Waste Management 172 (2023) 101–107

Contents lists available at ScienceDirect

Waste Management

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

Research Paper

Identification and verification of PCDD/Fs indicators from four typical large-scale municipal solid waste incinerations with large sample size in China

Lijun Liu [a,1], Xichao Chen [a,c,1], Wenhua Yin [a], Hao Wu [b], Junbin Huang [b], Yanyan Yang [a], Zhiqiang Gao [a], Jinqiong Huang [a], Jianping Fu [a], Jinglei Han [a,*]

[a] South China Institute of Environmental Sciences, Ministry of Ecology and Environment, Guangzhou 510000, China
[b] Shenzhen Energy Environment, Co., LTD, Shenzhen 518055, China
[c] State Key Laboratory of Organic Geochemistry, Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou 510640, China

ARTICLE INFO

Keywords:
MSWI
Emissions
Indicator
Traditional statistical method
Machine learning method

ABSTRACT

Monitoring PCDD/Fs emissions from municipal solid waste incinerations (MSWIs) is of paramount importance, yet it can be time-consuming and labor-intensive. Predictive models offer an alternative approach for estimating their levels. However, robust models specific to PCDD/Fs were lacking. In this study, we collected 190 PCDD/Fs samples from 4 large-scale MSWIs in China, with the average PCDD/Fs levels and TEQ levels of 0.987 ng/m^3 and 0.030 ng TEQ/m^3, respectively. We developed and evaluated predictive models, including traditional statistical methods, e.g., linear regression (LR) as well as machine learning models such as back propagation-artificial neural networks (BP ANN) and random forest (RF). Correlation analysis identified 2,3,4,7,8-PeCDF, 1,2,3,6,7,8-HxCDF, 2,3,4,6,7,8-HxCDF were better indicator congeners for PCDD/Fs estimation ($R^2 > 0.9$, $p < 0.001$). The predictive results favored the RF model, exhibiting a high R^2 value and low root mean square error (RMSE) and mean absolute error (MAE). Additionally, the RF model showed excellent prediction ability during external validation, with low absolute relative error (ARE) of 10.9 %-12.6 % for the three indicator congeners in the normal PCDD/F TEQ levels group (<0.1 ng TEQ/m^3) and slightly higher ARE values (13.8 %-17.9 %) for the high PCDD/F TEQ levels group (>0.1 ng TEQ/m^3). In conclusion, our findings strongly support the RF model's effectiveness in predicting PCDD/Fs TEQ emission from MSWIs.

1. Introduction

Rapid economic growth, industrialization, and urbanization in China have resulted in an unprecedented increase in municipal solid waste (MSW). Municipal solid waste incinerators (MSWIs) have become indispensable in reducing both the weight (up to 75 %) and volume (up to 90 %) of MSW, particularly in areas facing limited landfill space (Cheng and Hu, 2010; Domingo et al., 2015; Rovira et al., 2015). The demand for incineration has grown significantly, with the volume of MSW sent for incineration in China soaring from 46.3 million tons in 2013 to 146.08 million tons in 2020 (National Bureau of Statistics of China, 2020, National Bureau of Statistics of China (2013)). This underscores the importance of efficient waste management strategies, where MSWIs are essential to curb the environmental impact of mounting MSW volumes.

Polychlorinated dibenzo-p-dioxins and dibenzofurans (PCDD/Fs) primarily form through chlorine-involved combustion process such as waste incineration and 17 PCDD/Fs congeners with chlorine substitution in 2,3,7,8 positions have comparable or even higher toxicity than other congeners, inducing adverse health effects (e.g., reproductive effects, dermal toxicity, and carcinogenic effects) (Van den Berg et al., 1998; Pan et al., 2013; Zhou et al., 2018). Consequently, MSWIs have garnered increasing attention as significant sources of PCDD/Fs emissions. Despite China's adoption of the more stringent GB18485-2014 emission standard, MSWIs still face public opposition due to not-in-my-backyard (NIMBY) concerns (Sun et al., 2017). Therefore, continued monitoring and regulating PCDD/Fs emissions from MSWIs remain essential to mitigate their adverse impacts on public health and

* Corresponding author.
E-mail address: hanjinglei@scies.org (J. Han).
[1] These authors contributed equally to this work and should be listed as first author.

https://doi.org/10.1016/j.wasman.2023.10.016
Received 3 April 2023; Received in revised form 5 October 2023; Accepted 20 October 2023

the environment.

Accurately analyzing of PCDD/Fs involves substantial labor and financial expense, making it essential to identify indicator congeners capable of effectively quantifying and predicting toxic equivalent (TEQ) levels of PCDD/Fs emissions from MSWIs. Zhu et al. (2018) employed the linear regression (LR) model to conclude that 1,2,3,4,7,8,9-HpCDF, OCDF could serve as suitable indicator congeners for the prediction of PCDD/Fs TEQ levels based on the analysis of 57 flue gas samples. Li et al. (2017) utilized LR models and analyzed 36 flue gas samples to identify 1,2,3,7,8,9-HxCDF and 1,2,3,4,7,8-HxCDF as indicator congeners for the prediction of PCDD/Fs TEQ levels. The aforementioned studies have primarily relied on the LR model to establish relationship between congeners and PCDD/Fs with relatively sample size. However, the intricate relationship between congeners and PCDD/Fs TEQ levels may limit the effectiveness of the LR model in capturing this complexity. Furthermore, the impact of limited data on statistical power and applicability emphasizes the necessity for a model trained with a comprehensive dataset capable of predicting PCDD/Fs emission. Machine Learning (ML) presents another popular alternative for predictive modeling (Murphy, 2012). Artificial neural networks (ANN) and random forest (RF) were two prominent ML methods known for their ability to uncover intricate relationships between input and output variables. These models have shown promising performance in forecasting various pollutants such as trihalomethane, heavy metals, particulate matter (PM_{10} and $PM_{2.5}$), polycyclic aromatic hydrocarbons (PAHs) and water quality constituents (e.g., total nitrogen (TN), total phosphorus (TP) and total suspended solids (TSS)) in previous studies (Behrouz et al., 2022; Leng et al., 2017; Voukantsis et al., 2011; Wu et al., 2013; Xu et al., 2022). Bunsan et al. (2013) were among the first to apply ANN to model PCDD/Fs emission from an MSWI, utilizing variables such as hydrogen chloride and temperature in the mixing chamber. However, the challenge of establishing a highly accurate and robust model for predicting PCDD/Fs emission remains a pressing question.

In this study, we gathered 190 samples from four representative large-scale MSWIs in China. The aims were to explore characteristics of PCDD/Fs emissions in the flue gas, as well as evaluate potential indicator congeners for PCDD/Fs prediction. Subsequently, an extensive comparative analysis was conducted between the traditional method (LR) and two ML approaches, namely, back propagation (BP) ANN and RF, to determine the most appropriate prediction model for PCDD/Fs TEQ emissions. Furthermore, the model's applicability to ensure robust and reliable results was validated.

2. Materials and methods

2.1. Sample collection

A total of 190 flue gas samples were collected from four MSWIs in Southern China (depicted in Supplementary) using automatic isokinetic sampling system (TECORA, Italy). This system consists of a heated titanium probe, a sampling box equipped with a quartz fiber filter and a water-cooled trap. To capture particle-bound ΣPCDD/Fs and vapor-phase ΣPCDD/Fs, quartz fiber filters and Amberlite XAD-2 adsorbent resin (Supleco, Bellefonte, PA) were utilized (Mei et al., 2015). Prior to sampling, resins were spiked with 1000 pg/sample $^{13}C_{12}$-labeled Sampling Surrogate Standard (EPA 23 SSS) (Wellington Laboratories, Guelph, Canada) to assess the sampling campaigns. At each sampling site, three consecutive samples were collected, with each sample being collected for a minimum duration of 2 h. All samples were carefully wrapped in aluminum foil, sealed in polyethylene bags, and then stored at −20℃ before analysis.

2.2. PCDD/Fs extraction and analysis

Prior to pretreatment, 1000 pg of $^{13}C_{12}$ -labeled PCDD/Fs, serving as internal standards (EPA 23 IS), were added to the flue gas samples. The extraction of PCDD/Fs were performed using a mixture of dichloromethane and hexane (1:1, *v*/*v*) in a Soxhlet extractor for 24 h. The extracts were concentrated to approximately 1 mL and subjected to purification through a multilayer column, comprising anhydrous sodium sulfate, florisil, alkaline alumina, 33 % sodium hydroxide silica, neutral silica gel, 40 % sulfuric acid silica and anhydrous sodium sulfate from the bottom to the top. The final elute was further concentrated to approximately 20 μL under a gentle nitrogen stream and spiked with 400 pg $^{13}C_{12}$-labeled ΣPCDD/F injection standards, containing $^{13}C_{12}$-1,2,3,4-TCDD and $^{13}C_{12}$-1,2,3,7,8,9-HxCDD.

Seventeen PCDD/Fs congeners were separated using a high-resolution gas chromatography (HRGC) (Agilent 6890 N, Agilent Technologies, Santa Clara, CA) equipped with a DB-5MS (60 m × 0.25 mm × 0.25 μm) capillary column and analyzed with a high-resolution mass spectrometry (HRMS) (Micromass Autospec Ultima, Waters, Milford, MA). The gas chromatography (GC) temperature program was initiated at 140 °C (held for 2 min), and then it was increased to 220 °C at a rate of 8 °C/min. Subsequently, there was a further increase at a rate of 1.4 °C/min to 260 °C, followed by a final ramp to 320 °C (held for 4 min). The ion source temperature was set at 250 °C with an ionization energy of 35 eV. The HRMS was operated in the selected ion monitoring (SIM) mode, and the mass resolution was tuned to be over 10,000.

2.3. Quality assurance

A comprehensive quality assurance and quality control procedure was implemented in each batch of 10 field samples, comprising procedural blanks and ongoing precision recovery assessments. The recovery rates of the $^{13}C_{12}$ -labeled PCDD/Fs (EPA 23 SSS and EPA 23 IS) were carefully evaluated, showing ranges of 73 % to 102 % and 47 % to 120 %, respectively. The limits of detection (LOD) for PCDD/Fs in the flue gas samples were meticulously determined, with values ranging from 0.0015 to 0.036 ng/m^3 (Table S1). It is worth noting that procedural blanks exhibited PCDD/F levels below the LODs. To ensure the utmost accuracy, all the reported results were blank subtracted and corrected with the recoveries of spiked standards.

2.4. Statistical analysis

Statistical differences were assessed using the non-parametric Mann-Whitney *U* test. The normality of the data was examined using the Shapiro-Wilk (S-W) test. Pearson correlation was utilized for data conforming to a normal distribution, while Spearman correlation was chosen for non-normally distributed data. All statistical analyses were conducted using SPSS Statistics 29.0.

2.5. Development of prediction models

Indicator congeners were confirmed using Pearson correlation analysis, and to meet the assumption of normal distribution, the raw data was log-transformed before performing the correlation analysis. For those congeners that did not follow a normal distribution even after Log transformation, Spearman correlation analysis was chosen.

2.5.1. Linear regression (LR) model

LR model is a common choice for predicting PCDD/Fs TEQ levels, as demonstrated by several studies (Li et al., 2018, 2017; Oh et al., 2004; Zhu et al., 2018). The general liner regression equation, depicted Eq. (1), enables researchers to utilize selected independent variables to make predictions concerning outcomes.

$$PCDD/Fs TEQlevels = A \times indicatorcongener + B \tag{1}$$

2.5.2. Back propagation (BP) ANN model

ANN models have gained popularity for establishing reliable mapping relationship prediction model (Wu et al., 2020), with BP ANN being among the most widely used variants. Similar to ANN, the BP ANN structure comprises three layers: input, hidden, and output. The weights that describe the connection between three layers were adjusted during training to minimize the mean squared error (MSE) between the network's estimated and measured data (Mohamed, 2019). BP neural network effectively overcomes the local minimum defects of the original neural network and offers benefits such as rapid convergence and high accuracy. In this study, a BP ANN model for predicting PCDD/Fs TEQ levels was developed using the 'newrb' function (net = newff(inputn, outputn, hiddennum)) within MATLAB software (R2021b). The number of nodes in the hidden layer played a crucial role in the BP ANN model and was determined using the empirical formula (Eq. (2)).

$$Hiddenlayernodesnumber = \sqrt{(M+N} + A \tag{2}$$

where M represents the number of nodes in the input layer, N is the number of nodes in the output layer, and A is typically an integer ranging between 1 and 10.

2.5.3. RF model

RF is a potent machine learning model proposed by Breiman. (2001), designed specifically for regression problems. The fundamental principle of RF involves combining a set of binary decision trees. Each tree is constructed using a bootstrap sample extracted from the learning dataset, along with a randomly selected subset of features (input variables or predictors) at each node. The m_{try} (the number of explanatory variables sampled for splitting at each node) and n_{trees} (the number of trees grown) were the two key hyperparameters (Li et al., 2020). Our results indicated that an optimized combination of $m_{try} = 2$ and $n_{trees} = 10$ resulted in minimized errors across all indicators. The predictions from all individual decision trees are aggregated to make the final prediction. A notable advantage of the RF model lies in its capability to limit overfitting without substantially reducing prediction accuracy, as demonstrated in other studies (Zhou et al., 2020).

2.6. Evaluation criterion of the models

Four evaluation criteria of models were selected to comprehensively assess the prediction performance, considering insights from prior research (Lin et al., 2020; Chen et al., 2021). These criteria include the absolute relative error (ARE), correlation coefficient (R^2), root mean square error (RMSE) and mean absolute error (MAE). Superior prediction accuracy is indicated by high R^2 values and lower ARE, RMSE, and MAE values. The calculations for these criteria can be performed using Eqs. (3–6).

$$ARE = \left| \frac{PCDD/FsTEQ_{predicted} - PCDD/FsTEQ_{actual}}{PCDD/FsTEQ_{predicted}} \right| \tag{3}$$

$$R^2 = 1 - \frac{\sum_N (PCDD/FsTEQ^i_{predicted} - PCDD/FsTEQ^i_{actual})^2}{\sum_N (PCDD/FsTEQ^i_{predicted} - \overline{PCDD/FsTEQ^i_{actual}})^2} \tag{4}$$

$$RMSE = \sqrt{\frac{\sum_N (PCDD/FsTEQ^i_{predicted} - PCDD/FsTEQ^i_{actual})^2}{N}} \tag{5}$$

$$MAE = \frac{\sum_N \left| PCDD/FsTEQ^i_{predicted} - PCDD/FsTEQ^i_{actual} \right|}{N} \tag{6}$$

where N is the number of samples, PCDD/Fs $TEQ_{predicted}$ and PCDD/Fs TEQ_{actual} denoted the predicted and observed PCDD/Fs TEQ values, respectively.

3. Results and discussion

3.1. Characteristics of PCDD/Fs emissions

PCDD/Fs emissions of 4 typical large-scale MSWIs were extensively studied in 2018, and the detailed data were presented in Table S2 and Table S3. The average PCDD/Fs levels and TEQ levels were 0.987 ng/m^3 (0.108 to 5.188 ng/m^3) and 0.030 ng TEQ/m^3 (0.005 to 0.099 ng TEQ/m^3), respectively. Notably, there was no significant difference between the levels of PCDD/Fs congeners and similar statistical results were noted for TEQ levels. Compared to previous studies conducted in China before the 2014 promulgation of the new emission standard (Chen et al., 2017; Liu et al., 2013; Ni et al., 2009; Shi et al., 2008), we observed relatively lower PCDD/Fs levels (Table S4). This suggests that the advanced technology implemented in the large-scale MSWI's combustion chamber effectively controls PCDD/Fs emissions. Another contributing to factor for the reduced PCDD/Fs levels could be the full implementation of MSW classification, as specific preconditions can efficiently improve the heat value of incinerated MSW, thereby significantly reducing the formation of PCDDs/Fs in the combustion chambers (Shi et al., 2008; Zhang et al., 2015).

The proportion of each congener to PCDD/Fs levels in flue gas was shown in Table S5. The dominant congeners were 1,2,3,4,6,7,8-HpCDD and OCDD ($p < 0.05$), accounting for 66.69 % of the total concentration of PCDD/Fs in flue gas. Regarding TEQ levels, 2,3,4,7,8-PeCDF emerged as the most significant toxic congener, contributing to 29.13 % of total PCDD/Fs TEQ levels. These results were consistent with previous studies (Liu et al., 2013; Coutinho et al., 2015), reinforcing the critical role of 2,3,4,7,8-PeCDF in PCDD/Fs control.

3.2. Correlation analysis between congeners and PCDD/Fs TEQ levels

The S-W normality test revealed that log transformation of 1,2,3,7,8-PeCDF, 2,3,4,7,8-PeCDF, 1,2,3,4,7,8-HxCDF, 1,2,3,6,7,8-HxCDF, 2,3,4,6,7,8-HxCDF, 1,2,3,4,6,7,8-HpCDF, 1,2,3,4,7,8,9-HpCDF, and PCDD/Fs TEQ levels followed a normal distribution. Pearson correlation analysis demonstrated significant correlations ($p < 0.05$) between these congeners and PCDD/Fs TEQ levels (Table S6). However, for the remaining congeners (2,3,7,8-TCDF, 1,2,3,7,8,9-HxCDF, OCDF, 2,3,7,8-TCDD 1,2,3,7,8-PeCDD, 1,2,3,4,7,8-HxCDD, 1,2,3,6,7,8-HxCDD, 1,2,3,7,8,9-HxCDD, 1,2,3,4,6,7,8-HpCDD, OCDD), which did not conform to a normal distribution, Spearman correlation analysis was employed. The results demonstrated significant correlations ($p < 0.05$) between these congeners and PCDD/Fs TEQ levels as well (Table S6). Based on the strength of their correlations with the PCDD/Fs TEQ levels, we selected 2,3,4,7,8-PeCDF ($R^2 = 0.914$), 1,2,3,6,7,8-HxCDF ($R^2 = 0.908$) and 2,3,4,6,7,8-HxCDF ($R^2 = 0.904$) as the most robust indicators for predicting PCDD/Fs TEQ levels. Li et al. (2017) and Zhu et al. (2018) employed the LR model to conclude that congeners (1,2,3,7,8,9-HxCDF, 1,2,3,4,7,8-HxCDF, 1,2,3,4,7,8,9-HpCDF, OCDF) could serve as suitable indicators for prediction. The outcome in this study was inconsistent with those of previous studies, which may be attributed to the sample size.

3.3. Modelling key indicators for PCDD/Fs TEQ levels

The indicators for predicting PCDD/Fs TEQ levels in this study were selected based on a dataset comprising 190 samples. Out of these, 150 datasets (No. 1–150) were utilized for training the predictive models, while the remaining 40 datasets (No. 151–190) were reserved for testing the models' performance. The input variables used for predicting PCDD/Fs TEQ levels were three candidates (2,3,4,7,8-PeCDF, 1,2,3,6,7,8-HxCDF and 2,3,4,6,7,8-HxCDF), chosen due to their strong correlations with PCDD/Fs TEQ levels. The prediction models based on the linear regression (LR) were expressed mathematically as Eqs. (7–9).

Table 1
The evaluation parameters of the models in training datasets.

Model	Parameter	2,3,4,7,8-PeCDF	1,2,3,6,7,8-HxCDF	2,3,4,6,7,8-HxCDF	Average
LR	R^2	0.890	0.920	0.921	0.910
	MAE	0.145	0.145	0.148	0.146
	RMSE	0.187	0.185	0.193	0.188
	ARE	14.6 %	14.5 %	14.8 %	14.6 %
BP ANN	R^2	0.896	0.925	0.935	0.919
	MAE	0.142	0.130	0.129	0.134
	RMSE	0.181	0.165	0.165	0.170
	ARE	14.2 %	13.0 %	12.9 %	13.4 %
RF	R^2	0.941	0.968	0.970	0.960
	MAE	0.090	0.089	0.086	0.088
	RMSE	0.136	0.117	0.123	0.125
	ARE	9.00 %	8.90 %	8.60 %	8.80 %

Table 2
The evaluation parameters of the models in test datasets.

Model	Parameter	2,3,4,7,8-PeCDF	1,2,3,6,7,8-HxCDF	2,3,4,6,7,8-HxCDF	Average
LR	R^2	0.837	0.910	0.848	0.865
	MAE	0.180	0.185	0.196	0.187
	RMSE	0.219	0.250	0.227	0.232
	ARE	18.0 %	18.5 %	19.6 %	18.7 %
BP ANN	R^2	0.835	0.908	0.847	0.863
	MAE	0.172	0.181	0.189	0.181
	RMSE	0.211	0.236	0.222	0.223
	ARE	17.2 %	18.1 %	18.9 %	18.1 %
RF	R^2	0.913	0.928	0.929	0.923
	MAE	0.125	0.126	0.126	0.126
	RMSE	0.149	0.157	0.148	0.151
	ARE	12.5 %	12.6 %	12.6 %	12.6 %

$$\mathrm{Log10(PCDD/Fs\ TEQ)} = 0.919 \times \mathrm{Log10(2,3,4,7,8\text{-}PeCDF)} + 0.086 \quad (7)$$

$$\mathrm{Log10(PCDD/Fs\ TEQ)} = 0.896 \times \mathrm{Log10(1,2,3,6,7,8\text{-}HxCDF)} - 0.048 \quad (8)$$

$$\mathrm{Log10(PCDD/Fs\ TEQ)} = 0.826 \times \mathrm{Log10(2,3,4,6,7,8\text{-}HxCDF)} - 0.291 \quad (9)$$

Regarding the use of indicator congeners for predicting PCDD/Fs TEQ levels, previous studies primarily emphasized the accuracy and precision of predictive capabilities using R-squared values (Li et al., 2017; Zhu et al., 2018). In this study, the accuracy of the proposed methods was thoroughly evaluated through a combination of statistical analysis (Table 1 and Table 2) and graphical assessment (Fig. 1) on the training and test datasets. Across both sets, the congeners 2,3,4,7,8-PeCDF, 1,2,3,6,7,8-HxCDF, 2,3,4,6,7,8-HxCDF consistently exhibited low $RMSE_{train}$, $RMSE_{test}$, MAE_{train}, MAE_{test} values, along with high R^2_{train}, R^2_{test} value.

No significant differences were observed among the evaluation parameters of these indicator congeners ($p > 0.05$). Furthermore, the vast majority of absolute error percentages fell within the range of 0 and 30 %. Hence, these results clearly demonstrate that the inclusion of 2,3,4,7,8-PeCDF, 1,2,3,6,7,8-HxCDF, and 2,3,4,6,7,8-HxCDF as variables in LR, BP ANN, and RF models leads to accurate and reliable predictions of PCDD/Fs TEQ levels.

However, discernible differences were evident among the models. The mean $ARE_{training}$ values for LR, BP ANN and RF were 13.4 %, 14.7 %, and 8.8 %. Similarly, the average ARE_{test} values were 18.1 %, 18.7 % and 12.6 % for LR, BP ANN, and RF models, respectively. This suggested that the prediction performances of ML models exhibited a substantial superiority over the LR analysis. Considering the comprehensive evaluation of other parameters, the RF model stood out by showcasing exceptional performance with remarkable metrics. It achieved an average $RMSE_{train}$ of 0.125, $RMSE_{test}$ of 0.151, MAE_{train} of 0.088, MAE_{test} of 0.126, R^2_{train} of 0.960, and R^2_{test} of 0.923. These compelling results provided strong confirmation that the RF model was better suited for predicting the toxicity of PCDD/Fs in the flue gas emitted from the studied MSWIs.

3.4. The applicability of indicator congeners for PCDD/Fs emissions

The previously identified effective indicators for PCDD/Fs prediction were 1,2,3,4,7,8,9-HpCDF, OCDF, 1,2,3,4,7,8-HxCDF and 1,2,3,7,8,9-HxCDF, expressed by the LR prediction equations shown in Table 3 (Li et al., 2017; Zhu et al., 2018). The aforementioned studies did not evaluate the applicability of indicator congeners. When testing the applicability of these equations with our datasets, nearly 100 % mean percentage errors between predicted data and actual results were observed. This may be attributed to the limited sample size used to obtain these equations (Table S7). As a result, we proposed the use of the best ML model (RF model) for predicting PCDD/Fs TEQ levels, based on three indicator congeners (2,3,4,7,8-PeCDF, 1,2,3,6,7,8-HxCDF, and 2,3,4,6,7,8-HxCDF) obtained from a relatively large sample size.

To validate the practicality of these indicator congeners and selected model in predicting PCDD/Fs TEQ levels, we randomly selected 32 flue gas samples (Group A, our unpublished data) from other large-scale MSWIs in China, 14 flue gas samples (Group B) from Liu et al.'s (2013) study and 8 flue gas samples (Group C) from Wang et al.'s (2022) study. Fig. 2 displayed the ARE values of the three indicator congeners obtained from the RF model. For 2,3,4,7,8-PeCDF, 1,2,3,6,7,8-HxCDF and 2,3,4,6,7,8-HxCDF, the median ARE values were 12.3 %, 13.0 %, 12.8 %, respectively, with percentages of predictions having ARE < 30 % were 92.6 %, 90.7 %, 87.0 %. This indicates that using these three indicator congeners in the RF model is practical for predicting PCDD/Fs TEQ levels.

Additionally, PCDD/Fs TEQ levels of 15 flue gas samples in Groups A, B, and C, ranged from 0.110 to 2.810 ng TEQ/m^3, were classified as the high PCDD/F TEQ level group, while the remaining flue gas samples in Groups A, B, and C (<0.1 ng TEQ/m^3) were categorized as the normal PCDD/F TEQ level group. Median ARE values for the three indicator congeners (2,3,4,7,8-PeCDF, 1,2,3,6,7,8-HxCDF, and 2,3,4,6,7,8-

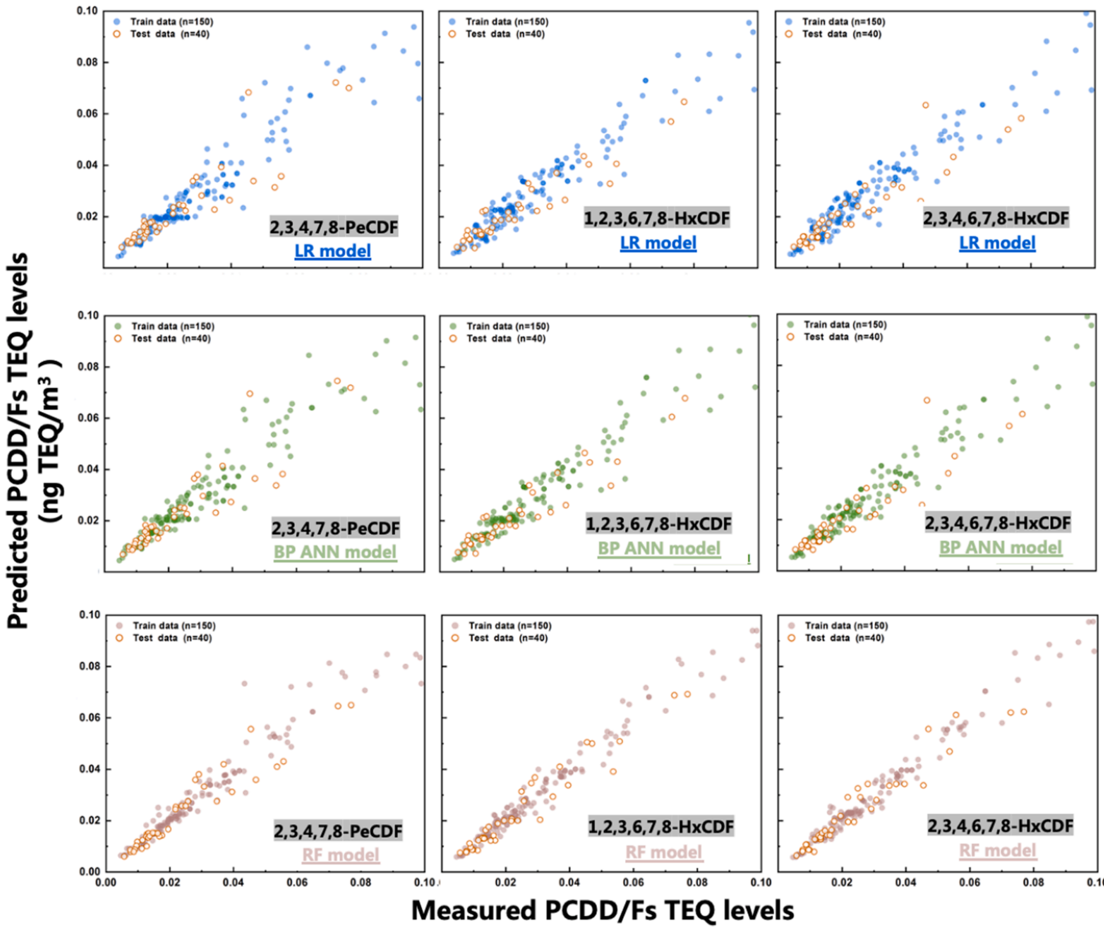

Fig. 1. The scatter plots of measured and predicted PCDD/Fs TEQ levels in train and test data based on the LR, BP ANN and RF model.

Table 3
The prediction models of PCDD/Fs TEQ from other studies.

Dependent variable	Slope	Independent variable	y-intercept	Reference
PCDD/Fs TEQ	0.0096	1,2,3,4,7,8,9-HpCDF	0.00002244	(Zhu et al., 2018)
	0.000989	OCDF	0.00000058	
	0.7766	1,2,3,4,7,8-HxCDF	0.1544	(Li et al., 2017)
	0.6377	1,2,3,7,8,9-HxCDF	0.1572	

HxCDF) were 17.9 %, 15.3 % and 13.8 % for the high PCDD/F TEQ value group, higher than those of the normal PCDD/F TEQ value group (10.9 %, 10.9 % and 12.6 %) (Fig. 3). This suggests that the accuracy of our method may be compromised when predicting high PCDD/F TEQ levels, which could be associated with the differences in the dominant formation mechanism between high and low emissions resulted from the distinct chlorine content in the organic matter and chlorine-containing components in waste (Xiong et al., 2021). In predicting PCDD/Fs TEQ levels in flue gas from MSWIs using indicator congeners, our research employed ML models, evaluation criteria (e.g., ARE, R^2, RMSE and MAE) and assessment of the applicability of models, demonstrating enhanced reliability compared to prior related research. In general, our results indicate that 2,3,4,7,8-PeCDF, 1,2,3,6,7,8-HxCDF and 2,3,4,6,7,8-HxCDF are suitable for predication and have strong resistance to disturbance when used in the RF model.

4. Conclusion

The indicator congeners (2,3,4,7,8-PeCDF, 1,2,3,6,7,8-HxCDF, and 2,3,4,6,7,8-HxCDF) were identified through Pearson analysis of 190 samples. The predictive performance of these indicators was compared using traditional method (LR) and machine learning methods (BP ANN and RF) for PCDD/Fs TEQ levels. Among these models, the RF method demonstrated superior performance over LR and BP ANN. To validate the practicality of the identified indicators and the RF model for predicting PCDD/Fs TEQ emissions, we tested the model with an additional 54 flue gas samples (32 samples determined in our other study and 22 samples from other studies). The results further confirmed the widespread applicability of these indicator congeners in predicting PCDD/Fs TEQ emissions. The application of advanced machine learning models in this research has significant potential for reducing the financial costs associated with waste management. Moreover, these predictive models can streamline labor and time-intensive sampling and analysis processes, providing valuable support for more informed and effective policy-making decisions. As the field of machine learning continues to advance, the findings of this study hold promise for optimizing waste management practices and promoting a more sustainable future.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

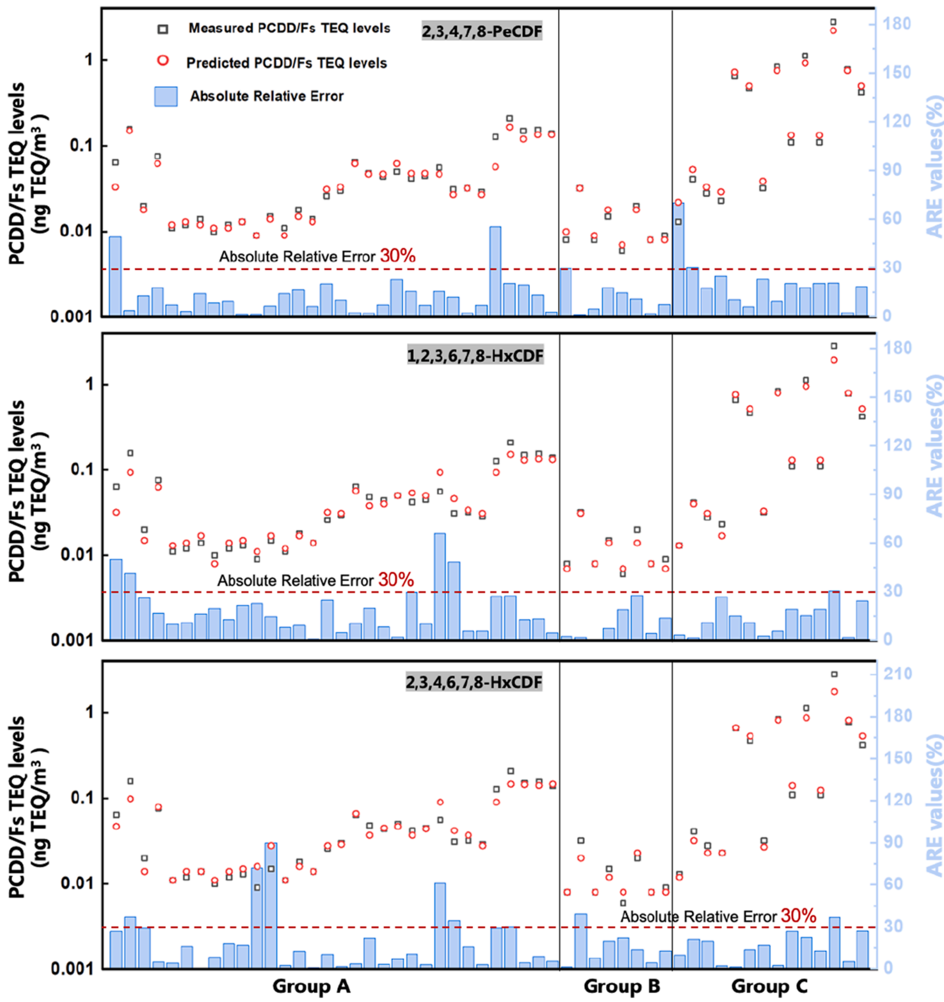

Fig. 2. ARE values of three indicator congeners based on the RF model with other studies' data. Data in group B was from Liu et al., 2013 and data in group C was from Wang et al., 2022.

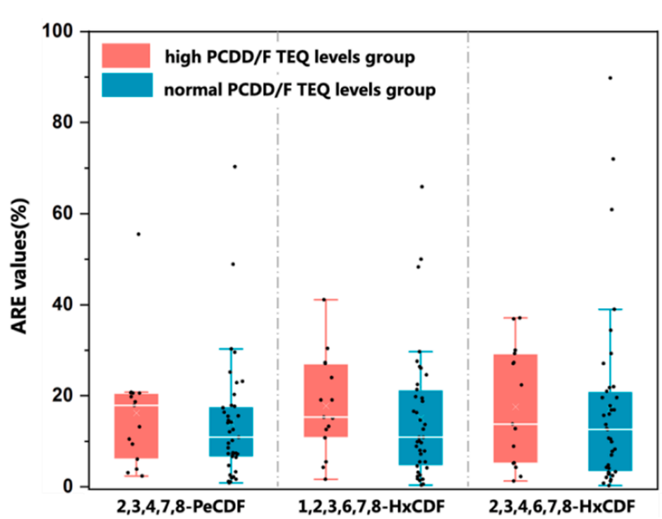

Fig. 3. Comparison between ARE values of high and normal PCDD/Fs TEQ levels group.

Data availability

Data will be made available on request.

Acknowledgment

The research was supported by Science and Technology Program of Shenzhen, China (KCXFZ20201221173402007).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wasman.2023.10.016.

References

Behrouz, M.S., Yazdi, M.N., Sample, D.J., 2022. Using Random Forest, a machine learning approach to predict nitrogen, phosphorus, and sediment event mean concentrations in urban runoff. J. Environ. Manage. 317, 115412. https://doi.org/10.1016/j.jenvman.2022.115412.

Bunsan, S., Chen, W.-Y., Chen, H.-W., Chuang, Y.H., Grisdanurak, N., 2013. Modeling the dioxin emission of a municipal solid waste incinerator using neural networks. Chemosphere 92, 258–264. https://doi.org/10.1016/j.chemosphere.2013.01.083.

Chen, K., Peng, Y., Lu, S., Lin, B., Li, X., 2021. Bagging based ensemble learning approaches for modeling the emission of PCDD/Fs from municipal solid waste incinerators. Chemosphere 274, 129802. https://doi.org/10.1016/j.chemosphere.2021.129802.

Chen, P., Xiao, X., Mei, J., Cai, Y., Tang, Y., Peng, P., 2017. Characteristic accumulation of PCDD/Fs in pine needles near an MSWI and emission levels of the MSWI in pearl river delta: A case study. Chemosphere 181, 360–367. https://doi.org/10.1016/j.chemosphere.2017.04.098.

Cheng, H., Hu, Y., 2010. Municipal solid waste (MSW) as a renewable source of energy: Current and future practices in China. Bioresour. Technol. 101, 3816–3824. https://doi.org/10.1016/j.biortech.2010.01.040.

Coutinho, M., Albuquerque, M., Silva, A.P., Rodrigues, J., Borrego, C., 2015. Long-time monitoring of polychlorinated dibenzo-p-dioxins and dibenzofurans over a decade in the ambient air of Porto, Portugal. Chemosphere 137, 207–213. https://doi.org/10.1016/j.chemosphere.2015.07.052.

Domingo, J.L., Rovira, J., Vilavert, L., Nadal, M., Figueras, M.J., Schuhmacher, M., 2015. Health risks for the population living in the vicinity of an Integrated Waste Management Facility: Screening environmental pollutants. Sci. Total Environ. 518–519, 363–370. https://doi.org/10.1016/j.scitotenv.2015.03.010.

Leng, X., Wang, J., Ji, H., Wang, Q., Li, H., Qian, X., Li, F., Yang, M., 2017. Prediction of size-fractionated airborne particle-bound metals using MLR, BP-ANN and SVM analyses. Chemosphere 180, 513–522. https://doi.org/10.1016/j.chemosphere.2017.04.015.

Li, R., Cui, L., Liang, J., Zhao, Y., Zhang, Z., Fu, H., 2020. Estimating historical SO2 level across the whole China during 1973–2014 using random forest model. Chemosphere 247, 125839. https://doi.org/10.1016/j.chemosphere.2020.125839.

Li, J., Lv, Z., Du, L., Li, X., Hu, X., Wang, C., Niu, Z., Zhang, Y., 2017. Emission characteristic of polychlorinated dibenzo-p-dioxins and polychlorinated dibenzofurans (PCDD/Fs) from medical waste incinerators (MWIs) in China in 2016: A comparison between higher emission levels of MWIs and lower emission levels of MWIs. Environ. Pollut. 221, 437–444. https://doi.org/10.1016/j.envpol.2016.12.009.

Li, J., Zhang, Y., Sun, T., Hao, H., Wu, H., Wang, L., Chen, Y., Xing, L., Niu, Z., 2018. The health risk levels of different age groups of residents living in the vicinity of municipal solid waste incinerator posed by PCDD/Fs in atmosphere and soil. Sci. Total Environ. 631–632, 81–91. https://doi.org/10.1016/j.scitotenv.2018.03.009.

Lin, H., Dai, Q., Zheng, L., Hong, H., Deng, W., Wu, F., 2020. Radial basis function artificial neural network able to accurately predict disinfection by-product levels in tap water: Taking haloacetic acids as a case study. Chemosphere 248, 125999. https://doi.org/10.1016/j.chemosphere.2020.125999.

Liu, W., Tian, Z., Li, H., Xie, H., Xiao, K., Li, C., Tang, C., Zheng, M., 2013. Mono-to octa-chlorinated PCDD/Fs in stack gas from typical waste incinerators and their implications on emission. Environ. Sci. Tech. 47, 9774–9780. https://doi.org/10.1021/es402136s.

Mei, W., Guorui, L., Xiaoxu, J., Wenbin, L., Li, L., Sumei, L., Minghui, Z., Jiayu, Z., 2015. Brominated dioxin and furan stack gas emissions during different stages of the secondary copper smelting process. Atmospheric Pollut. Res. 6, 464–468. https://doi.org/10.5094/APR.2015.051.

Mohamed, Z.E., 2019. Using the artificial neural networks for prediction and validating solar radiation. J. Egypt. Math. Soc. 27, 47. https://doi.org/10.1186/s42787-019-0043-8.

Murphy, K.P., 2012. Machine Learning: A Probabilistic Perspective. MIT Press.

National Bureau of Statistics of China, 2013. Collection, transport and disposal of consumption waste in cities. Available at: http://www.stats.gov.cn/sj/ndsj/2013/indexch.htm.

Ni, Y., Zhang, H., Fan, S., Zhang, X., Zhang, Q., Chen, J., 2009. Emissions of PCDD/Fs from municipal solid waste incinerators in China. Chemosphere 75, 1153–1158. https://doi.org/10.1016/j.chemosphere.2009.02.051.

Oh, J.-E., Touati, A., Gullett, B.K., Mulholland, J.A., 2004. PCDD/F TEQ indicators and their mechanistic implications. Environ. Sci. Tech. 38, 4694–4700. https://doi.org/10.1021/es034997s.

Pan, Y., Yang, L., Zhou, J., Liu, J., Qian, G., Ohtsuka, N., Motegi, M., Oh, K., Hosono, S., 2013. Characteristics of dioxins content in fly ash from municipal solid waste incinerators in China. Chemosphere 92, 765–771. https://doi.org/10.1016/j.chemosphere.2013.04.003.

Rovira, J., Vilavert, L., Nadal, M., Schuhmacher, M., Domingo, J.L., 2015. Temporal trends in the levels of metals, PCDD/Fs and PCBs in the vicinity of a municipal solid waste incinerator. Preliminary assessment of human health risks. Waste Manag. 43, 168–175. https://doi.org/10.1016/j.wasman.2015.05.039.

Shi, D.-Z., Wu, W.-X., Lu, S.-Y., Chen, T., Huang, H.-L., Chen, Y.-X., Yan, J.-H., 2008. Effect of MSW source-classified collection on the emission of PCDDs/Fs and heavy metals from incineration in China. J. Hazard. Mater. 153, 685–694. https://doi.org/10.1016/j.jhazmat.2007.09.026.

Sun, C., Meng, X., Peng, S., 2017. Effects of waste-to-energy plants on china's urbanization: Evidence from a hedonic price analysis in shenzhen. Sustainability 9, 475. https://doi.org/10.3390/su9030475.

Van den Berg, M., Birnbaum, L., Bosveld, A.T., Brunström, B., Cook, P., Feeley, M., Giesy, J.P., Hanberg, A., Hasegawa, R., Kennedy, S.W., Kubiak, T., Larsen, J.C., van Leeuwen, F.X., Liem, A.K., Nolt, C., Peterson, R.E., Poellinger, L., Safe, S., Schrenk, D., Tillitt, D., Tysklind, M., Younes, M., Waern, F., Zacharewski, T., 1998. Toxic equivalency factors (TEFs) for PCBs, PCDDs, PCDFs for humans and wildlife. Environ. Health Perspect. 106 (12), 775–792.

Voukantsis, D., Karatzas, K., Kukkonen, J., Räsänen, T., Karppinen, A., Kolehmainen, M., 2011. Intercomparison of air quality data using principal component analysis, and forecasting of PM10 and PM2. 5 concentrations using artificial neural networks, in Thessaloniki and Helsinki. Sci. Total Environ. 409 (7), 1266–1276.

Wang, P., Yan, F., Cai, J., Xie, F., Shen, X., Wei, X., Yang, G., Zhong, R., Huang, J., Li, Z., Zhang, Z., 2022. Emission levels and phase distributions of PCDD/Fs in a full-scale municipal solid waste incinerator: The impact of wet scrubber system. J. Clean. Prod. 337, 130468. https://doi.org/10.1016/j.jclepro.2022.130468.

Wu, G., Kechavarzi, C., Li, X., Wu, S., Pollard, S.J.T., Sui, H., Coulon, F., 2013. Machine learning models for predicting PAHs bioavailability in compost amended soils. Chem. Eng. J. 223, 747–754. https://doi.org/10.1016/j.cej.2013.02.122.

Wu, D., Zhang, D., Liu, S., Jin, Z., Chowwanonthapunya, T., Gao, J., Li, X., 2020. Prediction of polycarbonate degradation in natural atmospheric environment of China based on BP-ANN model with screened environmental factors. Chem. Eng. J. 399, 125878. https://doi.org/10.1016/j.cej.2020.125878.

Xiong, S., Lu, S., Shang, F., Li, X., Yan, J., Cen, K., 2021. Online predicting PCDD/F emission by formation pathway identification clustering and Box-Cox transformation. Chemosphere 274, 129780. https://doi.org/10.1016/j.chemosphere.2021.129780.

Xu, Z., Shen, J., Qu, Y., Chen, H., Zhou, X., Hong, H., Sun, H., Lin, H., Deng, W., Wu, F., 2022. Using simple and easy water quality parameters to predict trihalomethane occurrence in tap water. Chemosphere 286, 131586. https://doi.org/10.1016/j.chemosphere.2021.131586.

Zhang, D., Huang, G., Xu, Y., Gong, Q., 2015. Waste-to-energy in China: Key challenges and opportunities. Energies 8, 14182–14196. https://doi.org/10.3390/en81212422.

Zhou, X., Lu, P., Zheng, Z., Tolliver, D., Keramati, A., 2020. Accident Prediction Accuracy Assessment for Highway-Rail Grade Crossings Using Random Forest Algorithm Compared with Decision Tree. Reliab. Eng. Syst. Saf. 200, 106931. https://doi.org/10.1016/j.ress.2020.106931.

Zhou, Q., Yang, J., Liu, M., Liu, Y., Sarnat, S., Bi, J., 2018. Toxicological risk by inhalation exposure of air pollution emitted from China's municipal solid waste incineration. Environ. Sci. Tech. 52, 11490–11499. https://doi.org/10.1021/acs.est.8b03352.

Zhu, F., Li, X., Lu, J.-W., Hai, J., Zhang, J., Xie, B., Hong, C., 2018. Emission characteristics of PCDD/Fs in stack gas from municipal solid waste incineration plants in Northern China. Chemosphere 200, 23–29. https://doi.org/10.1016/j.chemosphere.2018.02.092.

National Bureau of Statistics of China, 2020. Collection, transport and disposal of consumption waste in cities. Available at: http://www.stats.gov.cn/sj/ndsj/2020/indexch.htm.