International Journal of Mechanical Engineering and Technology (IJMET)
Volume 9, Issue 9, September 2018, pp. 1516–1526, Article ID: IJMET_09_09_166
Available online at http://www.iaeme.com/ijmet/issues.asp?JType=IJMET&VType=9&IType=9
ISSN Print: 0976-6340 and ISSN Online: 0976-6359

THE SELECTION OF OPTIMAL CONTROL OF THE OPERATION MODES OF HETEROGENEOUS DUPLICATING EQUIPMENT BASED ON STATISTICAL MODELS WITH LEARNING

L.A. Mylnikov, M.V. Kulikov
Perm National Research Polytechnic University
614990 Perm, Russia, 29 Komsomolskiy Ave

B. Krause
Anhalt University of Applied Sciences
06366 Koethen, Germany, Lohmannstrasse 23

ABSTRACT

The article discusses the use of a statistical approach to select the mode of joint operation of a group of machines for the steam production for specific needs. The article discusses the optimal control model based on functionally defined dependencies characterizing the performance of boilers, obtained from statistical data. The resulting model is related to quadratic programming problems. The proposed approach, when compared with retrospective data, shows the possibility of saving water and gas consumption to produce the required amount of steam. It is shown that the statistical approach allows for building control systems without reference to the equipment type, the manufacturer and the model. At the same time, the obtained results can be improved by refining the characteristic curves characterizing the equipment used by applying more accurate methods and in the process of accumulating new data. Besides, the above approach allows for considering the changes in the equipment work characteristics associated with its repairs, configuration changes, and wear.

Key words: Optimal control, Technological process, Production system, Regression, Choice of operating mode, Boiler, Steam, Statistical model, Characteristic curve.

Cite this Article: L.A. Mylnikov, M.V. Kulikov, B. Krause, The Selection of Optimal Control of the Operation Modes of Heterogeneous Duplicating Equipment Based on Statistical Models with Learning, *International Journal of Mechanical Engineering and Technology* 9(9), 2018, pp. 1516–1526.
http://www.iaeme.com/IJMET/issues.asp?JType=IJMET&VType=9&IType=9

1. INTRODUCTION

At present, it is possible to observe a situation when equipment of different types, from different manufacturers, with significantly different characteristics is used to support activities. In such a situation, efficient control of the processes in which this equipment is involved, especially when it is complementary or interchangeable, becomes a complex and at the same time an actual technological task. Controlling such systems requires the use of intelligent control systems based on statistical and expert methods.

An example of such a task is the vapor pressure maintenance system for operating various devices and equipment (for example, in the production of pressure cardboard for corrugated units). The traditional approach to maintaining the required pressure [1] is associated with controlling the configuration of the steam distribution network when the equipment is operating in optimal conditions and using multiple steam generators operating in parallel mode. However, this approach does not work in the transmission networks of steam of small length and the absence or slight redundancy. In such networks, the factor of considering the characteristics becomes critical. An example of such systems is factory closed networks for which enterprises independently produce steam. Such networks are susceptible to changes in equipment operation: changes in its characteristics and operation modes.

The most promising competing approaches for solving such problems today are the theory of controlling dynamic objects using predictive models (Model Predictive Control (MPC)) [7] and the concept of multi-agent systems (MAS) [11]. Good results can be achieved using discrete control and probabilistic models and clustering and classification methods [20].

However, when considering factory networks, even a small error in the choice of solution leads to significant effects on the efficiency of the production system. In this regard, one should consider the issues of constructing a continuous model of the control object to solve the control problem since discrete systems will either give an unacceptably large error or have an unacceptably large dimension. Considering the significant influence on the functioning of the production system, even a slight deviation of parameters due to the small scale of such systems, it is necessary to build models that will take into account all the features of the equipment (the specifics of specific instances, wear during operation, the effect of maintenance and repairs).

Such models allow for building regression methods [10], neural networks [5] and machine learning and data mining methods [18], provided that the model is updated in the process of receipt of new statistical data accumulated in the process. This approach allows for moving away from the manufacturer's recommendations and for clarifying the equipment's performance. However, it requires the use of methods, the refinement of which will take no more time than the decision time.

2. DESCRIPTION OF THE CONTROL OBJECT AND THE STATEMENT OF THE CONTROL PROBLEM

Let us consider the control problem on the example of the installation of the production of paperboard, for which the production of steam is necessary. Historically, this facility has two boiler stations (station 1 consists of boilers 1 and 2, and station 2 from boiler 3), for steam production and two plants for the production of paper (corrugated units 1 and 2, see Fig. 1 in which W is water consumption volume, Gis gas consumption volume, Sis produced steam amount, P is produced goods amount).

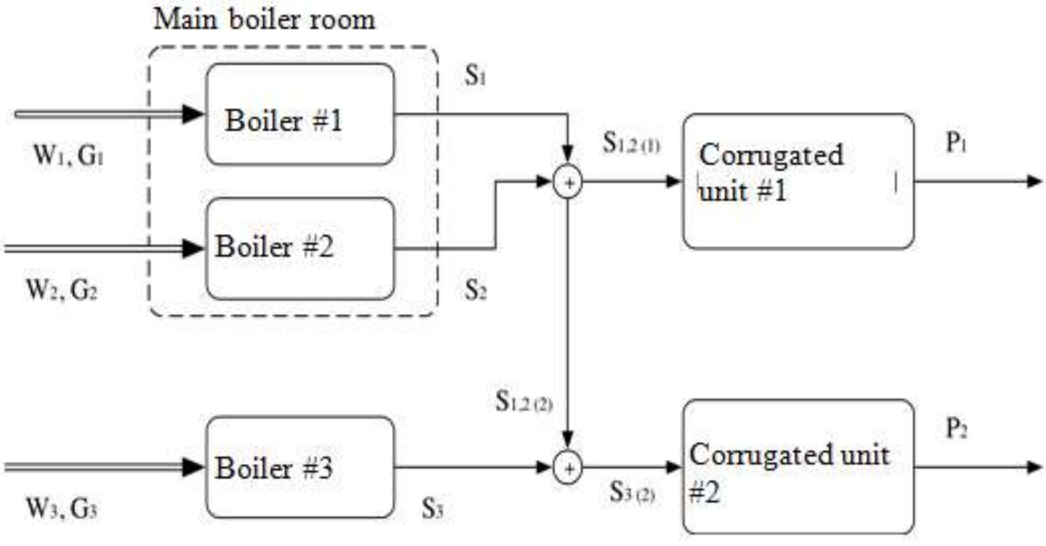


Figure 1. Mnemonic diagram of the control object.

The processing units production volume depends on the volume of the steam $P_1 = f(S_{1,2(1)})$, $P_2 = f(S_{3(2)}(t))$ supplied to them, and is limited by the maximum performance of these plants $P_1 \leq P_{1,max}$, $P_2 \leq P_{2,max}$ at which the protection elements in the pipelines with steam are activated (release of excessive pressure). The performance of the boilers also depends on the volumes of water, and steam consumption $S_1 = f(W_1,\ G_1)$,$S_2 = f(W_2,\ G_2)$, $S_3 = f(W_3,\ G_3)$and is limited by their maximum performance $S_1 \leq S_{1,max}$, $S_2 \leq S_{2,max}$, $S_3 \leq S_{3,max}$.

Based on the configuration of the scheme and the formulated conditions, an optimal control problem can be posed in the following statement:

$$C_w(W_1(t) + W_2(t) + W_3(t)) + C_G(G_1(t) + G_2(t) + G_3(t)) \ \rightarrow min,$$
$$P_1(t) = f(S_{1,2(1)}(t)),$$
$$P_2(t) = f(S_{3(2)}(t)),$$
$$P_1(t) + P_2(t) = P(t),$$
$$P_1(t) \leq P_{1,max}F_{P_1}(t),$$
$$P_2(t) \leq P_{2,max}F_{P_2}(t),$$
$$S_1(t) = f(W_1(t), G_1(t)),$$
$$S_2(t) = f(W_2(t), G_2(t)),$$
$$S_3(t) = f(W_3(t), G_3(t)),$$
$$S_1(t) \leq S_{1,max}F_{S_1}(t),$$
$$S_2(t) \leq S_{2,max}F_{S_2}(t),$$
$$S_3(t) \leq S_{3,max}F_{S_3}(t),$$
$$S_{3(2)}(t) = S_3(t) + S_{1,2(2)}(t)$$
$$S_1(t) + S_2(t) = S_{1,2(1)}(t) + S_{1,2(2)}(t),$$

where $C_w$is the price per cubic meter of water, $C_G$is the price per cubic meter of gas, P is the production plan, $t$is the time, $F(t) \in [0,1]$ is the schedule of preventive maintenance and equipment stops (if$F(t) = 0$ then the equipment does not work).

The problem can be considered fully formulated if the dependencies $P_1 = f(S_{1,2(1)})$,$P_2 = f(S_{3(2)}(t))$, $S_1 = f(W_1, G_1)$,$S_2 = f(W_2, G_2)$, $S_3 = f(W_3, G_3)$ are found. However, it is impossible to solve this problem analytically because there are nine unknowns and only six equations. The fact that gas and water consumption are interrelated quantities helps to get out of this situation. Then it is possible to build the three missing equations. $W_1 = f(G_1)$, $W_2 = f(G_2)$, $W_3 = f(G_3)$.

If, after the substitution of these dependencies, the problem is solved, then we can talk about solving the control problem by applying the algorithm shown in Fig. 2.

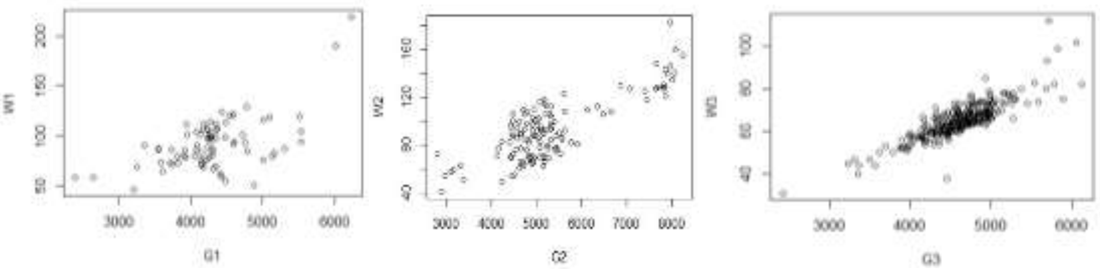
1.Getting the statistical data for P_1, P_2, S_1, S_2, S_3, G_1, G_2, G_3, W_1, W_2, W_3.
2. Making up the statistical models $P_1 = f(S_{1,2(1)})$,$P_2 = f(S_{3(2)}(t))$, $S_1 = f(W_1, G_1)$,$S_2 = f(W_2, G_2)$, $S_3 = f(W_3, G_3)$.
3. Cleaning up the statistical data.
4. Refinement of the statistical models $P_1 = f(S_{1,2(1)})$,$P_2 = f(S_{3(2)}(t))$, $S_1 = f(W_1, G_1)$,$S_2 = f(W_2, G_2)$, $S_3 = f(W_3, G_3)$.
5. Obtaining data on the production plan and data on the operability or limitations in the use of boilers ($P_1(t) \le P_{1,max} F_{P_1}(t)$, $P_2(t) \le P_{2,max} F_{P_2}(t)$, $S_1(t) \le S_{1,max} F_{S_1}(t)$, $S_2(t) \le S_{2,max} F_{S_2}(t)$, $S_3(t) \le S_{3,max} F_{S_3}(t)$).
6. Statement of the problem of optimal control (1) (determination of parameter values).
7. Solution of the optimal control problem (1) and determination of values G_1, G_2, G_3, W_1, W_2, W_3.
8. Providing control.
9. Retrieving system response data (new values P_1, P_2, S_1, S_2, S_3, G_1, G_2, G_3, W_1, W_2, W_3).
10. Return to step 4.

Figure 2. The control system algorithm.

Using the algorithm shown in Fig. 2 in addition to solving the problem of control allows for solving the problem of data storage by accumulating knowledge contained in them in the model [13]. As a result, there is no need to store all the accumulated data.

3. MAKING UP A MODEL OF THE CONTROL OBJECT

The model of the control object can be built based on the models of the elements of its components and data on their interaction. Data on the configuration and principles of interaction are laid down in the formulation of the optimal control problem (1). Then, the only thing left is to determine the following dependencies $P_1 = f(S_{1,2(1)})$, $P_2 = f(S_{3(2)}(t))$, $S_1 = f(W_1, G_1)$, $S_2 = f(W_2, G_2)$, $S_3 = f(W_3, G_3)$, $W_1 = f(G_1)$, $W_2 = f(G_2)$, $W_3 = f(G_3)$. These dependencies can be obtained through regression analysis methods based on statistical data. Using the statistics for 2014 and 2015, let us construct graphs for the listed dependencies (see Fig. 3).



a) b) c)

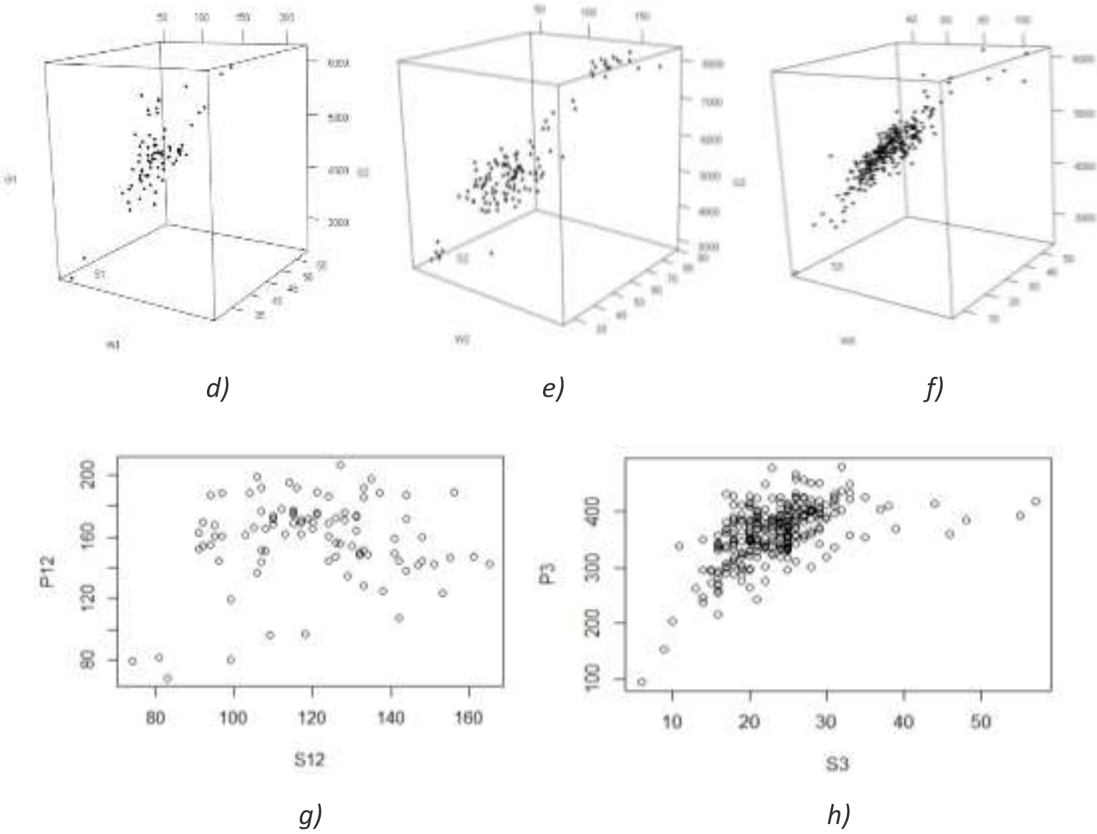


Figure 3. The collected statistics data on the operation of boilers and corrugated units: a) values $W_1 = f(G_1)$, b) values $W_2 = f(G_2)$, c) values $W_3 = f(G_3)$, d) values $S_1 = f(W_1, G_1)$, e) values $S_2 = f(W_2, G_2)$, f) values $S_3 = f(W_3, G_3)$, g) values $P_1 = f(S_{1,2(1)})$, h) values $P_2 = f(S_{3(2)}(t))$.

A visual analysis of the results obtained as a result of the construction suggests that the construction of regression models on the available statistics will be ineffective due to their high noise level [8]. It is possible to clean them based on considerations of the maximum efficiency of the equipment. Thus, the values should only grow (otherwise it turns out that the gas units are produced for a smaller volume of water that could be) to clear the dependence of water consumption on the volume of gas consumption. In this case, not all gas-water combinations will correspond to the maximum productivity with such filtration. Therefore, the regression curve will lie somewhat below those values that could be achieved and further calculations will not correspond to maximum productivity (there will be potential for further improving the quality of control). The specified value filtering allows for clearing values for dependencies $S = f(W, G)$ by using only the values corresponding to the remaining pairs $W = f(G)$. For dependencies of the productivity of corrugated units ($P = f(S(t))$) on steam volume, a similar approach is not acceptable since from visual analysis it can be seen that after a specific value of the steam volume, productivity starts to decline (due to the nature of the equipment or lack of data). However, it can be observed that one value S corresponds to several values P in the considered sample. In this case, the filtering of values may consist in choosing the maximum value P (maximum performance values). After the described procedure of cleaning the data, the results shown in Fig. 4 are obtained.

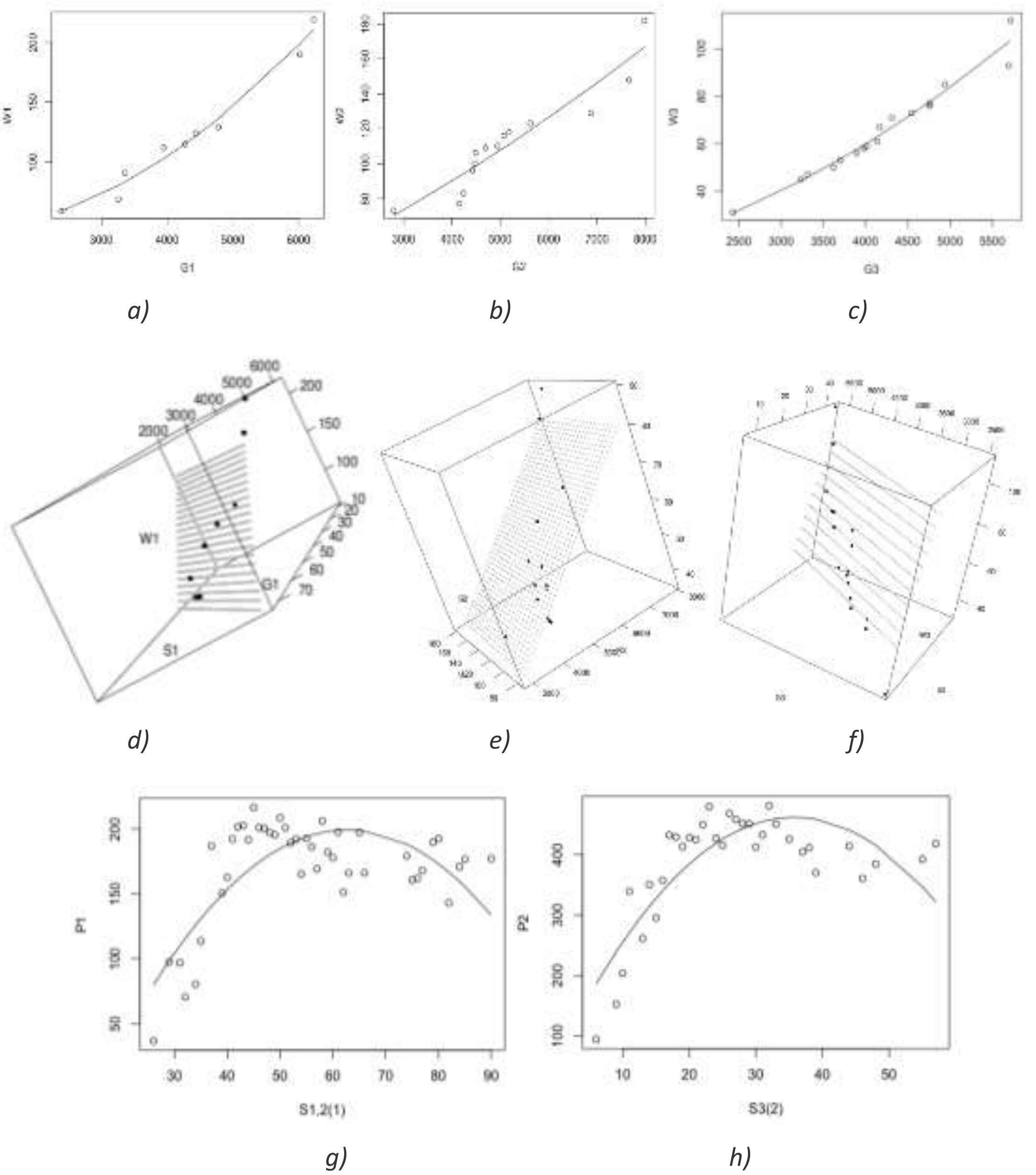


Figure 4. Filtered statistics and the resulting functional dependencies describing the operation of boilers and corrugated units: a) dependency $W_1 = f(G_1)$, b) dependency $W_2 = f(G_2)$, c) dependency $W_3 = f(G_3)$, d) dependency $S_1 = f(W_1, G_1)$, e) dependency $S_2 = f(W_2, G_2)$, f) dependency $S_3 = f(W_3, G_3)$, g) dependency $P_1 = f(S_{1,2(1)})$, h) dependency $P_2 = f(S_{3(2)}(t))$.

Many regression methods are known, which differ in the requirements for the amount of data, their purity, operation speed, obtained descriptions accuracy, etc. The simplest regression models are regression models based on polynomials as a series [2]. They have been well researched and widely used in practice. The advantage of this model type is the ability to solve collinearity problems. The method makes it easy to identify noise in data [15]. Thus, let us use this approach to construct regression models by restricting the degree of the polynomial to a square, so that problem (1) turns out to be analytically solvable [17] (see Fig. 4).

The mathematical formalization of the dependencies obtained will be as follows (all the descriptions obtained were tested for adequacy by the Pearson criterion):

$W_1(t) = 37.75903 - 0.02908416 \cdot G_1(t) + 0.000004921683 \cdot G_1(t)^2,$
$W_2(t) = 30.11489 + 0.01283592 \cdot G_2(t) + 0.0000005366 \cdot G_2(t)^2,$
$W_3(t) = 5.238735 + 0.005491781 \cdot G_3(t) + 0.000002047041 \cdot G_3(t)^2,$
$P_1 = -149.20678427 + 11.09998583 \cdot S_{1,2(1)}(t) - 0.08840261 \cdot S_{1,2(1)}(t)^2,$
$P_2 = 64.2436393 + 22.2134358 \cdot S_{3(2)}(t) - 0.3104346 \cdot S_{3(2)}(t)^2,$
$S_1(t) = 28.557991318 - 0.234945167 \cdot W_1(t) + 0.009753317 \cdot G_1(t),$
$S_2(t) = 8.019937969 - 0.007622630 \cdot W_2(t) + 0.009098759 \cdot G_2(t),$
$S_3(t) = -18.410381736 + 0.209038417 \cdot W_3(t) + 0.006672486 \cdot G_3(t).$

The substitution of the obtained results and the addition of the model (1) with the positivity conditions for the obtained results give an entry in the system for solving the LINGO optimization problems of [6] LINDO Systems Inc. shown in Fig. 5. The resulting problem has a linear criterion and quadratic constraints [17], therefore, is analytically solvable. In the above formalization, the price of a cubic meter of water is $C_w = 31.63$ rub., and gas is $C_G = 5$ rub. Additionally introduced restrictions for P_1 and P_2 set the planned values of cardboard production, as well as numerically set limits on the performance of steam boilers and corrugated units based on their technical characteristics.

```
MODEL:
MIN =         31.63*(W1+W2+W3) + 5*(G1+G2+G3);
W1            >=      0;
W2            >=      0;
W3            >=      0;
G1            >=      0;
G2            >=      0;
G3            >=      0;
S1            >=      0;
S2            >=      0;
S3            >=      0;
W1 =          37.75903        -0.02908416*G1       +0.000004921683*G1^2;
W2 =          30.11489        +0.01283592*G2       +0.0000005366*G2^2;
W3 =          5.238735        +0.005491781*G3      +0.000002047041*G3^2;
S1            <=      50;
S2            <=      90;
S3            <=      60;
S1  =         28.557991318    -0.234945167*W1      +0.009753317*G1;
S2  =         8.019937969     -0.007622630*W2      +0.009098759*G2;
S3  =         -18.410381736   +0.209038417*W3      +0.006672486*G3;
S121 + S122           =       S1 + S2;
S122 + S3   =         S32;
P1            >=      650;
P2            >=      250;
P1  =         -149.20678427   +11.09998583*S121    -0.08840261*S121^2;
P2  =         64.2436393      +22.2134358*S32      -0.3104346*S32^2;
END
```

Figure 5. The optimization task model, recorded in the LINGO system.

Changing the values of the plan according to the retrospective volume of production and making the appropriate calculations, it becomes possible to compare the model values for gas and water consumption with retrospective ones.

4. ANALYSIS OF THE RESULTS AND DISCUSSION

The solution of the received task will be the volumes of water W_1, W_2, W_3 and gas G_1, G_2, G_3 consumption, for each of the boilers and its planned volumes of steam production S_1, S_2, S_3. The obtained data can be recalculated into direct costs $(31.63 * \left(W_1(t) + W_2(t) + W_3(t)\right) + 5 * (G_1(t) + G_2(t) + G_3(t)))$ for production volume production $(P_1 + P_2)$ and compared with available statistical data (see Fig. 6). The data in the figure form two groups of values: values corresponding to the operation of the first corrugating installation and values corresponding to the operation of the second corrugation installation.

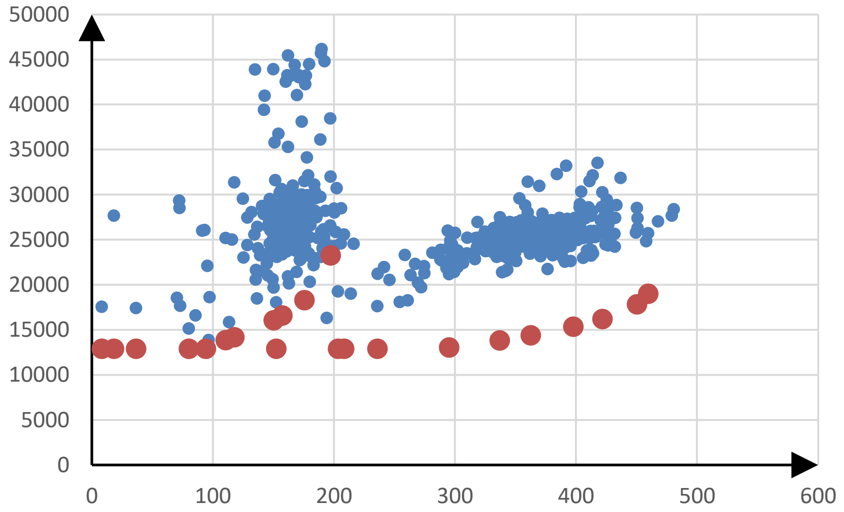


Figure 6. The dependence of the production volume on the price of resources used for it (gas and water for boilers): small blue dots mean retrospective data; big red dots mean the data obtained as a result of calculations on the model.

Fig. 6 shows that the results obtained on the model are the best for efficiency as compared with the operation modes chosen at the enterprise. It is possible to estimate the amount of money spent on purchasing gas and water over a specified period knowing the data of monthly production volumes and compare these costs with the model data. Based on these data, the average savings in one month can be about 600 thousand rubles (for example, for January 2014 it is 554,058.76 thousand rubles).

Even if it is assumed that the models are inaccurate, then as a result of their use, let us supply values corresponding to their maximum efficiency, and system response will be the result. These data according to the algorithm shown in Fig. 2 will be added to the statistical data, which is accordingly reduced to refine the model.

Another way to improve the accuracy of the model is to increase the degree of polynomials used or use other more complex methods [10]. However, in this case, it is needed to solve the problem using heuristic methods for some of which it is possible to set the requirements for the obtained solution accuracy and the time it is possible to spend on finding a solution. If heuristic methods are used to solve a problem, or combinatorial methods are supplemented with elements of heuristics, the proof of the completeness of the method used is complicated. However, today, some of them have convergence proved [4], unfortunately, while not saying anything about the amount of time it takes to search.

In general, two large groups can be distinguished among the heuristic methods of random search: learning randomized methods and evolutionary programming [16]. Practical application distinguishes the methods in the convergence speed and the iterations number necessary to find a feasible solution (some of the methods, for example, genetic algorithms allow for finding an extreme solution, but not necessarily the optimal one). The complexity of the selection problem is also related to the fact that their parameters determine the effectiveness of some methods of stochastic search (in particular, GA).

The approach to building a model described in the article opens up, with the accumulation of data, the possibility of transition to a dynamic predictive control of the described system by taking into account the inertia of the processes of changing parameters in boilers and corrugated units when changing the values of water, gas and steam at their input, which will allow to form in advance control impact and get additional energy savings [9].

Developing this approach, we face with the need to build a statistical model of installations in the form of multiple regression structurally which (for example, a boiler) may look as shown in Fig. 7.

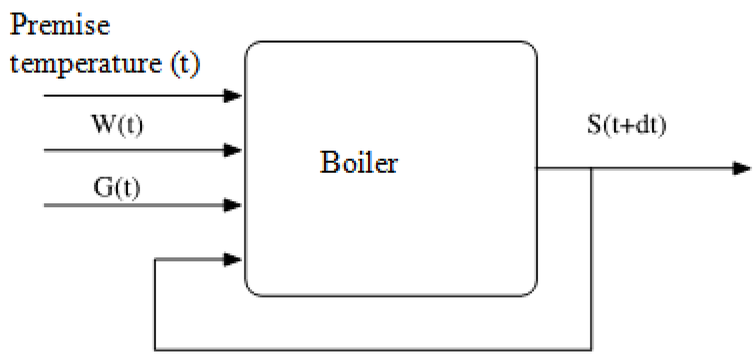


Figure 7. Structural scheme of data transmission for constructing a predictive model of a processing unit (steam boiler).

Such machine learning methods [18] are used to build these models: *Neural Network*, *Support Vector Machine* (*SVM*), *k-NN*, *PLS Regression*, *Lasso*, *Bayesian linear regression* (*Bayesian statistics*). Even though this direction has received great development in recent times, most of the methods have drawbacks, and their application is limited. The main disadvantage of the *SVM* method is its sensitivity to noises [14]. Identifying and removing noise components from feature vectors on a large dataset is not an easy task. Besides, the method is negatively influenced by high dimensionality (dependence of the target value on a large number of parameters and factors) [14]. *Neural Network* and the *k-NN* methods apply to models of any complexity, but when working with extensive data, the calculation is slower, and accuracy is deteriorating. By averaging the values, a good result is obtained, but in comparison with other methods of machine learning, the result does not justify itself. *k-Nearest Neighbor* is typically used in data mining applications. The value of *k*, which gives the highest accuracy, depends on the data set and must be configured in advance [12]. The downside of neural networks is the complexity of interpreting the principles of obtaining solutions because the model is necessarily a "black box," as well as a high dependence of the result on the data used and unsatisfactory results if the data have periodic processes (the data are often averaged). When using the Bayesian approach and the PLS regression method [3] in regression problems, there may be difficulties associated with obtaining estimates when using large amounts of data associated with loss of accuracy [19]. Thus, there is the challenge of choosing and possibly adapting the method for the process unit model.

5. CONCLUSIONS

In order to solve the problem considered in the article, it is possible to build a model that allows considering the peculiarities of the functioning of the equipment used, which allows not only obtaining direct savings through the optimal use of existing equipment but also to reduce its wear and tear (including accidents) through its use in the most efficient operating modes.

Initially, the solutions obtained are approximate; the model introduces an error due to the statement of the problem in a statistical formulation. However, the model itself is being refined in the process of work (it is trained as new data become available), which will lead to an improvement in the quality of the solutions obtained.

The use of statistical models in optimal control problems opens up new possibilities for studying the processes taking place in production systems and the methods of controlling these processes.

ACKNOWLEDGMENTS

The authors are grateful to the Government of Perm Territory for supporting research in the framework of the scientific project No. S-26/058.

REFERENCES

[1] Bungener, S.L., Van Eetvelde, G., Maréchal, F. Optimal Operations and Resilient Investments in Steam Networks. *Frontiers in Energy Research*, **4**, 2016.

[2] Dai, Q. Recent Advances in Data Mining Techniques and Their Applications in Hyperspectral Image Processing for the Food Industry: Recent advances in data mining techniques. *Comprehensive Reviews in Food Science and Food Safety*, **5**(13), 2014, pp. 891–905.

[3] Hair, J.F. A primer on partial least squares structural equation modeling (PLS-SEM). Los Angeles: Sage, 2017.

[4] Jaszkiewicz, A. Genetic local search for multi-objective combinatorial optimization. *European Journal of Operational Research*, **1**(137), 2002, pp. 50-71.

[5] Khoroshev, N.I., Kulikov, M.V. Increase of energy engineering equipment performance based on neural network modeling and cluster data analysis Saint Petersburg, Russia: IEEE, 2017.

[6] LINGO: The Modeler Language and Optimizer Chicago: LINDO Systems Inc., 2017.

[7] Mayne, D.Q. Constrained model predictive control: Stability and optimality. *Automatica*, **6**(36), 2000, pp. 789-814.

[8] Mylnikov, L. Intelligent data analysis in the management of production systems (approaches and methods). Moscow: BIBLIO-GLOBUS, 2017.

[9] Mylnikov, L.A. The use of dynamic predictive models for control of technical systems with inertness. *Perm National Research Polytechnic University Bulletin. Electrotechnics, information technologies, control systems*, **26**, 2018, pp. 77-91.

[10] Mylnikov, L.A., Seledkova, A.V., Krause, B. Forecasting characteristics of time series to support managerial decision making process in production-And-economic systems.

Proceedings of 2017 20th IEEE International Conference on Soft Computing and Measurements, SCM 2017 6 July 2017, pp. 853–855.

[11] Olfati-Saber, R., Fax, J.A., Murray, R.M. Consensus, and Cooperation in Networked Multi-Agent Systems. *Proceedings of the IEEE*, **1**(95), 2007, pp. 215-233.

[12] Ougiaroglou, S., Diamantaras, K.I., Evangelidis, G. Exploring the effect of data reduction on Neural Network and Support Vector Machine classification. *Neurocomputing*, **280**, 2018, pp. 101-110.

[13] Paprotny, A., Thess, M. Real-time data mining: self-learning techniques for recommendation engines. 2013.

[14] Shao, Z. A new electricity price prediction strategy using mutual information-based SVM-RFE classification. *Renewable and Sustainable Energy Reviews*, **70**, 2017, pp. 330-341.

[15] Shouval, R.. Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in SCT. 2013.

[16] Syswerda, G. Uniform crossover in genetic algorithms Morgan Kaufmann. 1989.

[17] Taha, H.A. Operations research: an introduction. Upper Saddle River, N.J: Prentice Hall, 2003.

[18] Witten, I.H., Frank, E. Data mining: practical machine learning tools and techniques. Amsterdam ; Boston, MA: Morgan Kaufman, 2005.

[19] Woody, N.A., Brown, S.D. Partial least-squares modeling of continuous nodes in Bayesian networks. *Analytica Chimica Acta*, **1-2** (490), 2003, pp. 355-363.

[20] Wu, X.. Top 10 algorithms in data mining. *Knowledge and Information Systems*, 1(14), 2008, pp. 1-37.