Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 100 (2016) 253 – 262

Procedia
Computer Science

Conference on ENTERprise Information Systems / International Conference on Project MANagement / Conference on Health and Social Care Information Systems and Technologies, CENTERIS / ProjMAN / HCist 2016, October 5-7, 2016

Regression models of the nuclear power unit VVER-1000 using data mining techniques

Tymur Foshch [1,2],Filipe Portela[1,3*], José Machado[1] and Maksim Maksimov[2]

[1]University of Minho, Portugal
[2]Odessa National Polytechnic University
[3]ESEIG, Porto Polytechnic, Portugal

Abstract

Due to plenty of changes in many interrelated processes at nuclear power plants there is the need to show which values of some parameters of the nuclear power plant with VVER-1000 are better. In this task data mining techniques can be introduced. In order to obtain regression models of nuclear power plant with VVER-1000 algorithms such as the Linear Regression, REPTree, and M5P were selected and the datasets were obtained by simulating two control programs in Simulink software. The study focused on such targets as the average temperature of the coolant in the first circuit and the output power of the power generator. This study demonstrates the good results of the correlation coefficients and the root relative squared error metrics in case of the improved compromise-combined control program in comparison with the control program with the constant average temperature of the coolant in the reactor core. In terms of the results the root relative squared error metric is less than 2.8% and the correlation coefficients had values higher than 99,95%. The use of these models can contribute to improving the understanding of the internal processes because using the best regression data mining models allows to see advantages of the improved compromise-combined control program.

Keywords: VVER-1000; Data Mining; Regression Models; Nuclear Power Plant.

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000 .
E-mail address: cfp@dsi.uminho.pt

doi:10.1016/j.procs.2016.09.151

1. Introduction

It goes without saying that nuclear is most often the lowest cost choice of power, but at the same time it is one of the most dangerous industries. However, the nuclear industry has defined excellence, in part, as safely generating electricity as evidenced by good capacity and generation metrics. In other words, 100% capacity and generation are the hallmarks of good power plant operations. This focus drives the culture for keeping the plants on line as "base load". The nuclear industry has a continuously improving culture, so there has been increasing preference to stay at 100%. Even so, grid controllers will ask plants to ramp given certain grid reliability needs, and the plant is required to respond and is in fact, capable of responding [1]. The delicate balance of generation and load on an electricity grid continuously fluctuates, often significantly and on varying timescales as industrial and household demands ebb and flow throughout a day, a week, or a season. Such problems produce new complex techniques and new automatic control systems every day, which allow to switch existing NPP in a load-following mode.

Moreover, the problems above mentioned have generated another issue, which consists of a selection of a control program, that needs to be used at nuclear power plants (NPP) in a load-following mode. This is another complex task because of the different conditions of NPP, types of reactors core and etc.

This study has shown the regression results of the two parameters of the mathematical multi-zone model NPP with VVER-1000 in a load-following mode using data mining (DM) techniques in frames of two control programs. The Russian abbreviation VVER stands for water-water energy reactor (i.e. water-cooled water-moderated energy reactor). This describes the pressurised water reactor (PWR) design. The main distinguishing features of the VVER compared to other PWRs are: Horizontal steam generators; Hexagonal fuel assemblies; No bottom penetrations in the pressure vessel; High-capacity pressurisers providing a large reactor coolant inventory.

Reactor fuel rods are fully immersed in water kept at 15 MPa pressure so that it does not boil at the normal (220 to over 300 C) operating temperatures. Water in the reactor serves both as a coolant and a moderator which is an important safety feature. Should coolant circulation fail, the neutron moderation effect of the water diminishes, reducing reaction intensity and compensating for loss of cooling, a condition known as negative void coefficient. Later versions of the reactors are encased in massive steel pressure shells. Fuel is low enriched (ca. 2.4–4.4% ^{235}U) uranium dioxide (UO_2) or equivalent pressed into pellets and assembled into fuel rods.

Reactivity is controlled by control rods that can be inserted into the reactor from above. These rods are made from a neutron absorbing material and, depending on depth of insertion, hinder the chain reaction. If there is an emergency, a reactor shutdown can be performed by full insertion of the control rods into the core.

A few of DM techniques such as LinearRegression, M5P and REPTree were explored. Although there is a plenty of DM techniques, the DM techniques were selected by 3 requirements such as actual (simulated) data values, availability and efficiency of use. Such variables were selected as an initial dataset for forecasting: the axial offset (AO,%), the energy release of the reactor (Q, %), the position of the 10th regulating group of control rods (hSUZ, cm), the concentration of ^{135}J (cJ, unit/cm^{-3}), the concentration of ^{135}Xe (cXe, unit/cm^{-3}), the input temperature of the coolant in the reactor core (Tin, r.u.), the average temperature of the coolant in the first circuit (Tav, r.u.), the output temperature of the coolant it the reactor core (Tout, r.u.), the steam pressure in the second circuit (P, r.u.), the steam flow in the second circuit (Gst, kg/sec), the output electric power of generator (PG, %). These 11 attributes were obtained by two control programs. The first program was the control program of the power unit with the constant average temperature of the coolant in the first circuit (Tav) and the second program – an advanced automatic control system that implements the improved compromise-combined control program (TinB).

The average temperature of the coolant in the first circuit and the output electric power of generator were chosen as main targets for DM analysis in case of both programs. The large numbers of samples of regression models were obtained and many of them were dropped because of bad results. The models in case of TinB control program show satisfactory results with the root relative squared error metric (RRSE) less than 2.8% in comparison with Tav control program, where the RRSE metric is more than 34%. The same result goes for the correlation coefficients of all scenarios. The correlation coefficients in case of TinB control program have values more than 0.9995 and in case of Tav control program – less than 0.7778. Besides the results there are only 6 satisfactory models using PG target.

The study consists of six sections. The first section addresses the problem presented. In the following section there are some addressed aspects related to nuclear energy, as also the concepts of two control programs and multi-zone model of VVER-1000 reactor. Then in section three is presented the work that created using the cross industry standard process for data mining and the design science research methodology. The fourth section presents CRISP-DM phases as data and business understanding, data preparation and modeling. The fifth section presents discussion of the best achieved results. Finally, the sixth section presents the most relevant conclusions.

2. Background

The development of nuclear power suggests the emergence of innovative projects of nuclear reactors whose safety is ensured largely due to the properties of self-protection. However, for most innovative projects the degree of realization of self-protection properties is not absolute and therefore requires a comprehensive assessment based on qualitative and quantitative analysis based on experience of existing reactor units.

Traditionally, nuclear power plants (NPP) have been considered as baseload sources of electricity as they rely on a technology with high fixed costs and low variable costs. In the beginning of the nuclear era, the share of nuclear power in the overall energy mix was usually small, and adjustments of electric load in response to variations in electricity demand could be left to technologies with different economic and technological characteristics, most notably low fixed cost and high variable cost gas plants. However, this simple state of affairs no longer applies in all countries. The share of nuclear power in the national electricity mix of some countries has become so large that the utilities have had to implement or to improve the maneuverability capabilities of their NPPs in order to be able to adapt electricity supply to daily, seasonal or other variations in power demand. This is the case in France where more than 75% of electricity is generated by NPPs, and where some nuclear reactors operate in load-following mode (see Figure 1) [1].

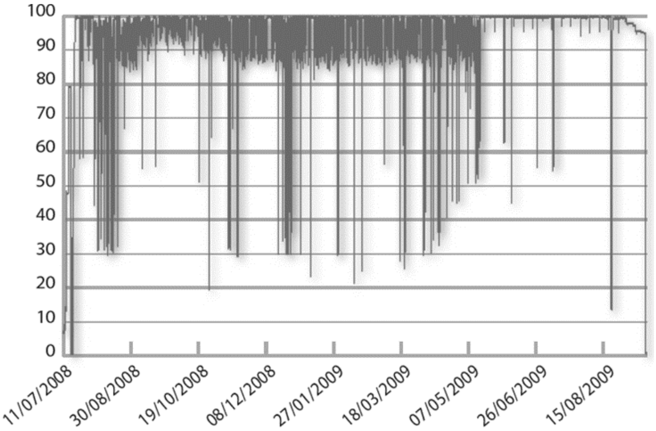

Figure 1. Typical power history during the Courtesy of Électricité de France (EDF) reactor cycle (in % of rated power).

Besides, changing the power of the reactor is substantially determined by its physical characteristics and especially by the effect of the periodic redistribution of power by volume of the core, the so-called xenon oscillations. Moreover, xenon transient processes affect the maneuverability of the NPP [2] and can lead to accumulation of cladding failure parameter [3].

2.1. Control programs of the power unit

A load-following mode of the power unit follows by static control programs, which represents the dependence of the basic installation parameters (coolant flow rate of the first and second circuits, pressure and temperature) of its power in the steady-state conditions. In this study a load-following mode means that NPP works in an experimental-mathematical modelling with changes of total power from 100% to 80% and back in 8 hours by Simulink of the software package Matlab. The choice of a control program affects the characteristics of the main equipment and it is a complex technical and economic problem. The main control programs of power units with VVER-1000 are:

- The control program with the constant average temperature of the coolant in the first circuit;

- The control program with the constant pressure in the second circuit;
- The control compromise program with the moderate change in the average temperature of the coolant in the first circuit and the pressure in the second;
- The control combined program with maintaining the constant pressure in the second circuit at low loads and constant average primary coolant temperature at high loads.

A lot of NPP are operating under the control program of the power unit with the constant average temperature of the coolant in the first circuit and it can be explained by good results of control in compare to other. Recently the advanced automatic control system that implements the improved compromise-combined control program was developed. The advanced automatic control system that implements the improved compromise-combined control program showed better results than the control program of the power unit with the constant average temperature of the coolant in the first circuit [4]. The results were obtained by Simulink of Matlab software package.

There are many different control programs of the power unit, but as a research object has been selected the control program of the power unit with the constant average temperature of the coolant in the first circuit, which is widely used in practice as it was mentioned above, and the previously developed the advanced automatic control system that implements the improved compromise-combined control program.

Figure 2 shows a diagram of the power control unit, which realizes the control program with the constant average temperature of the coolant in the reactor core (RC). Figure 3 shows a diagram of the advanced automatic control system that implements the improved compromise-combined control program. More detailed information about these control programs is described in the two related works [5, 6].

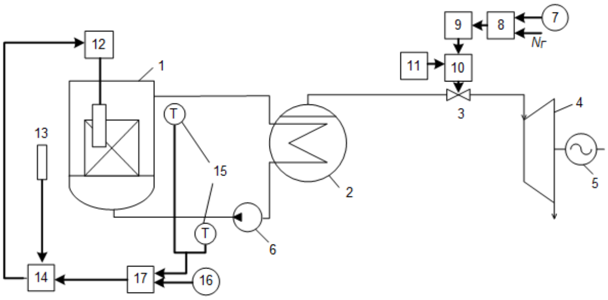

Figure 2. The diagram of the power unit with VVER–1000, which realizes the control program with the constant average temperature of the coolant in the RC.

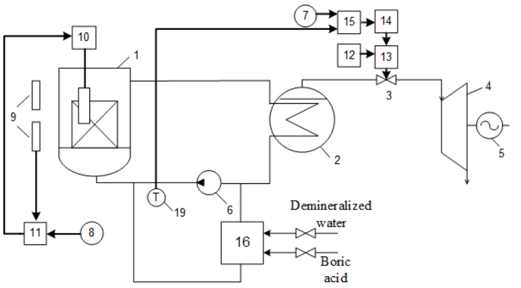

Figure 3. The advanced automatic control system that implements the improved compromise-combined control program.

In the figure 2 is used: 1 – reactor; 2 – steam generator; 3 – turbine control valves; 4 – turbine; 5 – generator; 6 – the main circulating pump; 7 – the setting device of the electric power generator; 8 – the power controller of the power unit; 9 – the control mechanism of the turbine; 10 – the servomotors; 11 – the rotational speed controller of the turbine; 12 – the control rod drive; 13 – the ionization chambers; 14 – the neutron power controller of the reactor; 15 – temperature sensors of the coolant in the first circuit; 16 – the setting device of the average temperature of the coolant in the first circuit; 17 – the controller of the average temperature of the coolant in the first circuit.

In the figure 3 is used: 1 – reactor; 2 – steam generator; 3 – turbine control valves; 4 – turbine; 5 – generator; 6 – the main circulating pump; 7 – the setting device of the temperature of the coolant at the input of the reactor core; 8 – the setting device of the AO; 9 – the ionization chambers; 10 – the control rod drives; 11 – the neutron power controller of the reactor; 12 – the rotational speed controller of the turbine; 13 – the servomotors; 14 – the control mechanism of the turbine; 15 – the power controller of the power unit; 16 – the system of feed and expulsion.

2.2. The mathematical multi-zone model of VVER-1000

In this paper the mathematical multi-zone model of VVER-1000 realized in the Simulink software of the package Matlab was used. This model allows getting sets of data for further analysis using DM. Such mathematical model considers all features associated with the dynamic processes which are described by nonlinear differential equations system. In order to solve and research some of the technical problems often use dot kinetic model, but this solution is not suitable for the establishment of an adequate NPP model, because of the inability to control the main parameters of the reactor, such as changes in temperature of the coolant in the reactor core, fuel temperature and etc. Thus, reactor model is developed, which has a core divided into ten sections by height. Each zone of the model has the same structure. The only difference in the geometric and thermo-hydraulic parameters of each zone and, accordingly, various static and dynamic properties. More detailed information about multi-zone mathematical model VVER-1000 and interpretation is described in the references [7, 8].

2.3. Data Mining and related processes

Data Mining (DM) is a process of detecting dependencies in data, which must be automatic or semiautomatic at least. Also, the processes, which are related to DM, are using different intelligent techniques of a statistical and mathematical analysis, which allow to get useful information and to determine dependencies of it from data-sets in varying degrees. The overall goal of the data mining process is to extract information from a dataset and transform it into an understandable structure for further use. At the same time, certain dependencies must demonstrate a directional advantage in something, as usual in the economic aspect [9]. Concerning this study, it is proposed to use knowledge discovery. Data mining is the analysis step of the "knowledge discovery" process. Knowledge discovery shows the process of automatically searching large volumes of data for patterns that can be considered knowledge about the data [10]. It is often described as deriving knowledge from the input data. Knowledge discovery developed out of the data mining domain, and is closely related to it both in terms of methodology and terminology.

A decision support system (DSS) is a computer-based information system that supports business or organizational decision-making activities. DSSs serve the management, operations, and planning levels of an organization (usually mid and higher management) and help people make decisions about problems that may be rapidly changing and not easily specified in advance—i.e. Unstructured and Semi-Structured decision problems. Decision support systems can be either fully computerized, human-powered or a combination of both. DSSs include knowledge-based systems. A properly designed DSS is an interactive software-based system intended to help decision makers compile useful information from a combination of raw data, documents, and personal knowledge, or business models to identify and solve problems and make decisions. Regression is a data mining (machine learning) technique used to fit an equation to a dataset. The simplest form of regression, linear regression, uses the formula of a straight line ($y = mx + b$) and determines the appropriate values for m and b to predict the value of y based upon a given value of x. Advanced techniques, such as multiple regression, allow the use of more than one input variable and allow for the fitting of more complex models, such as a quadratic equation.

3. Material and Methods

The process of DM is not a trivial task and consequently, the cross industry standard process for data mining (CRISP-DM) and design science research methodology (DSRM) were used to achieve the goals, which is a lightweight way for organizing, understanding, implementation and development analysis [9,11].

Design science research is a set of analytical techniques and perspectives for performing research in information systems. Design science research involves the design of novel or innovative artefacts and the analysis of the use and/or performance of such artefacts to improve and understand the behavior of aspects of Information Systems. This is achieved by the CRISP-DM methodology, which allows dividing the intellectual analysis to 5 phases. Thus, the current study is based on the CRISP-DM methodology. CRISP-DM stands for cross-industry process for data mining. The CRISP-DM methodology provides a structured approach to planning a data mining project. It is a robust and well-proven methodology. We do not claim any ownership over it. We did not invent it. We are however evangelists of its powerful practicality, its flexibility and its usefulness when using analytics to solve thorny business issues. It is the golden thread than runs through almost every client engagement. In this study, as an application of DM a free software Waikato Environment for Knowledge Analysis (Weka) was used. Weka is a set of visualization tools and algorithms for DM solutions and forecasting problems in conjunction with a graphical user shell for an access of them. Linear Regression, REPTree and M5P algorithms were used in the practical part of DM. After analysis each model shows such metrics as correlation coefficient (CC), mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE), root relative squared error (RRSE), total number of instances (TNI).

Datasets for intelligent analysis prepared by an experimental-mathematical modelling with the load-following of NPP from 100% to 80% and back in 8 hours by Simulink of the software package Matlab. As the data set are such variables as the axial offset (AO,%), the energy release of the reactor (Q, %), the position of the 10th regulating group of control rods (hSUZ, cm), the concentration of ^{135}J (cJ, unit/cm^{-3}), the concentration of ^{135}Xe (cXe, unit/cm^{-3}), the input temperature of the coolant in the reactor core (Tin, r.u.), the average temperature of the coolant in the first circuit (Tav, r.u.), the output temperature of the coolant it the reactor core (Tout, r.u.), the steam pressure in the second circuit (P, r.u.), the steam flow in the second circuit (Gst, kg/sec), the output electric power of generator (PG, %).

4. Study description

4.1. Business Understanding

The first phase of the methodology is focused on the understanding of the research object, main goals and requirements from a business point of view. The purpose of this work is to provide useful and relevant information to operators of the reactor facilities, which will show the benefits of using a particular control program.

The business goal of this paper is to show the different advantages and characteristics of VVER-1000 in two control programs for a load-following mode using DM. It will help to conclude what regression models of control programs are better to use at NPP. At the beginning it is necessary to mention that variables forecasting of a nuclear object is a complicated procedure as the object is non-linear and has a plenty of dependencies of various parameters. There is a mathematical multi-zone model of the VVER-1000, which was previously developed and it allows to obtain the necessary parameters for a specified period of operation time of the reactor facility by the mathematical modelling method, but this kind of model cannot predict the behavior of any of the parameters in this object [9].

The datasets were obtained for performing this study by simulating a nuclear power object VVER-1000 in load-following mode and the Simulink of software package Matlab was used for this. As the targets of DM were used such characteristics of a nuclear object as the average temperature of the coolant in the first circuit (Tav, r.u,) and the output electric power of generator (PG, %).

4.2. Data Understanding

The implementation of the DM models depends on data-sets, which were obtained by simulation. After the simulation of the object in a load-following mode with change the output electric power of generator from 100% to 80% and back in 8 hours, the samples of datasets were used. The sample dataset with using a control program, which realizes the average temperature of the coolant in the reactor core (Tav), was composed of 10356 records.

While the sample of data-set with using the compromise-combined program (TinB) was composed of 9848 records. Each sample of data-set was obtained during the defeneted period of simulation time. It was decided not to consider the option (simulation time) due to the fact that it did not bear any scientific value in this study. Thus this option was excluded in both control programs. For both control programs the following parameters were chosen:

- AO – the axial offset, %;
- Q – the energy release of the reactor core, %;
- hSUZ – the position of the 10th regulating group of control rods, cm;
- cJ – the concentration of ^{135}J, unit/cm^{-3};
- cXe – the concentration of ^{135}Xe, unit/cm^{-3};
- Tin – the input temperature of the coolant in the reactor core, r.u.;
- Tav – the average temperature of the coolant in the first circuit, r.u.;
- Tout – the output temperature of the coolant it the reactor core, r.u.;
- P – the steam pressure in the second circuit, r.u.;
- Gst – the steam flow in the second circuit, kg/sec;
- PG – the output electric power of generator, %.

The Table 1 and 2 demonstrate the statistical analysis for each parameter.

Table 1. Variables distribution of the control program, which realizes the constant average temperature of the coolant in the reactor core.

Tav control program	MIN	MAX	MEAN	stDEV	Unique
AO, %	-23.128	-3.415	-22.257	2.712	70%
Q, %	83.942	99.984	84.573	2.084	17%
hSUZ, cm	273.648	321.074	277.407	5.543	97%
cJ, unit/cm^{-3}	4.07e^{16}	4.46e^{16}	4.23e^{16}	1.15e^{15}	100%
cXe, unit/cm^{-3}	3.43e^{16}	3.48e^{16}	3.46e^{16}	1.32e^{14}	100%
Tin, r.u.	-0.009	2.399	2.89	0.31	11%
Tav, r.u.	-0.223	0.128	0	0.008	6%
Tout, r.u.	-2.378	-0.002	-2.288	0.311	11%
P, r.u.	0	0.403	0.386	0.052	10%
Gst, kg/sec	-81.718	-0.33	-78.715	10.697	10%
PG, %	80	99.91	80.707	2.62	10%

Table 2. Variables distribution of the control program which realizes the compromise-combined program.

TinB control program	MIN	MAX	MEAN	stDEV	Unique
AO, %	-3.413	-3.409	-3.411	0	3%
Q, %	87.223	100	90.375	3.015	99%
hSUZ, cm	318.091	319.307	318.49	0.448	67%
cJ, unit/cm^{-3}	4.18e^{16}	4.46e^{16}	4.36e^{16}	9.65e^{14}	100%
cXe, unit/cm^{-3}	3.42e^{16}	3.47e^{16}	3.44e^{16}	1.44e^{14}	100%
Tin, r.u.	-0.016	0.004	-0.001	0.003	20%
Tav, r.u.	-2.001	0	-1.508	0.471	98%
Tout, r.u.	-4.001	0	-3.014	0.942	99%
P, r.u.	-0.001	0.153	0.115	0.037	92%
Gst, kg/sec	-77.504	0.376	-58.347	18.322	99%
PG, %	81.004	100.031	85.708	4.504	99%

In the tables 1 and 2 you can see such parameter as unique. Unique is the percentage of instances in the data having a value for this attribute that no other instances have. For example, if dataset has such values as 2, 3, 5, 2, 5. The unique will be 3/5. Comparing each variable of both tables, it can be concluded that AO has more unique data in Tav program than in case of TinB program. Also, standard deviation (stDEV) of AO is zero in case of TinB program. It can be explained that AO is controlled by automation control system in case of TinB. The second variable Q of both tables has a large difference too. Q has more unique data in TinB program than in case of TinB program. The standard deviation (stDEV) of Q is almost the same in both control programs. hSUZ of Tav program has more unique data compared to hSUZ of TinB program, where 97% vs 67%. But, in case of stDEV, hSUZ of TinB program has less value than hSUZ of Tav, where 0.448% instead of 5.543%. It says that TinB program very rarely uses the control by control rods. As for cJ and cXe, the control programs have similar values and unique data is 100% in both cases. The sixth variable Tin of TinB program has the unique data 2 times more than Tin of Tav program and the stDEV is 100 times less. All rest variables such as Tav, Tout, P, Gst and PG of TinB have the unique data with value more than 90% and in the case of Tav program these variables have the unique data with value less than 12%. It says that TinB program has many changes in such variables while a load-following mode in contrast to the Tav program. As for the stDEV, Tav, Tout, P, Gst and PG of TinB program have a little bit higher values than in case of Tav program.

4.3. Data Preparation

As previously mentioned, the current study suggests making regression models of the average temperature of the coolant in the first circuit and the output electric power of generator parameters of the NPP with VVER-1000 in load-following mode. It needs to use the regression techniques to be sure that the data-sets, which will be used, can apply the appropriate DM approaches. However, the initial data, which was obtained by the simulation of the power unit in 8 hours of load-following mode from 100% to 80% and back, is not required to sorting the time data. Also, it is necessary to mention that the data has 11 attributes (i.e. variables). There is no need to forecast all the attributes, because only 2 attributes of the data are most important, although it is not allowed to completely exclude others. Concerning this study, such parameters were explored as the average temperature of the coolant in the first circuit, which is a special internal parameter of the reactor core and it objectively shows processes related to release energy; the output electric power of generator (PG), as well as the above-mentioned parameter, is one of the most important too, because it is very important to control, register and forecasting for economic and business points of view. From the above it follows that Tav (relative units) and PG (%) were used as forecasting targets.

4.4. Modelling

At this stage of the research it is very important to choose the best methods of DM. To solve this problem, regression algorithms of DM such as Linear Regression (LR), REPTree and M5P were chosen. Selection of these algorithms is based on the following characteristics: actual (simulated) data values, availability and efficiency of use [9,13,14]. The chosen algorithms completely correspond to above-mentioned characteristics. It should be mentioned one important aspect that the modelling stage is the implementation of testing mechanisms. The simulation was carried out by 10 Folds Cross Validation. A lot of references suggest using 10 Folds Cross Validation, because of good results [15].

The developed models can be represented by the following expression:

$$M_n = \{A_a; CP_b; V_c; TDM_d; T_e; SM_f\} \quad (1)$$

The model M_n belongs to an approach (A) regression and is composed by a type of control program (CP), variables (V), a DM technique (TDM) and a type of target (T) and a sampling method (SM):

$$A_a = \{Regression_1\} \quad (2)$$

$$CP_b = \{Tav_1, TinB_2\} \quad (3)$$

$$\begin{gathered} V_c = \{All_1, (AO, hSUZ, cJ, cXe, Tav, PG)_2, (Q, Tin, Tav, Tout, P, Gst)_3, (AO, Tav)_4, (Q, Tav)_5 \\ , (hSUZ, Tav)_6, (cJ, Tav)_7, (cXe, Tav)_8, (Tin, Tav)_9, (Tout, Tav)_{10}, (P, Tav)_{11}, (Gst, Tav)_{12}, (PG, Tav)_{13}, \\ (Q, Tin, Tout, P, Gst, PG)_{14}, (AO, PG)_{15}, (Q, PG)_{16}, (hSUZ, PG)_{17}, \\ (cJ, PG)_{18}, (cXe, PG)_{19}, (Tin, PG)_{20}, (Tav, PG)_{21}, (Tout, PG)_{22}, (P, PG)_{23}, (Gst, PG)_{24}\} \end{gathered} \quad (4)$$

$$TDM_d = \{LR_1, M5P_2, REPTree_3\} \quad (5)$$

$$T_e = \{Tav_1, PG_2\} \quad (6)$$

$$SM_f = \{10 - Folds\ Cross\ Validation_1\} \quad (7)$$

Using this notation for representing DM models it is possible to present a particular model implemented. For instance, the model (M_1) follows the regression approach using the data from Tav control program and all variables from the data, the technique LR with Tav target and the sampling method 10-Folds CV and it is expressed by:

$$M_1 = \{A_1; CP_1; V_1; TDM_1; T_1; SM_1\} \quad (8)$$

Accordingly, for each control program with one technique 26 combinations were designed. In total 156 models were obtained using the expression above.

5. Discussion

Plenty numbers of models were obtained during the experiment and the results were very satisfactory. More than 90% of the results with values of the RRSE metric were dropped and then the best 24 models were totally obtained. The results of these 24 models are shown in tables 3 and 4.

Table 3. The best results of Tav control program.

Scenario	Variables	Target	Algorithm	CC	MAE	RMSE	RAE	RRSE	TNI
1	All	Tav	LR	0.3809	0.0034	0.007	161.846	92.4518	10356
2	AO, hSUZ, cJ, cXe, Tav, PG	Tav	LR	0.3809	0.0034	0.007	161.846	92.4518	10356
3	Q, Tin, Tav, Tout, P, Gst	Tav	LR	0.0089	0.0021	0.0076	98.876	100.022	10356
4	All	Tav	M5P	0.9376	0.0004	0.0029	16.8479	38.1124	10356
5	AO, hSUZ, cJ, cXe, Tav, PG	Tav	M5P	0.7157	0.0006	0.0053	26.576	69.8365	10356
6	Q, Tin, Tav, Tout, P, Gst	Tav	M5P	0.2754	0.0013	0.0145	60.7543	191.152	10356
7	All	Tav	REPTree	0.8447	0.0004	0.0041	18.3429	53.5745	10356
8	AO, hSUZ, cJ, cXe, Tav, PG	Tav	REPTree	0.7953	0.0004	0.0046	19.9068	60.639	10356
9	Q, Tin, Tav, Tout, P, Gst	Tav	REPTree	0.7749	0.001	0.0048	47.1536	63.5455	10356

10	Tav, PG	PG	LR	0.017	1.2749	2.6204	99.9385	100.009	10356
11	Tav, PG	PG	M5P	0.7777	0.516	1.6509	40.4485	63.0077	10356
12	Tav, PG	PG	REPTree	0.9403	0.1445	0.8937	11.3278	34.1109	10356

Table 4. The best results of TinB control program.

Scenario	Variables	Target	Algorithm	CC	MAE	RMSE	RAE	RRSE	TNI
13	All	Tav	LR	1	0	0	0.0001	0.0001	9848
14	AO, hSUZ, cJ, cXe, Tav, PG	Tav	LR	0.9999	0.0032	0.0049	0.9591	1.0424	9848
15	Q, Tin, Tav, Tout, P, Gst	Tav	LR	1	0	0	0.0001	0.0001	9848
16	All	Tav	M5P	1	0.0005	0.0012	0.1553	0.2637	9848
17	AO, hSUZ, cJ, cXe, Tav, PG	Tav	M5P	1	0.0007	0.0024	0.2215	0.5161	9848
18	Q, Tin, Tav, Tout, P, Gst	Tav	M5P	1	0.0005	0.0013	0.161	0.2715	9848
19	All	Tav	REPTree	0.9997	0.0089	0.0106	2.7008	2.2417	9848
20	AO, hSUZ, cJ, cXe, Tav, PG	Tav	REPTree	0.9997	0.0092	0.0112	2.7976	2.381	9848
21	Q, Tin, Tav, Tout, P, Gst	Tav	REPTree	0.9997	0.0089	0.0106	2.7018	2.2421	9848
22	Tav, PG	PG	LR	0.9999	0.034	0.07	1.0761	1.5538	9848
23	Tav, PG	PG	M5P	0.9999	0.0271	0.0681	0.859	1.5124	9848
24	Tav, PG	PG	REPTree	0.9996	0.0979	0.1218	3.1038	2.7037	9848

The models in case of Tav control program show the RRSE metric more than 34% as we can see from tables 3 and 4. In case of TinB control program the same scenarios show very good results, where the RRSE metric is less than 2.8%. The correlation coefficients of all scenarios in case of TinB control program have values higher than 0.9995. The correlation coefficients in case of Tav control program are very bad in comparison with TinB control program, because the values of the CC are less than 0.7778. At the same time, few scenarios with numbers 4,7 and 12 in case of Tav control program have not bad values of the correlation coefficients, since scenario 4 has 0.9376, scenario 7 – 0.8447 and scenario 12 – 0.9403. Moreover, you can see that the tables 5 and 6 have only 6 models with PG target. This is because in both cases of control programs the models, which were obtained, have very bad results of the correlation coefficient as the RRSE metric. However, it seems to me that the results with PG target show us the advantages of TinB control program. In addition, the variables, which are mentioned in tables 3, 4 and have algorithms such as LR and M5P, have shown good regression values of CC and RRSE metrics. Where CC equals 1 and it equals 0.9999 only in two cases. It means that LR and M5P algorithms have shown best regression results. It can be explained by the fact that in LR and M5P algorithms are used linear functions to predict a target meanwhile REPTree is a fast decision tree learner, which builds a regression tree using information gain as the splitting criterion, and prunes it using reduced error pruning. As was already mentioned, data-sets were obtained by simulation with a load-following mode and it means that whole data-sets have a “linear” structure. Thus, any algorithms, in which a linear function is used, will be most suitable to these data-sets.

6. Conclusion and Future Work

The tables 3 and 4 show the best results with the CC and the RRSE metrics for the two control programs of NPP with VVER-1000. Based on the best results it can be concluded that the improved compromise-combined control program shows better results of regression models using DM techniques than the control program with the constant average temperature of the coolant in the reactor core. As it mentioned above the best results with PG target has in total 6 models, 3 models in case of the improved compromise-combined control program and 3 models in case of the constant average temperature of the coolant in the reactor core. It can be explained by the fact that almost all results were higher than 10% of the RRSE metric and thus a lot of models with PG target were dropped.

In this study, the best regression models are used to predict some parameters of NPP. Besides, it is one of easiest ways to predict parameters of the nuclear object as VVER-1000. The study is allowed to obtain good results of regression models with the cooperation of the multi-zone mathematical model simulation. That is why the best obtained regression models can be exploited in complicated calculations of expected state parameters and in control of NPP. The models are able to expect exactly state of a related parameter and it simplified further researches of a control in the nuclear area. Basically, the obtained regression models can be used further in deep researches related to a load-following mode at NPP. Also, the achieved results allow to obtain more detailed information about internal processes in frame of two explored control programs. Because only variables, which are mentioned in tables 3 and 4, have shown good regression metrics such as CC and RRSE. The applied DM technic such as M5P has shown excellent results of CC in comparison of other technics. However, to achieve provided results were used 3 DM technics and there is the need to use more regression in further research works because only 24 of 154 good models were obtained.

This study has shown that using DM techniques can demonstrate that forecasting of the average temperature of the coolant in the first circuit and the output electric power of generator parameters are better in case of the improved compromise-combined control program. Moreover, data-sets of both control programs of NPP with VVER-1000 in load-following mode were used to obtain the regression models. The free software Weka was used to extract useful data and as DM algorithms such as Linear Regression, REPTree and M5P were used too. Thus, this research work proves that the use of DM can say which regression models of the average temperature of the coolant in the first circuit and the output electric power of generator in case of two control program of NPP with VVER-1000 based on data-sets in a load-following mode, is better.

In the future some new experiments will be performed taking in consideration the use of ensembles and ubiquitouse data mining applied to Intelligent Decision Support Systems [15, 16]. At same time new agents can be developed for monitoring the entire process [17].

Acknowledgements

This work has been supported by FCT - Fundação para a Ciência e Tecnologia within the Project Scope UID/CEC/00319/2013.

References

1. Lokhov A. *Load-following with nuclear power plants* [Electronic resource] / A. Lokhov // NEA updates, NEA News, Is. 29(2), 2011.
2. Pelykh S., Maksimov M., Tsiselskaya T., Baskakov V. *A way of the WWER-1000 neutron field axial distribution stabilization at power manoeuvring.* Nuclear & Radiation Safety, Vol. 49, Is. 1., 2011.
3. Pelykh S., Maksimov M., Nikolsky M. *A method for WER fuel element cladding reliability prediction.* The Nuclear Physics and Atomic Energy journal, Vol. 15, Is. 1, 2014.
4. Todortsev Yu., Tsiselskaya T., Nikolskiy M. *Automated power control system for an NPP unit in power follow modes with constant coolant temperature at the reactor inlet.* The Nuclear & Radiation Safety journal, 2013.
5. Foshch T., Maksimov M., Nikolsky M. *Analysis of the influence of methods for controlling power unit with a pressurized water reactor for axial offset.* Eastern-European journal of enterprise technologies, Vol. 2, Is. 8(68), P.19-27, 2014.
6. Maksimov M., Pelykh S., Tsiselskaya T., Baskakov V. *Control method for water-type reactor nuclear power plant with changing reactor power or external load.* Patent number 59039. 1PC G21C 7/00, 2011.
7. Maksimov M., Beglov K., Tsiselskaya T. *A model of a power unit with VVER-1000 as an object of power control.* Odessa. Ukraine: Works of the Odessa polytechnic university, Vol. 38, Is. 1, 2012.
8. Maksimov M., Tsiselskaya T., Foshch T. *An improved automated system control of power unit PWR (VVER-1000) for use while manoeuvring in the daily cycle.* Informational technologies and automation. VI Ukrainian scientific-practical conference, 2013.
9. Ian H. Witten, Eibe Frank, Mark A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques (3rd ed.).* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2011.
10. Kuechler B, Vaishnavi V. *On theory development in design science research: Anatomy of a research project*. European Journal of Information Systems. Vol.17, Is.5, P.489–504, 2008.
11. David L. Olson Dr., Dursun Delen Dr. *Advanced Data Mining Techniques.* Springer-Verlag Berlin Heidelberg, 2008.
12. Zontul M., Aydin F., Dogan G., Sener S., Kaynar O. *Wind Speed Forecasting Using RepTree and Bagging Methods in Kirklareli – Turkey.* Journal of Theoretical and Applied Information Technology, 56 (1), pp. 17-29, 2013.
13. C. Zhan, A. Gan and M. Hadi. *Prediction of Lane Clearance Time of Freeway Incidents Using the M5P Tree Algorithm*. IEEE Transactions on Intelligent Transportation Systems. Vol. 12, no. 4, pp. 1549-1557, 2011.
14. Refaeilzadeh P., Tang L., Liu H. *Cross-Validation.* Encyclopaedia of Database Systems, 5 vols. Springer, 2009.
15. Filipe Portela, Manuel Filipe Santos, José Machado, Álvaro Silva and António Abelha. Pervasive and Intelligent Decision Support in Critical Health Care using Ensemble. Lecture Notes in Computer Science (LNCS) - Information Technology in Bio- and Medical Informatics. Volume 8060, 2013, pp 1-16. ISBN: 978-3-642-40093-3. Springer. DOI: 10.1007/978-3-642-40093-3_1. 2013.
16. Santos, Manuel · Portela, Filipe. Enabling Ubiquitous Data Mining in Intensive Care - Features Selection and Data Pre-Processing. ICEIS 2011 - International Conference on Enterprise Information Systems. pp 261-266. ISBN: 978-989-8425-53-9. China. SciTePress. (2011).
17. Luciana Cardoso, Fernando Marins, Carlos Portela, Manuel Santos, António Abelha and José Machado. The Next Generation of Interoperability Agents in Healthcare. International Journal of Environmental Research and Public Health. ISSN: 1660-4601. DOI: 10.3390/ijerph110505349 . MDPI. 2014