International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 1, pp 15-25, January-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.401003
ISSN (Online): 2583-7257

An Approach to Determining the Partition Coefficients between Solid and Air Phases for Organic Compounds Using the QSPR

Anwer Abdulwahhab Khalaf

Ministry of Electricity, Iraq

***Abstract* - An important metric for evaluating the emission of chemicals from solid materials and the subsequent human exposure is the partition coefficient between the material and air (K_{ma}). Current correlations for calculating K_{ma} are typically limited to a narrow range of chemical-material combinations and do not account for temperature variations. This study aims to expand the prediction of K_{ma} to a broader range of chemical-material pairings using a quantitative structure property relationship (QSPR) approach. A dataset comprising 991 measured K_{ma} values for I179 chemical compounds across 22 distinct material categories was compiled. The prediction of K_{ma} was achieved through a multiple linear regression model that incorporates several key variables, including temperature, material type, enthalpy of vaporization (ΔHv), and the chemical partition coefficient between octanol and air (Koa). Internal and external validation confirmed that the model is robust, reliable, and exhibits strong predictive performance ($R^2ext > 0.78$). Additionally, the model demonstrates a high degree of fit to the experimental data, along with a modified R^2 of 0.93. Moreover, a generalized QSPR model has been created to forecast K_{ma} based solely on temperature and chemical properties, with an adjusted $R^2 = 0.84$, thus eliminating the need to specify a particular material type. These QSPR-based correlations will enable Rapid analysis predictions of human exposure to chemicals in materials, with particular applicability to furniture and building materials.**

Keywords: correlation, organic chemicals, consumer exposure, indoor release, solid materials, partitioning.

I. INTRODUCTION

One significant contributor to passive emissions in indoor environments, as well as the transfer of substances into household dust and onto skin, is the presence of chemicals incorporated into solid materials. [1] Common examples of such chemicals include flame retardants used in furniture and plasticizers present in construction materials. The dimensionless partition coefficient between material and air (K_{ma}), which quantifies the ratio of a chemical's concentration in The ratio of a solid material's concentration to its concentration in air at equilibrium, is a crucial parameter for evaluating The emission of these chemicals from solid materials and the consequent exposure to consumers.[2] The concentration of chemicals at the surface of the material plays a significant role in human exposure, influencing pathways such as inhalation, skin contact, and dust ingestion, as well as the transfer of chemicals from solid materials into both air and household dust. The K_{ma} Varies for each chemical-material combination and depends on temperature. Experimental methods, including sorption tests for polymeric materials and chamber measurements for construction materials For example, vinyl flooring, [3] gypsum board, cement, and plywood, have been employed to determine a limited set of K_{ma} values. Recent studies have also reported K_{ma} values for textiles and apparel. However, due to the significant costs and time involved in conducting these studies, available K_{ma} values are restricted to a narrow range of chemical-material combinations. [4-6] Consequently, There is a growing need for quantitative relationships that can predict the partition coefficient between the material and air (K_{ma}) based on known physicochemical properties, particularly for chemicals for which experimental data are unavailable. This need is especially critical for high-throughput methodologies, which require the evaluation of numerous chemical-material combinations. [7][8]

Several correlation methods have been created to estimate K_{ma} based on the physical and chemical properties of various substances. For instance, numerous studies have analyzed data on volatile organic compounds (VOCs) present in building materials to explore the relationship between K_{ma} and the vapor pressure of the involved chemicals [4][9-11]. Additionally, research has demonstrated a connection between the partition coefficient between octanol and air (K_{oa}) and K_{ma}, especially in studies examining semi-volatile organic compounds (SVOCs) using passive sampling techniques. Furthermore [5][6][12][13] Holmgren et al. calculated K_{ma} for six sets of materials based on five Abraham solvation

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 1, pp 15-25, January-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.401003
ISSN (Online): 2583-7257

parameters; however, these parameters are not readily accessible. A major drawback of the previously mentioned methods is their specificity for specific chemical classes and materials, such as polycyclic aromatic hydrocarbons (PAHs) in low density polyethylene (LDPE). This specificity restricts their applicability to other combinations of chemicals and materials. To overcome this limitation, [11] Guo proposed a method for calculating K_{ma} based on the vapor pressure of chemicals, making it applicable across all materials and chemical classes. Nonetheless, the applicability of this approach to SVOCs is constrained due to its reliance on a limited dataset primarily focused on VOCs in construction materials. An additional limitation of previous studies is the inadequate consideration of temperature effects. While some research has accounted for temperature by adjusting predictors, others have presented separate correlation coefficients for specific temperatures. However, temperature correction is not always feasible due to the lack of available enthalpy data for phase transitions in many substances [15][16]. Additionally, commonly used physicochemical parameters, for example vapor pressure and K_{oa}, are typically reported only at 25°C. Although several studies have established a correlation between K_{ma} and temperature, these relationships have largely been validated through experimental methods.[17][18]

Current correlation methods for estimating K_{ma} fall short when it comes to addressing chemicals present in solid materials across varying ambient temperatures. In exposure science, there is an increasing focus on developing low-cost, high-throughput approaches to evaluate chemical exposure in consumer products, accounting for a diverse range of chemical-material combinations. This highlights the need for high-throughput estimations of K_{ma} across diverse material-chemical pairings. To meet the demand for more efficient exposure assessments, this study aims to develop a more reliable correlation method for estimating K_{ma} for various organic chemicals across different solid materials. Specifically, our goals are:

1. Perform a comprehensive literature review to gather experimental K_{ma} data covering a broad range of substances and materials.
2. Multivariate linear regression approaches to examine the association between K_{ma} and different predictor variables, such as temperature, material type, and physicochemical properties.
3. Conduct both internal and external validation to assess the accuracy and the predictive accuracy of the established correlations.

This IQSPR represents an advanced approach for estimating The K_{ma} values of organic compounds, offering broader applicability to various solid materials and chemicals while effectively incorporating temperature variations. Our research team has also developed a similar QSPR for estimating the internal diffusion rate within solid materials. Together, these QSPRs enable high-throughput evaluations of Chemical release and human exposure related to chemicals in solid materials. These advancements are valuable for various science-policy purposes, such as chemical alternatives assessment (CAA), life cycle assessment (LCA), and risk assessment, and, by providing accurate estimates of critical partitioning and diffusion parameters across diverse material-chemical combinations.

II. MATERIALS AND METHODS

Dataset

Data collection

Data on Measured material-air partition coefficients were sourced from 43 peer-reviewed scientific references. The partition coefficients were standardized to dimensionless values. For coefficients initially reported in mL/g or m^3/g, conversions were performed by multiplying the values by the mass density of the solid material. In cases where the coefficients were represented in meters, they were divided by the thickness of the material to derive dimensionless values. The original K_{ma} dataset contained 1008 entries, encompassing 75 distinct solid materials and 179 unique compounds.

Data curation

EPI Suite was used to obtain the molecular weight, water solubility, vapor pressure, and $\log K_{ow}$ at 25°C for each of the 179 distinct substances in the original K_{ma} dataset. In instances where experimental data were accessible, they were utilized for these physicochemical parameters; otherwise, values estimated by software were employed. Enthalpy of vaporization (ΔHv, J/mol) for each chemical is calculated using estimated values obtained from (www.chemspider.com).

Only a small subset of the 179 compounds in the dataset has experimental data for the partition coefficient between octanol and air ($\log K_{oa}$) at 25°C. To maintain consistency, we utilized the $\log K_{oa}$ values, computed by EPISuite 19 for all 179 compounds. The dimensionless logarithm of the air-water partition coefficient, $\log K_{aw}$, is estimated by utilizing the HenryWin and K_{ow}Win models, while $\log K_{oa}$ is derived in EPISuite by subtracting $\log K_{aw}$ from $\log K_{ow}$. As shown in Section S6, experimental $\log K_{oa}$ values were also gathered, and their effects on the QSPR were evaluated.

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 1, pp 15-25, January-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.401003
ISSN (Online): 2583-7257

The 75 original materials for K_{ma} were classified into 22 grouped material types, as described in Section S1, based on the material nomenclature and similarities in the regression coefficients. This classification was introduced to reduce the risk of overfitting in the QSPR model, ensuring that each consolidated material type included at least five data points and three distinct chemicals. Data points containing components that did not fit into the aforementioned categories were not included in further analysis.

991 data points with 179 distinct compounds across 22 aggregated material classes make up the final K_{ma} dataset. The K_{ma} was measured at temperatures between 15°C and 100°C. The supporting information contains the finished dataset.

Methods of modeling

Model of MLR (multiple linear regression)

MLR analysis was conducted to assess and Measure the impact of different parameters on the diffusion coefficient, with detailed information provided in our earlier publication on the quantitative structure-activity relationship (QSAR) for the diffusion coefficient. In summary, the MLR model is represented as follows:

$$\log_{10} K_{ma} = \alpha + \beta_1 . X_1 + \cdots + \beta_n . X_n + b_1 . M_1 + \cdots + b_n . M_n \qquad (1)$$

In this context, α represents the intercept, while log10Kma refers to the log of the dimensionless K_{ma}. The variables X_1 through X_n is independent variables associated with environmental or chemical properties, and β1 through βn are the regression coefficients associated with these variables. The dummy variables for the packing materials are labeled M_1 through Mm, with each dummy variable representing a specific type of material. For instance, M_1 equals (1) for material type 1 and (2) for material types 2 through m. Each dummy variable is assigned a value of 1 for the material type it signifies and 0 for all other types. The regression coefficients associated with these dummy variables, M_1 to Mm, are represented by b1 to bm. The total number of m corresponds to the number of material types analyzed, minus one, as the reference material type, PUether, which has the most extensive K_{ma} data, is used and does not require a dummy variable in the multiple linear regressions (MLR) analysis. The regression coefficients were determined using the least squares (LS) method, and all regression analyses were conducted with IBM SPSS Statistics version 23 (IBM Corporation, Armonk, NY).

Prior research has employed either the $\log K_{oa}$ [5][6][12][13] or the chemical vapor pressure [4][9][11] as predictors of the K_{ma} in a particular material. Holmgren et al.[14] also employed Abraham solvation parameters as predictors; however, because these parameters are not easily accessible, they are not taken into account here. According to preliminary regressions (Section S2), $\log K_{oa}$ predicts K_{ma} more accurately than vapor pressure. Therefore, the predictor variable representing chemical characteristics in Equation (1) was the chemical $\log K_{oa}$ at 25°C.

Consequently, the MLR model looks like this:

$$\log_{10} K_{ma} = \alpha + \beta_{logK_{oa}} . \log_{10} K_{oa} + \beta_T . T_{term} + b_1 . M_1 + \cdots + b_{21} . M_{21} \qquad (2)$$

Dependence of temperature

Van 't Hoff equation in thermodynamics are used to describe the relationship between the equilibrium constant and temperature, K_{eq}, varies with temperature:

$$\ln \frac{K_2}{K_1} = \frac{\Delta H_{phasechange}}{R} \left(\frac{1}{T_2} - \frac{1}{T_1} \right) \qquad (3a)$$

In this context, T_1 and T_2 represent the absolute temperatures measured in Kelvin (K), where R represents the ideal gas constant with a value of 8.314 J/(K·mol), and ΔH_{phase} change signifies the enthalpy associated with the phase change expressed in joules per mole (J/mol). The equilibrium constants at temperatures T1 and T2 are represented as K1 and K2, respectively.

The log10Koa of the chemical at 25°C (298.15 K) serves as a predictor variable in the multiple linear regression (MLR) model (Equation 2). Since K_{ma} is defined as an equilibrium constant, we hypothesize that the temperature dependence of K_{ma} Also follows the van't Hoff equation.

$$T_{term} = \log_{10} \frac{K_{ma.2}}{K_{ma.1}} = \frac{\Delta H_{ma}}{2.303 . R} \left(\frac{1}{T_2} - \frac{1}{298.15} \right) \qquad (3b)$$

The value of 2.303 serves as a conversion factor that relates log10K to lnK, while ΔH_{ma} represents the enthalpy associated with the partitioning process between the material and air, expressed in joules per mole (J/mol).

For different combinations of chemical materials, the enthalpy ΔH_{ma} is expected to exhibit variability. Utilizing the K_{ma} data collected between 15 and 95°C, Kamprad and Goss calculated the ΔHma values for 54 unique compounds in PU-ether. Consequently, a linear relationship was established to estimate ΔH_{ma} based on the chemical properties.

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 1, pp 15-25, January-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.401003
ISSN (Online): 2583-7257

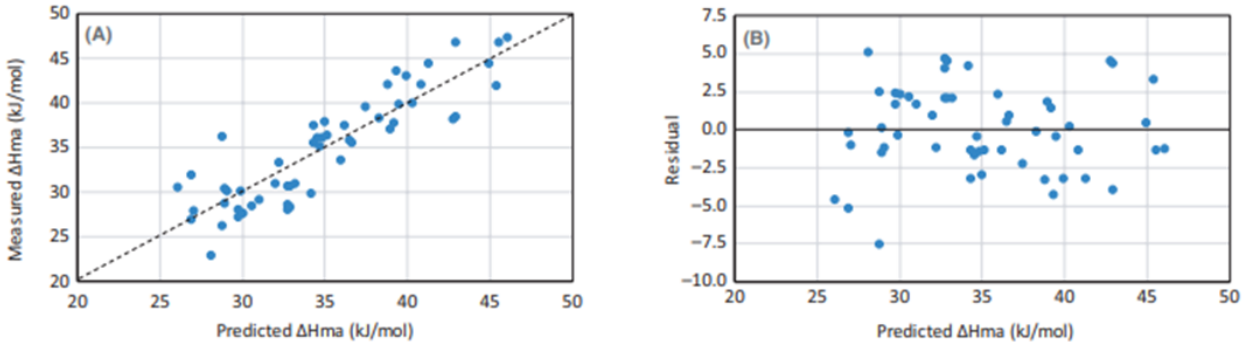

Figure 1A: Illustrates the recorded enthalpy of material-air partitioning (ΔH_3), while Figure 1B presents the residuals plotted against the predicted enthalpy of vaporization (ΔHv) derived from Equation 5. The dashed line in panel (A) indicates the 1:1 line

The ΔH_{ma} correlation previously established is utilized for all materials, when experimental ΔH_{ma} values for substances, other than PU-ether, are unavailable. Consequently, the ΔH_{ma} in our K_{ma}' regression model is chemically specific, yet not specific to any particular material. Therefore, the final multiple linear regression (MLR) model is presented as follows.

$$\log_{10} K_{ma} = \alpha + \beta_{\log_{K_{oa}}} . \log_{10} K_{oa} + \beta_T \frac{\Delta H_{ma}}{2.303.R}\left(\frac{1}{T_2} - \frac{1}{298.15}\right) + b_1 . M_1 + \cdots + b_{21} . M_{21} \quad (4)$$

Validation of the model

The QSARINS program created by Gramatica and colleagues, [20][21] was employed to validate the final multiple linear regression model (Equation 4).

Internal validation

The internal validation process assessed The IMLR model's capability to predict sections of the training dataset using two methods in QSARINS: the previously mentioned Y-scrambling and the leave more out (LMO) cross-validation,[1][21] 1000 iterations were employed, with a 20% omitted element percentage. Additionally, 1000 iterations were utilized for Y-scrambling.

External validation

The splitting technique was applied to separate the existing dataset of 991 data points into a training set and a prediction set. This method enabled us to evaluate the model's predictive accuracy Ion new datasets through external validation. The regression coefficients for the IMLR model were determined from the training dataset, and the model's predictive performance was subsequently assessed by applying it to the prediction set. Utilizing the built-in functionalities of the QSARINS software, three distinct types of splitting were executed: random proportion, ordered answer, and structural. Additionally, a fourth method, which involved splitting by studies, was implemented; in this case, the training set included all data points from specific studies, The prediction set included data points from external studies. To ensure representation of all consolidated material types in the IMLR model derived from the training set, any data points from a consolidated material type that came from a single study were assigned to the training set. The four splitting methods produced comparable sample sizes, with roughly 800 data points in the training set and 200 in the prediction set, as shown in Table S3.

III. OUTCOMES AND CONVERSATIONS

Dependence on temperature

The enthalpy associated with the material-air partitioning, denoted as ΔH_{ma} (J/mol), influences the temperature sensitivity of K_{ma}. A correlation has been established to estimate ΔH_{ma} based on the measured K_{ma} values for 54 compounds in PU-ether across a temperature range of 15 to 95°C (refer to Section S3 for the data):

$$\Delta H_{ma} = 1.37.\Delta H_V - 14.0 \quad (5)$$

N = 54, R^2 = 0.786, R^2_{adj} = 0.782, SE = 2.85, RMSE = 2.80

ANOVA: F = 191, df = 1, P < 0.0001

Enthalpy of vaporization (J/mol) of the chemical, sourced from Chem Spider, is denoted as ΔHv.

An adjusted R-squared value of 0.782 suggests that this simple linear model effectively represents the experimental ΔH_{ma} data, with a highly significant model fit indicated by an ANOVA P-value of less than 0.0001. Figure 1 displays the residual plot and the scatter plot comparing predicted versus measured ΔH_{ma} values. Both plots closely align with the 1:1 line and exhibit a random distribution of residuals throughout

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 1, pp 15-25, January-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.401003
ISSN (Online): 2583-7257

the dataset. These results indicate a linear correlation between ΔH_{ma} and ΔH_v in PU-ether, with ΔH_{ma} calculated for all other materials in the same way, using Equation (5) as the standard reference.

Final model fitting and QSPR

The final multiple linear regression model for estimating the partition coefficient between solid material and air is shown below, based on Equation (4) and the complete dataset of 991 data points:

Table 1: Regression coefficients for Equation (6)

Variable	Coefficient	SE[a]	P-value
$\log_{10}K_{oa}$	0.63	0.01	<0.001
Intercept	−0.38	0.05	<0.001
ΔH_{ma} (1/T-1/298.15)/2.303R	0.96	0.04	<0.001
Consolidated material types (coefficient b)			
Cellulose fabric (cotton, linen)	0.72	0.13	<0.001
Carpet	1.97	0.14	<0.001
Concrete	2.20	0.28	<0.001
Cement, Calcium silicate	1.11	0.10	<0.001
Ethylene Vinyl Acetate (EVA)	3.50	0.31	<0.001
Gypsum board	1.28	0.18	<0.001
Glass	1.11	0.29	<0.001
Latex and solvent-based paint	2.92	0.19	<0.001
Paper	0.14	0.10	0.16
Plywood	1.36	0.18	<0.001
Polyester fabric	0.60	0.14	<0.001
Polytetrafluoroethylene (PTFE)	2.08	0.28	<0.001
Polyether ether ketone (PEEK)	2.73	0.29	<0.001
Polyethylene (PE)	2.45	0.17	<0.001
Polypropylene (PP)	2.06	0.29	<0.001
PU-ester	−0.72	0.07	<0.001
PU-ether[b]	0.00	0.19	n/a
Stainless steel	2.07	0.29	<0.001
Rayon fabric	0.97	0.18	<0.001
PUF-undefined	1.06	0.15	<0.001
Vinyl flooring	2.26	0.10	<0.001
Woodenboards[c]	2.02	0.09	<0.001

[a]Standard error.

[b]Reference material.

[c]Includes oriented strandboard (OSB), particleboard, medium-density. board and high- density board.

$$\log_{10}K_{ma} = -0.38 + 0.63.\log_{10}K_{oa} + 0.96.\frac{\Delta H_{ma}}{2.303.R}\left(\frac{1}{T_2} - \frac{1}{298.15}\right) + b \qquad (6)$$

N = 991, R^2 = 0.934, R^2_{adj} = 0.933, SE = 0.62, RMSE = 0.62

ANOVA: F = 597, df = 23, P < 0.0001

For this case, T denotes the absolute temperature measured in Kelvin (K), while b refers to the coefficients specific to the material, as outlined in Table 1. The enthalpy associated with partitioning between the material and air, denoted as ΔH_{ma}, is measured in joules per mole (J/mol) and is calculated using Equation (5). The dimensionless solid material-air partition coefficient is denoted as K_{ma}. Additionally, Koa represents the octanol-air partition coefficient expressed in a dimensionless form for the chemical at 25°C. To facilitate practical use, Table 1 also presents the standard errors associated with the coefficient. With a standard error (SE) of 0.63 from the final model, the 95% confidence interval (CI) for the predicted log K_{ma} is given as the predicted value ± 1.22, suggesting that most predicted K_{ma} values lie within a factor of I16 of the observed K_{ma} values.

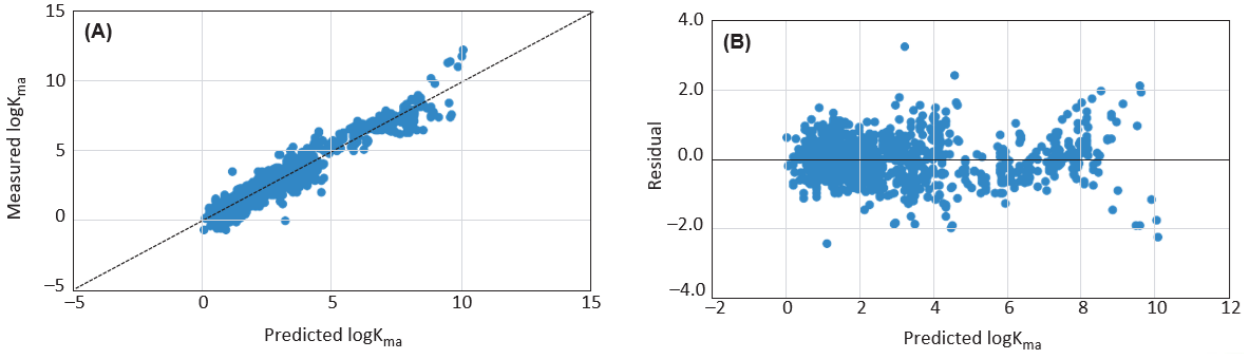

Figure 2: Shows the residuals plotted against the $logK_{ma}$ the values predicted by the final QSPRI (Equation I6) and the observed $logK_{ma}$ values are shown in panel A. The dotted line in panel A represents the 1:1 line

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 1, pp 15-25, January-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.401003
ISSN (Online): 2583-7257

The data in Table 1 indicate that the 95% confidence interval (CI) for the estimated $\log K_{ma}$ is computed as the predicted value ± 1.22, based on a standard error (SE) of I0.63 from the final model (refer to Equation 6). This suggests that most predicted K_{ma} values fall within a factor of 16 of the observed K_{ma} values.

The multiple linear regression (MLR) model exhibits a strong fit to the experimental data, demonstrated by a root mean square error (RMSE) of 0.62 and an adjusted LR-squared value of 0.93. The ANOVA P-value, which is less than 0.0001, indicates a highly significant model fit. The scatter plot comparing predicted and observed $\log K_{ma}$ values, shown in Figure 2A, closely follows the 1:1 line. Furthermore, the residual plot in Figure 1B reveals a consistent distribution of residuals across the dataset, with most residuals having absolute values below 2, further supporting the adequacy of the linear model in capturing the data.

Given the satisfactory fit of the model, it appears that the multiple linear regression (MLR) The model assumes a consistent relationship between $\log K_{ma}$ and the chemical $\log K_{oa}$ across all material types. However, using a single coefficient for $\log K_{oa}$, as applied in the current quantitative structure-activity relationship (QSAR) model, may not be ideal. This is reflected in the relationship between $\log K_{ma}$ and the chemical's $\log K_{oa}$, where the slopes of the fitted lines in Figure 3 are comparable, yet exhibit slight variations across different material types when plotted against $\log K_{ma}$ for selected materials. This might have been addressed by adding interaction terms between material types and $\log K_{oa}$, but doing so would have added 21 new terms to the model without significantly enhancing model fit (Section S5). The final QSPR model, therefore, excludes the interaction variables.

Due to the unavailability of experimental $\log K_{oa}$ values for every chemical in the dataset, the final multiple linear regression (MLR) model incorporates $\log K_{oa}$ values estimated by EPISuite as predictors, as detailed in the methods section. The results of MLR models developed using a combination of $\log K_{oa}$ values where experimental $\log K_{oa}$ values are applied when accessible, and EPISuite-estimated logKoa values are used otherwise were found to be comparable to those of the final MLR model, with an adjusted R^2 ranging is 0.930 - 0.931. This suggests that the availability of experimental $\log K_{oa}$ values had a negligible effect on the model's performance.

3.3 Each predictor's impact

The $\log K_{oa}$ of the chemical, along with the enthalpy of vaporization (ΔHv), the type of solid material, and temperature, are the key factors affecting the solid material-air partition coefficient, (Equation 6). The analysis reveals that the partition coefficient increases with an increase in $\log K_{oa}$, supported by a regression coefficient of 0.63 for $\log K_{ow}$, which is statistically significant ($P < 0.0001$).

The regression coefficient for the temperature variable is 0.96, suggesting a highly significant relationship ($P < 0.0001$), which indicates that K_{ma} decreases as temperature rises. The experimental results provided by Kamprad. Corroborate this observation, showing a reduction in K_{ma} with rising temperatures. It is also reasonable to conclude that a lower K_{ma} at elevated temperatures facilitates a more rapid transfer of chemicals from solid substances to the atmosphere. Furthermore, the impact of temperature on K_{ma} is influenced by ΔH_{ma}, which shows a linear increase in relation to the chemical's enthalpy of vaporization, ΔH_v.

The material dependency of the K_{ma} is illustrated through the inclusion of 21 dummy variables representing various material types. In this analysis, "PU-ether" (polyurethane-ether) serves as the reference material in the regression analysis, resulting in a coefficient b of 0 (as shown in Table 1). The coefficient b reflects the variation in $\log K_{ma}$ between PU-ether and each of the other material types. Chemicals contained within solid materials characterized by higher b values encounter greater difficulty in transitioning to air compared to those associated with lower b values. The three densest materials exhibiting the highest b coefficients are latex, solvent-based paint, polyether ether ketone (PEEK), and ethylene vinyl acetate (EVA).

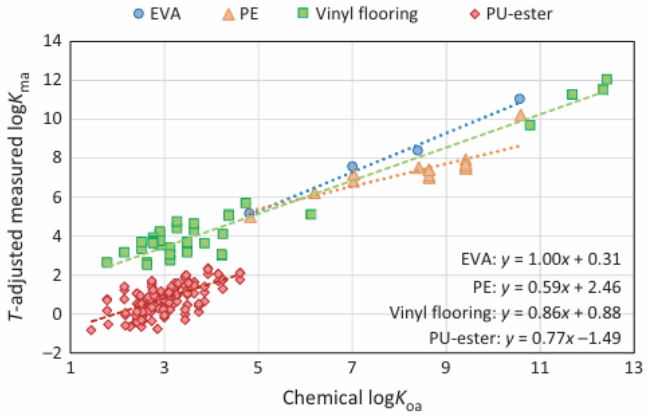

Figure 3: Displays the temperature-adjusted measured $\log K_{ma}$ as a function of 8$\log K_{oa}$ for several different material types, such as vinyl flooring, PE, EVA, and PU-ester

Polyurethane materials, specifically paper, PU-ether, and PU-ester, are generally characterized by their porous nature and exhibit the lowest b coefficients. It is important to note that the coefficients for different material types presented in Table 1 shows only the average composition and partition

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 1, pp 15-25, January-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.401003
ISSN (Online): 2583-7257

behavior for these materials. This is because the data for each combined material type were sourced from multiple studies, leading to potential variations in composition and properties across different research findings.

We systematically changed each predictor from its minimum to maximum value across the entire dataset, which includes 991 data points, while keeping the other predictors fixed. This approach was employed to assess the influence of each predictor on the material-air partition coefficient. Subsequently, we applied the regression coefficients from the final Quantitative Structure-Activity Relationship (QSPR) model (Equation 6) to assess the changes in $\log K_{ma}$. By using the lowest and highest values of ΔH_v in the dataset, we evaluated the effect of temperature at two extremes, as the chemical's ΔH_v determines ΔH_{ma}, thus altering the relationship between $\log K_{ma}$ and temperature. Among all predictors, the chemical's $\log K_{oa}$ exhibited the most significant Iimpact on $\log K_{ma}$, as illustrated in Figure 4. The effect of temperature on $\log K_{ma}$ is minimal, particularly when ΔH_v is at its lowest value of 22.3 kJ/mol. Conversely, with the highest ΔHv value of 75.6 kJ/mol, the influence becomes moderate. Consequently, For a chemical with a low enthalpy of vaporization, the variation in $\log K_{ma}$ with temperature is minimal, and Ithe opposite holds true for chemicals with high enthalpy of vaporization. Similar to the temperature effect at maximum ΔH_v, the material type also moderately affects $\log K_{ma}$. However, the overall effect of the influence of material type is relatively minor compared Ito the effect of the chemical's $\log K_{oa}$, suggesting that the type of solid material plays a limited role in the fluctuations of $\log K_{ma}$. This finding indicates the possibility of creating a generic QSPR model to predict $\log K_{ma}$ without relying on material-specific data.

Results of model validation

In-house verification

The average root mean square error for cross-validation (RMSEcv) is 0.63, while the average correlation coefficient for the LMO cross-validation is..., Q^2_{LMO}, is 0.93, with a range of 0.90 to 0.95 across 1000 iterations. The model shows internal stability, as the Q2 LMO and RMSEcv values align closely with the R2 and RMSE derived from the complete dataset, which are 0.93 and 0.62, respectively.

The analysis shows no relationship between the randomized responses and the predictors, as evidenced by the R^2_{Yscr}, Q^2_{Yscr}, and $RMSE_{Yscr}$ values obtained from 1000 iterations, with averages of 0.023, -0.028, and 2.37, respectively. These values are significantly different from the R^2, Q^2LMO, and RMSE of the original model. Therefore, the thorough internal validation confirms that the final QSPR model (Equation 6) is both stable and robust, ensuring that it is not merely the result of random correlation.

External validation

For external validation, four data-splitting methods were utilized, including splitting by ordered response, by structure, by studies, and by a random selection of 20%, as described in Section 2.3.2. The results of the six external validation criteria, previously outlined in references 1, 22, and 23, are presented in Table 2. For the first three splitting methods, the R^2_{ext} values exceed 0.9, and all five additional criteria meet or surpass the threshold of 0.9, this demonstrates the strong predictive ability of models developed using the training set data. Since the data points were chosen either randomly or alternately, the prediction sets generated by these three splitting methods largely fall within the application domain of the training sets.

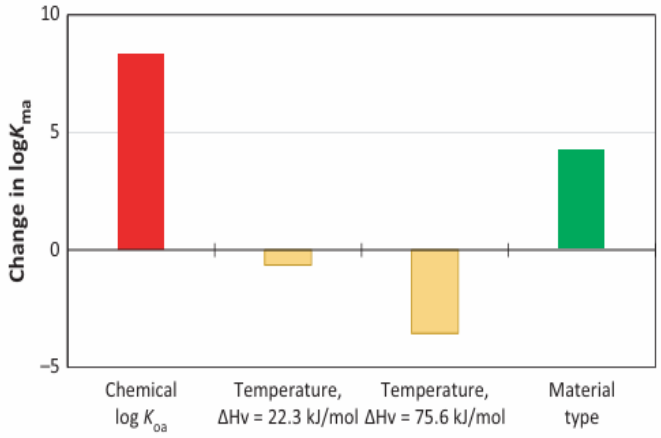

Figure 4: $\log K_{ma}$ change from minimum to maximum values across the entire dataset in relation to each predictor's change

The training set consisted of data from twenty studies, while the prediction set was formed from data of twenty-two studies in the study-by-study division. Since each data point from an individual study was assigned exclusively to either the training or the prediction set, this approach offers a more accurate representation of a truly "external" validation. The R^2 ext for this division decreased to 0.79, and the values for the other five criteria were lower compared to those from the three previously mentioned splitting methods, this suggests that the model derived from the training data exhibits reduced predictive accuracy. It is reasonable to infer that the prediction set may not align well within the acceptable domain defined by the training set, as variability across studies is greater than within individual studies. Nonetheless, all validation criteria for this separation meet the required thresholds, demonstrating a reasonable level of predictive performance (Table 2).

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 1, pp 15-25, January-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.401003
ISSN (Online): 2583-7257

Domain of applicability

Establishing the Application Domain (AD) of our Quantitative Structure-Activity Relationship (QSPR) model is essential, as it provides prospective users intending to utilize the model for new compounds with insights into the accuracy and dependability of the model's predictions. When new chemicals are included within the application domain (AD), the model's predictions are interpolated, leading to greater accuracy. In contrast, if the compounds fall outside the AD, the predictions are extrapolated, leading to reduced reliability [24].

The concluding QSPR model presented in Equation (6) is the model under evaluation for the assessment of the AD. Consequently, the training dataset encompasses the complete dataset, consisting of 991 data points. The scope of the model's predictors, the leverage strategy, and The principal component analysis (PCA) applied to the model's predictors - each of which has been thoroughly explained earlier—were used together to define the application domain (AD) of the K_{ma} QSPR.[25]

The model's range is defined by four predictors: $LogK_{oa}$, ΔHv, material type, and temperature. The application domain (AD) of the model is characterized by the values of $logK_{oa}$, ΔHv, and temperature from the training dataset, which range from 1.4 to 14.6, 22.3 to 75.6 kJ/mol, and 15 to 100°C, respectively. It is worth noting that the training set includes 22 distinct material types, with material type being classified as a categorical variable.

Table 2: External validation results

External validation criteria	R^2_{ext}	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	r^2_m	CCC
Threshold		>0.70	>0.70	>0.70	>0.65	>0.85
Splitting by random percentage	0.93	0.93	0.93	0.92	0.90	0.96
Splitting by ordered response	0.93	0.93	0.93	0.93	0.90	0.96
Splitting by ordered structure	0.94	0.94	0.94	0.94	0.91	0.97
Splitting by studies	0.79	0.86	0.78	0.86	0.71	0.89

R^2_{ext}: determination coefficient of the prediction set external data.

Q^2_{FI}: correlation coefficient proposed by Shi et al.

Q^2_{F2}: correlation coefficient proposed by Schuurmann et al.

Q^2_{F3}: correlation coefficient proposed by Consonni et al.

r^2m: determination coefficient proposed by Ojha et al.

CCC: concordance correlation coefficient proposed by Chirico and Gramatica.

Consequently, the model's acceptable domain (AD) is limited to these 22 material types. The AD is defined by h values that are smaller than h*, with the leverage technique identifying the critical value h* for the diagonal elements of the model's hat (h) matrix as 0.0727. Using Principal Component Analysis (PCA), the AD is characterized by the range of minimum and maximum scores for PC1 and PC2 from the training dataset, which are −4.39 to 2.04 and −4.52 to 2.22, respectively. For future users of the model, a new chemical should be classified as "within AD" if it is deemed within AD by all three methods, and as "outside AD" if it is considered outside AD by all three approaches. If the chemical does not fit into either category, it should be classified as "borderline.[25]

Generic QSPR

Using the same dataset, we developed a generic QSPR model that excludes material-specific factors, allowing for the prediction of K_{ma} without the need to assign material attributes. This model is as follows and solely uses temperature and chemical characteristics as predictors:

$$\log_{10}K_{ma} = -0.37 + 0.75.\log_{10}K_{oa} + 1.29\,.\,\frac{\Delta H_{ma}}{2.303.R}\left(\frac{1}{T}-\frac{1}{298.15}\right) \quad (7)$$

N = 991, R^2 = 0.80, R^2_{adj} = 0.80, SE = 1.08, RMSE = 1.08

ANOVA: F = 1943, df = 2 , P < 0.0001

This model demonstrates a robust alignment with the experimental data, as evidenced by its relatively high adjusted R-squared value of 0.80, particularly when juxtaposed with the regression model featuring a material coefficient of 0.93 (Equation 6) (Figure 5)., the impact of chemical properties on $logK_{ma}$ surpasses that of the solid material type, allowing for a reasonably accurate prediction of $logK_{ma}$ even without specifying the material type. Consequently, this generic Quantitative Structure-Activity Relationship (QSPR) provides a fairly reliable method for estimating K_{ma} across a diverse array of solid materials, particularly when classifying materials as outlined in Table 1 proves challenging. It

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 1, pp 15-25, January-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.401003
ISSN (Online): 2583-7257

significantly enhances high-throughput assessments of various chemical-material combinations by offering broader and more flexible coverage than the material-specific QSPR, albeit with slightly reduced accuracy. It is important to note that although this general model was constructed using experimental data from our dataset of 22 material types, material type was excluded as a predictor. As a result, while this general model is most effective for materials in Table 1 and other similar substances, it may produce significant inaccuracies for materials with unique characteristics. This includes materials exhibiting strong ionic interactions or pronounced pseudo-solvation effects, which can alter the structure of target adsorbate molecules due to ionization or tautomerization.

Limitations and Future work

Although the current model offers significant advantages, including the incorporation of temperature effects and Incorporating 22 consolidated materials (and potentially any solid substance), it does have several limitations. Firstly, the model does not account for ionization and interactions among chemicals within a solid material, potentially affecting their partitioning behavior between the material and air. Additionally, the model assumes that the correlation between ΔH_{ma} and the chemical's ΔHv is consistent across all material types, based on experimental ΔH_{ma} data from the material known as "PU-ether." To test this assumption or create separate ΔH_{ma}-ΔHv correlations for different material types, further experimental ΔH_{ma} data would be beneficial.

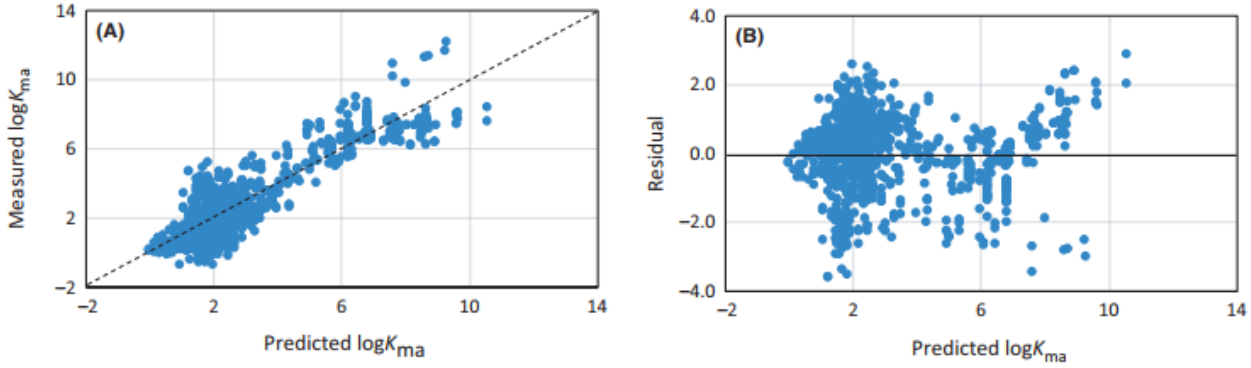

Figure 5: Displays the residuals as a function of the predicted $logK_{ma}$ values from the generic QSPR model (Equation 7) and the observed $logK_{ma}$ values (A). The 1:1 line is shown by the dotted line in (A)

The categorization of consolidated material types is mainly qualitative and depends solely on the names of the materials. This method can lead to significant variability in material properties within each category, as the original publications often fail to provide comprehensive descriptions of the properties for most K_{ma} datasets. Moreover, alterations in material structure can influence the partitioning behavior between the material and air, even when the chemical composition remains unchanged. To improve the model's accuracy for specific materials and expand its applicability to other material types not covered in the training dataset, it would be beneficial to include quantitative, the continuous properties of solid materials, including detailed descriptors of their composition and molecular structure, should be quantified and integrated into the model as numerical predictors. Additionally, utilizing quantitative variables for material types allows for the introduction of interaction terms between the $logK_{oa}$ of the chemical and the material type variables, thereby improving the model's fit without significantly increasing the number of parameters.

Many materials commonly encountered in indoor environments exhibit a lack of uniformity, such as plywood, carpet, gypsum board, paper, and concrete, which may possess varying qualities across different layers or sections. Therefore, it is likely that the K_{ma} values obtained from trials and the QSPR derived from them merely reflect the material properties throughout the experiments. Therefore, it is important to exercise caution when predicting K_{ma} with the current QSPR, particularly for extremely inhomogeneous materials. The difference between surface and bulk partitioning is another significant factor pertaining to heterogenicity. The surface properties can significantly influence the observed partitioning behavior, as the interaction between solid materials and air predominantly takes place at the material's surface. Consequently, for substances having a surface layer with unique characteristics, with materials that differ in surface/bulk structures but have the same composition, the current QSPR might not provide an accurate K_{ma} calculation. Either intrinsic properties that are time invariant or oxidative aging and soiling, which can alter over time, could be the cause of the different surface layer. These

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 1, pp 15-25, January-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.401003
ISSN (Online): 2583-7257

issues emphasize once more how crucial it is to use quantitative descriptions of Using the compositions and structures of materials as predictors in the IQSPR model

Lastly, the QSPR does not incorporate additional effect elements like relative humidity because it is uncertain how they work. More thorough investigation is needed to determine the possible chemical and material-dependent effects of relative humidity on K_{ma}.

IV. CONCLUSIONS

A multiple linear regression (MLR) model was developed to predict the solid material-air partition coefficients (K_{ma}) of organic compounds across various solid materials. This model was constructed using experimental K_{ma} data from 43 studies, incorporating one categorical variable, material type, along with three continuous predictors: chemical $logK_{oa}$, ΔHv, and absolute temperature. The model has demonstrated stability through both internal and external validation, robust, and accurate in its predictions. It is applicable to chemicals with $logK_{oa}$ values ranging from 1.4 to 14.6, ΔHv values from 22.3 to 75.6 kJ/mol, temperatures between 15°C and 100°C, and material types grouped into 22 distinct categories.

The main advantage of the existing model lies in its extensive applicability across diverse combinations of chemicals, materials, and temperature ranges. This characteristic makes it more comprehensive compared to the correlation methods developed in previous studies, which typically focused on room temperature and were tailored to individual solid substances. Furthermore, a generalized model has been established that can yield reasonably accurate estimations of K_{ma} without the necessity of specifying a particular material type. This feature enhances its suitability for high-throughput assessments of chemical emissions from solid materials and the subsequent consumer exposures.

REFERENCES

[1] Huang L, Fantke P, Ernstoff A, et al. A quantitative property relationship for the internal diffusion coefficients of organic compounds in solid materials. Indoor Air. 2017;27(6): 1128-1140.
[2] Huang L, Jolliet O. A parsimonious model for the release of chemicals encapsulated in products. Atmos Environ. 2016;127: 223-235.
[3] Liu Z, Ye W, Little JC. Predicting emissions of volatile and semi volatile organic compounds from building materials: a review. Build Environ. 2013;64:7-25.
[4] Kamprad I, Goss K-U. Systematic investigation of the sorption properties of polyurethane foams for organic vapors. Anal Chem.2007;79(11):4222-4227.
[5] Bartkow ME, Hawker DW, Kennedy KE, et al. Characterizing uptake kinetics of PAHs from the air using polyethylene-based passive air samplers of multiple surface area-to-volume ratios. Environ Sci Technol. 2004;38(9):2701-2706.
[6] Kennedy KE, Hawker DW, Müller JF, et al. A field comparison of ethylene vinyl acetate and low-density polyethylene thin films for equilibrium phase passive air sampling of polycyclic aromatic hydrocarbons. Atmos Environ. 2007;41(27):5778-5787.
[7] Morrison G, Li H, Mishra S, et al. Airborne phthalate partitioning to cotton clothing. Atmos Environ. 2015;115:149-152.
[8] Morrison G, Shakila N, Parker K. Accumulation of gas-phase methamphetamine on clothing, toy fabrics, and skin oil. Indoor Air. 2015;25(4):405-414.
[9] Zhao D, Little JC, Cox SS. Characterizing polyurethane foam as a sink for or source of volatile organic compounds in indoor air. J Environ Eng. 2004;130(9):983-989.
[10] Bodalal A, Zhang J, Plett E, et al. Correlations between the internal diffusion and equilibrium partition coefficients of volatile organic compounds (VOCs) in building materials and the VOC properties. ASHRAE Trans. 2001;107:789.
[11] Guo Z. Review of indoor emission source models. Part 2. Parameter estimation. Environ Pollut 2002;120(3):551-564.
[12] Chaemfa C, Barber JL, Gocht T, et al. Field calibration of polyurethane foam (PUF) disk passive air samplers for PCBs and OC pesticides. Environ Pollut. 2008;156(3):1290-1297.
[13] Shoeib M, Harner T. Characterization and comparison of three passive air samplers for persistent organic pollutants. Environ Sci Technol. 2002;36(19):4142-4151.
[14] Holmgren T, Persson L, Andersson PL, et al. A generic emission model to predict release of organic substances from materials in consumer goods. Sci Total Environ. 2012;437:306-314.
[15] Booij K, Hofmans HE, Fischer CV, et al. Temperature-dependent uptake rates of nonpolar organic compounds by semipermeable membrane devices and low-density polyethylene membranes. Environ Sci Technol. 2003;37(2):361-366.
[16] Adams RG, Lohmann R, Fernandez LA, et al. Polyethylene devices: Passive samplers for measuring dissolved hydrophobic organic compounds in aquatic environments. Environ Sci Technol. 2007;41(4):1317-1323.
[17] Liu Y, Zhou X, Wang D, et al. A prediction model of VOC partition coefficient in porous building materials

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 1, pp 15-25, January-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.401003
ISSN (Online): 2583-7257

based on adsorption potential theory. Build Environ. 2015;93:221-233.

[18] Zhang Y, Luo X, Wang X, et al. Influence of temperature on formaldehyde emission parameters of dry building materials. Atmos Environ. 2007;41(15):3203-3216.

[19] USEPA. Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.11. Washington, DC: United States Environmental Protection Agency; 2012.

[20] Gramatica P, Cassani S, Chirico N. QSARINS-chem: insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS. J Comp Chem. 2014;35(13):1036-1044.

[21] Gramatica P, Chirico N, Papa E, et al. QSARINS: a new software for the development, analysis, and validation of QSAR MLR models. J Comp Chem. 2013;34(24):2121-2132.

[22] Chirico N, Gramatica P. Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. J ChemInf Model. 2012;52(8):2044-2058.

[23] Chirico N, Gramatica P. Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. J ChemInf Model. 2011;51(9):2320-2335.

[24] Gramatica P. Principles of QSAR models validation: internal and external. QSAR Comb Sci. 2007;26(5):694-701.

[25] Cassani S, Gramatica P. Identification of potential PBT behavior of personal care products by structural approaches. Sustain Chem Pharm. 2015;1:19-27.

Citation of this Article:

Anwer Abdulwahhab Khalaf. (2025). An Approach to Determining the Partition Coefficients between Solid and Air Phases for Organic Compounds Using the QSPR. *International Current Journal of Engineering and Science - ICJES*, 4(1), 15-25. Article DOI: https://doi.org/10.47001/ICJES/2025.401003

****

