

Research



Cite this article: Li X, Yang C, Wang Y, Wang H, Zu X, Sun Y, Hu S. 2018 Compressor map regression modelling based on partial least squares. *R. Soc. open sci.* **5**: 172454. <http://dx.doi.org/10.1098/rsos.172454>

Received: 4 January 2018

Accepted: 3 August 2018

Subject Category:

Engineering

Subject Areas:

mechanical engineering/power and energy systems/mathematical modelling

Keywords:

diesel engine, performance modelling, compressor maps, partial least squares, regression modelling

Author for correspondence:

Yinyan Wang

e-mail: 1134898362@qq.com

Compressor map regression modelling based on partial least squares

Xu Li, Chuanlei Yang, Yinyan Wang, Hechun Wang, Xianghuan Zu, Yongrui Sun and Song Hu

College of Power and Energy Engineering, Harbin Engineering University, Harbin, People's Republic of China

id XL, 0000-0001-6119-7757; CY, 0000-0001-7468-5370; YW, 0000-0001-5665-5178; HW, 0000-0002-8690-0901; XZ, 0000-0002-2678-4750; YS, 0000-0002-5184-1534; SH, 0000-0003-1867-5869

In this work, two kinds of partial least squares modelling methods are applied to predict a compressor map: one uses a power function polynomial as the basis function (PLSO), and the other uses a trigonometric function polynomial (PLSN). To demonstrate the potential capabilities of PLSO and PLSN for a typical interpolated prediction and an extrapolated prediction, they are compared with two other classical data-driven modelling methods, namely the look-up table and artificial neural network (ANN). PLSO and PLSN are also compared with each other. The results show that PLSO and PLSN have a better prediction performance than the look-up table and the ANN, especially for the extrapolated prediction. The computational time is also decreased sharply. Compared with PLSO, PLSN is characterized by a higher prediction accuracy and shorter computational time than PLSO. It is expected that PLSN could save computational time and also improve the accuracy of a thermodynamic model of a diesel engine.

1. Introduction

The turbocharger is a vital component of the modern diesel engine. It allows the engine to increase its power density through the downsizing concept [1–3] while simultaneously decreasing fuel consumption. As is well known, the compressor and the turbine are the two most important components of the turbocharger itself. Consequently, compressor models play a significant role in the simulation and modelling of diesel engines, and, thus, it is crucial to accurately describe the compressor map. However, one of the most troublesome problems in the development of a compressor model is the strong nonlinearity of its pressure characteristics curves and efficiency characteristics curves. In general, many experimental studies on the various operating and environmental

Table 1. Nomenclature.

PLS	partial least squares
PLS-R	partial least squares regression
PLSO	partial least squares with a power function polynomial form as the basis function
PLSN	partial least squares with a trigonometric function polynomial form as the basis function
EPL	extrapolation performance at the lower part-load operating area
IPL	interpolation performance at the lower part-load operating area
EPU	extrapolation performance at the upper part-load operating area

conditions of a compressor are usually conducted to obtain the performance map. However, the experimental data in the full operational range are usually very limited because of the cost and restrictions of the compressor test (table 1).

In the published literature, there are several compressor modelling methodologies. The most popular methodology is look-up tables [4], which use the experimental compressor mass flow rate and efficiency maps. These maps are provided by compressor manufacturers in the form of lines of constant reduced speed and efficiency. The reduced speed and efficiency at each point are calculated by linearly interpolating and extrapolating the experimental maps [5]. As is well known, a compressor has strong nonlinear relationships among its reduced speed, pressure ratio, mass flow rate and efficiency. Therefore, the predictive accuracy of a compressor map using look-up tables is based on the quantity and quality of the experimental data. The prediction performance of look-up tables greatly deteriorates when the experimental data are very limited.

To improve the accuracy of the compressor's prediction, in [6], the elliptical curve fitting method was introduced to map the fitting process. In this method, the shape of the compressor map is expressed by the mathematical equation of an ellipse. According to the locations of the centre and axes of the ellipse, the elliptical curve fitting method is divided into three branches: a fixed centre without rational axes; a fixed centre with rational axes; and a flexible centre with rational axes. The method of a fixed centre without rational axes has the worst predicted accuracy, having the lowest number of subcoefficients (20 subcoefficients) of the three branches. The method of a flexible centre with rational axes has an acceptable predictive accuracy, with up to 100 subcoefficients, but this is too complicated for compressor map modelling.

Some intelligence algorithms are also applied to compressor map modelling. The well-known artificial neural network (ANN) algorithm is widely used in many areas [7] because of its ability to process nonlinearly and store massive amounts of experimental knowledge. Theoretically, an ANN can approximate any nonlinear model and develop the relationship between the input and output variables involved in a physical process without considering the underlying theories [8–10]. Hence, ANNs have become increasingly popular in recent years for predicting the performance of a compressor map. As a type of improved ANN, a feed-forward back-propagation neural network (BPNN) introduces feed-forward and back-propagation into the ANN. This means signals feed forward and there is a deviation between the experimental data and predictive value back-propagation. Consequently, BPNNs have the capability to rapidly converge the characteristics and nonlinear mapping. The network is first trained by experimental data, then the predictive results are regarded as experimental data in the second train. Although BPNNs are effective in predicting the compressor map via interpolation, their predictive accuracy is terrible in predicting the compressor map via extrapolation [11]. Moreover, it is important to note that a large number of experimental data are necessary to sufficiently train a BPNN to obtain a highly predictive accuracy and calculation stability. As a result, BPNNs are not suitable for compressor map modelling when the experimental data are very limited [12–15].

Computational fluid dynamics (CFD) simulations were applied to estimate the compressor map numerically in several studies [16–19]. This method has been proven to have a relatively high accuracy, but it is usually quite time-consuming, which makes it impossible to predict the compressor map in an available time frame. To shorten the computational time, the CFD simulation is converted into two-dimensional (2D) methodologies [20] and mean-line (one dimensional) methodologies [21] by using empirical or semi-empirical correlations. Accordingly, the predictive accuracy is inevitably sharply decreased. Another disadvantage of these methods is that they require some knowledge about the geometry of the compressor, which is not always available as this may be the manufacturer's proprietary information.

The partial least squares (PLS) regression method is a novel multiple statistical analysis method developed from the field of chemistry [22]. The PLS method has been introduced into compressor performance prediction because of the following advantages: it is strong and robust, it requires a low

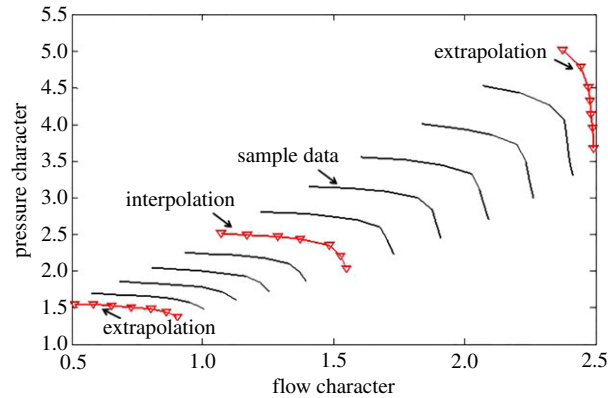


Figure 1. Compressor characteristics map generalization.

sample number and it has obvious physical meaning [23,24]. Even though the prediction accuracy of the PLS method is already at a high level, there is still ongoing work to improve it further. Tian *et al.* [25] combined an ANN and PLS to predict the performance of a scroll compressor, dividing the compressor performance into nonlinear and linear portions. The nonlinear portion was simulated using an ANN, while the linear one was modelled using the PLS method. The prediction results show that this method has a higher prediction accuracy than both an ANN and a normal PLS. However, this method is too complex, and it is difficult to distinguish the nonlinear and linear portions. Chu *et al.* [26] introduced a kernel function into the PLS method, referred to hereafter as the kernel partial least squares (KPLS). The KPLS improved the prediction accuracy owing to its diversity, but the prediction accuracy can greatly change when using different types of kernel functions.

As described above, completely modelling a successful engine's thermodynamic performance mainly depends on the method used, which must express well the component characteristic maps with good extrapolation and interpolation performance. In this work, we propose a new prediction model with a simple structure with a high prediction accuracy and short computational time. A new PLS prediction model (PLSN) is proposed based on the traditional PLS model by replacing the basic function of the PLS, a power function polynomial, with a trigonometric function. The results show that, when compared with a look-up table, the BPNN and the traditional PLS method (PLSO), PLSN achieves the best prediction accuracy while having the shortest computational time. Compared with other PLS models [25,26], the prediction accuracy is of the same magnitude, but the model structure of PLSN is much simpler. The remainder of this paper is organized as follows. In §2, the simulation problems of compressor characteristics maps and some relative solutions proposed in the literature are described. In §3, an application and an analysis of the novel method proposed by the authors are presented. Finally, §4 presents the conclusions.

2. Methodology

2.1. Problem description of compressor characteristics maps

One of the most troublesome problems in the development of a component-based diesel engine performance model is compressor modelling. Fortunately, a steady-state compressor map can be used in compressor modelling, because the air flow in the compressor can be assumed to be a quasi-steady flow. In general, the compressor is modelled as a lumped element in the thermodynamic performance model. The compressor performance can be represented by four characteristic parameters: reduced flow rate, isentropic efficiency, pressure ratio and reduced speed [27]. The characteristic parameters are usually represented as a flow characteristics map and an efficiency characteristics map. Once two arbitrary characteristic parameters are confirmed, the other two characteristic parameters are also confirmed through the flow characteristics map and efficiency characteristics map.

Usually, parts of the rotational speed characteristics curves of compressor maps, including the on-design operating point, are derived from a test bed or by means of CFD during actual thermodynamic performance modelling, and reasonable interpolation and extrapolation have to be performed in order to make good use of the flow characteristics and efficiency characteristics under a wider range of off-design characteristic areas. The experimental data shown in figure 1 are derived from the manufacturer, and the data are

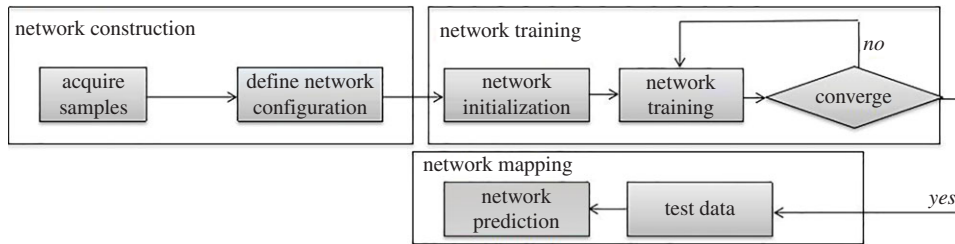


Figure 2. Neural network-based compressor performance modelling.

divided into three portions: sample data for building the predicting model, interpolation data and extrapolation data for verifying the accuracy of the prediction model.

2.2. Look-up table method

One of the most popular and simplest approaches is the look-up table method, in which the core algorithm is often linear or spline interpolation and extrapolation; this method has been widely used in almost every commercial thermodynamic calculation software, i.e. Krawal-modular, GT-ise, IPSEpro and Thermoflex. Owing to the sparsity of the compressor characteristics map data from the test bed or from CFD and discretization, which will reduce the amount of characteristic data further, a component characteristics table inserted into the thermodynamic performance model may cause information loss and cannot accurately represent the shape of the characteristics map. On the one hand, because of the different pressure ratio ranges for different constant rotational speed curves, the dimensional range of the pressure ratio in the table is different, which makes the interpretation of the table difficult. On the other hand, for a high, constant rotational speed curve, the flow characteristics curve is approximately vertical in the range of a comparatively small pressure ratio; however, if the data for discretization are incomplete, it is difficult to determine infinitesimal changes in the flow rate with the pressure ratio. Therefore, owing to the discretization process, a difference between the characteristics map in a table and the actual characteristics is simply unavoidable, so the sample data of compressor maps in the table should be dense and regular.

2.3. Artificial neural network method

The neural network method, which can approximate any nonlinear function by selecting the appropriate network, has become an effective solution for expressing compressor characteristics maps due to its highly nonlinear mapping capacity. ANN-based compressor performance modelling involves the following three steps: network construction, network training and network mapping. A detailed schematic diagram of the process is shown in figure 2.

A BPNN [28], which is a feed-forward BPNN and one of the foremost types in the ANN family, is used as a comparative algorithm. Here, the version of the BPNN that we use is from [29].

2.4. A new regression analysis method: partial least squares regression

The PLS regression (PLS-R) method is a novel multiple statistical analysis method developed from the field of chemistry and proposed by S. Wold and C. Albno in 1983. PLS-R combines the basic functions of a multiple linear regression analysis, a canonical correlation analysis and a principal component analysis, and integrates the data analysis method of modelling the prediction type with a non-model-based data analysis method organically [30]. In general, a PLS-R consists of the following steps.

Suppose the number of dependent variables is q and the number of independent variables is p . To study the statistical relationship between the dependent variables and independent variables, n number of sample points are observed to constitute the data sheet of the dependent variables and independent variables, i.e. $X = [x_1, x_2, \dots, x_p]_{n \times p}$ and $Y = [y_1, y_2, \dots, y_q]_{n \times q}$. Components t_1 and u_1 are extracted from X and Y , respectively (that is, t_1 is the linear combination of x_1, x_2, \dots, x_p and u_1 is the linear combination of y_1, y_2, \dots, y_q). During the extraction of these two components, two requirements should be noted as follows in order to accommodate the need for regression:

(i) t_1 and u_1 should carry the most mutation information about their own data sheets. This means the variances in t_1 and u_1 reach their respective maximum values.

(ii) The absolute value of the degree of the correlation of t_1 and u_1 achieves its maximum value.

These two requirements show that t_1 and u_1 should be as sufficient as possible to represent their own data sheet, while the component t_1 of X has the strongest explanatory power for component u_1 of Y . After extracting the first components t_1 and u_1 , PLS-R conducts the regression of X for t_1 and the regression of Y for u_1 . If the regression equation has reached a satisfactory accuracy, the calculation terminates; if not, the second round of component extraction is conducted, making use of the residual information of X being interpreted by t_1 and the residual information of Y being interpreted by u_1 , and so on until a satisfactory accuracy is reached. If m number of components t_1, t_2, \dots, t_m are finally extracted from X , PLS-R conducts the regression of y_k for t_1, t_2, \dots, t_m , and then, y_k is further expressed by a regression equation regarding the previous independent variables $x_1, x_2, \dots, x_p, k = 1, 2, \dots, q$. The basic mathematical procedure of PLS-R is as follows.

(i) Suppose $X = [x_1, x_2, \dots, x_p]_{n \times p}$, and $Y = [y_1, y_2, \dots, y_q]_{n \times q}$. Standardizing X gives

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad (2.1)$$

where $i = 1, 2, \dots, n, j = 1, 2, \dots, p$ and $s_j = \left[\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x}_j)^2 \right]^{1/2}$.

Then, $E_0 = [x_1^*, x_2^*, \dots, x_p^*]_{n \times p}$ is obtained.

Y is standardized in the same way, and, then, $F_0 = [y_1^*, y_2^*, \dots, y_j^*, \dots, y_q^*]_{n \times q}$ is obtained.

(ii) Extract component:

$$w_i = \frac{E'_{i-1} F_{i-1}}{\|E'_{i-1} F_{i-1}\|}, \quad t_i = E_{i-1} w_i, \quad p_i = \frac{E'_{i-1} t_i}{\|t_i\|^2} \quad (2.2)$$

and

$$E_i = E_{i-1} - t_i p_i' \quad (2.3)$$

where $i = 1, 2, 3, \dots, m, m \leq \text{rank}(E_0)$ and can be determined by cross-validation.

(iii) Solve regression coefficient vector:

$$r_i = \frac{F'_{i-1} t_i}{\|t_i\|^2}, \quad i = 1, 2, 3, \dots, m \quad (2.4)$$

and

$$w_i^* = \prod_{j=1}^{i-1} (I - w_j p_j') w_i. \quad (2.5)$$

(iv) Obtain the PLS-R model:

$$\begin{aligned} F_0 &= t_1 r_1' + t_2 r_2' + \dots + t_m r_m' + F_m = E_0 w_1^* r_1' + E_0 w_2^* r_2' + \dots + E_0 w_m^* r_m' + F_m \\ &= E_0 \left[\sum_{j=1}^m w_j^* r_j' \right] + F_m = E_0 B + F_m \end{aligned} \quad (2.6)$$

and

$$B = \sum_{j=1}^m w_j^* r_j', \quad (2.7)$$

where B is the regression coefficient vector of the PLS-R model, and F_m is the residual matrix.

(v) Transfer the standardized variables into original variables to obtain the final PLS-R model.

The PLS-R method was applied to obtain the compressor map regression model. As PLS-R is a multiple linear regression method, we take the speed character and pressure character, which can always be easily measured as independent variables, and the flow character and efficiency character as dependent variables, and then organize the two independent variables into a group to introduce more variables. Its combination form is as follows:

$$z = \sum_{i=1}^h \left(\sum_{k=0}^i (a \cdot v_1^{i-k} \cdot v_2^k) \right) + b, \quad (2.8)$$

where z is the flow character and efficiency character as dependent variables, v_1 and v_2 are the speed character and pressure character as independent variables, respectively, and h is the maximum fitting power of the polynomial.

Then:

$$i = 1, k = 0, \quad x_1 = v_1^{i-k} \cdot v_2^k = v_1, \quad (2.9)$$

$$i = 1, k = 1, \quad x_2 = v_1^{i-k} \cdot v_2^k = v_2, \quad (2.10)$$

$$i = 2, k = 0, \quad x_3 = v_1^{i-k} \cdot v_2^k = v_1^2, \quad (2.11)$$

.....

$$i = 1, k = m, \quad x \left[\sum_{i=1}^{i-1} \frac{(i+1)i}{2} + m + 1 \right] = v_1^{i-k} \cdot v_2^k = v_1^{1-m} \cdot v_2^m, \quad (2.12)$$

...

and
$$i = h, k = h, \quad x \left[\sum_{i=1}^h \frac{(i+1)i}{2} \right] = v_1^{i-k} \cdot v_2^k = v_2^h. \quad (2.13)$$

Here

$$p = h + \sum_{i=1}^h \left(\sum_{k=0}^i \frac{(k+1)k}{2} \right) \quad (2.14)$$

and

$$y_1 = z. \quad (2.15)$$

Then

$$\left. \begin{aligned} X &= [x_1, x_2, \dots, x_p]_{n \times p}, \\ Y &= [y_1]_{n \times 1}, \end{aligned} \right\} \quad (2.16)$$

where n is the number of sample data points.

Then, the coefficient vector a and constant b can be derived by applying X and Y into the PLS-R algorithm. Finally, the PLS-R model of the compressor characteristics maps is established.

Compared with a traditional regression method, the main features of PLS-R are as follows [30].

- (i) A strong correlation among independent variables is allowed during regression modelling.
- (ii) It has a good tolerance to measurement noise.
- (iii) Each regression coefficient for each independent variable can be interpreted in the PLS-R model.

To further match the nonlinearity of the compressor characteristics lines and to improve the prediction accuracy and shorten the calculation time, we chose the trigonometric function as the basis function of PLS ($\phi(x) = [1, \sin x, \cos x, \sin 2x, \cos 2x, \dots, \sin mx, \cos mx]^T$). This new prediction model is referred to as PLSN.

To compare the improvement of the prediction accuracy and computation time with the new model, the traditional power function polynomial-based PLS model ($\phi(x) = [1, x, x^2, \dots, x^{b-1}]^T$) is also listed in this paper, and is referred to as PLSO.

3. Application and analysis

Here, a certain compressor map obtained from a design manual of a centrifugal compressor, shown in figures 3 and 4 is used to test the effectiveness of the proposed method in which the sample data consist of nine speed lines. The speed lines of the test data are also shown in figures 3 and 4.

3.1. Flow characteristics map

To demonstrate the effectiveness of the proposed methods for the typical interpolated and extrapolated predictions, two other classical data-driven modelling methods, including the look-up table and feed-

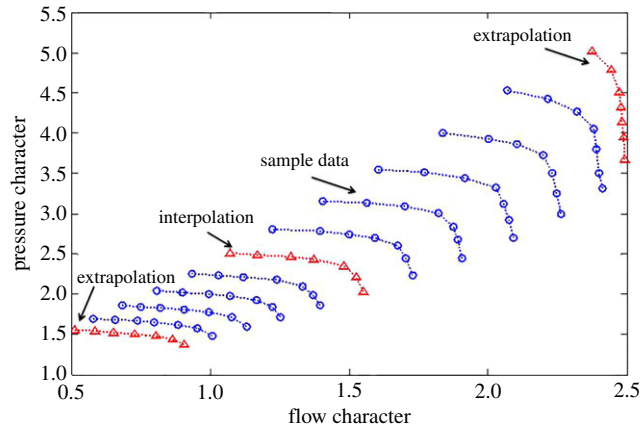


Figure 3. Flow characteristics map of a compressor.

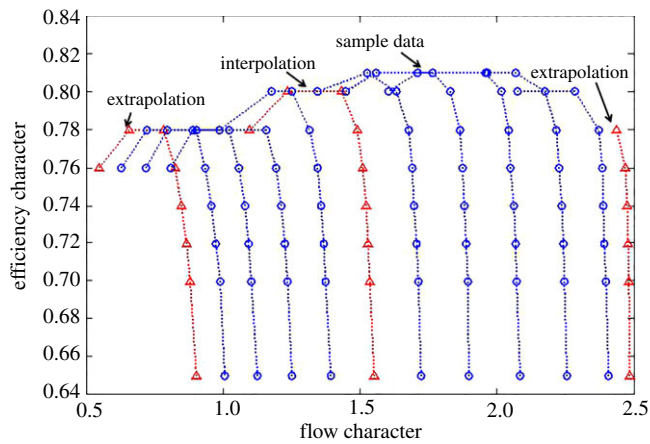


Figure 4. Efficiency characteristics map of a compressor.

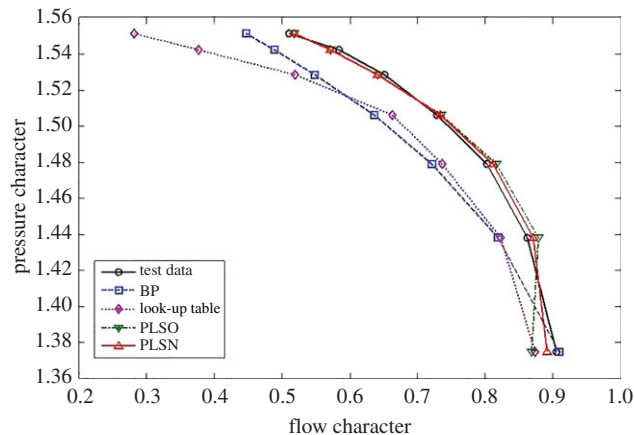


Figure 5. Extrapolation performance comparison at the lower part-load operating area.

forward BPNN, are compared from the perspective of prediction accuracy and time consumption. For the flow characteristics map, the extrapolation and interpolation performance from the traditional methods and the proposed method are shown in figures 5–7.

A criterion is shown in equation (3.1) where the root mean square error (RMSE) of the difference between the predicted data and the test data is introduced to check the interpolation and extrapolation performance of the proposed method. The comparative results are shown in tables 2 and 3.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^m [(z_{i, \text{predicted}} - z_{i, \text{test}})/z_{i, \text{test}}]^2}{m}}, \quad (3.1)$$

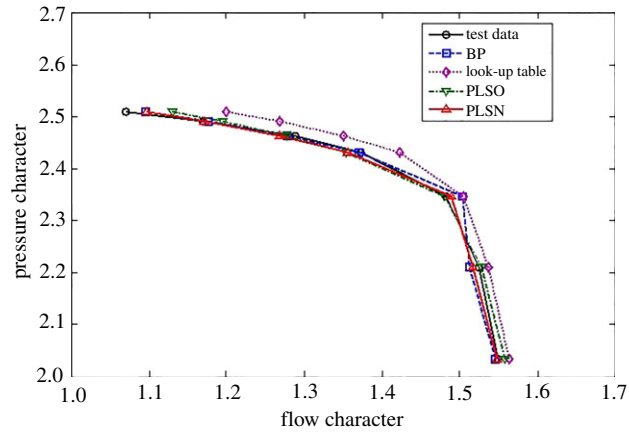


Figure 6. Interpolation performance comparison.

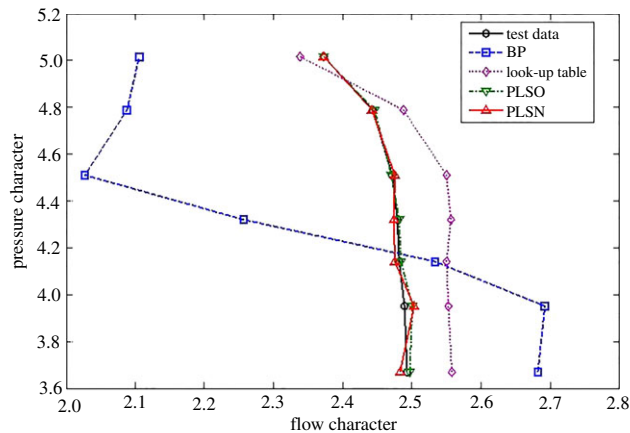


Figure 7. Extrapolation performance comparison at the upper part-load operating area.

Table 2. Extrapolation and interpolation performance comparison of the traditional methods and the proposed method for the flow characteristics map.

algorithm	RMSE			
	EPL (%)	IPL (%)	EPU (%)	total (%)
look-up table	13.3236	6.9334	6.3699	9.4193
BP	7.6496	1.4385	27.3835	16.4362
PLSO	1.7510	2.6251	0.4814	1.8429
PLSN	0.9856	1.4937	0.7414	1.1184

Table 3. Time consumption comparison of the traditional methods and the proposed method for calculating one operating point.

algorithm	look-up table	BP	PLSO	PLSN
time consumption (s)	8×10^{-4}	5.799×10^{-3}	5.1×10^{-5}	5×10^{-6}

where $z_{i,\text{test}}$ is the i th test data, $z_{i,\text{predicted}}$ is the corresponding prediction data of PLS-R model and m is the number of test data.

From figures 5–7 and table 2, the BPNN has a better interpolation performance than the other methods but also the worst extrapolation performance for the flow characteristics map, which is

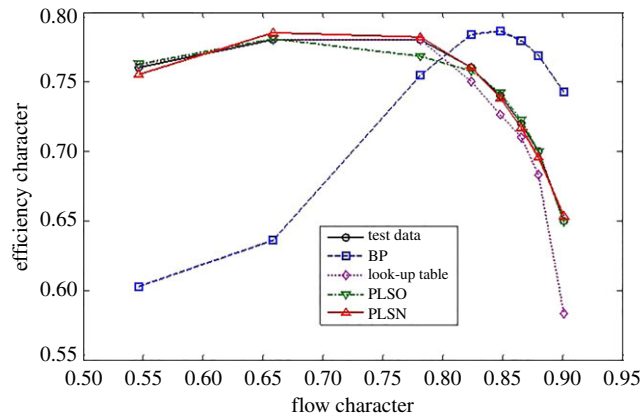


Figure 8. Extrapolation performance comparison at the lower part-load operating area.

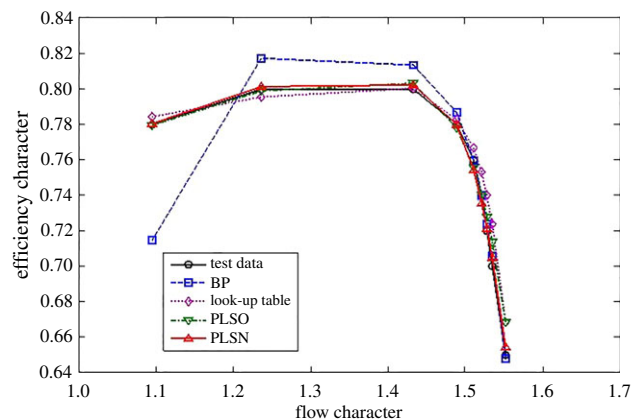


Figure 9. Interpolation performance comparison.

especially evident at the upper part-load operating area. In general, the proposed methods of PLSO and PLSN have an obvious improvement in both extrapolation and interpolation performance, especially for the extrapolation performance at the upper part-load operating area. From table 3, we can also see that the proposed method is very beneficial for engine real-time performance simulation and online monitoring. Furthermore, it is interesting to find that the PLS with the trigonometric function polynomial form as the basis function has an advantage over that with the polynomial form as the basis function, especially in computational time, which is reduced by an order of magnitude.

3.2. Efficiency characteristics map

For the efficiency characteristics map, the extrapolation and interpolation performance of the traditional methods and the proposed method are shown in figures 8–10.

The criterion shown in equation (3.1) is used to check the performance of the interpolation and extrapolation of the proposed method. The comparative results are shown in tables 4 and 5.

From figures 8–10 and table 4, the proposed methods of PLSO and PLSN have evident improvements in both the extrapolation and interpolation performance. For the interpolation performance, all four methods can predict the tendency of efficiency variations with the flow character. For the extrapolation performance, the BPNN lost the ability to predict the tendency of efficiency variations with the flow character in the lower part-load operating area, and only PLSN had the ability to predict the tendency of efficiency variations with the flow character in the upper part-load operating area. From table 5, we can also see that the proposed methods, especially PLSN, are very beneficial for engine real-time performance simulation and online monitoring because of their shorter computational time.

Furthermore, it is interesting to note that the PLS with the trigonometric function polynomial form as the basis function has several advantages over that with the power function polynomial form as the basis

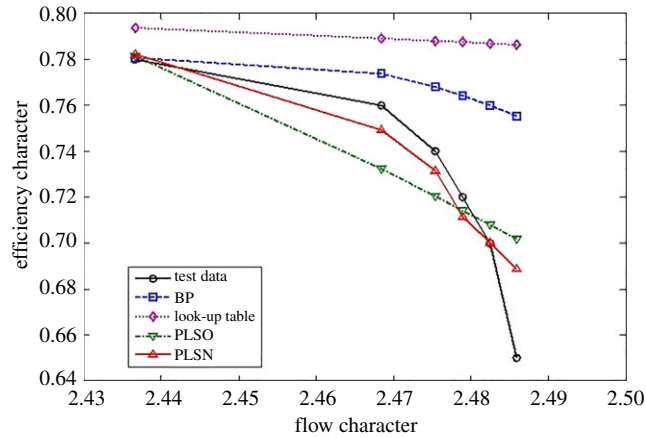


Figure 10. Extrapolation performance comparison at the upper part-load operating area.

Table 4. Extrapolation and interpolation performance comparison of the traditional methods and the proposed method for the efficiency characteristics map.

algorithm	RMSE			
	EPL (%)	IPL (%)	EPU (%)	total (%)
look-up table	2.5276	1.3200	7.5256	4.2045
BP	9.0518	2.3180	5.4155	6.1849
PLSO	0.4384	0.8336	2.5670	1.4345
PLSN	0.3500	0.3430	1.7169	0.9261

Table 5. Time consumption comparison of the traditional methods and the proposed method for calculating one operating point.

algorithm	look-up table	BP	PLSO	PLSN
time consumption (s)	1.12×10^{-4}	6.118×10^{-3}	5.0×10^{-5}	7×10^{-6}

function. One is that the PLSN has a higher prediction accuracy than the PLSO despite the interpolation and extrapolation; the other is that the PLSN can predict the tendency of efficiency variation with the flow character successfully at the upper part-load operating area where PLSO had lost this ability; even more surprising is that the computational time of PLSN is much shorter than that of PLSO.

4. Conclusion and discussion

A multiple linear regression method, i.e. the PLS-R method, was proposed in this paper and applied to obtain the compressor map regression model based on known compressor characteristics maps. The following meaningful conclusions were determined.

- (i) The proposed method has an evident improvement in both extrapolation and interpolation performance.
- (ii) PLS with the trigonometric function polynomial form as the basis function has the advantage of both expressive precision and time consumption over that with the power function polynomial form as the basis function.
- (iii) Compared with [25,26], the new PLS has a similar or even slightly improved prediction accuracy. It also avoids the complex structure of the ANN-PLS and complicated kernel function of KPLS.

- (iv) The PLS-R model may improve the real-time calculating performance during the dynamic performance simulation of a diesel engine performance model.
- (v) It can be expected that the application of PLS-R modelling has a certain reference value to improve the accuracy for thermodynamic performance modelling of a diesel engine.

Data accessibility. Our primary data can be found at the website: <http://www.jtp.com.cn/cpzx/zyq/jthxl/223820.htm>. Authors' contributions. X.L. and Y.W. designed this study. H.W. and C.Y. prepared all the sample data. X.L. collected and analysed the data. X.L., Y.W., H.W., C.Y., X.Z. and Y.S. interpreted the results and wrote the manuscript. S.H. and X.L. write this manuscript. All authors gave final approval for publication.

Competing interests. The authors declare that there is no conflict of interest regarding the publication of this article.

Funding. There is no financial funding support for our research.

Acknowledgements. We thank the Chongqing Jingjin Shipbuilding Industry Co., Ltd for supporting the essential data.

References

1. Emmenthaler KD, Hagemann G, Hucho WH. 1979 Turbocharging small displacement spark ignition engines for improved fuel economy. SAE Technical Paper 790311. See <https://www.sae.org/publications/technical-papers/content/790311/>.
2. Guzzella L, Wenger U, Martin R. 2000 IC-engine downsizing and pressure-wave supercharging for fuel economy. SAE Technical Paper 2000-01-1019. See <https://www.sae.org/publications/technical-papers/content/2000-01-1019>.
3. Petitjean D et al. 2004 Advanced gasoline engine turbocharging technology for fuel economy improvements. SAE Technical Paper 2004-01-0988. See <https://www.sae.org/publications/technical-papers/content/2004-01-0988/>.
4. Watson N, Janota MS. 1982 *Turbocharging the internal combustion engine*. London, UK: Macmillan Education.
5. Pesiridis A, Salim SIW, Martinez-Botas RF. 2012 Turbocharger matching methodology for improved exhaust energy recovery. In *Proc. of the 10th Int. Conf. on Turbocharging and Turbochargers*, pp. 203–218. See <https://bura.brunel.ac.uk/handle/2438/14007>.
6. Tsoutsanis E, Meskin N, Benammar M, Khorasani K. 2014 A component map tuning method for performance prediction and diagnostics of gas turbine compressors. *Appl. Energy* **135**, 572–585. (doi:10.1016/j.apenergy.2014.08.115)
7. Samarasinghe S. 2006 *Neural networks for applied sciences and engineering*, pp. 484–494. Boca Raton, FL: AUerbach Publications.
8. Zhao N, Wen X, Yang J, Li S, Wang Z. 2015 Modeling and prediction of viscosity of water-based nanofluids by radial basis function neural networks. *Powder Technol.* **281**, 173–183. (doi:10.1016/j.powtec.2015.04.058)
9. Wuhan A. 2004 New approach of characteristic map calculating of compressor based on neural network. *Comput. Digit. Eng.* **32**, 46–47.
10. Yu Y et al. 2007 Neural-network based analysis and prediction of a compressor's characteristic performance map. *Appl. Energy* **84.1**, 48–55. (doi:10.1016/j.apenergy.2006.04.005)
11. Sanaye S, Dehghandokht M, Mohammadbeigi H, Bahrami S. 2011 Modeling of rotary vane compressor applying artificial neural network. *Int. J. Refrig.* **34**, 764–772. (doi:10.1016/j.ijrefrig.2010.12.007)
12. Ławryńczuk M. 2007 Jet engine turbine and compressor characteristics approximation by means of artificial neural networks. In *Proc. of the 2007 Adaptive and Natural Computing Algorithms, International Conference, Icannga 2007, Warsaw, Poland, 11–14 April 2007*, pp. 143–152. Berlin, Germany: Springer-Verlag.
13. Fei J, Zhao N, Shi Y, Feng Y, Wang Z. 2016 Compressor performance prediction using a novel feed-forward neural network based on Gaussian kernel function. *Adv. Mech. Eng.* **8**, 1–14.
14. Ghorbanian K, Gholamrezaei M. 2009 An artificial neural network approach to compressor performance prediction. *Appl. Energy* **86**, 1210–1221. (doi:10.1016/j.apenergy.2008.06.006)
15. Gholamrezaei M, Ghorbanian K. 2010 Compressor map generation using a feed-forward neural network and rig data. *Proc. Inst. Mech. Eng. Part A J. Power Energy* **224**, 97–108. (doi:10.1243/09576509JPE792)
16. Rinaldi E, Pecnik R, Colonna P. 2014 Numerical computation of the performance map of a supercritical CO₂ radial compressor by means of three-dimensional CFD simulations. In *Proc. ASME Turbo Expo 2014: Turbine Technical Conference and Exposition*, pp. V03BT36A017–V03BT36A017. New York, NY: American Society of Mechanical Engineers.
17. Japikse D. 2009 Turbomachinery performance modeling. SAE Technical Paper 2009-01-0307. See <https://www.sae.org/publications/technical-papers/content/2009-01-0307>.
18. Palfreyman D, Martinez-Betas RF. 2004 The pulsating flow field in a mixed flow turbocharger turbine: an experimental and computational study. *J. Turbomach.* **127.1**, 144–155. (doi:10.1115/1.1812322)
19. Galindo D, Fajardo P, Navarro R. 2013 Characterization of a radial turbocharger turbine in pulsating flow by means of CFD and its application to engine modeling. *Appl. Energy* **103**, 116–127. (doi:10.1016/j.apenergy.2012.09.013)
20. Bohn D et al. 1998 A comparative throughflow analysis of axial flow turbines. *Proc. Inst. Mech. Eng. Part A J. Power Energy* **212**, 141–145.
21. Baines NC. 1998 A meanline prediction method for radial turbine efficiency. *IMECHE Conf. Trans.* **11**, 45–56.
22. Zhao CH, Yao Y, Gao FR, Wang FL. 2010 Statistical analysis and online monitoring for multimode processes with between-mode transitions. *Chem. Eng. Sci.* **65**, 5961–5975. (doi:10.1016/j.ces.2010.08.024)
23. Ying Y et al. 2014 Study on the regression method of compressor map based on partial least squares regression modeling. In *Proc. ASME Turbo Expo 2014: Turbine Technical Conference and Exposition*, pp. V01AT01A015–V01AT01A015. New York, NY: American Society of Mechanical Engineers.
24. Peng S et al. 2014 Accuracy research on the modeling methods of the gas turbine components characteristics. In *Proc. ASME Turbo Expo 2014: Turbine Technical Conference and Exposition*, pp. V03AT09A003–V03AT09A003. New York, NY: American Society of Mechanical Engineers.
25. Tian Z, Gu B, Yang L, Lu Y. 2015 Hybrid ANN–PLS approach to scroll compressor thermodynamic performance prediction. *Appl. Therm. Eng.* **77**, 113–120. (doi:10.1016/j.applthermaleng.2014.12.023)
26. Chu F, Wang F, Wang X, Zhang S. 2012 Performance modeling of centrifugal compressor using kernel partial least squares. *Appl. Therm. Eng.* **44**, 90–99. (doi:10.1016/j.applthermaleng.2012.03.043)
27. Ying Y, Cao Y, Li S, Li J. 2015 Nonlinear steady-state model based gas turbine health status estimation approach with improved particle swarm optimization algorithm. *Math. Probl. Eng.* **2015**, 1–12.
28. Li YG, Ghafir MFA, Wang L. 2012 Creep life prediction for aero gas turbine hot section component using artificial neural networks. *J. Eng. Gas Turbines Power* **136.3**, 195–205.
29. Wang ZT, Shu-Ying LI, Tan ZY. 2009 Comparing study on the fitting method of compressor characteristic curve based on BP and RBF neural network. *Turbine Technol.* **51**, 94–96.
30. Geladi P, Kowalski BR. 1986 Partial least-squares regression: a tutorial. *Anal. Chim. Acta* **185**, 1–17. (doi:10.1016/0003-2670(86)80028-9)