

Multiphase gas-flow model of an electrical submersible pump

Diana Marcela Martinez Ricardo^{1,2,*}, German Efrain Castañeda Jiménez^{1,2}, Janito Vaqueiro Ferreira¹, and Pablo Siqueira Meirelles¹

¹ Department of Computational Mechanics, UNICAMP-University of Campinas, São Paulo, Brazil

² Students-Agreement Graduate - PEC-PG, CAPES/CNPq Brazil

Received: 7 December 2016 / Accepted: 2 July 2018

Abstract. Various artificial lifting systems are used in the oil and gas industry. An example is the Electrical Submersible Pump (ESP). When the gas flow is high, ESPs usually fail prematurely because of a lack of information about the two-phase flow during pumping operations. Here, we develop models to estimate the gas flow in a two-phase mixture being pumped through an ESP. Using these models and experimental system response data, the pump operating point can be controlled. The models are based on nonparametric identification using a support vector machine learning algorithm. The learning machine's hidden parameters are determined with a genetic algorithm. The results obtained with each model are validated and compared in terms of estimation error. The models are able to successfully identify the gas flow in the liquid-gas mixture transported by an ESP.

1 Introduction

Different artificial lifting methods that ensure the flow of oil through a production line are used in oil and gas production systems to reduce production losses. One such method involves the use of Electrical Submersible Pumps (ESPs), which were developed in 1910 by Annals Arutunoff [1]. ESPs are centrifugal pumps that operate under multiphase flow conditions and are positioned at the end of the production piping. They are driven by a three-phase electric motor and are suitable for use in wells with high-viscosity fluids, high flow rates, high water content and a low gas-oil ratio. To ensure smooth operation of the pump, the gas-oil ratio should not exceed 10%. Exceeding this ratio can cause pumping to stop because of a phenomenon known as gas-lock, a major limitation of ESPs [2]. To avoid this phenomenon and consequent pump failure, the percentage of gas in the ESP when it is operating under multiphase flow conditions (specifically a gas-liquid mixture) must be known. One way to avoid the problem is by stopping the pump or changing the pump operating conditions. Various strategies for monitoring and analyzing the operation of ESPs have also been developed in an attempt to circumvent the gas-lock problem. A traditional approach is to use ammeter charts, which can be used to identify operating problems. By comparing an ammeter chart showing the equipment operating normally and another recorded when an inspection is being carried out, the state of the equipment during the inspection can be determined, as this is

reflected in changes in the current during the measurement period [1]. Another method for identifying a gas-lock is to monitor the output flow from the pump so that the pump can be turned off when there is no fluid in it [3]. Monitoring of this kind, however, requires premature shutdown of the pump, potentially reducing well productivity. Other strategies that have been developed involve characterizing the flow by analyzing signals from measuring instruments installed in the pipeline [4] and studying the fluid frequency response by measuring the vibration of the pump [5, 6]. However, these methods can suffer from drawbacks depending on where and how the instrumentation is installed in the pipeline.

As the gas-lock phenomenon in this context is not well understood and theoretical models are difficult to develop because of the complexity of the process, experimental studies are required to enable a model to be developed that allows the gas flow in these systems to be estimated. Such a model would also provide a better understanding of the real dynamics of the system of interest and be an important tool for estimating and analyzing the behavior of the system parameters and so preventing possible failures [7].

System identification is used to develop mathematical models based on a data set obtained from the system of interest. The methodology has been evolving for several decades [8]. In the 90s new areas of interest were considered, including system identification in the frequency domain, leading to nonparametric identification, which is currently widely used. The number of different methodologies being used is increasing steadily.

One of these new methodologies is based on machine learning, which involves various computational techniques

* Corresponding author: dmartinez628@gmail.com

that aim to increase the automation and efficiency of knowledge acquisition processes through the use of data processing. The main objective is to find the relationship between the system variables (input/output) using samples acquired from the system [9]. One of the many applications of machine learning is the identification of linear and nonlinear systems, a task at which Support Vector Machines (SVMs) are known to be particularly efficient compared with other learning methods [10]. Support Vector Machines are based on the structural risk minimization principle, which originated from statistical learning theory and was proposed by Vapnik. This principle has been shown to yield better results than the principle of empirical risk minimization used in neural networks [11].

In light of the increasing use of SVMs in different areas of research, the new methodologies that have been developed in an attempt to improve the operation and performance of SVMs and the good results obtained with these algorithms, an SVM was used here as the basis of a methodology for identifying the gas flow in an ESP.

2 Optimization of SVMR parameters with genetic algorithms

To develop a model that identifies the gas flow in an ESP operating with a two-phase (liquid-gas) fluid from data acquired directly from the system, a nonparametric identification method based on machine learning was developed. Specifically, a Support Vector Machine for Regression (SVMR) was used and the SVMR parameters were optimized with Genetic Algorithms (GAs).

Support Vector Machines are learning systems based on optimization tools that seek to minimize structural risk. They use a hypothesis space of linear functions in a high-dimensional space created by a kernel function. These transformations can be used in various learning problems, usually either classification or regression problems [12].

Support Vector Machine for Regression aims to estimate a function $f(x)$ where the output y_i will depend on an input x_i . Given a data approximation problem involving a set of training data x_i, y_i where $x \in \mathbb{R}^a$ and $x \in \mathbb{R}$, the objective is to find a linear function $f(x)$ that is an approximation to the system using a vector of minimum weights w ,

$$\begin{aligned} f(x) &= \langle w \cdot x \rangle + b \\ &= x^T w + b \\ &= \sum_{i=1}^n w_i x_i + b \end{aligned} \tag{1}$$

where the training data sets are $x = x_1, x_2, \dots, x_n$ and $y = y_1, y_2, \dots, y_n$, the weight vector is $w = w_1, w_2, \dots, w_n$ and b is the bias.

To achieve this objective, the vector w must be minimized using the Euclidean norm $\|w\|^2$. The problem can then be defined as the optimization problem

$$\min_w \frac{1}{2} \|w\|^2 \tag{2}$$

subject to

$$|y_i - (\langle w \cdot x_i \rangle + b)| \leq \varepsilon, \quad i = 1, 2, \dots, m \tag{3}$$

or

$$\begin{cases} y_i - (\langle w \cdot x_i \rangle + b) \leq \varepsilon, \\ y_i - (\langle w \cdot x_i \rangle + b) \geq -\varepsilon. \end{cases} \tag{4}$$

The system loss function is estimated by the parameter (ε), which introduces a degree of tolerance when samples are penalized. As the linear function $f(x)$ may not be able to fit all the training data, slack variables (ξ_i and ξ_i^*) are added to allow some errors subject to the following approximation conditions:

$$\begin{cases} y_i - (\langle w \cdot x_i \rangle + b) \leq \varepsilon + \xi_i, \\ y_i - (\langle w \cdot x_i \rangle + b) \geq -\varepsilon + \xi_i^*. \end{cases} \tag{5}$$

Finally, the SVMR problem can be described by equation (6), where the goal is to find w and b that minimize w [13]:

$$\begin{aligned} \min_w & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i + \xi_i^* \\ \text{subject to} & \begin{cases} y_i - (\langle w \cdot x_i \rangle + b) \leq \varepsilon + \xi_i \\ y_i - (\langle w \cdot x_i \rangle + b) \geq -\varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \tag{6}$$

where $C > 0$ is the parameter that penalizes permissible errors.

If the system is not linear, the training data $x \in \mathbb{X}$ belonging to space \mathbb{X} must be mapped to a higher-dimension space \mathbb{F} using a function ϕ called a kernel:

$$\phi : \mathbb{X} \rightarrow \mathbb{F}. \tag{7}$$

The problem discussed in this paper can be represented by a nonlinear system, and the kernel used to perform higher-dimensional mapping is a Gaussian function. The parameter to be optimized with a GA is σ^2 , which determines the width of the function. The kernel function is given by:

$$\phi(x, x_i) = e^{-\frac{(x-x_i)^2}{2\sigma^2}}. \tag{8}$$

Figure 1 shows an example of transformation of the mapping space by the kernel function to a larger space where the nonlinear system behaves as if it were linear.

Equation (6) can be solved using dual programming [14]. The goal is to build an objective function by adding a set of variables α_i called Lagrange multipliers, where the dual formulation is given by

$$\begin{aligned} \max & \left[-\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \phi(x_i, x_j) \right. \\ & \left. - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n y_i (\alpha_i + \alpha_i^*) \right] \\ \text{subject to} & \begin{cases} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \end{aligned} \tag{9}$$

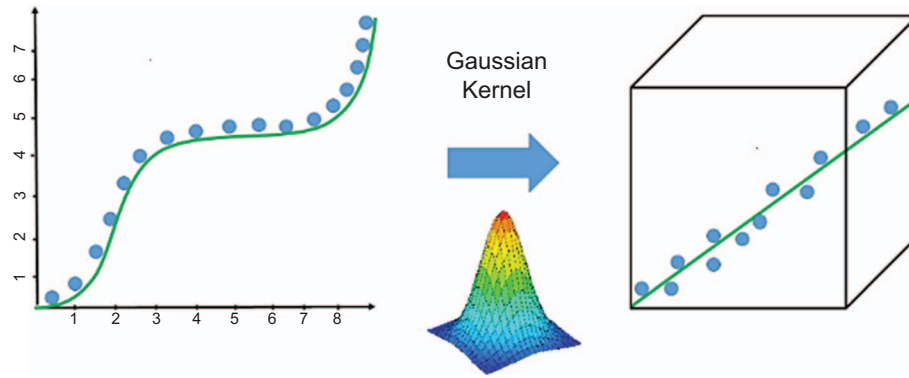


Fig. 1. Mapping the vectors into higher-dimensional space.

After solving the dual problem the optimal decision is given by

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \phi(x_i, x_j) + b. \quad (10)$$

Correct selection of the hidden parameters of an SVM ensures that the models developed using the SVM perform satisfactorily. The parameters are C , which determines the curvature of the margin penalizing permissible errors, ϵ , which defines the insensitivity of the margin, and σ^2 , which determines the width of the kernel function. In this article, the use of a GA is proposed to search for the best hidden parameters of the SVMR.

Genetic Algorithms provide a robust method for performing searches and optimization. They mimic the evolutionary processes of living things and can provide solutions for a wide range of problems. Because of their adaptability, GAs have been used by many researchers to look for the hidden parameters of SVMRs for various problems [15, 16]. The flowchart of the GA used in this paper is shown in Figure 2.

To find the best parameters (C, ϵ, σ^2) , the algorithm must perform the following four steps. The first step consists of creating a population of individual elements $p^k = \{x_1^t \cdots x_n^t\}$ or $p^k = \{(C, \epsilon, \sigma^2)_1^t \cdots (C, \epsilon, \sigma^2)_n^t\}$ that are candidates for a possible solution. In the second step, the adaptation of each of the individual elements to the problem being studied is measured using a fitness function. In this particular case, the validation error of each of the models built with the SVMR is used as the fitness criterion, and the objective function is given by

$$\text{Fitness} = \max \frac{1}{e_{\text{val}}} \quad (11)$$

$$\therefore e_{\text{val}} = \frac{1}{2n_{\text{val}}} \sum_{i=1}^{n_{\text{val}}} (f(x_{\text{val}})_i - y_{\text{val}})^2$$

where n_{val} is the number of validation samples.

In the third step, the individuals that are better adapted to the problem are selected. This is done by the roulette method, which not only ensures a higher probability of better-adapted individuals being selected, but also allows individuals that are less well adapted to the problem to

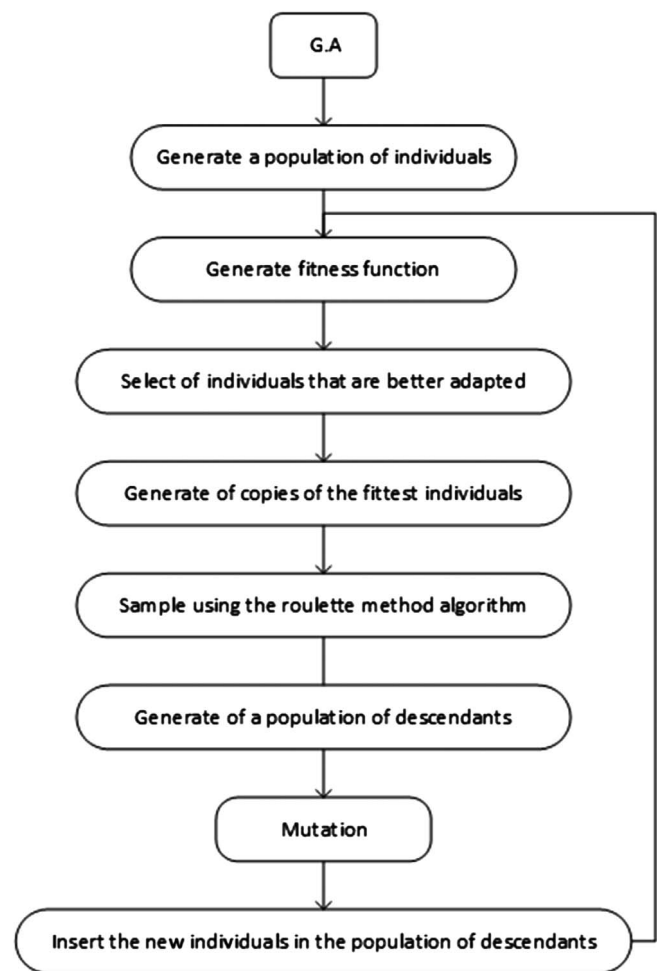


Fig. 2. Flowchart of the genetic algorithm.

be selected, helping to maintain the diversity of the population of solutions and so avoid convergence to local minima.

In the fourth step, a crossover is performed between two or more individuals selected in the previous stage (parents) to generate descendants with part of the genetic material from each parent. For the crossover to be performed, all individuals must be coded in binary representation using

the same number of bits for each parameter [*e.g.*, $P = (C, \epsilon, \sigma^2) = (2,4,3) \rightarrow P_{\text{bit}} = (0010 \ 0100 \ 0011)$]. Hence, if a parameter is represented by b bits, the genome size is $l = 3b$. In the crossover process, a random point k is selected and the bits to the right of this are swapped between the parents to create descendants [*e.g.*, if the parents are $P_1 = (0010 \ 0100 \ 0011)$ and $P_2 = (0101 \ 0001 \ 1111)$, then for $k = 5$ we have $S_1 = (0010 \ 0001 \ 1111)$ and $S_2 = (0101 \ 0100 \ 0011)$ as descendants]. This crossover process is repeated until the population of descendants is the same size as the initial population. The algorithm includes a final step called the mutation operator that is intended to maintain population diversity. This operator randomly selects some of the individuals generated and performs a mutation in their genomes by changing one bit at random [*e.g.*, a selected individual $S = (0010 \ 0001 \ 1111)$ becomes $S_{\text{mutate}} = (0010 \ 0001 \ 1011)$ after the mutation]. The above steps are repeated until the stop condition is satisfied.

3 Results

An SVMR was used to develop a nonparametric black-box model for identifying the gas flow in a J200N ESP based on different characteristic system parameters. Casing vibration, total flow, torque and elevation were used as inputs, and a graphical representation of estimated gas flow in the ESP was used as output, since these are the data that are available in the ESP circuit tests. Several experiments were performed in the ESP test circuit in the LABPETRO laboratory at Unicamp using the initial operating conditions in Table 1. The ESP system was tested at 1800 rpm and 3000 rpm with manometric inlet pressures of 100 bar and 200 bar, respectively, and gas flows of 0–3 kg h⁻¹.

After the data used to estimate the model parameters had been acquired, the signals were post-processed. This is necessary because the acquired ESP casing vibration signals contain noise and are difficult to predict in the time domain even though the initial test-circuit conditions are known. A spectral representation of the signals was therefore used so that they could be characterized by their power spectral density. This was done with a finite Fourier transform and the autocorrelation function, a technique typically used to analyze measurement signals when working with digital equipment and discrete algorithms, as in this case [17].

After the spectral representation had been generated, the autocorrelation of the power spectral density was calculated for specific positive frequencies corresponding to the pump speeds and their respective multiples. For example, for a pump speed of 1800 rpm, the frequencies evaluated were 30, 60, 90 and 120 Hz, as these are the frequencies at which there are significant changes in the spectral representation when the gas flow in the ESP changes. In this way, a new set of data is acquired that can be used as inputs in the system identification. The patterns of the spectral representations of the signals associated with the casing vibration variable can be seen in Figure 3.

Ten different models were built using a variety of signals as input, and once the parameters had been estimated the

Table 1. ESP test-circuit configuration for data acquisition.

Test	Pump speed (rpm)	Inlet pressure (bar)	Gas flow (kg h ⁻¹)
1	1800	100	0
		200	1
			2
2	3000	100	3
		200	0
			1
			2
			3

models were tested. The different sets of training samples used for each model are shown in Table 2. The objective of these 10 sets is to be able to select as input the best combination of the data acquired from the test system to be able to identify the gas flow in the EPS and thus obtain models that represent the variations of the percentage of gas in the pump.

Having defined and acquired the model inputs and outputs, we used a genetic algorithm to estimate the best parameters (C, ϵ, σ^2) and so identify the most representative model for each set of training samples. Various tests were performed for each of the ten models in Table 2. The configuration of the GA for each of the tests is shown in Table 3.

The results of the tests with each of the models using the GA configurations in Table 3 are shown in Table 4. Selection of the parameters to be used with the SVMR was based on analysis of the Mean Square Error (MSE) and autocorrelation coefficient. The parameters that yielded the lowest MSE and highest autocorrelation coefficient were selected and then used in the model to predict the ESP gas flow. Table 4 shows the configurations of the GA that yielded the best results and the corresponding SVMR parameters, MSE and autocorrelation coefficient for each model.

Once the input and output data for each model have been acquired and the SVMR parameters have been estimated by the GA, the gas flow in the ESP can be estimated using different characteristic system parameters. The power of generalization of each model was determined by measuring the performance of the model with a data set not included in the training data used to build the model. This test data set corresponded to 20% of the total data used and the rest of the data is used for training [18].

To evaluate the ten SVMR models, the actual gas flow was compared with the gas flow predicted by each model. For each of the four flows, a mean was estimated for all the test samples and the MSE was calculated for each model so that their performance could be compared. Table 5 shows the gas flows estimated by the SVMR models and the corresponding MSE.

To evaluate the performance of the models, the gas flows estimated by each model were compared. Figure 4 shows the flows estimated by each of the models for actual

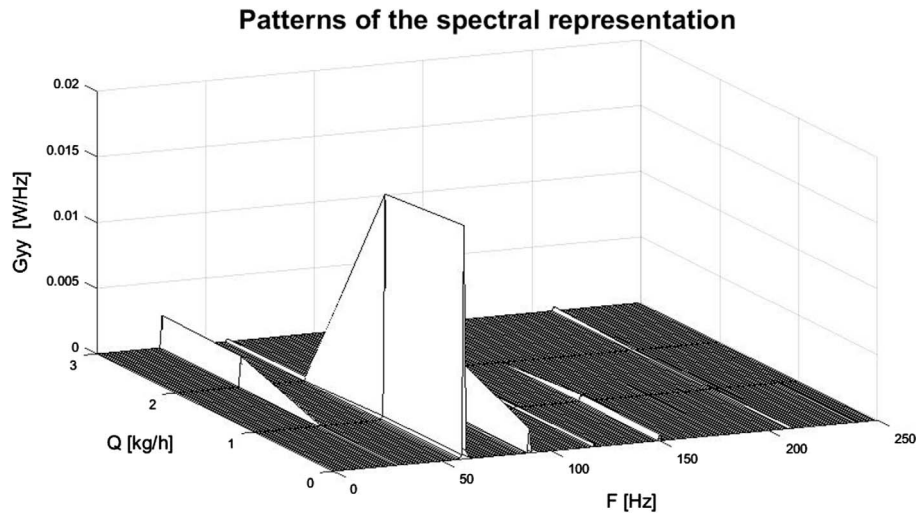


Fig. 3. Patterns of the spectral representation of casing vibration.

Table 2. Input and output data used with each model.

Model	Input	Output
1	Spectral representation of the ESP vibration signal and total flow	Gas flow
2	Spectral representation of the ESP vibration signal and torque	Gas flow
3	Spectral representation of the ESP vibration signal and elevation	Gas flow
4	Total flow and torque	Gas flow
5	Total flow and elevation	Gas flow
6	Torque and elevation	Gas flow
7	Spectral representation of the ESP vibration signal	Gas flow
8	Total flow	Gas flow
9	Torque	Gas flow
10	Elevation	Gas flow

Table 3. Configuration of the genetic algorithm in each test.

Config	Iter ^a	Ind ^b	Genes ^c	P_{cross} ^d	P_{mut} ^e
1	100	10	10	50	10
2	500	10	10	50	10
3	1000	10	10	50	10
4	100	100	10	40	20
5	500	100	10	40	20
6	1000	10	10	40	20
7	1000	100	10	60	10
8	50	10	10	60	10

^a No of iterations.

^b No of individuals.

^c No of genes.

^d Probability of crossing.

^e Probability of mutation.

Table 4. Selected SVMR parameters.

Model	SVMR parameters					
	GA_c^a	C	σ^2	ϵ	MSE ^b	r^c
1	7	20.698	0.127	0.0011	0.0183	0.98
2	7	15.934	0.125	0.0022	0.0844	0.93
3	3	473.793	0.345	0.0745	0.1023	0.89
4	2	547.638	30.990	0.0078	0.0761	0.96
5	4	14.322	1.667	1.6673	0.0027	0.99
6	2	55.237	2.193	0.0170	0.0582	0.95
7	4	4.061	0.463	0.0924	0.0227	0.98
8	4	8.756	30.857	0.2394	0.9309	0.59
9	6	777.785	31.495	0.2435	0.1610	0.83
10	7	903.983	1.226	0.2381	0.4413	0.71

^a GA configuration.

^b Training MSE.

^c Autocorrelation coefficient.

flows of 0–3 kg h⁻¹. These results show that all the models have a good power of generalization but that the model with the best power of generalization is model 5, which is based on total flow and elevation. The gas flow estimated with this model has the lowest error of all the models.

The model chosen to predict the gas flow in the ESP was the model with the lowest MSE and highest autocorrelation coefficient.

The results in Table 6 show that model 5, which uses total flow and elevation as input parameters, yielded the best results ($MSE = 0.003$ and autocorrelation coefficient = 0.997) and that the experimental data provided good information about changes in the gas flow in the pump, as the errors in the predicted values are small for most of the models. The findings also show that when the training data is of poor quality, the models do not produce good estimates.

Table 5. Gas flow estimated by each SVMR model.

Model	Actual gas flow	Gas flow estimated by the SVMR model	MSE
1	0	0.0414	0.0195
	1	0.8545	0.0370
	2	1.9475	0.0115
2	3	3.0507	0.0028
	0	-0.0659	0.0109
	1	1.2977	0.2408
3	2	1.8109	0.0388
	3	2.9372	0.0142
	0	0.1596	0.0395
4	1	1.1705	0.1425
	2	1.6571	0.1714
	3	3.0024	0.0001
5	0	0.2242	0.1019
	1	1.2975	0.1124
	2	2.1347	0.0409
6	3	2.9210	0.0062
	0	0.0010	0.0008
	1	0.9822	0.0059
7	2	2.0020	0.0006
	3	2.9639	0.0020
	0	0.0910	0.0658
8	1	0.7639	0.0994
	2	1.7985	0.0541
	3	3.0206	0.0009
9	0	0.1486	0.0226
	1	1.0421	0.0474
	2	1.9836	0.0002
10	3	2.9975	0.0069
	0	0.6648	1.1373
	1	1.0858	0.2538
1	2	3.2097	1.9340
	3	2.7668	0.054
	0	0.2432	0.0591
2	1	1.1819	0.2356
	2	1.8475	0.209
	3	2.7749	0.0513
3	0	0.3933	0.2680
	1	1.5125	0.4626
	2	1.8016	0.1280
4	3	2.2531	0.7857

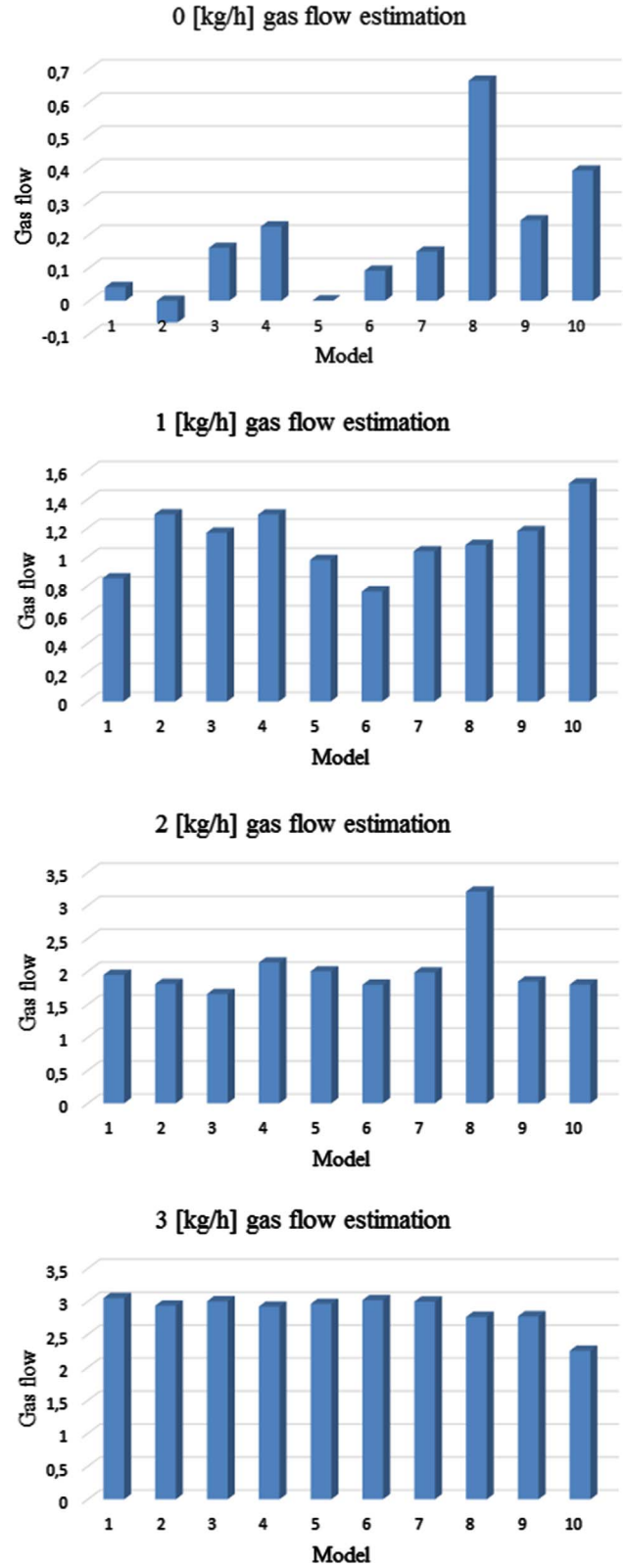


Fig. 4. Gas flow estimates for each model.

Table 6. MSE and autocorrelation coefficient of the SVMR models.

Model	MSE	Autocorrelation coefficient
1	1.84E ⁻⁰²	0.985
2	6.16E ⁻⁰²	0.954
3	0.102	0.896
4	0.076	0.967
5	0.0027	0.997
6	0.058	0.957
7	0.022	0.983
8	0.930	0.597
9	0.161	0.830
10	0.441	0.710

4 Conclusion

We have proposed different nonparametric VSMR models for estimating the gas flow in an ESP. The models use experimental vibration, total-flow, torque and elevation data collected from the pump and pipe system.

The hidden parameters of the SVMR were estimated using a GA, thus ensuring that the models have good power of generalization and, in turn, perform well. This was confirmed by testing the models with data that were not used during training. The models can be used to solve the problem of premature failure due to high gas flow in ESP systems. This problem arises because of insufficient information about the multiphase nature of the flow in the pump and can be easily avoided just by changing the pump operating conditions.

One of the most significant findings of this study is that total flow and elevation yielded the most accurate estimates of gas flow, as the models using these parameters had the lowest MSE (0.003) and highest autocorrelation coefficient (0.997).

References

- Tákač G. (2009) *Electrical submersible pumps manual: design, operations, and maintenance*, Gulf Professional Publishing.
- Varon M.P. (2013) Estudo de uma bomba centrífuga submersa (BCS) como medidor de vazão, *Msc Thesis*, FEM-UNICAMP, Campinas, Brazil.
- Snyder J.R., Dale R., Joe Haws H. (1989) Pump off/gas lock motor controller for electrical submersible pumps, EP Patent App. EP19,880,202,383, 03 May.
- Klein F.L., Selegheim, Jr. P., Eric H. (2004) Time-frequency analysis of intermittent two-phase flows in horizontal piping, *J. Braz. Soc. Mech. Sci. Eng.* **26**, 2, 174–179.
- Kolpak M.K., Rock T.J. (1996) Measuring vibration of a fluid stream to determine gas fraction, *U.S. Patent No. 5,524,475*, 11 June.
- Henry M.P., Richard R.P. (2014) Multiphase flow metering system, *U.S. Patent Application No 14/135,085*, 19 Dec.
- Nelles O. (2001) *Nonlinear system identification*, Springer-Verlag, Berlin Heidelberg.
- Ljung L. (1999) *System identification: Theory for the user*, Prentice Hall PTR, Upper Saddle River, NJ.
- Dietterich T.G. (1986) Learning at the knowledge level, *Mach. Learn.* **1**, 287–315.
- Rychetsky M. (2001) *Algorithms and architectures for machine learning based on regularized neural networks and support vector approaches*, Shaker Verlag GmbH, Germany.
- Vapnik V. (1995) *The nature of statistical learning theory*, Springer-Verlag, New York.
- Vapnik V., Golowich S., Steven E., Smola A. (1996) Support vector method for function approximation, regression estimation, and signal processing, *Advances in Neural Information Processing Systems 9*, pp. 281–287.
- Stoean R., Dumitrescu D., Preuss M., Stoean C. (2006) Evolutionary support vector regression machines, in: *SYNASC '06 Proceedings of the Eighth International Symposium on Symbolic and Numeric Algorithms for Scientific Computing*, IEEE Computer Society, Washington, DC, pp. 330–335.
- Zhang H., Yongmei L. (2013) BSP-Based Support Vector Regression Machine Parallel Framework, *Computer and Information Science (ICIS), 2013 IEEE/ACIS 12th International Conference on. IEEE, Niigata, Japan*, pp. 329–334.
- Yu Q., Liu Y., Rao F. (2009) Parameter selection of support vector regression machine based on differential evolution algorithm. Fuzzy systems and knowledge discovery, de FSKD '09. Sixth International Conference, pp. 596–598.
- Zhenyue H., Mei C. (2009) Soft sensor modeling using SVR based on genetic algorithm and akaike information criterion, intelligent human-machine systems and cybernetics, *Intelligent Human-Machine Systems and Cybernetics, 2009. IHMSC'09. International Conference on*, Vol. 2, IEEE.
- Blackman R.B., Tukey J.W. (1959) *The measurement of power spectra*, Dover publications, New York.
- Michael K. (1996) A bound on the error of cross validation using the approximation and estimation rates, with consequences for the training-test split, *Advances in Neural Information Processing Systems*, pp. 183–189.