using Microsoft Excel 2016, and is provided as a supplementary material (Excel file: REOCW-PCP-secure, along with the user manual).



575

576 Figure 17. A schematic diagram showing different modules of the developed DST, REOCW-PCP.

577 The proposed REOCW-PCP is composed of seven modules, which could be followed
578 sequentially. The first module displays the basic set up of REOCW-PCP, and contains the list
579 of key questions that could be explored using the tool. REOCW-PCP can be used by an
580 individual user or multiple users in the form of a group of different stakeholders (e.g.,
581 policy/decision makers, wastewater managers, researchers/scientists, and citizens). The key
582 questions are enumerated below.

583 **1.** What could be the removal efficiency of a certain EOC in CWs?

584 **2.** What is the likelihood of removal efficiency in falling under a certain quartile range (e.g.,
585 Q1: 0-25%; Q2: 26-50%; Q3: 51-75%; Q4: 76-100%)?

586 **3.** Which physicochemical properties are better predictor of the removal efficiency for a
587 certain EOC?

588 **4.** How the predicted and experimental removal efficiencies differ from each other?

589 **5.** What is the range of uncertainty in the available prediction and field observations?

590 **6.** What is the potential environmental risk posed by the EOCs of interest?

591 **7.** What lessons could be drawn based on the predictions and observations?

592 The second module contains the basic data of the EOCs including their names and
593 physicochemical properties. In this module, the user will select specific EOCs to examine. In
594 case REOCW-PCP is used by multiple stakeholders, they need to agree on the final list of the
595 EOCs to be examined. After selecting the EOCs from the given list, the tool automatically
596 does the required calculations to answer the key questions. The answers to these questions are
597 presented under module 7, and are ready for the review. However, it is recommended that the
598 users go through all the modules (3-7) in a sequential way to develop a good understanding of
599 the predictive models, details on calculations, and some key information provided in each
600 module.

601 The third module is composed of the novel predictive models developed in this research.
602 These models are used to make prediction of the removal efficiency, and used in making the
603 calculations required to answer the questions indicated under module 1. The fourth module
604 displays the results of the predicted removal efficiencies. The results are shown as a
605 deterministic value in the case of a single model application (e.g., best model) or mean and
606 standard deviation (e.g., including uncertainty range) when multiple models are used. The
607 fifth module enables further analysis of the results, mainly by comparing the predicted and
608 observed removal efficiency. This will help to triangulate the predictions with experimental
609 results, which will contribute to informed decision-making process. We have included
610 information on environmental risk posed by the EOCs in the sixth module, which is based on
611 our previous work (Ilyas et al., 2020; Ilyas and van Hullebusch, 2020c, 2020d). The results

612 are classified under various risk categories: high, medium, low, and no risk. The seventh
613 module offers the possibility to the users of this tool to find the answers to the key questions
614 and then discuss the lessons, and conclusions that could be drawn from the study. In this
615 module, standard answers to the questions are generated in Tabular and Graphical forms,
616 which will contribute to kick-off the discussion. Finally, the users may choose to finish the
617 session or choose to repeat the whole cycle (all modules) or a part of it (e.g., specific
618 modules) using a new set of EOCs.

619 # 5. Conclusions

620 The following conclusions are drawn from this study:

621 **1.** The physicochemical properties of EOCs are good predictors of their removal efficiency
622 in CWs. This was demonstrated by developing several plausible predictive models formed
623 by combining two to four physicochemical properties (MW, Log Dow, Log Koc, Log
624 Kow, and pKa) in a multiple linear regression model. Among the studied physicochemical
625 properties, Log Koc, Log Dow, and Log Kow emerged as the most significant predictors,
626 and combination of one or two of them with MW and/or pKa often improved the
627 reliability of the predictions.

628 **2.** The removal efficiency of PhCs, PCPs, and SHs in CWs could be predicted reasonably
629 well using their physicochemical properties: MW, Log Dow, Log Koc, Log Kow, and
630 pKa. In the case of PhCs, several novel models were formed to predict their removal
631 efficiency by combining three or four of these variables together in one model and the best
632 model was formed by combining MW, Log Dow, and Log Koc (R^2: 0.601; $p < 0.05$;
633 RMSEs: training set: 11%; test set: 27%). In the case of PCPs, the two plausible models
634 were developed, one with two independent variables: Log Kow and Log Koc (R^2: 0.644; p
635 < 0.1; RMSEs: training set: 14%; test set: 14), and the other with MW, Log Kow, and pKa

(R^2: 0.709 and $p \leq 0.1$; RMSEs: training set: 13%; test set: 21%). Similarly, the removal efficiency of SHs could be predicted very well by combining three of the five physicochemical properties and the reliability of the developed models was quite high, as demonstrated by a very good performance statistics for the three plausible models (R^2: $\geq$ 0.90; $p \leq 0.1$; RMSEs: training set: 3-4%; test set: 15-19 %).

3. The resulting regression statistics of a generic model by combing PhCs, PCPs, and SHs were not acceptable for the training set, as demonstrated by a very low performance statistics (R^2: < 0.3; *p*-values were much higher than 0.1). The lack of adequate predictive power of the combined model could be attributed to the distinct differences in the physicochemical properties of PhCs and PCPs as well as the differences in their dominant removal mechanisms in CWs. For example, most of the PCPs have much higher Log Kow, Log Dow, and Log Koc values and low WS compared with PhCs. Consequently, biodegradation (aerobic and/or anaerobic) is the most dominant removal mechanism for PhCs, and adsorption to the substrate and sorption onto organic surfaces for PCPs in CWs. These distinct features in their physicochemical properties lead to the differences in their dominant removal mechanisms and could limit the possibility to formulate a generic predictive model based on a combined data set of PhCs, PCPs, and SHs.

4. The adequate predictive power of the combined model of PhCs and SHs was achieved, which might be due to the similarities in some of the physicochemical properties of these compounds (e.g., Log Kow, Log Dow, and Log Koc) in most of the cases. The best performing model was formed by combining MW, Log Kow, Log Koc, and pKa, which demonstrated a very good performance for the training set (RMSE: 3%) and the test set (RMSE: 15%).

659 **5.** The combination of PCPs and SHs also resulted in reasonably good predictive models.
660 The adequate predictive power of the combined model of PCPs and SHs might be due to
661 the similarities in some of the physicochemical properties of these compounds (e.g., high
662 Log Koc > 2.50 and low WS < 500) in most of the cases. The best model was formed by
663 combining MW, Log Dow, and Log Koc, which yield quite good results for the training
664 set (RMSE: 15%) and the test set (RMSE: 23%).

665 **6.** It is very important to recognize the inherent uncertainties associated with the prediction
666 process. These uncertainties could be related to the input data (e.g., removal efficiencies
667 and physicochemical properties of EOCs) and the variability in the predictions made from
668 multiple plausible models. Therefore, these uncertainties should be recognized, and the
669 prediction results should be used with caution. It is advisable to estimate the mean and
670 standard deviation based on all the promising models, even though one estimate from the
671 best model could also serve the purpose. Where field data is available, a comparison with
672 observed values is advisable.

673 **7.** From the application point of view, the novel and plausible predictive models developed
674 in this study could be useful DSTs. In particular, the proposed models could serve as
675 screening tools to gain insights about the removal efficiency of a certain PhC, PCP or SH
676 in CWs. Moreover, the models could be included into existing DSTs such as inclusion
677 into the QSAR toolbox to extend its capabilities to offer the screening of EOCs removal
678 by CWs. There are good prospects to develop new DSTs using the predictive models
679 developed in this study. For example, a novel DST named as REOCW-PCP was
680 developed based on the findings of this research. In general, the information obtained by
681 the use of the proposed predictive models and developed REOCW-PCP could contribute
682 to enhance the knowledge and understanding for the design and application of CWs for
683 EOCs removal from the wastewater

References

Barbosa, M.O., Moreira, N.F.F., Ribeiro, A.R., Pereira, M.F.R., Silva, A.M.T., 2016. Occurrence and removal of organic micropollutants: an overview of the watch list of EU Decision 2015/495. *Water Res*. 94, 257-279. DOI: https://doi.org/10.1016/j.watres.2016.02.047

Breitholtz, M., Näslund, M., Stråe, D., Borg, H., Grabic, R., Fick, J., 2012. An evaluation of free water surface wetlands as tertiary sewage water treatment of micro-pollutants. *Ecotoxicol. Environ. Saf*. 78, 63-71. DOI: https://doi.org/10.1016/j.ecoenv.2011.11.014

Caliman, F.A., Gavrilescu, M., 2009. Pharmaceuticals, personal care products and endocrine disrupting agents in the environment-a review. *Clean* 37 (4-5), 277-303. DOI: https://doi.org/10.1002/clen.200900038

Carballa, M., Fink, G., Omil, F., Lema, J.M., Ternes, T., 2008. Determination of the solid–water distribution coefficient (Kd) for pharmaceuticals, estrogens and musk fragrances in digested sludge. *Water Res*. 287-295. DOI: https://doi.org/10.1016/j.watres.2007.07.012

Carranza-Diaz, O., Schultze-Nobre, L., Moeder, M., Nivala, J., Kuschk, P., Koeser, H., 2014. Removal of selected organic micropollutants in planted and unplanted pilot-scale horizontal flow constructed wetlands under conditions of high organic load. *Ecol. Eng.* 71, 234-245. DOI: https://doi.org/10.1016/j.ecoleng.2014.07.048

Chen, J., Liu, Y.-S., Deng, W.-J., Ying, G.-G., 2019. Removal of steroid hormones and biocides from rural wastewater by an integrated constructed wetland. *Sci. Total Environ.* 660, 358-365. DOI: https://doi.org/10.1016/j.scitotenv.2019.01.049

Chen, Y., Vymazal, J., Březinová, T., Koželuh, M., Kule, L., Huang, J., Chen, Z., 2016. Occurrence, removal and environmental risk assessment of pharmaceuticals and personal care products in rural wastewater treatment wetlands. *Sci. Total Environ.* 566-567, 1660-1669. DOI: https://doi.org/10.1016/j.scitotenv.2016.06.069

Choi, Y.-J., Kim, L.-H., Zoh, K.-D., 2016. Removal characteristics and mechanism of antibiotics using constructed wetlands. *Ecol. Eng.* 91, 85-92. DOI: https://doi.org/10.1016/j.ecoleng.2016.01.058

710 Dai, Y., Tao, R., Tai, Y., Tam, N.F., Dan, A., Yang, Y., 2017. Application of a full-scale newly
711 developed stacked constructed wetland and an assembled bio-filter for reducing phenolic
712 endocrine disrupting chemicals from secondary effluent. *Ecol. Eng.* 99, 496-503. DOI:
713 https://doi.org/10.1016/j.ecoleng.2016.11.007
714 Gogoi, A., Mazumder, P., Tyagi, V.K., Chaminda, G.G.T., An, A.K., Kumar, M., 2018. Occurrence
715 and fate of emerging contaminants in water environment: a review. *Groundw. Sustain. Dev.* 6,
716 169-180. DOI: https://doi.org/10.1016/j.gsd.2017.12.009
717 Gorito, A.M., Ribeiro, A.R., Almeida, C.M.R., Silva, A.M.T., 2017. A review on the application of
718 constructed wetlands for the removal of priority substances and contaminants of emerging
719 concern listed in recently launched EU legislation. *Environ. Pollut.* 227, 428-443. DOI:
720 https://doi.org/10.1016/j.envpol.2017.04.060
721 Herrera-Melián, J.A., Guedes-Alonso, R., Borreguero-Fabelo, A., Santana-Rodríguez, J.J., Sosa-
722 Ferrera, Z., 2018. Study on the removal of hormones from domestic wastewaters with lab-scale
723 constructed wetlands with different substrates and flow directions. *Environ. Sci. Pollut. Res.* 25
724 (21), 20374-20384. DOI: https://doi.org/10.1007/s11356-017-9307-8
725 Hijosa-Valsero, M., Matamoros, V., Sidrach-Cardona, R., Martin-Villacorta, J., Becares, E., Bayona,
726 J.M., 2010. Comprehensive assessment of the design configuration of constructed wetlands for
727 the removal of pharmaceuticals and personal care products from urban wastewaters. *Water Res.*
728 44, 3669-3678. DOI: https://doi.org/10.1016/j.watres.2010.04.022
729 Ilyas, H., Masih, I., van Hullebusch, E.D., 2020. Pharmaceuticals removal by constructed wetlands: a
730 critical evaluation and meta-analysis on performance, risk reduction and role of
731 physicochemical properties on removal mechanisms. *J. Water Health* 18 (3), 253-291. DOI:
732 https://doi.org/10.2166/wh.2020.213
733 Ilyas, H., van Hullebusch, E.D., 2019. Role of design and operational factors in the removal of
734 pharmaceuticals by constructed wetlands. *Water* 11 (11), 2356, 1-24.
735 DOI: https://doi.org/10.3390/w11112356
736 Ilyas, H., van Hullebusch, E.D., 2020a. Performance comparison of different types of constructed
737 wetlands for the removal of pharmaceuticals and their transformation products: a review.

738 *Environ. Sci. Pollut. Res.* 27 (13), 14342–14364. DOI: https://doi.org/10.1007/s11356-020-
739 08165-w
740 Ilyas, H., van Hullebusch, E.D., 2020b. The influence of design and operational factors on the removal
741 of personal care products by constructed wetlands. *Water* 12 (5), 1367, 1-20. DOI:
742 https://doi.org/10.3390/w12051367
743 Ilyas, H., van Hullebusch, E.D., 2020c. Performance comparison of different constructed wetlands
744 designs for the removal of personal care products. *Int. J. Environ. Res. Public Health* 17 (9),
745 3091, 1-26. DOI: https://doi.org/10.3390/ijerph17093091
746 Ilyas, H., van Hullebusch, E.D., 2020d. A review on the occurrence, fate, and removal of steroidal
747 hormones during treatment with different types of constructed wetlands. *J. Environ. Chem. Eng.*
748 8 (3), 103793. DOI: https://doi.org/10.1016/j.jece.2020.103793
749 Kumar, J.L., Zhao, Y.Q., 2011. A review on numerous modeling approaches for effective, economical
750 and ecological treatment wetlands. *J. Environ. Manage.* 92, 400-406. DOI:
751 https://doi.org/10.1016/j.jenvman.2010.11.012
752 Lee, S., Kang, S.-I., Lim, J.-L., Huh, Y.J., Kim, K.-S., Cho, J., 2011. Evaluating controllability of
753 pharmaceuticals and metabolites in biologically engineered processes, using corresponding
754 octanol-water coefficient. *Ecol. Eng.* 37 (10), 1595-1600. DOI:
755 https://doi.org/10.1016/j.ecoleng.2011.04.007
756 Luo, Y., Guo, W., Ngo, H.H., Nghiem, L.D., Hai, F.I., Zhang, J., Liang, S., Wang, X.C., 2014. A
757 review on the occurrence of micropollutants in the aquatic environment and their fate and
758 removal during wastewater treatment. *Sci. Total Environ.* 473-474. 619-641. DOI:
759 https://doi.org/10.1016/j.scitotenv.2013.12.065
760 Matamoros, V., Rodríguez, Y., Albaigés, J., 2016. A comparative assessment of intensive and
761 extensive wastewater treatment technologies for removing emerging contaminants in small
762 communities. *Water Res.* 88, 777-785. DOI: https://doi.org/10.1016/j.watres.2015.10.058
763 Matamoros, V., Rodríguez, Y., Bayona, J.M., 2017. Mitigation of emerging contaminants by full-scale
764 horizontal flow constructed wetlands fed with secondary treated wastewater. *Ecol. Eng*. 99, 222-
765 227. DOI: https://doi.org/10.1016/j.ecoleng.2016.11.054

766 Meyer, D., Chazarenc, F., Claveau-Mallet, D., Dittmer, U., Forquet, N., Molle, P., Morvannou, A.,
767 Pálfy, T., Petitjean, A., Rizzo, A., Campà, R.S., Scholz, M., Soric, A., Langergraber, G., 2015.
768 Modelling constructed wetlands: Scopes and aims-A comparative review. *Ecol. Eng.* 80, 205-
769 213. DOI: https://doi.org/10.1016/j.ecoleng.2014.10.031
770 Nivala, J., Kahl, S., Boog, J., Afferden, M., Reemtsma, T., Müller, R.A., 2019. Dynamics of emerging
771 organic contaminant removal in conventional and intensified subsurface flow treatment
772 wetlands. *Sci. Total Environ.* 649, 1144-1156. DOI:
773 https://doi.org/10.1016/j.scitotenv.2018.08.339
774 Park, N., Vanderford, B.J., Snyder, S.A., Sarp, S., Kim, S.D., Cho, J., 2009. Effective controls of
775 micropollutants in wastewater effluent using constructed wetlands under anoxic condition. *Ecol.*
776 *Eng.* 35, 418-423. DOI: https://doi.org/10.1016/j.ecoleng.2008.10.004
777 Reyes-Contreras, C., Hijosa-Valsero, M., Sidrach-Cardona, R., Bayona, J.M., Bécares, E., 2012.
778 Temporal evolution in PPCP removal from urban wastewater by constructed wetlands of
779 different configuration: a medium-term study. *Chemosphere* 88, 161-167. DOI:
780 https://doi.org/10.1016/j.chemosphere.2012.02.064
781 Rühmland, S., Wick, A., Ternes, T.A., Barjenbruch, M., 2015. Fate of pharmaceuticals in a
782 subsurface flow constructed wetland and two ponds. *Ecol. Eng.* 80, 125-139. DOI:
783 https://doi.org/10.1016/j.ecoleng.2015.01.036
784 Töre, G.Y., Meriç, S., Lofrano, G., Feo, G.D., 2012. Removal of trace pollutants from wastewater in
785 constructed wetlands. Lofrano, G., (Editor), *Emerging compounds removal from wastewater,*
786 *natural and solar based Treatment*. Springer Dordrecht Heidelberg New York London. pp (39-
787 58). DOI: https://doi.org/10.1007/978-94-007-3916-1_3
788 Tran, N.H., Reinhard, M., Gin., K.Y.-H., 2018. Occurrence and fate of emerging contaminants in
789 municipal wastewater treatment plants from different geographical regions-a review. *Water Res.*
790 133, 182-207. DOI: https://doi.org/10.1016/j.watres.2017.12.029
791 Verlicchi, P., Galletti, A., Petrovic, M., Barceló, D., Al Aukidy, M., Zambello, E., 2013. Removal of
792 selected pharmaceuticals from domestic wastewater in an activated sludge system followed by a

horizontal subsurface flow bed-analysis of their respective contributions. *Sci. Total Environ.* 454-455, 411-425. DOI: https://doi.org/10.1016/j.scitotenv.2013.03.044

Verlicchi, P., Zambello, E., Aukidy, M.A., 2015. Removal of Personal Care Products in Constructed Wetlands. *In Personal Care Products in the Aquatic Environment*; Díaz-Cruz, M.S., Barceló, D., Eds.; The Handbook of Environmental Chemistry; Springer: Cham, Switzerland, 2015; Volume 36, pp. 319-354. DOI: https://doi.org/10.1007/698_2014_282

Vo, H.-N.-P., Bui, X.-T., Nguyen, T.-M.-H., Koottatep, T., Bandyopadhyay, A., 2018. Insights of the removal mechanisms of pharmaceutical and personal care products in constructed wetlands. *Curr. Pollut. Rep*. 4 (2), 93-103. DOI: https://doi.org/10.1007/s40726-018-0086-8

Vymazal, J., Březinová, T., Koželuh, M., 2015. Occurrence and removal of estrogens, progesterone and testosterone in three constructed wetlands treating municipal sewage in the Czech Republic. *Sci. Total Environ*. 536, 625-631. DOI: https://doi.org/10.1016/j.scitotenv.2015.07.077

Vymazal, J., Březinová, T.D., Koželuh, M., Kule, L., 2017. Occurrence and removal of pharmaceuticals in four full-scale constructed wetlands in the Czech Republic – the first year of monitoring. *Ecol. Eng*. 98, 354-364. DOI: https://doi.org/10.1016/j.ecoleng.2016.08.010

Vystavna, Y., Frkova, Z., Marchand, L., Vergeles, Y., Stolberg, F., 2017. Removal efficiency of pharmaceuticals in a full scale constructed wetland in East Ukraine. *Ecol. Eng*. 108, 50-58. DOI: https://doi.org/10.1016/j.ecoleng.2017.08.009

Wang, Y., Yin, T., Kelly, B.C., Gin, K.Y.-H., 2019. Bioaccumulation behaviour of pharmaceuticals and personal care products in a constructed wetland. *Chemosphere* 222, 275-285. DOI: https://doi.org/10.1016/j.chemosphere.2019.01.116

Yi, X., Tran, N.H., Yin, T., He, Y., Gin, K.Y.-H., 2017. Removal of selected PPCPs, EDCs, and antibiotic resistance genes in landfill leachate by a full-scale constructed wetlands system. *Water Res*. 121, 46-60. DOI: https://doi.org/10.1016/j.watres.2017.05.008

Yin, T., Tran, N.H., Huiting, C., He, Y., Gin, K.Y.-H., 2019. Biotransformation of polyfluoroalkyl substances by microbial consortia from constructed wetlands under aerobic and anoxic conditions. *Chemosphere* 233, 101-109. DOI: https://doi.org/10.1016/j.chemosphere.2019.05.227

821 Yuan, C., Huang, T., Zhao, X., Zhao, Y., 2020. Numerical Models of Subsurface Flow Constructed
822 Wetlands: Review and Future Development. *Sustainability* 12, 3498, 1-16. DOI:
823 https://doi.org/10.3390/su12083498
824 Zhang, D., Gersberg, R.M., Ng, W.J., Tan, S.K., 2014. Removal of pharmaceuticals and personal care
825 products in aquatic plant-based systems: a review. *Environ. Pollut.* 184, 620-639. DOI:
826 https://doi.org/10.1016/j.envpol.2013.09.009
827 Zhang, D.Q., Gersberg, R.M., Hua, T., Zhu, J., Tuan, N.A., Tan, S.K., 2012a. Pharmaceutical removal
828 in tropical subsurface flow constructed wetlands at varying hydraulic loading rates.
829 *Chemosphere* 87, 273-277. DOI: https://doi.org/10.1016/j.chemosphere.2011.12.067
830 Zhang, D.Q., Gersberg, R.M., Zhu, J., Hua, T., Jinadasa, K.B.S.N., Tan, S.K., 2012b. Batch versus
831 continuous feeding strategies for pharmaceutical removal by subsurface flow constructed
832 wetland. *Environ. Pollut.* 167, 124-131. DOI: https://doi.org/10.1016/j.envpol.2012.04.004
833 Zhang, X., Jing, R., Feng, X., Dai, Y., Tao, R., Vymazal, J., Cai, N., Yang, Y., 2018. Removal of
834 acidic pharmaceuticals by small-scale constructed wetlands using different design
835 configurations. *Sci. Total Environ.* 639, 640-647. DOI:
836 https://doi.org/10.1016/j.scitotenv.2018.05.198

Graphical abstract

Prediction of the removal efficiency of emerging organic contaminants in constructed wetlands based on their physicochemical properties