Application of Data Mining to Optimize Settings for Generator Tripping and Shedding System (RPTC) in Emergency Control at Hydro-Québec

J.A. Huang*, S. Harrison**, G. Vanier*, A. Valette*, L. Wehenkel***

* IREQ, Hydro-Québec
1800 Lionel-Boulet
Varennes, Québec
Canada J3X 1S1

** TransEnergie, Hydro-Québec
Case Postale 10000
Montreal, Québec
Canada H5B 1H7

*** University of Liege
Dept. of EE & CS
Institut Montefiore
B-4000 Liege, Belgium

Abstract: This paper describes the on-going work done by Hydro-Québec to optimize the settings of automatic devices installed in its main power plants to maintain secure operation under extreme contingencies. The automatic generator tripping and load shedding system (RPTC) described in this paper is installed at the Churchill Falls hydroelectric power plant (5500 MW) in Labrador. Data mining techniques like decision trees (DT) and regression trees (RT) have been used. Real time snapshots of the Hydro-Québec power system collected over a 5 year period have been used to generate large amounts of results by transient stability simulations. The processing of these data has been done using software developed by the University of Liege. This approach gives the most relevant parameters and finds optimal settings for the RPTC system, minimizing the number of tripped generator units while maintaining the same performance in terms of security coverage. New operation rules can thus be established.

1 INTRODUCTION

Operating criteria at Hydro-Québec[1] ensure that the transmission network can support normal contingencies (Table 1) without interruption of electricity and without any assistance of Special Protection Systems (SPS). This set of contingencies is used to determine the secure power transfer limits. They comply with the Northeast Power Coordinating Council (NPCC) criteria.

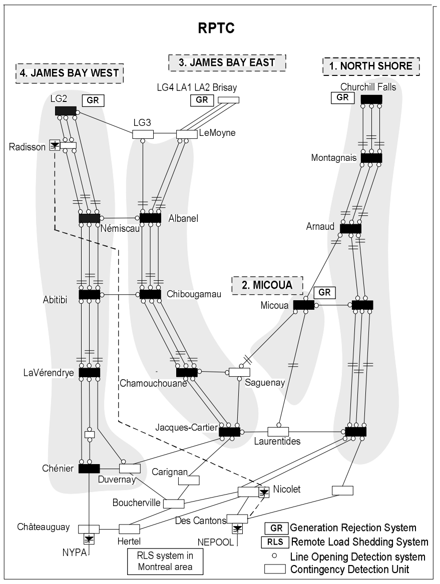
In addition to these requirements, Hydro-Québec has judged that it is also important for the system to be stable after certain extreme contingencies (Table 1)[2]. The system stability is maintained by a scheme of generator tripping and load shedding, called RPTC, such that transfer limits are not affected. Figure 1 and Figure 2 give an overview of the basic structure and the general operation of the RPTC system. Figure 1 illustrates RPTC systems distributed in fifteen 735 kV substations of the Hydro-Québec sys subsystems in a same corridor (or main axis) a into an independent group. There are a total shown in dark shaded areas in Figure 1. Ea RPTC systems performs, associated with indepen Protection Systems, the generation tripping sc particular generation site while the remote lo function is centralized. Figure 2 displays th diagram of RPTC. The RPTC system detects the (LOD) or the bypass of series compensation ban the transmission network. According to the location of the extreme contingencies, the Classification Unit (CCU) sends the information generation tripping, the remote load sheddi tripping of shunt reactors if required.

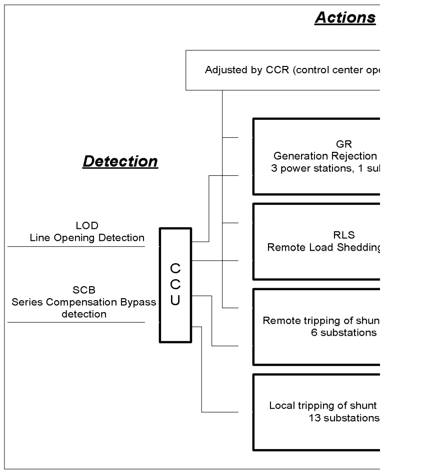
Table 1: Normal contingencies and extreme co

Normal Contingencies	Three phase fault with normal clearing
	Single line to ground fault with delayed clearin
	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 seri 735 kV line
	Loss of all 735 kV lines emanating from a subs
	Loss of all lines in a corridor
	Loss of two parallel 735 kV lines and bypass capacitors on the remaining line in the same co

The tuning of this Special Protection Syst complex: the proper level of generation trippi shedding must be programmed. There configurations to cover and the number of con large. Conventionally, the settings of these R were calculated using deterministic techniques cover the worst-case scenarios. Thus, these sett be optimized with respect to the number of tripped. A probabilistic approach applied to study seems to be particularly interesting.

The study described in this paper covers the F installed at the Churchill Falls hydroelectric (5500 MW) in Labrador. The purpose of thi optimize the settings of generation tripping or event. In order to extract useful information from the database and to cover a sufficiently diverse set of situations, a large amount of data cases were retrieved from the database where real time power system snapshots are stored. Various network states were simulated using the power system analysis software developed by the Hydro-Québec Research Institute (IREQ). These cases represent actual operating states collected over a 5-year period. By using the data mining technique [3]-[6], the most relevant parameters for this automatic device will be identified and effective settings will be determined.





- To suggest an algorithm to modulate the number of generation units to trip.

The methodology used is as follows:

- Extraction of 10 000 network cases spread over a period of 5 years;
- Filtering of those cases to select only topologies with 3 links between Churchill Falls and Montagnais;
- Creation of each case with a random fault duration and a random time of bypassing a capacitor;
- Simulations to find the minimum number of units to be tripped to ensure system stability after this event;
- Optimization of the number of units to be tripped with a decision tree.

3 DATABASE GENERATION

3.1 Methods for data base generation

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 described in section 2. The simulations have a 10 second time frame and are performed on a PC network using a transient stability program (Hydro-Québec ST600 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 envisioned:

- In the first approach, the generation of scenarios is done from a limited number of load 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 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 bee this work, seems more appropriate due to the operating cases used are real.

3.2 Data generation program

As shown in Figure 3, an extraction and a c data is first accomplished. The data conversion allow simulations of power system real snapsh the control center database.

For this task, CILEX [7] software is used. Québec in-house software is widely used snapshots for planning and operation planning en

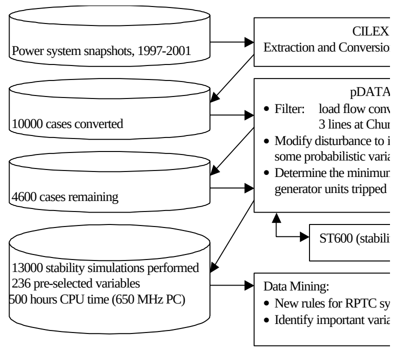


Figure 3: Data Generation Program

A control software (pData) was develop snapshot cases in order to keep just the r (cleaning process). For each filtered case, pD builds the disturbance to be simulated as a fu peculiarities of the studied case.

In this process, pData associates a random va parameters of the disturbance in order to take int effects of these variations on the results. Thes are the fault clearing time corresponding to the li time (breaker operation) and the series compens time.

From transient stability simulations (Hydro-Q program), pData determines, for each case, t number of units to be tripped by an iterative pro respecting security criteria.

Finally, pData extracts results and saves some pre-selected relevant parameters, which will be mining analysis.

3.3 Data coverage period and size of data case generated

From the 10000 extracted and converted snapshots, pData filter rejected more than half of the cases were, either because the resulting configuration did not have 3 lines connecting at Churchill Falls or data errors caused load flow non-convergence. In the remaining 4600 cases, 13000 load flow and transient stability simulations were run to find the minimum number of units to be tripped. The total simulation duration of the 13000 cases is in the range of 500 CPU hours on a 650 MHz PC.

4 DATA MINING

4.1 Data mining techniques

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].

Table 3: 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	100
2	157	0	157	314	2	100
1	58	0	58	58	1	100
0	370	370	0	0	0	0
Total	4560	683	3877	12270	2.69	44

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. Here, this paper is focused on using the decision tree type methodology to optimize RPTC system settings for generator tripping in emergency control at Hydro-Québec.

4.2 Decision (regression) trees

A decision tree (DT) is a map of the reaso This data mining technique is able to produ about a given problem in order to deduce inf new, unobserved cases. The DT has the hierarc a tree structured upside-down and is built on t Learning Set (LS). The LS comprises a num (objects). Each case consists of pre-classifi states (described by a certain number of para candidate attributes), along with its correct (called the goal attribute). The candida characterize the pre-disturbance operating point parameters which can be used to make decisio building process seeks to build a set of rules attributes to the goal attribute, so as to fit the lea well enough without over-fitting this data. The is tested on a different data set (test set) where t of the goal attribute by these rules is compared class (determined by simulation) for each tes classification error rate for the test set measures is successful or not.

There are many reasons to use decision trees their interpretability. A tree structure p information of how an output is arrived at. important asset is the ability of the method to id the candidate attributes the most relevant param problem. A last characteristic of decision computational efficiency. The particular decisi tree induction method used in this paper is details in [8].

5 RESULTS

5.1 Correlation studies

Using the generated database, correlation performed. Figure 4 shows the correlation of transfer margin on the Churchill Falls corridor generator unit tripping as determined by the simulations. The dashed horizontal line in Fig for example, that the currently used-rules req units if the transfer margin on the Churchill Fa less than 400 MW (see Table 2). From t database (4560 cases), there are about 2130 cas 3) in which the current rules require 8 units to b actually, we can see from Figure 4 that if the tr Churchill Falls corridor is less than 3300 vertical line), there is no need to trip generator u we count the number of cases for which the simulations found that it is necessary to trip 8 u only 205 cases. Thus, among the 2130 cases f

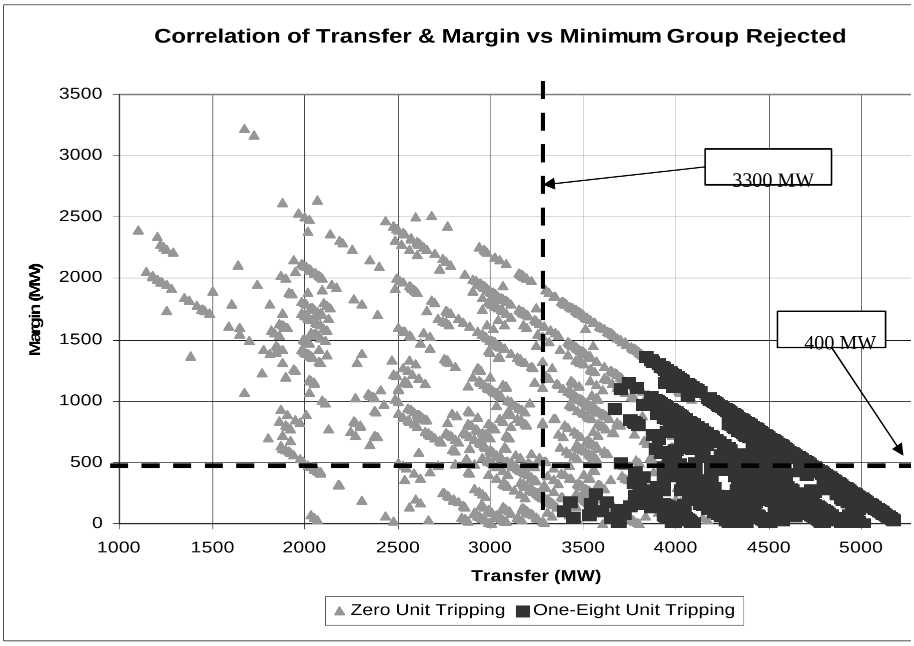


Figure 4: Correlation of Transfer and Margin at Churchill Falls Corridor

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.

Table 3 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 2.69 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.

5.2 *Applying a 5200 MW limit*

Table 2 takes into account the transfer margin (maximum transfer minus measured transfer on the Churchill Falls corridors) to determine the number of units to be tripped. The maximum transfer considers many network configurations (lines, series compensation, synchronous condenser) as well as circuit breaker configurations, but circuit breaker configurations have no impact o limit for the event used in this study. Therefore used in Table 2 are very conservative. To quicl impact of removing this restriction on the numb tripped in excess, we have rebuilt the table by account a margin computed on a maximum trans 5200 MW. The obtained results are surprising over-tripping to an average value of 1.72 per ca from 2.69 with the current rules. However, th cannot be used as it is because it doesn't take certain network configurations having an impact studied here. For these cases, the number of gro to the tripping is not enough to maintain stable o

5.3 *Regression tree*

Constructions of regression trees were carrie generated database using data mining softwa developed at the University of Liege. Among (cases), 2000 objects were selected as a learnin remaining 2560 objects were comprised as a t goal is to predict the minimum number of gene be tripped. Figure 5 Shows a constructed reg The tree is to read top-down: each internal node to a test on one of the candidate attributes and nodes correspond to decisions about the number

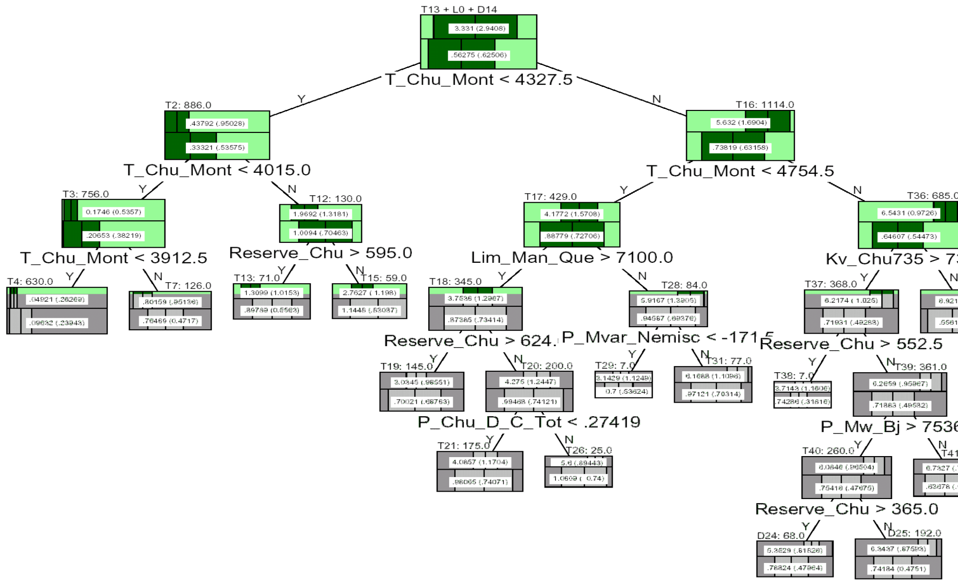


Figure 5: Regression Tree to Predict Number of Generator Unit Tripping

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 630 cases for which the expected number of units to be tripped is 0.04921. A case will be directed to this node if T_CHU_MONT < 3912.5 MW. On the other hand, the right-most terminal node (denoted T42) corresponds to an expected number of units to be tripped of 6.92, for the conditions of T_CHU_MONT > 4754.5 MW and KV_CHU735 > 739.5 KV. The parameter T_CHU_MONT represents the transfer on the Churchill Fall Corridors and KV_CHU735 is the voltage at the Churchill Falls.

Table 4: Average of Over-tripped Units per Case

Methods Used	Average
Current Rules	2.69
Margin Based on 5200 MW limits	1.72
Combination of Transfer and Margin	1.48
Regression Tree with Post-processing	1.01

Notice that among the 236 candidate attributes proposed to the tree building software, only 7 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 regres unbiased estimates of the true values; this mea errors are both negative and positive. In practic preferable to have rules which have less negativ few unit tripping – under-tripping) than positi many – over-tripping) because the cost of instab higher than the cost of unnecessarily tripping one units. Such a bias can be introduced as a post-p the regression tree output, for example by a positive constant to its predictions before round nearest integer.

5.4 Comparison of results

Table 4 lists the average of over-tripped un Although this value may not have a direct physi it is used here as an indication of the improve different methods tested. The first line of the ta the rules actually in use, and designed by deterministic method. The second line c modification of these rules when a constant tran 5200MW is used to compute the margin. The th the performance of another rule designed by hand into account both margin and transfer limit. Fir line gives the results obtained by rules of a post-processing of the regression tree of Figure 5. This 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 non-simulated unit tripping. A positive value means generator units are over-tripped while negative value means under-tripped. 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 5. 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 assess whether the risk of under-tripping is acceptable (with respect to other risks not taken into account in our study), and if not, to eliminate the few remaining cases of under-tripping.

6 CONCLUSIONS

The study described in this paper covers the RPTC automatic device installed at the Churchill Falls hydroelectric power station (5500 MW) in Labrador. The data mining technique was applied to the results of some 13000 network simulations. Various network states were taken from a real-time database and were simulated using the network analysis software developed by Hydro-Québec Research Institute (IREQ). The data cases represent actual operating states collected over a 5-year period. 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 developed at the University of Liege. 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. N
rules can thus be established and will be impleme

Following these very encouraging re
applications of these methods are being consider
Québec in the coming years.

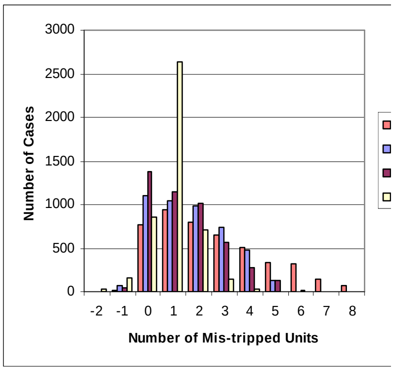


Figure 6: Comparison of Results of Different

7 REFERENCES

[1] J-P. Gingras, R. Mailhot, J. Lavergne, "Bilan de l
réseau d'Hydro-Québec avec compensation série,"
session 2000.

[2] G. Trudel, S. Bernard, G. Scott, "Hydro-Québec's defe
extreme contingencies," IEEE, PE-211-PWRS -0-06-19

[3] L. Wehenkel, M. Pavella, "Advances in Decision T
Power System Security Assessment," IEE 2nd Internati
on Advances in Power System Control, Operation an
December 1993, Hong Kong.

[4] S. Rovnyak, S. Kretsinger, J. Thorp, D. Brown, "Dec
Real-Time Transient Stability Prediction," IEEE Transa
Systems, Vol. 9, No.3, August 1994.

[5] N.D. Hatziargyriou, S.A. Papathanassiou, M.P.
"Decision Trees For Fast Security Assessment of Aut
Systems With a Large Penetration From Renewable," I
on Energy Conversion, Vol. 10, No.2, June 1995

[6] Y. Schlumberger, C. Levrevelec, M. De Pasquale, "
Security Analysis New Approaches Used at EDF,"
Meeting, July 21, 1999, Edmonton, Canada.

[7] L. Riverin, J.A. Huang, A. Valette, F. Lafrance, "Basic
a System of On-line Stability Limit Analysis at
International Conference on Power System Technology
1994, Beijing, China.

[8] L. Wehenkel, Automatic Learning Techniques in
Kluwer Academic, 1998