International Journal of Refrigeration 158 (2024) 345–352

Contents lists available at ScienceDirect

International Journal of Refrigeration

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

Mass flow characteristics prediction of refrigerants through electronic expansion valve based on XGBoost

Prédiction des caractéristiques de débit massique des frigorigènes à travers un détendeur électronique basé sur l'algorithme XGBoost

Anping Wan [a], Zhipeng Gong [a,b], Ting Chen [a], Khalil AL-Bukhaiti [a,c,*]

[a] *Department of Mechatronic Engineering, Hangzhou City University, Hangzhou, 310015, China*
[b] *Department of Mechanical Engineering, Zhejiang University of Technology, Hangzhou, 310014, China*
[c] *School of Civil Engineering, Southwest Jiaotong University, Chengdu, 610031, China*

ARTICLE INFO

Keywords:
Electronic expansion valve
Mass flow rate
XGBoost algorithm
Prediction accuracy
Refrigeration systems

Mots-clés:
Détendeur électronique
Débit massique
Algorithme XGBoost
Précision de la prédiction
Systèmes frigorifiques

ABSTRACT

In an effort to enhance the prediction accuracy of the mass flow characteristics of refrigerants through electronic expansion valves (EEVs), this study develops a mass flow model using the XGBoost machine learning algorithm. Utilizing experimental data from open literature for refrigerants R1233zd(E) and R245fa, the model aims to accurately predict the mass flow coefficient in EEVs, a crucial aspect for improving the performance and efficiency of refrigeration systems. The model's performance is evaluated using the coefficient of determination (R2) on the training and test datasets, revealing a minimal performance gap and no overfitting issue. Remarkably, our model's predictions for R1233zd(E) and R245fa datasets align consistently with experimental data, with 96.4 % and 90.6 % of the predicted data deviating from the experimental ones within ±5 %, respectively. Further, the root-mean-square error (RMSE) and coefficient of variation of the root mean square error (CV-RMSE) values are quite low, with 2.2 % and 2.33 % for R1233zd(E), and 4.23 % and 4.47 % for R245fa, indicating high prediction accuracy. Compared to the original models, the paper approach with the XGBoost algorithm significantly improves prediction accuracy. This advancement provides a promising direction for developing more efficient and reliable control strategies for refrigeration systems in various applications.

1. Introduction

In the ever-evolving technological landscape, refrigeration systems play a pivotal role across various industries and applications. These systems are essential for preserving perishable goods, maintaining comfortable indoor environments, facilitating specific processes in industries, and playing an increasingly important role in electric vehicle operation. Throttling devices are central to the functioning of these systems, which regulate the flow and pressure of refrigerants, thereby controlling the cooling process (Rongrong et al., 2018). Among the assortment of throttling devices, electronic expansion valves (EEVs) signify the latest generation. These advanced components represent a significant leap in technological innovation over traditional counterparts such as capillary tubes, short tubes, and thermal expansion valves. The advent of EEVs has heralded a transformative phase in refrigeration technology characterized by superior flexibility in the modulation of mass flow. Such flexibility facilitates precise control over the cooling process, leading to substantial improvements in the efficiency and performance of refrigeration systems.

In addition to their flexibility, EEVs are recognized for their rapid response times. This characteristic empowers these devices to swiftly adapt to shifting operational conditions, ensuring consistent performance under diverse scenarios. Further, the inherent versatility of EEVs allows them to be tailored to a broad spectrum of operating conditions, making them an adaptable choice for various applications. The advanced capabilities of EEVs have resulted in their widespread deployment in many heat pump and refrigeration applications. These encompass inverter heat pumps, electric vehicle air conditioning systems, battery cooling systems, and cold storage units, to name a few. The application of EEVs in such diverse technologies underscores their

* Corresponding author.
E-mail address: eng.khalil670@hotmail.com (K. AL-Bukhaiti).

https://doi.org/10.1016/j.ijrefrig.2023.12.011
Received 6 September 2023; Received in revised form 8 November 2023; Accepted 9 December 2023
Available online 12 December 2023
0140-7007/

significance in the contemporary landscape of refrigeration technology while simultaneously alluding to their potential for future advancements in this domain. Despite the numerous advantages of EEVs and their growing prevalence, harnessing their full potential presents several challenges. A primary challenge lies in the complex internal flow within the valve, which poses a significant barrier to the precise prediction of the mass flow characteristics of refrigerants. Achieving an in-depth understanding of these characteristics is vital for optimizing system performance. This knowledge is crucial for accurately synchronizing with the variable heat load and achieving real-time mass flow regulation. These factors contribute significantly towards enabling the refrigeration system to operate economically, efficiently, and reliably. Given the critical nature of understanding and predicting the mass flow characteristics of refrigerants in electronic expansion valves, this area has been the focus of substantial research. Therefore, domestic and international research on control strategies for electronic expansion valves is extensive (Yujia et al., 2016; Jianhua et al., 2020; Ekren et al., 2012; Fei et al., 2021; Shuai et al., 2016), and research on the mass flow characteristics of electronic expansion valves is actively being conducted. Park et al. (2007) developed an empirical relational equation for mass flow prediction of electronic expansion valves based on experimental data by modifying the orifice plate equation. Yu et al. (2014) gave a correlation equation for the mass flow of supercritical CO_2 through EEV based on Bernoulli's equation with expansion factor, in which the effect of expansion coefficient on mass flow was considered. Zhen et al. (2015) developed two models for predicting the mass flow rate of EEVs based on the Buckingham theorem and artificial neural network model. Saleh and Aly (2016) used the artificial intelligence network approach to develop a prediction model for each data set and a general model for the three data sets based on the data of three refrigerants. Xiang et al. (2016) used refrigerant inlet and outlet pressure, inlet subcooling, and EEV opening as the input of the neural network and refrigerant mass flow rate as the output of the neural network, an artificial neural network model for the mass flow rate of electronic expansion valve was developed. Ting et al. (2017); Ting et al. (2018, 2019) developed a power-law correlation model polynomial fitting relational and artificial intelligence network models to predict the mass flow rate of electronic expansion valves based on experimental data for two refrigerants, respectively. These studies have shed invaluable light on the dynamics of refrigerant flow and paved the way for developing increasingly accurate prediction models. However, despite these extensive research efforts, gaps remain in our understanding and predictive ability. Many existing models rely on traditional methods such as polynomial fitting and artificial neural networks. While these models have provided useful insights, they often fail in prediction accuracy. There is, therefore, a pressing need for the development of more accurate prediction models.

This study addresses the need for improved prediction models for the mass flow characteristics of refrigerants in electronic expansion valves in refrigeration systems. It does this by leveraging the XGBoost machine learning algorithm for handling complex non-linear relationships to develop a robust prediction model. This new model aims to outperform existing methodologies primarily based on polynomial fitting and artificial neural networks, facilitating more economical and reliable operation control strategies in heat pump and cooling systems. This study pushes the boundaries of current knowledge in refrigeration systems by combining in-depth theoretical analysis with practical experimentation. Overall, the accurate prediction of mass flow characteristics is critical for enhancing the efficiency and reliability of refrigeration systems, and this research, through advanced machine learning techniques like XGBoost, aims to contribute significantly to this field.

2. Analysis of mass flow characteristics in electronic expansion valves

The intricate structure of an electronic expansion valve, depicted in Fig. 1, enables modulation of the flow area through repositioning the needle valve body, actuated by a stepper motor. This mechanism is critical during the throttling process, wherein high-temperature, high-pressure refrigerant experiences a drastic reduction in flow area in its saturated liquid state. Consequently, the refrigerant transforms into a low-pressure, low-temperature liquid, generating a small quantity of flash gas. This phase transition facilitates heat exchange between the evaporator and the surrounding medium.

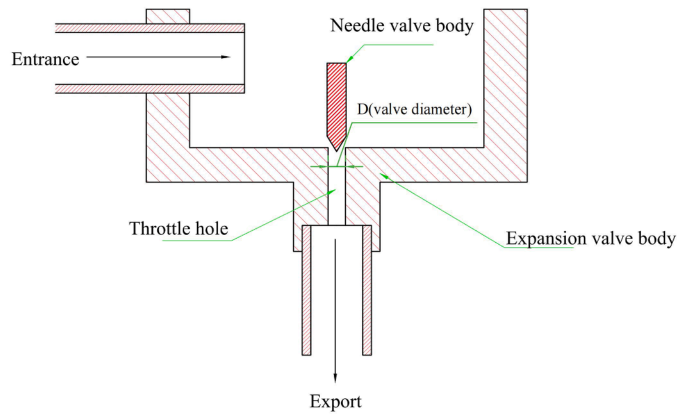

Fig. 1. Electronic expansion valve structure.

The mass flow coefficient is an important parameter of the electronic expansion valve, and the incompressible single-phase fluid equation derived from Bernoulli's equation is commonly used to calculate the mass flow of refrigerant through the electronic expansion valve (Saleh and Aly, 2016; Xiang et al., 2016; Zhifang et al., 2008). The equation is shown in Eq. (1).

$$\dot{m} = C_d \cdot A\sqrt{2\rho(P_{in} - P_{out})} \tag{1}$$

where C_d is the mass flow coefficient, ρ is the mixed density of the two-phase refrigerant as it exits the electronic expansion valve, accounting for both the liquid and gas phases present after throttling, P_{in} and P_{out} are the pressure of the refrigerant at the inlet and outlet of the electronic expansion valve, respectively. A is the flow area of the expansion valve port, which can be represented as:

$$A = \frac{1}{4}\pi(z \cdot D)^2 \tag{2}$$

where z is the valve opening, and D is the valve diameter. The C_d an incompressible single-phase fluid constant and the mass flow coefficient of an electronic expansion valve is usually measured by the manufacturer based on a standard of pure nitrogen. Still, the mass flow coefficient of a two-phase flow through an electronic expansion valve must be determined experimentally (Jinghui et al., 2008).

The mass flow characteristics of the electronic expansion valve are central to the functionality of the refrigeration system. Correct prediction of these characteristics can significantly improve the efficiency and reliability of the system. However, the complex internal flow within the electronic expansion valve presents a challenge to accurate prediction. This is where the XGBoost algorithm comes into play. XGBoost is a powerful machine learning algorithm capable of handling complex non-linear relationships, making it an ideal tool for this problem. By training the XGBoost model on experimental data, we can effectively capture the intricate dynamics of the refrigerant flow within the electronic expansion valve. This allows us to generate accurate predictions of the mass flow coefficient under various operating conditions, thereby addressing the key challenge in optimizing the performance of refrigeration systems. The accurate prediction of refrigerant mass flow through the electronic expansion valve is crucial for optimizing the performance and efficiency of refrigeration systems.

Controlling the mass flow rate precisely synchronizes the system with the changing heat load, allowing for real-time regulation. Precise mass flow prediction enables design and control strategies to maximize cooling capacity while minimizing power consumption. Small errors in the mass flow coefficient in control algorithms can compromise temperature regulation, lowering system effectiveness and reliability over time. Furthermore, accurately sizing expansion valves requires understanding how refrigerant properties and operating conditions affect mass flow. Correct sizing is important to avoid restrictiveness or oversizing that wastes energy. Lastly, mass flow prediction facilitates fault detection, as deviations from expected rates could indicate issues developing. Early detection permits timely maintenance before larger problems emerge. In summary, precisely determining the refrigerant's mass flow characteristics through the electronic expansion valve, as enabled by this study's model, is pivotal for designing efficient and dependable modern refrigeration systems.

3. XGBoost model

The XGBoost algorithm, for eXtreme Gradient Boosting, is an advanced implementation of the gradient boosting machine learning algorithm. It is designed to be highly efficient, flexible, and portable, making it a powerful tool for various predictive tasks. XGBoost is an open-source machine learning project developed by Tianqi Chen in 2016 and is considered one of the best algorithms for supervised learning for classification and regression problems. The main idea of the boosting algorithm is to keep adding trees and performing feature splits to grow a tree. At each iteration, a tree is added to the existing trees to fit the previous trees' residual between the predicted result and the true value. Currently, the XGBoost algorithm has been applied in several fields, such as disease prediction, urban fire incident prediction, runoff prediction, and heat load prediction (Xuehuai et al., 2017; Bingyue, 2018; Mingqi et al., 2020; Puning et al., 2019; Wei et al., 2019; Lingling et al., 2020).

In the context of this study, predicting the mass flow characteristics of refrigerants in electronic expansion valves is a complex task due to the non-linear relationships between various factors influencing the flow. Machine learning algorithms like linear regression or logistic regression may not effectively capture these non-linear relationships, which can lead to sub-optimal prediction performance. The XGBoost algorithm, on the other hand, is renowned for its ability to handle complex non-linear relationships. It achieves this by combining weak prediction models into a strong predictive model in an iterative manner, allowing it to learn from the residuals (errors) of the previous models. This approach enables the XGBoost algorithm to capture the intricate dynamics of refrigerant flow within the electronic expansion valve, thereby enhancing the accuracy of the predictions.

Another key advantage of XGBoost is its efficiency and speed. The algorithm has been engineered to maximize computational speed and performance, which is crucial for handling large datasets and performing multiple iterations. Given that the model in this study is trained on a large dataset of experimental data, this computational efficiency makes XGBoost an ideal choice. Finally, XGBoost offers a high degree of flexibility, with various customizable parameters that allow fine-tuning of the model based on the specific problem at hand. This flexibility enables us to optimize the model for predicting the mass flow characteristics of refrigerants in electronic expansion valves. The choice of XGBoost over other machine learning algorithms for this study is based on its ability to handle complex non-linear relationships, its computational efficiency and speed, and its high degree of flexibility. These attributes make it a powerful tool for predicting the mass flow characteristics of refrigerants in electronic expansion valves.

3.1. Principle of XGBoost

XGBoost is a tree integration model (shown in Fig. 2) that uses the sum of the predicted values of a sample from each of K (the total number of trees is K) trees as the prediction of that sample in the XGBoost system.

Given a dataset with n samples and m features:

$$D = (x_i, y_i)(|D| = n, x_i \in R^m, y_i \in R) \tag{3}$$

Define the function as follows:

$$\widehat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i) \tag{4}$$

$\widehat{y}_i$ is the predicted value and f_k is the weight function of the k_th tree. XGBoost is used for supervised learning problems, and usually, the objective function contains two components:

$$Obj(\theta) = L(\theta) + \Omega(\theta) \tag{5}$$

where $L(\theta)$ is the loss function that measures the error between the predicted and measured values:

$$L(\theta) = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 \tag{6}$$

$\Omega(\theta)$ is used to penalize the complexity of the model and to smooth the weights among the leaf nodes, thus avoiding overfitting:

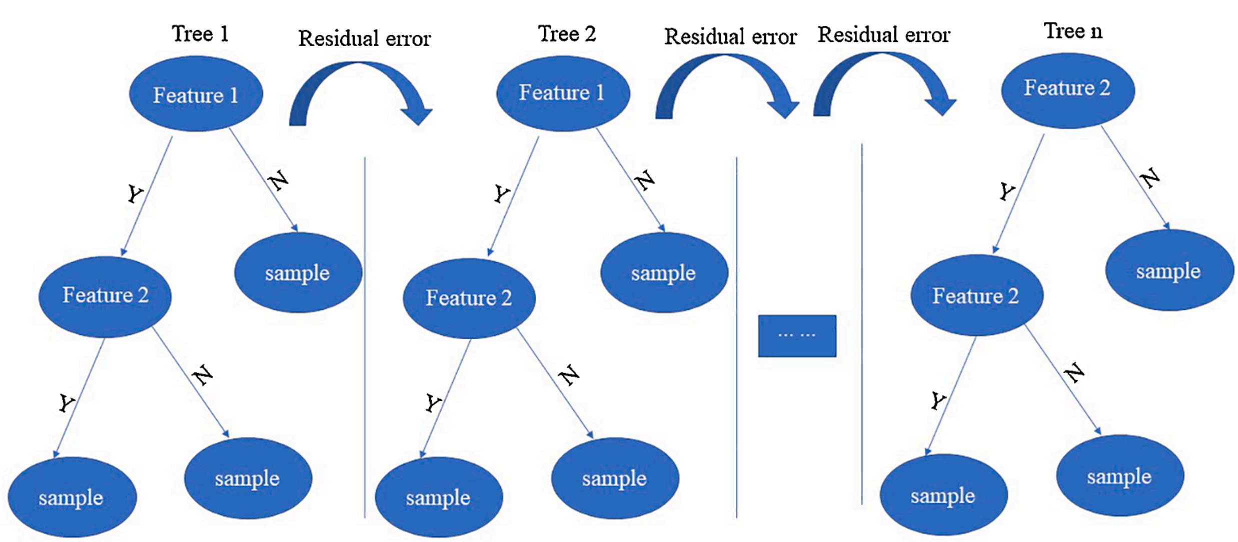

Fig. 2. XGBoost integrated tree model.

$$\Omega(\theta) = \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^{\,2} \tag{7}$$

T is the number of leaf nodes, the second part is the regularization term, w is the score of each leaf node, γ, and λ are the control factors used to avoid overfitting. The objective function can be written as:

$$Obj^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \sum_{i=1}^{t}\Omega(f_i) \tag{8}$$

$$l\left(y_i, \hat{y}_i^{(t)}\right) = l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) \tag{9}$$

When creating a new tree to fit the residuals left over from the previous iteration:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \tag{10}$$

where $\hat{y}_i^{(t)}$ is the predicted value of sample i at the t_th iteration, and $f_t(x_i)$ is the correlation function of the t_th tree. A second-order Taylor expansion of Eq. (8) yields:

$$Obj^t \sum_{i=1}^{n}\left[\left(y_i,\ \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)\right] + \Omega(f_t) + \sum_{i=1}^{t-1}\Omega(f_i) \tag{11}$$

For the t_th iteration, both $\sum_{i=1}^{t-1}\Omega(f_i)$ and $\sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)})$ are constants, so Eq. (11) can again be written as:

$$Obj^{(t)} = \sum_{i=1}^{n}\left[g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)\right] + \Omega(f_t) + \text{constant} \tag{12}$$

where: $g_i = \delta_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$ and $h_i = \partial^2_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$. Substituting Eq. (7) into Eq. (12), since each instance will end up in a leaf node, the traversal of the sample can be changed to a traversal for the leaf nodes, and the same terms can be combined as follows:

$$\mathrm{Obj}^{(t)} = \sum_{j=1}^{T}\left[\left(\Sigma_{i\in I_j} g_i\right) w_j + \frac{1}{2}\left(\Sigma_{i\in I_j} h_i + \lambda\right) w_j^2\right] + \gamma \mathrm{T} + \text{constant} \tag{13}$$

Set $G_j = \Sigma_{i\in I_j} g_i$ and $H_j = \Sigma_{i\in I_j} h_i$

$$\mathrm{Obj}^{(t)} = \sum_{j=1}^{\mathrm{T}}\left[\mathrm{G_j w_j} + \frac{1}{2}\left(\mathrm{H_j} + \lambda\right)\mathrm{w_j^2}\right] + \gamma \mathrm{T} + \mathit{constant} \tag{14}$$

Find the partial derivative of w_j, and make the partial derivative equal to 0. The optimal w_j and the objective functions are as follows:

$$w_1^* = -\frac{G_j}{H_j + \lambda}$$

$$Obj = -\frac{1}{2}\sum_{J=1}^{T}\frac{G_j^2}{H_j + \lambda} + \gamma T \tag{15}$$

Once the simplified formula is obtained, the next step is calculating the gain from the selected features to select the appropriate splitting features.

$$Gain = Obj_{(L+R)} - (Obj_L + Obj_R) \tag{16}$$

$$Gain = \frac{1}{2}\left[\frac{G_L{}^2}{H_L + \lambda} + \frac{G_R{}^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda}\right] - \gamma \tag{17}$$

Gain is the gain function, $Obj_{(L+R)}$ is the value before node splitting, and $(Obj_L + Obj_R)$ is the value after the node splits into left and right nodes. If the gain is positive, and the larger the value, the more worthy the cif ut; gain is negative, the less worthy the cut. XGboost uses the same method as the CART regression tree, using a greedy algorithm, traversing all feature division points for all features.

3.2. Build the XGBoost model

The key thermophysical properties of the two refrigerants under investigation are detailed in Table 1. This table provides the foundational data essential for further analysis of the flow characteristics exhibited by the refrigerants as they traverse the electronic expansion valve. The sources of data employed for this investigative analysis are outlined in Table 2.

The data sets utilized in this study are bifurcated into two groups. Group A comprises 149 data points, representing the experimental results from tests conducted on refrigerant R245fa passing through electronic expansion valves of four distinct valve diameters. In contrast, Group B consists of 112 data points, which denote the experimental results derived from testing refrigerant R1233zd (E) as it flowed through electronic expansion valves of three different valve diameters.

The terminology used for the inlet and outlet parameters in Table 2 and how it relates to the subcooling will be classified as follows:

- The "Entrance pressure" and "Entrance temperature" in Table 2 refer to the saturated refrigerant conditions at the inlet of the EEV.
- The "inlet pressure" used as an input parameter for the XGBoost model is the same as the "Entrance pressure" since it represents the saturated vapor pressure at the inlet.
- The subcooling directly affects the actual "inlet temperature" of the refrigerant entering the EEV by lowering it below the saturation temperature corresponding to the inlet pressure.
- The "Export pressure" and "Export temperature" refer to the outlet conditions from the EEV after expansion.
- The "outlet pressure" mentioned as a model input parameter is the same as the "Export pressure."

By clarifying these relationships, the role of subcooling is to reduce the refrigerant temperature entering the EEV below the saturation point associated with the inlet pressure. This subcooling effect is important to capture in the model as it influences the mass flow rate. It is important to note that the experimental conditions under which these tests were conducted for both refrigerants are also elaborated upon in Table 2. This comprehensive collation of data, encompassing varying valve diameters and experimental conditions, forms the bedrock for the ensuing analysis. By leveraging this data, the study aims to effectively train the XGBoost model to predict the mass flow characteristics of these refrigerants as they interact with the electronic expansion valve. Firstly, the experimental data were simply processed, and the mass flow coefficient of refrigerant flow through the electronic expansion valve was mainly influenced by inlet conditions, outlet conditions, and flow area as described in the literature (Shanwei et al., 2005; Qifang et al., 2007; Wenhua, 2013; Cichong et al., 2017). Therefore, in this study, the valve diameter, inlet pressure, outlet pressure, inlet temperature, subcooling degree, and valve opening in the data were set as the characteristic values, the mass flow coefficient was set as the target value, and the collated data were imported into python using the Pandas library in python. After importing the data, the original dataset is first reshuffled and randomly disrupted to improve the training effect of the model, after which the dataset is divided, and 70 % of the data is used as the

Table 1
Refrigerant parameters.

Parameters	R245fa	R1233zd(E)
Global Warming Potential	858	1
Ozone Depletion Potential	0	0
Flammability	None	None
Toxicity	Toxic	Low
Atmospheric life/year	7.6	0.07
Critical pressure/kPa	3651	3573
Critical temperature/ °C	154	165.6
Boiling point/ °C	15.1	18.3

Table 2
Data sources and operational conditions.

Data sources	Ting et al. (2018) Dataset A	Ting et al. (2019) Dataset B
Refrigerant type	R245fa	R1233zd(E)
Valve opening z/%	0 ~ 100	0 ~ 100
Valve diameter D/mm	1,1.4,1.8,2.5	1,1.4,1.8
Subcooling T_{sub}/ °C	5,10,15	5,10,15
Entrance temperature T_{in}/ °C	60,70,80	60,70,80
Entrance pressure P_{in}/kPa	463,610,789	391,511,658
Export temperature T_{out}/ °C	35	35
Export pressure P_{out}/ kPa	210	180

training set, and 30 % of the data is used as the test set. The training set is used to train the XGboost model, while the test set is used to evaluate the prediction accuracy of the established model. Because the magnitudes of the feature values are not the same, this can impact the results of the data analysis. To eliminate the influence of dimensionality among the indicators, this paper first uses Buckingham's π theorem to dimensional data process. Then, it uses the Z-score normalization method to normalize the data so that the processed data conforms to the standard normal distribution. According to the principle of the XGBoost model, some hyperparameters will greatly impact the performance of XGBoost. In this paper, choose the max_depth, learning_rate, and n_estimators as the adjustment parameters. In this paper, choose the max_depth, learning_rate, and n_estimators as the adjustment parameters because they have a significant impact on XGBoost model performance and tend to be the most influential where max_depth controls the number of layers (nodes) in each decision tree. Limiting depth prevents overfitting. learning_rate slows the rate at which the model learns, balancing underfitting and overfitting. Lower is safer but slower to converge. n_estimators set the number of boosted trees, a proxy for model complexity. More trees generally improve performance but risk overfitting. These hyperparameters needed tuning to find an optimal balance between bias and variance for the datasets - too low increases underfitting and too high risks overfitting. Tuning these key learning process hyperparameters was important to configure the XGBoost model for the best predictive accuracy given the specific task and datasets. GridSearchCV in sklearn is chosen as the tuning method, which can automatically adjust the parameters and give the optimal result parameters as long as the tuning range of the parameters is entered. This method is suitable for small data sets (Ranjan et al., 2019). The parameter adjustment range is set according to previous experience, and the tuning parameter range and results are shown in Tables 3 and 4.

After constructing the model, the predicted mass flow coefficients are compared with the actual mass flow coefficients in combination with statistical metrics (including RMSE, MAE, CV-RMSE, and R2 (defined below)) to assess the accuracy of the established XGBoost model. These four scores are the basic evaluation metrics for the regression task. The root mean square error (RMSE) represents the error between the predicted and actual values and is defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{y}_i - y_i\right)^2} \tag{18}$$

The mean absolute error (MAE) is the average of the absolute errors between the predicted and actual values and is determined by the following equation:

$$MAE = \frac{1}{n}\Sigma_{i=1}^{n} \parallel \widehat{y}_i - y_i \parallel \tag{19}$$

The CV-RMSE (coefficient of variation of the root mean square error) is a normalized measure of the model's prediction error, which allows comparisons across datasets with different units or variability. It is calculated as:

$$\mathrm{CV} - \mathrm{RMSE} = \mathrm{RMSE/mean(actual)} * 100\ \% \tag{20}$$

The coefficient of determination (R2) measures the bias in a data set and indicates the degree of goodness of fit of the regression.

$$R2 = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{y}_i - y_i\right)^2}{\frac{1}{n}\sum_{i=1}^{n}\left(\overline{y}_i - y_i\right)^2} \tag{21}$$

where y_i is the actual value of the mass flow coefficient, $\overline{y}_i$ is the average of the actual values and $\widehat{y}_i$ is the predicted value. The closer the R2 value is to 1, the smaller the error between the predicted and true values in the sample, indicating that the independent variables explain the dependent variable better in the regression analysis. Once the optimal combination of hyperparameters is determined, the best-estimated model is obtained, and then it is evaluated with the training and test sets, respectively. Taking the R1233zd (E) refrigerant data as an example, Fig. 3 shows the trend of R2 for the training and test data as the number of weak learners increases. The training curve starts with a large rise, reaches its maximum value of almost one after the cycle, and remains stable for the rest of the cycle time.

The test curve increases sharply to 0.52 in the first 50 cycles, after which the growth slope decreases steadily, and finally, the value of R2 stabilizes at about 0.932. According to the learning curve in Fig. 4, the proposed XGboost model converges rapidly in the first 100 iterations, reaches a minimum value of around 260 iterations, and remains constant. This coincides with the previous tuning results. Based on the above analysis, the proposed model in this paper proves feasible and reliable because of satisfactory evaluation scores in training and testing and because the performance gap between the training and testing sets is very small, indicating that no overfitting problem occurs.

Table 3
R1233zd(E) hyperparameter range of model adjustment.

Parameters	Adjusting range	Optimal parameters
max_depth	6~11	9
learning_rate	0.05~0.09	0.06
n_estimators	0~360	259

Table 4
R245fa hyperparameter range of model adjustment.

Parameters	Adjusting range	Optimal parameters
max_depth	6~11	8
learning_rate	0.05~0.09	0.06
n_estimators	0~360	69

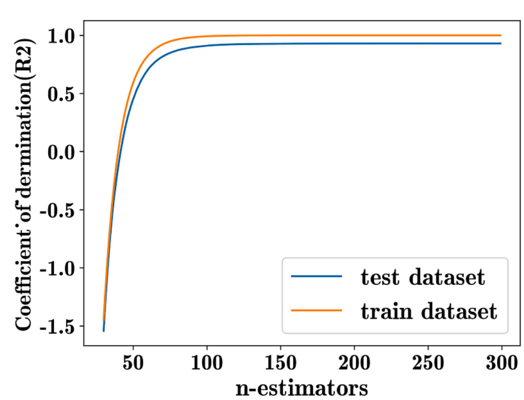

Fig. 3. R2 value change curve.

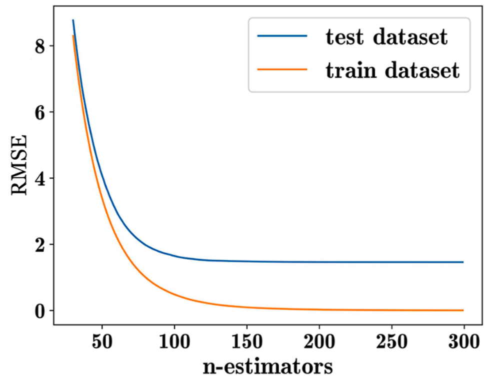

Fig. 4. RMSE change curve.

4. Analysis of the results

For the time-consuming and computational resources when the XGBoost models are performed in this study. Training time on a laptop (2.2 GHz Intel Core i7, 16GB RAM) was approximately 30 min for each refrigerant dataset. Prediction times were approximately 1 s for each test sample, suitable for near real-time applications. Polynomial fitting and ANN models may require fewer resources for simpler model fitting than iterative tree boosting in XGBoost. However, XGBoost's optimization improves computational efficiency compared to traditional gradient-boosting methods like GBRT. Even with higher resource requirements, XGBoost achieved the best predictive accuracy of any model evaluated over both refrigerants (up to 100 % within 10 % error vs 91–95.6 % for other models). While a direct runtime comparison is impossible, XGBoost demonstrated a significant accuracy improvement, warranting its greater computational costs compared to simpler approaches. After determining the final XGBoost model, the data for each of the two refrigerants were put into their respective prediction models, and the resulting predicted values of mass flow coefficients were transformed into mass flow rates $\dot{m}$ by Eq. (1). The relative error and root mean square error are applied to analyze the fitting effect of the predicted and measured values, where the relative error and root mean square error are defined in Eqs. (22) and (23).

$$RD = \frac{\dot{m}_p - \dot{m}_e}{\dot{m}_e} \times 100\% \tag{22}$$

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} RD_i^{\,2}} \tag{23}$$

Figs. 5 and 6 compare the predicted and measured values of the mass flow of the two refrigerants through the electronic expansion valve. As shown in Fig. 5, For the data of R1233zd(E) refrigerant, the errors between the predicted results of the XGBoost model and the experimental data are within the range of ±10 %, among which the errors of 96.4 % of the data are within the range of ±5 %.

The performance of the XGBoost model was evaluated using the root mean square error (RMSE), a commonly used metric for assessing the accuracy of prediction models. The overall RMSE for this model was determined to be 2.2 %. This low percentage indicates a high prediction accuracy, underscoring the efficacy of the XGBoost-based model in predicting the mass flow characteristics of refrigerants in electronic expansion valves. When evaluating the model's performance using the R245fa refrigerant data, represented in Fig. 6, it was observed that the prediction results were marginally inferior compared to those of R1233zd(E). However, the overall fit remained satisfactory. Upon further analysis, it was found that the errors for 97.3 % of the data fell within a margin of ±10 %. Even more impressive, the errors for 90.6 % of the data were contained within a narrower margin of ±5 %. Despite the slight decrease in prediction accuracy, the overall RMSE for the R245fa refrigerant data was determined to be 4.23 %. While this percentage is slightly higher than that for the R1233zd(E) refrigerant, it still denotes a high level of predictive accuracy. This indicates that, despite minor variations, the XGBoost-based model maintains a high performance across different refrigerants and operating conditions. These findings validate the model's robustness and generalizability, two crucial attributes for an effective prediction model in variable real-world scenarios. The CV-RMSE values for the XGBoost models are R1233zd(E) refrigerant data: 2.33 % and R245fa refrigerant data: 4.47 %. Including these CV-RMSE results provides a more complete picture of the predictive capability of the developed XGBoost models. The low CV-RMSE percentages, similar to the low RMSE values already reported, demonstrate the models achieved highly accurate predictions relative to the observed data ranges for both refrigerants.

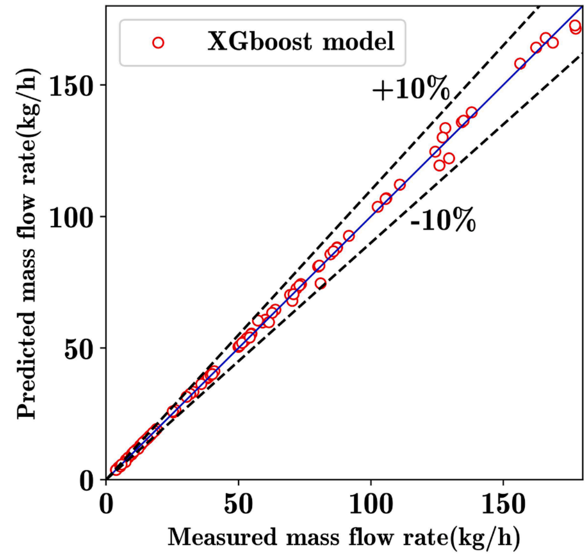

Fig. 5. R1233zd(E) comparison result of predicted value and measured value.

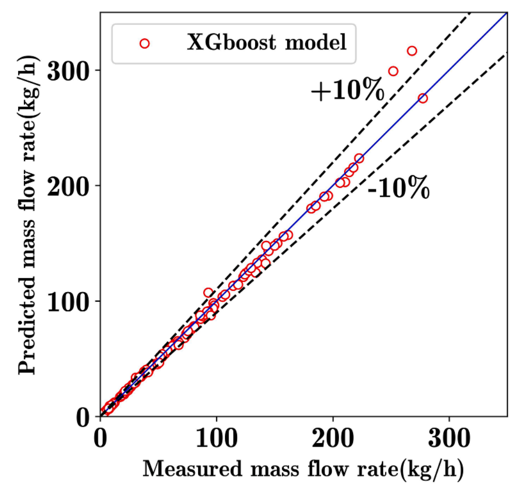

Fig. 6. R245fa comparison result of predicted value and measured value.

Fig. 7 compares the results of the R1233zd(E) refrigerant of the XGBoost model, predicting with those predicted by both power-law correlation and ANN models (Ting et al., 2019). The relative errors of the prediction results of the power-law correlation and ANN models within ±10 % as a percentage of the total data are 91 % and 95.6 %, respectively. In comparison, the prediction results of the XGBoost model can reach 100 % error control within ±10 %.

Similarly, Fig. 8 compares the relative errors of the three models

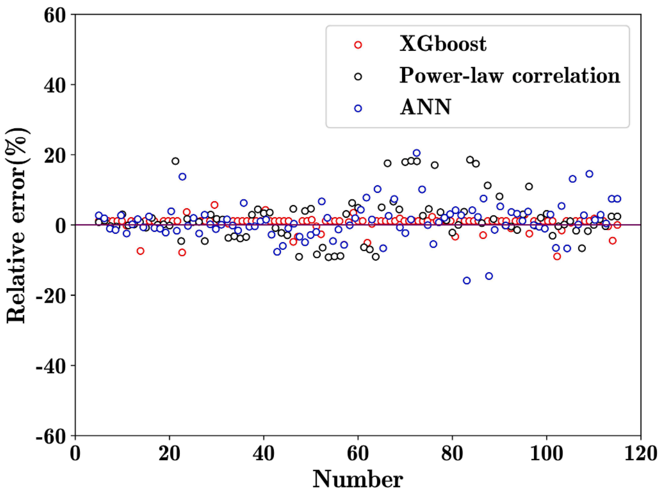

Fig. 7. R1233zd(E) comparison of the relative deviations of the three models.

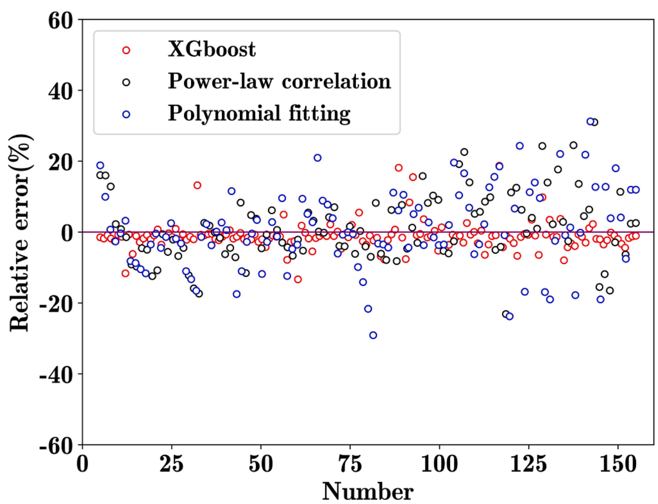

Fig. 8. R245fa comparison of the relative deviations of the three models.

regarding the refrigerant R245fa, and the results show that the errors of the XGBoost model fluctuate up and down around 0. There are data with relative errors within a percentage of 97.3 %, and very few values have errors exceeding 15 %. In the other two models (Ting et al., 2018), the best data with relative errors within 10 % for the prediction results are only 81 %. It is also clear from the figure that the fitting effect of this model is significantly better than the other two models.

Study results present a robust model using the XGBoost machine learning algorithm for predicting the mass flow characteristics of refrigerants in electronic expansion valves. The model's performance was evaluated and compared with existing models, revealing a significant improvement in prediction accuracy. The model's results align consistently with experimental data, with a high percentage of the predicted data deviating from the experimental data within a small margin. These results underscore the effectiveness of the XGBoost algorithm in handling the complex non-linear relationships inherent in the flow dynamics of the electronic expansion valve. However, despite the promising results, the model has limitations. One of the primary limitations is that the model's performance largely depends on the training data's quality and quantity. While we have used a large dataset from open literature for training the model, there may still be specific operating conditions under which the refrigerant's flow characteristics could change, which have not been represented in the training data. This limitation could affect the model's ability to make accurate predictions under these unrepresented conditions. Another challenge encountered during the study was the selection of the right hyperparameters for the XGBoost model. XGBoost has several hyperparameters that must be tuned to optimize the model's performance. While we utilized a systematic approach for hyperparameter tuning, finding the optimal set of hyperparameters was time-consuming and required significant computational resources. The main hyperparameters tuned for the XGBoost models were max_depth, learning_rate, and n_estimators. Where Max_depth controls the number of nodes in each tree, affecting the model complexity - larger trees overfit more easily. Max_depth tuned from 6 to 11. Learning_rate slows the contribution of each tree, lowering overfitting risk but slowing training. The learning_rate tuned from 0.05 to 0.09. N_estimators set the number of boosted trees, impacting performance - more trees generally improve accuracy, but overfitting is possible. n_estimators tuned from 0 to 360. GridSearchCV was used to identify the optimal values as max_depth = 8–9, learning_rate = 0.06, and n_estimators = 69–259 for R245fa and R1233zd(E). Proper tuning these hyperparameters was important to balance bias-variance to achieve strong predictive performance without overfitting. Additionally, while XGBoost has been effective in improving the prediction accuracy compared to existing models, it is a complex algorithm that can be computationally intensive, especially when dealing with large datasets. This complexity might challenge real-time applications where quick predictions are essential. In conclusion, while our model provides improved prediction accuracy for the mass flow characteristics of refrigerants in electronic expansion valves, future work should address these limitations. This could include using more comprehensive and diversified datasets for training the model and exploring more efficient techniques for hyperparameter tuning. Despite these challenges, the results of this study represent a significant step forward in predicting mass flow characteristics in electronic expansion valves.

5. Conclusion

This study has effectively developed a robust model using the XGBoost machine learning algorithm to predict the mass flow characteristics of refrigerants in electronic expansion valves. The significant improvement in prediction accuracy compared to existing models demonstrates the value of utilizing advanced machine-learning techniques in this domain. It presents an in-depth investigation into the mass flow characteristics of refrigerants R1233zd(E) and R245fa as they navigate through an electronic expansion valve. The XGBoost algorithm model, trained on open literature data, was employed to predict the refrigerant mass flow rate. The results underscore the high accuracy of the XGBoost-based model and a commendably low rate of overfitting. When juxtaposed with methodologies delineated in the literature (Ting et al., 2018, 2019), the predictive accuracy of the XGBoost-based model recorded a more significant improvement. Specifically, for refrigerant R1233zd(E), the model achieved a complete (100 %) prediction data error within a 10 % margin, with an impressive 96.4 % of data errors within a 5 % margin. The overall root means square error stood at a minimal 2.2 %. Concerning refrigerant R245fa, the model predictions demonstrated a relative error within ±10 % for 97.3 % of the total data and a relative error within ±5 % for 90.6 %. This performance noticeably surpasses the 81 % data error rate controlled within ±10 %, as reported in the literature (Ting et al., 2018). However, the model's predictive performance for refrigerant R245fa under large valve diameter conditions was less satisfactory, mirroring the results obtained in the literature (Ting et al., 2018). This limitation can be attributed to the relatively small volume of relevant data for these conditions. The XGBoost-based model has proven to be an effective tool for predicting the mass flow characteristics of refrigerants in electronic expansion valves, showing substantial improvements over existing methods. However, the model's predictive performance can be further enhanced by incorporating more comprehensive datasets that cover a broader range of experimental conditions. This study represents a significant step forward in optimizing refrigeration systems and paves the way for future advancements in this field.

Future work could focus on expanding and diversifying the training dataset, which could help improve the model's performance under a broader range of operating conditions. Additionally, more efficient techniques for hyperparameter tuning could be explored to further optimize the XGBoost model's performance. In terms of practical applications, this research has substantial implications for the design and control of refrigeration systems. The developed model could be used to design more efficient electronic expansion valves and develop more reliable control strategies, thereby improving the overall performance of refrigeration systems. Moreover, the approach used in this study could be applied to other components of refrigeration systems, potentially leading to broader improvements in system efficiency and reliability.

Data availability statement

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

Declaration of Competing Interest

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

Acknowledgment

This research is partially supported by the National Natural Science Foundation of China (52372420), and Scientific Research and Cultivation Fund Project of Hangzhou City University (204000-58162208).

References

Bingyue, Pan, 2018. Application of XGBoost algorithm in hourly PM2.5 concentration prediction. IOP Conf. Ser.: Earth Environ. Sci. 113 (1), 012127 (7pp).
Cichong, Liu, Ji, Zhang, Yongchao, Gui, et al., 2017. Mass flow characteristics and empirical modeling of R744 flow through electronic expansion device. Int. J. Refrig. 86, 82–88.
Ekren, O., Sahin, S., Isler, Y., 2012. Operation of Compressor and Electronic Expansion Valve via Different Controllers. InTech.
Fei, Wang, Wenming, Xu, Xinyi, Zhang, et al., 2021. Control method of electronic expansion valve based on PID algorithm of variable control parameters. Refrig. Air-condition. 21 (6), 95–97.
Jianhua, Wu, Jinbo, Li, Che, Wang, et al., 2020. Effects of system control strategies on startup characteristics of R290 room air conditioners. Appl. Therm. Eng., 172: 115166
Jinghui, Liu, Jiangping, Chen, Qifang, Ye, et al., 2008. A new model for depicting mass flow rate characteristic of electronic expansion valves. Exp. Therm. Fluid Sci. 32 (1), 214–219.
Lingling, Ni, Dong, Wang, Jianfeng, Wu, et al., 2020. Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. J. Hydrol. (Amst) 586, 124901.
Mingqi, Li, Xiaoyang, Fu, Dongdong, Li, 2020. Diabetes prediction based on XGBoost algorithm. IOP Conf. Ser. Mater. Sci. Eng. 768 (7), 72093.
Park, C., Cho, H., Lee, Y., et al., 2007. Mass flow characteristics and empirical modeling of R22 flowing through electronic expansion valves. Int. J. Refrig. 30 (8), 1401–1407.
Puning, Xue, Yi, Jiang, Zhigang, Zhou, et al., 2019. Multi-step ahead forecasting of heat load in district heating systems using machine learning algorithms. Energy 188 (Dec. 1): 116085.1-116085.15.
Qifang, Ye, Jiangping, Chen, Zhijiu, Chen, 2007. Experimental investigation of R407C and R410A flow through electronic expansion valve. Energy Convers. Manag. 48 (5), 1624–1630.
Ranjan, G., Verma, A.K., Radhika, S., 2019. K-Nearest Neighbors and grid search CV based real time fault monitoring system for industries. In: 2019 IEEE 5th International Conference for Convergence in Technology (I2CT). IEEE.
Rongrong, Zhang, Gen, Zhang, Edwin, S., et al., 2018. Application of electronic expansion valve in electric vehicle air-conditioning and battery cooling system. Refrig. Air-condition. 18 (005), 77–80.
Saleh, B., Aly, A.A., 2016. Artificial neural network models for depicting mass flow rate of R22, R407C, and R410A through electronic expansion valves. Int. J. Refrig. 63, 113–124.
Shanwei, Ma, Chuan, Zhang, Jiangping, Chen, et al., 2005. Experimental research on refrigerant mass flow coefficient of electronic expansion valve. Appl. Therm. Eng. 25, 2351–2366, 14-15.
Shuai, Yang, Wen, Meng, Wen, Xu, et al., 2016. Application of Fuzzy Control of Electronic Expansion Valve in Heat Pump System, 31. Automation and Instrumentation, pp. 42–45.
Ting, Chen, Bae, K.J., Kwon, O.K, 2018. Empirical correlation development of R245fa flow in electronic expansion valves. Int. J. Refrig. 88, 284–290.
Ting, Chen, Bae, K.J., Kwon, O.K, 2019. Mass flow rate prediction of R1233zd through electronic expansion valves based on ANN and Power-law correlation models. Appl. Therm. Eng. 158, 113826.
Ting, Chen, Cha, D.A., Kwon, O.K, 2017. Experimental investigation on mass flow characteristics of R245fa through electronic expansion valve. Appl. Therm. Eng. 125, 111–117.
Wei, Li, Yanbin, Yin, Xiongwen, Quan, et al., 2019. Gene expression value prediction based on XGBoost algorithm. Front. Genet. 10, 1077.
Wenhua, Li, 2013. Simplified modeling analysis of mass flow characteristics in electronic expansion valve. Appl. Therm. Eng. 53 (1), 8–12.
Xiang, Cao, Zeyu, Li, Liangliang, Shao, et al., 2016. Refrigerant flow through electronic expansion valve: experiment and neural network modeling. Appl. Therm. Eng. 92, 210–218.
Xuehuai, Shi, Qianmu, Li, Yong, Qi, et al., 2017. An accident prediction approach based on XGBoost. //. In: 2017 12th International Conference on Intelligent Systems and Knowledge Engineering. ISKE.
Yu, Hou, Changhai, Liu, Juanli, Ma, et al., 2014. Mass flowrate characteristics of supercritical CO2 flowing through an electronic expansion valve. Int. J. Refrig. 47, 134–140.
Yujia, Shang, Aiguo, Wu, Xing, Fang, et al., 2016. Dynamic simulation of electronic expansion valve-controlled refrigeration system under different heat transfer conditions. Int. J. Refrig. 72:41–52.
Zhen, Tian, Bo, Gu, Cheng, Qian, et al., 2015. Electronic expansion valve mass flow rate prediction based on dimensionless correlation and ANN model. Int. J. Refrig. 57, 1–10.
Zhifang, Xue, Lin, Shi, Hongfei, Ou, 2008. Refrigerant flow characteristics of electronic expansion valve based on thermodynamic analysis and experiment. Appl. Therm. Eng. 28 (2–3), 238–243.