Importance Sampling Based Decision Trees for Security Assessment and the Corresponding Preventive Control Schemes: the Danish Case Study

Liu, Leo; Rather, Zakir Hussain; Chen, Zhe; Bak, Claus Leth; Thøgersen, Paul

Published in:
Proceedings of the 2013 IEEE Grenoble PowerTech (POWERTECH)

DOI (link to publication from Publisher):
10.1109/PTC.2013.6652397

Publication date:
2013

Document Version
Early version, also known as pre-print

Importance Sampling Based Decision Trees for Security Assessment and the Corresponding Preventive Control Schemes: the Danish Case Study

Chengxi Liu[1], Zakir H. Rather[1, 2], Zhe Chen[1], Claus Leth Bak[1], Paul Thøgersen[2]

[1,]Dept. of Energy Technology Aalborg University
Aalborg Øst, DK-9220, Denmark

[2,]Innovation center kk-electronic a/s
Aalborg Øst, DK-9220, Denmark

***Abstract*— Decision Trees (DT) based security assessment helps Power System Operators (PSO) by providing them with the most significant system attributes and guiding them in implementing the corresponding emergency control actions to prevent system insecurity and blackouts. DT is obtained offline from time-domain simulation and the process of data mining, which is then implemented online as guidelines for preventive control schemes. An algorithm named Classification and Regression Trees (CART) is used to train the DT and key to this approach lies on the accuracy of DT. This paper proposes contingency oriented DT and adopts a methodology of importance sampling to maximize the information contained in the database so as to increase the accuracy of DT. Further, this paper also studies the effectiveness of DT by implementing its corresponding preventive control schemes. These approaches are tested on the detailed model of western Danish power system which is characterized by its large scale wind energy penetration and high proportion of distributed generation (DG). DIgSILENT PowerFactory is adopted for the power system simulation and Salford Predictive Modeler (SPM) is used for data mining.**

***Index Terms*—Classification and regression trees, data mining, decision trees, DIgSILENT PowerFactory, importance sampling.**

I. INTRODUCTION

The continuous demand for operation of modern power system, driven mainly by deregulated electricity market, at its maximum limit with higher efficiency have been forcing power systems to be operated close to their limit. Integration of high level of renewable energy to already burdened power system pushes it to be operated under highly stressed and unpredictable conditions. Modern power systems which are highly vulnerable should not only be protected by local protection schemes but also be controlled and protected by system level schemes, i.e. System Protection Schemes (SPS). The SPS based on security assessment helps in implementing the control and protection schemes when the system is in alert conditions. Wide area measurement systems (WAMS) using phasor measurement units (PMU) synchronized by Global Positioning System (GPS) are being widely applied in SPS. With the help of WAMS, the security assessment can be more accurate and intelligent so as to increase the reliability of the protection schemes.

Classification and Regression Trees (CART), a data mining algorithm was first developed by Breman *et. al* in the 1980s [1]. It has been now widely applied in many fields such as financial analysis, chemical constituent identification, and medical diagnostics. It is also applied in the field of power system to assess the power system security, to predict the power system stability and security and to provide control guidelines for preventing power system instability and severe blackouts. Decision Tree (DT), more accurately a binary tree structured classifier, should be based on a database which is created by a large number of offline time-domain simulations, and data mining algorithm aiming at uncovering the relationship between the power system operating conditions before the disturbances and the probabilities of insecurity after the disturbances. In addition, DT has ability to directly provide guidelines to PSO for implementation of preventive control by evaluating and ranking the thresholds of power flow dynamically and adaptively.

DT was first introduced into the field of power system by Wehenkel *et. al* in 1989 [2]. With the enhancing performance of computers, more and more researchers realized the applications of DT in large interconnected power system. The application of online dynamic security assessment scheme by adaptively updating DT is presented in [3]. In [4], multiple optimal DTs and corrective DTs are proposed to increase its accuracy to access the static voltage stability. In [5], DT based preventive and corrective control schemes are introduced to enhance the dynamic security of power system against contingencies causing transient instability. DT can also be used in controlled islanding [6], load shedding schemes [7], loss of synchronization detection [8] and quick restoration schemes [9].

The main criterion for successful CART is the accuracy of DT. The manners in which different operating conditions (OC) are investigated to create a database for training do affect the accuracy of DT. In this paper, importance sampling is adopted to maximize the information contained in the database and minimize the computation requirements. Contingency

This work was supported in part by Danish Strategic Research Centre under *DSF 09 -067255,* "Development of a Secure, Economic and Environmentally-friendly Modern Power System". (SEEMPS)

Authors can be reached at cli@et.aau.dk or zahus@kk-electronic.dk/zhr@et.aau.dk

dependent DT is also proposed to increase the accuracy, which is especially useful in a secure power system.

The rest of the paper is organized as follows. Section II introduces the contingency dependent DTs used in this approach. Section III presents the preparation of database using importance sampling to optimize the database. Section IV describes in detail the cooperative application of proposed two types of DTs to predict and prevent the potential insecure issues. Section V describes the application of preventive control scheme and shows the performance by a study case of western Danish power system. Conclusions and discussions are provided in Section VI.

II. CONTINGENCY DEPENDENT DECISION TREES

The DTs developed in this paper are based on the algorithms of Classification and Regression Trees (CART) [1], [10]. DT can predict the classification of an object (Secure or Insecure) by dropping a test case from the root node until a terminal node, as shown in Fig.1. Each non-terminal node has one parent node and two descendant nodes. The optimal classification is to select the split of a subset so that the data in each of the descendant subsets are "purer" than the data in its parent subset. The optimal split selection procedure can be thought of as a repeated attempt to minimize the overall tree impurity. Given a node t with estimated class probability $p(j|t)$, $j = 1,\ldots J$ is the number of classes, a measure of node impurity is defined as GINI index in (1).

$$i(t)=\sum_{i\neq j}p(i\,|\,t)p(j\,|\,t)=\left(\sum_{j}p^2(j\,|\,t)\right)-\sum_{j}p^2(j\,|\,t)=1-\sum_{j}p^2(j\,|\,t) \quad (1)$$

The training of DT should be based on a database which is built offline by selecting N_{OC} OCs and screening M_C *"n-1"* contingencies and *"n-k"* typical contingencies from power system operator's historical record or experience on each OC.

The conventional DT is based on a database containing $N_{OC} \times M_C$ cases considering all typical N_{OC} OCs and all potential M_C contingencies. However, only a small proportion of contingencies can result in the instability or insecurity of power system and only those most critical OCs violate the security criteria. So the percentage of insecure cases in database is very low. In (2), $N_{OC}(I)$ is the number of critical OCs leading to insecurity problem and $M_C(I)$ is the number of insecure contingencies.

$$M_C(I)\times N_{OC}(I) \ll M_C \times N_{OC} \quad (2)$$

The low percentage of insecure cases definitely increases the difficulty in finding and classifying the insecure cases which in turn decreases the accuracy of DT. In fact, most of cases in the database are not useful, especially for the assessment of a very secure power system.

In this paper, the contingency dependent DT is proposed to increase the accuracy of DT. The methodology is to first select the most critical OCs and find out the contingencies leading to instabilities by screening all contingencies under aforementioned critical OCs. The next step is to create multiple DTs and each DT is to evaluate only one contingency. A number of $M_C(I)$ DTs are created and for each DT the number of OCs can be increased to fractionize the OCs so as to increase the information contained in the database.

III. IMPORTANCE SAMPLING

The concept of entropy, defined in (3), commonly accepted in the field of information theory, is used here to evaluate the information contained in the database, where p_i is the proportion of training data S classified as class i [11][12].

$$Entropy(S) = -\sum_{i=1}^{c} p_i \log_2 p_i \quad (3)$$

Importance sampling is adopted to select the case containing more information by revising the sampling distribution function *f(x)* more toward the security boundary. As defined in (4),

$$g(x) = f(x) * I(h[g(x)] \le h[f(x)]) \quad (4)$$

where

$$I(h[g(x)] \le h[f(x)]) = \begin{cases} 1 & \text{if } h[g(x)] \le h[f(x)] \\ 0 & \text{if } h[g(x)] > h[f(x)] \end{cases}. \quad (5)$$

g(x) is the importance sampling distribution function and *h(x)* is the *n*-dimensional 2-norm distance from the sampled OC $\boldsymbol{x}$ to the security boundary $\boldsymbol{y}$, defined by (6)

$$h(x) = \left(\sum_{i=1}^{n} |x_i - y_i|^2\right)^{1/2}. \quad (6)$$

After implementation of this sampling approach, a number of OCs are selected and biased to the interested zone adjacent to the security boundary, as shown in Fig. 1. The database contains cases that the OCs are with relatively same granularity close to the security boundary, but the number of cases is significantly reduced.

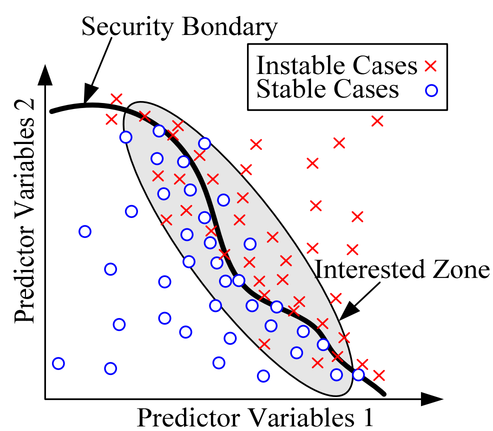


Figure 1. The selection of OCs using importance sampling.

IV. PREVENTIVE CONTROL SCHEME

If a power system at a given OC is detected to be vulnerable to a specific contingency, the operators take preventive actions to ensure its security following a contingency. Generation shift among central power plant (CPP) units is usually adopted to restore the system from insecure state to secure state.

T-D simulations for all the critical contingencies are carried out. Specific security criteria such as transient instability, transient voltage dip, insufficient damping for T-D simulation results are examined to determine the security classification for each case. Two DTs are trained for each contingency in this study, one for observation of insecure condition called as observation DT (ODT) and the other for prevention of insecure conditions called as prevention DT (PDT). ODTs specify the security boundary and the direction of change in generation is provided by PDTs.

A. Observation Decision Tree(ODT)

From database, each case includes a vector of measurements to present an OC before the disturbance which serves as predictors and the results of T-D simulations after the disturbance (Secure or Insecure) which serve as the target for aforementioned predictors. For each OC, synchronized data from PMU measurements, including voltage, current and power are able to provide accurate predictor values.

ODT is used to assess the security state by providing precise boundary of security level and providing the security margin of current OC. Since the DT is trained from all measurable values as predictors, the data mining algorithm is able to provide accurate security boundary in the space of dominant observable values.

B. Prevention Decision Tree(PDT)

In case of PDT, it selects only direct controllable parameters as predictors from measurements, e.g. outputs of generators, power exchange across HVDC links. Though PDT has relatively less accuracy than ODT due to the fewer number of predicting values, it is capable of searching out the most effective direct variables among all the measurable control parameters and providing the potential direction of control to draw the system back to a secure state.

C. Preventive Control Schme based on ODT and PDT

From online measurements, ODT is deployed to assess online DSA to identify the margins of predictors against their thresholds determined from DT training. If there is any violation of online measurements of predictors from thresholds, ODT would provide situational awareness on insecurity if that contingency really happens. On other hand, PDT would provide possible and feasible preventive control schemes to bring the state to a new OC where the contingency under question would not drive system to insecure state if it really happens. Therefore, parallel and cooperative utilization of PDT and ODT in the control schemes provides both situational awareness and preventive control against critical contingencies.

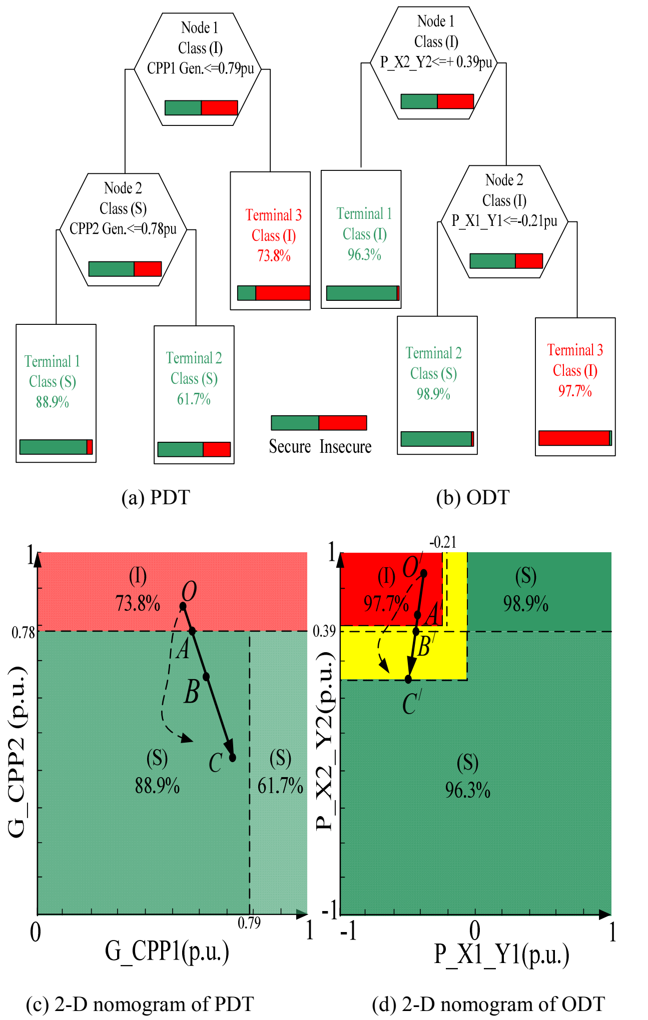


Figure 2. The ODT and PDT and their nomograms

Two 2-dimensional (2-D) nomograms shown in in fig.2 (c) and fig. 2 (d) are adopted to represent the secure operating regions determined by PDT in fig. 2(a) and the ODT in fig. 2(b) from the predictors and their thresholds, respectively. The direction of preventive control can be easily determined if the current OC *O* is identified in the insecure region, i.e. to reduce generation in CPP2 (G_CPP2), shown by the dashed line, in Fig. 2(c). Further, the variation in generators' output will definitely lead to the variation of the power flows in transmission lines, therefore the direction in PDT nomogram has its corresponding projection in ODT nomogram as shown by dashed line in Fig. 2(d).

Hence the process of preventive control can be represented by the trajectory shown in Fig.2(c) by the solid line in PDT nomogram. *O* and *O'* are the original insecure OC. *A* and *A'* represent the OC on the threshold of PDT, while *B* and *B'* represent the OC on the threshold of ODT. The thresholds in ODT are more accurate than the thresholds in PDT, so the region below *B* and *B'* can be deemed as secure region.

D. Prior Probability of DT

A compromise between dependability and security is used to adjust the control guidelines [1]. By adjusting the ratio of prior probabilities of secure cases π_s and insecure cases π_i from 0.5/0.5 to 0.99/0.01, one can find the security boundary in 2-D ODT nomogram with higher security. Conversely, by adjusting the ratio π_s/π_i from 0.5/0.5 to 0.01/0.99, the security boundary with higher dependability can be found in 2-D ODT nomogram. The yellow coloured region between these two boundaries is the fuzzy region as shown Fig. 8(d). The space below *C* and *C'* is the region with high security in which the probability of exception (i.e. insecure case) is below 0.01.

Therefore, by monitoring the OCs in ODT nomogram, the operators are informed of the exact amount of generation shift that should be adopted.

V. CASE STUDY

A. Danish Power System

The scheme proposed in this paper is tested on the western Danish power system detailed model with about 400 buses, 150 lines, 8 CPP units and 150 CHP units. Denmark, which currently produces around 28% of electricity from wind, plans to realize 50% wind share of electricity production by 2025 with wind turbines, especially offshore wind farms. Currently onshore and offshore wind farms are integrated in the system with capacities of 2232MW and 369MW respectively. A new offshore wind farm (Anholt) of 400MW capacity is expected to be commissioned in 2013. Further, about 40% of today's total installed capacity in Danish system is decentralized generation units, such as onshore wind turbines and CHP units. The geographical map of western Danish transmission system is shown in Fig. 3.

Cross border interconnections of western Danish power system to external grids are strong. To the north, the western Danish power system is connected to Norway and Sweden via LCC-HVDC links, with capacities of 1000MW and 750MW respectively. To the south, it is connected to the European Network of Transmission System Operators for Electricity (ENTSO-E) synchronous area via 2 400kV and 2 200kV AC transmission lines to Germany. To the east, the HVDC link---"Great Belt" with capacity of 600MW was commissioned in July 2010 interconnecting western Danish power system and eastern Danish power system. The abundance of hydro power generation in Norwegian and Swedish power systems can cooperate with the wind power generation in Denmark and Germany. The 400kV transmission system acts as the power transmission corridor, which is subjected to significant amount of active power transport.

Under circumstances that such a high penetration of wind generation and distributed generation, the wind forecast and distributed generation prediction are not highly accurate, which lead to considerable mismatch between predicted power flow and actual real time power flow patterns, so the online wide-area preventive control scheme in Danish power system has a great significance.

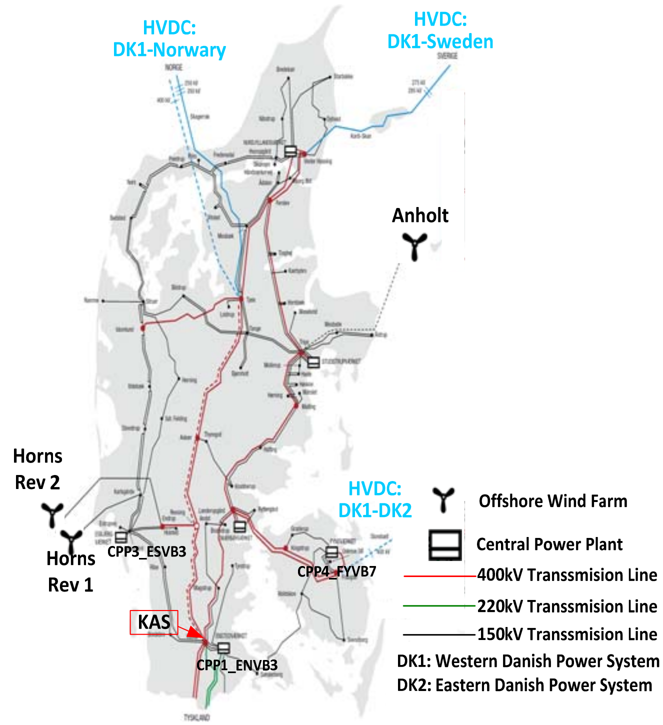


Figure 3. Western Danish power system [13].

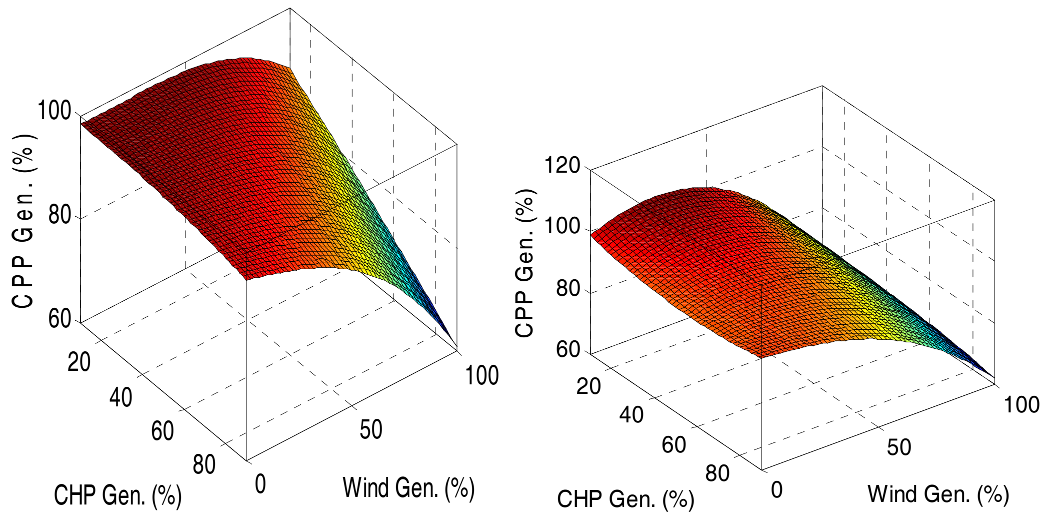
Bisection sampling method is adopted to preliminarily predict the security boundary. Fig. 4(a) and Fig. 4(b) show the predicted boundary of transient stability and voltage security respectively, which is based on the peak load. Fig. 5(a) and Fig. 5(b) show the selection of OCs based on importance sampling. Fig. 6(a) and Fig.6(b) show the PDT based on the 3-phase short circuit in a 400kV transmission line close to substation named Kassø (KAS) as marked in Fig.3, which combines both transient stability issue and short term voltage security issue.

B. Preventive Control Based on DTs

PDT in Fig. 6(a) can be reproduced to 2-D nomograms to depict the space and its regions, as shown in Fig. 6(b). Secure(S) and insecure (I) regions are shown by green and red regions respectively. PDT nomogram shown in Fig.6(b) is capable of informing the most influential CPPs (ENVB3, FYVB7 and ESVB3) that need to be controlled and their control direction, i.e., to reduce generation in ENVB3 and FYVB7 and to increase generation in ESVB3.

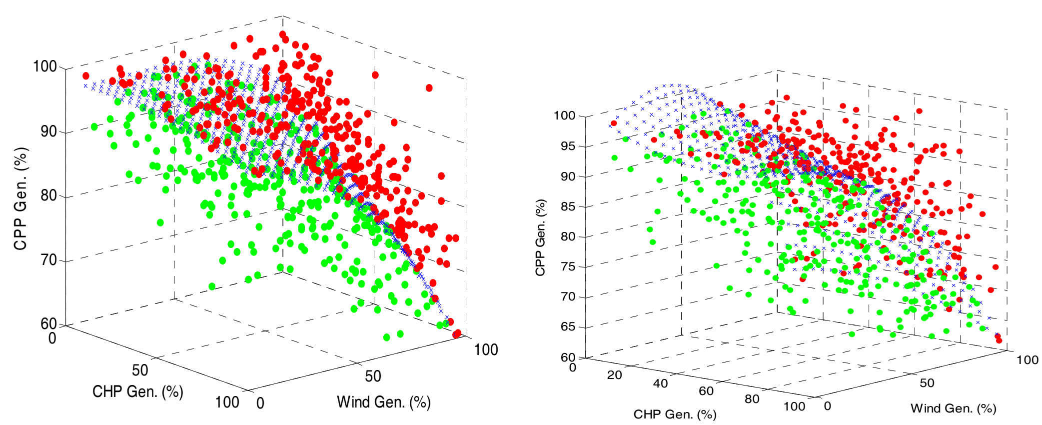
From the annual data of year 2011, an insecure state was detected at 09:00am, February 2nd, 2011, based on CPP generation, total CHP generation and total wind generation.

A surface boundary of constant generation can be found, within which the total generation is maintained constant, as shown in Fig. 7. The trajectories "A", "B" and "C" are some



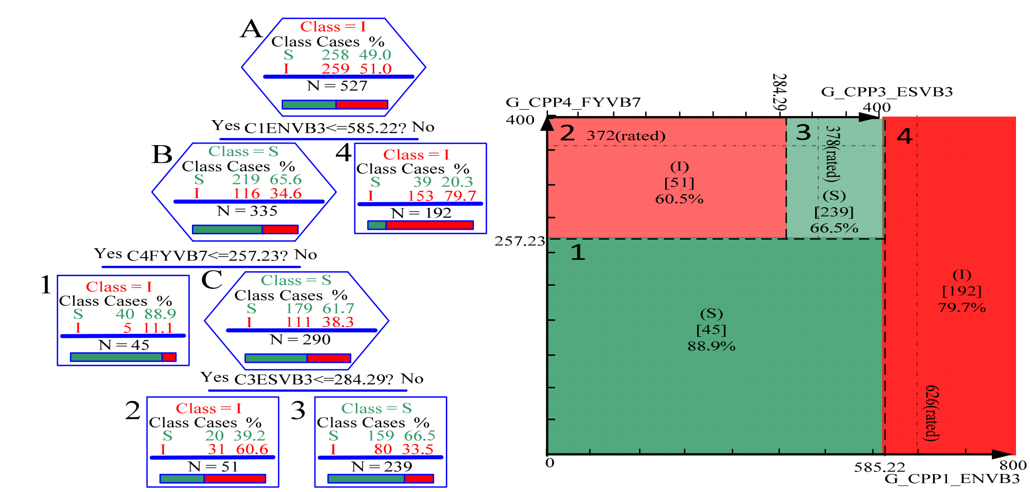
(a) Transient stability boundary (b) Voltage security boundary

Figure 4. The security boundary



(a) OC slection on transient stability boundary (b) OC slection on voltage stability boundary

Figure 5. Impotance sampling on security boundary.



(a) The PDT for contingency near KAS (b) The PDT for contingency near KAS

Figure 6. The decision trees created by CART.

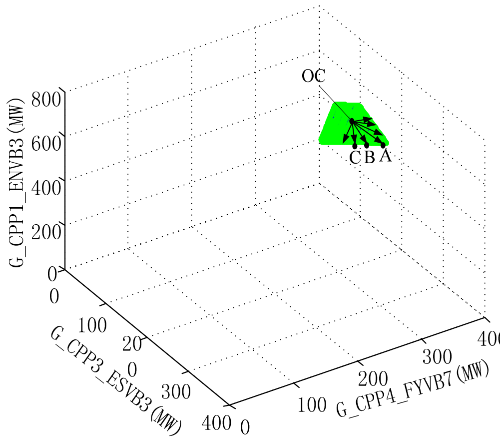


Figure 7. The allowed control surface.

of the feasible control directions on the surface which would lead the system to the secure region.

It can be observed that the trajectories "A", "B" and "C" are the control directions which are able to reliably draw the system to the secure region, as shown in Fig.8.

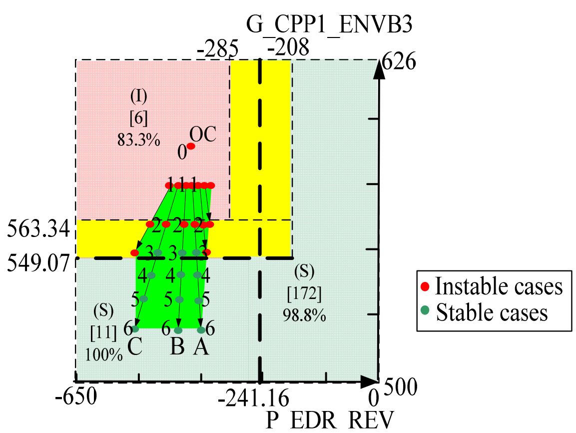


Figure 8. The T-D simulation results on ODT nomogram

VI. CONCLUSION

In this paper the importance of high accuracy and of DTs was discussed to highlight the need of efficient and high precision DT based approach. The DT based preventive control scheme was proposed and demonstrated in the operational model of western Danish power system in DIgSILENT PowerFactory. Further the database can be periodically updated considering new network topology, component availability and system conditions.

VII. ACKNOWLEDGEMENT

Authors would like to thank Mr. Per Lund, Chief Engineer Energinet.dk (Danish TSO) for providing data of western Danish Power System and valuable discussions from time to time during this research work.

REFERENCES

[1] L. Breiman, J. Friedman, R. A. Olshen and C. J. Stone, Classification and Regression Trees. Belmont, CA, Wadsworth, 1984.

[2] L. Wehenkel T. Van Cutsem and M. Ribbens-Pavella "An artificial intelligence framework for on-line transient stability assessment of power systems," *IEEE Trans. Power Syst.*, vol. 4, no. 2. pp. 789-800, May, 1989.

[3] K. Sun, S. Likhate, V. Vittal *et al.* "An online dynamic security assessment scheme using phasor measurement and decision trees," *IEEE Trans. Power Syst.*, vol. 22, no. 4. pp. 1935-1943, Nov, 2007.

[4] R. Diao, K. Sun, V. Vittal *et al*. "Decision tree-based online voltage security assessment using PMU measurements," *IEEE Trans. Power Syst.*, vol. 24, no. 2. pp. 832-839, May, 2009.

[5] I. Genc, R. Diao, V. Vittal *et al*. "Decision tree-based preventive and corrective control applications for dynamic security enhancement in power systems," *IEEE Trans. Power Syst.*, vol. 25, no. 3. pp. 1611-1619, Aug. , 2010.

[6] N. Senroy, G. Heydt, V. Vittal *et al*. "Decision tree assisted controlled islanding," *IEEE Trans. Power Syst.*, vol. 21, no. 4. pp. 1790-1797, Nov. , 2006.

[7] E. M. Voumvoulakis, A. E. Gavoyiannis, and N. D. Hatziargyriou, "Decision trees for dynamic security assessment and load shedding scheme," in *Proc. Power Eng. Soc. General Meeting*, Montreal, QC, Canada, Jun. 2006.

[8] K. Mei and S. M. Rovnyak, "Response-based decision trees to trigger one-shot stabilizing control," *IEEE Trans. Power Syst.,* vol. 19, no. 1, pp. 531–537, Feb. 2004.

[9] Y. Hou, C. C. Liu and K. Sun *et al*. "Computation of milestones for decision support during system restoration," *IEEE Trans. Power Syst.*, vol. 26, no. 3. pp. 1399-1409, Aug. , 2011.

[10] CART Tree-Structured Non-Parametric Data Analysis. [online]. Available: http://www.salfordsystem.com/cart.php

[11] V. Krishnan, J. D. McCalley, S. Henry *et al*. "Efficient database generation for decision tree based power system security assessment," *IEEE Trans. Power Syst.*, vol. 26 no. 4. pp. 2319-2327, Nov. , 2011.

[12] E. A. Unger, L. Harn, and V. Kumar, "Entropy as a measure of database information," in *Proc. 6th Annu. Computer Security Applications Conf.,*1990, pp. 80–87.

[13] System plan of Danish power system 2010. [online]. Available: http://www.energinet.dk