MULTIPLE-INPUT MULTIPLE-OUTPUT SOFT SENSORS BASED ON KPCA AND MKLS-SVM FOR QUALITY PREDICTION IN ATMOSPHERIC DISTILLATION COLUMN

Qi Li[1], Qun Du[1], Wei Ba[2] and Cheng Shao[1]

[1]School of Control Science and Engineering
Dalian University of Technology
No. 2, Linggong Road, Ganjingzi District, Dalian 116024, P. R. China
{ qili; qundu; cshao }@dlut.edu.cn

[2]Dalian Scientific Test and Control Technology Institute
Dalian 116013, P. R. China
baweidut@gmail.com

ABSTRACT. *In this paper a method based on kernel principal component analysis (KPCA) and mixed kernel least square support vector machine regression (MKLS-SVM) for online quality prediction in atmospheric distillation column is presented. Firstly, the KPCA is employed to reduce the input vector's dimensions of the multiple-input multiple-output (MIMO) soft sensor and created the data set which required training the MKLS-SVM. Then, considering that the characteristics of kernels have great impacts on learning and predictive results of LS-SVM, LS-SVM based on mixed polynomial kernel and RBF kernel is adopted to build the soft sensor model. The parameters of the MKLS-SVM are adaptively selected by the real-cord multi-population genetic algorithm (GA) with elitist strategy, migration operator, self-adaptive mutation and crossover operator. The modeling process is described with emphasis on data preprocessing and variables selection. Finally, the simulation results show that the MIMO soft sensors have good abilities of model generalization and the predicted values are in good agreement with lab measurements.*

Keywords: Kernel principal component analysis, Least square support vector machine, Genetic algorithm, Distillation column, Soft sensors

1. **Introduction.** The purpose of the industrial distillation process in refinery is to produce qualified products, and then the quality control of products becomes the core problem. In addition, in order to improve economic benefit of the industrial distillation process, the important process variables related to product quality should be strictly controlled. However, in most industrial distillation process, hardware sensors of such quality variables are not available. For example, dry point and flash point are very important quality values for the aviation kerosene which is one of the important products in the crude oil distillation units (CDU). The online analyzer and the lab measurement are available for measuring the aviation kerosene's quality. However, the online analyzer usually has significant time lag, high investment and maintenance costs. The test in lab is even slower and less frequent. The best way to solve this problem is building the composition estimators [1-3] or soft sensors [4]. The soft sensor has some merits like rapid response, continuously providing the information of the quality variables, low investment and maintenance costs.

Many approaches have been proposed to build the soft sensor for quality prediction online. We may summarize these methods into two categories. One method is to build the soft sensors from the mechanism model of the distillation column, such as equilibrium

stage model and nonequilibrium stage model of distillation column [5]. However, it is often difficult in refineries, due to the complexity of industrial distillation processes. Physical modeling can be very time-consuming and significant parameters are generally unknown. The other method is adopting empirical model or the black-box model for the soft sensors. There are many algorithms which have been proposed to build the soft sensor in this way, including using multivariate regression analysis [2,6,7], support vector machine (SVM) regression [8,9], artificial neural networks [10], etc. Among the techniques, neural networks and multivariate linear regression have been widely employed to develop such data-driven models. However, there are no guarantees of avoidance of local minima, the overfitting phenomenon and the number of hidden units in general neural networks are usually difficult to choose [9].

SVM regression is a newly developed method based on statistical learning theory and is widely used in nonlinear function regression [11]. Compared with the neural networks, SVM does not require defining the number of hidden neurons, and overcomes the local minimizing and inadequate statistic problem. By solving a set of linear equations instead of quadratic programming, LS-SVM has a good performance of nonlinear modeling. Using the same samples, LS-SVM has better abilities of model approach and generalization than traditional neural networks and less running time than traditional SVM regression method [12]. LS-SVM regression method has such good characteristics as simple computation, high approaching speed and excellent generalization. Taking the LS-SVM method to build the soft sensor online is an interesting problem [13].

The selection of parameters plays an important role in the efficient performance of the LS-SVM model. Most of the researchers use the cross validation for the tuning of parameters. The other optimization techniques such as genetic algorithm (GA), simulated annealing, Bayesian regularization, heuristic method, hybrid approach are widely used in industrial process optimization [9,14]. These methods are different in terms of speed, accuracy and computational complexity. Considering the global optimization, in this paper, we use the real-cord multi-population GA to select the parameters for the soft sensors.

It is well known that the satisfactory performance of soft sensors is likely to be achieved if only those secondary variables that are most sensitive to the primary variables are employed [15]. The traditional PCA is a kind of well-known linear method for feature extraction. By calculating the eigenvectors of the covariance matrix of the original inputs, PCA linearly transforms a high-dimensional input vector into a low-dimensional whose components are uncorrelated. A lot of redundant information is eliminated by transforming the original data set to principal components [15]. However, the traditional PCA fails to extract the complex nonlinear structure from input data set. Nonlinear PCA has been developed by using different algorithms. Specifically, KPCA firstly maps the original inputs into a high-dimensional feature space using the kernel method and then calculates PCA in the high-dimensional feature space. The linear PCA in the high-dimensional feature space is corresponding to a nonlinear PCA in the original input space. The KPCA almost ensures principal components model with arbitrary precision by extracting more principal components than PCA [13,16].

A soft sensor modeling mainly focuses on multiple-input single-output system (MISO), while the study about multiple-input multiple-output system (MIMO) is little [17]. However, there are many correlated qualities variables which should be estimated in distillation processes online. The correlated qualities variables are coupled and correlative with each other, thereby establishing the MISO soft sensor model of the quality variable firstly and then combining the individual models into a final MIMO model is unnecessary.

From the above discussion, it appears that a systematic approach for MIMO soft sensors in distillation column is still lacking. In this paper, we focus on the MIMO soft sensors for the atmospheric distillation column and adopt multiple KPCA-MKLS-SVM regression to establish the model for estimation of aviation kerosene quality; we use the real-cord multi-population GA to select the parameters. The simulation results show that the method not only has simple structure but also provide the necessary accuracy of estimation.

The remainder of this paper is organized as follows. In Section 2, the KPCA algorithms are discussed. In Section 3, the MKLS-SVM regression algorithm is described. In Section 4, the outlier detection and data smoothing algorithm are presented. In Section 5, the performance of the MIMO soft sensors are discussed. Finally, summary and discussion are given in Section 6.

2. Kernel Principal Component Analysis.

Given a set of centered input data set x_i, $\sum_{i=1}^{m} x_i = 0$, $i = 1, 2, \ldots, m$, $x_i \in R^n$. The general idea of KPCA is firstly to perform a non-linear transformation via a non-linear function φ to map the original input vector to a high dimensional feature space F, and then calculates linear PCA in $\varphi(x_i)$. We assume that we are dealing with the centered data, that is $\sum_{i=1}^{m} \varphi(x_i = 0)$.

The covariance matrix of the sample in feature space F as follows:

$$\bar{C} = \frac{1}{m} \sum_{i=1}^{m} \varphi(x_i)\varphi(x_i)^T \tag{1}$$

The corresponding eigenvalue problem is

$$\lambda V = \bar{C} V \tag{2}$$

where λ is nonzero eigenvalues set of $\bar{C}$; V is the corresponding nonlinear eigenvector set. V lies in the span of φ_1, ..., φ_m. So, we can establish an equation as

$$\lambda \varphi(x_i) \cdot V = \varphi(x_i) \cdot \bar{C} V \tag{3}$$

The corresponding eigenvector V can be represented as

$$V = \sum_{i=1}^{m} \alpha_i \varphi(x_i) \tag{4}$$

For all $k = 1, 2, \ldots, m$, we get

$$\lambda \sum_{i=1}^{m} \alpha_i \varphi(x_k)\varphi(x_i)^T = \frac{1}{m} \sum_{i=1}^{m} \alpha_i \left(\varphi(x_k) \sum_{j=1}^{m} \varphi(x_j) \right) (\varphi(x_j)\varphi(x_i)) \tag{5}$$

Defining a kernel function by

$$K_{ij} = \varphi(x_i)\varphi(x_j) \tag{6}$$

Equation (5) can be transformed to the eigenvalue problem as

$$m\lambda \alpha_i = K\alpha_i \tag{7}$$

where α_i is the corresponding eigenvector of K. According to (4), the eigenvector V can be calculated using α_i.

Finally, for arbitrary vector x of the original space, the principal components can be calculated by

$$V \cdot \varphi(x) = \sum_{i=1}^{m} \alpha_i \varphi(x_i) \cdot \varphi(x) = \sum_{i=1}^{m} \alpha_i \cdot K(x_i, x) \tag{8}$$

In order to eliminate different impact due to the dimension and range of different variables, raw data should be standardized. In addition, for making the sample input

vectors in $\varphi(x_i)$, centered $\sum_{i=1}^{m} \varphi(x_i) = 0$ in (8), the kernel matrix on the training set K and on the testing set K_t are respectively, modified by

$$\widetilde{K} = K - (I/m) \cdot K - K \cdot (I/m) + (I/m) \cdot K \cdot (I/m) \tag{9}$$

$$\widetilde{K}_t = K - (I_t/m_t) \cdot K_t - K \cdot (I/m) + (I_t/m_t) \cdot K_t \cdot (I/m) \tag{10}$$

where m and m_t is respectively the number of training data points and testing data points. I and I_t is $m \times m$ and $m_t \times m$ matrix whose elements are all ones, respectively.

From (8), it can be found that the maximal number of principal components in KPCA is m (the number of training data) while n (the number of variables) in PCA. KPCA almost ensure principal components model with arbitrary precision by extracting more principal components than PCA.

3. Mixed Kernel LS-SVM Regression Algorithm.

3.1. **LS-SVM regression.** LS-SVM is based on structural risk minimization principle and has shown powerful ability in learning with limited samples, nonlinear, high dimension and so on practical problems [12,18]. The algorithm of function approximation based on LS-SVM regression algorithm is described as follows. Let training sample data:

$$D = (x_i, y_i) \mid i = 1, 2, \cdots, n, \ \ x_i \in R^n, \ \ y_i \in R \tag{11}$$

where x_i is input data; y_i is output data. The optimization problem in weight ω space can be described as

$$\min J(\omega, e) = \frac{1}{2}\omega^T \omega + \frac{1}{2}\gamma \sum_{k=1}^{n} e_i^2$$

$$\text{s.t. } y_i = \omega^T \psi(x_i) + b + e_i \tag{12}$$

in which $\psi(\cdot) : R^n \to R^{n_h}$ is kernel space mapping function, weight vector $\omega \in R^{n_h}$, error variable $e_i \in R$, b is the bias, losing function J is the sum of SSE error and regularized variable, and γ is adjusted constant. The kernel mapping function is to extract characters from origin space, and map the sample in origin space to a vector in the high dimension characters space.

According to optimization (12), it defines Lagrange function as

$$L(\omega, b, e, \alpha) = J(\omega, e) - \sum_{i=1}^{n} \alpha_i \left[\omega^T \phi(x_i) + b + e_i - y_i\right] \tag{13}$$

in which Lagrange factor (viz. support vector) $\alpha_i \in R$. We can do optimization computation on the above equation:

$$\frac{\partial L}{\partial \omega} = 0 \longrightarrow \omega = \sum_{i=1}^{n} \alpha_i \phi(x_i)$$

$$\frac{\partial L}{\partial b} = 0 \longrightarrow \sum_{i=1}^{n} \alpha_i = 0$$

$$\frac{\partial L}{\partial e_i} = 0 \longrightarrow \alpha_i = re_i$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \longrightarrow \omega^T \phi(x_i) + b + e_i - y_i = 0 \tag{14}$$

Eliminating ω, e, we can get the matrix equation

$$\begin{bmatrix} 0 & l^T \\ l & \Omega + \frac{1}{\gamma}I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{15}$$

where $l = [1; 1; \cdots ; 1] \in R^n$, $\Omega = \phi(x_i)^T \phi(x_j)$.

According to Mercer condition, it exists mapping function ϕ and kernel function $K(\cdot, \cdot)$, satisfying

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \tag{16}$$

The function approximation of LS-SVM is

$$y(x) = \sum_{i=1}^{n} \alpha_i K(x, x_i) + b \tag{17}$$

There are some typical kernel functions [11]: 1) polynomial kernels: $K_{poly}(x, x_i) = (x \cdot x_i + 1)^q$; 2) radial basis function kernels: $K_{rbf}(x, x_i) = \exp\left(-\frac{\|x - x_i\|}{2p^2}\right)$; 3) sigmoid kernels: $K_{sig}(x, x_i) = \tanh(v(x, x_i) + c)$, where q is degree of polynomial function; p is the width parameter of the Radial basis function; v is scale, c is offset.

3.2. **MKLS-SVM regression.** As known that kernel method provides a way of avoiding the difficulty of dimensionality, but how to select a right kernel function or use proper kernels to construct a more effective kernel is another problem should be solved. According to structural risk minimization principle, the VC dimension of a function set is the most important criterion for evaluating the generalization ability of a learning machine. By adjusting the different nuclear parameters, implicitly change feature space dimension.

The above kernel functions possess their respective merits and have different effects on LS-SVM performance. Among these kernel functions, RBF kernel function is a local kernel function with a stronger learning ability but weaker ability. However, polynomial kernel function is a global kernel function which has a better dissemination ability and weaker learning ability. According to the theory of structural risk minimization, we should think about both the ability of learning and dissemination. If choose a Poly kernel and RBF kernel to make up a new kernel, it will make an improvement on LS-SVM [19,20]. We use the mixed kernel as follows:

$$K_{mix}(x, x_i) = \rho K_{poly}(x, x_i) + (1 - \rho) K_{rbf}(x, x_i) \tag{18}$$

where $0 \le \rho \le 1$. According to Mercer theorem and the above linear superposition formula, it can be known that the mixed K_{mix} is also a Mercer kernel. Based on the mixed kernel K_{mix}, the MKLS-SVM method has four parameters needed to be selected, penalty factor r, RBF kernel parameter p, Poly kernel parameter q, $0 \le \rho \le 1$.

4. **Data Smoothing and Filtering.** The collected sample data needed for soft sensing should be true signal, while the actual signal can potentially corrupted by various types of noises. Modeling with low precision measurement data or outliers may result in greatly decreasing of soft sensor measuring performance, even the failure of soft sensing.

Many types of filters are available for filtering or smoothing the noise, such as limiting filtering, median filtering, arithmetic mean filter, and moving average filtering. Some of them are not very much effective and some of them destroy the characteristics of raw objective data during filtering process. The Savitzky-Golay (S-G) filter is one of the filters which can smoothen out the signal without much destroying its original properties [21]. This method can retain distribution characteristics of signal such as relative maximum, minimum values and width.

S-G filtering principle is described as follows:

Filter the data with an unweighted linear least squares fit using a polynomial of the specified degree. Assume that $2m + 1$ data points are positioned symmetrically about the origin (i.e., midpoint)

$$x = [x_{-m}, x_{-m+1}, \cdots, x_m]^T \tag{19}$$

Now, use an n $(n \leq 2m + 1)$ order polynomial to fit the set of data points

$$f_i = \sum_{k=0}^{n} b_{nk} i^k = b_{n0} + b_{n1}i + b_{n2}i^2 + \cdots b_{nn}i^n \tag{20}$$

According to least-squares technique, coefficients $b_{n1}, b_{n2}, \cdots, b_{nn}$ are needed to define in order to keep error squares minimum. It describes as follows:

$$\min E = \sum_{i=-m}^{m} [f_i - x(i)]^2 = \sum_{i=-m}^{m} \left[\sum_{k=0}^{n} b_{nk} i^k - x(i) \right]^2 \tag{21}$$

Use convolution coefficients of Savitzky-Golay polynomial convolution table to calculate the coefficient $b_{n1}, b_{n2}, \cdots, b_{nn}$. This method is convenient and fast [22].

5. MIMO Soft Sensor for Product Quality Predicting.

5.1. **The atmospheric distillation column.** The schematic diagram of the atmospheric distillation column is shown in Figure 1 which is one unit of the crude oil distillation units (CDU). B is the atmospheric distillation column and the fractionation occurs on the trays, separating the crude oil into desired fractions according to their boiling points, from lighter to heavier. The gasoline, aviation kerosene, light diesel oil, heavy diesel oil and other products are trapped out of the atmospheric distillation column and then steam stripped to remove light hydrocarbons. The aviation kerosene is got from the B1 stream in Figure 1.

The dry point and flash point are very important quality value for the aviation kerosene which cannot be online measured by hardware sensors. They can be got from the lab measurements which need a long time. As the quality parameters can not be measured, the indirectly control scheme is adopted in the refinery. The temperature and the flow constitute the cascade control to control the aviation kerosene quality in Figure 1. However, it is difficult to keep the aviation kerosene quality at its set-point by using this control scheme. So building the dry point and flash point soft sensors and realizing prediction online would be beneficial to help the control room personnel to make timely adjustment to the process to keep the control variables within limits.

5.2. **Selection of the secondary variables and data collection.** The choice of the secondary variables is one of the key techniques of soft sensing. Proper secondary variables choice can make the soft sensing be built based on the right relation of the input and output sample data. The fractionation occurs on the trays, separating the crude oil into desired fractions according to their boiling points, from lighter to heavier. The pressure and temperature of the trays is the main factors influence boiling point of liquid. So, the selection of the secondary variables should focus on temperature and pressure. In addition, the fluctuation of flow can also influence the working state of the distillation process through changing the related temperature, such as the backflow in the top of the atmospheric distillation column and the intermediate reflux. From the above discussion, in order to design the required soft sensor, the 23 secondary variables have been chosen. The considered 23 secondary variables are listed in Figure 1.

The control system of the atmospheric distillation column is the DeltaV distributed control system (DCS), which is made by Fisher-Rosemount, Emerson enterprise. DeltaV DCS has the function of OLE for Process Control (OPC). The sample data collection of the 23 secondary variables is reading the real-time data through OPC software. By the way, the aviation kerosene dry point and flash point is gotten from the lab which measured every 8 h in this factory. We collected the 80 groups of sample data which sorted by time order. Figure 2 is the sample data of the 23 secondary variables. Figure 3 shows the dry point and flash point values of aviation kerosene.
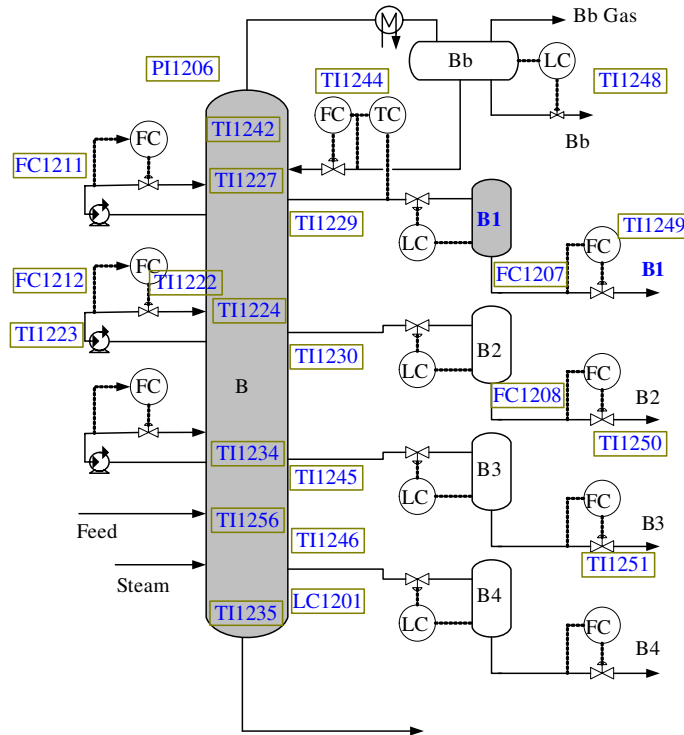
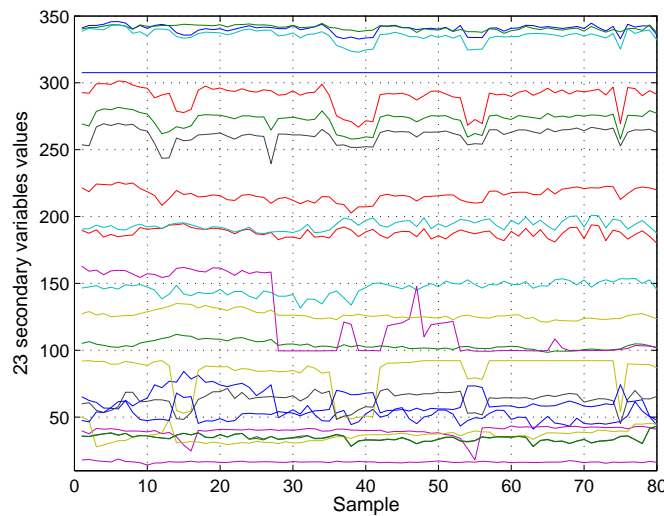FIGURE 1. Schematic diagram of the atmospheric distillation column

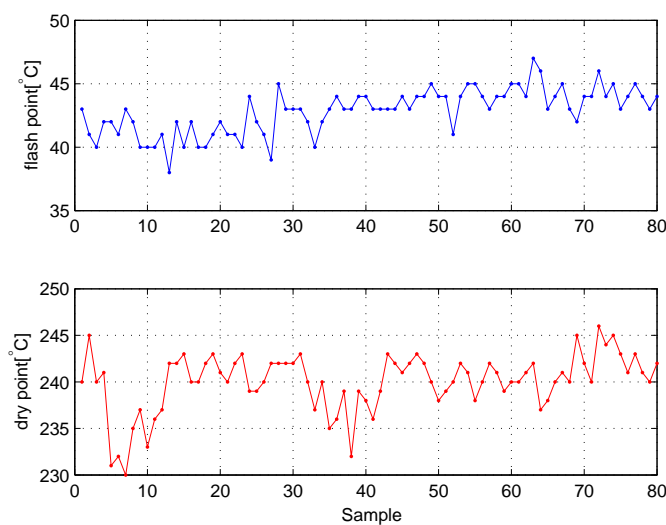FIGURE 2. The original data of the 23 secondary variables

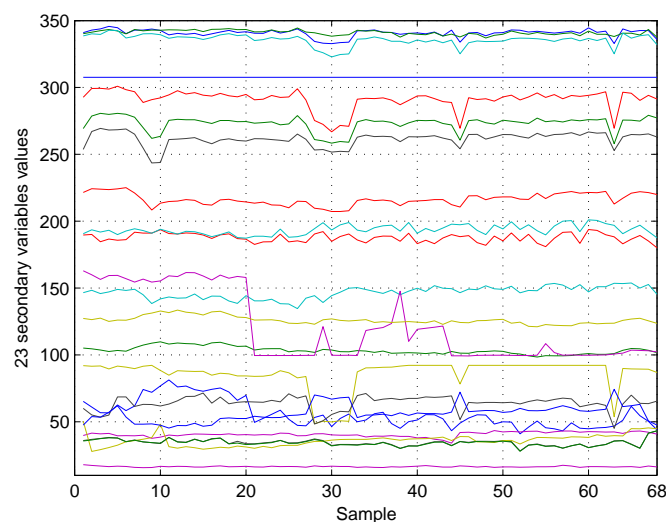FIGURE 3. The dry point and flash point values of aviation kerosene

FIGURE 4. Secondary variables values by sig-sigma rule

5.3. **Data handling.** Firstly, the sig-sigma rule was used to detect outliers and remove it form the sample data. Using the method, most outliers can be removed and complex computation can be avoided, the 68 groups of data are left. Figure 4 is the 23 secondary variables values processed by sig-sigma rule. Then the S-G filter was used to reduce the noise. In contrasting, the moving average filtering method is also used. Figure 5 and Figure 6 are original data and smoothed data of flash point and dry point values. The simulation results show that S-G filter can smoothen out the signal without much destroying its original properties and more effective than moving average filtering.

5.4. **Dimension reduction using KPCA.** The secondary variables selection should be based on the principle of downsizing. Some of the secondary variables may be interrelated, thus they cannot be directly input variables for the soft sensor model. KPCA was applied to choose the nonlinear principal component of the model in input data space; the effective
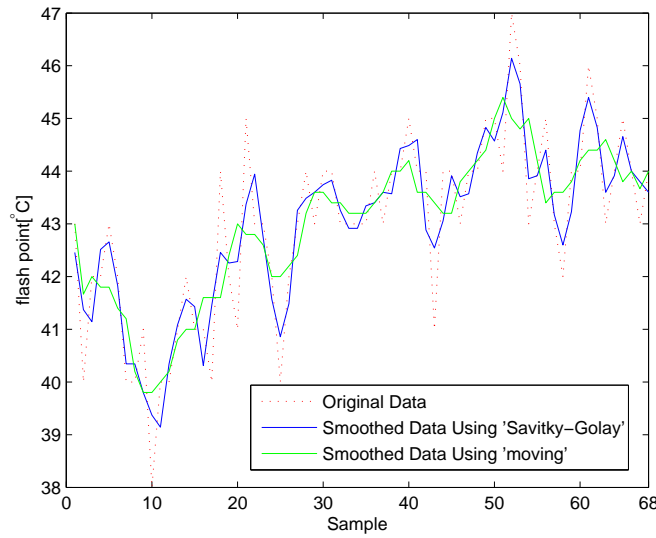
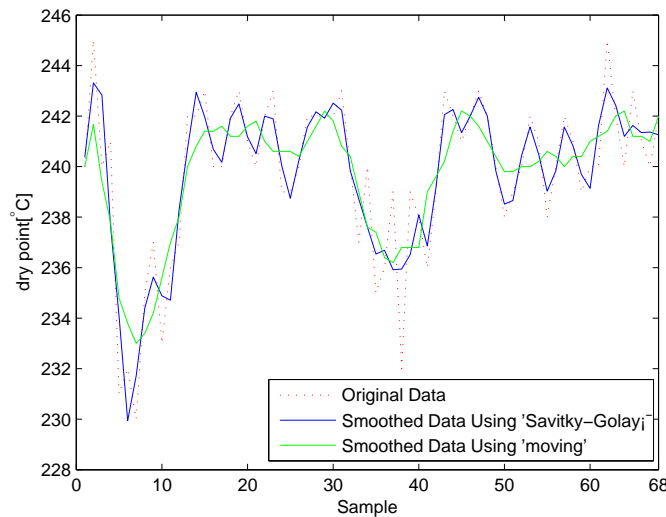FIGURE 5. Original data and smoothed data of flash point values

FIGURE 6. Original data and smoothed data of dry point values

information was extracted to eliminate redundancy variables. The results of KPCA are shown in Figure 7.

The first five principal components cumulative contribution percent of variance is 87%>85% which can be drawn from Figure 7. Thereby, the 23 secondary variables as input vector of the MIMO soft sensors have been reduced to the 5 variables which are uncorrelated.

5.5. **Parameters optimization based on multi-population GA.** The GA is to mimic the natural selection principle of survival of the fittest and make an evolutionary computing strategy to explore the relevant search space in order to find an optimal solution. The GA is not limited by the type of the objective function. Unlike many traditional optimization algorithms, there is no requirement for the objective function to be differentiable or continuous. GA is often trapped into local optimums during the
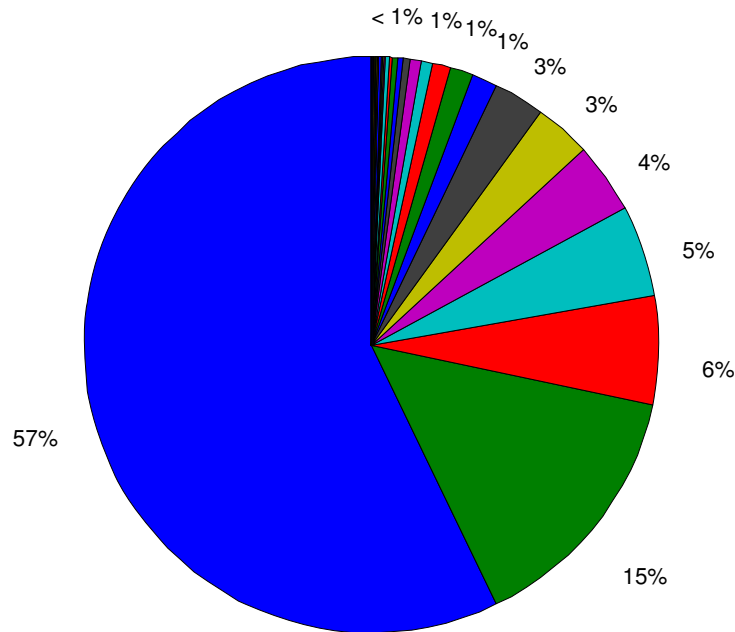
FIGURE 7. Principal components contribution percent of variance

optimization procedure [23,24]. To prevent premature convergence and to obtain near global optimal solutions, the real-cord multi-population GA with elitist strategy, migration operator and self-adaptive mutation and crossover operator is adopted to optimize the parameters in MKLS-SVM.

A multi-population GA divides a single population into smaller subpopulations. Each subpopulation evolves by genetic operations in parallel with the other, while maintaining a limited but powerful interaction between all subpopulations. The most important advantage of subpopulations is to enhance diversity among the subpopulations [24]. A migration mechanism, which exchanges chromosomes among the subpopulations, exchanges information during the joint optimization to maintain diversity and thus avoid a systematic premature convergence toward a single local optimum.

GA consists of five main stages: evaluation, selection, crossover, mutation and migration, which called genetic operations, embody the particular property of GA. The main process of parameters optimization is described in the following.

1) Parameters coding

To reflect the property of problem, floating point numbers are applied to code the gene. Moreover, floating point number code has higher precision, larger scope for search and avoids decoding and coding repeatedly which binary code need. For a real coded GA, a chromosome corresponds to a vector of real parameters, a gene corresponds to a real number, and an allele corresponds to a real value.

2) Population initialization

The genetic operators and the parameters used for the GA for evolving each individual population are as follows:

Number of subpopulations Nsubpop = 8; Number of individuals per subpopulation Nchild = 30; Generation gap Ggap = 0.8; Maximum number of generations Maxgen = 300; Value for termination Vter = 1e-4; Insertion rate INSR = 0.7, how many of the offspring are inserted due to INSR; Migration rate between subpopulations Migr = 0.2.

Optimum parameter scope $\gamma \in [0.001, 1000]$, $p \in [0.001, 1000]$, $q \in [0.1, 100]$, $\rho \in [0.01, 1]$ in MKLS-SVM.

Setting the genetic operators and the parameters, randomly generate an initial population of chromosomes which present the values of parameters in MKLS-SVM.

3) Fitness function setting

Each resulting MKLS-SVM is trained and its performances are evaluated by the mean of a root mean squared error (RMSE) calculated on a validation set. Parents of the next generation are selected according to a fitness function that depends on the RMSE. Individuals with larger fitness value have greater possibility of being selected as parents. The RMSE and the fitness function is defined as

$$RMSE(\gamma, p, q, \rho) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \frac{(y_i - \widehat{y_i})^2}{y_i}} \tag{22}$$

$$fitness = \frac{1}{RMSE(\gamma, p, q, \rho)} \tag{23}$$

where y_i and $\widehat{y_i}$ represent the actual and validation values respectively; n is the number of validation set.

4) Genetic operation

a) Selection operation: Populations are sorted by fitness values in descending order. Select excellent chromosomes with higher fitness values as transitional generation by means of stochastic universal sampling. According to the generation gap make sure individual number of selection. Filter out individuals with lower fitness and retain the best individuals. If the ith individual's fitness is f_i, then its probability of selection becomes $P_i = f_i(\sum_{k=1}^{m} f_k)^{-1}$, m represents the population size.

b) Crossover operation and mutation operation: In contrast to GA, the researchers have recently been proposing new adaptive approaches in the GA for the mutation and crossover operators to increase the probability of capturing the global optimum, to enhance the performance of the GA and to relieve the user from the burden of having to determine sensitive parameters existed in the GA [25]. The adaptive approaches for mutation and crossover operator of the GA are as follows [25]:

$$p_c = \begin{cases} (f_{\max} - f')/(f_{\max} - f_{ave}); & f' \geq f_{ave} \\ 1.0; & f' < f_{ave} \end{cases} \tag{24}$$

$$p_m = \begin{cases} 0.5(f_{\max} - f)/(f_{\max} - f_{ave}); & f \geq f_{ave} \\ (f_{\max} - f)/(f_{\max} - f_{\min}); & f < f_{ave} \end{cases} \tag{25}$$

Here, f is the fitness of individual, f_{ave} is average fitness value of the population. $f_{\max}$, $f_{\min}$ are maximum and minimum fitness value of the population respectively. f' is the larger of the fitness values of the solutions to be crossed. The adaptive approaches are able to adjust themselves automatically during the evolutionary process. Therefore, the algorithm does not need any predefined parameters.

We use multiple-point crossover of discrete recombination to recombine selected individuals. For real-coded GA, discrete recombination is the process in which exchange variable value between individuals, then produce new chromosome. Set a crossing point in any two adjacent genes of chromosome. Crossing points is set to 3, so that generate all possible individuals which produced by the parent recombination. For each variable of offspring, select the father which contributes to offspring according to crossover probability p_c. Two individuals exchange each variable between them.

Gaussian mutation has usually been used to produce offspring for the real-coded GA. A Gaussian mutation operator requires two parameters [26]: the mean which is often set to zero, and the standard deviation σ which can be interpreted as the mutation step size.
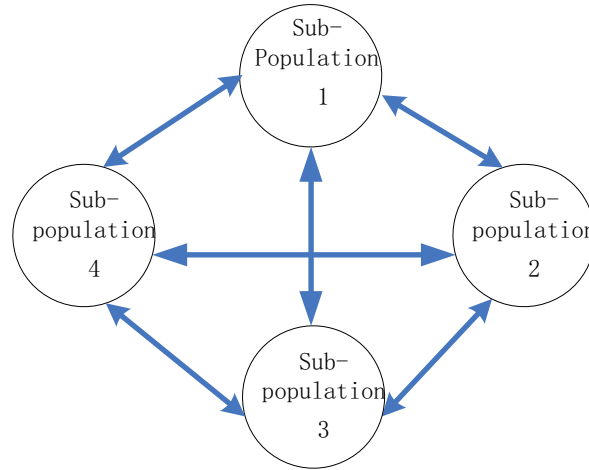
FIGURE 8. The complete network structure of migration

Then, mutations are realized by replacing components of the vector x by $v_i' = v_i + N(0, \sigma)$, where $N(0, \sigma)$ is a random Gaussian number with mean zero and standard deviation.

c) Reinsertion (Elitist strategy): To restore population size, insert best transition generation replacing worst parents after calculating every individual's fitness f. The Insertion rate defines the insertion number. Through reinsertion the optimal individuals of father generations have been reserved.

d) Migration mechanism: Select randomly some individuals (the number is Nchild* Miggen) from one subpopulation and replaced by the best individuals of other subpopulations based on individual's fitness f.

All subpopulation will repeat the process. Figure 8 is the complete network structure diagram of migration.

The evolutionary process continues until stop conditions are satisfied. When a termination criterion is met, the individual having the best fitness defines the optimal parameters.

5.6. **MIMO soft sensor modeling.** The MKLS-SVM based on polynomial kernel and RBF kernel is adopted to build the MIMO soft sensors. The 61 groups of data are used for training samples, the other 7 groups are used for testing samples in the soft sensing. The hyper-parameters are adaptively evolved by multi-population GA according to the training samples and testing samples. Figure 9 and Figure 10 are the prediction results of KPCA and MKLS-SVM soft sensors for training data and testing data.

In order to contrast the performance of the MKLS-SVM method, the LS-SVM based on RBF kernel, the BP neural network and the RBF neural network methods are also adopted. We use the root mean square error (RMSE) as the performance criteria of these models. The results of these models with the optimal selected parameters obtained by multi-population GA for flash point and dry point are summarized in Table 1 and Table 2 respectively. The considered optimal selected parameters of models are also listed in Table 1 and Table 2, where DF is the number of the hidden neurons; HN is the number of Hidden layer, $Goal$ is the performance indicators of training error.

Next, we present a comparison between the PCA and the KPCA method based soft sensors. Table 3 and Table 4 respectively showed the performance of the four corresponding PCA combined models with the optimal parameters for flash point and dry point. The considered parameters of models are also listed in Table 3 and Table 4.

Simulation result indicates that the KPCA-MK-LS-SVM regression has better abilities of approach and generalizing overhead than KPCA-LS-SVM, KPCA-BP, KPCA-RBF
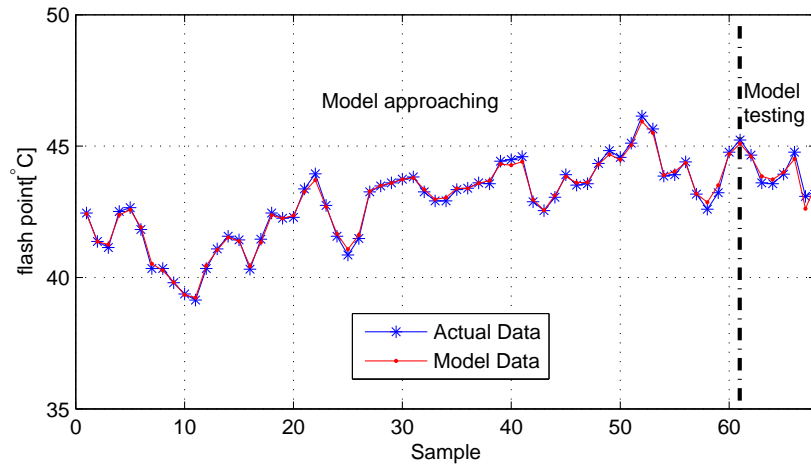
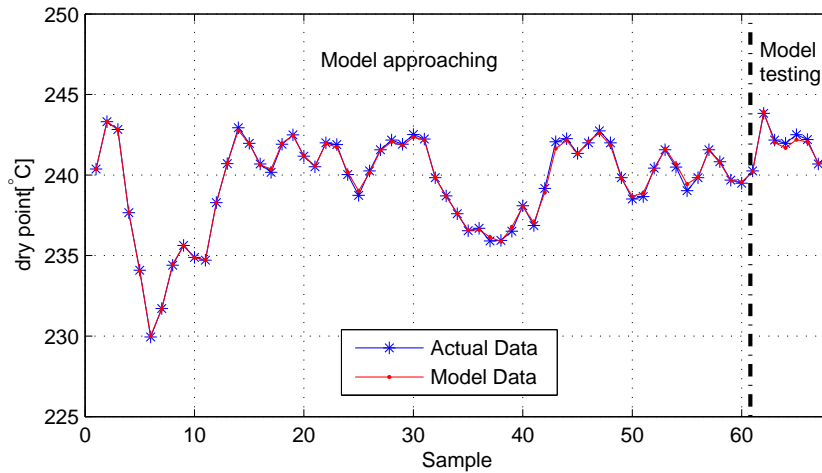FIGURE 9. KPCA-MKLS-SVM method for the flash point data

FIGURE 10. KPCA-MKLS-SVM method for the dry point data

TABLE 1. KPCA combined models for flash point values

	Training volume	Testing volume	Optimal selected parameters
MKLS-SVM	0.11123	0.25533	$\gamma = 0.01848$, $p = 0.01$, $q = 5.6$, $\rho = 0.966$
LS-SVM	0.36395	0.41426	$\gamma = 0.03$, $p = 0.315$
RBF	0.47431	0.55026	$DF = 26$, $p = 1$, $Goal = 0.3$
BP	0.55123	0.59307	$DF = 6$, $HN = 1$, $Goal = 0.001$

methods. The soft sensor's computation errors of training and generalization can be controlled well based on MK-LS-SVM method. Comparisons between the PCA and KPCA based models are shown in Tables 1-4, the simulation results show that KPCA own more powerful ability in choosing the nonlinear principal component of the model input data space and feature extraction.

In summary, the accuracy of soft sensors based on MKLS-SVM method is higher than that based on LS-SVM with RBF kernel under the same condition. Compared with LS-SVM with RBF kernel, MKLS-SVM possesses the better dissemination ability and approach ability by absorbing the advantages of RBF kernel and polynomial kernel function.

TABLE 2. KPCA combined models for dry point values

	Training volume	Testing volume	Optimal selected parameters
MKLS-SVM	0.13286	0.26503	$\gamma = 0.01548$, $p = 0.01$, $q = 25.2$, $\rho = 0.966$
LS-SVM	0.3967	0.96414	$\gamma = 0.00087$, $p = 0.25$
RBF	0.44929	1.1762	$DF = 41$, $p = 0.8$, $Goal = 0.07$
BP	0.7765	0.92075	$DF = 15$, $HN = 1$, $Goal = 0.003$

TABLE 3. PCA combined models for flash point values

	Training volume	Testing volume	Optimal selected parameters
MKLS-SVM	0.35805	0.45389	$\gamma = 0.002$, $p = 4.2848$, $q = 1$, $\rho = 0.9666$
LS-SVM	0.37283	0.59881	$\gamma = 0.15$, $p = 3.1848$
RBF	0.45955	0.60492	$DF = 18$, $p = 8.5$, $Goal = 0.3$
BP	0.3802	0.66558	$DF = 6$, $HN = 1$, $Goal = 0.001$

TABLE 4. PCA combined models for dry point values

	Training volume	Testing volume	Optimal selected parameters
MKLS-SVM	0.11251	0.84632	$\gamma = 0.015$, $p = 0.01$, $q = 1$, $\rho = 0.9666$
LS-SVM	0.41385	1.0248	$\gamma = 0.01$, $p = 2.71$
RBF	0.6434	1.2561	$DF = 54$, $p = 1.6$, $Goal = 0.01$
BP	0.78841	1.3338	$DF = 20$, $HN = 1$, $Goal = 0.003$

Using appropriate kernel functions, the methods based on KPCA show better performance than the methods based on PCA. The computational complexity has nothing with the dimension of input variable, so soft sensor model combined with KPCA methods will have better performance than PCA methods.

6. **Conclusions.** This paper does research on MIMO soft sensor for aviation kerosene quality prediction based on KPCA and MKLS-SVM regression. The simulation results show that this method has good abilities of model generalization and the predicted values are in good agreement with lab measurements. In addition to modeling, additional factors such as the selection of secondary variables, data pretreatment and model parameters optimization make a big difference to realize the soft sensors successfully.

In this paper, the mechanism analysis combined with KPCA is adopted to select the secondary variables. Compared with the traditional statistical methods PCA, KPCA used nonlinear extraction method shows good effect in fast speed of feature extraction and sufficient feature information retained. The S-G filter method is used to smooth and filter the original data, S-G filter can smoothen out the signal without much destroying its original properties and more effective than moving average filtering. Because the GA shows efficient global parallel search ability and it is especially suitable for solving optimization problems of complex systems with multi-target and nonlinear. To prevent premature convergence and to obtain near global optimal solutions, the real-cord multi-population GA with elitist strategy, migration operator and self-adaptive mutation and crossover operator is adopted to optimize the parameters in MKLS-SVM.

Further research in this direction may be needed, to gain better understanding of the number of the historical datasets, the noise distribution of the sample data, the real-time performance of the model and the number of the input variables. In particular, the number of the relevant input variables may be optimized to provide superior generation

performance. In summary, successful installation of the soft sensing in a refinery will ensure better product quality control with higher productivity.

REFERENCES

[1] B. Roffel, B. H. L. Betlem and R. M. de Blouw, A comparison of the performance of profile position and composition estimators for quality control in binary distillation, *Computers & Chemical Engineering*, vol.27, no.2, pp.199-210, 2003.

[2] E. Zamprogna, M. Barolo and D. E. Seborg, Estimating product composition profiles in batch distillation via partial least squares regression, *Control Engineering Practice*, vol.12, no.7, pp.917-929, 2004.

[3] T. Chatterjee and D. N. Saraf, On-line estimation of product properties for crude distillation units, *Journal of Process Control*, vol.14, no.1, pp.61-77, 2004.

[4] D. Wang, J. Liu and R. Srinivasan, Data-driven soft sensor approach for quality prediction in a refining process, *IEEE Transactions on Industrial Informatics*, vol.6, no.1, pp.11-17, 2010.

[5] R. Taylor, R. Baru and R. Krishna, Dynamic behaviour of reactive distillation columns described by a nonequilibrium stage model, *Chemical Engineering Science*, vol.56, no.6, pp.2085-2102, 2001.

[6] M. Kano, K. Miyazaki, S. Hasebe and I. Hashimoto, Inferential control system of distillation compositions using dynamic partial least squares regression, *Journal of Process Control*, vol.10, no.2-3, pp.157-166, 2000.

[7] X. F. Yan, Modified nonlinear generalized ridge regression and its application to develop naphtha cut point soft sensor, *Computers & Chemical Engineering*, vol.32, no.3, pp.608-621, 2008.

[8] P. Jain, I. Rahman and B. D. Kulkarni, Development of a soft sensor for a batch distillation column using support vector regression techniques, *Chemical Engineering Research & Design*, vol.85, no.A2, pp.283-287, 2007.

[9] W. W. Yan, H. H. Shao and X. F. Wang, Soft sensing modeling based on support vector machine and Bayesian model selection, *Computers & Chemical Engineering*, vol.28, no.8, pp.1489-1498, 2004.

[10] S. Yu, K. Zhu and F. Diao, A dynamic all parameters adaptive BP neural networks model and its application on oil reservoir prediction, *Applied Mathematics and Computation*, vol.195, no.1, pp.66-75, 2008.

[11] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.

[12] J. A. K. Suykens, J. De Brabanter, L. Lukas and J. Vandewalle, Weighted least squares support vector machines: Robustness and sparse approximation, *Neurocomputing*, vol.48, pp.85-105, 2002.

[13] M. Liu, D. P. Huang, Z. H. Sun and Z. S. Chen, Combining KPCA with LSSVM for the Mooney-Viscosity forecasting, *Proc. of the 2nd International Conference on Genetic and Evolutionary Computing*, Jingzhou, China, pp.522-526, 2008.

[14] S. Huang and G. Xu, LS-SVM with genetic algorithm for forecasting of runway incursion events, *Proc. of 2009 International Conference on Information Engineering and Computer Science*, Wuhan, China, pp.1-4, 2009.

[15] E. Zamprogna, M. Barolo and D. E. Seborg, Optimal selection of soft sensor inputs for batch distillation columns using principal component analysis, *Journal of Process Control*, vol.15, no.1, pp.39-52, 2005.

[16] L. J. Cao, K. S. Chua, W. K. Chong, H. P. Lee and Q. M. Gu, A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine, *Neurocomputing*, vol.55, pp.321-336, 2003.

[17] I. Goethals, K. Pelckmans, J. A. K. Suykens and B. De Moor, Identification of MIMO Hammerstein models using least squares support vector machines, *Automatica*, vol.41, no.7, pp.1263-1272, 2005.

[18] J. A. K. Suykens, J. Vandewalle and B. De Moor, Optimal control by least squares support vector machines, *Neural Networks*, vol.14, no.1, pp.23-35, 2001.

[19] J. Luts, G. Molenberghs, G. Verbeke, S. Van Huffel and J. A. K. Suykens, A mixed effects least squares support vector machine model for classification of longitudinal data, *Computational Statistics & Data Analysis*, vol.56, no.3, pp.611-628, 2012.

[20] N. Segata and E. Blanzieri, Operators for transforming kernels into quasi-local kernels that improve SVM accuracy, *Journal of Intelligent Information Systems*, vol.37, no.2, pp.155-186, 2011.

[21] M. A. Awal, S. S. Mostafa and M. Ahmad, Performance analysis of Savitzky-Golay smoothing filter using ECG signal, *IJCIT*, vol.1, no.2, pp.24-29, 2011.

[22] A. Savitzky and M. J. E. Golay, Smoothing and differention of data by simplified least squares procedures, *Analytical Chemistry*, pp.1627-1639, vol.36, no.7, 1964.

[23] L. Chen, Real-coded genetic algorithm optimization of longterm reservoir operation, *Journal of the American Water Resources Association*, vol.39, no.5, pp.1157-1165, 2003.

[24] L. Chen and F. J. Chang, Applying a real-coded multi-population genetic algorithm to multi-reservoir operation, *Hydrological Processes*, vol.21, no.5, pp.688-698, 2007.

[25] V. Togan and A. Daloglu, Optimization of 3d trusses with adaptive approach in genetic algorithms, *Engineering Structures*, vol.28, no.7, pp.1019-1027, 2006.

[26] R. Hinterding, Gaussian mutation and self-adaptation in numeric genetic algorithms, *Proc. of 1995 IEEE International Conference on Evolutionary Computation*, Perth, WA, USA, pp.384-389, 1995.