IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 25, 2020, accepted August 10, 2020, date of publication August 14, 2020, date of current version August 25, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3016690

Machine Learning Assisted Prediction Algorithm of Atmospheric Ion Mobility

YONG YI[1], (Member, IEEE), ZHENGYING CHEN[2], AND LIMING WANG[2], (Senior Member, IEEE)

[1]School of Automation, Guangdong University of Technology, Guangzhou 510006, China
[2]Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China

Corresponding author: Yong Yi (yiyongshanxi@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51807103 and Grant 51677101.

ABSTRACT Atmosphere ion mobility is an important electrical parameter related to the filamentary ion flow field of high voltage direct current (HVDC) transmission lines and other characteristics of the glow and streamer corona discharge. Many machine learning (ML)-based algorithms have already been widely used in the prediction of air discharge. This article proposes a genetic algorithm (GA)-support vector regression (SVR) combined with kernel principal components analysis (PCA) to predict the ion mobility, involving dimensionality reduction, feature selection, parameter optimization of SVR. Kernel PCA could reduce the dimension of data and GA with adaptive probability parameters is employed to optimize the parameters of SVR model. An improved parallel-plates ion generator is employed to produce corona discharge and then measure the saturation ion current density and then obtain the training data and testing data of ion mobility. The prediction results show that the proposed algorithm outperforms the other methods in terms of mean relative error and mean squared error criteria. In addition, the parameters of model and data features have a major influence on the performance of the prediction algorithm. Based on the measured data and reference data, the prediction result obtained on ion mobility under different humidity shows a satisfactory generalization and effectiveness of proposed model for the prediction.

INDEX TERMS Ion mobility, support vector regression, electric field, parallel plates, coaxial cylinder, relative humidity, prediction model.

I. INTRODUCTION

Corona discharge occurs on the energized conductors of high voltage direct current (HVDC) power transmission lines. Compared to the narrow ionization region (plasma region) closing to the conductor, the drift region occupying the most inter-electrode gap became more concerned for the ion flow field of unipolar or bipolar lines [1], [2]. In order to calculate the space electric field profiles and ambient electromagnetic environment accurately, complex corona phenomenon have comprehensively been studied by the theoretical model [3]–[5]. In the filamentary ion flow field model, the atmosphere ion mobility is a crucial physical parameter, which refers to an average mobility for the various ionic species in the ion drift region. The atmospheric ion mobility is usually considered as a constant proportionality coefficient in the realm of electro-hydrodynamics and electromagnetism [6]–[8]. In fact, the mobility is a multivariable function and determined by many impacts, including the atmospheric composition, temperature, humidity, atmospheric pressure, atmospheric pollutants, etc.

In the previous literatures, it was well established that ion mobility had been measured under different electric field strength. The point-to-plane electrode had been used to generate the severe non-uniform electric field and the ion mobility in sulfur hexafluoride (SF_6) was obtained by the correlation between the voltage-current curve and the ion mobility [9], [10]. However, whether it can be applied to air at atmospheric pressure needs to be further verified. In addition, the drift tube method was used to measure mobility spectra of atmospheric ions. The temporal decay of the current and the ion drift time were monitored during the test [11], [12]. However, in the earlier studies, the measurement of ion mobility had mainly focused on SF_6 or ion mobility had been measured at low temperature. Ion mobility spectrometry was

The associate editor coordinating the review of this manuscript and approving it for publication was Canbing Li.

the most widely used approach to measure the ion mobility in the atmosphere; meanwhile, the ironic species could be identified [13]–[16]. However, the ion mobility spectrum was often operated at a temperature of hundreds of degrees Celsius which is inconsistent with the actual atmospheric temperature. In addition, parallel plate ion current generator was employed to measure the ion mobility and a calibration method was proposed to calibrate the ion mobility [17], [18]. However, computational prediction research on the atmospheric ion mobility and the effect of humidity impacts appear sparse in literature.

The correlation between the atmosphere conditions and atmospheric ion mobility had been researched by the ion mobility spectrometry. The major environmental impact factors were focused on temperature and atmosphere pressure. The involved charge carriers were organic macromolecules and their derivatizations, dimeric polymer, hydrates, etc. They changed with different atmosphere environment. In addition, the previous measurements of the ion mobility spectrometry were in the range of 100 ° to 300 °, which was higher than normal temperature. The effect of the humidity on the ion mobility was mainly through the ion mobility mass spectrometry in the realm of chemistry. However, due to the water vapor, the spectrum of the ion mobility is dramatically changed and the selectivity and sensibility of the spectrometry apparatus decreased.

In the present paper, a new prediction algorithm of atmospheric ion mobility based on genetic algorithm (GA)-support vector regression (SVR) combined with kernel principal component analysis (PCA) is proposed. The electric field features and charge number density features are extracted and regarded as input of SVR model. The parameters of SVR are optimized by GA. The training and testing data are from experimental method which is that an improved parallel-plates ion current generator is placed in the environment controlled chamber. Finally, through comparing the predictions with the reference data, prediction results obtained on ion mobility show a satisfactory generalization and effectiveness of proposed model for the prediction.

II. PREDICTION ALGORITHM BASED ON GA-SVR COMBINED WITH KERNEL PCA

A. KERNEL PRINCAIPAL COMPONENTS ANALYSIS

For linear case, PCA is suitable for unsupervised dimensionality reduction of unlabeled data, especially for small samples. The kernel PCA is used to preprocess the non-linear and high-order correlation datasets [19]. The idea of kernel PCA is that a non-linear mapping function is introduced to map the raw data to the high-dimensional space. In this high-dimensional space, the linearly inseparable data become linear separable, and then the PCA algorithm can be used. Kernel PCA is a widespread unsupervised feature extraction algorithm and used for dimensionality reduction. The detailed kernel PCA algorithm is presented in Algorithm 1.

Algorithm 1 Kernel PCA

Input: Given the feature set $\boldsymbol{X} = \{x_1, \ldots\ldots, x_N\}$, $x_i \in R^M$, by mapping function Φ, $\Phi(x_i) \in R^F$. $\tilde{\Phi}(x_i) = \Phi(x_i) - \frac{1}{N}\sum_{j=1}^{N}\tilde{\Phi}(x_j)$.The covariance matrix $\boldsymbol{C}$ in F space, $\boldsymbol{C} = \frac{1}{N}\sum_{i=1}^{N}\tilde{\Phi}(x_i)\,\tilde{\Phi}(x_i)^T$.

1. Given kernel function $k(x, y) = \exp\left(-\|x-y\|^2/\gamma^2\right)$, γ is the kernel parameter, and kernel matrix is $\boldsymbol{K} = [\Phi(x_1), \ldots\ldots\Phi(x_N)]^T\,[\Phi(x_1), \ldots\ldots\Phi(x_N)]$
2. Compute new kernel matrix $\tilde{\boldsymbol{K}} = \boldsymbol{K} - \frac{1}{N}\boldsymbol{I}(\boldsymbol{K}) - \frac{1}{N}\boldsymbol{K}\boldsymbol{I} - \frac{1}{N^2}\boldsymbol{I}\boldsymbol{K}\boldsymbol{I}$
3. Compute the eigenvector and eigenvalue $[\boldsymbol{\alpha}, \lambda] = eig\left(\tilde{\boldsymbol{K}}\right)$
4. Vector normalization $\boldsymbol{\alpha}^T\boldsymbol{\alpha} = \frac{1}{\lambda}$
5. Compute the eigenvector $[\boldsymbol{v}] = eig(\boldsymbol{C}) = \sum_{i=1}^{N}\alpha_i\tilde{\Phi}(x_i)$
6. Compute the principle components $y = \left(\boldsymbol{v}, \tilde{\Phi}(x_i)\right) = \sum_{i=1}^{N}\alpha_i\tilde{\Phi}(x_i)^T\,\tilde{\Phi}(y_j)$

Return: The new feature vector $\tilde{\boldsymbol{X}} = \left\{\tilde{x}_1, \ldots\ldots, \tilde{x}_p\right\}$

B. GENETIC ALGORITHM-SUPPORT VECTOR REGRESSION

SVMs are machine learning method for performing different learning tasks, which could be applied for the regression problem. Given a training dataset, $\{(\boldsymbol{x}_1, y_1), \ldots\ldots, (\boldsymbol{x}_N, y_N)\}$, the SVR formulation is [20]

$$min\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right) \tag{1}$$

$$\boldsymbol{s.t.}\ \begin{cases}\boldsymbol{w}^T\psi(\boldsymbol{x}_i) + b - y_i \le \varepsilon + \xi_i, & \xi_i \ge 0\\ y_i - \boldsymbol{w}^T\psi(\boldsymbol{x}_i) - b \le \varepsilon + \xi_i^*, & \xi_i^* \ge 0,\ i = 1, \ldots, N\end{cases} \tag{2}$$

where C is a regularization parameter which represents the trade-off of large margin and noise tolerance; ξ donates the slack variable; Feature vector $\boldsymbol{x}_i$ is mapped into a higher dimensional space by function $\psi(\boldsymbol{x}_i)$; $\boldsymbol{w}$ is the vector variable.; ε is the insensitive loss function. The solution of SVR is quadratic programming problem. Solving SVR can be changed to solving dual problem by Lagrange multipliers and Karush-Kuhn-Tucker (KKT) conditions. Dual formulation of SVR [21] is

$$min\ \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\left(\alpha_n - \alpha_n^*\right)\left(\alpha_m - \alpha_m^*\right)K(\boldsymbol{x}_m, \boldsymbol{x}_n) + \sum_{n=1}^{N}\left((\varepsilon - y_n)\,\alpha_n\right)\left((\varepsilon + y_n)\,\alpha_n^*\right) \tag{3}$$

$$\boldsymbol{s.t.}\ \sum_{n=1}^{N}\left(\alpha_n - \alpha_n^*\right) = 0; \quad 0 \le \alpha_n, \alpha_n^* \le C \tag{4}$$

In the paper, Gaussian kernel is used due to its high flexibility and high dimensional space mapping of data. Gaussian kernel takes the form

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{\sigma^2}\right) \tag{5}$$

where $K(\boldsymbol{x}_m, \boldsymbol{x}_n) = \psi(\boldsymbol{x}_m)^T \psi(\boldsymbol{x}_n)$ is the kernel function; α is the Lagrange multiplier. The high efficiency SVR is depending on the selection of kernel function and model parameters. σ is the tuning kernel parameter.

The parameters (C, σ) must be specified to train the model. In order to train SVR model, parameter (C, σ) should be specified. The appropriateness of the parameters directly affects the quality of the SVR model. The genetic algorithm is employed to optimize the cross-validation accuracy. GA is a search algorithm of parallel computing and global optimization and without derivative, which is widely used in many fields. In the current study, the crossover probability P_c and the mutation probability P_m are constant values in the evolution period of population, which may delay the convergence of model and cause the long training time. Therefore, the adaptive p_c and p_m of GA is given by the fitness value f. The maximum crossover and mutation probability are P_{cmax}, P_{mmax}. The minimum crossover and mutation probability are P_{cmin}, P_{mmin}. The detailed GA is presented in Algorithm 2.

Algorithm 2 GA for Parameter Optimization

Input: Given the parameter range $C \in [0, 1000]$, $\sigma \in [0, 64]$, $\varepsilon \in [0.0001, 0.1]$, $P_c \in [0.4, 0.99]$, $P_m \in [0.0001, 0.1]$. L is the total evolutionary algebra, l is current population evolution algebra. Size of population is 400.

1. Chromosomes initialization $\boldsymbol{Z}_0 = \{C_0, \sigma_0, \varepsilon_0\}$
2. Compute the fitness of chromosomes by fitness function and k-fold cross validation
   $MSE_{cv} = \frac{1}{N}\sum_{i=1}^{N}(z_i - \tilde{z}_i)^2$
3. Stochastic universal sampling is used to select chromosome
4. Roulette wheel selection algorithm
5. Two point crossover and create new offspring. The probability of crossover is computed by [22]
   $P_c = \frac{\left(\left(\frac{L-l}{L}P_{cmax} + \frac{l}{L}P_{cmin}\right) + \frac{P_{cmax}}{f_{max}}f_{min}\right)}{2}$
6. Simple mutation and the probability of mutation is computed by [22]
   $P_m = \frac{L-l}{L}P_{mmax} + \frac{l}{L}P_{mmin}$
7. Stop criterion. If the number of epochs equals 150 or $f_{max}^{l} - f_{max}^{l-1} < 0.001$, then the optimal chromosome is presented as a solution; otherwise go back to Step 2

Return: The new chromosome $\tilde{\boldsymbol{Z}} = \{C, \sigma\}$

Algorithm 3 Prediction Method

Input: Ion mobility (training data and testing data) $\boldsymbol{\mu} = \{\mu_1, \ldots, \mu_N\}$, $x_i \in R^M$, by mapping function Φ, $\Phi(x_i) \in R^F$.

Begin training model

1. Compute electric field distribution of training data
2. **For** $(C_0, \sigma_0, \varepsilon_0)$ **do**
   compute parameters of SVR by genetic algorithm
   end for
3. Features extraction through kernel PCA
4. Compute SVR dual problem and then output the approximate function and prediction value $\tilde{\mu}$
   Testing the SVR model
5. Compute electric field distribution of testing data
6. Features extraction through kernel PCA
7. Compute ion mobility through SVR model

Return: The prediction ion mobility $\tilde{\boldsymbol{\mu}} = \{\tilde{\mu}_1, \ldots, \tilde{\mu}_N\}$

C. GA-SVR COMBINED WITH KERNEL PCA ALGORITHM FOR ION MOBILITY PREDICTION

The architecture of the prediction algorithm based on GA-SVR and kernel PCA is presented in Algorithm 3. The feature extraction, weighting parameter and SVR parameter optimization are proposed in the paper. The prediction algorithm first computes electric field and number density features of coaxial cylinder electrode model and optimizes the parameters of SVR by genetic algorithm, which has adaptive crossover probability and mutation probability depending on fitness value and population evolution algebra. And then, kernel PCA technique is used to extract the features and SVR dual problem is solved by quadratic problems with linear constraint. Finally, according to the optimal parameters and feature subset, the prediction model is used in the testing dataset.

III. MACHINE LEARNING FEATURES FROM ELECTRIC FIELD AND SPACE CHARGE NUMBER DENSITY OF COAXIAL CYLINDRICAL ELECTRODE

A. CONFIGURATION OF COAXIAL CYLINDRICAL ELECTRODE

The prediction algorithm of ion mobility is primarily based on GA-SVR. The feature vectors (or called descriptors) of data are generated by calculation. The coaxial cylinder electrode model is used to calculate the electric field features and space charge features. The features of data include electric field features and space charge features of coaxial cylinder electrode model. These features are closely related to the atmospheric ion mobility. The diameter of the inner conductor is r_0 and the diameter of the outer electrode is r_R, which is grounded.

B. CALCULATION OF ELECTRIC FIELD FEATURES AND SPACE CHARGE FREATURES FROM HYDRODYNAMIC MODEL OF STEADY-STATE

In the positive and negative corona discharge, the hydrodynamic models for the steady-state positive and negative

corona in one dimension symmetry cylindrical are defined by following equations.

Continuity equations of the electrons, the positive ions and the negative ions [23]–[25] are given by

$$\frac{d\left(rn_{\mathrm{e}}u_{\mathrm{e}}\right)}{rdr}-\frac{d\left(rD\frac{dn_{\mathrm{e}}}{dr}\right)}{rdr}=\mp\left[\left(\alpha-\eta\right)n_{\mathrm{e}}u_{\mathrm{e}}-n_{\mathrm{e}}n_{\mathrm{p}}\beta+S\right] \tag{6}$$

$$\frac{d\left(rn_{\mathrm{p}}u_{\mathrm{p}}\right)}{rdr}=\pm\left(\alpha n_{\mathrm{e}}u_{\mathrm{e}}-\beta n_{\mathrm{e}}u_{\mathrm{p}}-\beta n_{\mathrm{p}}u_{\mathrm{n}}+S\right) \tag{7}$$

$$\frac{d\left(rn_{\mathrm{n}}u_{\mathrm{n}}\right)}{rdr}=\mp\left(\eta n_{\mathrm{e}}u_{\mathrm{e}}-\beta n_{\mathrm{p}}u_{\mathrm{n}}\right) \tag{8}$$

Poisson equation is given by

$$\frac{d}{rdr}\left(r\frac{d\varphi}{dr}\right)=\pm\frac{e\left(n_{\mathrm{p}}-n_{\mathrm{n}}-n_{\mathrm{e}}\right)}{\varepsilon} \tag{9}$$

where n_{p}, n_{n} and n_{e} are the positive ion, negative ion and electron number density, respectively, cm^{-3}; u_{p}, u_{n} and u_{e} are the positive ion, negative ion and electron velocity, respectively, cm/s, α, η are the ionization and attachment coefficient, respectively, cm^{-1}; β is recombination coefficient, cm^3/s; D_{e} is diffusion coefficient, cm^2/s; E is the electric field, V/cm; ε is the permittivity of air; φ is the potential. S is the photoionization term. The above symbol of $\mp$ and $\pm$ represent positive polarity. The electron ionization and attachment coefficient are determined by the reduced electric field and the expressed as [26]–[28].

The boundary conditions include the conductor surface, plasma boundary and outer electrode boundary. The positive or negative ion density on the conductor surface is considered to be zero. The positive or negative ion density on the outer plasma boundary is assumed to be zero for the negative and positive corona mode. The corona current is given by

$$I=I_{\mathrm{p}}+I_{\mathrm{n}}+I_{\mathrm{e}} \tag{10}$$

In the process of solving the coaxial cylinder electrode model, the Runge–Kutta technique and the central finite difference method are used to calculate the equations. The relative error of the potential gradient and space charge density are the convergent conditions in the numerical iterative calculation. In the process of calculation, the ion mobility is considered to be constant and independent of the electric field, which is the same as the previous hypotheses used in the calculation of ion flow field.

C. ELECTRIC FIELD AND SPACE CHARGE FREATURES EXTRACTION

In the gap of coaxial cylinder electrode, forty-one features are defined and the feature set of electric field and space charge as shown in TABLE 1.

The electric field features are defined as follows: Maximum value of electric field: E_{m}; Average value of electric field: E_{a};Electric field distortion: $E_{\mathrm{d}}=\left(E_{\mathrm{m}}-E_{\mathrm{a}}\right)/E_{\mathrm{a}}$; Maximum value of electric field gradient: E'_{m}; Average value of electric field gradient: E'_{a}. The second subscript p or i represent plasma region or ion flow region. $E_{\mathrm{x\%}}$ represents the ratio of area to total area which exceed x% of the maximum electric field.

TABLE 1. Feature set of electric field and space charge.

Physical quantities	Features
Electric field	E_{mp} E_{mi} E_{ap} E_{ai} $E_{90\%}$ $E_{70\%}$ $E_{50\%}$ $E_{30\%}$ $E_{10\%}$ E_{dp} E_{di} E'_{mp} E'_{mi} E'_{ap} E'_{ai}
Positive ion density	n_{pmp} n_{pmi} n_{pap} n_{pai}
Negative ion density	n_{nmp} n_{nmi} n_{nap} n_{nai}
Electron density	n_{emp} n_{emi} n_{eap} n_{eai}
Current density	J_{pmp} J_{pmi} J_{pap} J_{pai} J_{nmp} J_{nmi} J_{nap} J_{nai} J_{emp} J_{emi} J_{eap} J_{eai} J_{mtotal} J_{atotal}

The space charge features are defined as follows: Maximum and average positive ion density: n_{pm} and n_{pa}; Maximum and average negative ion density: n_{nm} and n_{na}; Maximum and average electron density: n_{em} and n_{ea}. The third subscripts p or i represents plasma region or ion flow region.

The current density features are defined as follows: Maximum and average positive current density: J_{pm} and J_{pa}; Maximum and average negative current density: J_{nm} and J_{na}; Maximum and average electron current density: J_{em} and J_{ea}; Maximum and average total current density: J_{mtotal} and J_{atotal}; The third subscripts p or i represents plasma region or ion flow region.

IV. PREDICTION OF ION MOBILITY

The training data and testing data of ion mobility are sourced from the experimental measurement. The ion mobility is measured by parallel-plate system.

A. STRUCTURE OF PARALLEL-PLATE SYSTEM

The schematic view of a parallel-plate system, which is recommended to produce uniform field with space charges and calibrate the electric field strength meter in IEEE Standard [29], is shown in Fig. 1 and Fig. 2. The apparatus used in this article is a circular coaxial geometry and different from the previous apparatus [18], which is a square coaxial configuration and avoids the inhomogeneous electric field effect. The space electric field of circular geometry is more uniform and isotropic. The apparatus has five layers, from the top to down, corona ring, corona wires layer, first screen, top plate (second screen) and bottom plate (grounded plate). The corona wires layer consists of the equidistant lined up copper wires and the interval of the wires is 5 cm. The diameter of the copper wire is 1 mm. The first screen is made by the square grids 0.5 × 0.5 cm with 0.2 mm-diameter copper wires and

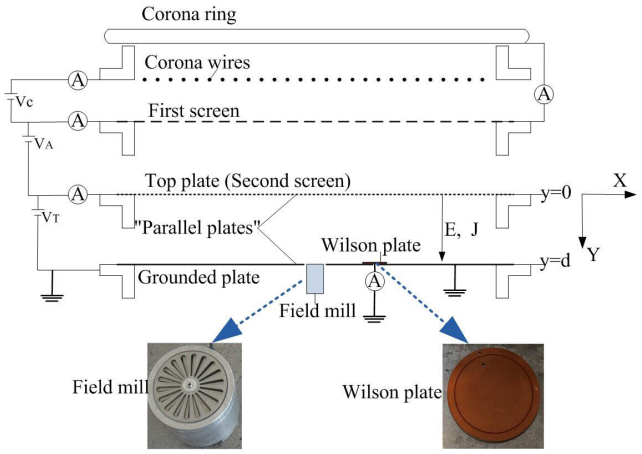


FIGURE 1. Schematic of parallel plate apparatus.

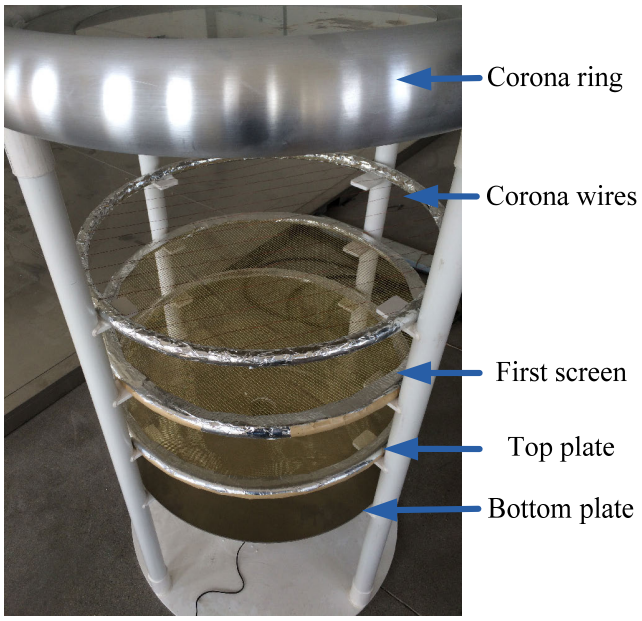


FIGURE 2. Picture of parallel plate apparatus.

fixed on the 1 m-diameter copper ring. The top plate (second screen) comprises square grids 1 × 1 mm with 0.1 mm-diameter copper wires and is also fixed on the 1 m-diameter copper ring. The Wilson plate is placed on the grounded metal plate with 7.9-cm-diameter current sense area and 1-cm-wide guard band. An electrometer is mounted between the Wilson plate and the ground for measuring the ion current. A hole is in the center of the grounded plate for placing the field mill.

The corona ring is employed to uniform space electric field and used to suppress the fringing effect. The corona wires layer has copper wires, which is applied voltage to generate corona discharge, and the corona-originated ions move downward to the first screen. The ions pass through the first screen and then entered into the region of the parallel plates, which is similar to the two layer metal parallel-plates electrode.

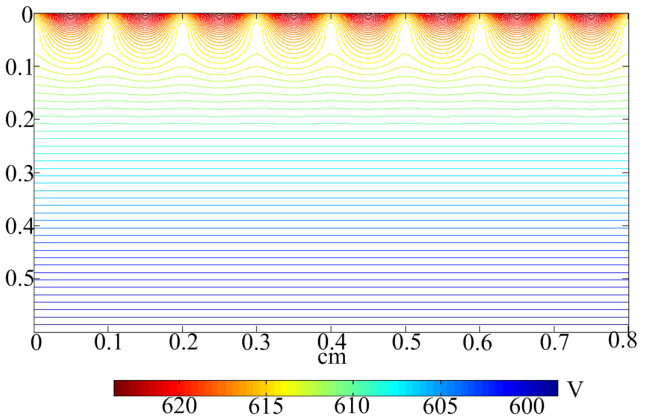


FIGURE 3. Potential distribution between the top plate and the bottom plate.

The second screen and the metal grounded plate generate an approximately uniform electric field. The ions are driven by the field. With the increase of the controlled voltage (V_T), the ion current density becomes saturation gradually and the space ions gathered at the second screen. Then, the electric field at the second screen becomes zero due to the space charge with same polarity and the ion mobility is obtained under the saturation condition.

Fig. 3 shows the potential distribution of the parallel plates. The potential of along Y axis become uniform gradually. The effect of top plate grids on the potential of the most space between the parallel plates was rather limited. So, the one-dimensional model can be used and only vertical component need to be considered.

B. TRAINING DATA AND TESTING DATA

The experiment has been carried out in the electromagnetic environment laboratory, which is relatively isolated from the outside environment. The temperature and atmosphere pressure is constant. The parallel-plates ion generator is placed in the sealed PMMA (polymethyl methacrylate) chamber with moisture content adjustable. The relative humidity (RH) is employed to describe the moist degree of the certain temperature and more intuitive than absolute humidity. The positive or negative ion mobility has been obtained with the applied voltage between ±200V and ±3000V.

TABLE 2 shows a small portion of the training data and testing data of ion mobility. TABLE 3 shows a small portion of the training data and testing data of ion mobility under different relative humidity (RH).

After the model training is completed, the prediction function can be applied to predict the value of testing data. The prediction results can be evaluated by the following error analysis method.

Mean squared error (MSE)

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(f\left(\boldsymbol{x}_i\right) - y_i\right)^2 \tag{11}$$

TABLE 2. Training data and testing data of ion mobility.

	P=101.3kPa	T=20℃	RH%=36%	
	Training data d=10cm		Testing data d=9.5cm	
J_s 10^{-6}(A/m^2)	V_T (V)	μ 10^{-4}(m^2/V·s)	V_T' (V)	μ′ 10^{-4}(m^2/V·s)
Positive				
0.129	325.1	1.28	314.1	1.33
0.510	622.1	1.34	613.2	1.33
1.131	920.5	1.34	912.3	1.33
1.961	1220.1	1.33	1211.5	1.32
3.050	1520.5	1.31	1510.6	1.32
4.215	1811.3	1.29	1809.7	1.27
5.551	2110.0	1.28	2108.8	1.27
Negative				
0.192	-345.5	1.68	-332.5	1.71
0.702	-645.1	1.69	-631.7	1.73
1.518	-950.1	1.70	-930.8	1.72
2.621	-1255.0	1.71	-1229.9	1.71
4.055	-1505.5	1.71	-1529.1	1.71
4.999	-1845.3	1.52	-1828.2	1.47
6.644	-2145.8	1.48	-2127.3	1.45

TABLE 3. Training data and testing data of ion mobility under different humidity.

	T=20℃					
	Training data (10^{-4}m^2/V·s)			Testing data (10^{-4}m^2/V·s)		
RH (%)	Pressure(kPa)			Pressure(kPa)		
	99.5	100	101	98.5	99.5	102
Positive		μ			μ′	
14.15	1.70	1.68	1.67	1.75	1.73	1.65
28.55	1.41	1.40	1.38	1.44	1.42	1.37
36.70	1.28	1.27	1.26	1.32	1.30	1.25
44.38	1.19	1.17	1.17	1.22	1.21	1.16
56.22	1.09	1.08	1.07	1.12	1.11	1.07
67.03	1.03	1.02	1.02	1.06	1.04	1.01
75.89	1.01	0.98	0.91	1.04	1.02	9.84
84.33	0.98	0.97	0.96	1.01	0.99	9.63
96.51	0.95	0.94	0.93	0.97	0.96	9.31
Negative						
14.15	2.12	2.12	2.08	2.19	2.15	2.06
28.55	1.78	1.77	1.76	1.85	1.82	1.75
36.70	1.64	1.63	1.62	1.69	1.67	1.60
44.38	1.53	1.52	1.51	1.57	1.55	1.49
56.22	1.40	1.38	1.37	1.43	1.41	1.36
67.03	1.32	1.30	1.29	1.34	1.33	1.28
75.89	1.28	1.26	1.25	1.30	1.29	1.24
84.33	1.27	1.24	1.23	1.28	1.26	1.22
96.51	1.26	1.23	1.22	1.26	1.25	1.21

Squared correlation coefficient

$$r^2 = \frac{\left(N\sum_{i=1}^{N} f(x_i)\, y_i - \sum_{i=1}^{N} f(x_i) \sum_{i=1}^{N} y_i\right)^2}{\left(N\sum_{i=1}^{N} f(x_i)^2 - \left(\sum_{i=1}^{N} f(x_i)\right)^2\right)\left(N\sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i\right)^2\right)} \tag{12}$$

When $0 < |r| < 1$, there is a certain degree of linear correlation between the two variables. And the closer the $|r|$ approaches 1, the closer the linear relationship between the two variables is.

Mean relative error (MRE)

$$\mathrm{MRE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{f(\boldsymbol{x}_i) - y_i}{y_i}\right| \tag{13}$$

where $f(\boldsymbol{x}_i)$ is the prediction and y_i is the measurement.

C. EXPERIMENTAL RESULTS AND DISCUSSION

TABLE 4 shows the optimized parameters and performance between GA-SVR-Kernel PCA and other models. The parameters $\{C, \sigma\}$ of SVR model are optimized by grid search method (SG), genetic algorithm (GA) with constant probability parameters and genetic algorithm (GA) with adaptive probability parameters. The performance of GS-SVR, least squares-SVR and GA-SVR are measured by MSE, MRE and squared correlation coefficient r. Compared with other models, due to the introduction of dimensionality reduction and parameter optimization techniques, the time consuming of GA-SVR-kernel PCA is higher than other models and increases by approximately 50% to 300%. When comparing the various parameter optimization algorithms of SVR models, MSE and MRE of GA-SVR is lower and r of GA-SVR is higher. Compared with GA-SVR with constant probability parameters, due to the introduction of adaptive probability parameters, MRE of GA-SVR decreases by 27.8%. For the above four models, the trend of MSE and MRE are decreasing, which indicates that the parameter optimization and dimensionality reduction has a significant influence on the performance of models. The MRE of GA-SVR-kernel PCA is lower than that of other models. It can be observed that the generalization performance of GA-SVR-kernel PCA is better than that of other models. While the feature is optimized and reduced from 41 dimensions to 11 dimensions by kernel PCA, MRE and MSE value decreases gradually and then increases. By comparing MRE and MSE value, GA-SVR model with 20 dimensions has a satisfactory accuracy.

V. PREDICTION OF ION MOBILITY UNDER DIFFERENT HUMIDITY

In order to validate the effectiveness of the proposed model, the ion mobility of different humidity and air pressure is predicted and compared with data in the reference. Fig. 4 and Fig. 5 show the comparisons of measured and predicted of positive and negative ion mobility under different relative humidity. These two pictures illustrate the prediction results intuitively. The GA-SVR-kernel PCA model with Gauss kernel function coefficient $\sigma = 0.9003$ and penalty coefficient $C = 88.7001$ for positive polarity, and with Gauss kernel function coefficient $\sigma = 0.6113$ and penalty coefficient $C = 68.8821$ for negative polarity are applied to predict the value under different relative humidity at 20°. The measured data of positive and negative ion mobility are from reference [30]. The predicted values are consistent with the measured values. Compared with measured data, the maximum deviation is 0.4015 m^2/Vs for positive ion mobility and 0.4231 m^2/Vs for negative ion mobility.

TABLE 4. Comparisons of optimized parameters and performance between ga-svr-kernel pca and other models.

Models \ Parameters	Positive						Negative					
	C	σ	MSE	r	t	MRE	C	σ	MSE	r	t	MRE
GS-SVR (41d)	413.11	0.335	0.588	0.77	19	8.22%	337.15	0.554	0.501	0.71	19	9.10%
LS-SVR (41d)	331.19	0.412	0.531	0.72	18	4.51%	309.32	0.661	0.530	0.62	18	4.82%
GA-SVR (41d)	109.12	0.334	0.526	0.81	22	4.88%	129.06	0.543	0.511	0.78	23	4.66%
(P_c P_m is constants)												
GA-SVR (41d)	167.89	1.398	0.483	0.83	42	3.52%	103.34	1.013	0.463	0.81	42	3.70%
(with adaptive P_c P_m)												
GA-SVR- kernel PCA(39d)	110.22	0.729	0.479	0.87	67	4.33%	133.89	1.122	0.459	0.83	66	3.61%
GA-SVR--kernel PCA(33d)	130.68	0.848	0.433	0.89	68	3.99%	99.64	0.884	0.421	0.88	68	3.23%
GA-SVR-kernel PCA (20d)	77.28	0.883	0.419	0.91	68	3.45%	63.22	0.692	0.403	0.92	68	3.11%
GA-SVR-kernel PCA (11d)	83.99	0.922	0.442	0.88	67	4.15%	83.09	1.003	0.441	0.86	68	3.44%

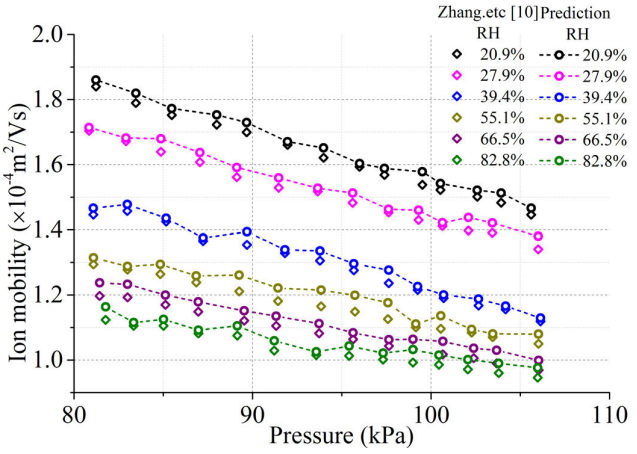


FIGURE 4. Comparisons of the measurements and predictions of positive ion mobility under different humidity.

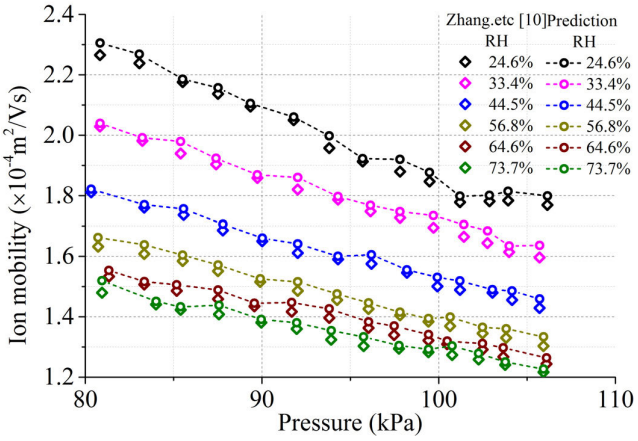


FIGURE 5. Comparisons of the measurements and predictions of negative ion mobility under different humidity.

For the positive ion mobility, the maximum mean relative error of different RH (20.9%-82.8%) is 2.7661%, 2.984%, 3.235%, 4.441%, 4.041% and 4.031%, respectively, as show in Fig. 6. For the negative ion mobility, the maximum mean relative error of different RH (24.6%-73.7%) is 2.128%, 2.433%, 3.199%, 3.221%, 3.098% and 2.877%, respectively.

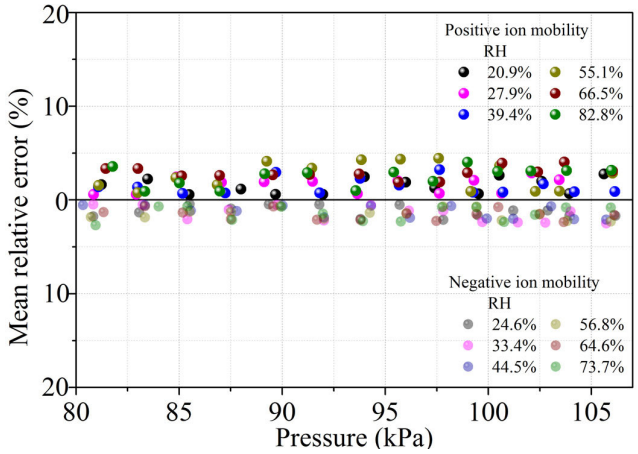


FIGURE 6. Mean relative error of the measurements and predictions of negative and positive ion mobility under different humidity.

It is observed that prediction model has a high accuracy in the ion mobility estimation and effectiveness of GA-SVR combined with kernel PCA and numerical stability. The prediction results show that the feature extraction and parameter selection are appropriate. The prediction model presents a satisfactory generalization ability.

VI. CONCLUSION

In this article, an ion mobility prediction algorithm based genetic algorithm (GA) - support vector regression (SVR) combined with kernel PCA is proposed. The electric field features and space charge features of coaxial cylinder electrode model are extracted as the input of model and complicated physical process need not be considered. Kernel-PCA technique is used for dimensionality reduction. In addition, GA with adaptive probability parameters is an effective strategy and used to search the optimal penalty parameter and kernel parameter of prediction model.

An improved parallel-plates ion current generator has been developed for measuring the atmosphere ion mobility under different humidity. The prediction and measurement result show that proposed model offers higher prediction accuracy for ion mobility prediction. Besides, comparisons between

prediction values and the publication data, the proposed prediction model offers a better performance. MRE of prediction values is below 4.441% and 3.221% for positive and negative ion mobility. The prediction results obtained on ion mobility show the satisfactory generalization and effectiveness.

REFERENCES

[1] P. S. Maruvada, *Corona in Transmission Systems: Theory, Design and Performance*. Johannesburg, South Africa: Eskom Holdings, 2011, pp. 10–80.

[2] P. S. Maruvada, *Corona Performance of High Voltage Transmission Lines*. London, U.K.: Taylor & Francis, 2000, pp. 5–55.

[3] M. Sarma and W. Janischewskyj, "Analysis of corona losses on DC transmission lines part II–Bipolar lines," *IEEE Trans. Power App. Syst.*, vol. PAS-88, no. 10, pp. 1476–1491, Oct. 1969.

[4] T. Takuma and T. Kawamoto, "A very stable calculation method for ion flow field of HVDC transmission lines," *IEEE Trans. Power Del.*, vol. PWRD-2, no. 1, pp. 189–198, Jan. 1987.

[5] M. Abdel-Salam and Z. Al-Hamouz, "A finite-element analysis of bipolar ionized field," *IEEE Trans. Ind. Appl.*, vol. 31, no. 3, pp. 477–483, May 1995.

[6] N. Farnoosh, K. Adamiak, and G. S. Castle, "Numerical calculations of submicron particle removal in a spike-plate electrostatic precipitator," *IEEE Trans. Dielectrics Electr. Insul.*, vol. 18, no. 5, pp. 1439–1452, Oct. 2011.

[7] B. Zhang, W. Li, J. He, and R. Zeng, "Analysis of ion flow field of UHV/EHV AC transmission lines," *IEEE Trans. Dielectrics Electr. Insul.*, vol. 20, no. 2, pp. 496–504, Apr. 2013.

[8] V. A. Chirkov, Y. K. Stishkov, and S. A. Vasilkov, "Characteristics of electrohydrodynamic pump of the dissociation type: Low and high voltage ranges," *IEEE Trans. Dielectrics Electr. Insul.*, vol. 22, no. 5, pp. 2709–2717, Oct. 2015.

[9] A. G. Arson and I. M. Bortnik, "Mobility of ions in SF6," in *Proc. 6th Int. Conf. Gas Discharges*, 1980, pp. 165–167.

[10] R. T. Waters, O. Farish, and O. Ibrahim, "Positive and negative mean ion mobilities in corona discharges in SF6 and mixtures," in *Proc. 7th Int. Conf. Gas Discharges Their Appl.*, 1982, pp. 251–253.

[11] R. G. Stearns, "Ion mobility measurements in a positive corona discharge," *J. Appl. Phys.*, vol. 67, no. 6, pp. 2789–2799, Mar. 1990.

[12] Y. Liu, S. Huang, and L. Zhu, "Influence of humidity and air pressure on the ion mobility based on drift tube method," *CSEE J. Power Energy Syst.*, vol. 1, no. 3, pp. 37–41, Sep. 2015.

[13] G. A. Eiceman, Z. Karpas, and H. H. Hill, Jr., *Ion Mobility Spectrometry*. Boca Raton, FL, USA: CRC Press, 2013, pp. 25–60.

[14] H. H. Hill, Jr., W. F. Siems, and R. H. St. Louis, "Ion mobility spectrometry," *Anal. Chem.*, vol. 62, no. 23, pp. 1201A–1209A, Dec. 1990.

[15] D. C. Collins and M. L. Lee, "Developments in ion mobility spectrometry-mass spectrometry," *Anal. Bioanal. Chem.*, vol. 372, no. 1, pp. 66–73, Jan. 2001.

[16] A. B. Kanu, P. Dwivedi, M. Tam, L. Matz, and H. H. Hill, Jr., "Ion mobility-mass spectrometry," *J. Mass. Spectrometry*, vol. 43, no. 1, pp. 1–22, 2008.

[17] R. S. Withers, J. R. Melcher, and J. W. Richmann, "Charging, migration and electrohydrodynamic transport of aerosols," *J. Electrostat.*, vol. 5, pp. 225–239, Sep. 1978.

[18] M. Misakian, "Generation and measurement of DC electric fields with space charge," *J. Appl. Phys.*, vol. 52, no. 5, pp. 3135–3144, May 1981.

[19] B. Scholkopf, S. Mika, C. J. C. Burges, P. Knirsch, K.-R. Muller, G. Ratsch, and A. J. Smola, "Input space versus feature space in kernel-based methods," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1000–1017, Sep. 1999.

[20] C. Cortes and V. Vapnik, "Support-vector network," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[21] V. Vapnik, *Statistical Learning Theory*. New York, NY, USA: Wiley, 1998.

[22] P. Tao, Z. Sun, and Z. Sun, "An improved intrusion detection algorithm based on GA and SVM," *IEEE Access*, vol. 6, no. 5, pp. 13624–13631, Mar. 2018.

[23] J. H. Chen and J. H. Davidson, "Electron density and energy distributions in the positive DC corona: Interpretation for corona-enhanced chemical reactions," *Plasma Chem. Plasma Process.*, vol. 22, no. 2, pp. 199–224, Jun. 2002.

[24] J. H. Chen and J. H. Davidson, "Model of the negative DC corona plasma: Comparison to the positive DC corona plasma," *Plasma Chem. Plasma Process.*, vol. 23, no. 1, pp. 199–224, Mar. 2003.

[25] Y. Yi, L. Wang, and Z. Chen, "A comprehensive numerical model of steady-state negative corona for coaxial cylindrical electrode in atmosphere air," *IEEE Trans. Plasma Sci.*, vol. 45, no. 2, pp. 200–207, Feb. 2017.

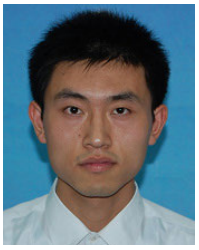
[26] M. Abdel-Salam, "Positive wire-to-plane coronas as influenced by atmospheric humidity," *IEEE Trans. Ind. Appl.*, vol. IA-21, no. 1, pp. 35–40, Jan. 1985.

[27] G. E. Georghiou, A. P. Papadakis, R. Morrow, and A. C. Metaxas, "Numerical modelling of atmospheric pressure gas discharges leading to plasma production," *J. Phys. D, Appl. Phys.*, vol. 38, pp. 1377–1382, Sep. 2012.

[28] L. Fouad and S. Elhazek, "Effect of humidity on positive corona discharge in a three electrode system," *J. Electrostat.*, vol. 35, no. 1, pp. 21–30, Jul. 1995.

[29] *IEEE Guide for the Measurement of DC Electric-Field Strength and Ion Related Quantities*, IEEE Standard 1227-1990(R2010), 2010.

[30] B. Zhang, J. He, and Y. Ji, "Dependence of the average mobility of ions in air with pressure and humidity," *IEEE Trans. Dielectrics Electr. Insul.*, vol. 24, no. 2, pp. 923–929, Apr. 2017.



YONG YI (Member, IEEE) was born in Shanxi, China. He received the Ph.D. degree from the Department of Electrical Engineering, Tsinghua University, in 2017. He is currently working as a Postdoctoral Researcher at the Tsinghua–Berkeley Shenzhen Institute (TBSI). His research interests include electromagnetic environment and corona discharges, streamer simulation, and dielectric materials.

ZHENGYING CHEN was born in Fujian, China, in 1993. He received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2015. His research interests include electromagnetic environment and corona discharges of HVDC systems.

LIMING WANG (Senior Member, IEEE) was born in Zhejiang, China, in November 1963. He received the B.S., M.S., and Ph.D. degrees in high voltage engineering from the Department of Electrical Engineering, Tsinghua University, Beijing, China, in 1987, 1990, and 1993, respectively. He has been working at Tsinghua University, since 1993. His major research interests include high-voltage insulation and electrical discharges, flashover mechanisms on contaminated insulators, and applications of pulsed electric fields.