



City Research Online

City St George's, University of London

Citation: Zhang, Y., Wang, Z., Kuang, H., Fu, F. & Yu, A. (2023). Prediction of Surface settlement in Shield Tunneling Construction Process using PCA-PSO-RVM Machine Learning. *Journal of Performance of Constructed Facilities*, 37(3), 04023012. doi: 10.1061/jpcfev.cfeng-4363

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/29752/>

Link to published version: <https://doi.org/10.1061/jpcfev.cfeng-4363>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

Prediction of Surface settlement in Shield Tunneling Construction Process using PCA-PSO-RVM Machine Learning

Yan Zhang¹, Zicheng Wang², Hwei Kuang³, Feng Fu⁴, CEng, F.ASCE, Aiping Yu⁵

Abstract:

Surface settlement is one of the key engineering issues during shield construction process. In order to accurately predict surface settlement, this paper proposes a new machine learning method based on Relevance Vector Machine (RVM), Principal Component Analysis (PCA) and Particle Swarm Optimization (PSO). Taking Beijing Metro Line 6 as an case study, the PCA-PSO-RVM model is used to make the prediction and compared with the prediction results of the RVM model using the same samples. In order to evaluate the reliability of the model, three evaluation indexes including mean relative error (MRE), root mean square error (RMSE) and Theil inequality coefficient (TIC) were calculated, and sensitivity analysis was carried out on them. The results show that the minimum relative error between PCA-PSO-RVM and the actual value is only 0.06%. The calculated MRE, RMSE and TIC are 0.17%, 0.0714 and 0.027% respectively, which shows that PCA-PSO-RVM model has higher prediction accuracy, smaller deviations and higher reliability compared with other three models. Through sensitivity analysis, it is found that the weighted average internal friction angle (φ) has the most significant impact on the surface settlement, which should be focused on in relevant research.

Keywords: shield tunneling; surface settlement; principal component analysis; particle swarm optimization; correlation vector machine; prediction model

1. Yan Zhang Professor, College of Civil and Architectural Engineering, Guilin University of Technology, Guilin 541004. zhangyan@glut.edu.cn;

2. Zicheng Wang, Master Degree Candidate, College of Civil and Architectural Engineering, Guilin University of Technology, Guilin 541004. 1271229848@qq.com;

28 3. Hwei Kuang Master Degree Candidate, College of Civil and Architectural Engineering, Guilin University of
29 Technology, Guilin 541004. chenxuandong@glut.edu.cn;
30 4. Feng Fu, Associate Professor, Department of Civil Engineering, School of Mathematics, Computer Science &
31 Engineering, City, University of London, Northampton Square, London, EC1V 0HB, U.K. College of Civil and
32 Architectural Engineering, Adjunct Professor Guilin University of Technology, Guilin 541004. (Corresponding
33 author), feng.fu.1@city.ac.uk
34 5 Aiping Yu ¹, Professor, College of Civil and Architectural Engineering, Guilin University of Technology, Guilin
35 541004. apyu@glut.edu.cn;

36
37

38 **1 Introduction**

39 With the development of urbanization, the urban traffic congestion increases day by day, and the
40 development and utilization of underground space has become high demand in urban traffic development
41 (Azhdar and Nazemi, 2020; Du and Zheng, 2020). The shield tunneling method is widely used in subway
42 construction due to its advantages of high precision, good safety and self-control feature. During the
43 tunneling process, the stratum suffers different degrees of displacement and settlement, which may cause
44 serious tunnel damage, ground subsidence, surface pipeline damage, and surrounding buildings damage
45 (Zhou *et al.*, 2019; Liu *et al.*, 2020). Therefore, the prediction of surface settlement during shield
46 construction has become an important research topic. The effective assessment of surface settlement
47 provides guidance for subway construction and provides basis for protection measures for surrounding
48 buildings (Singh *et al.*, 2018; Zhang *et al.*, 2021).

49 The surface settlement of shield tunneling is affected by many factors. Each factor interacts and
50 influences each other. It is difficult to accurately determine the surface settlement. Many scholars have
51 paid attention to the complexity of this problem and have carried out a series of related Research (Kasper
52 and Meschke, 2006; Chen *et al.*, 2019). The research methods of surface settlement of shield tunneling
53 construction mainly include empirical method (Sharghi *et al.*, 2017), theoretical analysis (Verruijt and
54 Booker, 1998; Chou and Bobet, 2002) and neural network prediction using machine learning (Wang *et al.*,

2013; Ocak and Seker, 2013). Peck (1969) proposed in 1969 that the morphological distribution of the settlement tank in tunnel follows normal distribution. Based on the theory of ground loss (Li *et al.*, 2021), a formula for estimating the surface settlement of a circular tunnel is proposed, which is currently the most widely used. On the basis of this formula, many theoretical studies and settlement formula calculations have been developed and widely used. However, the geological conditions during shield construction are complicated and various parameters during the excavation process will affect the surface settlement, and these parameters are uncertain, and it is difficult to use a simple formula to establish the relationship between surface settlement and influencing factors.

In recent years, with the rapid development of computer technologies, machine learning methods based on artificial intelligence stand out and are widely used in the study of nonlinear problems in various engineering fields. Salimi *et al.*, (2016) used two different artificial neural network models to predict the working efficiency of TBM tunnel construction in hard rock based on actual engineering projects. The results show that artificial neural network has good adaptability and accuracy. On this basis, the use of optimization algorithms to improve the accuracy and operating efficiency of existing neural network models (Zhang *et al.*, 2022) has gradually become one of the research hotspots. Hao *et al.*, (2015) proposed a differential evolution ant colony wavelet neural network with relative entropy as the optimization standard and verified the accuracy of the model through the measured data of surface settlement during the shield construction of Beijing Metro Line 6, and achieved good results. There are still some imperfections in the neural network method itself. When the number of training samples is too small, the prediction accuracy cannot be guaranteed. When the number of training samples is too large, it is difficult to normalize the prediction results. Therefore, seeking a more economical, accurate and efficient prediction model for the surface settlement of shield construction is essential.

Tipping M. E. (2001a; 2001b) proposed Relevance Vector Machine (Relevance Vector Machine,

78 RVM) on the basis of Support Vector Machine (Yao *et al.*, 2013; Borthakur and Dey, 2020) (Support
79 Vector Machine, SVM). RVM includes the advantages of SVM and improves the shortcomings of SVM.
80 The kernel function has a large degree of freedom. and fewer parameters, the correlation vector of the
81 model is reduced. It has the characteristics of high sparsity in probabilistic model. Therefore, improves
82 the prediction efficiency and can better handle the regression problem. However, the RVM model is
83 slightly insufficient in the screening of influencing factors and the determination of kernel function
84 parameters, which affects the diagnostic accuracy and generalization ability of RVM.

85 Therefore, this paper introduces Principal Component Analysis (PCA) (Su *et al.*, 2021; Wang *et al.*,
86 2022) and Particle Swarm Optimization (PSO) methods (Marini and Walczak, 2015; Yan *et al.*, 2022).
87 PCA filters out the principal components by dimensionality reduction, and PSO optimizes the relevant
88 parameters of the model to obtain the optimal parameters. Based on them, this paper establishes a PCA-
89 PSO-RVM for shield construction surface settlement prediction model with the actual case study of shield
90 construction of Beijing Metro Line 6, and obtains a non-mapping relationship between surface settlement
91 and principal components.

92 Under the same conditions, the RVM model, the PCA-RVM model, the PSO-RVM model and the
93 PCA-PSO-RVM model were all used to predict the settlement, and the results obtained by the various
94 models were compared for assessment of the prediction accuracy, dispersion and balance degree analysis.
95 The comparative analysis of the indicators determines the sensitivity of the influencing factors, verifies
96 the accuracy and reliability of the PCA-PSO-RVM shield construction surface settlement prediction model
97 proposed in this paper, and provides a new way to obtain the shield construction surface settlement.

98 **2 PCA-PSO-RVM model related theory**

99 ***2.1 Principle of PCA***

100 PCA is a process of computing the principal components and using them to perform a change of basis

101 on the data, sometimes using only the first few principal components and ignoring the rest. It is used in
 102 this paper to reduce the dimensions of the multiple influencing factors affecting the surface settlement of
 103 shield construction through PCA projection. Assuming that there are n samples in the data set of surface
 104 settlement, and each sample has p values of influencing factors, an $n \times p$ order matrix is constructed.

$$105 \quad \mathbf{X}_{n \times p} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (1)$$

106 Among them, x_{n1} represents the first influencing factor affecting the surface settlement, and x_{np}
 107 represents the p th influencing factor.

108 (1) In order to avoid the error caused by the different dimensions of each influencing factor, it is
 109 necessary to standardize the original data.

$$110 \quad x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sqrt{\text{Var}(x_j)}} \quad (i = 1, \dots, n; j = 1, \dots, p) \quad (2)$$

$$111 \quad \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (i = 1, \dots, n; j = 1, \dots, p) \quad (3)$$

$$112 \quad \text{Var}(x_j) = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \quad (i = 1, \dots, n; j = 1, \dots, p) \quad (4)$$

113 In the formula, \bar{x}_j and $\text{Var}(x_j)$ represent the sample mean and variance of the j th factor,
 114 respectively.

115 (2) Find the correlation coefficient matrix \mathbf{M} :

$$116 \quad \mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1p} \\ m_{21} & m_{22} & \cdots & m_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ m_{p1} & m_{p2} & \cdots & m_{pp} \end{bmatrix} \quad (5)$$

$$117 \quad m_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 (x_{kj} - \bar{x}_j)^2}} \quad (6)$$

118 In the formula, $m_{ij} = m_{ji}, m_{ii} = 1(i, j = 1, \dots, p; k = 1, \dots, n - 1)$

119 (3) Find the eigenvalues and eigenvectors of \mathbf{M} :

120 According to the characteristic equation $|\mathbf{M} - \lambda \mathbf{I}| = 0$ of \mathbf{M} , use the Jacobi method to obtain the p
121 eigenvalues $\lambda_j (j = 1, 2, \dots, p)$ of \mathbf{M} , and arrange $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ according to the size, and obtain the
122 corresponding orthogonal unitized eigenvector $\mathbf{e}_1, \mathbf{e}_2 \dots \mathbf{e}_p$.

123 (4) Determine the number of principal components q :

124 The contribution rate of variance and the cumulative contribution rate of the top q factors are
125 $\lambda_i / \sum_{i=1}^p \lambda_i$ and $\sum_{i=1}^q \lambda_i / \sum_{i=1}^p \lambda_i (i = 1, 2, \dots, p)$, respectively. The number of principal components is
126 selected according to the cumulative contribution rate. Generally, the cumulative contribution rate is
127 greater than 85%, and the corresponding first q principal components contain the information provided by
128 the p original factors.

129 (5) Find the principal components

130 The original influencing factor is x_1, x_2, \dots, x_p , and the principal component after PCA dimensionality
131 reduction is $y_1, y_2, \dots, y_q (q \leq p)$.

$$\begin{cases} y_1 = c_{11}x_1 + c_{12}x_2 + \dots + c_{1p}x_p \\ y_2 = c_{21}x_1 + c_{22}x_2 + \dots + c_{2p}x_p \\ \vdots \\ y_q = c_{q1}x_1 + c_{q2}x_2 + \dots + c_{qp}x_p \end{cases} \quad (7)$$

133 In the formula, c_{ij} and y_i are uncorrelated, and $c_{i1}^2 + c_{i2}^2 + \dots + c_{ip}^2 = 1$. y_i is the one with the largest
134 variance among all linear combinations of x_1, x_2, \dots, x_p , and y_1, y_2, \dots, y_q is uncorrelated with each
135 other, thus reducing the number of variables and achieving the effect of dimensionality reduction. The
136 principle is shown in Fig. 1.

137 2.2 Principle of RVM

138 2.2.1 Model description

139 RVM (Ma and Hanson, 2020; Galuzio *et al.*, 2020) is a sparse probability model based on Bayesian
140 principle proposed by American scholar Michael E. Tipping in 2000. As a new supervised learning method,
141 it can train the model quickly. It uses the weighted combination of kernel functions to apply to regression
142 and other problems. At the same time, machine learning based on Bayesian principle is used to ensure the
143 sparsity of the model.

144 Let the training sample data set be $\{x_n, t_n | n=1, 2, \dots, N\}$, x_n represents the input training sample
145 vector value, and t_n represents the output target value. Suppose t_n is independently distributed with
146 Gaussian white noise ξ_n , and establish a functional relationship about t_n :

$$148 \quad t_n = y(x_n; \boldsymbol{\omega}) + \xi_n \quad (8)$$

$$149 \quad y(\mathbf{x}, \boldsymbol{\omega}) = \sum_{n=1}^N \omega_n K(\mathbf{x}, x_n) + \omega_0 \quad (9)$$

150 Among them, $\boldsymbol{\omega}$ represents the weight vector, $\boldsymbol{\omega} = [\omega_0, \omega_1, \dots, \omega_N]^T$, $K(\mathbf{x}, x_n)$ represent the
151 kernel function, and ω_0 is the bias. ξ_n represents the additional Gaussian noise satisfying
152 $\xi_n \sim N(0, \sigma^2)$, and the variance σ^2 is an unknown quantity, which needs to be obtained by iteration.
153 Because the Gaussian kernel function is stable and has strong linear interpolation ability, this paper uses
154 the Gaussian kernel function.

$$155 \quad K(\|y - y_c\|) = \exp\left\{-\frac{\|y - y_c\|^2}{2W^2}\right\} \quad (10)$$

156 where y_c is the center of the kernel function, and W is the width of the Gaussian kernel.

157 Assuming that t_n are distributed independently of each other, the likelihood function of the dataset of
158 training samples can be expressed as:

159
$$p(\mathbf{t} | \boldsymbol{\omega}, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2} \|\mathbf{t} - \boldsymbol{\Phi}\boldsymbol{\omega}\|^2\right\} \quad (11)$$

160 Among them, $\mathbf{t} = (t_1, \dots, t_N)^T$ is the target vector, $\boldsymbol{\omega} = [\omega_0, \omega_1, \dots, \omega_N]^T$ is the parameter vector,

161 $\boldsymbol{\Phi}$ is the $N \times (N+1)$ matrix composed of the kernel function, and $\boldsymbol{\phi} = [\boldsymbol{\phi}(x_1) \boldsymbol{\phi}(x_2) \dots \boldsymbol{\phi}(x_N)]^T$,

162 $\boldsymbol{\phi}(x_n) = [1, K(x_n, x_1), K(x_n, x_2), \dots, K(x_n, x_N)]^T$. In order to avoid the occurrence of over-learning phenomena,

163 certain mandatory conditions can be attached to some parameters.

164 The Bayesian perspective method is applied in the correlation vector machine, and the size of each

165 weight parameter ω_n is set to zero mean in the Gaussian prior distribution, which constitutes a simple

166 function about ω , and such a function is in the zero mean Gaussian. The prior distribution is widely used.

167
$$p(\boldsymbol{\omega} | \boldsymbol{\alpha}) = \prod_{n=0}^N N(\omega_n | 0, \alpha_n^{-1}) \quad (12)$$

168 The parameters in the formula are all independently distributed, and the complexity of the prior

169 function distribution has been greatly alleviated. $\boldsymbol{\alpha}$ is the $N+1$ -dimensional hyperparameter that

170 determines the prior distribution of the weight $\boldsymbol{\omega}$, and the hyperparameter vector $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_N)$.

171 In order to obtain the final function, the scale function parameter σ^2 also needs to be introduced.

172 2.2.2 Parameter inference and regression prediction

173 According to the prior probability distribution, the posterior probability distribution of the training

174 samples can be worked out based on Bayesian theory.

175
$$P(\boldsymbol{\omega}, \boldsymbol{\alpha}, \sigma^2 | \mathbf{t}) = \frac{P(\mathbf{t} | \boldsymbol{\omega}, \boldsymbol{\alpha}, \sigma^2) P(\boldsymbol{\omega}, \boldsymbol{\alpha}, \sigma^2)}{P(\mathbf{t})} \quad (17)$$

176
$$P(\mathbf{t}) = \int P(\mathbf{t} | \boldsymbol{\omega}, \boldsymbol{\alpha}, \sigma^2) P(\boldsymbol{\omega}, \boldsymbol{\alpha}, \sigma^2) d\boldsymbol{\omega} d\boldsymbol{\alpha} d\sigma^2 \quad (18)$$

177 Since the posterior probability distribution $P(\boldsymbol{\omega}, \boldsymbol{\alpha}, \sigma^2 | \mathbf{t})$ cannot be directly calculated through

178 integration, it is decomposed into

179
$$P(\boldsymbol{\omega}, \boldsymbol{\alpha}, \sigma^2 | \mathbf{t}) = P(\boldsymbol{\omega} | \mathbf{t}, \boldsymbol{\alpha}, \sigma^2) P(\boldsymbol{\alpha}, \sigma^2 | \mathbf{t}) \quad (19)$$

180 The posterior distribution of the available weight vector $\boldsymbol{\omega}$ is:

$$P(\omega|t, \alpha, \sigma^2) = \frac{P(t|\omega, \sigma^2)P(\omega|\alpha)}{P(t|\alpha, \sigma^2)} = (2\pi)^{-(N+1)/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(\omega - \mu)^T \Sigma^{-1}(\omega - \mu)\right\} \quad (20)$$

Among them, it can be concluded that the probability distribution obeys the multivariate Gaussian model, the posterior probability distribution mean $\mu = \sigma^{-2} \Sigma \Phi^T t$, the covariance $\Sigma = (\sigma^{-2} \Phi^T \Phi + A)^{-1}$ represents the uncertainty of the model prediction, and $A = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$ represents the diagonal matrix.

In the process of estimating hyperparameters, the maximum likelihood estimation parameters α_{MP} and σ_{MP}^2 can be obtained according to parameter inference. Assuming that the sample to be tested is x^* , the predicted value t^* is distributed as follows:

$$P(t^*|t, \alpha_{MP}, \sigma_{MP}^2) = \int P(t^*|\omega, \sigma_{MP}^2) P(\omega|t, \alpha_{MP}, \sigma_{MP}^2) d\omega \quad (21)$$

$$P(t^*|t, \alpha_{MP}, \sigma_{MP}^2) = N(t^*|y^*, \sigma_*^2) \quad (22)$$

Among them, expected value $y^* = \mu^T \phi(x^*)$, variance $\sigma_*^2 = \sigma_{MP}^2 + \phi(x^*)^T \Sigma \phi(x^*)$. Therefore, the distribution of the predicted value t^* of the sample x^* to be tested, the mean value $y^* = (x^*; \mu)$. In order to facilitate understanding, Fig. 2 shows the visual structure of the model, which contains the input layer, hidden layer and result output layer of the data.

2.3 Principle of Particle Swarm Optimization (PSO)

Kennedy and Eberhart proposed PSO in 1995 based on the simulation of bird flock foraging behavior. It is the use of particle swarm stochastic intelligence optimization characteristics, using the relationship between individuals and groups, through the individual particle swarm in the group of competition and cooperation generated by the group intelligence, and ultimately guide the optimization of search. PSO is widely used in pattern recognition and parameter optimization due to its advantages of less parameters, intelligent optimization, and fast convergence.

203 The particles in the PSO represent the answer to the problem to be solved, the coordinate vector
 204 $x_i = (x_1, x_2, \dots, x_d, \dots, x_D)$ of each particle, the flying speed $v_i = (v_{i1}, v_{i2}, \dots, v_{id}, \dots, v_{iD})$ of the particle,
 205 the historical optimal coordinate $P_i = (P_{i1}, P_{i2}, \dots, P_{id}, \dots, P_{iD})$ of the i th particle, the optimal coordinate
 206 $P_g = (P_{g1}, P_{g2}, \dots, P_{gd}, \dots, P_{gD})$ experienced by each particle, and the particle swarm is flying during the
 207 flight process. continuously updated.

$$208 \quad v_{id}^{k+1} = \omega v_{id}^k + c_1 r_1 (P_{id} - x_{id}^k) + c_2 r_2 (P_{gd} - x_{id}^k) \quad (23)$$

$$209 \quad x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \quad (i = 1, 2, \dots, m; d = 1, 2, \dots, D) \quad (24)$$

211 Among them, m is the particle swarm size, D is the particle swarm dimension, v_{id}^k is the iteration
 212 offset, k is the number of iterations, ω is the inertia weight, c_1 and c_2 are the acceleration factors,
 213 and r_1 and r_2 are based on random number between $[0,1]$. The overall schematic diagram is shown
 214 in Fig. 3.

215 For the selection of the acceleration parameter, it is known from the literature (Wang and Ma, 2017)
 216 that in general $c_1 = c_2$, which can be chosen as any constant between 1.8 and 2.0. In the manuscript, we
 217 define the acceleration parameter as $c_1 = c_2 = 2$. For the choice of population size, the population size N is
 218 determined by the complexity of the problem. Too small a population size will make the results less
 219 accurate, while too large a population size will make the processing slower, computationally more
 220 expensive and time-consuming. A more appropriate value needs to be selected by synthesizing between
 221 the degree of accuracy and the cost effective. of calculation, which can be done by summarizing the
 222 existing references (Wen and Liu, 2004) and selecting the appropriate population size, so the population
 223 size $N=20$ is selected in this paper.

224 For the maximum number of iterations of the particle swarm algorithm, the larger the maximum
 225 number of iterations, the better, under the premise of satisfying the minimum error. Considering the

226 existing literature, the maximum number of iterations chosen in this paper is $M=1000$.

227 ***2.4 Comparison of the merits and demerits of the models***

228 PCA can make RVM more efficient and convenient in analyzing sample data, and PSO makes the problem of
229 kernel function parameters of RVM solved. The PCA-PSO-RVM model proposed in this paper is more advanced
230 and computationally more powerful than the RVM model, PSO-RVM model and PCA-RVM model. The results of
231 comparing the above models are shown in Table 1 below:

232 **3 Building the predictive models**

233 ***3.1 Background of the project***

234 In this paper, the ground settlement data for the shield construction from East New Town Station to Dong
235 Xiaoying Station of Beijing Metro Line 6 are used. The section of the interval passes through green areas and
236 Songlang intersection along the line, where there are several rainwater, power and telecommunication pipelines at
237 Songlang Road intersection and Canal East Street Southeast, with a total length of 842.95m. Shield construction
238 interval tunnel design section is circular, the outer diameter is 6.0m, the inner diameter 5.4m, the ground elevation
239 along the interval 19.5m to 19.8m, interval tunnel bottom buried depth 13.7m to 21m, through the stratum mainly
240 includes a layer of fine powder sand, medium and coarse sand layer, local sandwich powder clay layer.

241 ***3.2 data samples***

242 The surface settlement of shield construction is affected by the interaction of soil parameters and
243 construction parameters, and there are often a series of problems such as uncertainty and randomness when
244 selecting influencing factors. In this paper, PCA is used to reduce the dimension of multiple influencing
245 factors to obtain new principal component variables, and then use the PSO-RVM model to predict.

246 Based on the analysis of soil characteristics and shield machine parameters during the actual
247 construction of Beijing Metro Line 6, the jack thrust (F), grouting pressure (P), overburden thickness (H),
248 weighted average compression modulus (E_s), weighted average cohesion (C), weighted average natural
249 density (ρ), and weighted average internal friction angle (φ), a total of seven conventional physical
250 parameters, were comprehensively selected. These seven parameters were used as the relevant influencing
251 factors of surface settlement (S), and the relevant data are shown in Table 2. Equation (6) was calculated
252 for the seven influencing factors in the 51 sets of data after standardization in Table 2, and the correlation
253 coefficient matrix was obtained as shown in Table 3.

254 It can be seen from Table 3 that the absolute values of the correlation coefficients between the seven
255 influencing factors, such as jack thrust, grouting pressure, and covering soil thickness, are all between 0
256 and 1, and there is correlation between each factor, and the closer the correlation coefficient is to one
257 factor. the greater the correlation. In order to further explore the specific influence value of each factor,
258 the score diagram of 51 groups of data is obtained based on the PCA principle Fig. 4 (a), and the
259 contribution rate and cumulative contribution rate of each factor are calculated Fig. 4 (b).

260 Fig. 4(b) shows that the contribution rate of jack thrust (F), grouting pressure (P) and covering soil
261 thickness (H) is the largest, and the contribution rates of other influencing factors decrease in turn. The
262 cumulative contribution rate of the first four influencing factors is 86.539% and exceeds 85%, indicating
263 that it contains the amount of information represented by the seven factors. According to the cumulative
264 contribution rate, four principal component variables are extracted. Each principal component variable is
265 equal to the product of the seven influencing factors and their corresponding seven component score
266 coefficients.

267 ***3.3 Surface settlement model of shield tunneling construction***

268 In this paper, 7 influencing factors are dimensionality reduced into 4 principal component variables

269 through PCA, and the 4 principal component variables are selected as the variable input layer of the shield
270 construction surface settlement, and the surface settlement is used as the output layer. The optimal
271 parameters were automatically retrieved through PSO, and then a PCA-PSO-RVM correlation model
272 based on four principal component variables was established according to the principle of the PCA-PSO-
273 RVM regression prediction model.

274 ***3.4 prediction steps***

275 (1) The input data of the sample is the four principal components of the surface settlement of the
276 shield construction, and the output data is the surface settlement, and the influencing factors are
277 standardized.

278 (2) Initialize the position and velocity of the particle swarm, determine the size of the particle
279 swarm, and update the parameters of the PSO model according to the formula.

280 (3) It is judged whether the termination condition is met, and the optimal kernel function parameters
281 of the RVM model are further calculated, and the PSO-RVM prediction model is established.

282 (4) The predicted value and the corresponding measured value are compared and analyzed for
283 multiple indicators to verify the accuracy and reliability of the model. The overall process is shown in Fig.
284 5.

285 **4 Validation**

286 In this paper, the four principal component variables processed by PCA and the relevant data of
287 surface settlement are used to verify the accuracy of the PSO-RVM prediction model, and the optimal
288 solution is found after many times of learning. The prediction results of PCA-PSO-RVM model and RVM
289 model, PCA-RVM model and PSO-RVM model in surface settlement were compared under the same
290 sample conditions, as shown in Table 4.

291 It can be seen from Table 4 that in the relative error of the PCA-PSO-RVM machine model, the
 292 smallest sample No. 51 and No. 46 are only 0.06%, while in the PSO-RVM, PCA-RVM and RVM models,
 293 the minimum relative errors are 0.21%, 0.56%, 0.97%. In the comparison of the maximum relative error,
 294 the prediction error of the PCA-PSO-RVM model is also the smallest, only 0.73%. The overall analysis
 295 PCA-PSO-RVM model is the most accurate, and Fig. 6 more intuitively shows the distribution
 296 characteristics of the prediction results of the four models.

297 It can be seen from Fig. 6 that the prediction results of PCA-PSO-RVM are closer to the true value
 298 than the prediction results of the PSO-RVM model, the prediction results of the PCA-RVM model and the
 299 RVM model, and the fitting degree is significantly higher than that of the other three prediction models.
 300 The predicted value of each sample of the RVM model has the largest dispersion, especially the predicted
 301 value of No. 46, 47 and 50 samples obviously deviates from the measured value. In order to better compare
 302 the overall prediction accuracy, dispersion and balance of the four models, the average relative error MRE,
 303 root mean square error RMSE, and Theil inequality coefficient TIC of the four models were calculated
 304 respectively (Murray_smith, 1998). Calculated as follows:

$$305 \quad \text{MRE} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - y'_i|}{y_i} \times 100\% \quad (27)$$

$$307 \quad \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2} \quad (28)$$

$$308 \quad \text{TIC} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2}}{\sqrt{\sum_{i=1}^n y_i^2 + \sum_{i=1}^n (y'_i)^2}} \quad (29)$$

309 Among them, n is the number of samples; y_i is the actual monitoring value; y'_i is the model
 310 predicted value. The specific distribution of the average relative error MRE, the root mean square error
 311 RMSE, and the Theil inequality coefficient TIC is shown in Fig. 7.

312 Fig. 7(a) highlights the comparison of the degree of error of the models. The average relative errors
 313 of PCA-PSO-RVM and PSO-RVM are 0.17% and 1.25%, while the average relative errors of PCA-RVM
 314 model and RVM model are 1.63% and 3.76%. The relative degree of PCA-PSO-RVM is the smallest. Fig.
 315 7(b) highlights the dispersion of the model prediction results, the rms of PCA-PSO-RVM and PSO-RVM

316 are 0.0714 and 0.4744, and the PCA-RVM model and RVM model are 0.5981 and 1.9184. Fig. 7(c)
 317 highlights the degree of balance of the model prediction results, with the Theil inequality coefficients of
 318 0.027% and 0.183% for PCA-PSO-RVM and PSO-RVM, 0.231% and 0.775% for PCA-RVM model and
 319 RVM model, PCA-PSO-RVM has the least volatility. The comparison of the calculation results of the
 320 three indices shows that the PCA-PSO-RVM model has a greater advantage. The PCA-PSO-RVM model
 321 proposed in this paper has higher overall prediction accuracy, less discreteness and higher reliability.

322 **5 Sensitivity analysis**

323 Jack thrust (F), grouting pressure (P), overburden thickness (H), weighted average compression
 324 modulus (E_s), weighted average cohesion (C), weighted average natural density (ρ) and weighted average
 325 internal friction angle (φ) affect the surface settlement of shield tunneling in varying degrees. In order to
 326 explore the sensitivity of the seven influencing factors, on the basis of MRE, RMSE and TIC calculated
 327 by each prediction model, the impact degree of the factors is evaluated and compared, which is helpful for
 328 researchers to determine and pay attention to the parameters of surface settlement during shield
 329 construction. The formula is as follows:

$$330 \quad R_{1j} = \text{MRE}_j / \text{MRE} \quad (30)$$

$$331 \quad R_{2j} = \text{RMSE}_j / \text{RMSE} \quad (31)$$

$$332 \quad R_{3j} = \text{TIC}_j / \text{TIC} \quad (32)$$

333 In the formula, R_{1j} , R_{2j} , R_{3j} are the ratios of the three indicators (MRE, RMSE and TIC),
 334 respectively, and the results obtained by the new model are compared with those obtained by the original
 335 model. The value of j ($=1, 2, \dots, 7$) corresponds to the seven influencing factors (F, P, H, E_s, C, ρ and φ)
 336 not considered by the new model in turn, RMSE, TIC, and MRE are the missing j th factor, respectively.
 337 Standard deviation, root mean square error, and Theil inequality coefficient. MRE_j , RMSE_j , and TIC_j
 338 represent the three indicators of the initial prediction model, respectively. The size of R_{ij} is proportional
 339 to the sensitivity of surface settlement.

340 Table 5 shows the calculated values of the three indicators of the model prediction results without
341 corresponding factors, as well as the comparison results of the three indicators. Compared with the
342 calculation results of Fig. 7, the three indicators calculated by the six-factor model are all larger than the
343 indicators calculated by the original model, emphasizing the importance of the 7-factor model and paving
344 the way for easy identification of the sensitivity of the factors. The comparison values of the three
345 indicators are all greater than 1, indicating that each factor will affect the prediction results of Surface
346 settlement to varying degrees. The ranking of sensitive factors further compares the degree of influence
347 of the factors. The above shows that the seven factors selected in this paper are reasonable. In order to
348 intuitively show the influence of each factor on the surface settlement, the radar chart in Fig. 8 is used to
349 represent the distribution of multiple indicators.

350 The size of the radar chart reflects the quality of the evaluation object, which can be used as a basis
351 to diagnose and control the evaluation object. For indicators that are closer to the center, the more measures
352 to be taken to improve. It can be seen that F , P , H , E_s and C are the closest to the center, and corresponding
353 measures should be taken in the study of the surface settlement of shield construction. Among the factors
354 affecting the surface settlement, the weighted average natural density (ρ) and the weighted average internal
355 friction angle (φ) have the largest sensitivity factor index to the surface settlement of shield construction,
356 indicating that these two factors are closely related to the surface settlement. Then the jack thrust (F),
357 grouting pressure (P), soil cover thickness (H), weighted average compressive modulus (E_s) and weighted
358 average cohesion (C) need to be improved in the study.

359 **6 Conclusion**

360 Establishing an accurate prediction model for the surface settlement of shield construction can help
361 control the shield construction process and reduce the adverse effects of surface settlement caused by the

362 construction process. Based on the shield construction data of Beijing Metro Line 6, a PCA-PSO-RVM
363 prediction model is established. The main conclusions are as follows:

364 (1) The surface settlement of shield tunneling is affected by multiple factors, and there is an intricate
365 nonlinear mapping relationship between each factor and the surface settlement. The PCA-PSO-RVM
366 prediction model proposed in this paper can accurately establish the nonlinear mapping relationship
367 between surface settlement and influencing factors, simplify complex problems and facilitate the
368 establishment of prediction models.

369 (2) Examples show that the prediction of surface settlement for shield construction using the PCA-
370 PSO-RVM model yields better results than the RVM model, PCA-RVM model and PSO-RVM model,
371 and that the PCA-PSO-RVM model has a clear advantage for problems with a small number of learning
372 samples for prediction. The 7 influencing factors were reduced into 4 linearly independent principal
373 components by PCA, and the redundant information among the influencing factors was eliminated.
374 Through sensitivity factor analysis, the precise sensitivity and discrete sensitivity of the main influencing
375 factors are further understood, and it is known that the weighted average internal friction angle (φ) has
376 the greatest influence among the influencing factors.

377 (3) In the field inspection, combined with the method proposed in this study, the RVM model can be
378 used to collect more extensive information to screen out the factors that have a greater impact on the
379 surface settlement of shield construction, and to summarize a more complete nonlinear mapping
380 relationship. Then a more optimized PCA-PSO-RVM model is obtained, which improves the accuracy
381 and applicability of model prediction. At the same time, the parameters and influencing factors can be
382 adjusted reasonably according to the actual problems in the shield construction site and the valuable
383 opinions put forward by the researchers, so that the model has a wider scope of application.

384 **7 Data Availability**

385 Some or all data, models, or code that support the findings of this study are available from the
386 corresponding author upon reasonable request.

387 **8 Acknowledgements**

388 The authors would like to acknowledge the financial support from the National Natural Science
389 Foundation of China under Grant No.52068016. The work in this paper was also supported by the Guangxi
390 Key Laboratory of Geomechanics and Geotechnical Engineering (Grant No.19-Y-21-9, 20-Y-XT-01), the
391 High Level Innovation Team and Outstanding Scholar Program of Universities in Guangxi Province.
392 (Grant No. 202006) and the Guangxi Natural Science Foundation under Grant Nos.
393 2020GXNSFAA297118 and 2020GXNSFAA159125.

394

395 **References:**

- 396 Azhdar, R. and Nazemi, A. (2020) Modeling of incentive-based policies for demand management for the Tehran subway.
397 *Travel Behaviour and Society* **20**: 174-180, <https://doi.org/10.1016/j.tbs.2020.03.014>.
- 398 Borthakur, N. and Dey, A. (2020) Evaluation of Group Capacity of Micropile in Soft Clayey Soil from Experimental
399 Analysis Using SVM-Based Prediction Model. *International Journal of Geomechanics* **20(3)**:
400 [https://doi.org/10.1061/\(ASCE\)GM.1943-5622.0001606](https://doi.org/10.1061/(ASCE)GM.1943-5622.0001606).
- 401 Chou, W. and Bobet, A. (2002) Predictions of ground deformations in shallow tunnels in clay. *Tunnelling and*
402 *Underground Space Technology* **17(1)**: 3-19, [https://doi.org/10.1016/S0886-7798\(01\)00068-2](https://doi.org/10.1016/S0886-7798(01)00068-2).
- 403 Chen, R., Zhang, P., Kang, X., Zhong, Z., Liu, Y. and Wu, H. (2019) Prediction of maximum surface settlement caused
404 by earth pressure balance (EPB) shield tunneling with ANN methods. *Soils and Foundations* **59(2)**: 284-295,
405 <https://doi.org/10.1016/j.sandf.2018.11.005>.
- 406 Du, R. and Zheng, S. (2020) Agglomeration, housing affordability, and new firm formation: The role of subway network.
407 *Journal of Housing Economics* **48**: 101668, <https://doi.org/10.1016/j.jhe.2020.101668>.
- 408 Galuzio, P., de Vasconcelos Segundo, E., Coelho, L. and Mariani, V. (2020) MOBOpt — multi-objective Bayesian
409 optimization. *SoftwareX* **12**: 100520, <https://doi.org/10.1016/j.softx.2020.100520>.
- 410 Hao, R., Ji, Y. and Ni, Z. (2015) Study on predicting the surface settlement for shield tunneling based on DEACO—
411 WNN. *Journal of Railway Engineering Society* **32(1)**: 12-16.
- 412 Kasper, T. and Meschke, G. (2006) On the influence of face pressure, grouting pressure and TBM design in soft ground
413 tunnelling. *Tunnelling and Underground Space Technology* **21(2)**: 160-171,
414 <https://doi.org/10.1016/j.tust.2005.06.006>.
- 415 Kennedy, J. and Eberhart, R. (1995) Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on*
416 *Neural Networks. IEEE.* **4**: 1942-1948, <https://doi.org/10.1109/ICNN.1995.488968>
- 417 Li, S. H., Zhang, M. J., and Li, P. F. (2021) analytical solutions to ground settlement induced by ground loss and
418 construction loadings during curved shield tunneling. *Journal of Zhejiang University-SCIENCE A*, **22(4)**, 296-313,
419 <https://doi.org/10.1631/jzus.A2000120>
- 420 Liu, B., Yu, Z., Han, Y., Wang, Z., Zhang, R. and Wang, S. (2020) Analytical solution for the response of an existing
421 tunnel induced by above-crossing shield tunneling. *Computers and Geotechnics* **124**: 103624,
422 <https://doi.org/10.1016/j.compgeo.2020.103624>.
- 423 Murray_smith, D. (1998) Methods for the external validation of continuous system simulation models:a review.

- 424 *Mathematical and Computer Modelling of Dynamical Systems* **4(1)**: 5-31,
 425 <https://doi.org/10.1080/13873959808837066>.
- 426 Marini, F. and Walczak, B. (2015) Particle swarm optimization (PSO). A tutorial. *Chemometrics and Intelligent*
 427 *Laboratory Systems* **149**: 153-165, <https://doi.org/10.1016/j.chemolab.2015.08.020>.
- 428 Ma, Z. and Hanson, T. (2020) Bayesian nonparametric test for independence between random vectors. *Computational*
 429 *Statistics & Data Analysis* **149**: 106959, <https://doi.org/10.1016/j.csda.2020.106959>.
- 430 Ocak, I. and Seker, S. (2013) Calculation of surface settlements caused by EPBM tunneling using artificial neural network,
 431 SVM, and Gaussian processes. *Environmental Earth Sciences* **70(3)**: 1263-1276, [https://doi.org/10.1007/s12665-](https://doi.org/10.1007/s12665-012-2214-x)
 432 [012-2214-x](https://doi.org/10.1007/s12665-012-2214-x).
- 433 Peck, R. B. (1969) Deep excavations and tunneling in soft ground. Proceeding of 7th international conference on soil
 434 mechanics and foundation engineering. *Mexico City: State of the Art Report*: 225-290.
- 435 Salimi, A., Rostami, J., Moormann, C. and Delisio, A. (2016) Application of non-linear regression analysis and artificial
 436 intelligence algorithms for performance prediction of hard rock TBMs. *Tunnelling and Underground Space*
 437 *Technology* **58**: 236-246, <https://doi.org/10.1016/j.tust.2016.05.009>.
- 438 Sharghi, M., Chakeri, H. and Ozcelik, Y. (2017) Investigation into the effects of two component grout properties on
 439 surface settlements. *Tunnelling and Underground Space Technology* **63**: 205-216,
 440 <https://doi.org/10.1016/j.tust.2017.01.004>.
- 441 Singh, D., Aromal, V. and Mandal, A. (2018) Prediction of surface settlements in subway tunnels by regression analysis.
 442 *International Journal of Geotechnical Engineering* **14(7)**: 836-842,
 443 <https://doi.org/10.1080/19386362.2018.1477294>.
- 444 Su, M., Liu, Y., Xue, Y., Cheng, K., Ning, Z., Li, G. and Zhang, K. (2021) Detection method of karst features around
 445 tunnel construction by multi-resistivity data-fusion pseudo-3D-imaging based on the PCA approach. *Engineering*
 446 *Geology* **288**: 106127, <https://doi.org/10.1016/J.ENGGEOL.2021.106127>.
- 447 Tipping, M.(2001) Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*
 448 **1(3)**: 211-244.
- 449 Tipping, M.(2001) The relevance vector machine. *Advances in Neural Information Processing Systems* (**12**): 652-658.
- 450 Verruijt, A. and Booker, J. (1998) Surface settlements due to deformation of a tunnel in an elastic half plane.
 451 *Géotechnique* **48(5)**: 709-713, <https://doi.org/10.1680/geot.1998.48.5.709>.
- 452 Wang, F., Gou, B. and Qin, Y. (2013) Modeling tunneling-induced ground surface settlement development using a wavelet
 453 smooth relevance vector machine *Computers and Geotechnics* **54**: 125-132.
- 454 Wang, G. and Ma, Z. (2017) Hybrid particle swarm optimization for first-order Reliability Method. *Computers and*
 455 *Geotechnics*, **81**: 49–58, <https://doi.org/10.1016/j.compgeo.2016.07.013>.
- 456 Wang, J., Mohammed, A., Macioszek, E., Ali, M., Ulrikh, D. and Fang, Q., (2022) A Novel Combination of PCA and
 457 Machine Learning Techniques to Select the Most Important Factors for Predicting Tunnel Construction Performance.
 458 *Buildings* **12(7)**: 919, <https://doi.org/10.3390/BUILDINGS12070919>.
- 459 Wen, Z. and Liu, Y. (2004) Reactive power optimization based on PSO in a practical power system. *IEEE Power*
 460 *Engineering Society General Meeting, 2004*, 239-243, <https://doi.org/10.1109/pes.2004.1372792>.
- 461 Yao, Y., Liu, Y., Yu, Y., Xu, H., Lv, W., Li, Z. and Chen, X. (2013) K-SVM: An Effective SVM Algorithm Based on K-
 462 means Clustering. *Journal of Computers* **8(10)**: 2632-2639, <https://doi.org/10.4304/jcp.8.10.2632-2639>.
- 463 Yan, H., Liu, K., Xu, C. and Zheng, W. (2022) A novel method for identifying geomechanical parameters of rock masses
 464 based on a PSO and improved GPR hybrid algorithm. *Scientific Reports* **12(1)**:1-18, [https://doi.org/10.1038/s41598-](https://doi.org/10.1038/s41598-022-09947-7)
 465 [022-09947-7](https://doi.org/10.1038/s41598-022-09947-7).
- 466 Zhou, C., Kong, T., Zhou, Y., Zhang, H. and Ding, L. (2019) Unsupervised spectral clustering for shield tunneling

467 machine monitoring data with complex network theory. *Automation in Construction* **107**: 102924,
468 <https://doi.org/10.1016/j.autcon.2019.102924>.

469 Zhang, W., Li, H., Wu, C., Li, Y., Liu, Z. and Liu, H. (2021) Soft computing approach for prediction of surface settlement
470 induced by earth pressure balance shield tunneling. *Underground Space* **6(4)**: 353-363,
471 <https://doi.org/10.1016/j.undsp.2019.12.003>.

472 Zhang, W., Gu, X., Tang, L., Yin, Y., Liu, D. and Zhang, Y. (2022) Application of machine learning, deep learning and
473 optimization algorithms in geoenvironment and geoscience: Comprehensive review and future challenge. *Gondwana*
474 *Research* **109**: 1-17, <https://doi.org/10.1016/j.gr.2022.03.015>.

475

Table 1. Comparison of the merits and demerits of the models

Name of algorithm	RVM	PCA—RVM	PSO—RVM	PCA—PSO—RVM
Merits	<p>The operational efficiency of the kernel function is improved and the sparsity is enhanced.</p> <p>Reduced for the number of sample data, suitable for small sample data prediction.</p>	<p>The complexity of the raw data is reduced and the overall speed and accuracy of the model is improved compared to RVM. It can quickly find the focus for the analysis of data with multiple influencing factors.</p>	<p>As a probabilistic global optimization algorithm, there are more opportunities to solve the global optimal solution.</p> <p>The kernel function parameter problem of RVM is solved and the model reliability is enhanced.</p>	<p>The model has strong universality and fast convergence.</p> <p>The influence of data samples and RVM kernel function parameters on the calculation results is reduced, and the overall accuracy and running speed are increased.</p>
Demerits	<p>The choice of kernel function parameters has a large impact on the results.</p>	<p>There are some limitations in using PCA for eigenvalue decomposition, such as the transformation matrix must be a square matrix. Kernel function parameters have great influence on the results.</p>	<p>For functions with multiple local extreme points, it is easy to fall into local extreme points.</p>	<p>The data samples should not be selected too small, otherwise overfitting may occur.</p>

Table 2. Surface settlement for shield tunneling data set

Sample no.	F (kN)	P (MPa)	H (m)	E_s (MPa)	C (kPa)	ρ ($g \cdot cm^{-3}$)	φ ($^\circ$)	S (mm)
1	18600	0.20	13.60	15.50	8.60	1.84	29.20	-19.70
2	22000	0.19	14.70	14.20	7.20	1.79	24.30	-26.70
3	24600	0.19	15.30	13.30	8.10	1.74	22.50	-11.20
4	17100	0.19	18.20	19.30	26.80	1.78	22.10	-26.30
5	18900	0.18	16.20	13.50	8.80	1.75	22.40	-37.60
6	16100	0.21	13.50	15.40	8.80	1.80	29.30	-69.30
7	21500	0.20	13.10	19.20	11.90	1.90	31.60	-58.00
8	21200	0.19	21.00	29.40	5.30	1.75	37.60	-4.80
9	17700	0.19	19.30	15.60	9.70	1.95	37.80	-25.40
10	15100	0.18	18.80	10.80	20.40	1.70	18.40	-34.10
11	22200	0.18	19.80	38.30	7.60	1.85	39.10	-45.90
12	21800	0.19	14.30	17.60	9.40	1.80	25.10	-22.70
13	21300	0.22	14.60	24.80	13.40	1.90	26.20	-30.20
14	17100	0.20	23.20	41.30	2.90	1.75	41.30	-2.30
15	17900	0.19	18.50	20.40	35.80	1.64	20.60	-25.50
16	16100	0.21	19.10	13.80	7.60	1.74	35.30	-50.10
17	20500	0.20	16.70	10.30	20.20	1.66	18.10	-39.50
18	24800	0.22	15.60	40.40	51.20	1.80	23.10	-14.30
19	19200	0.20	16.50	20.80	35.40	1.70	20.20	-20.90
20	14600	0.19	13.40	19.10	11.30	1.90	31.40	-7.71
21	17600	0.19	15.10	16.20	15.80	1.75	23.10	-15.80
22	20600	0.20	17.30	13.10	7.40	1.75	35.20	-35.60
23	20800	0.19	16.20	9.30	9.20	1.60	19.80	-27.10
24	21800	0.21	14.10	24.30	13.80	1.90	26.20	-22.20
25	19400	0.20	16.50	22.00	22.30	1.80	24.40	-33.50
26	17100	0.21	21.20	47.60	3.90	1.85	43.20	-3.70
27	21200	0.20	21.60	25.10	3.20	1.70	33.30	-47.10
28	17500	0.19	17.90	14.60	12.40	1.65	20.60	-33.30
29	23800	0.20	15.90	16.50	20.40	1.70	21.40	-29.70
30	20600	0.21	22.90	47.20	3.80	1.85	43.20	-4.90
31	22900	0.19	14.40	17.80	9.80	1.80	25.70	-24.50
32	21900	0.22	13.80	12.30	6.40	1.75	28.10	-9.70
33	25300	0.19	14.90	46.80	43.50	1.90	25.50	-20.80
34	18200	0.21	17.00	15.30	9.80	1.85	37.60	-37.60
35	19000	0.19	17.60	11.80	5.30	1.70	29.30	-35.20
36	20600	0.19	20.50	41.80	3.20	1.75	41.80	-5.30
37	21000	0.20	21.00	23.60	4.80	1.75	30.50	-57.20
38	21800	0.21	14.10	18.50	10.40	1.80	33.20	-55.60
39	16600	0.21	19.40	12.80	6.40	1.75	35.20	-39.30
40	18800	0.20	13.60	14.00	7.20	1.80	30.10	-43.70
41	24800	0.22	15.40	39.80	49.10	1.80	23.40	-16.50*
42	22100	0.19	15.20	15.80	8.80	1.80	28.80	-44.10*
43	20100	0.19	16.40	9.80	8.30	1.65	20.40	-27.50*
44	18200	0.21	16.80	13.30	9.00	1.75	22.80	-39.20*
45	21100	0.22	14.90	25.10	14.00	1.85	25.50	-29.20*
46	21300	0.20	13.80	18.60	12.50	1.90	33.60	-69.30*
47	15300	0.19	12.90	17.40	11.80	1.90	35.10	-54.70*
48	16600	0.21	13.80	16.60	10.80	1.83	28.50	-23.30*
49	21800	0.19	14.50	15.60	7.80	1.85	25.10	-34.80*
50	21300	0.20	15.80	10.40	20.90	1.70	18.40	-39.50*
51	16600	0.21	13.90	15.90	10.00	1.85	29.60	-20.20*

Table 3. Correlation coefficient matrix

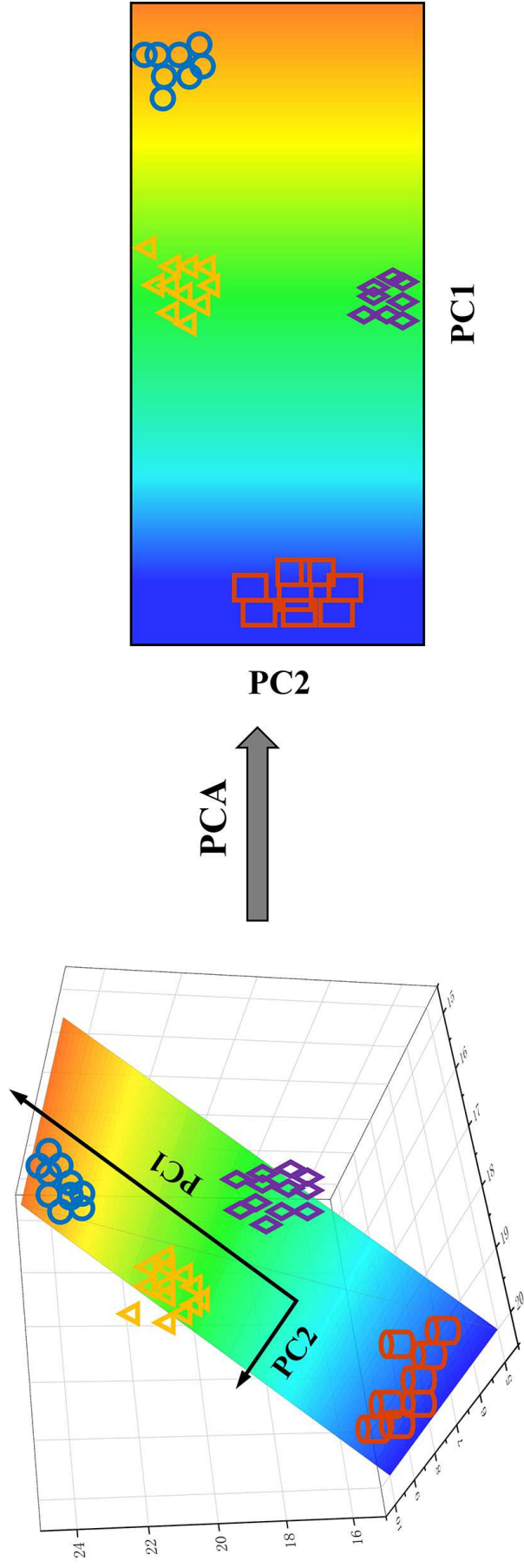
	F	P	H	E_s	C	ρ	φ
F	1.000	0.099	-0.148	0.287	0.304	0.020	-0.196
P	0.099	1.000	-0.127	0.200	0.150	0.226	0.140
H	-0.148	-0.127	1.000	0.438	-0.206	-0.313	0.450
E_s	0.287	0.200	0.438	1.000	0.257	0.332	0.480
C	0.304	0.150	-0.206	0.257	1.000	-0.038	-0.512
ρ	0.020	0.226	-0.313	0.332	-0.038	1.000	0.462
φ	-0.196	0.140	0.450	0.480	-0.512	0.462	1.000

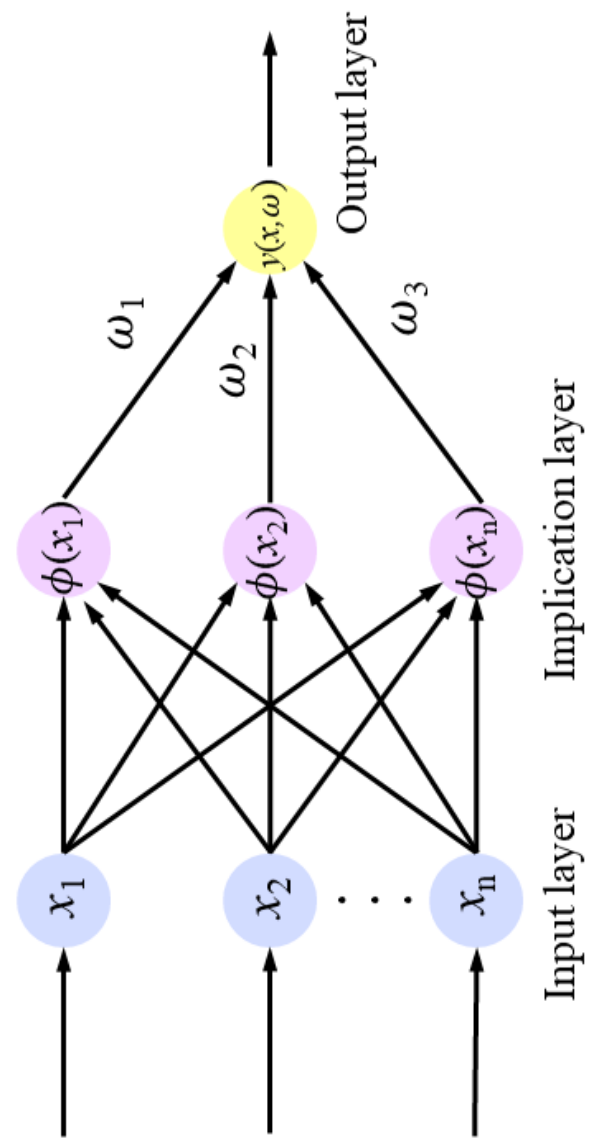
Table 4. Prediction results of different methods

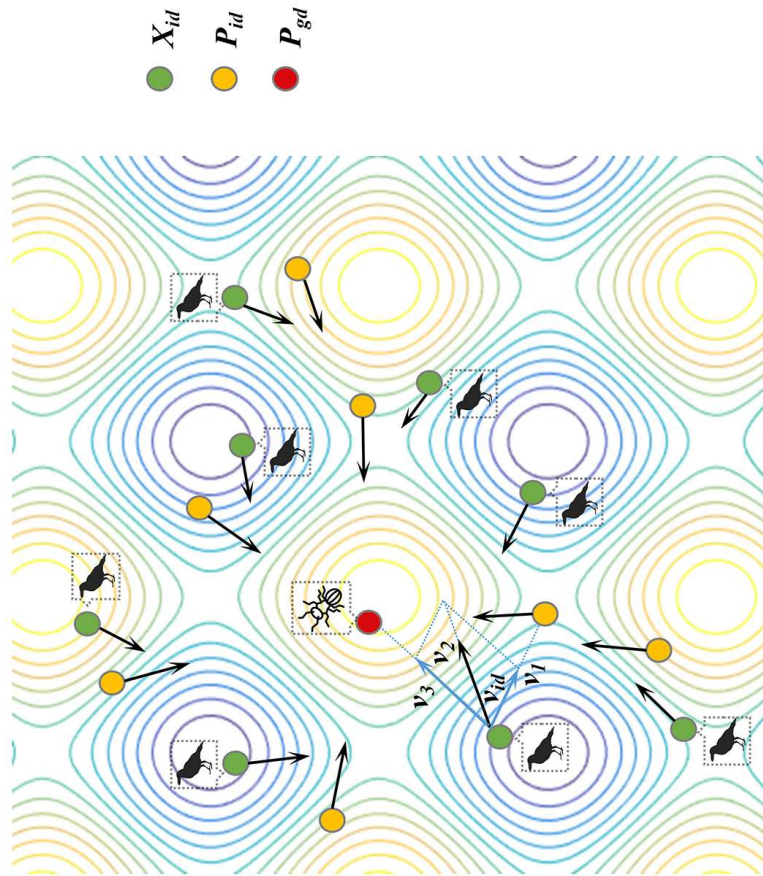
Sample no.	Measured value (m/s)	RVM		PCA-RVM		PSO-RVM		PCA-PSO-RVM	
		Predictive value (mm)	Relative error (%)	Predictive value (mm)	Relative error (%)	Predictive value (mm)	Relative error (%)	Predictive value (mm)	Relative error (%)
41	-16.50	-15.815	4.15	-16.267	1.41	-16.465	0.21	-16.556	0.34
42	-44.10	-43.509	1.34	-43.606	1.12	-43.725	0.85	-44.070	0.07
43	-27.50	-26.810	2.51	-27.346	0.56	-27.016	1.76	-27.298	0.73
44	-39.20	-37.487	4.37	-40.027	2.11	-39.757	1.42	-39.170	0.08
45	-29.20	-28.082	3.83	-28.654	1.87	-28.567	2.17	-29.132	0.23
46	-69.30	-65.031	6.16	-68.392	1.31	-68.565	1.06	-69.256	0.06
47	-54.70	-51.533	5.79	-54.005	1.27	-54.164	0.98	-54.663	0.07
48	-23.30	-22.543	3.25	-22.762	2.31	-22.988	1.34	-23.276	0.10
49	-34.80	-34.462	0.97	-34.490	0.89	-35.023	0.64	-34.769	0.09
50	-39.50	-37.110	6.05	-40.187	1.74	-39.014	1.23	-39.470	0.08
51	-20.20	-19.608	2.93	-19.517	3.38	-19.778	2.09	-20.187	0.06

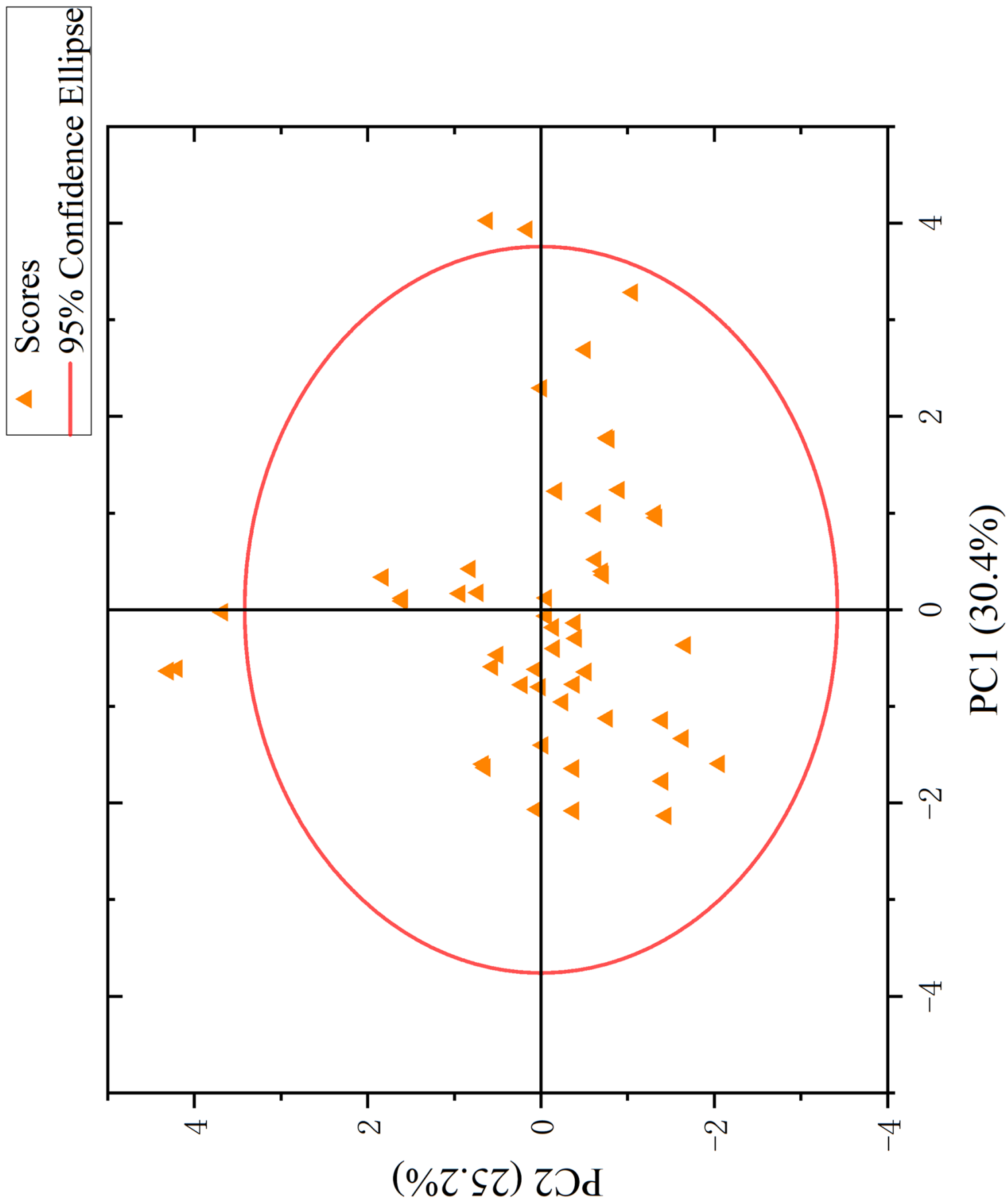
Table 5. Sensitive factor comparison result

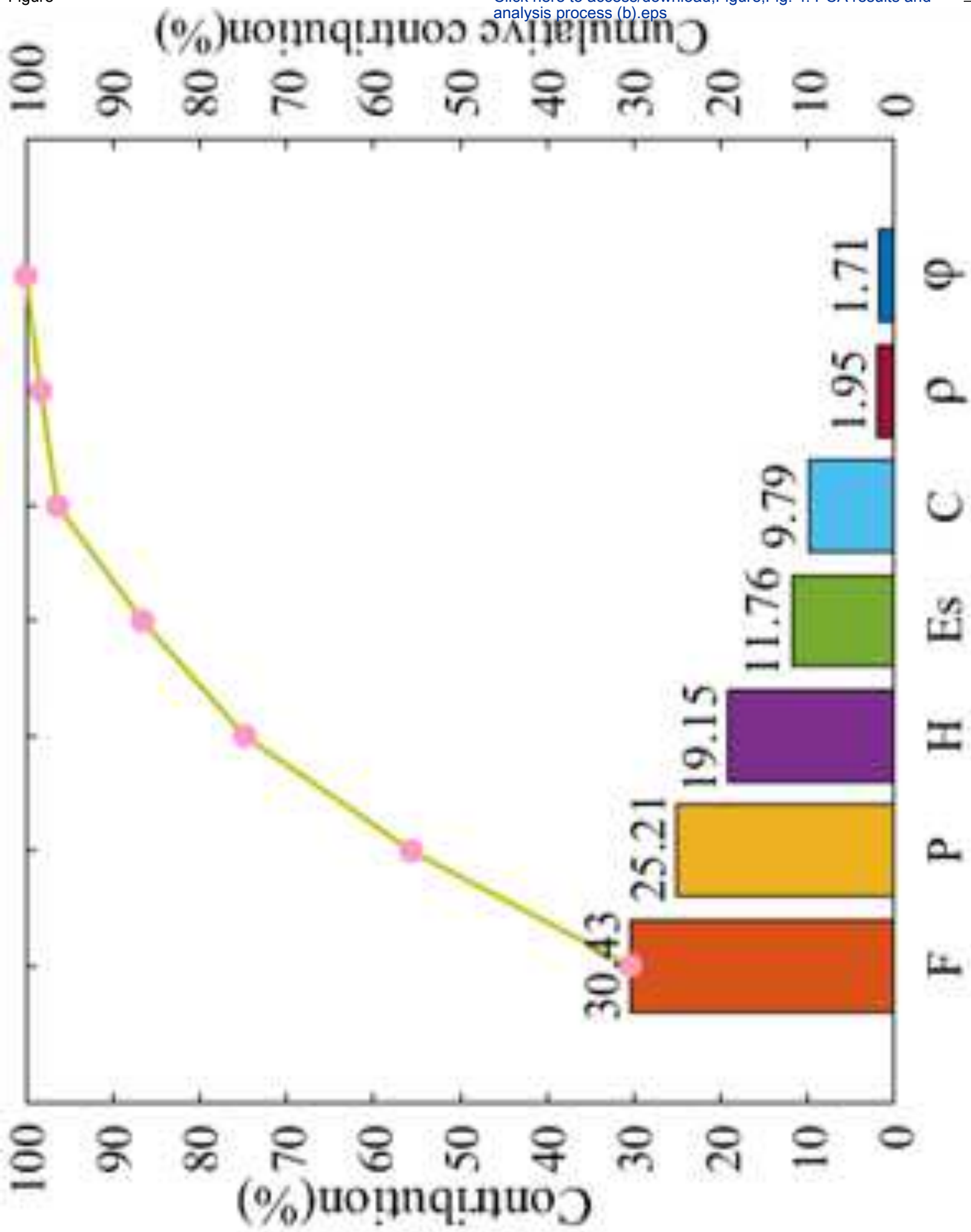
Research object		F	P	H	E_s	C	ρ	φ
Indicators	MRE(%)	0.20	2.20	0.21	1.20	0.95	5.23	8.23
	RMSE	0.082	0.712	0.099	0.476	0.308	2.068	3.217
	TIC(%)	0.032	0.275	0.038	0.182	0.119	0.804	1.189
Ratio	R_{1j}	1.18	12.94	1.24	7.06	5.59	30.76	48.41
	R_{2j}	1.15	9.97	1.39	6.67	4.31	28.96	45.06
	R_{3j}	1.19	10.19	1.41	6.74	1.61	10.86	16.07
Sensitivity factor order		7	3	6	4	5	2	1

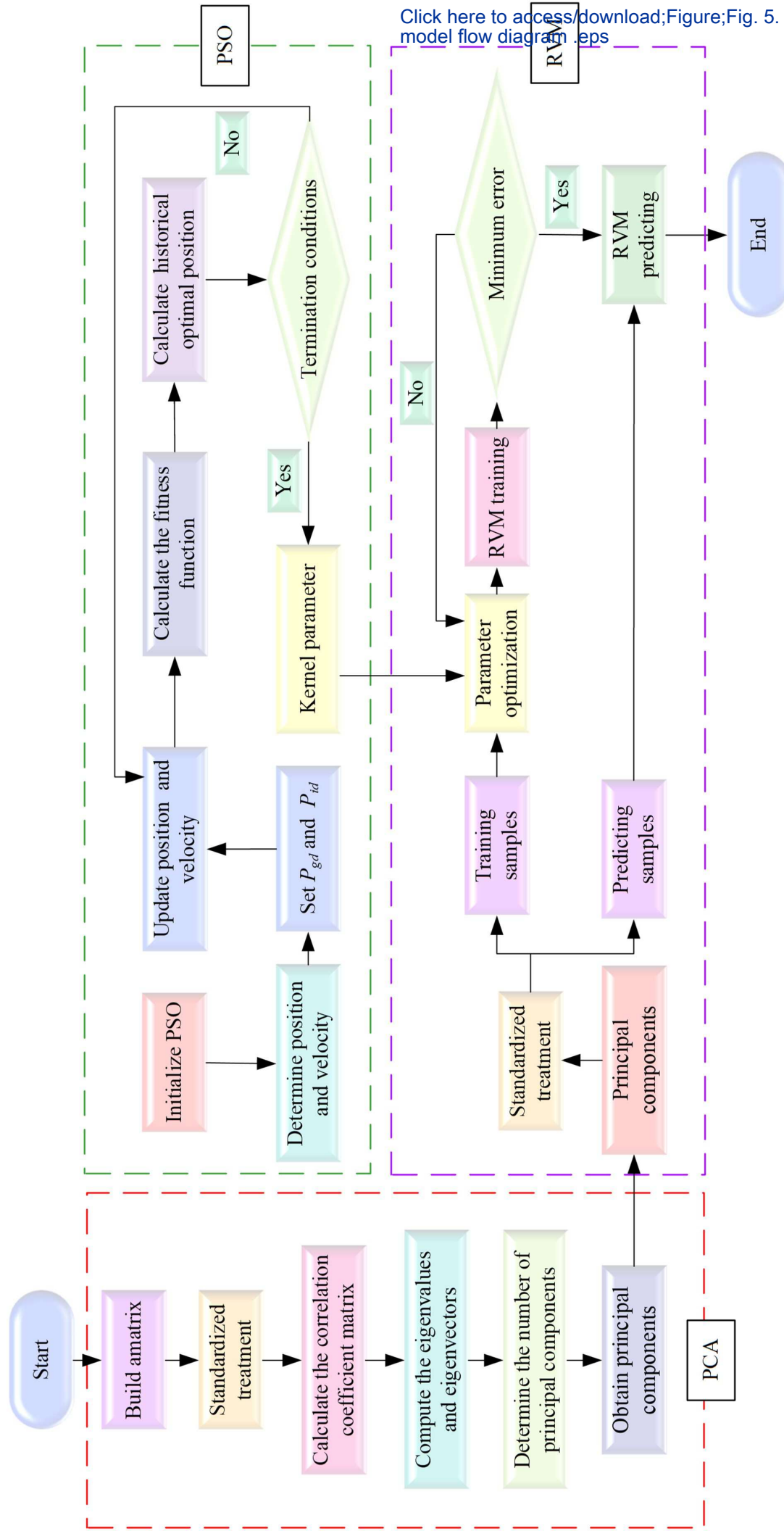


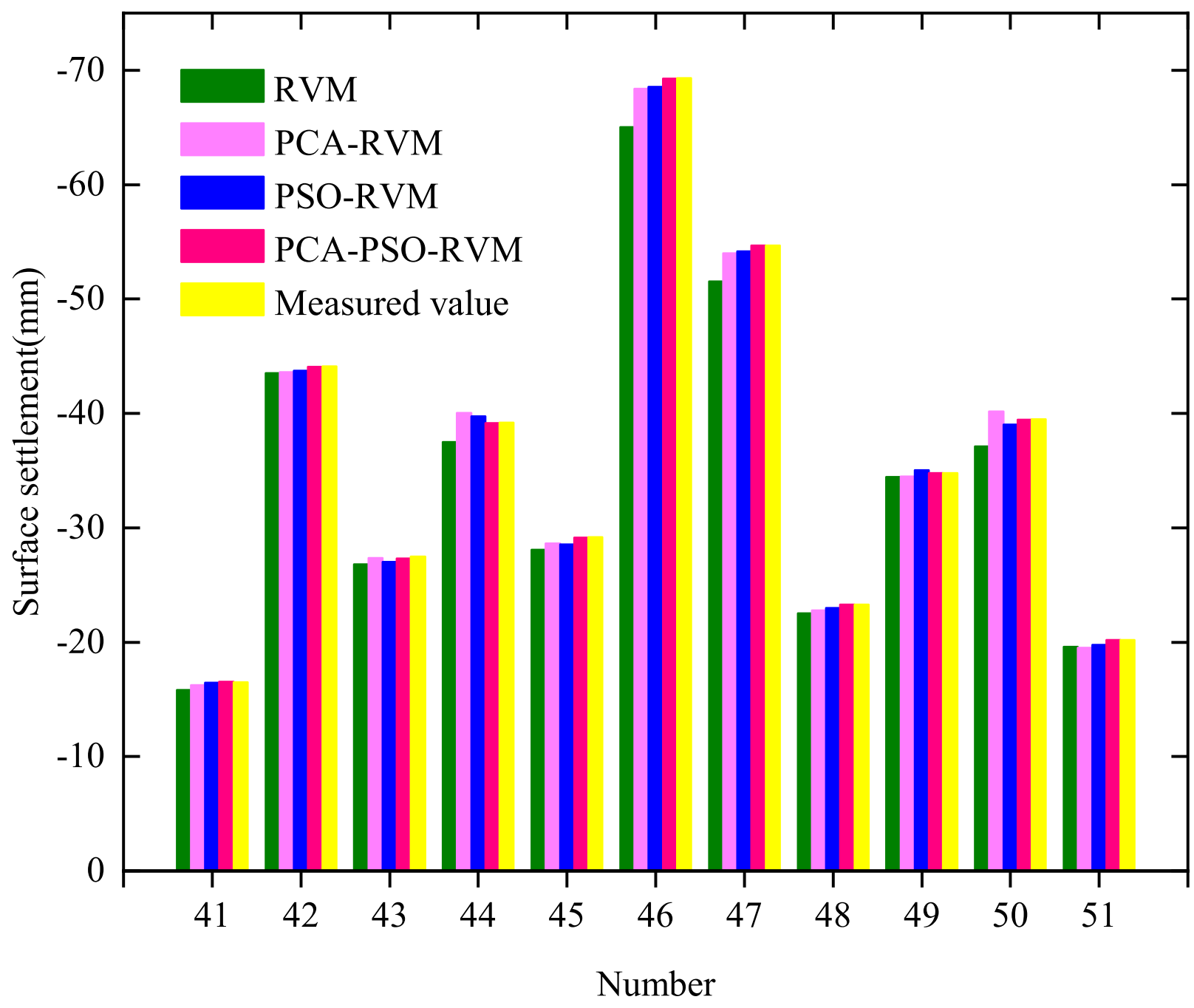


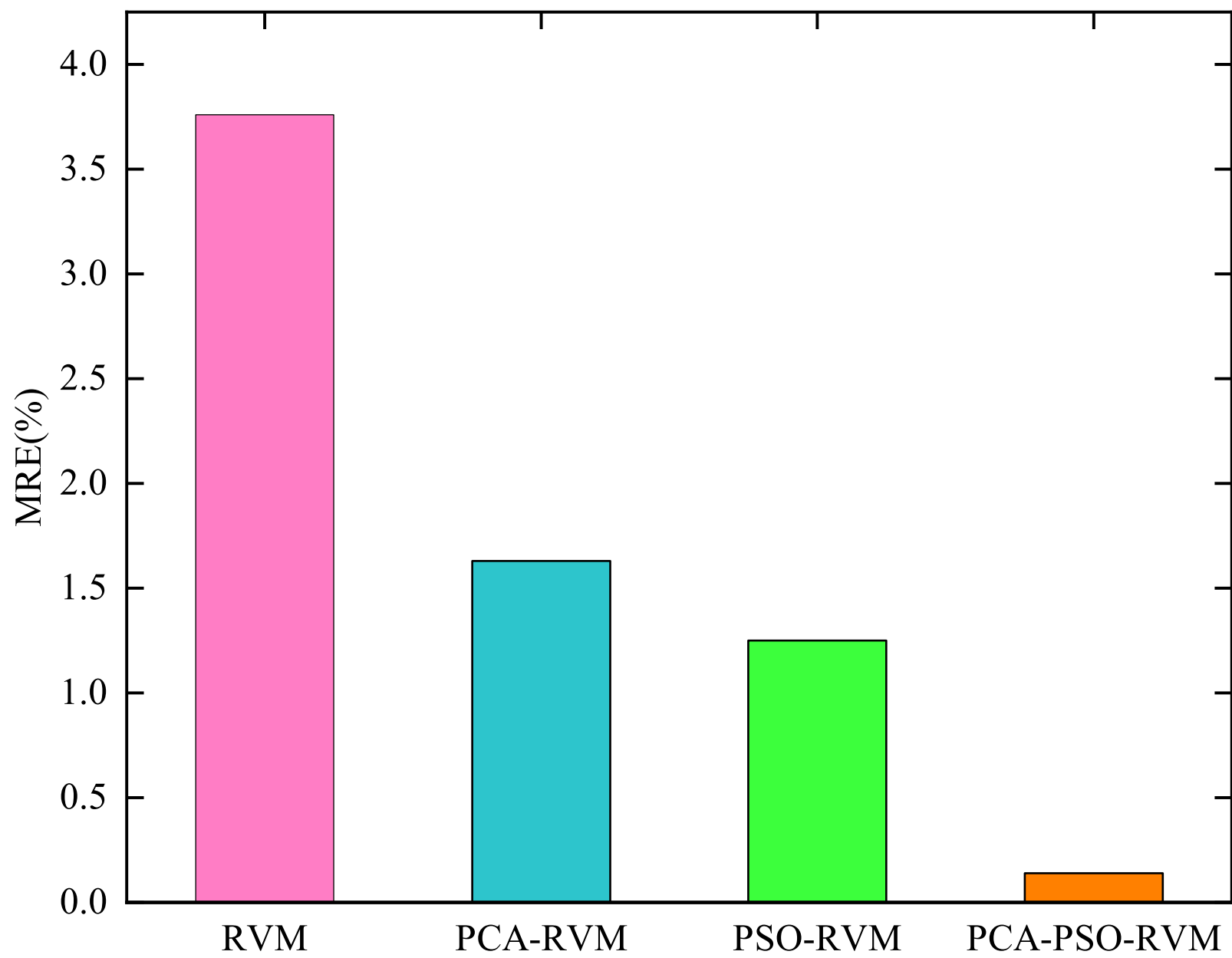


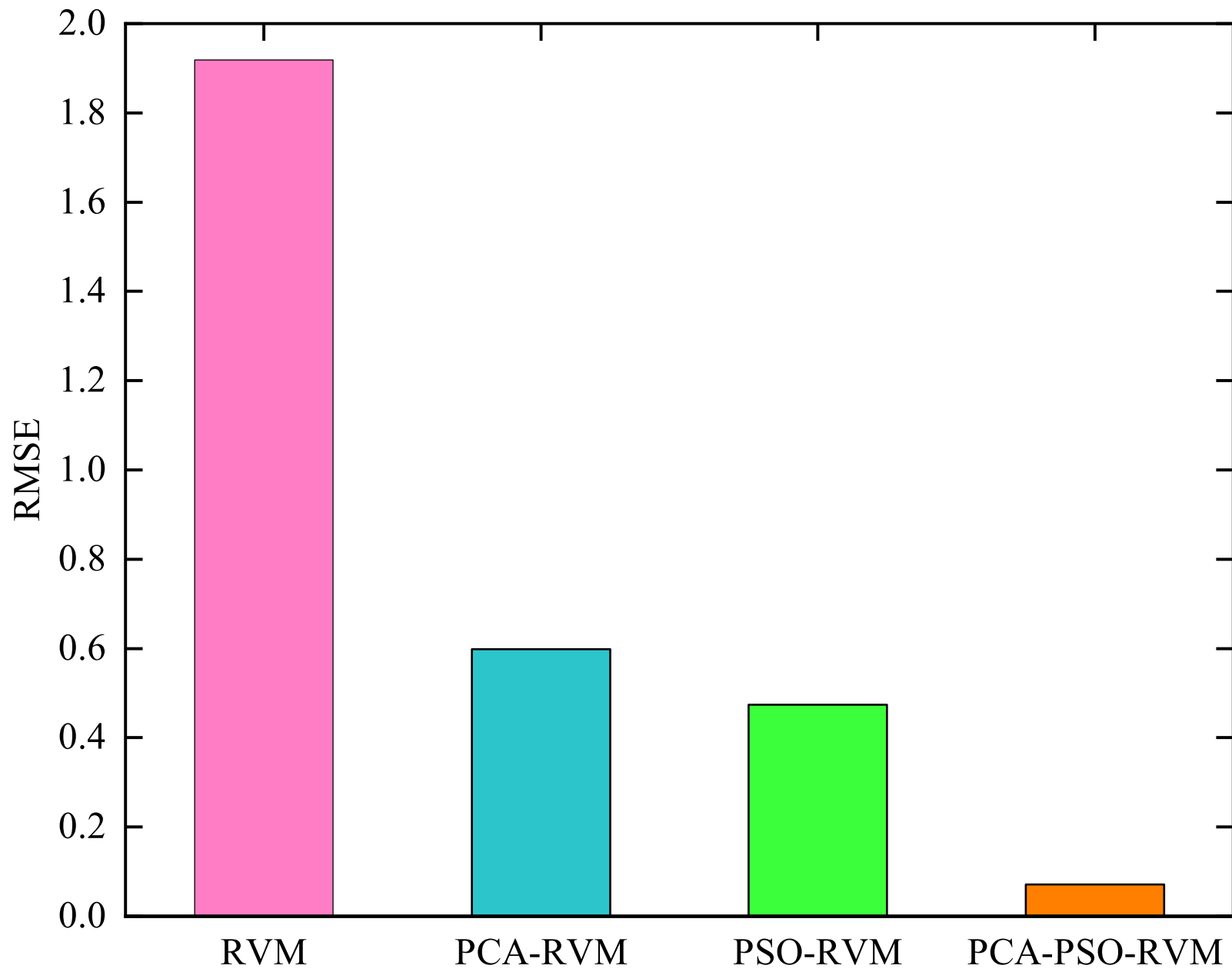


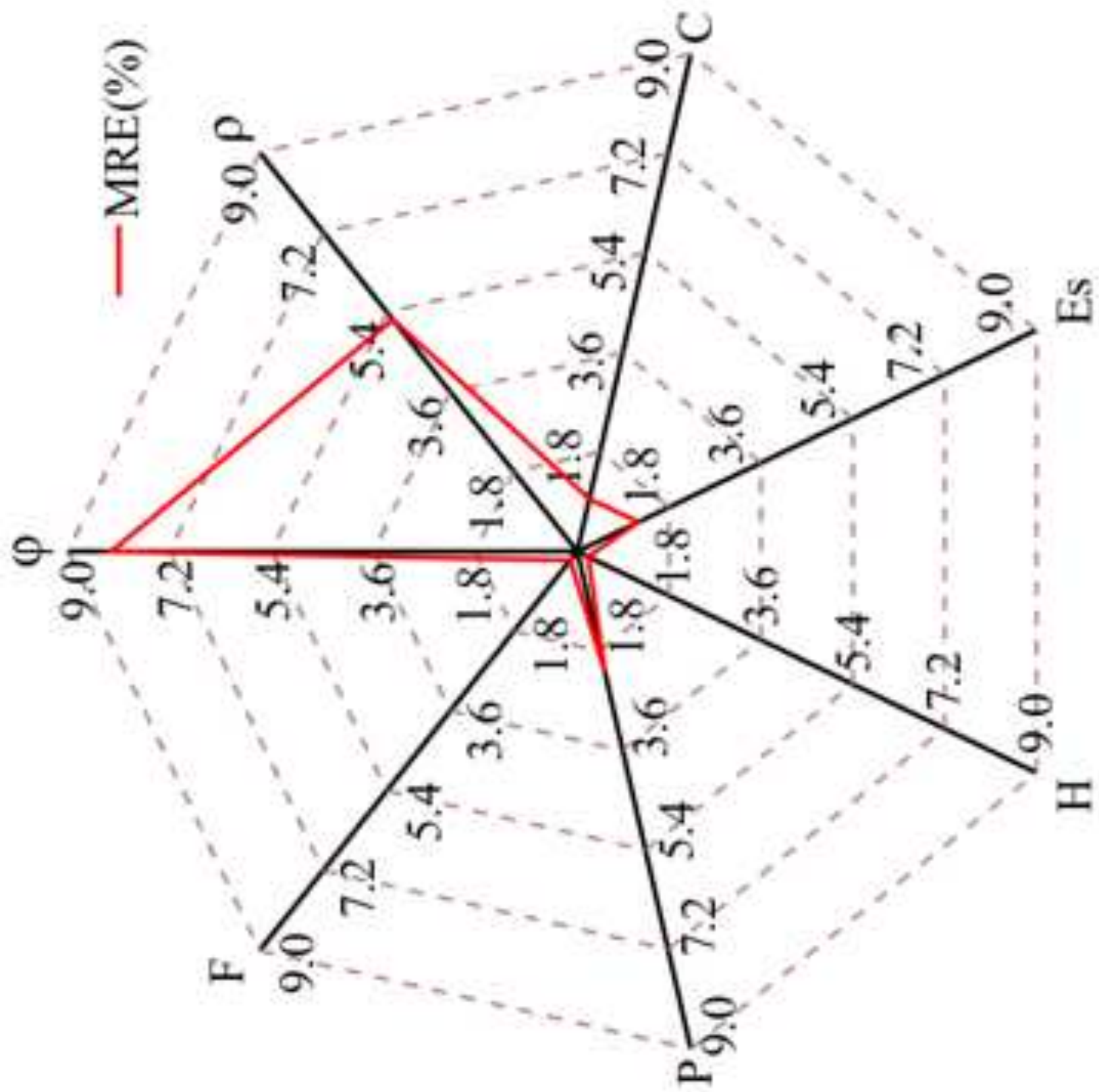


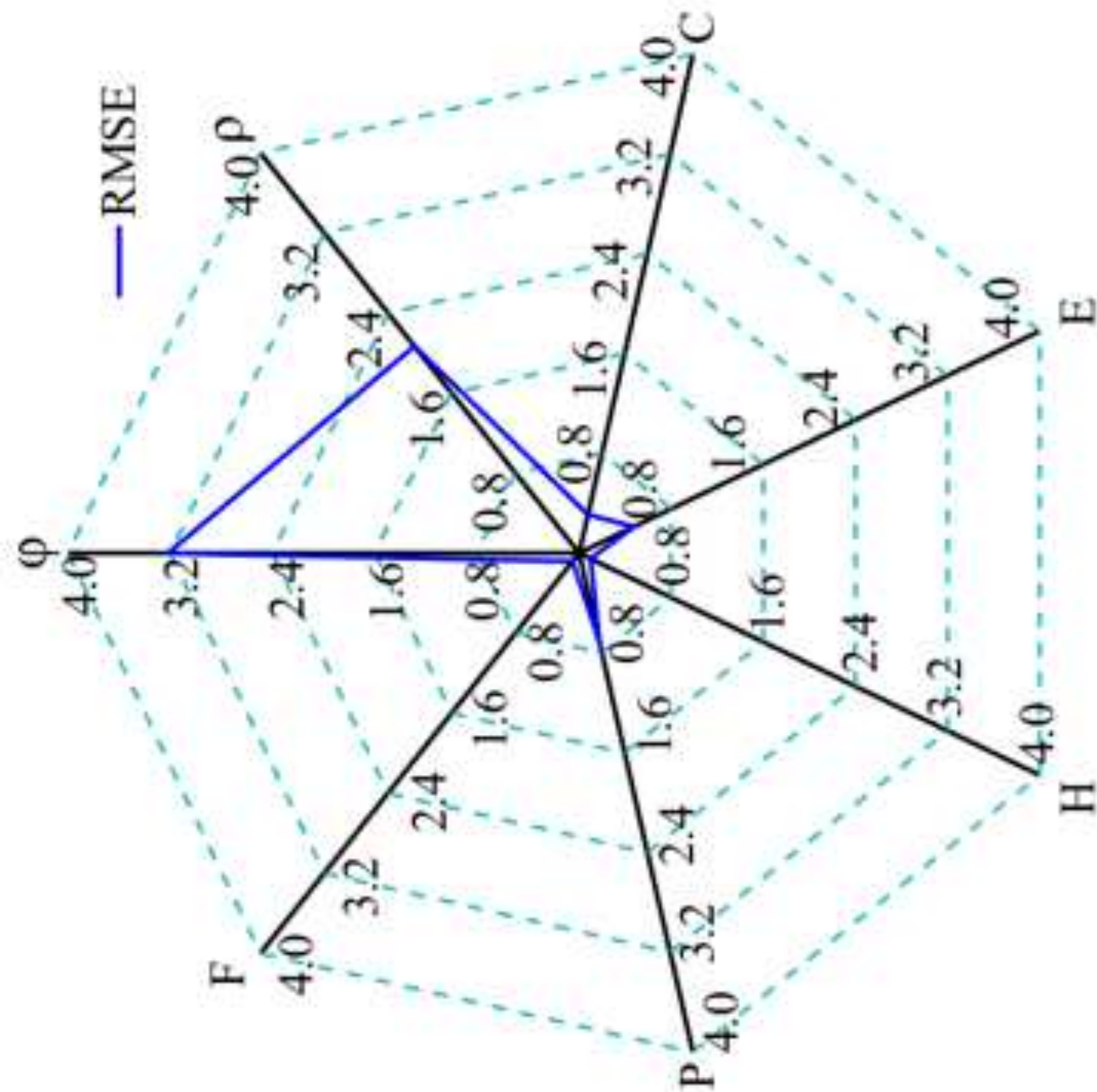












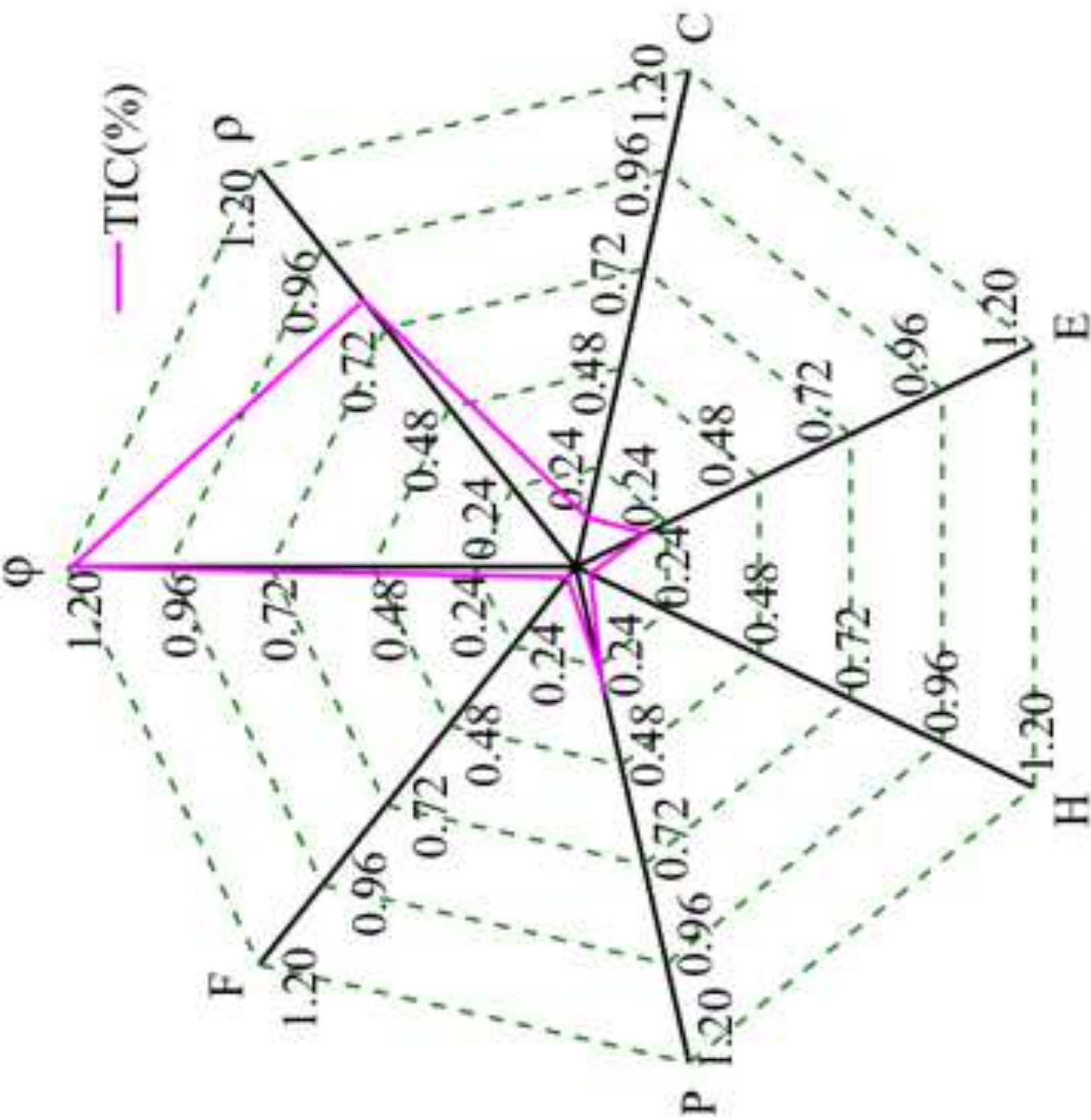


Figure Captions

Fig. 1. Schematic diagram of PCA

Fig. 2. Schematic diagram of RVM algorithm

Fig. 3. A schematic of PSO for solving the global optimization problem

Fig. 4. PCA results and analysis process: (a) PCA results; (b) Contribution rate of each factor and cumulative contribution rate.

Fig. 5. PCA-PSO-RVM model flow diagram

Fig. 6. Different methods for predicting results

Fig. 7. Comparison of MRE, RMSE and TIC in different methods: (a) MRE; (b) RMSE; (c) TIC.

Fig. 8. Sensitivity factor analysis: (a) MRE; (b) RMSE; (c) TIC.