



## Supplementary Materials

**Table S1.** Best subset regression at different size for up to six variables.

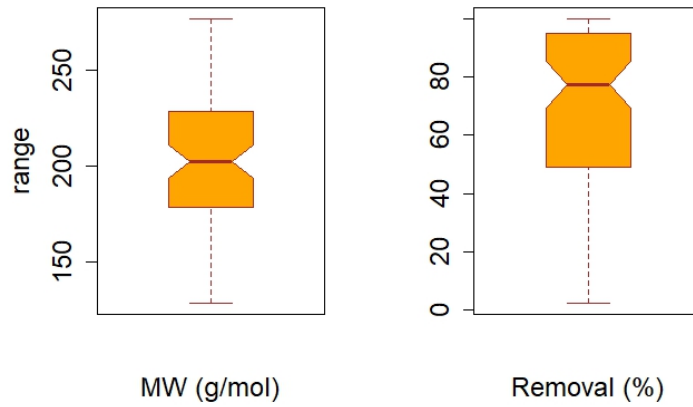
No. of variables	Temperature	Time	Soil/water ratio	SOM	MW	Solubility
1		√				
2	√		√			
3	√	√	√			
4	√	√	√		√	
5	√	√	√	√	√	
6	√	√	√	√	√	√

\* S/W ratio – soil/water ratio; SOM – soil organic matter; MW – molecular weight

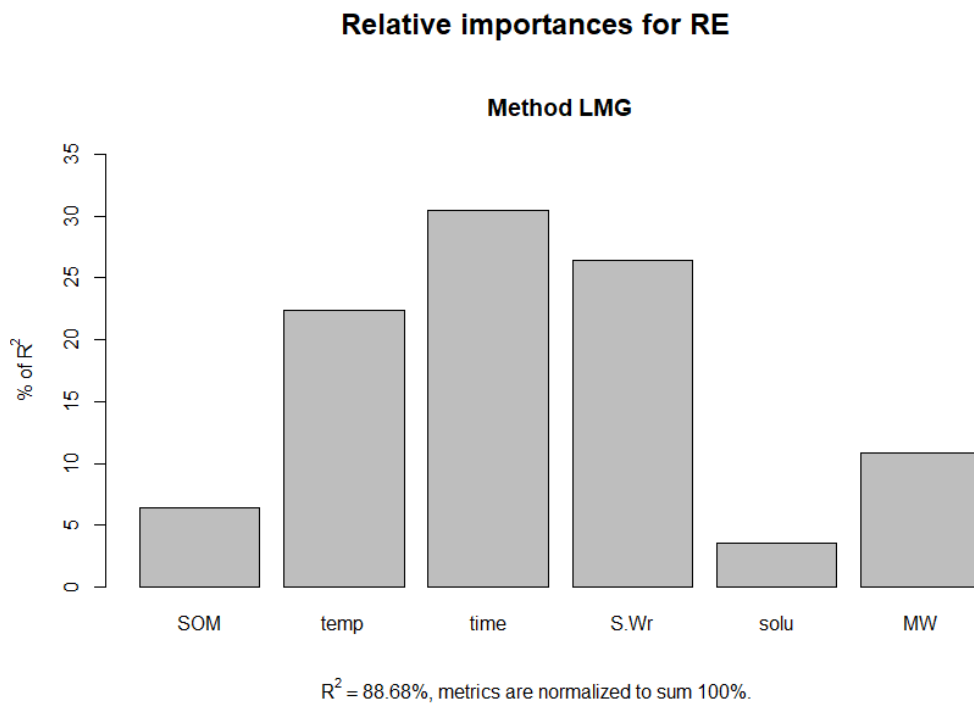
**Table S2.** Principal component analysis (PCA) results with the variables loading percentage.

Principle components	Eigenvalue	Variance %	Cumulative variance %	Main contributors
PC1	2.786	46.44	46.44	SOM 28.32% MW 19.49%
PC2	1.435	23.93	70.37	Time 51.64% S/W ratio 24.69%
PC3	0.947	15.79	86.16	Temperature 35.62%

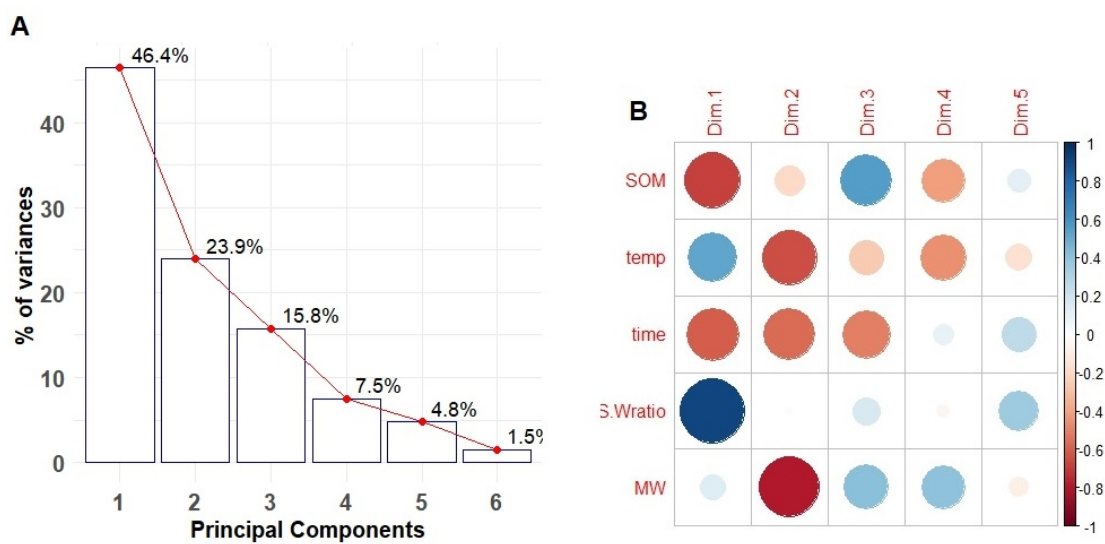
\* S/W ratio – soil/water ratio; SOM – soil organic matter; MW – molecular weight



**Fig. S1.** Data visualization of the range of the investigated polycyclic aromatic hydrocarbons (PAHs) (a) and their removal rate by the subcritical water extraction (SCWE) process (b).



**Fig. S2.** Relative contribution of the variables to the global model.



**Fig. S3.** (a) Principal component analysis (PCA) of the influence of the input variables on the prediction of the PAHs removal rate by the multiple linear regression (MLR) model and (b) variable loading of each PC. Collectively, the first three PCs explained 86% of the variance in the dataset.