



Article

Prediction of CO₂ solubility in deep eutectic solvents using random forest model based on COSMO-RS-derived descriptors

Jingwen Wang^{a,b}, Zhen Song^b, Lifang Chen^b, Tao Xu^{a,*}, Liyuan Deng^c, Zhiwen Qi^{b,*}

^a Academy of Building Energy Efficiency, School of Civil Engineering, Guangzhou University, Guangzhou, 510006, China

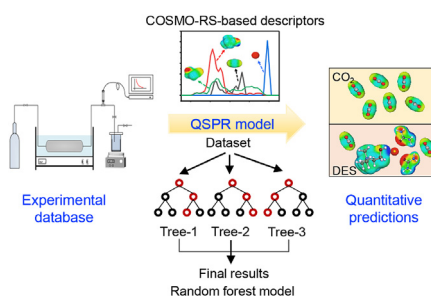
^b State Key Laboratory of Chemical Engineering, School of Chemical Engineering, East China University of Science and Technology, 130 Meilong Road, Shanghai, 200237, China

^c Department of Chemical Engineering, Norwegian University of Science and Technology, Sem Sælandsvei 4, 7491, Trondheim, Norway

HIGHLIGHTS

- A QSPR model is developed for predicting CO₂ solubility in DESs.
- Efficiency of the input variables is explored by a multiple linear regression model.
- Importance of the input variables in the QSPR model is ranked by random forest model.
- CO₂ solubility predictions by QSPR model are compared to the COSMO-RS model.
- QSPR model is suggested to be a reliable tool for selecting DESs for CO₂ capture.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:

CO₂ solubility prediction
 Deep eutectic solvents
 Quantitative structure-property relationship model
 COSMO-RS-derived descriptors
 Random forest

ABSTRACT

This work presents the development of molecular-based mathematical model for the prediction of CO₂ solubility in deep eutectic solvents (DESs). First, a comprehensive database containing 1011 CO₂ solubility data in various DESs at different temperatures and pressures is established, and the COSMO-RS-derived descriptors of involved hydrogen bond acceptors and hydrogen bond donors of DESs are calculated. Afterwards, the efficiency of the input variables, i.e., temperature, pressure, COSMO-RS-derived descriptors of HBA and HBD as well as their molar ratio, is explored by a qualitative analysis of CO₂ solubility in DESs using a simple multiple linear regression model. A machine learning method namely random forest is then employed to develop more accurate nonlinear quantitative structure-property relationship (QSPR) model. Combining the QSPR validation and comparisons with literature-reported models (i.e., COSMO-RS model, traditional thermodynamic models and equations of state methods), the developed QSPR model with COSMO-RS-derived parameters as molecular descriptors is suggested to be able to give reliable predictions of CO₂ solubility in DESs and could be used as a useful tool in selecting DESs for CO₂ capture processes.

1. Introduction

As a new generation of ionic liquids (ILs) analogues, deep eutectic

solvents (DESs) bear similar physicochemical properties as ILs, such as negligible vapor pressure, high chemical/thermal stability and easily tunable character [1]. Moreover, compared to ILs, DESs offer two

* Corresponding authors.

E-mail addresses: beexu@gzhu.edu.cn (T. Xu), zwqi@ecust.edu.cn (Z. Qi).

<https://doi.org/10.1016/j.gce.2021.08.002>

Received 24 June 2021; Received in revised form 20 July 2021; Accepted 8 August 2021

Available online 10 August 2021

2666-9528/© 2021 Institute of Process Engineering, Chinese Academy of Sciences. Publishing services by Elsevier B.V. on behalf of KeAi Communication Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

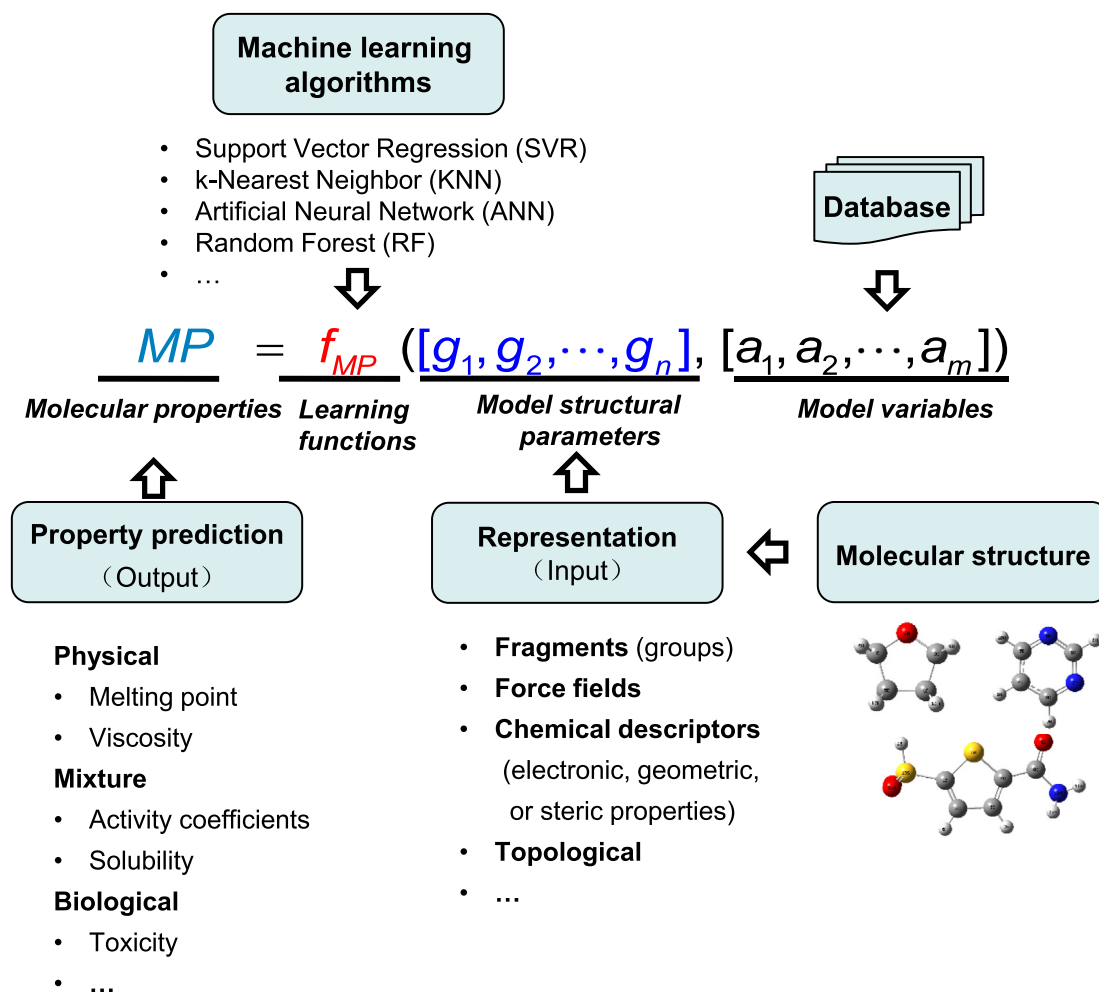


Fig. 1. Framework for establishing QSPR model based on machine learning method.

advantages: (1) DESs can be prepared easily by mixing a hydrogen bond acceptor (HBA) with a hydrogen bond donor (HBD), avoiding complex synthesis and purification steps for ILs; (2) A large number of cheap and renewable compounds can act as the HBA (e.g., ammonium and phosphonium salts) or HBD (e.g., organic alcohols and acids), making DESs more affordable and sustainable over ILs [2,3]. Due to these advantages, DESs have been trust into the limelight in various chemical processes [4, 5], among which CO₂ absorption is extensively concerned [6–10].

So far, most DES-based CO₂ absorption studies have mainly focused on the experimental measurement of CO₂ solubility in DESs, which have covered only a small proportion of DES candidates considering the large quantity of possible HBA-HBD combinations at different ratios [11–14]. To research more optimal DESs over a large space, experimental trial-and-error method is expensive, time-consuming and even unrealistic to explore all the possibilities. In this context, a reliable theoretical model for predicting CO₂ absorption capacity in DESs is highly desirable [15, 16].

In recent years, traditional thermodynamic models (e.g., NRTL and UNIQUAC) and equations of state methods (e.g., PC-SAFT, soft-SAFT and PR-EOS) have been successfully extended to DES-containing systems as reliable predictive tools for gas solubility [17–21]. Nevertheless, these methods require a number of experimentally fitted molecule-specific and mixing parameters, which limits the application space for novel systems. Additionally, COSMO-based thermodynamic models (i.e., COSMO-RS and COSMO-SAC), merely relying on quantum-chemically derived molecular descriptors, have been widely applied for predicting CO₂

solubility in IL solvents with acceptable accuracy [22–27]. However, these models are found to overpredict or underpredict the gas solubilities (CO₂, CO, CH₄, H₂ and N₂) severely in DESs [28]. Beyond the above-mentioned thermodynamic approaches, molecular simulation methods, such as molecular dynamics simulations and Monte Carlo simulations have been demonstrated to be reliable for predicting thermo-physical properties or phase equilibrium of DESs, including the gas solubility in solvents [29,30]. However, such methods require high computation-demanding and thus are not impractical for covering the huge diversity of gas-in-DES solubilities.

Alternatively, molecular descriptor-based quantitative structure-property relationship (QSPR) approach has attracted increasingly significant interests, which could for one thing act as an accurate and cost-effective tool for evaluating DES properties, and for another offer useful insights to understand the relationships between molecular-level interactions of DESs and their macroscopic properties [31–33]. As the prerequisite for QSPR models, COSMO-RS-based descriptors, such as charge distribution area (S_σ -profile), have been proven to be reliable molecule-specific input parameters for predicting IL and DES properties [34–38]. For instance, based on S_σ -profile, Lemaoui et al. [39] recently developed a QSPR model for the prediction of electrical conductivity of DESs, which can be used as a useful guideline in selecting DESs with the desired electrical conductivity for industrial applications. Similar studies were also observed by Benguerba and his co-workers [34] for the density and viscosity prediction of DESs, demonstrating the excellent performances of the molecular COSMO-RS-derived descriptors as QSPR input

parameters at predicting the properties of DESs. Therefore, COSMO-RS-based parameters will be explored for establishing the QSPR model for CO₂ solubility prediction in this study.

With respect to the other requisite for QSPR models, proper mathematical algorithms are also of great significance to correlate the objective function (i.e., CO₂ solubility in DESs of this work) with the input parameters [40–42]. Considering the fact that many properties cannot be directly described by linear models with high accuracy, the machine learning (ML) algorithms have been developed and are popular for building complex nonlinear QSPR model for estimating the phase behaviors of IL and DES-involved systems [43–47]. Among the diverse ML algorithms, random forest method has attracted considerable attention for data classification and regression due to its evident advantages of high prediction accuracy, no need to pre-select descriptors, and suppressed overfitting characteristics [48,49]. The predictive ability of such method has been well evidenced in previous development of QSPR approaches for solubility modeling. For example, Saghafi et al. [50] developed a QSPR model for the prediction of CO₂ solubility in aqueous solution of diethanolamine plus methyldiethanolamine based on random forest algorithm, with an average absolute relative deviation of 3.74%. More recently, random forest based QSPR models have been constantly reported to establish the relationships between CO₂ solubility and IL molecular structures, and the results show that all of the models can give reliable predictions [51–54]. Considering the promising performance aforementioned, the random forest is thus employed here for the prediction of CO₂ solubility in various DESs under different temperatures and pressures.

Taking account of the essential aspects mentioned above, this work aims to develop a predictive QSPR model for CO₂ solubility in DESs with random forest based on the COSMO-RS-derived descriptors. To begin with, a comprehensive database consisting of 1011 CO₂ solubility data

measured in various types of DESs at wide temperature and pressure ranges is first established. The COSMO-RS-derived descriptors of involved HBAs and HBDs of DESs are then addressed by quantum chemistry method. Based on the database and input parameters, the QSPR model is established and validated, offering guides for the selection of DES components in terms of CO₂ capacity.

2. Method description

The whole framework for establishing machine learning based QSPR model is illustrated in Fig. 1, which mainly consists of four steps:

Step 1. Collect the available experimental data as the basis for QSPR model and specify the input variables that have effects on the target property.

Step 2. Optimize the involved molecular structures and calculate the structural descriptors as input parameters for QSPR model.

Step 3. Employ an efficient machine learning algorithm to establish the structure-property relationship.

Step 4. Output the target predictions and assess the accuracy and reliability of the developed QSPR model.

In the following, more details on the development of QSPR model for CO₂ solubility in DESs are elaborated.

2.1. CO₂ solubility dataset

In this contribution, 1011 data points on the CO₂ solubility in 59 different DESs (solubility range of 0.0084–6.4231 mol kg⁻¹) covering a wide range of temperatures (293.15–343.15 K) and pressures

Table 1

List of DES compositions, experimental temperature and pressure as well as CO₂ solubility ranges collected in this study.

HBA	HBD	Molar ratio	T range/K	P range/MPa	CO ₂ solubility/(mol kg ⁻¹)
AChCl	Guaiacol	1:3, 1:4, 1:5	293.15–323.15	0.0537–0.5358	0.0118–0.1960
ATPPB	Diethylene glycol	1:4, 1:10, 1:16	303.15	0.16–1.954	0.2272–7.2872
ATPPB	Triethylene glycol	1:4, 1:10, 1:16	303.15	0.142–1.957	0.1437–6.4231
TBPC	Glycerol	1:12	298.15	1	0.4686
TBPB	Ethylene glycol	1:12	298.15	1	0.6018
ChCl	Glycerol	1:2	303.15–333.15	0.187–6.347	0.0538–3.6929
ChCl	Urea	1:2	303.15–343.15	0.299–5.911	0.1562–3.5592
ChCl	Ethylene glycol	1:2	303.15–343.15	0.236–6.323	0.0716–3.1265
ChCl	Phenol	1:2, 1:3, 1:4	293.15–303.15	0.099–0.5291	0.0222–0.2108
ChCl	Diethylene glycol	1:3, 1:4	293.15–303.15	0.1104–0.5269	0.0203–0.1852
ChCl	Triethylene glycol	1:3, 1:4	293.15–303.15	0.1093–0.5203	0.0225–0.1941
ChCl	Urea	1:2	309–329	0.0405–0.1535	0.0105–0.0406
ChCl	Ethylene glycol	1:2	309–329	0.036–0.156	0.0084–0.0387
ChCl	Guaiacol	1:3, 1:4, 1:5	293.15–323.15	0.0469–0.5459	0.0109–0.1885
ChCl	Levulinic acid	1:3, 1:4, 1:5	303.15–333.15	0.06–0.5874	0.0201–0.2869
ChCl	Furfuryl alcohol	1:3, 1:4, 1:5	303.15–333.15	0.0652–0.5864	0.0168–0.2276
ChCl	1,4-butanediol	1:3, 1:4	293.15–323.15	0.1058–0.5259	0.0217–0.1624
ChCl	2,3-butanediol	1:3, 1:4	293.15–323.15	0.1071–0.5288	0.0251–0.1915
ChCl	1,2-propanediol	1:3, 1:4	293.15–323.15	0.1044–0.5256	0.0206–0.1884
ChCl	Urea	1:1.5, 1:2, 1:2.5	313.15–333.15	1–12.73	0.3715–5.1673
ChCl	Urea	1:4, 1:2.5	298.15	1	0.2604–0.3237
ChCl	Glycerol	1:3, 1:8	298.15	1	0.3242–0.4574
ChCl	Ethanolamine	1:6	298.15	1	0.2604
ChCl	Diethanolamine	1:6	298.15	1	0.4574
DEA-HCl	Guaiacol	1:3, 1:4, 1:5	293.15–323.15	0.0451–0.5223	0.0197–0.2321
TMPB	Ethanolamine	1:6, 1:7, 1:8	298	1	1.436–1.6286
TBAB	Ethanolamine	1:6	298	1	1.3838
TBAB	Diethanolamine	1:6	298	1	0.8487
TBAB	Triethanolamine	1:3	298	1	0.4702
TBAC	Lactic acid	1:2	308–318	0.093–1.992	0.0369–0.9546
TEAC	Lactic acid	1:2	308–318	0.094–1.993	0.0270–0.6783
TMAC	Lactic acid	1:2	308–318	0.095–1.993	0.0239–0.6472

(0.036–12.73 MPa) are collected from previous works. The involved DESs consist of 11 HBAs including choline chloride (ChCl), acetylcholine chloride (AChCl), allyltriphenylphosphonium bromide (ATPPB), tetrabutylphosphonium chloride (TBPC), dimethylaminopropane hydrochloride (DEA·HCl), tetramethylphosphonium bromide (TMPB), tetrabutylammonium bromide (TBAB), tetrabutylammonium chloride (TBAC), tetrabutylphosphonium bromide (TBPB), tetraethylammonium chloride (TEAC) and tetramethylammonium chloride (TMAC), and 16 HBDs containing 1,4-butanediol, 2,3-butanediol, diethanolamine, diethyleneglycol (DEG), ethanolamine, ethyleneglycol, glycerol, lactic acid, furfuryl alcohol, guaiacol (Gau), levulinic acid, phenol, propanediol, triethanolamine, triethyleneglycol (TEG), and urea. The detailed information of the collected CO₂ solubility, DES compositions, the experimental temperatures and pressures together with the corresponding references are given in Supporting Information, Table S1, and a summarization of these data points is listed in Table 1.

2.2. Calculation of COSMO-RS-derived descriptors

The structural parameters of HBAs and HBDs are generated by quantum chemistry calculations using the Gaussian 09 software package (version D.01). First, geometric optimizations of all the structures of HBAs and HBDs are carried out at the B3LYP/6–31++G** theoretical level in the ideal gas phase, and vibrational frequency analysis is conducted to ensure the configuration in global energy minimum. Then, the COSMO files of the stable conformers can be acquired based on single-point quantum COSMO calculation using the BP86/TZVP level of theory. Here, it should be noted that different anion locations around the cation are taken into consideration for the initial configurations of the involved HBAs. Optimized HBA and HBD structures with COSMO surfaces are shown in Fig. 2, where the red part and the blue part represent positive COSMO charge density and negative charge distribution, respectively.

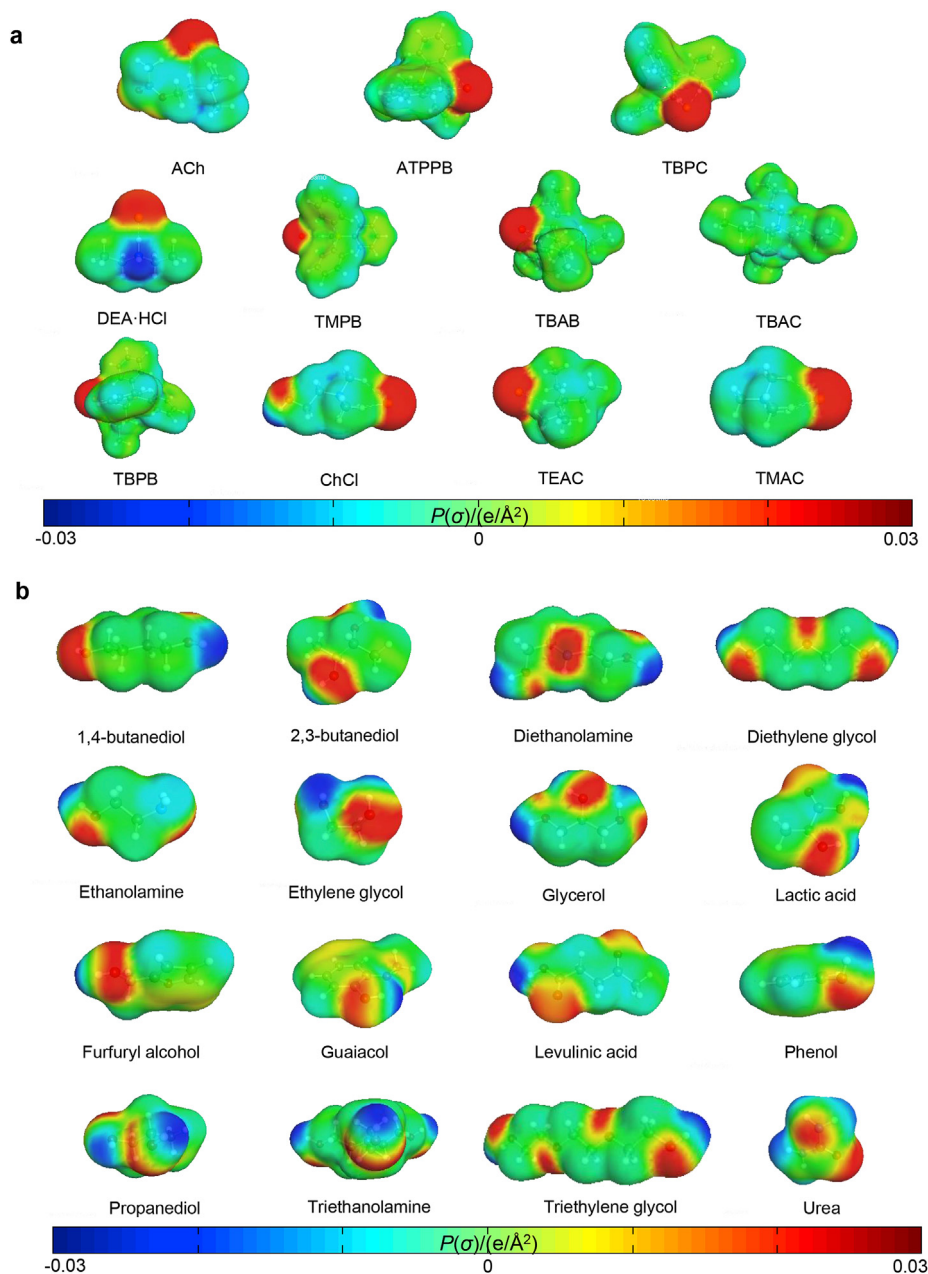


Fig. 2. Optimized (a) HBA and (b) HBD structures with COSMO surfaces.

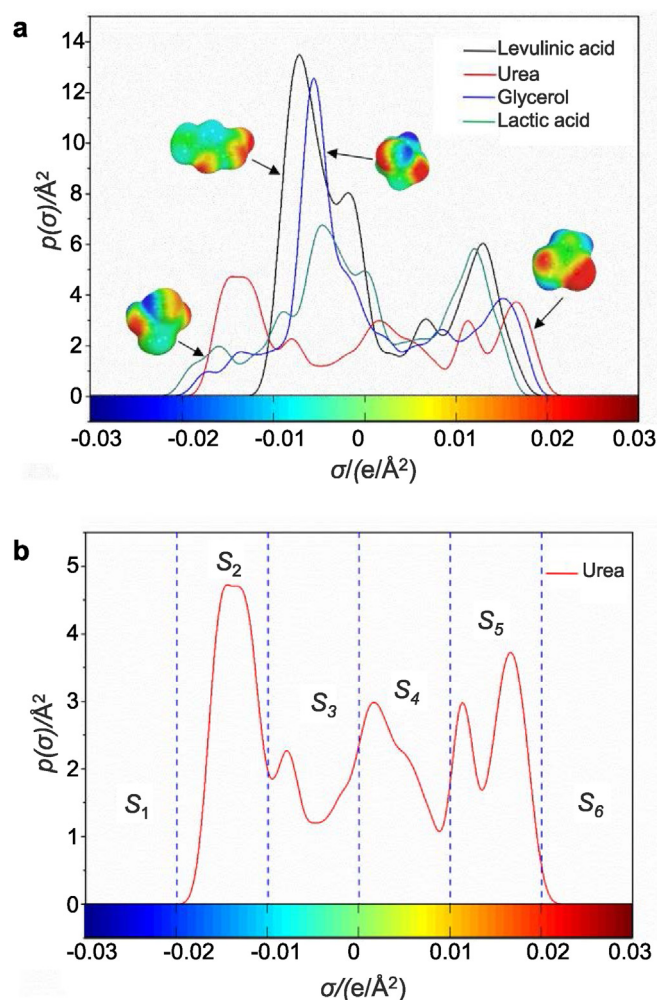


Fig. 3. (a) σ -Profiles of four representative HBD solvents: levulinic acid, urea, glycerol and lactic acid; and (b) COSMO areas defined by the σ -profile (urea for illustration).

In order to generate the desired COSMO-RS-derived descriptors, the obtained COSMO files are directly imported into the COSMOthermX software (version C30_1401) to output the σ -profiles of HBAs and HBDs. Fig. 3a displays the σ -profiles of four HBDs, namely levulinic acid, urea, glycerol and lactic acid, where the σ -profile of each compound is composed of 61 elements with a screening charge density range of -0.03 – 0.03 $e/\text{\AA}^2$. As shown, the σ -profile distribution in hydrogen bond donor and acceptor regions as well as its area of the above four HBDs vary remarkably, indicating the molecule-specific characteristic of σ -profile. To decrease the vector of the input parameters for QSPR model, each σ -profile of HBAs and HBDs are divided into 6 regions by integrating

those curves over σ . As exemplified by urea in Fig. 3b, the 6 under-curve areas include: S1: -0.03 $e/\text{\AA}^2 \leq \sigma \leq -0.02$ $e/\text{\AA}^2$, S2: -0.02 $e/\text{\AA}^2 \leq \sigma \leq -0.01$ $e/\text{\AA}^2$, S3: -0.01 $e/\text{\AA}^2 \leq \sigma \leq 0$ $e/\text{\AA}^2$, S4: 0 $e/\text{\AA}^2 \leq \sigma \leq 0.01$ $e/\text{\AA}^2$, S5: 0.01 $e/\text{\AA}^2 \leq \sigma \leq 0.02$ $e/\text{\AA}^2$ and S6: 0.02 $e/\text{\AA}^2 \leq \sigma \leq 0.03$ $e/\text{\AA}^2$. For the COSMO volumes (V_{COSMO}) of HBAs and HBDs, they can be directly obtained from the information in the COSMO files. The detailed COSMO information, i.e., S1–S6 and V_{COSMO} , of the involved HBAs and HBDs in this work is listed in Table 2 and Table 3, respectively.

2.3. Random forest algorithm

Random forest, first proposed by Breiman [55] in 2001, is an ensemble training algorithm that constructs multiple decision trees. The main features of random forest are listed as follows: (1) Each decision tree in random forest is independent and can be processed in parallel during the data classification or regression, and thus reduce the load and time demands; (2) As the same way done in bagging, the random forest is able to suppress overfitting to the training samples by random selection of training samples for tree construction, which makes it robust against noise; (3) The random selection of features to be used at splitting nodes enables fast training of this algorithm, even in the case of large dimensionality of the feature vector; (4) Random forest can assess the importance of the descriptors to the QSPR model by investigate the deterioration in model quality with the certain descriptor reduced. Because of the advantages mentioned above, the random forest has been extensively applied in many fields, and the detailed description can be referred to previous works [48,49].

In this study, an open resource of RandomForestRegressor method in the scikit-learn module of Python programming software (version 3.2) is employed for the QSPR modeling, and the involved parameter values in this case for CO_2 solubility prediction in DESs are summarized in Table 4. It should be highlighted that the whole data set in Table 1 are randomly divided into a training set (711 data points, about 70% of the data) to establish the QSPR model and a test set of the remaining 300 data points to evaluate the predictive accuracy of the obtained model. The information of the data points contained in the training set and test set are shown in Supporting Information Tables S2–S3, respectively.

2.4. Model validation and performance

For providing intuition into the predictive power of the developed model, different statistical parameters, including the coefficient of determination (R^2), average absolute relative deviation (AARD), mean-square error (MSE) and root-mean-square error (RMSE) are determined as follows:

$$R^2 = \frac{\sum_{i=1}^{N_p} (y_i - \bar{y}_m)^2 - \sum_{i=1}^{N_p} (y_i^{\text{cal}} - y_i)^2}{\sum_{i=1}^{N_p} (y_i - \bar{y}_m)^2} \quad (1)$$

Table 2

Calculated COSMO-RS-derived descriptors of the involved HBAs.

HBA	S1	S2	S3	S4	S5	S6	V_{COSMO}
AChCl	0	23.277	114.265	26.692	48.661	0	221.182
ATPPB	0	14.188	183.196	114.328	40.439	0	414.4741
TBPC	0	5.182	205.556	136.604	34.196	0	460.7552
TBPB	0	4.772	214.109	122.572	38.839	0	438.8105
ChCl	0.03	22.81	93.368	19.736	44.037	0	175.9708
DEA-HCl	0.034	10.085	96.2	18.973	33.867	0	147.4631
TMPB	0	6.96	183.749	105.459	39.204	0	377.8873
TBAB	0	1.885	250.102	73.352	40.782	0.107	393.383
TBAC	0	1.07	249.257	73.749	35.676	0.931	386.9221
TEAC	0	4.363	143.139	23.69	36.1	0	224.0953
TMAC	0	13.133	91.788	11.195	37.451	0	146.3961

$$AARD (\%) = 100 \times \sum_{i=1}^{N_p} \left| \frac{(y_i^{\text{cal}} - y_i)}{y_i} \right| / N_p \quad (2)$$

$$MSE = \sum_{i=1}^{N_p} (y_i^{\text{cal}} - y_i)^2 / N_p \quad (3)$$

$$RMSE = \sqrt{\sum_{i=1}^{N_p} (y_i^{\text{cal}} - y_i)^2 / N_p} \quad (4)$$

where N_p is the total number of the whole set; y and \bar{y}_m represent the experimental CO₂ solubility in DESs and the average of the experimental data; and the superscript “cal” denotes the calculated CO₂ solubility by random forest model.

3. Results and discussion

3.1. Qualitative analysis of CO₂ solubility in DESs

As demonstrated in previous studies, the CO₂ solubility in the DESs grows with increasing pressure (P) while decreases with increasing temperature (T). Moreover, it is found that both the type and molar ratio of HBA and HBD have notable effects on the CO₂ solubility under identical conditions [9]. Taking account of these aspects, the variables of T , P , HBA:HBD molar ratio, together with the COSMO-RS-derived descriptors of HBAs and HBDs in Tables 2 and 3 are employed as the input parameters for the QSPR model for predicting CO₂ solubility in DESs.

Before developing the QSPR model using the random forest algorithm, the efficiency of the input variables is inspected beforehand. For this purpose, multiple linear regression (MLR) is used to build a simple linear model using the above-mentioned variables based on the 1011 collected experimental data, and the MLR model is expressed as follows:

$$y = 8.0821 - 16.4733S1 + 0.1674S2 - 0.0064S3 - 0.0394S4 - 0.2791S5 - 0.7656S6 + 0.0241V_{\text{COSMO}} - 0.5000S1' - 0.0641S2' - 0.0396S3' - 0.0115S4' + 0.0483S5' + 0.5288S6' + 0.0246V'_{\text{COSMO}} + 0.2211Ratio - 1.5913T + 0.0324P \quad (5)$$

where y represents the CO₂ solubility in DESs; P and T denote the pressure and temperature, respectively; $S1 - S6$ and V_{COSMO} are the COSMO-RS-derived descriptors of HBAs while $S1' - S6'$ and V'_{COSMO} for the HBDs; and $Ratio$ means the molar ratio of HBA to HBD.

In Eq. (5), the plus and minus sign before the variables suggest the

Table 4

Parameter values of the random forest algorithm used in this work.

Parameters	Value	Parameters	Value
n_estimators	500	max_leaf_nodes	None
criterion	mse	min_impurity_decrease	0
max_features	auto	bootstrap	True
max_depth	None	oob_score	True
min_samples_split	2	n_jobs	-1
min_samples_leaf	1	random_state	1
min_weight_fraction_leaf	1	verbose	0

positive and negative effect of the variables on CO₂ solubility in DESs, respectively. As displayed, the minus and plus signs in front of temperature and pressure indicate that the decrease of T and increase of P are favorable for CO₂ capacity in DESs, which is agree well with the gas-insolvent dissolution characteristics [12–14]. Moreover, by comparing the absolute t value of the introduced COSMO-RS-derived descriptors in Eq. (5), $S1$, $S6$, $S1'$ and $S6'$ are found to be the most important molecular structure descriptors, which correspond to the polar regions of HBAs and HBDs. Such finding shows that the DES polarity has a larger effect on the CO₂ solubility than the non-polarity of DES, accounting for the molecular interactions between HBA and HBD of DES [56]. Generally, lower polarity of the HBA and HBD molecules allow weaker HBA-HBD intermolecular interactions in DES systems, which is believed to be beneficial for easier expansion of the lattice and more CO₂ insertion into the interstitial space (free volume) of IL-type solvents [57]. The negative correlation between the DES polarity and CO₂ solubility is in good accordance with the dominant minus sign before the polar region descriptors (i.e., $S1$, $S6$, $S1'$ and $S6'$). In addition, the HBA and HBD type are suggested to have larger effect on the CO₂ solubility in DESs compared to the HBA:HBD molar ratio by the higher absolute t value before the COSMO-RS-derived descriptors over $Ratio$ from Eq. (5), which is also consistent with the previous studies [12].

From above, Eq. (5) can give qualitative correlation of the CO₂ solubility in DESs, indicating the feasibility for establishing QSPR model using the proposed input variables. However, the prediction results based on such model is still far from adequate from the quantitative point of view. To be specific, the R^2 value of the MLR model for the total set is only 0.7498 with the $AARD$ up to 113.89%. To this end, the MLR model is considered to be “qualitative at best” but is not sufficient enough as a quantitatively predictive model for CO₂ solubility in DESs. Thus, a nonlinear model is developed to obtain more accurate quantitative results in the next section.

Table 3

Calculated COSMO-RS-derived descriptors of the involved HBDs.

HBD	S1	S2	S3	S4	S5	S6	V_{COSMO}
1,4-butanediol	0	14.42	76.864	25.817	20.875	0	119.182
2,3-butanediol	0	13.174	62.691	40.315	16.863	0	118.3448
Diethanolamine	0	15.464	83.682	31.256	21.617	1.034	134.1709
Diethylene glycol	0	15.292	81.126	30.835	23.825	0	130.5433
Ethanolamine	0	12.397	53.584	20.206	14.719	2.414	83.03382
Ethylene glycol	0	13.877	47.2	17.918	18.443	0	78.58532
Furfuryl alcohol	0.006	9.826	57.113	55.757	9.014	0	117.3356
Glycerol	0	15.032	60.236	29.498	20.279	0	107.9718
Guaiaicol	0	4.013	79.641	68.406	5.874	0	149.3569
Levulinic acid	1.064	7.052	87.705	32.21	25.549	0	138.0178
Phenol	0.129	7.534	58.8	57.839	7.259	0	119.1968
Propanediol	0	10.608	60.947	29.059	15.437	0	98.35612
Triethanolamine	0	22.197	100.873	40.486	31.335	0	184.6856
Triethylene glycol	0	15.289	117.831	41.676	28.461	0	182.4924
Urea	0	27.009	19.404	23.644	20.404	0	70.58779
Lactic acid	1.508	11.947	53.342	33.602	18.709	0	102.3986

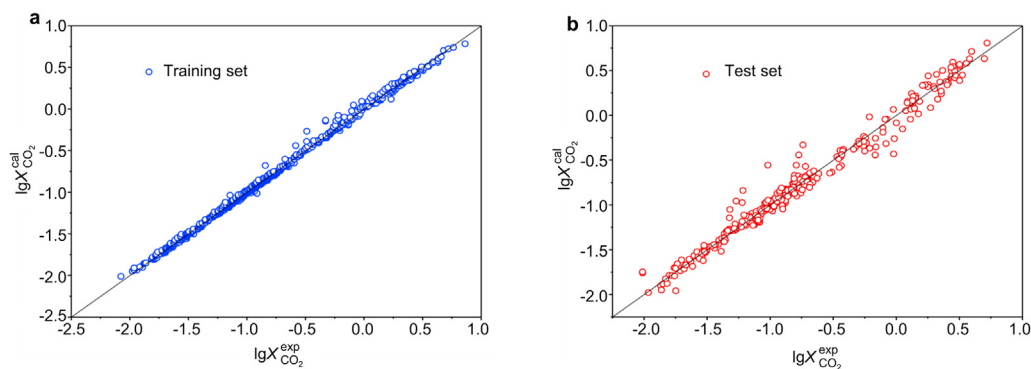


Fig. 4. Comparisons between the experimental and predicted CO₂ solubility in DESs of the (a) training and (b) test set.

3.2. Results of the QSPR model based on random forest algorithm

3.2.1. Validation of the QSPR model

In order to establish a more accurate model, the random forest algorithm is employed to build nonlinear model based on the same input parameters used in the MLR model. Fig. 4 illustrates the comparison of experimental and predicted CO₂ solubility in the training and test set, and the statistical parameters for the QSPR model including R^2 , AARD, MSE, and RMSE are listed in Table 5. As depicted in the parity plot Fig. 4a, the predictions for the training set are in very agreement with the experimental ones with a R^2 value of 0.9903. Despite a larger deviation between the QSPR predictions and experimental data in Fig. 4b, the overall data in the test set still fall almost evenly in a small range close to the diagonal with a R^2 of 0.9500. These results correspond well to the statistical parameters in Table 5, where the statistical values of the test set are slightly larger than those of the training set, suggesting that no overfitting occurs in this work. For the total data set, the R^2 and AARD are 0.9758 and 7.76%, respectively, which are all in a very desirable level. Moreover, the majority of the overall data (81%) present AARDs lower than 10% while for only 8% of them, the AARDs go beyond 20%. Such fact clearly demonstrates the good CO₂ solubility prediction accuracy of the developed QSPR model.

To further estimate the reliability of the developed QSPR model, the predicted effects of the input variables (i.e., temperature, pressure, the type of HBA and HBD and their molar ratio) on CO₂ solubility are investigated and compared to the experimental measurements. As exemplified in Fig. 5a, in the temperature arrange of 293.15–323.15 K, the CO₂ solubility in the ChCl:glycerol (1:3) grows almost linearly by increasing pressure with the extrapolation line basically passing through the origin of the coordinate, and meanwhile decreases significantly with the increase of temperature. These predictions are consistent with the experimental results.

Fig. 5b illustrates the QSPR calculated effects of HBD type and HBA:HBD molar ratio on the CO₂ solubility in DESs. As previously reported by Ghaedi [58], by mixing with the same HBA of ATPPB at the certain mole ratio, the TEG-based DES has a higher CO₂ solubility than the DEG-based one for its larger free volume and stronger van der Waals molecular interaction with CO₂ molecules. The same results are again observed from Fig. 5b, where the CO₂ solubility in the ATPPB-based DESs follow the ranking ATPPB: TEG (1:10/1:16) > ATPPB:DEG (1:10/1:16). Moreover, from Ghaedi's work, it was found that the CO₂ solubility in ATPPB:TEG/DEG types of DESs decreases slightly as the HBA:HBD molar ratio changes from 1:10 to 1:16, which is also supported by the QSPR predicted results in Fig. 5b.

For analyzing the effect of HBA type, the predicted CO₂ solubility in the AChCl- and ChCl-based DESs are compared when mixing with guaiacol under identical ratios. The main CO₂ solubility difference between these two DESs lies in the –COO group in AChCl, which has been experimentally confirmed to be beneficial for improving the CO₂ capacity due to its strong intermolecular interactions with CO₂ [57].

Therefore, the higher CO₂ solubility in the AChCl-based DES in Fig. 5c than that in the ChCl-based one again indicates the rationality of the proposed QSPR model.

From above, the established QSPR model can not only give accurate CO₂ solubility predictions, but also well interpret the effects of the input variables. Therefore, it can be concluded that the proposed QSPR model in this work can be employed as a reliable tool for predicting CO₂ absorption capacity in DESs.

3.2.2. Importance order of the variables in QSPR model

Apart from outputting the target CO₂ solubility by the QSPR model, the significance of involved variables also can be sorted by comparing the prediction accuracy before and after reducing the certain variable. Fig. 6 displays the importance ranks of the input variables, where the importance of HBA and HBD type is identified as the sum of its corresponding COSMO-RS-derived descriptors (i.e., $S1 - S6$ and V_{COSMO} for HBA; $S1' - S6'$ and V'_{COSMO} for HBD). As shown, as the headmost one, the pressure has the notably largest influence (0.7867) on CO₂ solubility in DESs for two main reasons: (1) Most of the data set concerns the CO₂ solubility at low pressures, where the CO₂ solubility changes linearly following the pressure. (2) Compared to the certain set of temperature, HBA and HBD type and their molar ratio, the pressure of the collected data distributes at a significantly wider range from 0.036 to 12.73 MPa, and thus the developed QSPR is more sensitive to the variable of pressure.

Beyond the experimental conditions, the importance of DES compositions is in the order HBA (0.0835) > HBD (0.0731) > HBA:HBD molar ratio (0.0339), again suggesting the larger influence of HBA and HBD type over their molar ratio on the CO₂ solubility. Moreover, the dominant role of the HBA agrees well with the molecular dynamics simulation results in our recent work, where the HBA is found to have higher contribution to the DES-CO₂ interactions [9]. Overall, the importance order of the above variables could offer useful guides for the selection of promising DES and operating conditions for CO₂ capture.

3.2.3. Comparison between the developed QSPR and literature-reported models

The above discussions have demonstrated the performance of the proposed QSPR model. For the sake of comparison, the CO₂ solubility prediction accuracy derived from QSPR model is first compared to that of COSMO-RS model, which is widely used as predictive thermodynamic tool in IL-containing systems. Here, the COSMO-RS calculations are performed with the COSMOthermX software (version C30_1401), where

Table 5
Statistical parameters for the developed QSPR model.

Data set	No.	R^2	AARD, %	MSE	RMSE
Training set	711	0.9903	4.44	0.006902	0.08303
Test set	300	0.9500	14.59	0.04788	0.2188
Total set	1011	0.9758	7.76	0.01906	0.1381

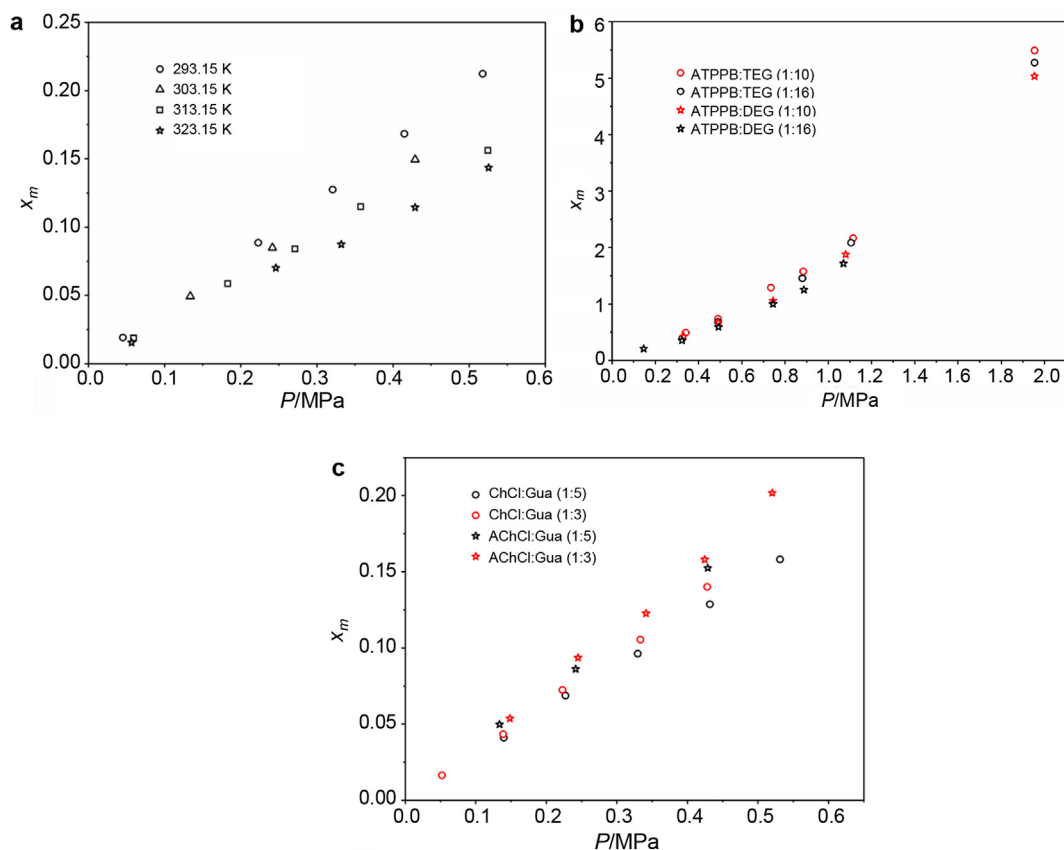


Fig. 5. Predicted CO_2 solubility in (a) ChCl:glycerol (1:3) under different temperatures and pressures; (b) ATPPB-based DESs and (c) ChCl and AChCl-based DESs using the developed QSPR model.

the only required input, COSMO files of HBAs and HBDs, are the same as the origins for COSMO-RS-derived descriptors for the developed QSPR model. As summarized in Table 6, the AARD of the studied QSPR model ranges from 2.47%–21.90% for each DES system, significantly lower than that of the COSMO-RS model (42.91%–88.17%). For the total 1011 data points, the average AARD of the QSPR model is only 7.76%, whereas up to 64.81% in the case of COSMO-RS model, demonstrating the higher reliability of the QSPR model.

Furthermore, the predicted CO_2 solubility predictions are compared between the QSPR model and traditional models [17–20]. As seen from Table 7, although the AARD of the developed QSPR model (7.76%) is

found to be slightly higher than the traditional thermodynamic models (i.e., NRTL (3.10%), UNIQUAC (2.90%)) and equations of state methods (i.e., PC-SAFT (3.97%), PR-EOS (0.80%) and soft-SAFT (0.22%–4.95%)), prediction accuracy of the QSPR model for CO_2 solubility in DESs is also in an acceptable level. Considering the inapplicability of the traditional model for novel systems, the QSPR model established in this contribution based on the random forest and COSMO-RS-derived descriptors is more suitable for predicting CO_2 solubility in DESs over wide temperature and pressure ranges.

4. Conclusions

A QSPR model based on COSMO-RS-derived descriptors as inputs has been developed for the prediction of the CO_2 solubility in different types of DESs at various temperatures and pressures using the random forest algorithm. The qualitative analysis from multiple linear regression shows that the variables, including the COSMO-RS-derived descriptors of HBAs and HBDs, molar ratio of HBA to HBD as well as the temperature and pressure, are efficient input parameters for predicting CO_2 solubility in DESs. Then, the nonlinear QSPR model established using random forest algorithm is demonstrated to have good predictive performance with high R^2 (0.9758) and low AARD (7.76%) values, and can well interpret the influences of such input variables. Meanwhile, the importance of the involved variables in the QSPR model is ranked as pressure > HBA type > HBD type > HBA:HBD molar ratio > temperature. Finally, the comparison between the CO_2 solubility predictions by the QSPR model and literature-reported models (COSMO-RS model, traditional thermodynamic models and equations of state methods) again indicates the reliability of the developed QSPR model. In conclusion, the obtained QSPR model can provide fast estimations of CO_2 solubility in DESs in the cases where experimental measurements are difficult or costly. On the

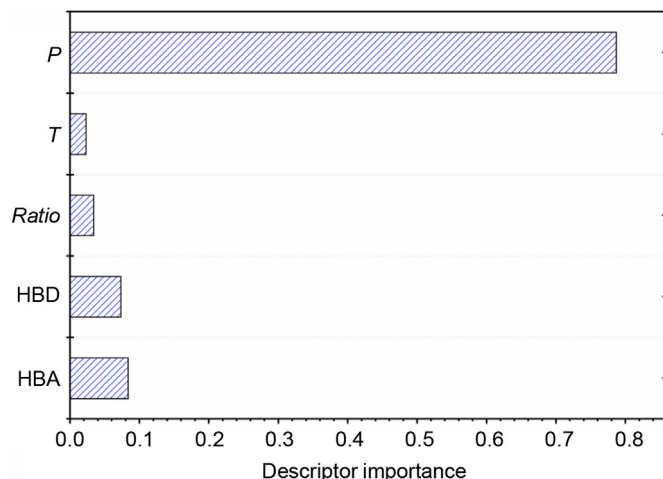


Fig. 6. Importance order of the input variables for the developed QSPR model.

Table 6

Comparison of random forest-based QSPR and COSMO-RS models for CO₂ solubility in DESs.

HBA	HBD	No. of data points	AARD, %	
			Random forest	COSMO-RS
AChCl	Guaiacol	72	3.72	71.48
ATPPB	Diethylene glycol	24	2.81	47.89
ATPPB	Triethylene glycol	24	8.76	65.49
TBPC	Glycerol	1	7.91	67.04
TBPB	Ethylene glycol	1	9.70	63.79
ChCl	Glycerol	40	15.72	69.35
ChCl	Urea	34	21.90	50.43
ChCl	Ethylene glycol	40	12.28	66.99
ChCl	Phenol	60	8.06	74.10
ChCl	Diethylene glycol	40	10.01	81.08
ChCl	Triethylene glycol	40	6.15	69.87
ChCl	Urea	21	4.92	75.63
ChCl	Ethylene glycol	18	8.12	75.27
ChCl	Guaiacol	72	5.39	76.16
ChCl	Levulinic acid	72	7.28	73.04
ChCl	Furfuryl alcohol	72	2.47	58.07
ChCl	1,4-butanediol	40	13.80	63.93
ChCl	2,3-butanediol	40	8.23	62.42
ChCl	1,2-propanediol	40	10.24	59.53
ChCl	Urea	68	5.26	61.88
ChCl	Urea	2	3.10	72.52
ChCl	Glycerol	2	4.17	53.89
ChCl	Ethanolamine	1	7.31	88.17
ChCl	Diethanolamine	1	9.30	78.27
DEA·HCl	Guaiacol	72	8.13	49.32
TMPB	Ethanolamine	3	5.14	64.10
TBAB	Ethanolamine	1	10.20	70.91
TBAB	Diethanolamine	1	11.24	67.81
TBAB	Triethanolamine	1	7.42	71.03
TBAC	Lactic acid	28	7.30	42.91
TEAC	Lactic acid	40	5.51	59.42
TMAC	Lactic acid	40	5.35	60.21
Total		1011	7.76	64.81

other hand, this model can be incorporated into a solvent design framework to identify suitable DESs for carbon capture processes.

It is worth mentioning that the experimental solubility data, especially the cases of very low solubilities, may have large uncertainties. Usually, these uncertainties should be taken into account in model development. However, since most of the available literature has not reported the data uncertainties, they are not considered in the present work. In the future, an extension and modification of the database should be carried out to further improve the prediction accuracy of the QSPR model. Of course, alternative structural presentations could also be

Table 7

Comparison of random forest-based QSPR and traditional thermodynamic models as well as equations of state methods.

Models	DESs	AARD, %
NRTL [17]	ChCl/TEAC/AChCl/TEAB/TBAC/TBAB:levulinic acid (1:3), ChCl:glycerol (1:1, 1:2, 1:3, 1:4), ChCl:ethylene glycol/urea (1:2), ChCl:malonic acid/thiourea (1:1)	3.10
UNIQUAC [17]	ChCl/TEAC/AChCl/TEAB/TBAC/TBAB:levulinic acid (1:3), ChCl:glycerol (1:1, 1:2, 1:3, 1:4), ChCl:ethylene glycol/urea (1:2), ChCl:malonic acid/thiourea (1:1)	2.90
PC-SAFT [18]	TMAC/TEAC/TBAC:lactic acid (1:2), TBAC:lactic acid (1:3)	3.97
PR-EOS [19]	ChCl:urea/ethyleneglycol (1:2), ChCl:malic acid:ethylene glycol (1.3:1:2.2)	0.80
Soft-SAFT [20]	ChCl:ethylene glycol/lactic acid/glycerol (1:2)	0.22–4.95
QSPR of this work	As listed in Table 1 in this manuscript	7.76

utilized as inputs for QSPR model to enable faster prediction of DES properties, including CO₂ solubility.

Declaration of interests

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.

Acknowledgement

The financial support from National Natural Science Foundation of China (21861132019 and 21776074) is greatly acknowledged.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.gce.2021.08.002>.

List of Symbols

σ	Surface screening charge density, e/Å ²
S_i	Surface area with a charge density of σ , Å ²
V_{COSMO}	Surface volume, Å ³
P	Pressure, MPa
T	Temperature, K
<i>Ratio</i>	Molar ratio of HBA to HBD
y	CO ₂ solubility in DESs, mol kg ⁻¹

Abbreviations Used

ChCl	choline chloride
AChCl	acetylcholine chloride
ATPPB	allyltriphenylphosphonium bromide
TBPC	tetrabutylphosphonium chloride
DEA·HCl	dimethylaminopropane hydrochloride
TMPB	tetramethylphosphonium bromide
TBAB	tetrabutylammonium bromide
TBAC	tetrabutylammonium chloride
TBPB	tetrabutylphosphonium bromide
TEAC	tetraethylammonium chloride
TMAC	tetramethylammonium chloride
DEG	diethyleneglycol
TEG	triethyleneglycol
Gau	guaiacol

References

- [1] Q. Zhang, K. De Oliveira Vigier, S. Royer, F. Jerome, Deep eutectic solvents: syntheses, properties and applications, *Chem. Soc. Rev.* 41 (2012) 7108–7146.
- [2] B.B. Hansen, S. Spittle, B. Chen, D. Poe, Y. Zhang, J.M. Klein, A. Horton, L. Adhikari, T. Zelovich, B.W. Doherty, B. Gurkan, E.J. Maginn, A. Ragauskas, M. Dadmun, T.A. Zawodzinski, G.A. Baker, M.E. Tuckerman, R.F. Savinell, J.R. Sangoro, Deep eutectic solvents: a review of fundamentals and applications, *Chem. Rev.* 121 (2021) 1232–1285.
- [3] E.L. Smith, A.P. Abbott, K.S. Ryder, Deep eutectic solvents (DESs) and their applications, *Chem. Rev.* 114 (2014) 11060–11082.
- [4] H.Y. Cheng, Z.W. Qi, Applications of deep eutectic solvents for hard-to-separate liquid systems, *Separ. Purif. Technol.* 274 (2021) 119027.
- [5] R.Z. Wang, H. Qin, J.W. Wang, H.Y. Cheng, L.F. Chen, Z.W. Qi, Reactive extraction for intensifying 2-ethylhexyl acrylate synthesis using deep eutectic solvent [Im: 2PTSA], *Green Energy Environ.* 6 (2021) 405–412.
- [6] G. Garcia, S. Aparicio, R. Ullah, M. Atilhan, Deep eutectic solvents: physicochemical properties and gas separation applications, *Energy Fuels* 29 (2015) 2616–2644.
- [7] Y. Liu, Z. Dai, Z. Zhang, S. Zeng, F. Li, X. Zhang, Y. Nie, L. Zhang, S. Zhang, X. Ji, Ionic liquids/deep eutectic solvents for CO₂ capture: reviewing and evaluating, *Green Energy Environ.* 6 (2021) 314–328.
- [8] Z. Song, X. Hu, H. Wu, M. Mei, S. Linke, T. Zhou, Z. Qi, K. Sundmacher, Systematic screening of deep eutectic solvents as sustainable separation media exemplified by the CO₂ capture process, *ACS Sustain. Chem. Eng.* 8 (2020) 8741–8751.
- [9] J. Wang, H. Cheng, Z. Song, L. Chen, L. Deng, Z. Qi, Carbon dioxide solubility in phosphonium-based deep eutectic solvents: an experimental and molecular dynamics study, *Ind. Eng. Chem. Res.* 58 (2019) 17514–17523.

- [10] Y. Zhang, X. Ji, X. Lu, Choline-based deep eutectic solvents for CO₂ separation: review and thermodynamic analysis, *Renew. Sustain. Energy Rev.* 97 (2018) 436–455.
- [11] X. Li, M. Hou, B. Han, X. Wang, L. Zou, Solubility of CO₂ in a choline chloride + urea eutectic mixture, *J. Chem. Eng. Data* 53 (2008) 548–550.
- [12] Y. Chen, N. Ai, G. Li, H. Shan, Y. Cui, D. Deng, Solubilities of carbon dioxide in eutectic mixtures of choline chloride and dihydric alcohols, *J. Chem. Eng. Data* 59 (2014) 1247–1253.
- [13] R.B. Leron, A. Caparanga, M.-H. Li, Carbon dioxide solubility in a deep eutectic solvent based on choline chloride and urea at T = 303.15 – 343.15 K and moderate pressures, *J. Taiwan Inst. Chem. Eng.* 44 (2013) 879–885.
- [14] R.B. Leron, M.H. Li, Solubility of carbon dioxide in a eutectic mixture of choline chloride and glycerol at moderate pressures, *J. Chem. Thermodyn.* 57 (2013) 131–136.
- [15] J. Wang, Z. Song, H. Cheng, L. Chen, L. Deng, Z. Qi, Computer-aided design of ionic liquids as absorbent for gas separation exemplified by CO₂ capture cases, *ACS Sustain. Chem. Eng.* 6 (2018) 12025–12035.
- [16] C.H. Jiang, H.Y. Cheng, Z.X. Qin, R.Z. Wang, L.F. Chen, C. Yang, Z.W. Qi, X.C. Liu, COSMO-RS prediction and experimental verification of 1,5-pentanediamine extraction from aqueous solution by ionic liquids, *Green Energy Environ.* 6 (2021) 422–431.
- [17] R. Haghbakhsh, S. Raeissi, Modeling vapor-liquid equilibria of mixtures of SO₂ and deep eutectic solvents using the CPA-NRTL and CPA-UNIQUAC models, *J. Mol. Liq.* 250 (2018) 259–268.
- [18] L.F. Zubeir, C. Held, G. Sadowski, M.C. Kroon, PC-SAFT modeling of CO₂ solubilities in deep eutectic solvents, *J. Phys. Chem. B* 120 (2016) 2300–2310.
- [19] N.R. Mirza, N.J. Nicholas, Y. Wu, K.A. Mumford, S.E. Kentish, G.W. Stevens, Experiments and thermodynamic modeling of the solubility of carbon dioxide in three different deep eutectic solvents (DESs), *J. Chem. Eng. Data* 60 (2015) 3246–3252.
- [20] E.A. Crespo, L.P. Silva, J.O. Lloret, P.J. Carvalho, L.F. Vega, F. Lovell, J.A.P. Coutinho, A methodology to parameterize SAFT-type equations of state for solid precursors of deep eutectic solvents: the example of cholinium chloride, *Phys. Chem. Chem. Phys.* 21 (2019) 15046–15061.
- [21] M.B. Haider, R. Kumar, Solubility of CO₂ and CH₄ in sterically hindered amine-based deep eutectic solvents, *Separ. Purif. Technol.* 248 (2020) 117055.
- [22] Y.S. Zhao, R. Gani, R.M. Afzal, X.P. Zhang, S.J. Zhang, Ionic liquids for absorption and separation of gases: an extensive database and a systematic screening method, *AIChE J.* 63 (2017) 1353–1367.
- [23] J.L. Han, C.N. Dai, G.Q. Yu, Z.G. Lei, Parameterization of COSMO-RS model for ionic liquids, *Green Energy Environ.* 3 (2018) 247–265.
- [24] Y.Y. Zhang, X.Y. Ji, Y.J. Xie, X.H. Lu, Screening of conventional ionic liquids for carbon dioxide capture and separation, *Appl. Energy* 162 (2016) 1160–1170.
- [25] J. Yang, Z. Hou, G. Wen, P. Cui, Y. Wang, J. Gao, A brief review of the prediction of liquid-liquid equilibrium of ternary systems containing ionic liquids by the COSMO-SAC Model, *J. Solut. Chem.* 48 (2019) 1547–1563.
- [26] J. Zhang, L. Qin, D. Peng, H. Cheng, L. Chen, Z. Qi, COSMO-descriptor based computer-aided ionic liquid design for separation processes. Part II: Task-specific design for extraction process, *Chem. Eng. Sci.* 162 (2017) 364–374.
- [27] D. Peng, J. Zhang, H. Cheng, L. Chen, Z. Qi, Computer-aided ionic liquid design for separation processes based on group contribution method and COSMO-SAC model, *Chem. Eng. Sci.* 159 (2017) 58–68.
- [28] A. Kamgar, S. Mohsenpour, F. Esmailzadeh, Solubility prediction of CO₂, CH₄, H₂, CO and N₂ in choline chloride/urea as an eutectic solvent using NRTL and COSMO-RS models, *J. Mol. Liq.* 247 (2017) 70–74.
- [29] H.S. Salehi, R. Hens, O.A. Moulton, T.J.H. Vlucht, Computation of gas solubilities in choline chloride urea and choline chloride ethylene glycol deep eutectic solvents using Monte Carlo simulations, *J. Mol. Liq.* 316 (2020) 113729.
- [30] R. Hens, T.J.H. Vlucht, Molecular simulation of vapor-liquid equilibria using the Wolf method for electrostatic interactions, *J. Chem. Eng. Data* 63 (2018) 1096–1102.
- [31] B. Sepehri, A review on created QSPR models for predicting ionic liquids properties and their reliability from chemometric point of view, *J. Mol. Liq.* 297 (2020) 112013.
- [32] A. Khajeh, M. Shakourian-Fard, K. Parvaneh, Quantitative structure-property relationship for melting and freezing points of deep eutectic solvents, *J. Mol. Liq.* 321 (2021) 114744.
- [33] T. Lemaoui, N.E.H. Hammoudi, I.M. Alnashef, M. Balsamo, A. Erto, B. Ernst, Y. Benguerba, Quantitative structure properties relationship for deep eutectic solvents using S_p-profile as molecular descriptors, *J. Mol. Liq.* 309 (2020) 113165.
- [34] Y. Benguerba, I.M. Alnashef, A. Erto, M. Balsamo, B. Ernst, A quantitative prediction of the viscosity of amine based DESs using S_p-profile molecular descriptors, *J. Mol. Struct.* 1184 (2019) 357–363.
- [35] X. Kang, C. Liu, S. Zeng, Z. Zhao, J. Qian, Y. Zhao, Prediction of Henry's law constant of CO₂ in ionic liquids based on SEP and S_p-profile molecular descriptors, *J. Mol. Liq.* 262 (2018) 139–147.
- [36] Y. Zhao, J. Gao, Y. Huang, R.M. Afzal, X. Zhang, S. Zhang, Predicting H₂S solubility in ionic liquids by the quantitative structure-property relationship method using S_p-profile molecular descriptors, *RSC Adv.* 6 (2016) 70405–70413.
- [37] D. Peng, F. Picchioni, Prediction of toxicity of ionic liquids based on GC-COSMO method, *J. Hazard Mater.* 398 (2020) 122964.
- [38] X. Kang, Z. Chen, Y. Zhao, Assessing the ecotoxicity of ionic liquids on *Vibrio fischeri* using electrostatic potential descriptors, *J. Hazard Mater.* 397 (2020) 122761.
- [39] T. Lemaoui, A.S. Darwish, N.E.H. Hammoudi, F. Abu Hatab, A. Attoui, I.M. Alnashef, Y. Benguerba, Prediction of electrical conductivity of deep eutectic solvents using COSMO-RS sigma profiles as molecular descriptors: a quantitative structure-property relationship study, *Ind. Eng. Chem. Res.* 59 (2020) 13343–13354.
- [40] Z. Song, H. Shi, X. Zhang, T. Zhou, Prediction of CO₂ solubility in ionic liquids using machine learning methods, *Chem. Eng. Sci.* 223 (2020) 115752.
- [41] N.A. Kovdienko, P.G. Polishchuk, E.N. Muratov, A.G. Artemenko, V.E. Kuz'min, L. Gorb, F. Hill, J. Leszczynski, Application of random forest and multiple linear regression techniques to QSPR prediction of an aqueous solubility for military compounds, *Mol. Inform.* 29 (2010) 394–406.
- [42] G. Chen, Z. Song, Z. Qi, K. Sundmacher, Neural recommender system for the activity coefficient prediction and UNIFAC model extension of ionic liquid-solute systems, *AIChE J.* 67 (2021) e17171.
- [43] P. Chaudhari, N. Ade, L.M. Pérez, S. Kolis, C.V. Mashuga, Quantitative Structure-Property Relationship (QSPR) models for Minimum Ignition Energy (MIE) prediction of combustible dusts using machine learning, *Powder Technol.* 372 (2020) 227–234.
- [44] S. Chinta, R. Rengaswamy, Machine learning derived quantitative structure property relationship (QSPR) to predict drug solubility in binary solvent systems, *Ind. Eng. Chem. Res.* 58 (2019) 3082–3092.
- [45] N. Meftahi, M.L. Walker, B.J. Smith, Predicting aqueous solubility by QSPR modeling, *J. Mol. Graph. Model.* 106 (2021) 107901.
- [46] D.A. Saldana, L. Starck, P. Mouglin, B. Rousseau, B. Creton, Prediction of flash points for fuel mixtures using machine learning and a novel equation, *Energy Fuels* 27 (2013) 3811–3820.
- [47] S. Yuan, Z. Jiao, N. Quddus, J.S. Kwon II, C.V. Mashuga, Developing quantitative structure-property relationship models to predict the upper flammability limit using machine learning, *Ind. Eng. Chem. Res.* 58 (2019) 3531–3537.
- [48] S.P. David, M.O. Noel, C.G. Robert, B.O.M. John, Random forest models to predict aqueous solubility, *J. Chem. Inf. Model.* 47 (2007) 150–158.
- [49] Y. Mishina, R. Murata, Y. Yamauchi, T. Yamashita, H. Fujiyoshi, Boosted random forest, *IEICE Trans. Info Syst.* E98.D (2014) 1630–1636.
- [50] H. Saghaei, M.M. Ghiasi, A.H. Mohammadi, Analyzing the experimental data of CO₂ equilibrium absorption in the aqueous solution of DEA plus MDEA with random forest and leverage method, *Int. J. Greenh. Gas Con.* 63 (2017) 329–337.
- [51] V. Venkatraman, B.K. Alsberg, Predicting CO₂ capture of ionic liquids using machine learning, *J. CO₂ Util.* 21 (2017) 162–168.
- [52] M. Aghaie, S. Zendejboudi, Estimation of CO₂ solubility in ionic liquids using connectionist tools based on thermodynamic and structural characteristics, *Fuel* 279 (2020) 117984.
- [53] T. Wu, W.L. Li, M.Y. Chen, Y.M. Zhou, Q.Y. Zhang, Prediction of Henry's law constants of CO₂ in imidazole ionic liquids using machine learning methods based on empirical descriptors, *Chem. Pap.* 75 (2020) 1619–1628.
- [54] F. Yusuf, T. Olayiwola, C. Afagwu, Application of Artificial Intelligence-based predictive methods in ionic liquid studies: a review, *Fluid Phase Equilib.* 531 (2021) 112898.
- [55] L. Breiman, Random forests, *Mach. Learn.* 45 (2001) 5–32.
- [56] S. Sarmad, Y. Xie, J.P. Mikkola, X. Ji, Screening of deep eutectic solvents (DESs) as green CO₂ sorbents: from solubility to viscosity, *New J. Chem.* 41 (2017) 290–301.
- [57] S. Zeng, X. Zhang, L. Bai, X. Zhang, H. Wang, J. Wang, D. Bao, M. Li, X. Liu, S. Zhang, Ionic-liquid-based CO₂ capture systems: structure, interaction and process, *Chem. Rev.* 117 (2017) 9625–9673.
- [58] H. Ghaedi, M. Ayoub, S. Sufian, A.M. Shariff, S.M. Hailegiorgis, S.N. Khan, CO₂ capture with the help of phosphonium-based deep eutectic solvents, *J. Mol. Liq.* 243 (2017) 564–571.