Probabilistic design of power-system special stability controls[1]

L. Wehenkel[a],*, C. Lebrevelec[b], M. Trotignon[b], J. Batut[b]

[a] *Department of Electrical Engineering, University of Liège Sart Tilman B28, B-4000 Liège, Belgium*
[b] *Electricité de France, R&D Division 1, avenue du Général de Gaulle, 92141 Clamart Cedex, France*

Abstract

A probabilistic approach to the design of power-system special stability controls is presented here. Using Monte-Carlo simulations, it takes into account all the potential causes of blackouts, slow and fast dynamics, and modeling uncertainties. A large number of scenarios are simulated in parallel by time-domain numerical integration, and the relevant parameters of the resulting system trajectories are stored in a database. Data-mining tools are used to identify the most important system weaknesses and possible improvements. The approach is tested on a large-scale study on the South–Eastern part of the extra-high-voltage system of Electricité de France. © 1999 Elsevier Science Ltd. All rights reserved.

1. Introduction

The security or stability of a power system refers to its ability to perform in a satisfactory way under the occurrence of various kinds of adverse events, such as external disturbances or internal failures of its components. While 'security' refers to the robustness of the system in a given state, 'reliability' refers to the expected value of security over a longer period, say several years. Under normal conditions, modern interconnected power systems are rather secure, and they are able to absorb in a satisfactory way the most likely adverse events, i.e. with no or only very localized consequences. However, it is always possible that an (unlikely) combination of events would lead the system to an overall, system-wide failure. While these types of situations are of very low probability, they may lead to tremendous consequences. On the other hand, they are very difficult to anticipate, in particular due to the complexity of power systems and the wide range of dynamic phenomena which may be involved. For example, the following five types of phenomena may be involved: (i) cascading tripping of overloaded lines, (ii) loss of voltage stability, (iii) loss of transient (angle) stability of synchronous generators, (iv) frequency collapse, and (v) undamped oscillations. It is also true that in system-wide disturbances several of these phenomena are generally involved at the same time, and interact in a way that depends on the structure of the particular power system under consideration.

Economic and environmental pressure, together with competition, make modern power systems operate closer to their limits and at the same time be less predictable than in the past. Recent experience in North America has shown that these circumstances increase the probability of blackouts; therefore the need for systematic methods to study these phenomena and mitigate their consequences is strongly felt.

In order to increase the reliability and security of a bulk power system, there are essentially three possible options:

(i) expand the power system, either by increasing its transmission capabilities (e.g., by building new lines) or by reducing the need for long-distance transmission (e.g., by building power plants close to the load centers),

* Corresponding author. Tel.: + 32 4 366 2688; fax: + 32 4 366 29 84; e-mail: lwh@montefiore.ulg.ac.be.

(ii) operate the system in such a way that its security margins are large enough,

(iii) improve the real-time emergency (or stability) control strategies that act automatically on the system to minimize the consequences of important disturbances (i.e., make the system more intelligent).

However, since it is becoming increasingly difficult to expand transmission systems within reasonable delays (e.g., due to ecological constraints), and since the operation with high security margins conflicts with economic efficiency, the present trend in power systems is to rely more and more on special emergency control schemes.

The design of such special emergency control schemes implies the identification of the main failure modes of the system, the determination of appropriate mitigating control actions, and the design of automatic triggering devices. These latter must be able to detect in real time when the system is in the process of losing its stability, and send appropriate control signals quickly enough to avoid dramatic consequences. In particular, these systems must be robust with respect to changing operating conditions and other uncertainties related to measurement and modeling errors. To enhance system performance (i.e., to reduce failure risks), existing emergency control systems could be improved or replaced by new, more effective ones. It is also believed that the coordination of various such devices, acting in different time frames or in different geographical areas of a power system, should be improved.

Given the fact that power-system failures are (luckily) very rare events, actual experience is slim, and the engineers in charge of the design of emergency control schemes rely mainly on simulation. Their approach thus classically consists of building up simulation scenarios in a manual trial-and-error fashion, and analyzing the simulation results by hand. This is an intricate and time-consuming activity, which strongly relies on human expertise. Generally, problems are decomposed into elementary subproblems (e.g., voltage vs. angle stability, or region-wise) which are treated separately, often by different engineers. However, due to the faster changes occurring in today's power systems, the need for more regular studies is increasing, while human expertise may quickly become obsolete and even misleading. Moreover, the fact that systems tend to operate closer to their limits may lead to more-complex interactions among the different phenomena, making it even more difficult to find globally effective solutions.

But, above all, the principal shortcoming of this *deterministic* approach is its inability to take into account the stochastic nature of the causes of power system failures, and the unavoidable modeling uncertainties. Indeed, not all failure modes are equally likely to happen, their consequences are highly variable in practice, and all simulation models are wrong to a certain extent.

In order to circumvent these difficulties, while taking advantage of existing computing power and modern simulation tools, this paper proposes a smarter *probabilistic* approach, to help engineers carry out such studies in a more systematic way. It is based on Monte-Carlo simulations screening a representative sample of security scenarios, and exploits automatic learning tools to extract useful information from the simulation results (Wehenkel, 1998). In this methodology, uneven prior probabilities of possible causes that could lead to extreme conditions are taken into account, the economic impact of their consequences is also evaluated, and various uncertainties about less well-known system components are integrated in the decision-making process (Cigré, 1997). Note that this approach can identify the most likely failure modes of a system, while in the deterministic approach this information must be known prior to the study. Although the probabilistic approach needs additional information and modeling work in order to define the random sampling probabilities, it is able to provide a more objective and a more anticipatory view of the possible failure modes of a system.

The approach consists of three steps:

(i) modeling probabilistically the causes potentially leading to extreme conditions (weakened operating conditions, abnormal operation of protections, multiple disturbances) as well as uncertainties (load behavior, external systems …),

(ii) using parallel Monte-Carlo simulations to sample scenarios according to this information (random combinations of operating conditions, dynamic modeling hypotheses (e.g., relay settings, malfunctions, external systems, load …), and disturbances),

(iii) building up a database of simulation results, collecting key variables and their temporal behavior, and using data-mining techniques to extract from this synthetic information about the main breakdown modes and possible ways to improve emergency control schemes.

While the approach and the developed tools are general, this paper focuses on their application to the study of the French power system under extremely disturbed conditions, with the aim of evaluating the risks of various types of failure modes (e.g., voltage collapse vs. loss of synchronism) so as to identify the weak points and improve protection schemes. Note also that, in the study, a detailed global dynamic model of the French power system was first set up, comprising more than 11 000 state variables. Furthermore, given the size (several GBytes) of a simulation results database, and the complexity of its information (temporal curves describing the dynamics of a large-scale non-linear system), a prototype data-mining tool was developed for the research, incorporating a number of graphical and automatic learning tools (clustering algorithms, decision trees, nearest neighbor …).

The paper is organized as follows: Section 2 describes the probabilistic approach, and Section 3 the software tools developed. Section 4 discusses the study on the South-Eastern part of the French extra-high-voltage (EHV) system, focusing on modeling and database generation aspects, and Section 5 provides the data-mining results obtained, and discusses further work in progress. Section 6 draws some first conclusions, and gives some guidelines for future work. Throughout the paper, the applicability of the approach to other control-system design problems will be pointed out.

2. Overview of the probabilistic approach

2.1. Overall framework

The overall framework is depicted in Fig. 1. The approach proceeds by iterating through the following elementary steps:

Study specification: setting up a detailed probabilistic model of the possible *causes* of insecurity: multiple disturbances, bad coordination and/or mal-operation of protective devices, over-optimistic preventive security strategies due to uncertainties in modeling parameters.

Database generation: sampling representative combinations of these causes, and carrying out extensive simulations to determine the behavior of the power system under these assumptions.

Data mining: analyzing the database of dynamic simulation results to identify *a posteriori* the main weaknesses of the system, and possible ways to improve its reliability.

Evaluation: evaluating the most effective countermeasures, e.g. in the form of new or modified special stability-control systems, and validating them through a cost/benefit analysis on the scenarios stored in the database.

2.2. Study specification

The first step of the proposed methodology consists of specifying the range of scenarios that the particular study will address. While the approach aims at enabling the engineers to carry out broader studies, more systematically covering all kinds of situations and dynamic phenomena, the scope of a particular study should be clearly defined *a priori* in terms of security scenarios (Jacquemart et al., 1996).

A scenario has three components: (i) its starting *operating point* (available lines and generators, load level and distribution, generation dispatch, voltage/var schedule …), (ii) its *dynamic modeling hypothesis* (e.g., control-loop gains, protection settings, malfunctions, load behavior, measurement noise …), and (iii) a sequence of *disturbances* (e.g., short-circuits, outages, load trends …).

Starting with the existing expertise and problem statement, the random sampling specifications are set up for each one of these components, generally through a sequence of discussions among experts in different fields, such as power-system dynamics, protection and economic questions. Some base cases are selected, and a catalog of variable parameters that are important for the study under consideration is set up. Then, for each parameter, a probability distribution is chosen, in order to screen its possible values; constraints among these parameters may also be defined in order to filter out unrealistic scenarios. For example, among the many parameters that were handled in this way in the study described in Section 4, one could mention line overload protections (the thresholds and triggering delays of which were randomized around their normal values) and load models (whose sensitivities to voltage were randomized).

The other part of the database specification concerns the choice of the *attributes* that will be extracted from the simulations and stored in the database: these are scalar and temporal parameters describing the dynamic behavior of the system, and will be used as inputs (e.g., candidate measurements) or outputs (e.g., evaluation of consequences) to the data-mining tools, in order to extract synthetic information from the database. Examples of extracted attributes given in Section 4.3 are EHV voltages and generator rotor speeds.

Thus, the first time a new problem is considered, the initial investment in the database specification is quite important. However, when similar problems are subsequently considered, their adaptation is much more straightforward.

2.3. Database generation

The database generation process is carried out in a fully automatic way by a dedicated software tool (see Section 3.1). It consists of two successive steps: (i) random
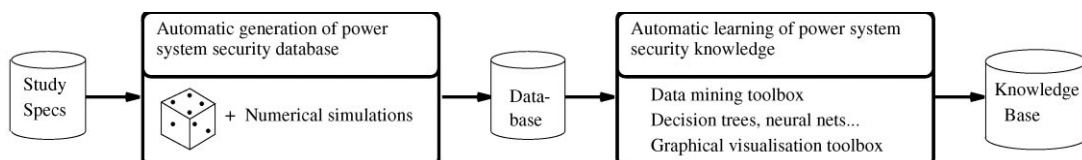

Fig. 1. Overall probabilistic framework.

sampling of individual scenario specifications, and (ii) numerical simulation of the sampled scenarios.

2.3.1. Random sampling of scenarios

This is essentially a sequential process, building a so-called *a priori* data base of individual scenario specifications. Only trivial computations are carried out here, and even for very complex problems, the computing time required by the random sampling process is negligible with respect to the time needed for the simulation of the scenarios.

2.3.2. Scenario simulation

While the random sampling builds up the *a priori* database sequentially, the scenario simulation engine extracts successive specifications and launches their simulations in parallel, using a cluster of available workstations as simulation servers. A master process collects the results of the servers, and gradually builds up the so-called *a posteriori* database with their results.

Notice that, depending on the size of the power system and the degree of refinement of the dynamic model being used, the computational burden of these simulations may be rather high, which justifies their parallelization. Moreover, the raw outputs of the simulation engine have to be post-processed in order to extract useful information in a format which is compatible with the data-mining tool, and to save disk space if necessary by using appropriate compression techniques. For example, in the study described below, the total amount of raw data generated was about 250 GB; thanks to extraction and compression, the *a posteriori* database was finally only about 800 MB large.

Notice also that some of the scenarios generated by random sampling may not result in successful simulation, for example due to physical non-feasibility or numerical convergence problems. Thus, an *a posteriori* database generally corresponds to a proper subset of the corresponding *a priori* database.

Once a sufficient number of scenarios has been simulated successfully, a database release is created which can be exploited by the data-mining module. Possibly, after having carried out some analyses on this database, the database can be completed by specifying new random sampling conditions and carrying out further simulations, thus yielding successive database releases which will be used successively in a trial and error fashion during the study and design process.

2.4. Data mining

The data-mining process is itself composed of successive steps, aimed at extracting more and more refined information from the simulation database. During this process a variety of visualization and statistical data exploration and modeling techniques are used.

Step 'zero' in the data mining process consists of validating the scenarios that are contained in the database release. In particular, since some of specified scenarios may have been filtered out (e.g., because they could not be simulated properly due to physical non-feasibility or numerical convergence problems) it is necessary to check the representativity of the database. If validation is successful, data mining can be started in effect; otherwise database specifications may be modified and the database generation process is restarted in order to improve representativity.

The first step would be to rank the scenarios according to their severity. For example in the study carried out below, various measures of scenario severity were used, quantifying various undesirable consequences (number of transmission elements lost, amounts of lost generation, amount of lost load, variation in power flows in important tie lines). Some interesting (e.g., very severe) scenarios can be identified and analyzed by 'hand' using the graphical visualization tools incorporated in the data-mining software. For example, the temporal evolution of voltages, power flows, and other interesting variables may be displayed, and it is possible to step through the various discrete events corresponding to the action of protections by using a one-line diagram automatic playback.

The next objective of data mining will be to determine the main weak points of the power system being studied (i.e., the relative frequency of different types of behaviors: stable, voltage collapse, cascading line tripping, loss of synchronism, frequency collapse, oscillations, dynamic instabilities...). This can be achieved using automatic unsupervised learning techniques, as illustrated in Section 5.

Finally, further steps will consist of assessing how the different variables used to characterize the scenarios are interrelated, and will automatically identify those that are useful to detect various undesirable phenomena. In particular, supervised learning and correlation analysis may be used to screen automatically a large number of variables and scenarios relevant for each task. Notice that at this step the engineer can try out various possibilities, being limited only by his imagination and the information contained in the database. In order to find efficient ways to predict system failures, he may try out different ways of decomposing the overall problem into subproblems, and test different sets of measurements and detection logics. Such detection rules may be defined either manually or automatically using supervised learning methods.

Eventually, the data-mining process will lead to some changes in some existing special stability controls, or to the design of new ones. If this is the case, the suggested improvements may be evaluated by incorporating them in the simulation model and generating a new "variant" of the database to simulate their effect. The comparison of this latter with the original database will allow the

engineer to assess whether the system works, and if not, to suggest some improvements until satisfactory results are obtained.

Thus the overall design process is essentially similar to the classical trial-and-error process used presently. However, the main difference is that at each step actions are evaluated in parallel on the same statistically representative sample of scenarios. This makes comparisons among different possibilities much easier and more meaningful, while a lot of flexibility is gained. Moreover, the engineer spends his time carrying out interesting investigations, rather than merely running simulations and managing their input and output files.

Also, the fact that all analyses are carried out on a representative sample makes the studies more transparent, makes it easier to exchange information among various engineers, and allows one to refresh the results easily if some parameter is changed (e.g., if a new FACTS device comes into operation, or if the dynamic models are modified).

3. Software tools

In order to carry out this research, a set of software tools were specifically developed for database generation and data mining. These tools may be considered as advanced research prototypes of yet-to-be-developed industrial-grade software tools.

3.1. Database generation tool

The software tool developed for the automatic database generation is depicted in Fig. 2. It uses the *Eurostag* simulation engine (see Meyer and Stubbe, 1992) and is composed of two main modules (Lebrevelec et al., 1997):

(i) Random sampling.
   Input: specification of the study scope, in terms of probability distributions, Eurostag data files, number of scenarios to be generated, random number seeds.
   Output: database describing the sampled scenarios.

(ii) Simulation scheduler (master/slave, see Fig. 2).
   (a) Eurostag file builder (master).
       Input: scenario description, Eurostag data files.
       Output: a set of modified Eurostag data files.
   (b) Task dispatch (master).
   (c) Task simulation and extraction of relevant information (slaves).
       Input: Eurostag data files; specification of attributes to extract.
       Output: Extracted attribute files.
   (d) Database builder (master).
       Input: result files for each scenario.
       Output: result database.

The master and slaves are Unix workstations, exchanging data through files. As soon as a slave becomes idle, it receives from the master a scenario to simulate. In the study described in Section 4, these simulations typically take several hours of CPU time, and generate several hundreds of Mbytes of raw output. From these latter the slave extracts only the relevant information, i.e. a subset of interesting attributes, sub-sampled at an appropriate (variable) time step. Note that this time step is in practice about 20 times larger than the (variable) time-step used by the numerical integration techniques for their simulations (see below, Section 3); this allows for significant space savings in the results database. The latter is organized hierarchically, one directory per scenario containing about 30 compressed flat ASCII files (UNIX compression facility).

Notice that the software could be easily adapted in order to replace Eurostag by another simulation engine.

3.2. Data-mining tool

A dedicated data-mining tool was developed from an earlier advanced research prototype used in other investigations (Wehenkel et al., 1994a, b, 1995) . The features added concern the representation and manipulation of the (large amounts of) temporal data encountered in this particular problem (curves reflecting the variation over
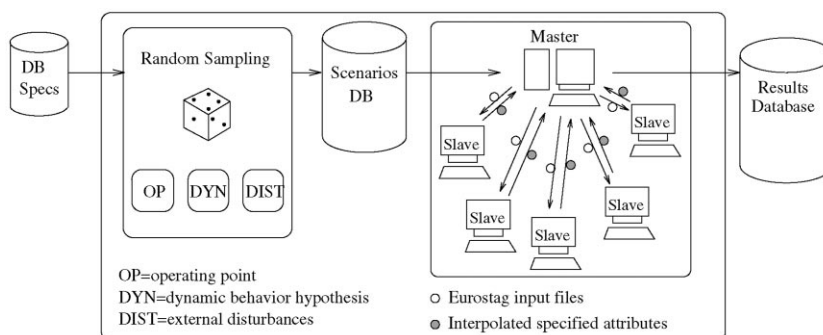

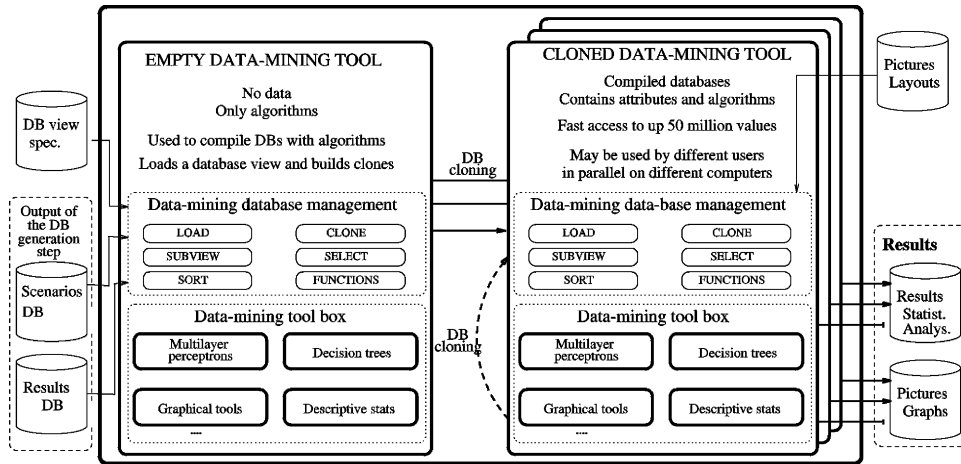Fig. 2. Database generation overview.

Fig. 3. Data-mining tool.

time of different variables describing the scenarios, and sequences of events corresponding to the action of various protections).

The overall architecture of this tool is depicted in Fig. 3. It is composed of two main components.

3.2.1. Data-mining database management module

This accesses data from the generated database flat files, while allowing one to select a subset of the scenarios and a subset of attributes, yielding a particular *view* of the databases. These data are then compiled into an internal representation, and an executable image is cloned, containing the viewed data and the data-mining algorithms. On a workstation with 256MB of main memory it is thus possible to access simultaneously up to 50 million attribute values: about 1500 scenarios, described by 200 temporal attributes, with an average number of 150 time steps per scenario.

The database management module allows the user to define additional attributes as functions of the basic ones extracted from the simulations, and compiles them into efficient code. It provides also various subset and variable sorting/selection facilities, allowing one to focus on a particular *sub-view* of the loaded database.

3.2.2. Data-mining tool box

This module contains both low-level and high-level algorithms, in order to allow the user to extract knowledge from a database. The former are basic statistical summarizations in text and/or graphical forms (means, standard deviations, correlations, bar diagrams, curves, scatter plots ...). The latter are automatic learning algorithms, deemed useful for security assessment applications, e.g. unsupervised learning (clustering of similar types of scenarios or attributes), top–down induction of decision trees, and smooth non-linear regression (multilayer perceptrons).

The data-mining module also contains a graphical scenario inspector *VisioNet*, which enables the user to step through time while seeing the effect of protections and changes in electrical parameters on a one-line diagram representing the power system (Geurts and Wehenkel, 1998a).

4. Case study: south-eastern EDF subsystem

4.1. Brief description of existing problems

In July 1995 a study was started to apply the methodology described in Section 2 to a case study on the Electricité de France (EDF) system under highly disturbed conditions, so as to yield a representative test-bench, and to assess the behavior of this system in extreme conditions. The primary objective was to study interactions among slow and fast dynamics and their corresponding protection systems. Consequently, EDF experts defined a study region in the Provence/Alpes/Riviera subsystem, which was already known to present rather diverse failure modes: cascading line tripping, plant and area mode loss of synchronism, and voltage collapse. This South-Eastern part of the EDF system (see Fig. 4) generally exports large amounts of power to the rest of France and to foreign countries (Italy, Switzerland and, indirectly, Spain). In the very extreme of South-Eastern part it is weakly meshed and deficient in generation, and thus liable to voltage-collapse phenomena. It is also interesting to notice that this subsystem (1998 projection) is equipped with various automatic emergency control systems, in order to mitigate various types of failure modes: the coordinated defense plan (regional islanding scheme in case of area mode loss of synchronism); generation shedding to alleviate specific line overloads; automatic tap-changer blocking scheme to mitigate voltage collapse, etc.
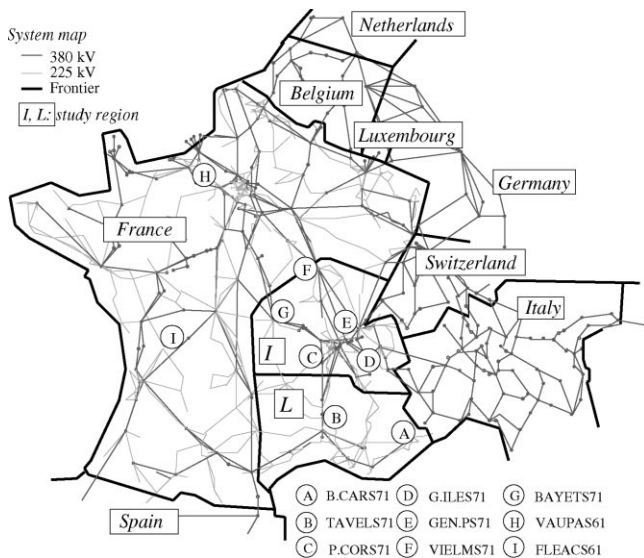
Fig. 4. One-line diagram (study region and surroundings).

4.2. Dynamic system model used

To enable the simulation of both fast and slow phenomena, while taking into account the operation of the relevant protections and special stability-control systems, a rather detailed dynamic model was first set up.

This model, comprising all in all more than 11000 state variables, is summarized in Table 1.

4.3. Database generation

During the first year of the research project the database specifications were set up and an appropriate dynamic model was developed, together with the software for generating and analyzing the databases. A first preliminary database of a few hundred scenarios was generated in early 1996, and, in August 1996, the generation of the final database was started, using the tool described in Section 2. By end of October 1996, 1500 scenarios had been simulated.

4.3.1. Operating points
Three different operating points were used, corresponding to three different situations of import/export of power from the study region. They were chosen manually, in order to provide interesting scenarios to validate the approach, and are representative of realistic N-1 secure operating conditions of the 1998 system projection.

4.3.2. Dynamic modeling hypotheses
In order to take account of the effect of uncertainty and/or errors in protection settings (delays, thresholds...),

Table 1
Dynamic system model. (Lebrevelec et al., 1997)

System	Nb. buses[a]		Nb. lines		Nb. units[b]		Installed pwr	
EDF	380 (kV)	225 (kV)	380 (kV)	225 (kV)	Therm.	Hydr.	Gen. (MW)	Load (MW)
Study regn.	51	217	74	252	22	48	26 000	20 400
Rest	306	646	259	810	61	12	58 800	55 200
Equiv.								
Belgium	22	1	32	1	2		11 400	12 200
Germany	13	3	32		10		64 500	65 800
Switzerld.	31	16	35		4		13 400	15 600
Italy	81	22	124	32	20		15 700	18 200
Spain/Port.	11	2	39	3	9		33 400	34 000
Total	515	907	595	1098	128	60	223 200	221 400

Special control systems
Secondary Voltage Control (EDF system)
UnderFrequency load shedding (EDF and foreign systems)
Generation shedding line overload mitigation (study region)
First step of EDF's coordinated defense plan [Trotignon et al., 1992]
Local line tripping scheme on loss of synchronism (study regn.)
OLTCs blocking scheme in case of voltage collapse (study regn.)
Busbar low-voltage protection (study regn.)

Other protections
Line overload (all 380 kV/225 kV lines of the EDF system)
Under/over voltage/frequency (all units of the EDF system)

[a] 193 loads in the study region are connected to the EHV system through a cascade of two OLTCs.
[b] Generators modeled in detail, with AVR, governors and PSS.

these were systematically randomized in the database generation. Similarly, the load model ($P = P_0(V/V_0)^\alpha$, $Q = Q_0(V/V_0)^\beta$) was also randomized in order to represent the uncertainty and variability of load behavior. Moreover, in order to represent what may happen in real life, in each scenario there is a random selection of some protections that are supposed to mal-operate: untimely generator tripping for over/under frequency or voltage protections, untimely line tripping for overload protections, or partial non-operation of underfrequency load-shedding protections.

4.3.3. External disturbances

These are composed of two consecutive contingencies in the study region, chosen randomly (probabilities more or less reflecting their relative likelihood in real life) among the following ones:

Generator loss: loss of one unit (thermal or nuclear); loss of some units of the same plant (thermal and hydro units); loss of a plant.

Faults on lines: temporary fault on a line, permanent fault on a line; permanent fault on parallel circuits; two temporary or permanent faults separated by about 100ms on geographically close lines (lightning storm).

Faults in substations: (bus bar fault) fault inside a substation leading to the loss of part of the substation; fault outside (but close to) a substation combined with protection failure and leading to partial loss of the substation; loss of the whole substation after a major fault inside.

The two external disturbances are applied in sequence within a rather short time-slot (less than 10 min), and it is supposed that there is no operator action in between. Then, the scenarios are simulated during the 40 mins after the second external disturbance, in order to evaluate the consequences.

The scenarios were simulated by Eurostag (Meyer and Stubbe, 1992); this simulation program is used in dynamic security assessment studies at EDF and, with its variable integration time step, is able to simulate slow dynamic phenomena (e.g., voltage collapse) as well as faster ones (e.g., loss of synchronism). The simulations were carried out in parallel on a cluster of 12 workstations, available at night and during the week-ends.

4.3.4. Attributes

About 800 temporal attributes are used to describe the scenarios in the results database. They are:

- voltage (magnitude and phase angle) of defined buses (130 attributes);
- rotor angle, velocity, and acceleration as well as excitation current and mechanical power of defined units (315 attributes);

- total active and reactive load and mean voltage in defined load areas (54 attributes);
- mean transformation ratio of 'on-load tap changers' in defined areas (13 attributes);
- equivalent voltage angle and frequency for the regions of the defense plan (6 attributes);
- reactive generation of the units participating in secondary voltage control (19 attributes);
- active and reactive power flows of all 380 kV lines of the EDF system, and some important 225 kV lines in the study region (234 attributes);
- a listing of the discrete events happening in the system during the simulation.

Some of these attributes are to be used in order to define the severity of the scenarios, i.e. to measure the *consequences* in terms of loss of load and generation. The others are to be used as input parameters to criteria for characterizing the dynamic behavior of the scenarios and to predict their severity as accurately and as early as possible. Applying data-mining algorithms to the database will allow one to find those that are most strongly correlated to scenario severity. These latter may then be used to predict the future behavior for real-time system monitoring, together with appropriate decision rules extracted from the database.

4.4. General overview of the database

Table 2 provides a first glance at the diversity of the information in the database.

First of all, the *database physics data*:

- the *CPU time* of a scenario (about 11 h) is quite important, which is due to the complexity of the dynamic model used and the long simulation time. In future database generations, this time could be reduced significantly by simplifying the simulation model outside the study region, and by taking advantage of faster computers (a factor 3 within 2 yr);
- the number of *time steps after interpolation* is rather small if compared to the number of steps used by the simulator; consequently, the **size** of a scenario stored in the results database is small enough (540 K bytes) to allow the storage (on available cheap mass-storage devices) of databases comprising up to to 5000 scenarios.

Thus, a results database can be easily built and stored. Concerning the **contents** of the database, the following preliminary comments are appropriate:

- The *number of 380/225 kV lines tripped* shows the amplitude of the phenomena in the database, going from a simple cascading line tripping, to the action of the coordinated defense plan, opening all the frontier lines between the study region and the rest of the EDF and the foreign systems.

Table 2
Result DB statistics (Lebrevelec et al., 1997)

Object	Min	Max	Mean	σ
CPU simulation time (s)	0	99 000	38 000	22 000
Time steps before interpolation	89	46 800	3800	3071
Time steps after interpolation	4	1728	145	137.2
Size (Mb) (raw)	4	2140	174	140
Size (Kb) (after extraction)	32.4	3720	540	460
Nb of lines lost				
380 kV	0	48	5	6.3
225 kV	0	149	9.6	18.53
Thermal units				
Nb of units lost	0	15	1.15	2.1
Mechanical power lost (MW)	0	13 000	617	1213
Mech. power variation (EDF, MW)	− 20 800	546	− 1265	2445
Hydro units				
Nb of units lost	0	32	2.7	5.3
Mechanical power lost	0	2952	271	584
Mech. power variation (EDF, MW)	− 3039	60.5	− 332	604
Load variation (*MW*)				
I region	− 9046	654	864	1714
L region	− 8944	288	− 1194	2368
EDF system	− 22000	426.8	− 2417	4323
Exportation variation (*MW*)				
EDF system to Belgium	− 1527	1568	22	460
EDF system to Germany	− 1832	1607	17	476
EDF system to Switzerland	− 3301	830	− 150	400
EDF system to Italy	− 2974	1609	− 48	463
EDF system to Spain	− 1644	1253	− 148	435
EDF system to all foreign systems	− 8470	4946	− 305	1626
Voltage in buses at end of simulation (*see Fig. 4*)				
A: B.CARS71 (380 kV)	0	424	327	137
B: TAVELS71 (380 kV)	0	467	366	98
C: P.CORS71 (380 kV)	0	463	394	60
D: G.ILES71 (380 kV)	0	487	395	71
E: GEN.PS71 (380 kV)	0	469	395	57
F: VIELMS71 (380 kV)	0	449	397	45
G: BAYETS71 (380 kV)	0	416	387	55
H: VAUPAS61 (225 kV)	0	288	211	59
I: FLEACS61 (225 kV)	0	255	235	36

- Generator (under/over voltage/frequency) protections installed on units in the EDF system are often activated.
- The different *voltages* on widespread buses in the study region demonstrate the variety of the voltage collapse areas, even if some areas of the system are more deeply concerned than others.
- Among the scenarios generated randomly, not all could be simulated until the end. Thus, the analysis below concentrates on 1140 scenarios, which were simulated completely. The other scenarios will be analyzed separately.

5. Application of data mining

Some of the results obtained with data-mining methods are illustrated here, using scalar attributes derived from the temporal ones, so as to give a flavor of the type of information that can be extracted, and of the diversity of the dynamic scenarios stored in the database. To save space, the reader is referred to specialized literature for the description of the algorithms used. The reader who is interested in automatic learning theory is recommended to consult (Vidyasagar, 1997), which also provides a control-systems perspective on this subject.

5.1. Automatic identification of failure modes

The analysis starts by using unsupervised learning to identify different families of dynamic behavior: stable ones, and various types of breakdown scenarios. To this end 39 synthetic attributes are used, describing the 1140 scenarios in terms of consequences (see also Table 2): overall number of (225 and 400 kV) lines and generators

tripped, amount of load reduction in various areas, variation of active power flows through interfaces with neighboring systems, voltages at various 400 kV buses. Using the K-means clustering algorithm to automatically define a certain number of classes of scenarios (Duda and Hart, 1973), it was found that the database can be partitioned into six main classes, described below in increasing order of severity.

(1) *702 stable scenarios*. These are characterized by a mean loss of 150 MW load and six lines; voltages close to nominal at the end of the simulation.
(2) *77 local losses of synchronism*. These are characterized by a loss of synchronism of some hydro plants in region I (2000 MW lost generation, in the mean) and the tripping of some 225 kV lines. In terms of load loss, the consequences remain rather limited (1200 MW in the mean, mainly in region I close to the hydro plants), and voltages in the 400 kV system remain normal. The difference between the lost generation and the lost load is compensated for by a decrease in the power exported towards foreign systems.
(3) *176 local losses of load*. These scenarios correspond mainly to cascading losses of various transmission lines, and to almost no loss of generation. In the mean, about 2100 MW of load is lost (50% in region I, and 50% in region L) and 300 MW of generation. Thus the overhead is partially compensated for by primary frequency control in France (900MW) and an increase in exportation (900 MW). Further clustering analysis allows the identification of two subclasses: 93 scenarios with loss of load in the Rhne valley and stable voltages; 83 others yielding voltage collapse and loss of load in the South-Eastern part of region L.
(4) *113 region L voltage collapses*. In the mean, 44 EHV lines are lost, 2000 MW of generation, 6400 MW of load in region L and only 1000 MW in region I. At the end of the simulation, voltages are very low throughout region L, while they remain stable everywhere else. The large difference between the loss of load and generation is compensated for by primary frequency control (about 2000 MW) and an increase of power export (3000 MW).
(5) *33 region L + I voltage collapses*. In the mean, 17 000 MW of load is lost in France (5500 MW in region I, 6700 MW in region L, and 4800 MW outside the study region). Most of the load variation is compensated for by primary frequency control in Europe. 400 kV voltages drop to very low values within the whole study region: 170kV in the South-East and 230 kV in the North.
(6) *17 Regional loss of synchronism*. Loss of synchronism extends to the whole study region, and the defense plan trips lines to isolate it from the system, and activates load shedding in France. Thus, very large amounts of generation are lost: in the mean, 7000 MW (9 thermal and 12

hydro units). As a byproduct, voltage collapses in region L, where most of the load is lost (7000 MW in the mean). In region I, on the other hand, voltages are low (365 kV) but stable, and only 2550 MW of load is lost in the mean.

5.2. Searching for a specific kind of scenario

For example, one could search in the database for a voltage collapse in region L, with the following specification (see Fig. 4) : (i) voltages in subst. A and B close to 200 kV; (ii) voltage in subst. E close to 400 kV; (iii) loss of load of 8000 MW evenly spread among regions L and I; (iv) loss of 9 lines in the 400 kV system and 60 lines in the 225 kV system.

Using the nearest-neighbor method (Duda and Hart, 1973) it transpires that scenario No. 1439 (class 4) fits these specifications best: loss of 9400 and 65225 kV lines, 5 thermal units, 6 hydro units, 3100 MW of load in region I and 4700 MW in region L; increase in total exportation of 3600 MW; voltage in substations A, B and C of 153, 266 and 408 kV, respectively.

A brief description of the chronology of events happening within this scenario is as follows:

Initialization ($t = 0$–505 s). At $t = 20$ s a first (bus bar) fault appears in the 400 kV substation B, leading to immediate tripping of 225 kV lines (overload protections working improperly). The fault is cleared at $t = 20.1$ by permanent tripping of breakers in the substation, leading to the loss of the two 400 kV lines connected to the bus bar. At $t = 40$ s 12 tap-changer blocking devices act within the 225 kV subsystem close to the hydro plants. Nothing else happens until the occurrence of the second fault, at $t = 504$ s: this is a three-phase short-circuit on a critical 400 kV line in the Rhône valley, leading to the loss of two major 400 kV lines (a parallel line, due to overload protection maloperation, and the faulty line due to normal operation of the protections).

Intermediate stage ($t = 505$–2000 s). Given the number of circuits lost, 9 (400 and 225 kV) lines are overloaded in the study region, leading to their successive tripping between $t = 1650$ and 2000 s. During the same period, tap changers start reacting, first in region I, then in region L.

Voltage collapse in region L ($t = 2000$–2904 s). Upon loss of the last three lines the critical point is passed in the Eastern part of region L: tap changers continue raising their taps, but load decreases. At about the same time, low-voltage protections start disconnecting 56 further (mainly 225 kV) lines around Lyon, leading to the loss of 3100MW of load in region I. At $t = 2230$ s a second wave of 30 tap-changer blocking devices are activated throughout region L, unfortunately too late. At $t = 2310$ s a further 400 kV line in the North-Eastern part
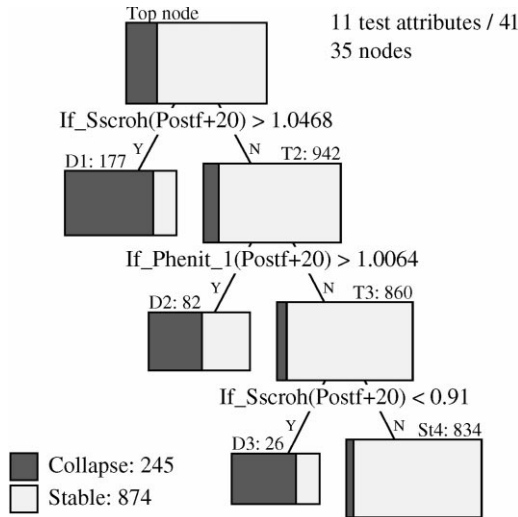
Fig. 5. Voltage collapse prediction DT (partial view).

a rather low non-detection rate (6%), and a rather high false-alarm rate (19%).

To see whether it would be possible to improve the classification accuracy at the price of being less predictive, another tree was built, using the same candidate attributes at a later time (1000 s before the end of the simulation). The resulting tree is indeed simpler and more accurate: lower non-detection rate (2%) and reduced false alarm rate (6%). Note that it has selected field currents of eight different machines in region L.

The preceding example illustrates the contradictory requirements between degree of anticipation and accuracy of emergency control criteria, and how data mining may be used in order to evaluate different types of compromises. Before drawing conclusions about possible criteria, many further investigations will be carried out, considering different types of attributes and different compromises. But the preliminary results already obtained look very promising.

5.4. Ongoing R&D

Research is proceeding along two directions, in order to take the best advantage of the probabilistic methodology.

The first direction is problem-oriented. Among others, it is aimed at defining objective scenario severity criteria, so as to identify the most dangerous modes of blackout. Another objective is to study the interactions among different types of instabilities, and the corresponding emergency-control schemes. Still another topic consists of combining fast voltage and transient stability screening tools with detailed time-domain simulations, so as to carry out efficiently global security assessment studies at the full system level.

The second research direction is aimed at developing machine-learning methods that are able to exploit the temporal nature of the databases more efficiently. The first step of this work consists of extending the decision-tree induction method, in order to build automatically, instability detection rules that offer a good compromize between selectivity and degree of anticipation (Geurts and Wehenkel, 1998b).

of region I trips on overload protection, and two seconds later some generators in the Southern part trip on under-voltage protections. About ten seconds later, some other generators are lost due to overspeed protections; two further lines are tripped on overload at $t = 2650$ and $t = 2900$ s, without notable consequences. At $t = 2904$ s the system stabilizes, with very low voltages throughout region L, and normal ones in region I.

5.3. Predicting voltage collapse in region L

Since region L is weak in terms of voltage stability, and tap-changer blocking may act too late to avoid voltage collapse, it would be interesting to find an anticipatory criterion to trigger emergency control. Thus, one can build a decision tree using information about machines' field currents to predict a voltage collapse in the South-Eastern part of region L. Thus, a scenario is classified as a collapse if, at the end of the simulation, the voltage in substation A is below 304kV. As attributes, the field currents of 41 candidate machines shortly after the second fault occurrence will be used (i.e., as early as possible after the damping of fast transients). Note that, of the 1119 scenarios, 245 are voltage collapses (230 with voltages lower than 250 kV), some local and some extending to the whole study region.

Fig. 5 shows the first levels of the tree built with the method described in [Wehenkel and Pavella, 1993]; it has 35 nodes, and selects from among the 41 machines the 11 most important ones. Note that the most discriminating attribute selected at the top-node (see Fig. 5) corresponds to a hydro unit in the South-Eastern part. Note also that, in order to avoid non-detections of collapse scenarios, the tree was biased during its building by enhancing the weights of collapse scenarios. Its reliability (estimated by cross-validation) thus corresponds to

6. Conclusions

In this paper a novel probabilistic approach to complex design problems under uncertainties was presented. It relies on Monte-Carlo sampling to generate automatically a database of detailed simulations of system performance under diverse hypotheses, as well as the systematic exploitation of this database by data-mining tools, in particular automatic learning algorithms such as decision trees or neural nets, but also simple statistical summaries and graphics.

While this approach is very general, and flexible enough to yield a broad variety of applications, this paper has focused on the design of electric power-system protection schemes against blackouts. The end-goal of this research is to develop a probabilistic methodology that will enable engineers to appraise the likelihood and the consequences (i.e., the risk) of various types of breakdown modes of their system, so as to choose the best countermeasures, and to validate them in the study environment before going towards actual field testing.

Admittedly, this goal is very ambitious and needs further research to be fully reached. The objective of the paper was therefore threefold: (i) to give a prospective discussion of the overall approach; (ii) to describe the tools developed to make it practicable; (iii) to show its technical feasibility on the basis of first results coming from a real-life study on a large-scale system.

As far as future work is concerned, the first step will consist of taking full advantage of the database. Customization of some of the automatic learning algorithms (decision trees, in particular) is under way so as to handle the temporal information more effectively. In the longer term, the approach should lead to a complete probabilistic methodology for security assessment and design of power systems.

Acknowledgements

References

Cigré Task Force 38.03.12, Marceau, R., & Endreyeni, J. (convenors) (1997). Power system security assessment : A position paper. *Electra*, *175*, 49–78.

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis.* New York: Wiley.

Geurts, P., & Wehenkel, L. (1998a). Visualizing Dynamic Power System Scenarios for Data Mining. *Proc. of LESCOPE'98, Halifax (Ca), 7–9 June.* To appear.

Geurts, P., & Wehenkel, L. (1998b). Early prediction of electric power system blackouts by temporal machine learning (submitted for publication).

Jacquemart, Y. et al. (1996). Statistical approaches to dynamic security assessment: The data base generation problem. *Proc. SIPOWER'95, 2nd IFAC Symp. on Cont. of Power Plants and Power System.* (pp. 243–246).

Lebrevelec, C., et al. (1997). A new probabilistic approach to improve security assessment. *EPRI-IERE Workshop on Future Trends in Power System Reliability Assessment.*

Meyer, B., & Stubbe, M. (1992). EUROSTAG, a single tool for power system simulation. *Transmission Distribution Int.*, *3*(1), 47–52.

Trotignon, M., et al. (1992). Defense plan against major disturbances on the EHV power system: present realization and prospects of evolution. *CIGRE Report 39-306, Paris,* August.

Vidyasagar, M. (1997). *A theory of learning and generalization.* Berlin: Springer.

Wehenkel, L., Pavella, M. (1993). Decision tree approach to power system security assessment. *Electr Power Energy Systems*, *15*(1), 13–36.

Wehenkel, L., et al. (1994a). Decision tree based transient stability method – a case study. *IEEE Trans. Power Systems PWRS-9*, 459–469.

Wehenkel, L., et al. (1994b). Machine learning, neural networks and statistical pattern recognition for voltage security: a comparative study. *Engng Int. System Electr. Engng Commun.*, *2*(4), 233–245.

Wehenkel, L., et al. (1995). Automatic learning approaches for on-line transient stability preventive control of the Hydro-Québec system – Part I. Decision tree approaches. *Proc. of SIPOWER'95, 2nd IFAC Symp. on Cont. of Power Plants and Power System* (pp. 231–236).

Wehenkel, L. (1998). *Automatic learning techniques in power systems.* Dordrecht: Kluwer Academic.