CIGRÉ 2004

OPERATION RULES DETERMINED BY RISK ANALYSIS FOR SPECIAL PROTECTION SYSTEMS AT HYDRO-QUÉBEC

J.A. Huang*
G. Vanier
A. Valette

Institut de Recherche d'Hydro-Québec
(Canada)

S. Harrison
F. Lévesque

Hydro-Québec TransÉnergie
(Canada)

L. Wehenkel

University of Liege
(Belgium)

Abstract: This paper describes a new approach used by Hydro-Québec to determine the rules of automatic devices installed in its main power plants to maintain secure operation under extreme contingencies. An example of application of this approach is given to illustrate how to apply data mining technique for the rules of the automatic generator rejection and remote load shedding system (RPTC: **R**ejet de **P**roduction et **T**élédélestage de **C**harges in French) installed at the Churchill Falls hydroelectric power plant (5500 MW) in Labrador. Real time snapshots of the Hydro-Québec power system collected over several years data have been used to generate large amounts of results (database) by transient stability simulations. The database is processed by the data mining software developed by the University of Liege to construct the decision trees. This approach gives the most relevant parameters and finds optimal settings for the RPTC system at the Churchill Falls, minimizing the number of generator rejection while maintaining the same performance in terms of security coverage. New operation rules have thus been established.

Keywords: Protection systems - Automatisms - Dynamic Security - Risk - Data Mining - Stability.

1 INTRODUCTION

The operation criteria [1] at Hydro-Quebec require that the power system remains stable without any assistance of Special Protection Systems (SPS) following normal contingencies (Table i)[2]. These normal contingencies are hence used to determine the secure transfer limits for the various corridors of the system, in compliance with the Northeast Power Coordinating Council (NPCC) criteria.

Table i: Normal and extreme contingencies

Normal Contingencies	Three phase fault with normal clearing
	Single line to ground fault with delayed clearing
	Breaker fault with normal clearing
	Loss of a bipolar dc line
	Loss of double-circuit line
	Loss of any element without fault
Extreme Contingencies	Single line to ground fault with loss of two series or parallel 735 kV line
	Loss of all 735 kV lines emanating from a substation
	Loss of all lines in a corridor
	Loss of two parallel 735 kV lines and bypass of all series capacitors on the remaining line in the same corridor

In addition to these requirements, Hydro-Québec considers that it is also important for the system to remain stable after certain extreme contingencies (Table i)[2]. The system stability is maintained by Special Protection Systems. According to the event and the configuration of the power system, the special protection system – the RPTC system ("**R**ejet de **P**roduction et **T**élé délestage de **C**harge" in French) activates the generator rejection and remote load shedding scheme. Figure 1 gives an overview of the basic structure and the general operation of the RPTC system. RPTC systems are installed in fifteen 735 kV substations of the Hydro-Québec system. The subsystems in a same corridor (or main axis) are combined into an independent group. There are a total of 4 groups shown in dark shaded areas in Figure 1. Each group of RPTC systems performs, associated with independent Special Protection Systems, the generation rejection scheme at one particular generation site while the remote load shedding function is centralized.

*Email: huang.jinan@ireq.ca

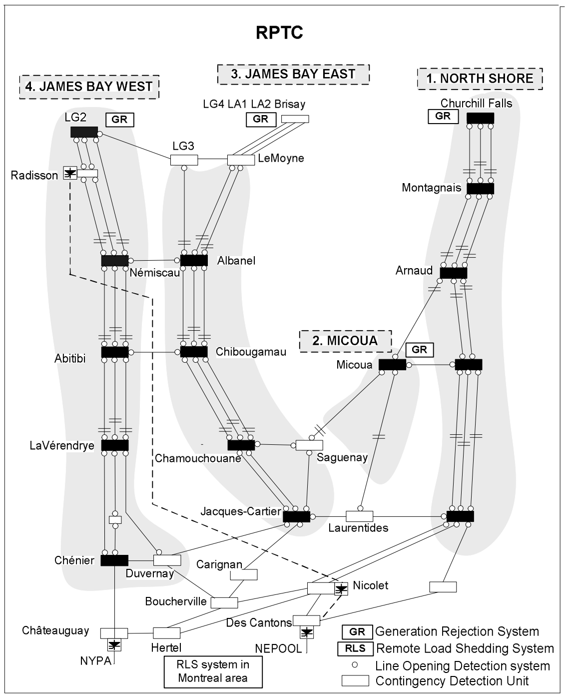

Figure 1: RPTC system

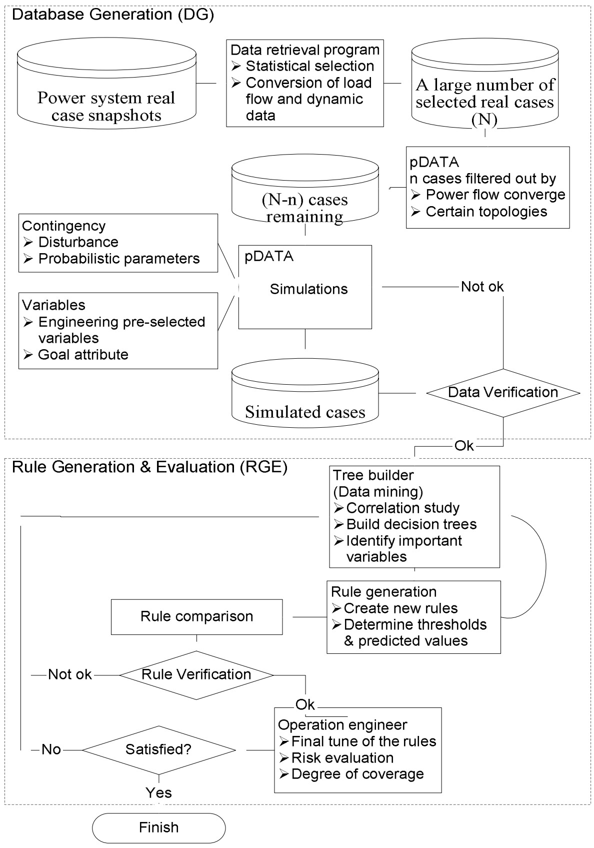

Figure 2: Flow chart of rule generation for SPS

eliminating some tree nodes. Results of rules are compared and validated by their precision and degree of coverage for stability. The one rule which gives better results is chosen for the special protection system.

This last set of rules is reviewed by the operation planning engineers in charge of these devices and is used to elaborate the final set of rules to be implemented in the field. Some considerations (like reliability of the relevant parameters or variables inputs or like future addition of new equipment or removal of equipment for maintenance, not covered adequately by the data mining process) are then taken into account to make some final adjustments to the rules. This final set of rules is then validated with a large independent set of real case snapshots by simulation and compared with the rules used in the field. The risk level is also evaluated by degree of coverage for stability. If the results are satisfactory for the operation planning engineers, then the operation rules are programmed and updated in the control centre.

The construction of decision trees has been achieved by using data mining software [8].

3 DATABASE GENERATION

3.1 Database generation approach

The statistical approach used in this work requires the processing of a very large quantity of results generated by numerous scenario simulations. Each scenario is composed of a power flow snapshot of the network with the disturbances. The simulations have typically a 10 second time frame and are performed on a PC network using an in-house transient stability program. The approach has to generate pessimistic scenarios in order to cover adequately the critical situations where the RPTC automatic device operates with a good variance on the critical parameters and variables.

Two approaches have been initially envisioned:

- In the first approach, the generation of scenarios is done from a limited number of power flow base cases corresponding to real operating situations. These cases are then modified according to certain rules and the corresponding scenarios are simulated in order to create many critical situations for the RPTC system.
- In the second approach, the scenarios are generated from snapshots of real operating cases taken periodically over a long period of time (years) and disturbances critical for the RPTC system are simulated.

Particular care has to be taken in the data generation process to avoid overrepresentation of non-relevant cases.

The results from the first approach are biased due to the overrepresentation of critical situations with in fact very low probability. This could be corrected only if probability data are available on disturbances and/or operating conditions. Therefore, the second approach, which has been retained in this work, seems more appropriate due to the fact that all operating cases used are real and can be selected to cover adequately the envisioned power system conditions.

3.2 Database generation program

As shown in Figure 2, an extraction and a conversion of data is first accomplished. The data conversion is needed to allow simulations of power system real snapshots stored in the control center database.

For this task, in-house data retrieval program [7] is used to generate snapshots for planning and operation planning engineers.

Control software (pData) was developed to filter snapshot cases in order to keep just the relevant ones (cleaning process). For each filtered case, pData software builds the disturbance to be simulated as a function of the peculiarities of the studied case.

In this process, pData associates a random value to certain parameters of the disturbance in order to take into account the effects of these variations on the results. These parameters are the fault clearing time corresponding to the line re-closing time (breaker operation) and the series compensation by-pass time.

From transient stability simulations, pData determines by an iterative procedure, for each case, the goal variable. The goal variable is depending on the type of studies, which could be the minimum number of units to be tripped, the amount of loads to be shed, or both.

Finally, pData extracts results and saves some engineering pre-selected relevant parameters and variables to simulated database which will be used for Rule Generation and Evaluation. It is important to validate the correctness and validity of the data. Any errors in the data will lead to wrong results. Quite often, results have to be re-simulated due to missing information or errors.

4 RULE GENERATION AND EVALUATION

Tree construction is based on data mining techniques. Classification induction methods are used to construct decision trees. Based on the information provided by decision trees, rules are generated.

4.1 Data mining technique

Data mining refers to the extraction of high-level synthetic information (knowledge) from databases containing large amounts of low-level data. It is also called Knowledge Discovery in Databases (KDD). Data mining has received a wide range of applications in recent decades, for example in medical diagnosis, in character recognition, as well as in financial and marketing problems. The main reason for the important breakthrough is the tremendous increase in computing power. This makes possible the application of the often very computationally intensive data mining algorithms to practical large-scale problems. Nowadays, data mining techniques are also used in solving power system problems such as security assessment [3]-[5].

Data mining involves an integration of techniques from multiple disciplines such as database technology, statistics, machine learning, high-performance computing, pattern recognition, neural networks and so on. Many methods have been developed in the field of data mining. The present paper focuses on using the decision/regression tree type methodology to optimize RPTC system rules for generator tripping in the context of emergency control at Hydro-Québec.

4.2 Decision/regression trees

A decision tree (DT) is a map of the reasoning process. This data mining technique is able to produce classifiers about a given problem in order to deduce information for new, unobserved cases. The DT has the hierarchical form of a tree structured upside-down and is built on the basis of a Learning Set (LS). The LS comprises a number of cases (objects). Each case consists of pre-classified operating states (described by a certain number of parameters called candidate attributes), along with its correct classification (called the goal attribute). The candidate attributes characterize the pre-disturbance operating points in terms of parameters which can be used to make decisions. The tree building process seeks to build a set of rules relating these attributes to the goal attribute, so as to fit the learning set data well enough without over-fitting this data. The resulting tree is tested on a different data set (test set) where the prediction of the goal attribute by these rules is compared with the true class (determined by simulation) for each test case. The classification error rate for the test set measures if the method is successful or not.

There are many reasons to use decision trees. The first is their interpretability. A tree structure provides the information of how an output is arrived at. Another very important asset is the ability of the method to identify among the candidate attributes the most relevant parameters for each problem. A last characteristic of decision trees is its computational efficiency. The particular decision/regression tree induction method used in this paper is described in details in [8].

4.3 Rule evaluation

As mentioned earlier, decision trees can identify the most relevant parameters among the candidate attributes. Based on this information, many trees are constructed by selecting different set of relevant parameters. These trees are then translated into rules (if else) and put into excel sheets for tuning and comparing the performance. Performances of the rules are evaluated by their precision and risk level. The precision is measured differences between predicted variables and simulated variables. The risk level is the degree of coverage for stability.

Rules given directly by decision trees are unbiased estimates of the true value. Results could lead in unstable cases. Therefore, it is necessary to tune the rules in order to eliminate the unstable cases as many as possible while maintaining the precision. Tuning process is done by adjusting thresholds for each parameters and modifying values at terminal nodes of trees. Some weighting factors can be used to penalize the case where the predicted variables are far from simulated variables. If rules are not performed well, then the process has to go back to the tree builder to construct more trees.

The risk level is evaluated by the percentage of number of unstable cases given by the rule over total number of cases in the database containing simulation results. High level coverage for stability means high percentage of stable cases given by the rule.

The rate of acceptable coverage varies from event to event. It has to consider the consequences and the probability of occurrence of studied events. If the event leads to a loss of synchronism for the main grid, a high coverage rate will be required. Further more, if this event has already occurred in the past, an even higher coverage rate shall be required. On the other hand if the event does not lead to a loss of synchronism on the main grid, but to a frequency variation within acceptable limits, a lower coverage rate can be accepted. These considerations are qualitative and are based on the judgment and the experience of operation planning engineers. Generally, the coverage rate given by the new rule should be higher than the existing rule.

Table ii: Generator unit tripping scheme

Modulation: 3 links at Churchill Falls	
Margin At Churchill Falls:	Units to be tripped
400 MW and less	8
401 to 700 MW	7
701 to 1000 MW	6
1001 to 1200 MW	5
1201 to 1400 MW	4
1401 to 1600 MW	3
1601 to 1800 MW	2
1801 to 2000 MW	1
2001 MW and more	0

5 EXAMPLE

5.1 Event used for this study

The event used for this study is detected by the RPTC at the Churchill Falls substation and is particularly severe for the network. The following describes the disturbances:

- Single line to ground fault at the Churchill Falls substation;
- Tripping of two parallel lines between the Churchill Falls and Montagnais substations;
- Bypass of the series compensation bank of the remaining line in the same corridor.

To ensure network security and to avoid an unnecessary generation tripping, the number of generation units associated with the generation tripping scheme is adjusted to the loading and the configuration of the Churchill Falls – Arnaud corridor (North Shore indicated in Figure 1). Table ii presents the existing rules determined by the conventional approach. The margin represents the difference between the maximum power transfer considering normal contingencies and the measured power transfer. The table gives the number of units to be tripped based on the margin. For example, if the margin is 400 MW and less, then tripping of 8 units is required. If the margin is above 2001 MW, then no unit tripping is required.

In this study, a method of analysis based on the probabilistic approach will permit:

- To establish the coverage of the current rules;
- To establish the most sensitive variables that affect the network stability for this extreme event;
- To optimize the current rules;
- To suggest an algorithm to modulate the number of generation units to trip.

Table iii: Statistical data of unnecessary tripped units from the generated database

# Of Units to Trip	Number Of Cases			Over-tripping		
	With Current Rules	Of Matched Tripping	With Over-tripping	# of Units	Average per case	Rate (%)
8	2130	205	1925	5643	2.65	33
7	647	93	554	1818	2.81	40
6	626	15	611	2288	3.65	61
5	135	0	135	625	4.62	93
4	278	0	278	1047	3.77	94
3	159	0	159	477	3.00	100
2	157	0	157	314	2.00	100
1	58	0	58	58	1.00	100
0	370	370	0	0	0.00	0
Total	4560	683	3877	12270	2.70	44

The methodology used is as follows:

- Extraction of 10 000 network cases spread over several years;
- Filtering cases to select only ones with topologies of 3 lines between Churchill Falls and Montagnais. 4600 cases are remaining;
- Creation of each case with a random fault duration and a random time of bypassing a capacitor;
- About 13 000 stability simulations were performed to find the minimum number of units to be tripped to ensure system stability after this event for each remaining cases, and some 236 variables are pre-selected and saved for the database used in the data mining;
- Optimization of the number of units to be tripped with a decision tree.

5.2 Correlation studies

Using the generated database, correlation studies were performed. Figure 3 shows the correlation of transfer and transfer margin on the Churchill Falls corridor vs minimum generator unit tripping as determined by the time-domain simulations. The dashed horizontal line in Figure 3 shows, for example, that the currently used-rules require to trip 8 units if the transfer margin on the Churchill Falls corridor is less than 400 MW (see Table ii). From the generated database (4560 cases), there are about 2130 cases (see Table iii) in which the current rules require 8 units to be tripped, but actually, we can see from Figure 3 that if the transfer on the Churchill Falls corridor is less than 3300 MW (dashed vertical line), there is no need to trip generator units at all. If we count the number of cases for which the time domain simulations determine it is necessary to trip 8 units, we find only 205 cases. Thus, among the 2130 cases for which the current rules tell us to trip 8 units, there are 1925 cases which need less than eight (down to zero) unit tripping. More synthetically, if we count among these 2130 cases the difference between the number of units required to trip by the current rules and the actual number of units necessary to be tripped according to the simulations, it is found that 33% of generator units (5643 units) are unnecessarily tripped with the currently used 8 units tripping rules.

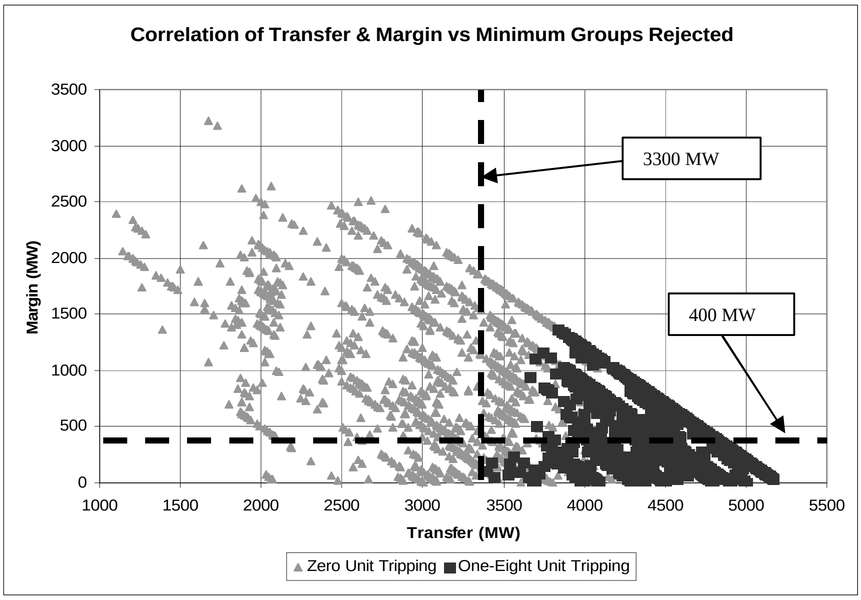

Figure 3: Correlation of transfer and margin at Churchill Falls corridor

Table iii shows some other statistical data from the generated database. It can be observed that among 4560 cases the average of generator units over-tripped is about 2.70 per case and 44% of generator units are over-tripped. From this, it can be seen that the current rules are highly conservative and could possibly be improved by taking into account not only the margin but also the total power transfer through the Churchill Falls corridor in their formulation.

5.3 Regression tree

Constructions of regression trees were carried out on the generated database. Among 4560 objects (cases), 2000 objects were selected as a learning set and the remaining 2560 objects were comprised as a test set. The goal is to predict the minimum number of generator units to be tripped. Figure 4 shows a constructed regression tree. The tree is to read top-down: each internal node corresponds to a test on one of the candidate attributes and the terminal nodes correspond to decisions about the number of units to be tripped. These nodes are sorted left to right by increasing number of units to be tripped. For example the left-most terminal node (denoted T4) corresponds to 600 cases for which the expected number of units to be tripped is 0.04167. A case will be directed to this node if T_CHU_MONT < 3899 MW. On the other hand, the right-most terminal node (denoted T33) corresponds to an expected number of units to be tripped of 7.032 with the following 4 conditions:

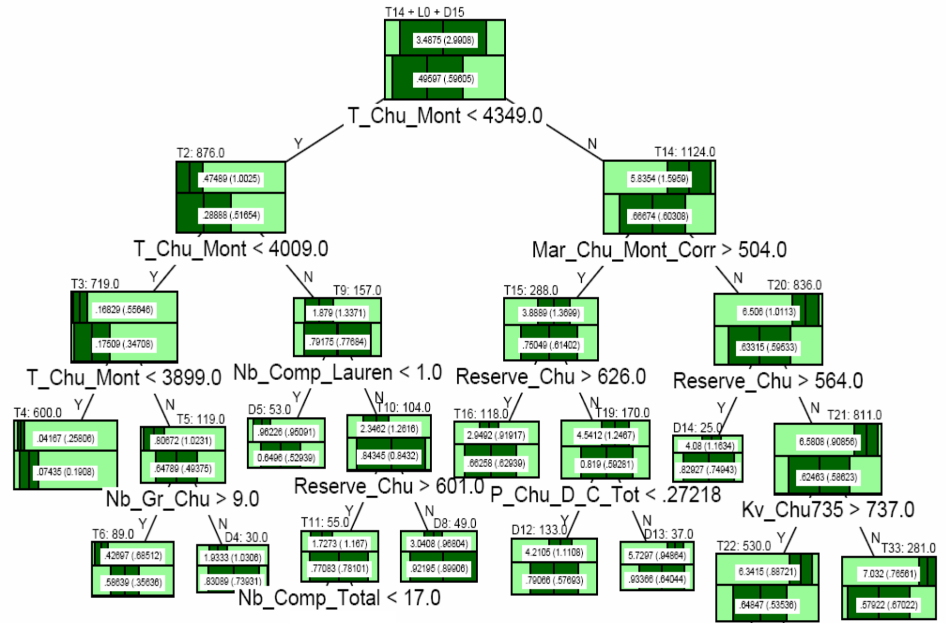

Figure 4: Regression tree to predict number of generator unit tripping

1) The transfer on the Churchill Fall Corridors T_CHU_MONT > 4349 MW;
2) The transfer margin on the Churchill Falls corridor Mar_Chu_Mont_Corr < 504 MW;
3) The spinning reserve at the Churchill fall Reserve_Chu < 564 MW;
4) The voltage at the Churchill Falls KV_CHU735 < 737 kV.

Notice that among the 236 candidate attributes proposed to the tree building software, only 8 attributes were identified as important variables to decide on the number of units to be tripped. Notice also that in order to translate the rules provided by the tree into decision rules it is obviously necessary to convert the fractional predictions into integer numbers (e.g. by rounding up to the nearest larger integer value – ceiling function).

By construction, the predictions of the regression tree are unbiased estimates of the true values; this means that their errors are both negative and positive. In practice, it may be preferable to have rules that have less negative errors (too few unit tripping – under-tripping) than positive ones (too many – over-tripping) because the cost of instability is much higher than the cost of unnecessarily tripping one or two more units. Such a bias can be introduced as a post-processing of the regression tree output, for example by adding some positive constant to its predictions before rounding up to the nearest integer.

Table iv: Average of over-tripped units per case

Methods Used	Average
Current Rules	2.70
Regression Tree with Post-processing	1.08
New Rules with two variables	1.62

5.4 New Rules

Although the regression tree gives much better results over the existing rules (see in section 5.5), the under-tripping rate is still too high and cannot be accepted by operation planning engineers. From the regression tree, it is observed that some important variables appear on the tree such as the transfer margin on the Churchill Falls corridor and the spinning reserve at Churchill Falls generation station. The new rules based on these two variables are established and shown in Table v. The new rules can also be illustrated in Figure 5. The reason to choose the margin and spinning reserve is that the margin represents two variables: transfer and limit which implies the network topology, while the spinning reserve represents both the number of generators on line and the amount of power generated in the plant. It is also believed that rules expressed in terms of these two variables are more relevant for operation planning engineers since they should be more robust with respect to future addition of new equipments or removal of equipments for maintenance.

Table v: The new rules combining two variables

Rules for spinning reserve at Churchill		Rules for transfer margin		New rules combining the two variables
Values of reserve	Number of gr. to trip (A)	Values of margin	Number of gr. to trip (B)	
≤ 500	8	≤ 600	8	Min(A,B)
501 to 600	7	601 to 800	7	
601 to 700	6	801 to 900	6	
701 to 900	5	901 to 1000	5	
901 to 1000	4	1001 to 1100	4	
1001 to 1100	3	1101 to 1200	3	
1101 to 1200	2	1201 to 1300	2	
1201 to 1300	1	1301 to 1600	1	
≥ 1301	0	≥ 1601	0	

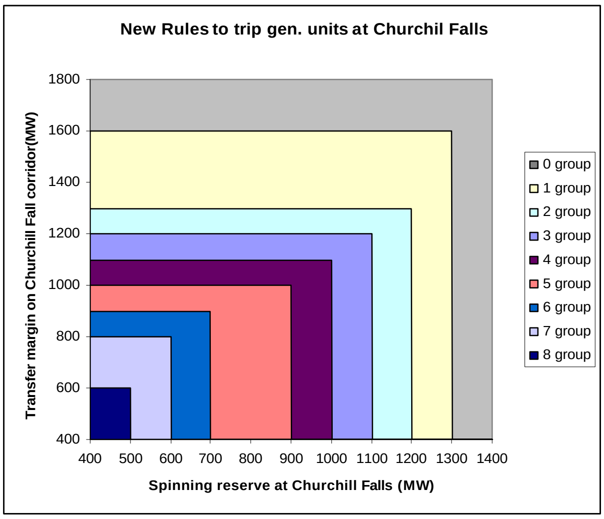

Figure 5: New rules combining two variables

5.5 Comparison of results

Table iv lists the average of over-tripped units per case. Although this value may not have a direct physical meaning, it is used here as an indication of the improvement of the different methods tested. The first line of the table refers to the rules actually in use, and designed by the classical deterministic method. The second line gives the results obtained by rules of a post-processing of the regression tree of Figure 4. Third line gives the performance of the new rules designed by hand from taking into account both margin and spinning reserve at the Churchill Falls power plant. The post-processing consists in adding a positive bias of 0.45 to the predictions of the tree and round up to the nearest integer (ceiling function). It can be seen that the regression tree has the least average value of over-tripped generator units per case. This means that if the regression tree rules are implemented, the number of generator unit tripping will be closest to their minimum among all other methods.

Figure 6 shows the frequency diagram of mis-tripped generator units for different methods. The term of “mis-tripped” unit is defined as the difference between simulated optimal unit tripping and the number of units tripping prescribed by a rule. A positive value means generator units are over-tripped while negative value means under-tripped with respect to the value determined by simulations. From this diagram, one can observe that the distribution of mis-tripped units by the current rules is widely spread while that of the regression tree is concentrated. In most cases, the regression tree gives one generator unit over-tripping while the current rules sometimes gives 8 generator units over-tripping. The reason that the regression tree mis-tripping is concentrated around one is, as mentioned previously, that the post-processing is applied to the regression tree of Figure 4. This post-processing adds a positive value of 0.45 to the predictions of the tree before applying the ceiling function in order to eliminate most of the generator unit under-tripping. Therefore, it appears that in most cases the regression tree settles on one generator unit over-tripping. The results from the regression tree are very promising, but more efforts have to be made to eliminate under-tripping cases.

The new rules combining two variables further eliminate the under-tripping cases. Despite it sacrifices the accuracy of generator tripping; the results are more reliable and acceptable by operation planning engineers. It also reduces the error of over-tripping cases. Furthermore, the new rules are quite simple and easy to be implemented.

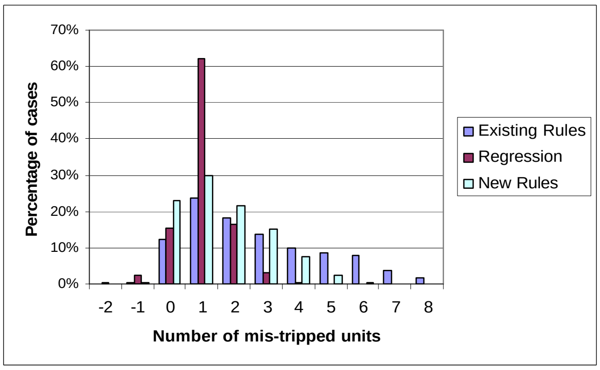

Figure 6: Comparison of results of different methods

6 CONCLUSIONS

A new approach to improve settings of special protection devices used to protect Hydro-Québec power system against extreme contingencies has been introduced. This approach applies data mining technique to real case snapshots of Hydro-Quebec power system in a probabilistic way.

This approach is illustrated here with a detailed description of its application to an emergency control automat, the RPTC SPS device installed at the Churchill Falls hydroelectric power station (5500 MW) in Labrador. The data mining technique was applied to the results of some 13 000 network simulations. Various network states were taken from a real-time database and were simulated using power system analysis software. The data cases represent actual operating states collected over several years. By using the data mining technique, the most relevant parameters for this automat were identified and effective settings were determined.

A correlation analysis and the construction of regression trees were carried out on the results of these simulations using data mining software. This analysis made it possible to minimize, in particular, the number of generators tripped by the RPTC system for a large number of network conditions, while maintaining the same performance in terms of security coverage. New operation rules can thus be established by operation planning engineers and will be implemented.

Following these very encouraging results, other applications of these methods are being considered at Hydro-Québec in the coming years.

7 REFERENCES

[1] J-P. Gingras, R. Mailhot, J. Lavergne, “Bilan de l’exploitation du réseau d’Hydro-Québec avec compensation série,” CIGRÉ, 38-101 session 2000.
[2] G. Trudel, S. Bernard, G. Scott, “Hydro-Québec’s defence plan against extreme contingencies,” IEEE, PE-211-PWRS -0-06-1998
[3] L. Wehenkel, M. Pavella, “Advances in Decision Trees Applied to Power System Security Assessment,” IEE 2nd International Conference on Advances in Power System Control, Operation and Management, December 1993, Hong Kong.
[4] S. Rovnyak, S. Kretsinger, J. Thorp, D. Brown, “Decision Trees For Real-Time Transient Stability Prediction,” IEEE Transactions on Power Systems, Vol. 9, No.3, August 1994.
[5] N.D. Hatziargyriou, S.A. Papathanassiou, M.P. Papadopoulos, “Decision Trees For Fast Security Assessment of Autonomous Power Systems With a Large Penetration From Renewable,” IEEE Transaction on Energy Conversion, Vol. 10, No.2, June 1995
[6] Y. Schlumberger, C. Levrevelec, M. De Pasquale, “Power Systems Security Analysis New Approaches Used at EDF,” IEEE Summer Meeting, July 21, 1999, Edmonton, Canada.
[7] L. Riverin, J.A. Huang, A. Valette, F. Lafrance, “Basic Requirement for a System of On-line Stability Limit Analysis at Hydro-Québec,” International Conference on Power System Technology, October 18-21, 1994, Beijing, China.
[8] L. Wehenkel, Automatic Learning Techniques in Power System, Kluwer Academic, 1998