

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

THÈSE PRÉSENTÉE À
L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

COMME EXIGENCE PARTIELLE
À L'OBTENTION DU
DOCTORAT EN GÉNIE
Ph. D.

PAR
Anissa KACI

ANALYSE ET ÉVALUATION DE LA SÉCURITÉ DYNAMIQUE DU RÉSEAU
ÉLECTRIQUE PAR LE FORAGE DES MESURES SYNCHRONES

MONTRÉAL, LE 31 MARS 2016



Anissa Kaci, 2016



Cette licence [Creative Commons](#) signifie qu'il est permis de diffuser, d'imprimer ou de sauvegarder sur un autre support une partie ou la totalité de cette œuvre à condition de mentionner l'auteur, que ces utilisations soient faites à des fins non commerciales et que le contenu de l'œuvre n'ait pas été modifié.

PRÉSENTATION DU JURY

CETTE THÈSE A ÉTÉ ÉVALUÉE

PAR UN JURY COMPOSÉ DE :

M. Louis-A. Dessaint, directeur de thèse
Département de génie électrique à l'École de technologie supérieure

M. Innocent Kamwa, codirecteur de thèse
Institut de recherche d'Hydro-Québec

M. Roger Champagne, président du jury
Département de génie logiciel et des TI à l'École de technologie supérieure

Mme. Ouassima Akhrif, membre du jury
Département de génie électrique à l'École de technologie supérieure

M. Pierre Jean Lagacé, membre du jury
Département de génie électrique à l'École de technologie supérieure

M. Gérard-André Capolino, membre externe indépendant
Université de Picardie, Amiens, France

ELLE A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 09 MARS 2016

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

Je tiens à exprimer mes vifs remerciements et mes sincères reconnaissances à toutes les personnes qui m'ont permis, de près ou de loin, d'une façon ou d'une autre, par leur collaboration, leur soutien et leur avis judicieux, de mener à bien ce travail.

Je remercie M. Louis-A. Dessaint, mon directeur de thèse, pour l'honneur qu'il m'a fait en acceptant de diriger ce travail ainsi que pour sa confiance qu'il a affiché envers moi concernant l'achèvement de ma mission, malgré les nombreuses embûches, liées aux aléas de la vie, que j'ai rencontrées durant la réalisation de ce travail.

Je tiens à exprimer mes remerciements ainsi que ma reconnaissance à M. Innocent Kamwa, mon codirecteur, Professeur associé aux Universités de Mc-Gill et Laval et Chef expertise réseaux électriques et mathématiques de l'Institut de Recherche d'Hydro-Québec (IREQ), pour l'honneur qu'il m'a fait en me proposant ce travail. Je le remercie de m'avoir encadré, orienté, aidé et conseillé. Je le remercie pour sa disponibilité, malgré ses nombreuses préoccupations, pour ses encouragements et sa mise en ma disposition les moyens nécessaires pourachever ce travail et accomplir ma tâche. Il m'a écouté avec modestie et tolérance ce qui m'a permis d'élaborer ce travail dans les meilleures conditions. Grâce aux discussions que nous avons eues et son soutien aux moments les plus difficiles, j'ai amélioré mon bagage d'un point de vue théorique, profité de sa longue expérience, de ses riches connaissances et de sa vision profonde, rigoureuse et clairvoyante. Ses qualités humaine et scientifique m'ont, considérablement, motivées et aidées à surmonter les difficultés inhérentes à certaines étapes lors de la réalisation de ce travail. Le travail avec lui est un grand honneur et un plaisir. Permettez-moi M. Innocent de vous témoigner mes sincères respects et considérations les plus distinguées.

Je remercie les membres de jury qui ont évalué ma thèse : Prof. Roger Champagne (École de technologie supérieure), Prof. Ouassima Akhrif (École de technologie supérieure), Prof. Pierre Jean Lagacé (École de technologie supérieure) et Prof. Gérard-André Capolino

(Université de Picardie, Amiens). Qu'ils trouvent ici mes sincères remerciements pour l'intérêt affiché à mon travail.

Je voudrais, en particulier, exprimer ma reconnaissance à M. Sébastien Guillon, expert en Planification et Stratégie du Réseau Principal à Hydro-Québec-TransÉnergie, pour ses écrits, ses suggestions, ses remarques constructives qu'il a formulées, et ce en se basant sur sa riche expérience et son expertise dans l'analyse des réseaux d'Hydro-Québec. Par ailleurs, il sied de noter sa contribution à l'interprétation, d'un point de vue pratique, de certains résultats obtenus au chapitre 4.

Je tiens aussi à remercier tout le corps de chercheurs et scientifiques de la division Expertise réseaux électriques et mathématiques, ainsi que le personnel de la bibliothèque de l'IREQ pour leur sympathie et les bons moments passés ensemble.

Je prends aussi un réel plaisir à remercier chaleureusement toutes les personnes que j'ai rencontrées au cours de la réalisation de cette thèse. Nos rencontres et nos discussions ont laissé une empreinte particulière sur ma volonté d'aller jusqu'à l'aboutissement de ma mission. Je cite particulièrement K. Dragon, D. Martin et H. Annissa, M. Fatima, D. Hanane et L. Jean-Marc.

Enfin, je souhaite dédier cette thèse à toute ma famille. Je voudrais tout particulièrement exprimer ma reconnaissance à ma mère, mon père, mes sœurs, mes frères, mon époux et mes enfants.

À tous ces intervenants, je présente mes remerciements, mon respect et ma gratitude.

ANALYSE ET ÉVALUATION DE LA SÉCURITÉ DYNAMIQUE DU RÉSEAU ÉLECTRIQUE PAR LE FORAGE DES MESURES SYNCHRONES

Anissa KACI

RÉSUMÉ

Les synchrophaseurs (*Phasor Measurement Units PMUs*) sont considérés aujourd'hui comme des appareils de mesure les plus importants dans l'avenir des réseaux électriques. Dans l'optique de surveiller la sécurité du réseau de transport d'énergie et spécifiquement de fournir à l'opérateur une visibilité étendue du réseau de transport (*wide-area grid visibility*), les synchrophaseurs ont connu une utilisation importante en Amérique du nord. Ainsi, aux États-Unis seulement, l'installation de ces appareils est passée de 166 à 1 043 de 2010 à 2014.

Cette thèse a pour objectif de répondre à la problématique du procédé d'utilisation des mesures des synchrophaseurs pour tirer leur plein potentiel lors de leur intégration dans l'environnement opérationnel de la sécurité du réseau de transport. En effet, en vertu des qualités de mesure des synchrophaseurs telles que la précision, la latence et la visibilité, il y a le grand volume de données en temps réel à savoir gérer. À cet effet, une procédure systématique, en trois étapes, de développement des modèles de surveillance de la sécurité du réseau par le biais de phaseurs synchrones est proposée. Cette démarche s'inscrit dans le cadre d'apprentissage statistique. Le jeu de données analysé est extrait de la base de données EMS/SCADA d'Hydro-Québec (HQ) et s'échelonne sur quatre ans, à un intervalle d'une minute.

En premier lieu, suite à une étude purement statistique de la base de données, nous avons identifié les angles de phase de tension comme prédicteurs de stabilité et en conséquence, défini des limites des angles de phase où le réseau de transport semble stable. En second lieu, avec l'algorithme des forêts aléatoires (FA), nous avons établi des modèles de prévision des marges de sécurité et des transits actifs à base des prédicteurs définis. Puis une étude complète de validation a été effectuée. En troisième abord, avec l'algorithme GLM (modèle linéaire généralisé), nous avons élaboré des modèles de prévision des angles de phase de tension en fonction de leurs états antérieurs : une heure, quatre heures ou vingt-quatre heures au préalable. Le mappage de ces angles prédits sur les modèles établis par les FA nous a permis d'anticiper l'évaluation de nouvelles valeurs limites de stabilité une heure, quatre heures ou vingt-quatre heures en avance.

Notre première contribution est axée principalement sur l'utilisation d'une base de données réelle du réseau de transport d'HQ pour la conception et la validation de nos modèles. Notre deuxième contribution concerne l'utilisation des techniques de forage de données pour proposer une solution à l'évaluation rapide des limites de stabilité dynamique d'un réseau de transport. Notre troisième contribution consiste en l'identification des écarts angulaires en tant que prédicteurs de transit de puissance ainsi que des limites de stabilité dynamique associées. Par ailleurs, la quatrième contribution se résume à la conception d'une nouvelle

approche reflétant le comportement dynamique réel du réseau de transport à partir des données de synchrophaseurs. Finalement, notre apport d'un point de vue économique consiste en la bonification de l'approche conventionnelle de surveillance de la sécurité dynamique par simulation avec l'apport des mesures synchronisées.

Ce travail ouvre plusieurs perspectives de recherches, à savoir : généraliser l'approche adoptée à des réseaux maillés ou non ; introduire de nouvelles fonctionnalités aux écarts angulaires (réserves réactives, injections de puissance active et réactive, les barres critiques, etc.) ; combiner les FA avec la logique floue afin d'améliorer la prévision des angles de phase de tension ; et enfin réaliser une étude similaire pour identifier d'autres signatures de défauts.

Mots clés : Sécurité des réseaux électriques, Évaluation de la sécurité dynamique (*DSA*), Unité de mesure synchrophaseur (*PMU*), La fouille de données, Forêts Aléatoires, Modèles linéaires généralisés (*GLM*).

ANALYSIS AND EVALUATION ELECTRICAL NETWORK DYNAMIC SECURITY BY SYNCHROPHASOR MEASURES DATA MINING

Anissa KACI

ABSTRACT

Phasor Measurement Units (PMUs) are nowadays considered as the most important measuring devices in the future of power networks. In order to monitor the energy security and specifically to provide a wide-area grid visibility to the operator, the PMUs have seen significant use in North America. In the United States alone, the number of PMUs is increased from 166 networked devices in 2010 to 1,043 in 2014.

This thesis aims to address the issue of the process of using PMUs measures to achieve their full potential during their integration in the operational environment of transmission power system security. Indeed at PMU measures qualities such as accuracy, latency and visibility, there is the large volume of real-time data to know manage. To this end, based on phasor measurements, a systematic procedure development of security monitoring models in three steps is proposed.

This process is in the context of a statistical learning framework. The analyzed data set is extracted from the database EMS/SCADA of Hydro-Québec (HQ), and spread over four years, at an interval of one minute. First, following a statistical study of the database, we identified the voltage phase angles as predictors of stability and consequently, defined the limits of the phase angles where the transmission power system appears stable. Second, with random forests (RF) algorithm in the statistical learning framework, we established forecast models security margins and power transfers with defined predictors. Then, a comprehensive validation study was performed. Third, with a generalized linear model (GLM) algorithm, we developed forecast models of voltages phase angles from day-ahead to hour-ahead time frames. The mapping of these voltages phase angles predicted on models established by the FA, allowed us to anticipate the computation of new stability limit values from day-ahead to hour-ahead time frames.

Our first contribution is mainly focused on the use of a real database of HQ's network for design and validation of our models. Our second contribution is the use of data mining techniques to propose a solution for the rapid assessment of the limits of dynamic stability. Our third contribution is the identification of the angular deviations as power flow predictors as well as associated dynamic stability limits. Furthermore, the fourth contribution is related to the design of a new approach reflecting real dynamic behavior of the transmission network from Synchrophasors data. Finally, from an economic viewpoint, this thesis contributes to the improvement of the conventional approach based on the simulation of the monitoring of the dynamic security with synchronized measurements inputs.

Our work provides many research perspectives, namely: generalize the used approach to mesh networks or not; introduce new features to the angular deviations (reactive reserves, active and reactive power injections, critics 'bars, etc.); combine RF with fuzzy logic in order to improve the prediction of phase angles voltage; and finally conduct a similar study to identify other defected models.

Key Words: Power system security, Dynamic security assessment (DSA), Phasor measurement unit (PMU), Data mining, Random Forest (RF), General linear models (GLM).

TABLE DES MATIÈRES

	Page
INTRODUCTION.....	1
CHAPITRE 1 REVUE DE LITTÉRATURE.....	5
1.1 Introduction.....	5
1.2 L'évaluation de l'analyse de la sécurité du réseau électrique	6
1.3 Les politiques de la fiabilité et la planification des opérations	8
1.4 L'évaluation de la sécurité dynamique en ligne dans la littérature scientifique	9
1.4.1 Intégration numérique	10
1.4.2 Méthode de Lyapunov	10
1.4.3 Méthodes probabilistes	11
1.4.4 Méthodes Système expert	11
1.4.5 Méthodes de reconnaissance des formes	12
1.5 L'évaluation de la sécurité dynamique en ligne sur le plan opérationnel.....	13
1.5.1 Les critères de l'évaluation de la sécurité dynamique	13
1.5.2 Réflexions sur les méthodes d'évaluation de la sécurité dynamique.....	15
1.6 La synergie entre les nouvelles technologies et le progrès de la DSA en ligne.....	16
1.6.1 La DSA en temps réel	16
1.6.2 Les synchrohaseurs (<i>PMUs</i>)	17
1.6.3 La mesure à grande échelle.....	18
1.6.4 L'utilisation des synchrophaseurs pour détecter la vulnérabilité d'un réseau électrique de transport.....	19
1.7 Le processus de sécurité du réseau de transport d'HQ	20
1.7.1 Généralités	20
1.7.2 Le logiciel LIMSEL (LIMit SElection).....	23
1.8 Présentation de la problématique	24
1.8.1 Objectif	25
1.8.2 La méthodologie	26
1.8.3 Notre contribution	27
CHAPITRE 2 LE CONTEXTE DE L'ÉTUDE ÉLABORÉE.....	29
2.1 Introduction.....	29
2.2 Description du réseau électrique d'HQ.....	29
2.3 Les écarts angulaires en tant que prédicteurs de la sécurité dynamique du réseau électrique de transport.....	32
2.4 Originalité de notre travail	33
2.5 Base de données des synchrophaseurs	35
2.6 SCADA vs. synchrophaseur	35
2.7 Base de données EMS/SCADA	37
2.8 Prétraitement des données.....	38
CHAPITRE 3 LE FORAGE DE DONNÉES	41

3.1	Introduction.....	41
3.2	Le forage de données	41
3.3	L'apprentissage statistique et la modélisation	43
3.4	Le logiciel R.....	44
3.5	Arbre de régression	48
	3.5.1 Le concept.....	48
	3.5.2 Mise en œuvre sous R.....	49
3.6	Les forêts aléatoires	55
	3.6.1 Généralités	55
	3.6.2 Sélection de variables	56
	3.6.3 La prévision avec les forêts aléatoires	56
	3.6.4 Le traitement des données catégorielles	57
	3.6.5 Le traitement des valeurs manquantes	58
	3.6.6 Mise en œuvre sous R.....	58
3.7	Les réseaux de neurones	63
	3.7.1 Principe	63
	3.7.2 Mise en œuvre sous R.....	65
3.8	Les modèles linéaires généralisés	66
	3.8.1 Théorie	66
	3.8.2 Mise en œuvre sous R.....	68
3.9	Discussion	70
3.10	Conclusion	74
 CHAPITRE 4 L'APPROCHE STATISTIQUE DE L'ÉVALUATION DE LA SÉCURITÉ DYNAMIQUE DU RÉSEAU ÉLECTRIQUE DE TRANSPORT		
4.1	Introduction.....	75
4.2	Fondements théoriques	75
	4.2.1 Relation entre les limites de puissance et les phaseurs	76
	4.2.1.1 Modèles linéaires (faible écart d'angle de tension)	77
	4.2.1.2 Modèle non-linéaire (indépendamment des écarts angulaires de tension)	79
	4.2.2 L'évaluation de la capacité de charge des lignes de transport à haute tension et/ou des corridors de transport	82
4.3	Les données de l'EMS-SCADA	86
4.4	Les variations des angles de phase de tension des postes à 735kV	92
	4.4.1 Diagramme en boîte (<i>Box plot</i>).....	92
	4.4.2 Interprétation.....	93
4.5	La dépendance mutuelle des postes	96
	4.5.1 Matrice de corrélation des angles de phase de tension des postes à 735kV.....	97
	4.5.2 Interprétation.....	98
4.6	Les angles de phase de tension versus les transferts et les marges de puissance.....	102
4.7	Les prédicteurs de la stabilité dynamique	104
	4.7.1 L'identification des angles de phase clefs par PAM.....	111

4.7.2	Calibration statistique des angles de phase clefs	114
4.8	Conclusion	125
CHAPITRE 5 PRÉVISION DES LIMITES DE SÉCURITÉ DYNAMIQUE PAR LES FORÊTS ALÉATOIRES		127
5.1	Introduction.....	127
5.2	Prévision des transits et des marges de sécurité par les forêts aléatoires.....	127
5.3	Résultats et commentaires.....	129
5.4	La consistance de l'approche de prévision de la sécurité dynamique par les angles de phase de tension des barres à à 735kV avec la méthode des FA.....	145
5.5	Conclusion	156
CHAPITRE 6 PRÉVISIONS ANTICIPÉES DES ÉCARTS ANGULAIRES DES TENSIONS ET LEURS INTÉGRATION DANS LA CONDUITE DU RÉSEAU		159
6.1	Introduction.....	159
6.2	Modélisation des angles de phase de tension des barres du réseau de transport	159
6.3	Prévisions des angles de phase de tension des postes à 735kV du réseau de transport	161
6.4	Les critères d'évaluation de la qualité des modèles de prévisions.....	163
6.5	Les performances des différents algorithmes de régression	163
6.6	Prévisions anticipées des angles de phase de tension par GLM	166
6.7	Le couplage entre les approches à base de GLM et les modèles à base de FA	177
6.8	Évaluation de prévisions anticipées des transferts actifs et des marges de sécurité ..	179
6.9	Conclusion	186
CONCLUSION		187
RECOMMANDATIONS.....		191
ANNEXE I	PROGRAMMER AVEC R	193
ANNEXE II	LES PARAMÈTRES DES MODÈLES G06_24H ET G06_4H	197
ANNEXE III	RÉSULTATS DE PRÉVISIONS	203
LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES		209

LISTE DES TABLEAUX

	Page
Tableau 3.1	Les modèles linéaires générales 66
Tableau 3.2	Les familles des GLM 68
Tableau 3.3	Comparaison des différentes méthodes 71
Tableau 3.4	Résumé des caractéristiques statistiques des erreurs de prévisions 73
Tableau 4.1	Description des variables d'entrée 88
Tableau 4.2	Description des variables de sortie 89
Tableau 4.3	La corrélation du poste Boucherville avec le reste des postes à 735kV .. 100
Tableau 4.4	Les regroupements suite à l'algorithme PAM 114
Tableau 4.5	Caractéristiques statistiques des angles de phase de tension de quelques postes pendant les saisons typiques 122
Tableau 5.1	L'importance relative des prédicteurs lors de la formation des modèles de prévision des transits actifs 130
Tableau 5.2	L'évaluation de la performance des modèles de FA (Base de données : l'année 2010 par 3 minutes) 131
Tableau 5.3	Caractéristiques statistiques de l'ensemble test extrait de la base de données de 2010 à 3 minutes 144
Tableau 5.4	Caractéristiques statistiques de l'ensemble prédit par les FA en fonction de tous les angles de la base de données de 2010 à 3 minutes 144
Tableau 5.5	Caractéristiques statistiques des erreurs de prévisions en fonction de tous les angles par les modèles de FA établis sur l'ensemble d'apprentissage de données de l'année 2010 à 3 minutes 144
Tableau 5.6	Caractéristiques statistiques de l'ensemble prédit par les FA en fonction des angles de phase des barres médiods de la base de données de 2010 à 3 minutes 145
Tableau 5.7	Caractéristiques statistiques des erreurs de prévisions en fonction des angles des barres médiods par les modèles de FA établis sur l'ensemble d'apprentissage de données de l'année 2010 à 3 minutes 145

Tableau 5.8	L'évaluation des modèles de prévisions par les FA des transits actifs et des marges de sécurité à base de données des années 2009 à 2012.....	146
Tableau 5.9	Caractéristiques statistiques de l'ensemble test extrait de la base de données de 3 ans à 9 minutes	151
Tableau 5.10	Caractéristiques statistiques de l'ensemble prédit par les FA en fonction de tous les angles à base de données de 3 ans à 9 minutes	151
Tableau 5.11	Caractéristiques statistiques des erreurs de prévisions en fonction de tous les angles par les modèles de FA établis sur l'ensemble d'apprentissage de données de 2010 à 2012 à 9 minutes	151
Tableau 6.1	Les variables des modèles des angles de phase de tension des barres à 735kV	161
Tableau 6.2	Un aperçu sur le modèle G06.1H avec la méthode GLM.....	167
Tableau 6.3	Le test d'ANOVA du modèle G06.1H avec la méthode GLM.....	169
Tableau 6.4	Les mesures d'évaluation des modèles de prévisions des transits de puissance et des marges de sécurité	179
Tableau 6.5	Les prévisions par le mappage et les valeurs réelles.....	185

LISTE DES FIGURES

	Page	
Figure 0.1	Le déploiement des synchrophaseurs à travers l'Amérique du Nord Tirée de DOE (2014, p. 3)	2
Figure 1.1	Les huit régions de NERC Adaptée de NERC (2014, p. 4).....	22
Figure 2.1	Le réseau de transport d'Hydro-Québec TransÉnergie	31
Figure 2.2	Base de données utilisée pour l'étude	39
Figure 3.1	Le R sous Windows	46
Figure 3.2	L'interface de R Commander	46
Figure 3.3	L'interface de Rattle	47
Figure 3.4	L'interface de RStudio	47
Figure 3.5	Les 21 premières observations du fichier de données Exemple	50
Figure 3.6	L'arbre de régression	51
Figure 3.7	Impression sous forme textuelle de l'arbre de régression.....	52
Figure 3.8	Les règles générées avec rpart	53
Figure 3.9	L'évaluation du modèle de prévision par l'arbre de régression -Exemple-	54
Figure 3.10	La sortie du modèle de forêt aléatoire.....	60
Figure 3.11	Taux d'erreurs en fonction du nombre d'arbres.....	61
Figure 3.12	Les règles du premier arbre de la forêt aléatoire.....	62
Figure 3.13	L'évaluation du modèle de prévision par la FA -Exemple-	63
Figure 3.14	Représentation d'un neurone formel	64
Figure 3.15	Exemple de perceptron multicouche élémentaire	64
Figure 3.16	L'évaluation du modèle de prévision par le réseau de neurones -Exemple-	65

Figure 3.17	Les paramètres du modèle GLM - Exemple-	69
Figure 3.18	L'évaluation du modèle de prévision par GLM	70
Figure 3.19	Les séries chronologiques des erreurs de prévisions	72
Figure 3.20	Les erreurs de prévisions avec les F.A en fonction du paramètres "m"	74
Figure 4.1	Le modèle simplifié du réseau de transport entre deux postes	77
Figure 4.2	La ligne de transport avec les deux équivalents de Thevenin Tirée de Gutman, Marchenko et Dunlop (1979, p. 2).....	83
Figure 4.3	Les caractéristiques de capacité de charge d'une ligne de transport Tirée de Gutman (1988, p. 4).....	84
Figure 4.4	La marge de stabilité à l'état stable Tirée de Gutman (1988, p. 5)	85
Figure 4.5	Les postes du réseau de transport.....	87
Figure 4.6	Séries chronologiques à partir des archives CILEX et EMS (a) Les écarts angulaires des tensions de quelques barres (b) Les transits actifs à travers les cinq interfaces.....	90
Figure 4.7	Box plot des angles de phase de tension des barres à 735kV (175 000 échantillons).....	92
Figure 4.8	Matrice de corrélation des angles de phase de tension des 37 postes.....	98
Figure 4.9	Matrice de corrélation de la variation des angles de phase de tension ...	101
Figure 4.10	La corrélation des angles de phase de tension des barres avec : (a) Les transferts de puissance, (b) Les marges de sécurité	103
Figure 4.11	Le niveau maximum de corrélation des angles avec les transferts, les marges et les limites de puissance sur 3 ans par minutes	104
Figure 4.12	Les tendances mensuelles et horaires en 2010 des marges de transferts vs. les angles de phase de tension	105
Figure 4.13	Les nomogrammes de transits (en haut) à travers l'interface SUD et (en bas) à travers l'interface BJT à base des phases de 2010.....	106
Figure 4.14	Les variations quotidiennes des angles de phase de tension en 2010	108
Figure 4.15	Profilage saisonnier des angles de phase de tension	109

Figure 4.16	Comparaison des profils des angles de phase de tension pendant les mois de janvier et d'août de 2010	110
Figure 4.17	Les tendances mensuelles -la marge vis-à-vis l'angle de phase-	111
Figure 4.18	La visualisation de la distance de Manhattan utilisée par la matrice de dissimilarité.....	112
Figure 4.19	La répartition des angles de phases de tension la plus optimale.....	113
Figure 4.20	Les courbes des dépassements saisonniers en périodes de pointe (en pointillé) et hors pointe (en continu).....	115
Figure 4.21	La variation quotidienne des angles de tension en été 2010	116
Figure 4.22	La variation quotidienne de l'angle de phase de la tension D64 en périodes de pointe et hors pointe de l'été 2010	117
Figure 4.23	La variation quotidienne de l'angle de phase de la tension D60 en périodes de pointe et hors pointe de l'été 2010	118
Figure 4.24	La variation quotidienne de l'angle de phase de la tension D06 en périodes de pointe et hors pointe de l'été 2010	119
Figure 4.25	La variation quotidienne de l'angle de phase de la tension D50 en périodes de pointe et hors pointe de l'été 2010	120
Figure 4.26	La variation quotidienne de l'angle de phase de la tension D30 en périodes de pointe et hors pointe de l'été 2010	121
Figure 4.27	Le dépassement du seuil des angles.....	123
Figure 4.28	Détection des évènements à base des dépassements des angles	124
Figure 5.1	Le principe de prévision des transits actifs et des marges de sécurité à base des FA	128
Figure 5.2	L'erreur de prévision en fonction de nombres d'arbres de régression.....	129
Figure 5.3	L'évaluation du modèle de prévision avec la méthode des FA.....	132
Figure 5.4	Les séries chronologiques des erreurs de prévisions avec les modèles de FA de la marge de sécurité du corridor BZN.....	134
Figure 5.5	Les séries chronologiques des erreurs de prévisions avec les modèles de FA de la marge de sécurité du corridor BJS	135

Figure 5.6	Les séries chronologiques des erreurs de prévisions avec les modèles de FA de la marge de sécurité du corridor MQ	135
Figure 5.7	Les séries chronologiques des erreurs de prévisions avec les modèles de FA de la marge de sécurité du corridor CHU	136
Figure 5.8	Les séries chronologiques des erreurs de prévisions avec les modèles de FA de la marge de sécurité du corridor SUD	136
Figure 5.9	Les séries chronologiques des erreurs de prévisions par les FA des transferts de puissance de l'année 2010 par 3 minutes à travers les interfaces BJT et BJS.....	137
Figure 5.10	Les séries chronologiques des erreurs de prévisions par les FA des transferts de puissance de l'année 2010 par 3 minutes à travers les interfaces CHU, MQ et SUD	138
Figure 5.11	Les densités des erreurs de prévisions des transits actifs du sous-ensemble test en fonction de tous les angles par les modèles de FA établis sur l'ensemble d'apprentissage de données de l'année 2010 à 3 minutes	140
Figure 5.12	Les densités des erreurs de prévisions des marges de sécurité du sous-ensemble test en fonction de tous les angles par les modèle de FA établis sur l'ensemble d'apprentissage de données de l'année 2010 à 3 minutes	141
Figure 5.13	Les densités des erreurs de prévisions des transits actifs du sous-ensemble test en fonction des angles des barres médiods par les modèles de FA établis sur l'ensemble d'apprentissage de données de l'année 2010 à 3 minutes	142
Figure 5.14	Les densités des erreurs de prévisions des marges de sécurité du sous-ensemble test en fonction des angles des barres médiods par les modèles de FA établis sur l'ensemble d'apprentissage de données de l'année 2010 à 3 minutes.....	143
Figure 5.15	Les erreurs de prévisions par les FA du transit actif à travers l'interface BJT et de la marge de sécurité associée sur une période de 3 ans par 9 minutes.....	148
Figure 5.16	Les erreurs de prévisions par les FA du transit actif à travers l'interface BJS et de la marge de sécurité associée sur une période de 3 ans par 9 minutes.....	149
Figure 5.17	Les erreurs de prévisions par les FA du transit actif à travers l'interface CHU et de la marge de sécurité associée sur une période de 3 ans par 9 minutes.....	149

Figure 5.18	Les erreurs de prévisions par les FA du transit actif à travers l'interface MQ et de la marge de sécurité associée sur une période de 3 ans par 9 minutes.....	150
Figure 5.19	Les erreurs de prévisions par les FA du transit actif à travers l'interface SUD et de la marge de sécurité associée sur une période de 3 ans par 9 minutes.....	150
Figure 5.20	Les distributions des erreurs de prévisions des transits actifs du sous-ensemble test en fonction de tous les angles par les modèles de FA établis sur l'ensemble d'apprentissage de données de trois ans à 9 minutes	152
Figure 5.21	Les densités des erreurs de prévisions des transits actifs du sous-ensemble test en fonction de tous les angles par les modèles de FA établis sur l'ensemble d'apprentissage de données de trois ans à 9 minutes	153
Figure 5.22	Les distributions des erreurs de prévisions des marges de sécurité du sous-ensemble test en fonction de tous les angles par les modèles de FA établis sur l'ensemble d'apprentissage de données de trois ans à 9 minutes	154
Figure 5.23	Les densités des erreurs de prévisions des marges de sécurité du sous-ensemble test en fonction de tous les angles par les modèles de FA établis sur l'ensemble d'apprentissage de données de trois ans à 9 minutes	155
Figure 6.1	Les méthodes NNET, rpart et GLM et la caractéristique MAE	165
Figure 6.2	Les méthodes NNET, rpart et GLM et la caractéristique R.....	165
Figure 6.3	Les prévisions anticipées des angles médoids	170
Figure 6.4	Prévisions anticipées des angles de phase de tension	171
Figure 6.5	Les performances des modèles de prévisions une heure en avance.....	172
Figure 6.6	Les corrélations entre les ensembles prévus et l'ensemble de mesures réelles de l'angle de phase de tension du poste Micoua sur le même ensemble test.....	173
Figure 6.7	Évaluation de la performance du modèle GLM avec le critère R (haut) et les résidus (bas).....	174
Figure 6.8	Les prévisions anticipées des angles médoids	175
Figure 6.9	Les erreurs brutes de prévisions anticipées des angles médoids.....	176

Figure 6.10	Construction des modèles de prévisions des transits actifs et des marges de sécurité à base des angles de phase réels par la méthode de FA.....	178
Figure 6.11	Le mappage de l'ensemble des angles prévus par GLM sur les modèles de prévisions déjà établis des transits actifs et des marges de sécurité	178
Figure 6.12	Les modèles des FA des transits actifs à travers les différentes interfaces à base des angles de phase réels	181
Figure 6.13	Les prévisions anticipées des transits en MW à travers le corridor SUD.....	182
Figure 6.14	Les prévisions anticipées des transits en MW à travers le corridor BJR.....	182
Figure 6.15	Les prévisions anticipées des transits en MW à travers le corridor CHU	183
Figure 6.16	Les prévisions anticipées des transits en MW à travers le corridor MQ	183
Figure 6.17	Les prévisions anticipées des marges de sécurité de l'interface SUD	184
Figure 6.18	La marge prévue et la marge réelle du corridor SUD	185

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

DOE	US Departement Of Energy
DSA	Dynamic security assessment : l'évaluation de la sécurité dynamique
EMS	Energy Management System
EMS/SCADA	Energy Management System (Systèmes de contrôle de stabilité)
ERCOT	Electric Reliability Council of Texas
FA	Forêts aléatoires
FERC	Federal Energy Regulatory Commission
GLM	Modèles linéaires généralisés
GPS	Global Positioning systems
HQ	Hydro-Québec
HQTE	Hydro-Québec TransÉnergie : Hydro-Québec Transport
IREQ	Institut de recherche d'Hydro-Québec
MAIS	Mancœuvre Automatisée d'Inductances Shunt
NA	Valeurs manquantes
NPCC	Northeast Power Coordinating Councils
NERC	North American Electric Reliability Council
PAM	Partitioning Around Medoids
PDC	Phasor Data Concentrator
PMU	Phasor Measurement Unit: synchrophaseur
<i>P</i>	Puissance active (MW)
<i>Q</i>	Puissance réactive (MVAR)
R	Le logiciel de statistique R

RPTC	Rejet de Production et Télé délestage de Charge
SBSA	Synchrophasor-Based Situational Awareness : Éveil situationnel à base de synchrophaseurs
SCADA	Supervisory Control And Data Acquisition
SMDA	Système de Mesure du Décalage Angulaire
SPS	Special Protection System : grands automatismes de sauvegarde
SSA	Static Security Assessment : l'évaluation de la sécurité statique
TSA	Transient Security Assessment : l'évaluation de la sécurité transitoire
VSA	Voltage Security Assessment : l'évaluation de la sécurité de tension
WASA	Wide-Area Situational Awareness : Éveil situationnel à grande échelle
WAMS	Wide-Area Measurement Systems : système de mesure à grande échelle
WECC	Western Electricity Coordinating Council

INTRODUCTION

La technologie des synchrophaseurs a été inventée dans les années 1980. Sa présence dans les réseaux électriques a été essentiellement limitée à des applications de recherche. Une synthèse d'état d'art sur l'intérêt et l'utilité des PMUs (*Phasors Measurement Units*) dans le domaine de réseaux électriques a été rapportée par les auteurs de (Aminifar et al., 2015).

L'ampleur des améliorations que peut mener l'intégration des synchrophaseurs dans un environnement opérationnel, est due fondamentalement à la qualité supérieure de leurs mesures telles que la précision, la latence et la visibilité. À cet effet, afin d'améliorer la sécurité des réseaux électriques, l'utilisation des synchrophaseurs a été fortement recommandée en 2004, principalement pour fournir en temps réel une visibilité globale de l'ensemble des réseaux interconnectés (la grille). Depuis, leur installation et leur utilisation ont connu une croissance phénoménale comme en témoigne la Figure 0.1.

Pour atteindre le plein potentiel des synchrophaseurs lors de leur intégration dans un environnement opérationnel, tel qu'il est mentionné dans le rapport (DOE, 2014), d'autres points essentiels s'imposent :

1. De nombreux synchrophaseurs devraient être installés dans le but de créer une «masse critique» de capteurs qui pourrait vraiment caractériser les opérations de réseaux ;
2. Des systèmes de communication capables de transmettre de grandes quantités d'informations à une échelle de temps suffisamment courte afin qu'ils soient utiles pour les opérateurs de réseaux ;
3. Des normes de données et de communications pour permettre le partage de l'information des synchrophaseurs à travers tout le réseau ;

4. Des systèmes de gestion et de traitement des données pour traiter les grands volumes de données des synchrophaseurs en temps réel ;
5. Des logiciels d'application pour utiliser les informations des synchrophaseurs afin d'améliorer la modélisation de prévision et le contrôle des réseaux (Kezunovic, Le et Grijalva, 2013).

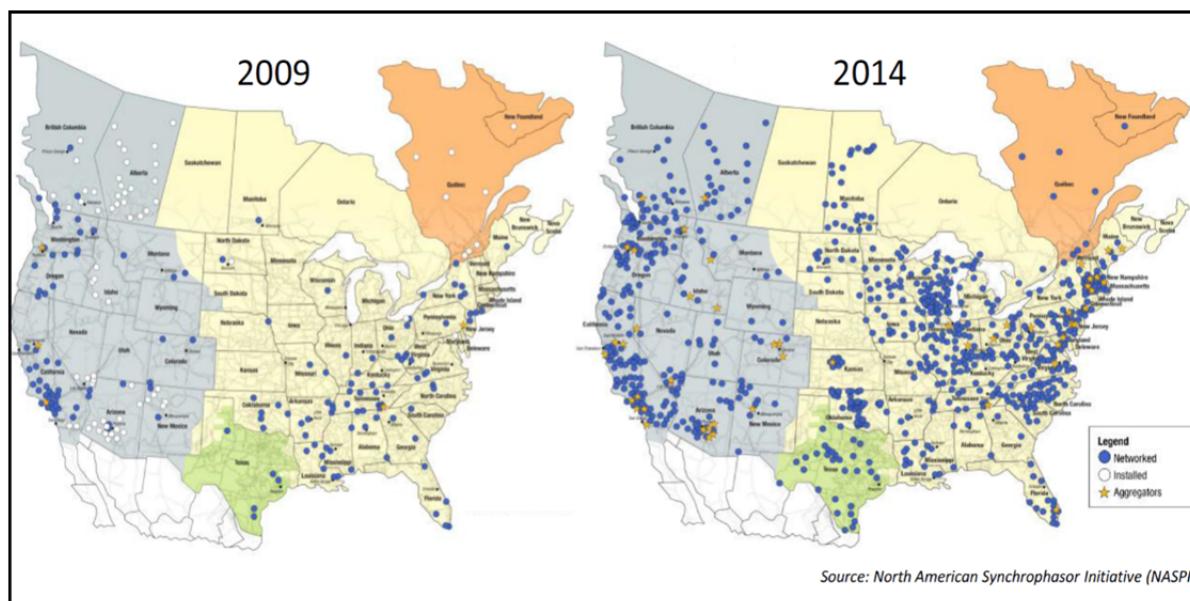


Figure 0.1 Le déploiement des synchrophaseurs à travers l'Amérique du Nord
Tirée de DOE (2014, p. 3)

Le travail élaboré dans le cadre de cette thèse se situe dans le même contexte exposé précédemment. Ainsi, notre contribution consiste à suggérer des solutions aux deux derniers points soulevés. À cet effet, et à défaut d'une base de données des synchrophaseurs, nous travaillons avec des données extraites de l'EMS/SCADA (Energy Management System), caractérisant le comportement du réseau de transport d'Hydro-Québec, sur une durée de quatre ans. Cette option de travailler avec des données réelles des réseaux électriques engendre plusieurs avantages. En particulier, elle permet de mesurer les performances réelles des méthodes proposées (Chengxi et al., 2014).

Tout au long de ce travail, lors de forage de données, nous utilisons le logiciel **R** (Gareth et al., 2013) pour entreprendre les différentes études d'apprentissage statistique envisagées. Nous faisons appel en premier lieu à l'algorithme des forêts aléatoires (FA). Cette méthode se montre très précise devant les jeux de données de formation d'une taille imposante. Plus particulièrement, avec des variables d'entrée en grand nombre (des centaines voire même des milliers), comparativement à d'autres méthodes comme démontré dans les travaux (Kamwa, Samantaray et Joos, 2010; 2012). L'avantage premier des forêts aléatoires est leur résistance aux variances et aux biais. Par conséquent, nous les utilisons pour la prévision des marges de sécurité et les transits actifs sur les interfaces clefs du réseau de transport et ceci dans le but final d'anticiper la prévision des limites de stabilité dynamique du réseau. Le second algorithme d'apprentissage statistique utilisé est l'algorithme des Modèles linéaires généralisés, *Generalized Linear Models* (GLM). Ils nous ont servi à modéliser les angles de phase de tension des barres du réseau de transport, définis comme prédicteurs de stabilité.

Ce manuscrit est structuré en six chapitres suivis par une conclusion, trois annexes et une bibliographie. Le premier chapitre est consacré à la revue de la littérature ayant trait à la sécurité des réseaux électriques. Un exposé assez exhaustif qui permet de situer l'originalité et la contribution de notre travail. Le sujet de recherche dont la problématique, les objectifs et la méthodologie, sont définis dans la dernière section de ce chapitre.

Le second chapitre présente le contexte global de l'analyse entreprise, en débutant par les idées de base de cette étude, et passant aux hypothèses, au réseau considéré et aux données de l'étude.

Le troisième chapitre aborde la thématique de forage de données. Il expose les méthodes d'apprentissage statistique répondant aux besoins de forage de données ainsi que leur mise en œuvre sous l'environnement **R**.

Le quatrième chapitre aborde le thème du sujet, à savoir les angles de phase de tension comme prédicteurs de stabilité, par une approche statistique. Une analyse des données à

partir des différentes représentations statistiques, reflétant le comportement du réseau de transport, est décrite d'une façon détaillée.

Le cinquième chapitre aborde la méthode des forêts aléatoires (FA) et son application pour la prévision des marges de sécurité et des transits actifs des corridors en fonction des prédicteurs de stabilité tels que les angles de phase de tension.

Le sixième chapitre est constitué de deux parties. En premier lieu, il propose des modèles élaborés de type GLM (modèles linéaires généralisés), pour la prévision des prédicteurs définis. Puis, une évaluation détaillée des résidus de ces modèles de prévision est scrutée. En second lieu, il présente le couplage de l'approche des FA avec l'approche de GLM pour prévoir les transits et les marges à base des angles de phase de tension estimés par prévision.

Enfin, une conclusion récapitule les principaux résultats établis, ainsi que des suggestions recommandées mettant l'accent sur les perspectives et les directions de recherche induites par les résultats obtenus dans le cadre de ce travail.

CHAPITRE 1

REVUE DE LITTÉRATURE

1.1 Introduction

La sécurité d'un réseau électrique se réfère au degré de risque de la capacité de celui-ci à survivre aux contingences sans interruption de service à la clientèle (Morison, Wang et Kundur, 2004). Autrement, elle se rapporte à la robustesse du réseau électrique face aux contingences et, par conséquent, dépend des conditions de fonctionnement ainsi que de la probabilité conditionnelle de contingences. Historiquement, assurer une exploitation sécuritaire (Secure) et économique du réseau électrique a toujours été d'une importance primordiale. Dans les années vingt, mené des outils d'analyse des réseaux électriques existant sous de nombreuses formes, le gestionnaire du réseau arrive à contrôler et assurer le bon fonctionnement du réseau. Peu de temps après, l'analyseur de réseau était largement utilisé et la simulation numérique était étudiée (Ackerman, 1930; Stagg et El-Abiad, 1968; Stedman, 1931). L'introduction formelle du concept de la sécurité dans le cadre de planification et d'exploitation des réseaux électriques a émergé en 1967 (Liacco, 1967). Les outils existants étaient efficaces à cette époque et les réseaux avaient tendances à être plus sécuritaires que dans ces dernières décennies, et ceci pour plusieurs nombre de raisons. Tout d'abord, les réseaux électriques étaient conçus, construits et exploités par des institutions (généralement des organismes gouvernementaux et des services publics). La planification intégrée assure que les installations de production et de transport suivent le rythme de la croissance de la charge, limitant ainsi le problème de surcharge et les défaillances des équipements qui causent des perturbations des réseaux. Les programmes d'entretien étaient en général aussi rigoureux. D'un point de vue opérationnel, la prévision des conditions de réseau était plus simple, vu qu'il y avait moins de propriétaires de production et de transport et qu'ils opéraient d'une manière soigneusement planifiée et concertée. Pour cette raison, les réseaux étaient exposés à moins de perturbations en comparaison avec ceux d'aujourd'hui. Ils étaient plus robustes dans leurs réponses aux perturbations et ils étaient plus prévisibles dans leurs modes de fonctionnement. Cependant, au cours des deux dernières décennies, l'évolution de

l'industrie d'énergie électrique vers l'ouverture des marchés a mis en place un certain nombre de facteurs qui ont augmenté les possibilités de perturbations du réseau. De ce fait, ces derniers ont réduit la robustesse des réseaux face aux contingences ainsi que la prévisibilité de son fonctionnement (Morison, Wang et Kundur, 2004). Dans ce nouvel environnement concurrentiel, des évaluations plus rigoureuses de la sécurité du réseau et des nouvelles exigences relatives aux capacités des outils d'analyse de réseau électrique deviennent nécessaires. Cette exigence de déterminer rapidement et avec plus de précision les limites de sécurité du réseau est d'une importance économique et une question de fiabilité. Pour s'assurer qu'un réseau électrique soit suffisamment fiable et que la probabilité de son bon fonctionnement à long terme soit acceptable, le réseau électrique doit être :

- conçu avec une considération primordiale pour la sécurité;
- suivi pendant son fonctionnement pour assurer en tout temps l'existence d'une marge de sécurité suffisante (même dans le cas où les conditions de fonctionnement ou les contingences sont différentes de celles planifiées).

Le succès de la réalisation de ces objectifs repose largement sur l'application de bonnes méthodes qui, à leur tour, font recours à l'utilisation de nouveaux outils pour une analyse rapide de l'état stationnaire et du comportement dynamique du réseau électrique après une contingence.

1.2 L'évaluation de l'analyse de la sécurité du réseau électrique

L'évaluation de la sécurité du réseau électrique est l'analyse effectuée pour déterminer si et dans quelle mesure le réseau, dans son fonctionnement, est sécurisé face aux différentes perturbations qui peuvent surgir. Les court-circuits électriques, les changements de configuration de réseau de transport dus aux défauts, la perte des composants du réseau, les actions de commutation de ligne et les augmentations de charge soudaines sont des exemples de ces perturbations. Pour la planification et l'exploitation, avoir un système sécurisé signifie qu'après une perturbation, le réseau électrique (1) survivra au transitoire survenu et se

déplacera vers un état d'équilibre acceptable, et (2) dans ce nouvel état d'équilibre, tous les composants fonctionneront dans les limites établies (Taylor, 1994). L'analyse utilisée pour la première condition est l'analyse transitoire TSA (*Transient security assessment*). L'analyse utilisée pour la deuxième condition est l'évaluation de la sécurité statique (SSA). En outre, le soutien de la tension est aussi une préoccupation majeure, d'où la nécessité d'une troisième analyse qui est l'évaluation de la sécurité de tension (VSA), bien que cette analyse chevauche souvent avec SSA et TSA (Sauer et al., 2004).

Typiquement, la SSA est effectuée en premier, suivi par la TSA. Traditionnellement, elles sont exécutées hors ligne en utilisant des procédés de simulation dans le domaine temporel pour récupérer la réponse du réseau suite à une contingence. La SSA évalue l'état d'équilibre post-contingent du réseau, en négligeant le comportement transitoire et d'autres variations temporelles. La VSA est basée principalement sur des méthodes statiques ou pseudo-dynamiques accompagnée de fonctionnalités en ligne. Le problème de la stabilité de tension est généralement associé à l'élévation de la charge dans les lignes de transport longues et l'insuffisance d'alimentation réactive locale. Une chute de tension progressive initiale, suivie d'un effondrement, caractérise ces types de phénomènes.

L'évaluation de la sécurité dynamique (*Dynamic security assessment*) DSA évalue la transition en fonction du temps de l'état précontingent à l'état postcontingent déterminant la stabilité du réseau pour les petites et les grandes perturbations. Les deux problèmes dynamiques, qui sont la stabilité transitoire et l'écoulement de tension, doivent être considérés lors de l'exécution d'une étude de sécurité dynamique. L'évaluation de la stabilité transitoire est la préoccupation majeure lors de l'analyse de DSA pour les réseaux électriques à plusieurs machines. Dans un tel réseau, un défaut ou une perte d'un grand générateur peut donner lieu à de grandes oscillations électromécaniques entre les unités de production, conduisant à une perte de synchronisme dans le système.

1.3 Les politiques de la fiabilité et la planification des opérations

Pour un réseau électrique réel, l'évaluation de la sécurité dynamique se réfère à l'analyse nécessaire pour déterminer si oui ou non un tel réseau peut répondre dans les délais aux critères de la fiabilité et de la sécurité spécifiques aux états d'équilibre et transitoire pour toutes les contingences crédibles.

Le réseau électrique actuel, connecté aux autres réseaux, est caractérisé par la réduction progressive de ses marges de sécurité suite à la pression des marchés. Dans ce contexte rempli d'incertitudes, pour maintenir sa fiabilité, des critères de fiabilité WECC (*Western Electricity Coordinating Council*) ont été établis afin de coordonner son fonctionnement. La sécurité du réseau dépend de la coopération des différentes entités interconnectées qui travaillent pour établir un environnement typique pour le fonctionnement du réseau. Pour cela, les entités doivent respecter les critères de fiabilité WECC. Ces lignes directrices sont les résultats des expériences sur le terrain et des nombreuses études sur le fonctionnement du réseau. Elles précisent les effets admissibles que peuvent avoir les réseaux voisins l'un sur l'autre. Les objectifs de ces critères soulignent le fait que chaque système de WECC doit, en premier lieu, assurer la continuité de service pour ses clients et, en second lieu, préserver le fonctionnement interconnecté au cours d'une perturbation. Les critères de fiabilité sont définis et mesurés en termes de niveau de performance d'un réseau dans des conditions de stress. Chaque critère est en fonction du type de contingence.

D'une manière générale, chaque étude de planification des opérations doit se préoccuper de la sécurité statique, la sécurité de tension et la sécurité dynamique. Pour un état hors service de tout équipement significatif et une condition de charge donnée, la réponse à toutes les contingences crédibles et majeures est étudiée. Cependant, la prévision de performance nécessite beaucoup de simulations parce que les tests effectués sur les systèmes existants ne sont pas exploitables. Pour la VSA, la méthode de PV (marge MW) ou la méthode VQ (marge MVar) est utilisée pour déterminer une marge de sécurité de tension adéquate (Taylor, 1994). Les marges admissibles et les critères de fiabilité associés sont basés sur les lignes directrices du WECC (Sauer et al., 2004). Les critères de stabilité de la tension

s'appliquent aux études des interfaces et des zones de charge. Les interfaces incluent les principaux corridors de WECC, les lignes d'interconnexion avec les réseaux voisins et les corridors critiques du réseau. Ces critères sont spécifiés en termes de marges de puissance active et réactive pour chaque niveau de perturbation. Les marges pour chaque scénario sont déterminées et documentées dans des nomogrammes ou des tables consultatives (Abed, 1999). Les nomogrammes illustrent graphiquement les limites du réseau pour certains paramètres de quelques scénarios, tels que les scénarios impliquant de grands transferts ou des pannes de matériel. L'opérateur va alors se baser sur cette information pour des décisions en temps réel.

Les tables consultatives ont un inconvénient évident de rigidité. Elles sont sujettes à des erreurs car les opérateurs doivent rechercher le scénario le plus pertinent dans les tables. Les nomogrammes, d'autre part, procurent un peu plus de souplesse car ils offrent des compromis dans les conditions de fonctionnement, par exemple entre une condition de charge et un transfert à travers une interface clé. Cependant, les nomogrammes ne parviennent pas à saisir pleinement toutes les informations contenues dans les études hors ligne et n'ont pas la capacité de gérer des situations plus variées. Cette approche pratique diffère sensiblement des approches décrites dans la littérature concernant les études de la sécurité en ligne qui se focalisent sur la sélection de contingences et les méthodes rapides de calcul de la sécurité. Dans la pratique, les limites de sécurité sont rassemblées hors ligne comme décrit ci-dessus. Les gestionnaires de réseau, plus ou moins en fonction de l'utilité, ont tendance à s'appuyer fortement sur les limites identifiées par les planificateurs des opérations et de faire un usage relativement limité des outils de sécurité en ligne.

1.4 L'évaluation de la sécurité dynamique en ligne dans la littérature scientifique

Les recherches sur la sécurité DSA en ligne ont généralement porté sur le déploiement des méthodes efficaces pour déterminer rapidement la stabilité du réseau, y compris la simulation en domaine temporel des indices de stabilité transitoire et les méthodes de fonction d'énergie (Wang et Morison, 2006). Ceci est en contraste frappant avec les longues études détaillées utilisées dans la planification opérationnelle. Les techniques classiques d'analyse de la

sécurité dynamique ont besoin de temps de calcul excessif. Par conséquent, ceux-ci sont généralement indésirables pour des fins en ligne.

La littérature scientifique n'a cessé d'exposer des algorithmes pour l'évaluation de la sécurité dynamique en ligne des réseaux électriques. On retrouve l'intégration numérique, la deuxième méthode de Lyapunov, les méthodes probabilistes, l'arbre de décision (*Decision Tree* ; DT), réseaux de neurones artificiels (*Artificial Neural Networks* ; ANNs) (Tiako, Jayaweera et Islam, 2011) et les forêts aléatoires.

1.4.1 Intégration numérique

Les algorithmes d'intégration numérique sont utilisés pour résoudre le système d'équations différentielles du premier ordre qui décrit la dynamique d'un modèle du réseau (Dommel et Sato, 1972). Les méthodes les plus utilisées sont les méthodes prédictives et prédicteur-correcteur de Runge-Kutta. L'intégration numérique fournit des solutions exactes se rapportant à la stabilité du réseau en fonction du détail des modèles utilisés. Ces approches sont largement utilisées dans les environnements hors ligne, mais pour leur exploitation en ligne nécessite des calculs fastidieux.

1.4.2 Méthode de Lyapunov

Cette approche est désignée comme la méthode de fonction d'énergie transitoire (TEF). L'idée consiste à remplacer l'intégration numérique par un critère de stabilité. Ce dernier peut être désigné par une fonction de Lyapunov V calculée à l'instant de la dernière commutation dans le réseau et comparée à la valeur critique V_{cr} déterminée auparavant. Si V est inférieur à V_{cr} , le processus transitoire post-défaut est stable. Cette méthode se caractérise par son efficacité concernant la capacité de calcul du degré de la stabilité et la vitesse de calcul, lorsque les variables du réseau sont simplifiées. Dans la pratique, il y a encore quelques inconvénients et des problèmes non résolus concernant cette approche (Bose, 1984), (Ribbens-Pavella et Evans, 1985).

1.4.3 Méthodes probabilistes

Avec ces méthodes, l'analyse de la stabilité est considérée comme un problème probabiliste car les facteurs de perturbation (type et l'emplacement de défaut) et l'état du réseau (la configuration et la charge) sont de nature probabiliste. Par conséquent, cette méthode vise à déterminer les distributions de probabilité pour la stabilité du réseau électrique. Elle évalue la probabilité pour que le réseau reste stable lorsqu'une contingence spécifique se produit. Afin d'avoir des résultats statistiquement significatifs, un temps important est nécessaire pour considérer un grand nombre de défauts à différents endroits et à différents régimes de compensation (Patton, 1974). Par conséquent, cette méthode se montre appropriée pour la planification mais inappropriée pour une application en ligne (Anderson et Bose, 1983). Ces méthodes seraient intéressantes pour des applications en ligne si elles sont combinées avec les techniques de reconnaissance des formes.

1.4.4 Méthodes Système expert

Dans cette approche, l'expertise est codée par un programme conçu à base des règles (Chen-Ching et Tomsovic, 1986). Un système expert est composé de deux parties, une base de connaissances ou de faits et un ensemble de règles d'inférence. Généralement, l'expertise de la base de connaissances provient des opérateurs qui possèdent une vaste expérience sur un réseau particulier et des informations obtenues par l'analyse de la stabilité hors ligne (Fouad, Vekataraman et Davis, 1991). Ces méthodes Système expert, comme celles basées sur la théorie des ensembles flous, ont surmontés les difficultés rencontrés par l'utilisation des anciennes méthodes d'évaluation de la sécurité dynamique (Kamwa, Pradhan et Joos, 2007; Samantaray et al., 2010). Leur principal avantage est qu'elles reflètent le fonctionnement réel des réseaux électriques, ce qui n'est pas le cas avec d'autres méthodes heuristiques basées seulement sur des expériences. L'inconvénient évident est qu'il est devenu de plus en plus difficile de comprendre les limites des réseaux caractérisés par un nombre historiquement élevé de transactions sous les conditions actuelles du marché.

1.4.5 Méthodes de reconnaissance des formes

Les méthodes de reconnaissance des formes sont SVM (*Support Vector Machine*) (Kalyani et Swarup, 2011), les réseaux de neurones ANNs (*Artificial Neural Networks*) et les arbres de décision DTs (*Decision trees*). L'objectif de ces méthodes consiste à établir une relation fonctionnelle entre les caractéristiques sélectionnées et l'état du réseau vis-à-vis de la limite de la région de stabilité (Hakim, 1992). Au cours des dernières années, il y a eu plusieurs méthodes, à base de données et de classement de patterns, introduites pour construire des bons estimateurs de limites de sécurité (El-Keib et Ma, 1995). L'idée essentielle est : 1) la génération hors-ligne d'un ensemble de caractéristiques (comme les flux de ligne, les charges, et les limites de générateur) représentatives d'états de fonctionnement stables et instables, appelé ensemble de formation ; 2) la détermination de la fonction de classificateur (règle de décision) en utilisant l'ensemble de formation (le jeu d'apprentissage) des patterns étiquetés. Cette fonction est utilisée pour classer l'état réel de fonctionnement face à une contingence donnée. Souvent, l'étape de la classification est mise en œuvre en utilisant ANNs (Alves da Silva, 2005; Liqiang et al., 2000). La formation d'un tel estimateur est prévue pour généraliser aux cas similaires non étudiés. Les méthodes de réseaux de neurones artificiels (ANN) semblent les plus prometteuses vu leur grande capacité de généralisation et leur rapidité d'exécution (Al-Masri et al., 2013; Heqing, Feng et Min, 2010; Kassabaliidis, El-Sharkawi et Marks, 2002; Kumar et al., 1991; Liqiang et al., 2000; Mansour et al., 1997; Pao et Sobajic, 1992; Sawhney et Jeyasurya, 2004; Sobajic et Pao, 1989). La principale limite de ces approches se résume dans leur incapacité de considérer pleinement les études spécifiques du réseau utilisées dans la pratique (topologie, niveau de charge, les différents états de fonctionnement). Ainsi, des difficultés surgissent dans la sélection des caractéristiques et le développement de jeu d'apprentissage, ce qui limite leur application en ligne.

Dans l'approche des arbres de décision, un grand nombre de points de fonctionnement avec leurs critères de stabilité correspondants, appelé ensemble d'apprentissage, est généré et classé hors ligne. À partir d'une liste d'attributs de candidats, les plus sensibles au comportement de la stabilité transitoire sont les plus pertinents et sont sélectionnés sur la base des informations qu'ils portent, en évaluant leur corrélation dans le but de classification

(Wehenkel, Van Cutsem et Ribbens-Pavella, 1989). L'évaluation de la sécurité en ligne implique un passage à travers l'arbre de décision pour une condition de fonctionnement donnée. L'outil fonctionne bien pour une application en ligne. Cependant, en raison de la mauvaise sélection des attributs de candidats, le pourcentage des conditions de fonctionnement insécurisées classées par erreur comme sécurisées est un risque pour la sécurité du réseau.

1.5 L'évaluation de la sécurité dynamique en ligne sur le plan opérationnel

Sur le plan opérationnel, dans presque tous les cas, les calculs actuels de la sécurité sont exécutés dans un environnement de fonctionnement où ils mettent seulement l'accent sur l'évaluation de la sécurité statique en ligne (Lei, Jianzhong et Haoen, 2012). Lorsque les problèmes de l'écroulement de tension ou de stabilité transitoire sont soupçonnés, l'analyse est transférée aux planificateurs des opérations pour une étude détaillée. Certains opérateurs sont guidés par des nomogrammes (abaques) saisonniers qui ont été classés selon les charges de corridors, ou en « lignes directrices de l'opérateur ». Dans plusieurs cas, le personnel des opérations réalisent des simulations couteuses dans le domaine temporel lors de la planification des opérations d'un jour en avance. Ce manque d'activité d'analyse de sécurité dynamique est dû à un manque d'outils efficaces et intégrés. Les problèmes d'évaluation de la DSA ne peuvent être résolus rapidement ou sans investissements économiques importants. En conséquence, les opérateurs, face à des défis croissants en matière de sécurité, sont à la recherche de solutions innovantes pour améliorer le fonctionnement du réseau en temps opportun et abordable. Les méthodes DSA ne sont pas encore pleinement opérationnelles en ligne, avec quelques approches testées et évaluées actuellement. Une telle solution est de prendre un snapshot de l'état du réseau, puis effectuer une évaluation complète de la sécurité en temps quasi réel, et fournir aux opérateurs lors de situations anormales des avertissements ainsi que de recommandations pour des mesures correctives.

1.5.1 Les critères de l'évaluation de la sécurité dynamique

Dans la pratique, l'évaluation de la sécurité dynamique repose sur un ensemble spécifique de contraintes à surveiller. L'identification de cet ensemble de critères est basée sur l'expérience,

la connaissance du réseau et le jugement des ingénieurs de planification et d'exploitation. (Fouad et al., 1988; Sauer et al., 2004). Ces critères de l'évaluation de la sécurité dynamique (DSA) incluent :

1. Le critère de la stabilité d'inertie : concerne principalement l'évolution des angles et des fréquences des machines ;
2. La variation de tension (chute ou hausse) au-delà du niveau de seuil et de durée spécifiés : ceci inclut séparément les pairs seuil/durée de la variation tension pour la haute tension, la basse tension, et les seuils des variations instantanées des maximum/minimum ;
3. Le critère des marges de relais. Celles-ci sont définies pour les conditions de pré-contingence et post-contingence. Si après une perturbation, la marge de relais est violée durant un laps de temps supérieur au temps maximum spécifié, le réseau est identifié comme non sécurisé ;
4. Le critère d'amortissement minimum : pour une liste de contingences désignée, si le réseau à l'état post-contingence présente des oscillations, elles doivent être amorties positivement (diminution d'amplitude).

En règle générale, l'objectif de la DSA est de déterminer quelles sont les contingences qui peuvent provoquer des violations des limites du réseau électrique ou de l'instabilité du réseau. Le but final est de générer les directives opérationnelles pour définir des zones d'exploitation sécurisées. La génération des lignes directrices d'exploitation comprend la sélection de contingences, la réalisation d'une étude détaillée de la stabilité et l'analyse des résultats des violations. De nouvelles méthodes pour l'étude de la DSA sont exploitées.

1.5.2 Réflexions sur les méthodes d'évaluation de la sécurité dynamique

Afin de combler l'écart entre les modalités pratiques employées pour déterminer les limites des interfaces du réseau électrique et les différentes méthodes proposées pour la sécurité en ligne, il est nécessaire d'examiner un certain nombre de facteurs :

1. Les méthodes utilisées lors de la planification ne sont pas en mesure d'identifier toutes les conditions de fonctionnement possibles qui peuvent se produire et sont généralement trop lentes à répéter en ligne lorsque les conditions non étudiées du réseau sont présentes;
2. Les opérateurs n'ont pas un accès complet à toutes les hypothèses détaillées et utilisées dans une étude hors ligne ni les études des cas traités. Ils ont accès seulement aux conclusions de ces dernières à savoir les limites de transfert réelles ;
3. Quelques méthodes de sécurité en ligne proposées sont des méthodes rapides pour déterminer la sécurité, mais ne sont pas aussi efficaces pour déterminer une limite pratique de l'exploitation. Les différentes méthodes de sécurité en ligne proposées fonctionnent bien dans certaines conditions mais pas dans d'autres situations qui ne sont pas étudiées ;
4. La plupart des méthodes d'évaluation de la sécurité en ligne ne considèrent pas les détails des exigences de fiabilité employés par les différents conseils régionaux ;
5. En pratique, l'évaluation de la sécurité du réseau a toujours un certain degré de considérations relatives au réseau concerné qui ne se prêtent pas à une analyse plus formelle ;
6. La plupart des opérateurs ne font pas confiance aux approches «boîte noire» à cause de manque de connaissance des techniques de calcul employées dans l'analyse de la sécurité du réseau.

1.6 La synergie entre les nouvelles technologies et le progrès de la DSA en ligne

1.6.1 La DSA en temps réel

Les systèmes spéciaux de protection (SPS) ont été conçus afin de détecter les conditions anormales et mener des actions correctives permettant d'atténuer les conséquences possibles et d'assurer une performance du réseau. Cependant, ces systèmes de protection sont généralement configurés pour opérer lorsque les conditions de fonctionnement spécifiques préétablies sont atteintes, et ils sont incapables d'opérer pour des contingences inconsidérées qui pourraient déclencher des événements en cascade. En vertu de ces considérations, la surveillance du réseau et le déclenchement de la protection doivent être ajustés en fonction de la progression des événements en temps réel.

Récemment, il a été avancé que la DSA en temps réel pourrait résoudre la vulnérabilité des méthodes proposées pour la DSA en ligne, telle que l'incertitude de la prévision des conditions d'exploitations futures dans un environnement concurrentiel. La DSA en ligne et en temps réel se base sur des mesures de la condition de fonctionnement courante du réseau obtenues à partir des snapshots. Elle améliore l'analyse de la sécurité pendant un temps très proche du réel et passe l'information aux systèmes de contrôle ou à l'opérateur afin de réagir si une contingence analysée s'avère potentiellement dangereuse. Comme cette approche effectue des analyses sur un snapshot de l'état réel, l'incertitude incorporée dans l'analyse hors ligne en utilisant la condition prédite est largement éliminée. Cette approche fournit un mécanisme de type radar qui balaie en permanence le réseau pour des problèmes potentiels qui pourraient résulter si une contingence de type (n-1) ou tout autre de type (n-x) se produise.

D'énormes progrès ont été réalisés ces dernières années vers la mise en œuvre de technologies pour la DSA en temps réel pour développer la fonctionnalité de la DSA en ligne (Wang, 2012). Le développement des logiciels sophistiqués, très automatisés, et fiables, combiné avec l'amélioration sans fin dans la performance du matériel informatique, a conduit à l'installation de DSA en temps réel dans un certain nombre de sites dans le monde, en Amérique du Nord, en Chine (Guangxi Electric Power Compagnie) (Powertech-Labs), etc.

D'une façon fiable et crédible, les contingences peuvent être triées avec n'importe quel critère de sécurité et montré aux opérateurs de réseau. Si une contingence est identifiée comme insécurisée, la DSA en ligne évaluera les mesures de contrôle d'urgence disponibles et suggérera aux opérateurs une action de contrôle appropriée. En outre, les conditions clés du réseau ainsi que les résultats de la DSA sont affichés sur le site intranet, permettant aux utilisateurs autorisés d'examiner ces informations en temps réel en utilisant le navigateur Web à n'importe quel endroit à l'intérieur de l'intranet.

1.6.2 Les synchrophaseurs (*PMUs*)

Le phaseur est un nombre complexe qui représente à la fois l'amplitude et l'angle de phase de la tension et du courant sinusoïdaux à un instant précis. Les synchrophaseurs sont des mesures synchronisées de certains paramètres du réseau électrique et qu'on peut avoir à partir des appareils de mesure de surveillance du réseau appelés *Phasor Measurement Units (PMUs)*. Les synchrophaseurs mesurent la tension, le courant ainsi que la fréquence et calculent les phaseurs. Cet ensemble de données synchronisées de l'état du réseau est appelé les données phaseurs. Chaque mesure de phaseur est antidiée par rapport au temps universel GPS (*Global Positioning System*) : lorsqu'une mesure phaseur est antidiée, elle est appelée un synchrophaseur (DOE, 2013; IEEE, 2006).

L'échantillonnage des synchrophaseurs est de 30 observations par seconde, comparé aux technologies, comme SCADA, qui mesure une fois toutes les deux à quatre secondes. Les opérations de base nécessaires pour le fonctionnement des synchrophaseurs, y compris les techniques de traitement du signal et de filtrage concernées, ont été décrites pour la première fois en 1983 (Phadke, Thorp et Adamiak, 1983). Le premier prototype de synchrophaseur a été développé en 1988 (Phadke, 2002). Peu de temps après que ces nouveaux appareils de mesure ont été décrits, l'application des synchrophaseurs à l'éveil situationnel (*situational awareness*) du réseau a commencé à être étudiée.

Des recherches approfondies sur l'application des informations des synchrophaseurs afin d'améliorer l'éveil situationnel du réseau ont été menées depuis leur introduction, y compris

les applications sur l'estimation d'état (Ming et al., 2006; Thorp, Phadke et Karimi, 1985; Weiqing, Vittal et Heydt, 2007), l'évaluation de la sécurité dynamique (Kai et al., 2007; Kamwa, Beland et McNabb, 2006; Khatib et al., 2004), et la visualisation (Klump, Wilson et Martin, 2005; Zhang, Hirsch et Lee, 2007).

Les données phaseurs sont utilisées pour entraîner deux principaux types d'applications : les applications en ligne et hors-ligne. Pour des applications en ligne (ou proche du temps réel) les données phaseurs sont utilisées pour l'exploitation des réseaux. Elles comprennent la visualisation des zones étendues (*wide-area*) accompagnées de l'éveil situationnel du réseau (FERC, 2009b), la surveillance des tensions, de fréquence ainsi que leurs tendances, la détection d'importants changements dans les conditions de réseau, la surveillance des oscillations du réseau, l'amélioration de l'estimation d'état, l'intégration des énergies renouvelables, l'évaluation de la dynamique et la gestion de la congestion, la restauration de pannes, et le contrôle de l'équipement automatisé.

Pour les applications hors-ligne, les données phaseurs ne sont pas utilisées pour l'exploitation mais plutôt pour analyser les événements et les pannes après leur survenance, valider les modèles et les améliorer (Kamwa, Samantaray et Joos, 2009). Elles comprennent la conception du cas de référence (*baselining*) des performances du réseau électrique, la validation des modèles (modèles statiques et dynamiques pour l'ensemble des interconnexions, pour les centrales électriques individuelles et les classes de centrales), l'analyse des événements, la caractérisation de la charge, la conception des systèmes de protection spéciaux et d'îlotage, la conception des limites de fonctionnement du réseau et des points alarmants pour soutenir les opérations dans la salle de contrôle en temps réel (Silverstein et Dagle, 2012).

1.6.3 La mesure à grande échelle

L'intérêt porté à l'intégration et l'implémentation des synchrophaseurs dans les réseaux électriques a augmenté depuis que le rapport d'enquête de panne US-Canada en 2004, a reconnu que beaucoup de grandes pannes d'Amérique du Nord ont été causées par

l'ignorance ou le manque de connaissance des opérateurs du réseau de l'état de celui-ci. En effet, l'utilisation de technologies de synchrophaseurs pour fournir une visibilité de la zone étendue du réseau en temps réel est l'une des recommandations soulignées par cette investigation. Récemment, NERC (*the North American Electric Reliability Corporation's (NERC's) 2008 Real-Time Tools Best Practices Task Force*) recommande que les outils opérationnels en temps réel devraient avoir des capacités à grande vitesse, à la fois en termes d'accès aux données et de traitement de ces dernières, afin d'assurer que les réseaux électriques futurs soient fiables (Baldick et al., 2008). Depuis, un nouveau procédé de surveillance des réseaux électriques basé sur les mesures de synchrophaseurs, appelé système de mesure à grande échelle (*Wide-Area Measurement Systems, WAMS*) a été lancé dans certains pays afin de surmonter les obstacles précédents par une meilleure couverture et supervision de tout le réseau (Jihun et al., 2013; Kamwa et Grondin, 2002).

Un système de mesure à grande échelle est une infrastructure de données phaseurs qui comprend généralement des synchrophaseurs, un système de collecte de données constitué de concentrateurs de données phaseurs PDCs (Phasor Data Concentrators) pour synchroniser et archiver les données, les réseaux de communication de différentes technologies et vitesses pour délivrer les données phaseurs des synchrophaseurs vers les PDCs et les unités de gestion afin de les traiter, les interpréter et les utiliser pour des fins en ligne et hors ligne (DOE, 2013).

1.6.4 L'utilisation des synchrophaseurs pour détecter la vulnérabilité d'un réseau électrique de transport

La vulnérabilité (ou le stress) d'un réseau électrique est l'un des premiers concepts introduits pour évaluer la sécurité dynamique des réseaux électriques. Elle fait référence au niveau de sécurité du réseau, exprimé par des indices de sécurité, et à sa tendance de muter vers des conditions de fonctionnement critiques (Kamwa, Beland et McNabb, 2006). Un réseau vulnérable fonctionne avec un niveau réduit de la sécurité qui le rend vulnérable aux effets cumulatifs d'une série de perturbations modérées (Fouad, Qin et Vittal, 1994). Une zone vulnérable est une section spécifique du réseau où la vulnérabilité commence à se

développer. La survenance d'une contingence dans de telles zones définit un système sur le point d'effondrement. Les zones vulnérables sont caractérisées par quatre différents symptômes de stress : l'instabilité de tension, l'instabilité angulaire, les écarts de fréquence en dehors des limites, et les surcharges (McGillis et al., 2006).

La plupart de ces applications en temps réel ont été focalisées seulement sur l'analyse des phénomènes électriques, ou la reconnaissance des signaux qui suggèrent un état d'équilibre non-sécurisé (*Insecure*) possible. Ce type d'évaluation de la vulnérabilité mène l'opérateur à prendre des mesures appropriées pour conduire le réseau à un état de fonctionnement plus sécurisé (contrôle préventif). Néanmoins, l'utilisation de synchrophaseurs a un grand potentiel d'évaluer la vulnérabilité post-contingence qui pourrait être utilisée pour déclencher les SPS pour mettre en œuvre des commandes correctives (Baldick et al., 2009; Kamwa et al., 2009).

Les analyses indiquent que le taux de variation de la différence d'angle de phase entre les postes du réseau de transport est un indicateur important de la croissance du stress du réseau électrique. L'augmentation de l'angle de phase ou un grand déphasage angulaire est utilisé comme la base d'alerte aux opérateurs de transport. Un moyen pour connaître l'état du réseau à base de synchrophaseurs est de montrer la tendance des angles de phase par rapport aux limites des angles de phase. Ainsi, les opérateurs sont avertis lorsque la contrainte augmente. Un tel outil offre des renseignements à l'opérateur du réseau électrique. Lorsque les angles de phase dépassent les limites critiques, les opérateurs peuvent effectuer des actions correctives (DOE, 2013; NASPI, 2009; Parashar et Mo, 2009; Patel et al, 2010; Vaiman et al., 2010).

1.7 Le processus de sécurité du réseau de transport d'HQ

1.7.1 Généralités

À son origine, le réseau de transport d'Hydro-Québec a été conçu selon un critère de performance peu sévère. Le niveau de fiabilité était jugé satisfaisant vis-à-vis : des événements les plus susceptibles de se produire, des attentes de la clientèle et des investissements à consentir. Au fil des ans, divers événements imprévisibles sont survenus,

causant des pannes majeures influant ainsi l'exploitation du réseau. Au début des années quatre-vingts, Hydro-Québec était amenée à réévaluer la fiabilité de son réseau. Au cours des études menées, elle a noté une évolution significative des attentes de sa clientèle en regard de sa dépendance à l'énergie électrique et de même une hausse importante des coûts économiques et sociaux reliés aux interruptions de service (Québec). Depuis, elle a décidé d'investir dans une plus grande continuité de service en élevant le niveau de fiabilité de son réseau de transport. De ce fait, elle s'est unie à l'organisme Northeast Power Coordinating Council (NPCC) et elle y est devenue un membre à part entière. Le NPCC est l'un des huit conseils régionaux du North American Electric Reliability Council (NERC), Figure 1.1. Comme le NPCC, le NERC a pour mission de promouvoir la fiabilité du service électrique fourni par les principaux réseaux de transport du continent. Son rôle consiste, par ailleurs, à établir des normes et encadrements de base en matière de planification, d'exploitation et de maintenance des réseaux de transport.

Le contexte de la stratégie suivie par TransÉnergie HQ pour assurer la sécurité du réseau de transport principal d'HQ est un ensemble de :

1. Limites de transits à respecter sur les grands axes du réseau et sur les installations à courant continu;
2. Actions relatives à l'exploitation des équipements et des grands automatismes.

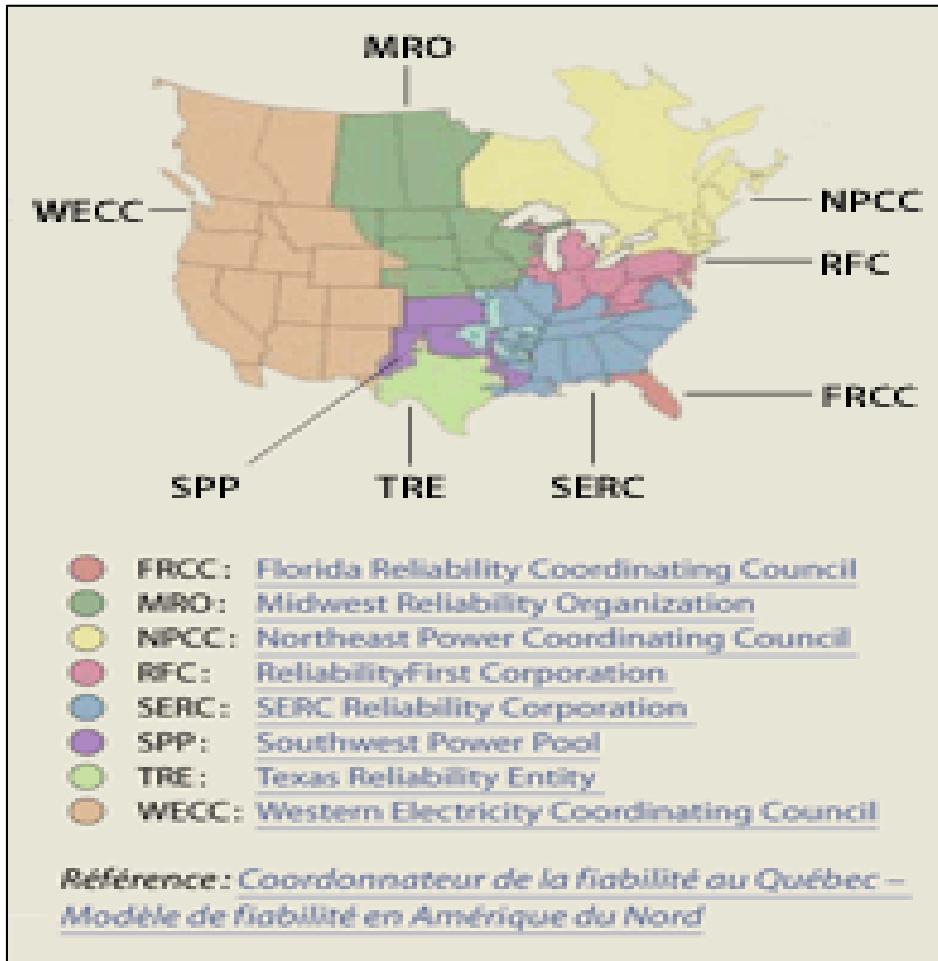


Figure 1.1 Les huit régions de NERC
Adaptée de NERC (2014, p. 4)

Ces stratégies visent à optimiser l'utilisation du réseau de transport et à couvrir une liste des configurations pouvant résulter de retraits d'équipements pour entretien et/ou d'indisponibilités imprévues.

Le réseau de transport d'HQ est exploité de manière à ce qu'il puisse supporter les événements les plus susceptibles de se produire en situation normale et en situation d'urgence. Ces événements sont spécifiés par le NPCC et TransÉnergie et ils sont pris en considération lors de l'établissement des limites de transit. Ainsi en tout temps suite à ces événements, la stabilité du réseau doit être maintenue et la tension soutenue, en appelant les

actions qui leurs sont autorisées comme la sollicitation des grands automatismes de sauvegarde du réseau (*Special Protection Systems SPSs*) tels que : le système Rejet de Production et Télé délestage de Charge (RPTC) ou *generation rejection and/or remote load-shedding system, under frequency load-shedding system (UFLS)*, Mancœuvre automatisée d'Inductances Shunt (MAIS) ou *shunt reactor automatic switching system*, Télé Délestage de charge Sous Tension (TDST) ou *under-voltage load-shedding system (UVLS)* et SPSR ou *permanent solution network separation* (Heniche, Kamwa et Dobrescu, 2013). Ces automatismes sont conçus et ajustés en ayant comme objectif de contrer les événements servant à fixer les limites de transit, et d'autres événements qui ne le sont pas. Dans ce dernier cas, toutes les actions et tous les automatismes sont permis pour assurer la stabilité du réseau. Malgré cela dans certaines conditions dégradées, certains de ces événements peuvent ne pas être couverts. Ainsi, aucune action ne sera prise puisque le répartiteur ne sera pas informé de cette situation.

Ces stratégies sont élaborées en mode différé. Elles sont ensuite intégrées au logiciel LIMSEL qui affiche les stratégies appropriées aux situations vécues en temps réel ou en prévisionnel (Loud et al., 2010).

1.7.2 Le logiciel LIMSEL (LIMit SELection)

À partir des études de sécurité dynamiques complétées, les résultats sont présentés sous forme de tables et sont intégrées dans le système LIMSEL. LIMSEL est un ensemble de tables fournissant les limites de transits selon les conditions du réseau. Ces tables donnent tous les seuils de transits associés aux différentes topologies du réseau et à l'indisponibilité d'autres équipements stratégiques tels les compensateurs synchrones ou statiques, etc. Ce système fournit également un résumé des éléments sensibles et leur impact sur les limites ou les stratégies d'exploitation. Ce logiciel a pour but de faciliter la mise en œuvre et le suivi en temps réel des stratégies et limites de sécurité. C'est un système de sécurité en temps réel efficace pour la surveillance et le contrôle de la fiabilité du réseau électrique (Guillon, 2005).

Au fil des années, le réseau d'Hydro-Québec a connu de l'expansion : l'ajout d'éléments nouveaux - lignes de transport et lignes d'attache entre les différents corridors, compensateurs, bancs de compensation série, lien multi-terminal à courant continu et automatismes de sauvegarde-. Face à cette évolution constante du réseau, ce logiciel a connu une refonte complète axée sur la facilité et la rapidité d'implantation des nouvelles stratégies pour tenir compte d'un nombre plus élevé de configurations et de plus de points de surveillance de transits (Huang et al., 2012; Valette et al., 2008). Ce logiciel effectue d'autres tâches, en plus de ses fonctions premières de sélection des limites dans les corridors de transport, les centrales et compensateurs.

[...] le logiciel LIMSEL déclenche immédiatement une alarme s'il relève une configuration de réseau interdite et émet un message à l'intention du répartiteur pour l'informer des mesures à prendre pour y remédier. [...] LIMSEL transmet en tout temps les consignes de rejets à mettre en œuvre en cas de contingences éventuelles. [...] LIMSEL maintient un lien direct avec le programme d'analyse de contingences de l'estimateur d'état, afin de repérer toute surcharge des bancs de compensation série ou des transformateurs lors des pertes de lignes. (Doyle et al., 2004, p. 6)

Actuellement, LIMSEL n'est pas seulement un outil majeur d'HQ pour les opérateurs dans les fonctions de gestion en temps réel. Il est aussi un outil pour les ingénieurs dans leurs activités de planification telles que l'étude des modes de fonctionnement futurs, planification des programmes de maintenance et planification des échanges avec les réseaux voisins (Valette et al., 2009).

1.8 Présentation de la problématique

L'évaluation actuelle de la sécurité dynamique du réseau, comme on l'a vu dans les sections précédentes, s'effectue par différents logiciels de simulation. Cette conception est la doctrine du gestionnaire de réseau depuis la révolution des processeurs de calculs. En outre, depuis le développement des synchrophaseurs, le gestionnaire a pris conscience de la nécessité d'étendre l'utilisation de cette nouvelle technologie à la surveillance de l'état du réseau de transport électrique vu la précision, la latence et la visibilité de ses mesures. Cependant, ces capteurs avancés donnent accès à de trop grands volumes d'informations de haute qualité sur

l'état présent du réseau. Ces données peuvent pourtant permettre à l'opérateur d'améliorer la fiabilité du réseau de transport. Devant cette nouvelle avalanche de données, désormais le logiciel de simulation qui est nécessaire pour l'évaluation de la sécurité devient insuffisant. En effet, cette nouvelle technologie de synchrophaseur exige de nouvelles méthodes et de nouveaux outils afin de réunir le grand volume de données et de les présenter d'une manière utile aux opérateurs, pour l'analyse, la surveillance et le contrôle de la sécurité. Ainsi, changer le paradigme d'évaluation de la sécurité dynamique DSA de la simulation vers les mesures est indispensable pour améliorer la fiabilité et assurer la sécurité du réseau.

La problématique posée est comment tirer parti de cette technologie pour habiliter cette évolution et comment utiliser plus efficacement cette ressource de données.

1.8.1 Objectif

Notre objectif est de suggérer une démarche systématique qui répond à une question fondamentale à savoir ; comment convertir ces mesures de synchrophaseurs, si précises et si abondantes, en informations sur lesquelles l'opérateur peut se baser, dans les conditions normales et les conditions d'urgence, pour surveiller le fonctionnement du réseau de transport et assurer la sécurité de ce dernier.

Nos objectifs spécifiques considèrent :

1. L'investigation des données de notre base de données;
2. L'identification et la légitimation des écarts angulaires comme prédicteurs de la stabilité dynamique du réseau de transport;
3. La modélisation en fonction des prédicteurs, par les forêts aléatoires, dans le but de prédire les transits de puissance sur les couloirs du réseau de transport d'Hydro-Québec;

4. La mise au point d'une approche de prévision des écarts angulaires comme des séries chronologiques similaires à la demande;
5. Enfin, la prévision en temps réel des transits de puissance et de leurs limites en fonction des nouveaux prédicteurs, sans avoir à les évaluer par des logiciels de DSA.

À long terme, ces outils pourront permettre certainement à opérer le réseau automatiquement, ce qui contribuera à l'amélioration des conditions d'exploitation.

1.8.2 La méthodologie

Pour atteindre ces objectifs, nous avons procédé en suivant les étapes ci-dessous :

1. Première étape : créer ou obtenir des données représentatives du problème à partir des données réelles d'exploitation du réseau, tirées de l'EMS/SCADA;
2. Deuxième étape : Explorer les données de base de ces séries chronologiques réelles s'échelonnant sur 4 ans à la minute, en utilisant une approche statistique;
3. Troisième étape : Mise au point de modèles prédictifs permettant d'utiliser les écarts angulaires pour prédire les transits de puissance et les limites de stabilité dynamique associées, en utilisant des forêts aléatoires (FA);
4. Quatrième étape : Mise au point des modèles des prédicteurs de stabilité, en utilisant les modèles linéaires généralisés (GLM);
5. Cinquième étape : Prédire en temps réel et avec beaucoup d'anticipation les limites de stabilité dynamique par le couplage des deux modèles, celui de FA avec celui de GLM.

1.8.3 Notre contribution

1. Exploitation des techniques de fouille de données pour proposer une solution à la problématique de l'évaluation rapide des limites de stabilité dynamique en temps réel :
 - A. Utilisation des données réelles, sous forme de séries chronologiques, du réseau de transport d'HQ, de l'année 2010 à 2013, pour la conception et la validation de notre méthode;
 - B. Identification des écarts angulaires comme prédicteurs de transit de puissance ainsi que des limites de stabilité dynamique associées.
2. Conception d'une nouvelle approche reflétant le comportement dynamique réel du réseau de transport à partir des données de synchrophaseurs et bonification de l'approche conventionnelle de surveillance de la sécurité dynamique par simulation avec l'apport des mesures synchronisées.

CHAPITRE 2

LE CONTEXTE DE L'ÉTUDE ÉLABORÉE

2.1 Introduction

Le présent chapitre regroupe les idées de base qui ont contribué à l'élaboration de l'approche proposée dans ce travail. Le but de celle-ci est de refléter le comportement dynamique du réseau de transport d'Hydro-Québec, à partir des mesures des synchrophaseurs. Cette approche met en évidence la pertinence et l'efficience des écarts angulaires instantanés des tensions au niveau des barres pour la prévision de la stabilité dynamique du réseau de transport. Cependant, cette notion n'est pas suffisamment justifiée d'un point de vue théorique, par contre elle est soutenue fortement par les chercheurs et éprouvée par les gestionnaires des réseaux. La section 2 est consacrée à la description du réseau étudié. Les informations relatées concernant ce dernier sont tirées du site <http://www.hydroquebec.com/publications/en/corporate-documents/annual-report.html> et des références citées. Dans la section 3, nous mettons en exergue les références qui approuvent cette méthodologie d'approche. Pour la mise en œuvre de cette dernière, nous avons utilisé l'analyse par le forage de données qui repose sur le traitement d'une base de données qui regroupe les observations de toutes les variables explicatives qui définissent la problématique. Dans notre cas, nous avons exploité la base de données SCADA/EMS d'HQ comme alternative à l'inexistence de celle des synchrophaseurs. À travers la section 5, nous abordons l'intérêt d'acquérir la base de données des synchrophaseurs et la raison pour laquelle nous avons travaillé avec les données de SCADA/EMS. Les sections 7 et 8 résument la démarche suivie dans la construction d'une base de données propre à la mise au point de notre application.

2.2 Description du réseau électrique d'HQ

Le réseau électrique de transport d'Hydro-Québec (HQ) est l'un des plus complexes (spécifiques) dans le monde. Sa spécificité ou sa complexité réside dans la superficie qu'il couvre et aussi dans les conditions climatiques auxquelles il est soumis. Ces mêmes

contraintes qui le rendent complexes ont joué un rôle capital dans son développement et sa modernisation.

Le réseau de transport d'HQ, comme il est montré à la Figure 2.1, est constitué de grands couloirs à 735kV couvrant la majeure partie du territoire québécois. La production locale d'Hydro-Québec est caractérisée par sa mixte production hydraulique, thermique et éolienne. Ce réseau assure en premier lieu l'acheminement de la totalité de cette production des centrales vers des centres de consommation et en second lieu les échanges de puissance avec les réseaux voisins. Les plus importantes centrales de production sont situées au nord du Québec et elles alimentent une charge localisée à plus de 1000 km au sud, dont la charge de pointe est supérieure à 37000 MW (en janvier 2013). Le réseau de transport principal d'HQ est composé de plusieurs lignes à 735 kV et d'une ligne à 765kV, réparties sur deux axes nord-sud reliant les grandes centrales de production éloignées : Manicouagan, Churchill Falls et Baie James, aux centres de consommation situés dans la vallée du Saint-Laurent entre les villes de Montréal et Québec. De plus, en parallèle avec l'axe Baie James à courant alternatif, le réseau dispose d'un lien multi-bornes à courant continu exploité à 450 kV servant aussi de lien d'interconnexion avec la nouvelle Angleterre (Trudel, Gingras et Pierre, 2005). Le réseau de transport d'Hydro-Québec n'est pas synchronisé avec les réseaux voisins tels que les réseaux de l'Ontario, du Nouveau-Brunswick et des États-Unis, ce qui fait qu'il est moins vulnérable aux perturbations de ces derniers. Les échanges d'électricité avec ces réseaux voisins se font par l'intermédiaire d'interconnexions à courant continu. Le réseau de transport comprend également des lignes exploitées à des niveaux de tension allant de 44 kV à 345 kV. En tout, le réseau comporte plus de 32 000 kilomètres de lignes et plus de 500 postes de sectionnement, de transformation ou d'élévation de tension aux centrales (Doyle et al., 2004; Huang et al., 2012; Loud et al., 2010).

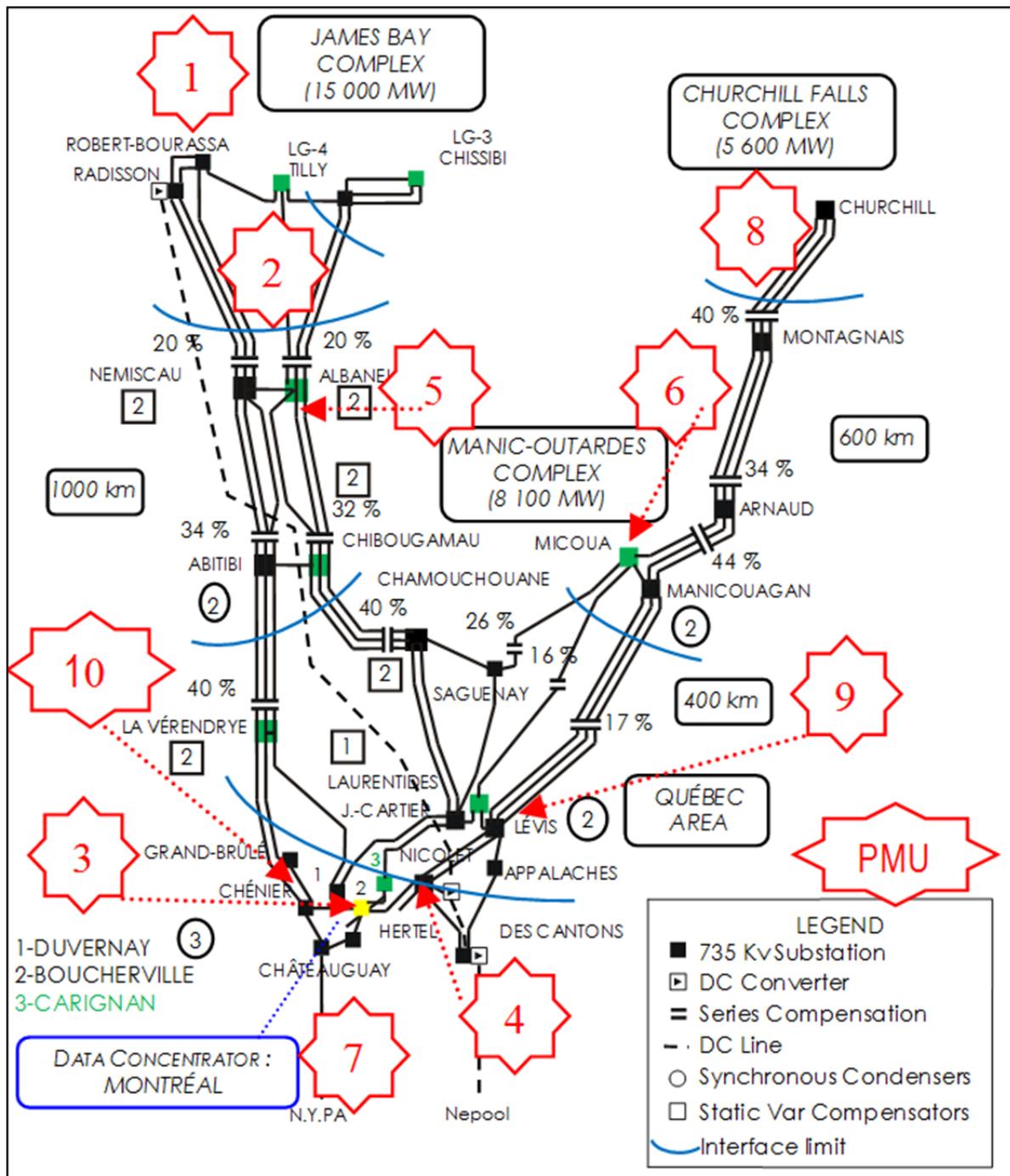


Figure 2.1 Le réseau de transport d'Hydro-Québec TransÉnergie

2.3 Les écarts angulaires en tant que prédicteurs de la sécurité dynamique du réseau électrique de transport

Dans le contexte de la sécurité dynamique de fonctionnement du réseau électrique de transport, l'éveil situationnel (*Situational Awareness*) signifie être constamment au courant de l'état d'évolution des conditions de fonctionnement du réseau électrique. Autrement, comprendre l'environnement courant et être en mesure d'anticiper avec précision les problèmes futurs pour permettre des actions efficaces (Giri et al., 2012; Jun et al., 2011). Cette fonction en temps réel à base des synchrophasateurs (*Syncrophasor-based Situational Awareness (SBSA)*), exploite la perspicacité de ces données qui est due à la vitesse élevée d'échantillonnage et à la nouvelle granularité des angles de phase. Même si cette approche n'est pas suffisamment maîtrisée, elle est recommandée par le FERC (FERC, 2009a) pour le réseau de transport nord-américain et par UCTE pour le réseau de transport européen. On dénote cet intérêt dans plusieurs travaux comme dans (DOE, 2013; NASPI, 2009; Parashar et Mo, 2009; Patel et al, 2010; Vaiman et al., 2010) où les chercheurs sont conscients que le développement de *Wide-Area-Visualization* en temps réel contribue beaucoup à l'amélioration de l'éveil situationnel du réseau. Cette clairvoyance est perçue lors des analyses et des études de différentes pannes que les réseaux de transport ont connues. Patel (Patel et al, 2010) a souligné que le manque *Wide-Area-Visualization* empêchait l'identification précoce de la panne d'électricité du Nord du 14 Août 2003. Le rapport d'enquête sur la panne des États-Unis-Canada émit l'hypothèse que si un système de phasateurs était en service à ce moment-là, les préconditions du blackout, en particulier les problèmes de croissance de tension dans l'Ohio et aussi la divergence significative de l'angle de phase entre Cleveland et le Michigan dans les dernières minutes avant la cascade, auraient été identifiés et compris tôt dans la journée (NERC, 2004). Subséquemment, l'information rapportée par le déphasage angulaire entre deux zones (*phase angle separation*) pour évaluer la fiabilité du réseau est bien documentée (Allen, Santoso et Muljadi, 2013 ; Parashar et Jianzhong, 2009; Samantaray, Kamwa et Joos, 2010). Dans certains travaux qui rapportent la manifestation de cette grandeur, on retrouve des expressions telles que :

Angle differences are indicative of the steady state stress of the system. The angle difference across a transmission corridor increases either because of

increasing power flow or because of weakening of the corridor through loss of transmission lines. Angle difference is in some cases a better measure of the corridor's capability than power flow because it is related to both the steady state and dynamic limitations of the corridor. (Giri et al., 2012, p. 29)

Toujours dans le même contexte et en connaissance de cause, en plus de la stratégie de *Wide-Area-Visualization*, les chercheurs de PITT (Planning Implementation Task Team) et OITT (Operational Implementation Task Team) de NASPI ont accordé une priorité plus élevée à la calibration statistique (*baselining*) des angles de phase (NASPI, 2009). En effet, vu ce lien incontestable perçu via les mesures entre les déphasages angulaires et la dynamique du réseau électrique et, en même temps qui n'est pas formalisé par un théorème avéré, la calibration statistique s'avère le meilleur procédé à entreprendre pour traiter cette notion de lien. Ainsi, des travaux dans cet axe sont amorcés pour transplanter cette approche dans la salle de contrôle (Parashar et Jianzhong, 2009) comme dans le cas de mesure de stress et de la capacité de charge d'une ligne par les variations d'angles (Gutman, 1988; Khan, 2008; Reinhard, Sauer et Dominguez-Garcia, 2013; Sauer, 2011; Vaiman et al., 2010). Parmi ces travaux, on cite les deux références (Ferryman et Amidan, 2012; Patel, 2011), où une utilisation simple des différences d'angle entre deux régions pourrait consister en comparant des valeurs mesurées à partir de PDC avec des seuils résultants des études des opérations de planification et des référentiels historiques des modèles des écarts angulaires.

2.4 Originalité de notre travail

Notre travail suggère une démarche qui se place au-delà des moyens standards utilisés pour l'évaluation de la sécurité des réseaux de transport d'HQ. Notre but principal consiste à saisir l'importance et le rôle des déphasages angulaires dans l'évaluation de la sécurité dynamique des réseaux électriques, à savoir ce que cette mesure nous révèle sur le comportement dynamique du réseau.

Notre travail se distingue par rapport à ceux déjà réalisés par :

1. En plus d'utiliser des données réelles, nous étudions le réseau de transport d'HQ en son état normal, sur une période de trois ans. En effet, beaucoup de travaux à base de données des synchrophaseurs étaient déployés pour l'analyse du réseau après des perturbations contrairement aux efforts investis dans l'interprétation des informations des phaseurs dans l'état normal du réseau. Parmi les rares travaux, on cite les références (Ferryman et Amidan, 2012; Patel, 2011; Sauer, 2011; Shrestha, 2012), qui ont abordé cette thématique qui est pourtant très importante pour deux raisons : 1) elle est la situation la plus courante adressée par un gestionnaire de réseau. 2) une portion importante des blackouts à l'échelle mondiale s'est passée lors d'un fonctionnement normal du réseau, loin des moments de stress où la demande électrique est forte (Lu, 2009) ;
2. L'application d'approche de cas de référence (*baseline*) qui combine des statistiques descriptives classiques avec le datamining prédictif pour évaluer l'état de la stabilité ($n-1$) du réseau électrique ;
3. Dans le travail (Allen, Santoso et Grady, 2010), les auteurs établissent un exemple de lien entre la production et le déphasage angulaire entre deux barres. Dans notre cas, nous établissons le lien entre les transits à travers cinq couloirs et 37 barres ;
4. Dans le travail (Vaiman et al., 2010), les auteurs se sont intéressés aux limites de stabilité (thermique et stabilité de tension). Notre étude s'est basée sur les limites dynamiques ;
5. En plus d'un objectif et d'une démarche similaires à ceux de (Ferryman et Amidan, 2012; Patel, 2011) dans la première phase de notre travail, nous avons utilisé, dans la seconde phase, les forêts aléatoires pour la prévision.

2.5 Base de données des synchrophaseurs

Depuis la reconnaissance des synchrophaseurs comme une source de mesures de haute qualité, également en résolution et en latence, comme détaillé dans (Patel et al, 2010), l'installation de cette technologie est en voie de progression continue. Les services publics et les organismes gouvernementaux ont massivement déployé les synchrophaseurs pour des applications EMS et le développement des systèmes de SBSA (Synchrophasor-based situational awareness) qui s'intéressent à l'évaluation de la marge de fiabilité des éléments et des zones du réseau. Aux États-Unis par exemple, le nombre des synchrophaseurs installés est passé de 166 à 1 043, seulement de 2010 à 2014, Selon (DOE, 2013). Par conséquent, le stockage à long terme des données de mesures de ces appareils, leur récupération et leur traitement pour les interpréter constituent une procédure qui s'impose d'elle-même vu que, le but ultime est d'exploiter ces données de synchrophaseurs dans des mises en œuvre réelles (Basu et al., 2014).

De même, Hydro-Québec s'est intéressée et s'est donnée la mission de développer son propre système de surveillance en zone étendue (Kamwa et al., 2006). Le système utilisé actuellement pour une supervision centralisée du comportement dynamique de son réseau de transport est le système de mesure de décalage angulaire (SMDA). Il est constitué de dix synchrophaseurs répartis aux points clés du réseau à 735kV et d'une unité centrale d'acquisition et de traitement de données (Data Concentrator), comme le montre la Figure 2.1. Les mesures des synchrophaseurs sont acheminées à cette unité, puis traitées, coordonnées et transférées pour différentes applications, (Basu et al., 2014). Actuellement, la base de données des synchrophaseurs est en voie d'élaboration et de restitution. En conséquence, elle a un historique limité en raison du nombre limité d'appareils de mesure et d'autres problèmes de technologies récentes.

2.6 SCADA vs. synchrophaseur

Pour une meilleure compréhension du fonctionnement du réseau, la qualité de mesures est d'une importance suprême dans la modélisation, la simulation et la prévision : « *The wealth*

of operational data provides significant opportunity to improve our models and understanding of the grid. ». (DOE, 2011, p. 34)

Avec l'avancement de la technologie des synchrophaseurs, la comparaison entre les mesures de SCADA et les mesures de synchrophaseur est très vite établie : « *Phasor measurement units (PMUs) with data sampling rates in excess of 30 Hz, a 100x improvement of over legacy supervisory control and data acquisition systems, will allow users to observe and analyze the bulk power system in real-time.* ». (DOE, 2011, p. 33)

Pour le système SCADA:

The SCADA system, which monitors current flow through lines, transformers, and other components into delivery points, typically makes measurements every 2 to 6 seconds. While that sampling rate has proven adequate for steady-state operations, it is insufficient to understand (or even detect) the details of transient phenomena with timescales of milliseconds (one 60 Hz cycle is 16 milliseconds). (DOE, 2012, p. 148)

Alors que pour les synchrophaseurs:

Synchrophasor technologies and systems use monitoring devices called phasor measurement units (PMUs) that measure the instantaneous voltage, current, and frequency at specific locations in an electric power transmission system (or grid). The sampling of these parameters takes place 20 or more times per electrical cycle which is 1200 or more times per second. PMUs convert the measured parameters into phasor values, typically 30 or more values per second. The PMUs also add a precise time stamp (using a well-defined format known as IEEE C37.118) to these phasor values turning them into synchrophasors. (DOE, 2013, p. 1)

Bien que la qualité de données des synchrophaseurs prenne le dessus sur celle de SCADA, la préservation de la source de données SCADA est préconisée. Par mesure de sécurité du réseau, les politiques de NERC (Patel et al, 2010) et de FERC (FERC, 2009b) ont recommandé de développer et de mettre en œuvre le WASA à base des synchrophaseurs séparément de SCADA. En effet, l'acquisition de différentes sources de données assurera

toujours la continuité de la fonction de surveillance du réseau si l'une d'elle est endommagée. Effectivement, dans le cas de la panne de l'Est en 2003, où il y avait une perturbation du processus SCADA (Doyle et al., 2004), une seconde base de données aurait été utile pour assurer le rôle de ce dernier. Les deux sources de données sont des alternatives, l'une à l'autre, sur lesquelles le WASA se base. Cela sous-entend que leurs données concordent, malgré que l'une soit plus volumineuse que l'autre, en raison de la faible latence des synchrophaseurs.

Dans notre cas, avec l'utilisation de système de mesure à grande échelle illustré par la Figure 2.1, cette équivalence entre les deux sources nous permet de surmonter l'entrave rencontrée en travaillant avec les données de la SMDA, provenant des synchrophaseurs installés, en raison du manque de données enregistrées. Donc, l'alternative offerte est d'utiliser les données historiques d'EMS/SCADA. En effet, nous avons élaboré notre travail sur l'hypothèse que les écarts angulaires fournis par l'estimateur d'état et l'interface CILEX sont en bon accord avec les données de synchrophaseurs. Des résultats similaires ont été rapportés par PJM (Patel, 2011) et ERCOT (Shrestha, 2012).

2.7 Base de données EMS/SCADA

Les logiciels CILEX et LIMSEL EMS d'Hydro-Québec (Doyle et al., 2004; Huang et al., 2002; Valette et al., 2009) assurent conjointement un accès à une énorme base de données instantanées en temps réel du réseau électrique (flux de charge et données dynamiques), minute par minute et archivée sur plusieurs années. Ces données ont été acquises au cours de la mise au point d'un nouveau simulateur de commande du réseau électrique pour l'étude de l'intégration des éoliennes (De Montigny et al., 2013). Il contient, pour cinq ans civils, les tensions et les déphasages angulaires au niveau des barres à 735 kV, sept limites d'interface avec les transits actifs et réactifs correspondants, charge de la zone d'équilibrage, ses prévisions en temps réel, les réserves d'exploitation, AGC, la production de la compensation réactive, et autres.

L'idée est de construire une base de données propre à l'étude entreprise. Elle est constituée des séries chronologiques de toutes les grandeurs physiques dont on a besoin, à savoir : les angles de phase de tension de toutes les barres à 735kV, les limites de sécurité dynamique des couloirs principaux du réseau, les transits actifs le long de ces couloirs et la charge. Les limites de transfert dynamiques considérées sont celles définies dans (Valette et al., 2009). Elles sont des fonctions complexes, dépendantes de plusieurs configurations du réseau électrique ainsi que de la charge. Elles sont calculées hors lignes, des mois en avance.

2.8 Prétraitement des données

Les années 2010, 2011 et 2012 ont été choisies pour cette étude. Toutes les variables ont été initialement rangées avec un écoulement de puissance convergent puis adaptées en agrégation en minute par minute exigé de l'EMS/SCADA (De Montigny et al., 2013; Kamwa et al., 2012). Lorsqu'une charge convergente est manquante pour une minute spécifique, toutes les variables leurs sont attribuées un NA et traitées comme des valeurs manquantes. Pour réduire le volume de données dans certaines situations, où on y fera référence au fur et à mesure tout au long de ce document, ces données ont été filtrées en sous-échantillons de fréquence d'échantillonnage de 3 min, ce qui entraîne 175 000 échantillons sur un an.

Dans notre étude, nous avons utilisé les variables suivantes : les écarts angulaires des tensions, par rapport au poste Boucherville, des 37 postes de à 735 KV du réseau de transport, les limites de transfert dynamiques des cinq couloirs du réseau, à savoir le couloir Baie-James Nord (BJN), Baie-James Sud (BJS), Churchill (CHU), Manic-Québec (MQ) et Sud, les transits actifs le long de ces cinq corridors et la charge prédictive une heure en avance. Nous avons aussi introduit cinq nouvelles variables pour chaque couloir qui sont les marges de limites de transfert (c'est-à-dire la différence entre la limite LIMSEL et le transfert présent). Notre base de données est constituée de 54 séries chronologiques. Autrement dit, c'est une base de données de dimension (1 578 240 observations x 54 variables) comme le montre la Figure 2.2.

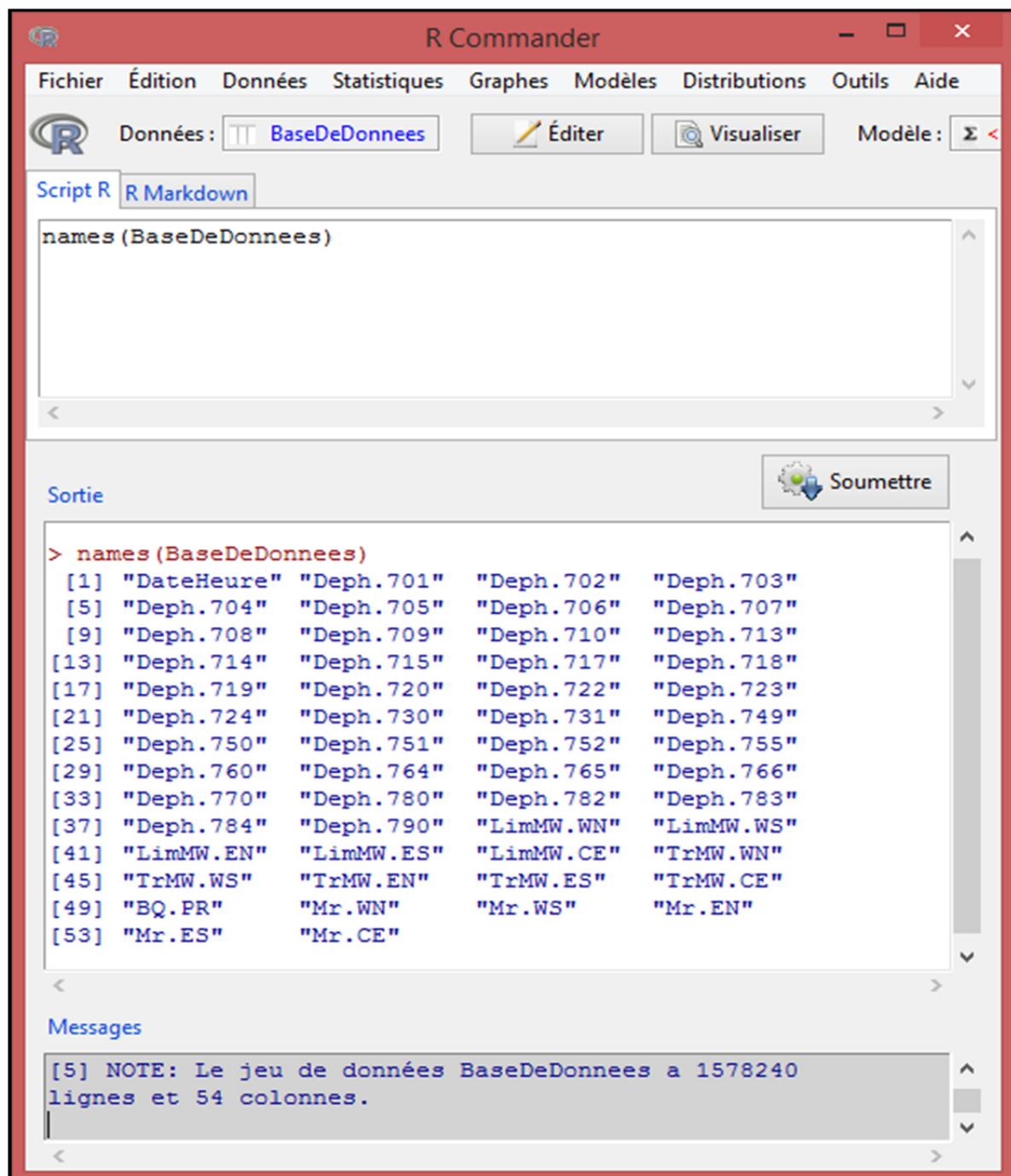


Figure 2.2 Base de données utilisée pour l'étude

CHAPITRE 3

LE FORAGE DE DONNÉES

3.1 Introduction

Tout le long des chapitres suivants, on parle de la variable cible et des variables explicatives. La variable cible est la variable dont on cherche à connaître la valeur. Dans la littérature, on lui attribue aussi la nomination de variable à expliquer, variable réponse, variable dépendante, variable endogène. C'est la variable « en sortie ». Les variables explicatives sont les variables utilisées pour construire le modèle. On parle également de variables prédictives ou de prédicteurs et de variables indépendantes. Ce sont les variables « en entrée ».

Au cours de ce chapitre, nous introduisons un bref historique de la locution du forage de données qui constitue une nouvelle approche pour résoudre les problèmes, suivi par une section où, on aborde le concept de l'apprentissage statistique. Dans la section 4, nous présentons les outils et le matériel informatique utilisés pour atteindre nos objectifs de recherche. Puis dans les sections subséquentes à cette dernière, nous introduisons les quatre méthodes utilisées dans cette thèse basées sur le forage de données. Nous concluons ce chapitre par un commentaire et une conclusion se rapportant directement à la méthode des forêts aléatoires.

3.2 Le forage de données

Le forage de données ou la fouille de données, en anglais « *Data Mining* » est une expression récente qui date de 1995. La fouille de données est l'ensemble des méthodes et techniques destinées à l'exploration et l'analyse de bases des données, souvent grandes, de façon automatique ou semi-automatique, en vue de détecter dans ces données des règles, des associations, des tendances inconnues ou cachées, ou des structures particulières restituant l'essentiel de l'information utile tout en réduisant la quantité de données (Tufféry, 2005).

Dans (Fayyad, Piatetsky-Shapiro et Smyth, 1996, p. 83), le forage de données est désigné par : “*Data Mining is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data*”.

Le forage de données se distingue par son but final en deux catégories :

1. Le forage de données descriptif (ou exploratoire) vise à mettre en évidence des informations présentes mais cachées par le volume des données ;
2. Le forage de données prédictif (ou explicatif) vise à généraliser de nouvelles informations à partir des informations présentes. Ces nouvelles informations peuvent être qualitatives ou quantitatives.

L'approche de forage de données s'est développée rapidement et est devenue un outil scientifique performant. La première raison de ce progrès est le faible coût de stockage de données ou la saisie automatique de transaction (code à barres, click données de localisation GPS, internet) résultant du développement de la Technologie de l'information et l'augmentation de la puissance de calculs des ordinateurs. D'amples exemples et applications dans différents domaines sont détaillés sur le site <http://www.kdnuggets.com/>. Toutes les applications, quels que soient leurs domaines, reposent sur un squelette bien identifié. Les points indispensables de celui-ci s'énumèrent en : comprendre et analyser les objectifs de l'application, créer une base de données pour la mise au point de l'application, effectuer le prétraitement et le nettoyage des données, analyser les données statistiquement, identifier le type de problème (discrimination, clustering, etc.) et choisir l'algorithme, évaluer la performance de l'algorithme, réitérer les étapes précédentes si nécessaire et déployer l'application. De nos jours, les techniques de forage de données sont également utilisées dans la résolution des problèmes des réseaux électriques comme l'évaluation de la sécurité (Cepeda, Colome et Castrillon, 2011; Hatziargyriou, Papathanassiou et Papadopoulos, 1995; Kai et al., 2007; Rovnyak et al., 1994).

Le forage de données implique une intégration des techniques provenant de diverses disciplines telles que la technologie de base de données, les statistiques, l'apprentissage automatique machine, le calcul haute performance, la reconnaissance des formes, les réseaux de neurones et d'autres encore (Tufféry, 2010).

3.3 L'apprentissage statistique et la modélisation

L'apprentissage statistique est la méthode utilisée pour la mise au point d'un projet de forage de données. Il sert à extraire des règles interprétables à partir de très grandes masses d'exemples et de scénarios simulés. Il fournit un aperçu physique du comportement non linéaire du système à travers les règles extraites. Ces dernières expriment les propriétés spécifiques du problème et elles peuvent être facilement évaluées, critiquées et finalement adoptées par les ingénieurs (Besse, 2009).

La notion de modèle, dans la tradition statistique, est une relation entre des variables exprimables sous une forme analytique ou algorithmique. Le choix du modèle est guidé par des critères d'ajustement et des décisions de validité et où l'interprétation du rôle de chaque variable explicative est prépondérante. Cette démarche essaie de réaliser un bon ajustement de données afin d'approcher la réalité ou le vrai modèle basé sur un théorème physique dont la finalité est explicative.

En revanche, la théorie de l'apprentissage (Saporta, 2011) montre qu'un modèle a un cadre différent et les majorations d'erreur requièrent une autre approche (Besse, Garivier et Loubes, 2014). Avec l'apprentissage statistique l'objectif du modèle est essentiellement prédictif. Le meilleur modèle est celui qui donne de bonnes prévisions. Autrement, celui qui a la capacité de prédire sur des nouvelles observations (généralisation) et non pas celui qui ajuste le mieux les données (prévision du passé). Le choix du modèle dans ce cas est basé sur des critères de qualité de prévision visant à la recherche d'un modèle parcimonieux, pour lequel la complexité est limitée et l'interprétation passe au deuxième plan.

3.4 Le logiciel R

Pour mener notre travail à terme, nous avons exploité le logiciel **R**. Ainsi, tous les calculs et les simulations sont élaborés par le biais du logiciel **R**. Ce dernier est un logiciel de statistique. Il est à la fois un langage informatique et un environnement de travail : les commandes sont exécutées grâce à des instructions codées dans un langage relativement simple, les résultats sont affichés sous forme de texte et les graphiques sont visualisés directement dans une fenêtre qui leur est propre. Le **R** est un logiciel gratuit et à code source ouvert (*open source*). Il fonctionne sous UNIX, Windows et Macintosh. Il peut être téléchargé à partir du site (R, 2001), où, on peut suivre toutes ses évolutions et avoir des réponses aux questions fréquemment posées. Ce logiciel est comparable à Matlab à certains égards, mais se révèle beaucoup plus puissant dans le domaine des traitements statistiques (Lafaye de Micheaux, Drouilhet et Liquet, 2014). Il est très semblable au logiciel Splus qui est un logiciel payant. Il permet entre autre :

- la manipulation de données, les opérations mathématiques classiques (cos, sin, etc.), les nombres complexes, les données discrètes, l'algèbre linéaire ;
- les techniques statistiques : la régression, les modèles GLM, la régression non-linéaire, le Bootstrap, les arbres hiérarchiques, les arbres de régression, les techniques multivariées, les nombres aléatoires et les distributions, les séries chronologiques, etc. ;
- les représentations graphiques : histogrammes, diagrammes en bâton, boxplot, camemberts, lissage, arbres, dessins 2D et 3D, courbes de niveaux, dessins conditionnels, etc. ;
- l'utilisation d'une grande variété de librairies (ou packages), qui offrent de multiples fonctions.

Pour le forage de données, le logiciel **R** nous offre plusieurs algorithmes : l'analyse discriminatoire linéaire (*Linear Discriminant Analysis, lda*), la régression logistique

(*LogisticRegression*, *glm*), les arbres de décision (*DecisionTrees*, *rpart*, *wsrpart*), les forêts aléatoires (*RandomForests* (*randomForest*, *wsrf*)), (*BoostedStumps*, *ada*), les réseaux de neurones (*Neural Networks*, *nnet*), le (*Support Vector Machines*, *kernlab*), etc.

Concernant l'utilisation de **R**, on trouve de nombreux documents de référence et d'aide comme par exemple (Husson, Sébastien et Jérôme, 2009; Lafaye de Micheaux, Drouilhet et Liquet, 2011). On trouve également de nombreux supports en ligne, sur le site officiel du logiciel **R** par exemple <http://cran.r-project.org/manuals.html> et <http://cran.r-project.org/other-docs.html>. Le travail sous **R**, dont l'interface est montrée par la Figure 3.1, s'effectue par des lignes de commande. Il existe d'autres interfaces plus conviviales qui rendent le travail avec le logiciel **R** plus souple comme : l'interface RCommander montrée par la Figure 3.2, l'interface Rattle présentée par la Figure 3.3 qui est très utile pour l'apprentissage statistique (Williams, 2011) et l'interface de RStudio montrée par la Figure 3.4.

Avec le **R**, nous réaliserons la succession d'opérations suivantes : charger le fichier, le scinder en échantillons d'apprentissage, d'évaluation et de test, définir le rôle des variables (cible vs. prédictives), réaliser quelques statistiques descriptives et graphiques pour appréhender les données, construire les modèles prédictifs sur l'échantillon d'apprentissage, les jauger sur l'échantillon test à travers les outils usuels d'évaluation.

L'échantillon d'apprentissage sert à construire le modèle. Celui de validation sert à estimer l'erreur de ce modèle. Le troisième échantillon sert à estimer l'erreur de généralisation du modèle final.

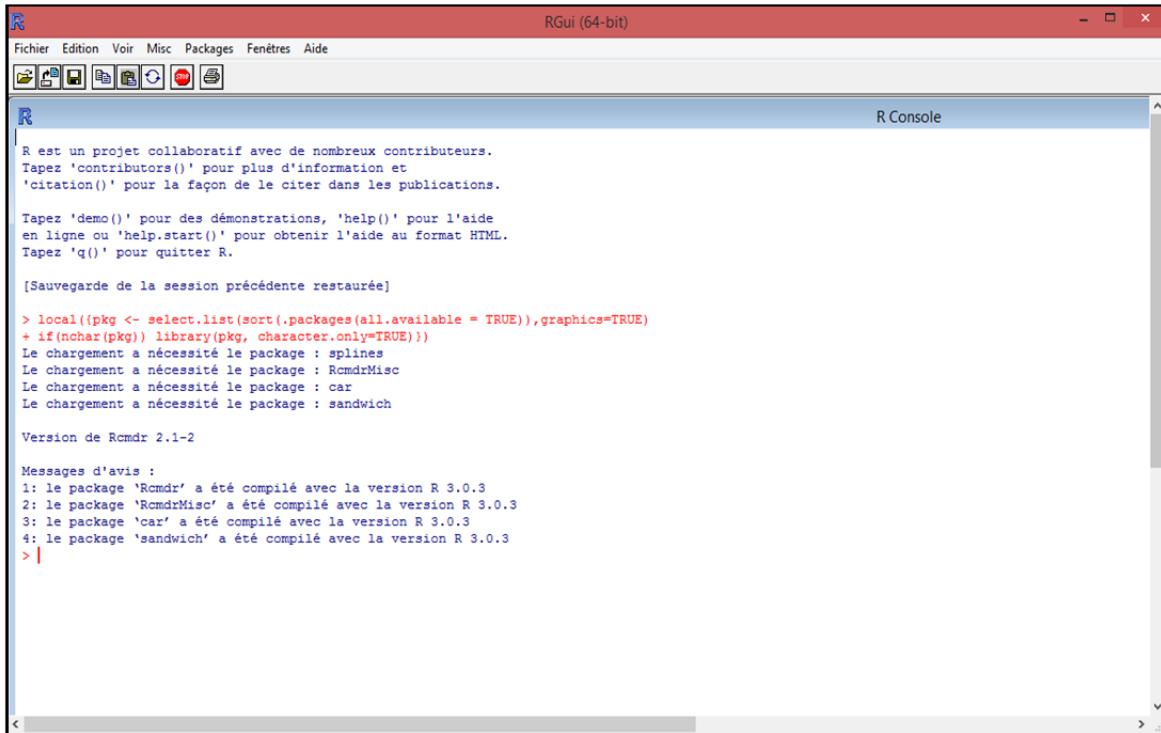


Figure 3.1 Le R sous Windows

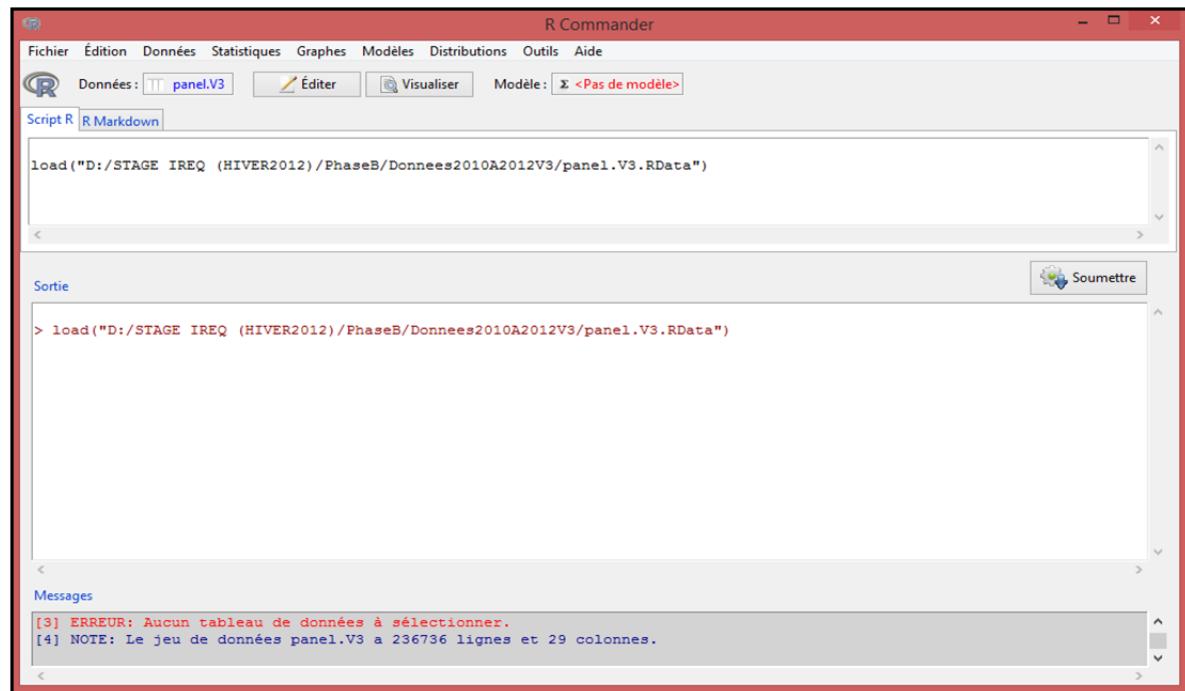


Figure 3.2 L'interface de R Commander

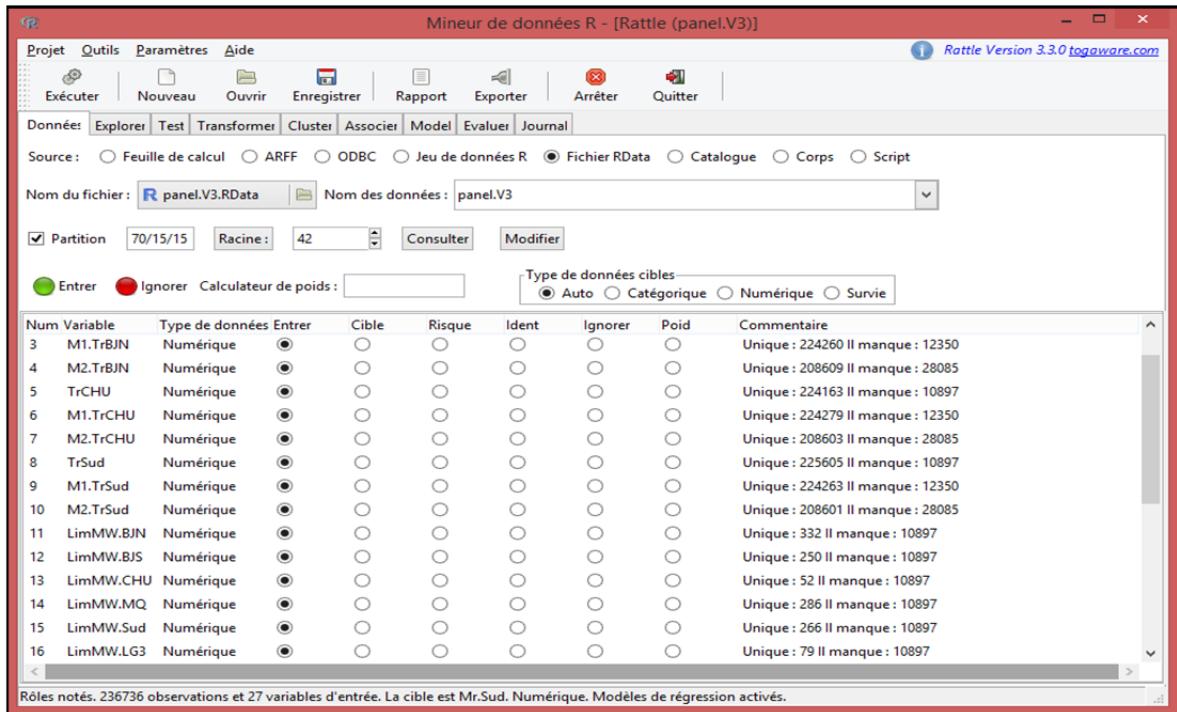


Figure 3.3 L'interface de Rattle

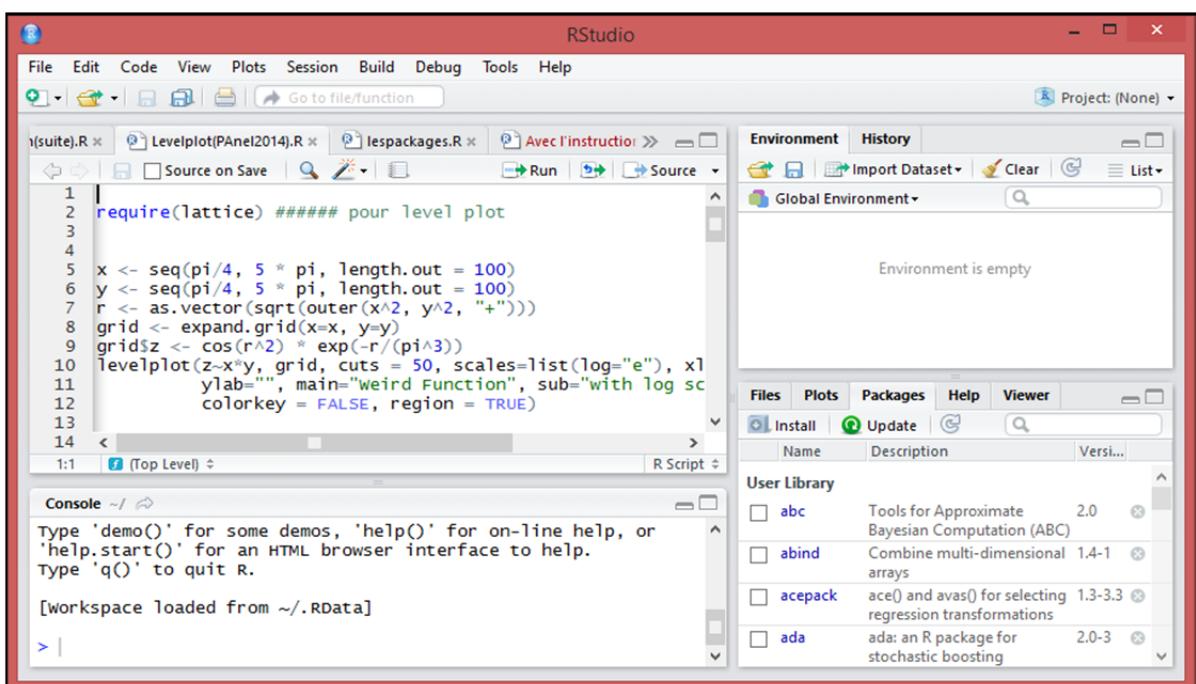


Figure 3.4 L'interface de RStudio

3.5 Arbre de régression

3.5.1 Le concept

L’arbre de décision est une méthode classique d’apprentissage statistique. Le but de celle-ci est de créer un modèle qui prédit la valeur d’une variable cible depuis les valeurs des variables d’entrée. L’apprentissage se fait par partition récursive des instances, selon des règles sur les variables explicatives (Breiman et al., 1984).

Les arbres de décision se distinguent en deux principaux types, selon la variable à expliquer :

- arbres de classification (*Clustering Tree*) : pour modéliser les variables de type nominal (facteur).
- arbres de régression (*Regression Tree*) pour modéliser des variables de type numérique.

Construire un arbre de régression consiste à définir une suite de nœuds par une procédure itérative :

1. Partager l’ensemble des observations initiales, dit racine de l’arbre, en deux sous-ensembles, dits nœuds descendants, sur la base d’une des variables explicatives. Parmi tous les partages possibles explorés sur toutes les variables explicatives et tous les seuils, on retient celui qui minimise, la somme des carrés des écarts à la moyenne suivant :

$$\min_t \sum_k (y_k - \bar{y}_t)^2 .$$

Tel que, \bar{y}_t est la moyenne des y_k observations présentes dans le nœud t ;

2. Une fois, ce critère est satisfait et le partage est effectué, on recommence la même procédure de partage appliquée aux nœuds descendants;
3. Définir quand s’arrête le découpage, en définissant un nœud terminal (feuille). Le critère de l’arrêt du découpage du nœud est de ne pas découper le nœud si le nombre

d'observations dans l'échantillon d'apprentissage contenus dans le nœud est inférieur à un certain nombre ;

4. Élaguer l'arbre quand le nombre de nœuds devient trop important en sélectionnant un sous arbre optimal à partir de l'arbre maximal. L'élagage signifie faire grandir l'arbre au maximum puis supprimer les nœuds inintéressants ;
5. Attribuer au nœud terminal la valeur la plus probable. Chaque feuille (ou nœud terminal de l'arbre) représente une valeur de la variable cible ;
6. Valider l'arbre à partir d'une validation croisée ou d'autres techniques.

Les avantages de la méthode des arbres de régression sont multiples : simple, aisée à comprendre, facile à utiliser et à interpréter. En outre, c'est une méthode qui prend en considération des interactions entre les variables. Toutefois, c'est une méthode instable qui présente une grande variance. Pour remédier à ce problème, Breiman (2001) introduit les forêts aléatoires. Cette dernière consiste à construire de manière aléatoire un très grand nombre d'arbres de régression avant d'en faire la moyenne.

3.5.2 Mise en œuvre sous R

Pour construire un arbre de régression sous **R**, on charge le package **rpart** (Recursive Partitioning and Regression Trees) puis, on fait appel à la fonction **rpart** en exécutant l'instruction **rpart(formula, data, rpart.control)**. Ainsi, on forme l'arbre de régression à base des éléments du jeu de données **data**. On décrit le modèle voulu par la formule **formula** en fonction de la variable cible et les variables prédictives. On règle les paramètres de l'arbre de régression avec **rpart.control**.

Sous le logiciel **R**, on obtient l'arbre construit, sous forme textuelle ou sous forme graphique. Nous présentons un exemple de construction d'un arbre de régression pour visualiser les différents résultats qui s'affichent suite à l'exécution de l'instruction précédente.

Soit le jeu de données **Exemple** extrait de notre base de données, établie pour entreprendre notre travail de thèse. Ce sont les données de 2010 à un pas de 30 minutes. Le jeu de données **Exemple** contient 11 variables et 17 520 observations. Autrement dit, il est constitué de 11 séries chronologiques.

On veut modéliser la variable cible TrMW.CE, le transit actif le long du couloir (CE), en fonction des 10 variables qui restent. Ainsi, la variable TrMW.CE est la variable à prédire et les autres variables de la Figure 3.5 constituent les prédicteurs.

	Deph.702	Deph.703	Deph.704	Deph.706	Deph.708	Deph.717	Deph.719	Deph.722	Deph.750	Deph.751	TrMW.CE
1	0.856796	7.107448	8.156891	20.89167	-0.3449552	8.872883	0.3892278	41.41045	41.48413	41.48399	10331.44
2	1.010957	7.366337	8.570712	21.69734	-0.41695	9.374523	0.04159045	43.15777	43.23233	43.23285	10613.79
3	0.9994732	7.352202	8.528221	21.60431	-0.4167569	9.308356	0.04523271	42.78267	42.85782	42.85744	10543.88
4	0.8651117	7.216817	8.356273	21.30026	-0.5486475	9.107099	-0.256871	42.05507	42.12828	42.12855	10405.26
5	0.8304116	7.119623	8.246049	21.19014	-0.5103726	8.967399	-0.1082767	41.19159	41.26508	41.26567	10271.37
6	0.7620679	6.761479	7.888309	20.62881	-0.4933762	8.622774	0.09506719	40.35608	40.42953	40.42955	10031.38
7	1.028081	6.564681	7.740528	19.89483	-0.3407535	8.579108	0.3601371	40.66268	40.73764	40.73708	9683.478
8	0.9579138	6.326812	7.514565	19.69409	-0.5186398	8.365479	0.1201298	40.1784	40.25342	40.25299	9619.703
9	0.967995	6.195151	7.394063	19.5023	-0.6096361	8.269232	-0.04461887	39.98915	40.06321	40.06286	9538.574
10	1.294988	6.634025	7.886312	20.29001	-0.54457	8.807282	-0.2352629	41.76929	41.84577	41.84616	9842.149
11	1.332397	6.652531	7.923323	20.46104	-0.5408763	8.874403	-0.2223528	42.28983	42.36797	42.36736	9924.423
12	1.119971	6.760632	8.00654	21.42832	-0.593852	8.855837	-0.1780981	40.15127	40.22404	40.22424	9851.03
13	1.022266	6.711752	7.970951	21.64269	-0.5753058	8.841953	0.0444232	40.24802	40.30156	40.32384	9917.883
14	1.111944	7.119299	8.336491	22.215	-0.3524648	9.161717	0.448036	40.85853	40.91417	40.93225	10135.9
15	1.131411	7.045643	8.246759	22.109	-0.3857493	9.074209	0.4681005	40.60461	40.65951	40.67809	10053.09
16	1.019424	7.077583	8.213628	22.61561	-0.543368	8.928643	0.2245962	39.2757	39.33113	39.34882	10055.23
17	1.051927	7.096564	8.245312	22.67076	-0.5307201	8.96547	0.1978585	39.44708	39.50316	39.52209	10097.22
18	1.106756	7.147624	8.304837	22.73461	-0.4808017	9.042605	0.2226657	39.58944	39.64589	39.66473	10083.36
19	1.142744	7.400055	8.578622	23.325	-0.5212716	9.320214	0.03835732	40.42015	40.47738	40.49641	10290.3
20	1.100937	7.660095	8.826352	23.92098	-0.5851924	9.531136	-0.1185857	40.99201	41.04955	41.06849	10548.74
21	1.129192	7.876788	9.09888	24.23699	-0.5694913	9.844589	-0.1257064	42.1219	42.17933	42.19806	10788.24

Figure 3.5 Les 21 premières observations du fichier de données **Exemple**

L’arbre de régression est construit par l’instruction :

```
Exemple.rpart=rpart(TrMW.CE~.,data,control=rpart.control(minsplit = 20, minbucket = round(minsplit/3), cp = 0,01, maxdepth = 30))
```

minsplit=20 : le nombre minimum d’observations qu’un nœud peut avoir pour pouvoir le segmenter.

minbucket : le nombre minimum d'observations dans tout nœud terminal (nœud enfant ou feuille).

cp = 0,01 : le paramètre de complexité.

Le résultat de l'arbre de régression sous forme graphique est donné par la Figure 3.6.

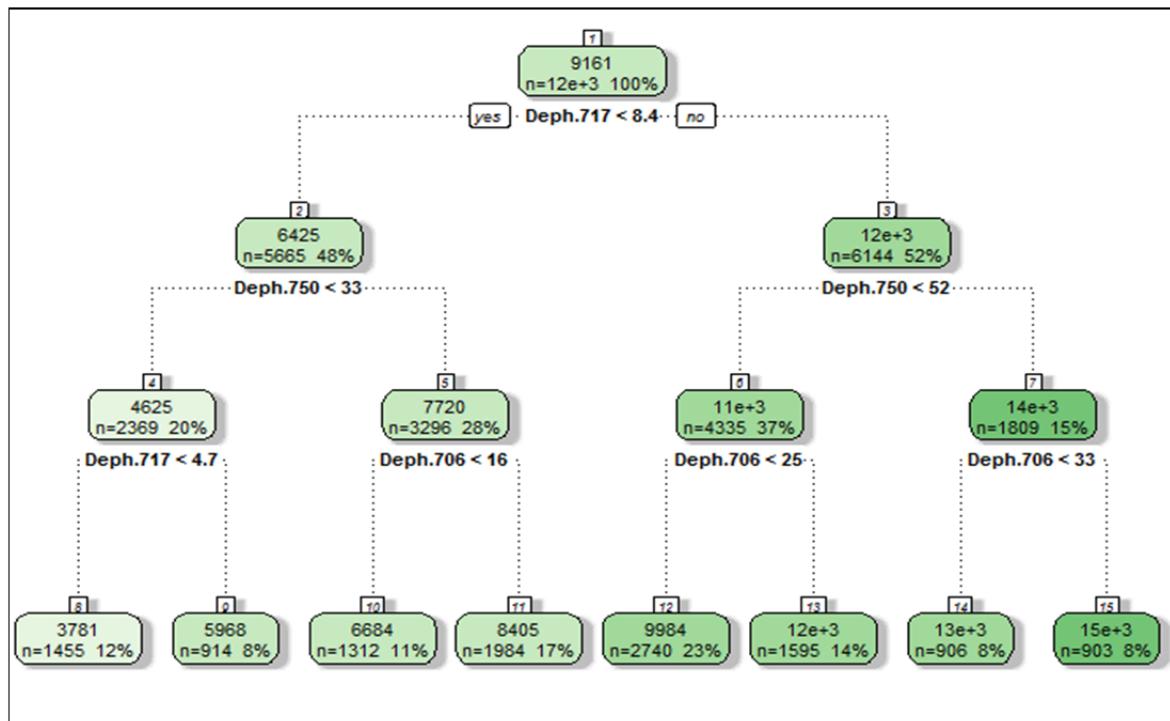


Figure 3.6 L'arbre de régression

Sous forme textuelle, le résultat est illustré par la Figure 3.7. On peut lire, sur cette dernière les paramètres suivants :

- *node* : le nœud,
- *split* : le critère de division,
- *n* : le nombre d'instance de chaque nœud,
- *deviance* : la déviance,
- *yval* : la valeur prédite de la variable à prédire.

```

node), split, n, deviance, yval
  * denotes terminal node

1) root 11809 134434400000 9160.644
  2) Deph.717< 8.410833 5665 23368680000 6425.317
    4) Deph.750< 32.95119 2369 4808556000 4624.521
      8) Deph.717< 4.670834 1455 1420404000 3780.614 *
      9) Deph.717>=4.670834 914 702365600 5967.941 *
    5) Deph.750>=32.95119 3296 5356093000 7719.638
    10) Deph.706< 15.86392 1312 1123766000 6683.705 *
     11) Deph.706>=15.86392 1984 1893257000 8404.691 *
  3) Deph.717>=8.410833 6144 29599010000 11682.720
    6) Deph.750< 51.62601 4335 7534408000 10577.540
    12) Deph.706< 24.90503 2740 2959630000 9984.286 *
    13) Deph.706>=24.90503 1595 1953854000 11596.660 *
    7) Deph.750>=51.62601 1809 4081258000 14331.130
    14) Deph.706< 32.74875 906 875937500 13183.060 *
   15) Deph.706>=32.74875 903 813041700 15483.010 *

```

Figure 3.7 Impression sous forme textuelle de l'arbre de régression

Exemple au nœud 2, on a :

- le numéro de nœud 2);
- le critère de split (ou root pour la racine) : Deph.717< 8,410833;
- le nombre total d'instances pour le nœud : 5 665;
- la déviance, c'est-à-dire la somme des carrés des écarts à la valeur prédictive pour les valeurs de toutes les instances du nœud : 23 368 680 000;
- la valeur prédictive de la variable à prédire : 6 425,317;
- une '*' si c'est un nœud terminal.

On peut aussi visualiser les règles générées lors de la construction d'arbre de régression. La Figure 3.8 présente les huit dernières règles parmi les 15 générées concernant les 15 nœuds.

```

Arbre comme règles :

Rule number: 12 [TrMW.Sud=9984.2856814873 cover=2740 (23%)]
Deph.717>=8.411
Deph.750< 51.63
Deph.706< 24.91

Rule number: 11 [TrMW.Sud=8404.6908104806 cover=1984 (17%)]
Deph.717< 8.411
Deph.750>=32.95
Deph.706>=15.86

Rule number: 13 [TrMW.Sud=11596.6607519751 cover=1595 (14%)]
Deph.717>=8.411
Deph.750< 51.63
Deph.706>=24.91

Rule number: 8 [TrMW.Sud=3780.61397255511 cover=1455 (12%)]
Deph.717< 8.411
Deph.750< 32.95
Deph.717< 4.671

Rule number: 10 [TrMW.Sud=6683.70455773293 cover=1312 (11%)]
Deph.717< 8.411
Deph.750>=32.95
Deph.706< 15.86

Rule number: 9 [TrMW.Sud=5967.9407603059 cover=914 (8%)]
Deph.717< 8.411
Deph.750< 32.95
Deph.717>=4.671

Rule number: 14 [TrMW.Sud=13183.0619769909 cover=906 (8%)]
Deph.717>=8.411
Deph.750>=51.63
Deph.706< 32.75

Rule number: 15 [TrMW.Sud=15483.0058044917 cover=903 (8%)]
Deph.717>=8.411
Deph.750>=51.63
Deph.706>=32.75

```

Figure 3.8 Les règles générées avec rpart

Pour estimer la qualité du modèle de prévision obtenu, on a évalué ce modèle de prévision sur un ensemble de données test dont les observations n'ont pas contribué dans la formation du modèle. L'échantillon test « newdata » représente 15% de l'ensemble de données constituant la base de données **Exemple**. Ces 2 628 observations sont choisies aléatoirement.

Pour la généralisation, on exécute l'instruction :

```
Exemple.pre = predict(Exemple.rpart, newdata)
```

Sur la Figure 3.9, on peut lire le résultat de prévision du modèle à base de la méthode d'arbre de régression sur l'ensemble **test**. La courbe affiche les points Y prédits en fonction des points X réels en abscisse. La ligne bleue en continue représente un ajustement linéaire simple entre les valeurs prédites et les valeurs observées : $Y=aX$. La ligne noire en pointillées est une diagonale, représentant une corrélation parfaite.

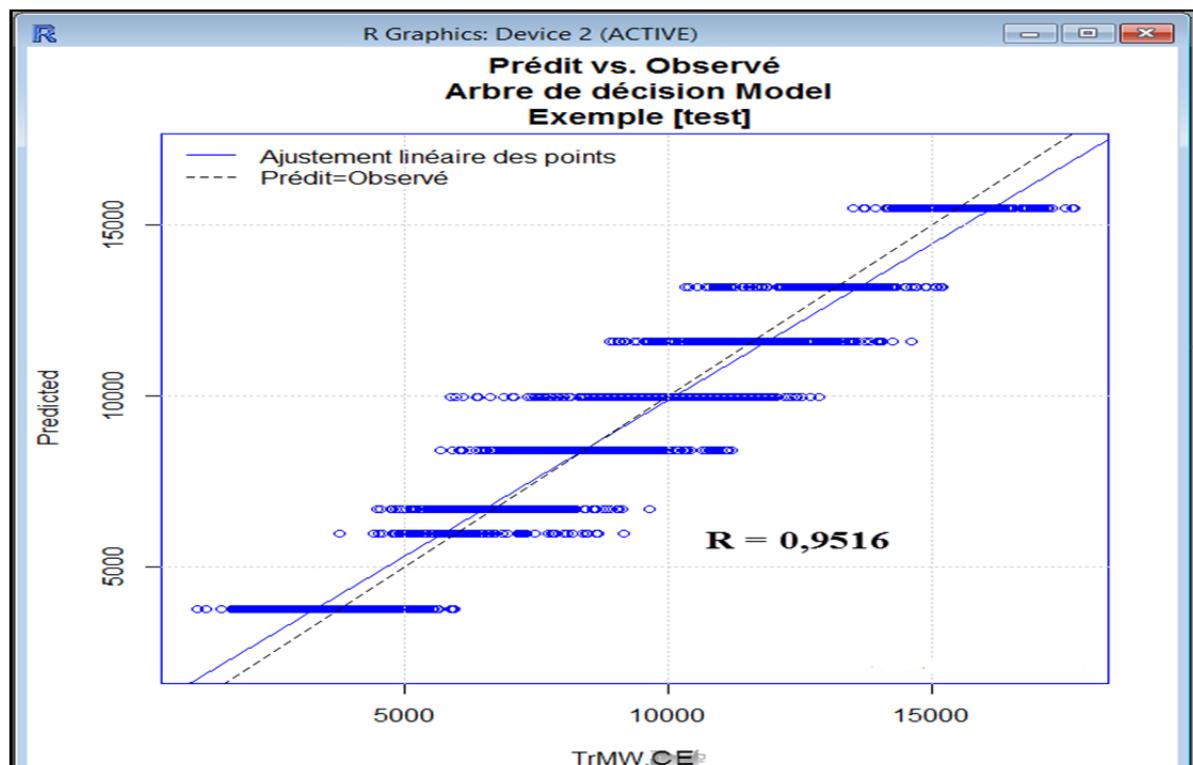


Figure 3.9 L'évaluation du modèle de prévision par l'arbre de régression -Exemple-

Le paramètre R. représente le facteur de corrélation entre les vecteurs des données réelles et des données prédites par le modèle. Il est estimé à 95,16%.

3.6 Les forêts aléatoires

3.6.1 Généralités

Une forêt aléatoire (FA) est un ensemble d'arbres de décision non-élagués. Son algorithme consiste à faire tourner en parallèle un grand nombre (≈ 400 ou plus) d'arbres de décisions construits aléatoirement, avant de les moyenner. L'échantillonnage Bootstrap (OOB : Out-Of-Bag) est utilisé en FA pour avoir une meilleure estimation de la distribution de l'ensemble des données d'origine (Breiman, 2001).

Le Bootstrapping signifie, plutôt que d'utiliser toutes les données pour construire les arbres, on choisit aléatoirement pour chaque arbre un sous-ensemble des données. En termes statistiques, si les arbres sont non corrélés, cela permet de réduire la variance de prévisions (Gareth et al., 2013).

Donc, une forêt aléatoire est une multitude d'arbres de décision non corrélés de telle sorte que chaque arbre dépend d'un sous vecteur sélectionné au hasard du vecteur complet des prédicteurs. Dans le cas de régression, la sortie du modèle des FA est la moyenne d'un grand nombre d'arbres. Ainsi, le terme forêt aléatoire vient du fait que les prédicteurs individuels sont des prédicteurs par arbre, et du fait que chaque arbre dépend d'une variable aléatoire.

Les forêts aléatoires sont souvent utilisées avec les jeux de données de formation d'une taille importante et plus particulièrement avec des variables d'entrée en grand nombre (des centaines voire même des milliers). L'algorithme fonctionne selon un nombre important de variables dans la mesure où il répète le groupement en sous-ensembles des variables disponibles.

L'avantage premier des forêts aléatoires est leur résistance aux variances et aux biais. Le taux d'erreurs des modèles à base des FA diminue quand le nombre d'arbres augmente.

3.6.2 Sélection de variables

Dans le **R**, la package *randomForest* comporte une fonctionnalité nommée « *importance* ». Cette dernière nous permet de déterminer les variables les plus importantes dans un modèle. Cette sélection de variables importantes est basée sur le critère de l'erreur moyenne au carré (MSE). Tout d'abord, pour chaque arbre, l'erreur de prévision sur la partie OOB des données est calculée (taux d'erreur pour la classification, MSE pour la régression). Ensuite, pour chaque variable prédictive, l'erreur est également calculée de la même façon suite à la permutation au hasard des valeurs de la variable en question parmi les données de la partie OOB. La différence (erreur après permutation – erreur avant permutation) est ensuite calculée et puis moyennée sur tous les arbres, et normalisée par l'écart-type des différences.

Cette mesure de la diminution de l'erreur de prévision est ainsi obtenue pour chaque variable explicative et permet de déterminer, selon l'importance de la diminution, lesquelles de ces variables sont les plus utiles dans le modèle.

3.6.3 La prévision avec les forêts aléatoires

On utilise l'algorithme des forêts aléatoires dans le cas de régression afin de prévoir une variable dépendante continue, et de classification, afin de prévoir une variable dépendante catégorielle.

Une forêt aléatoire de type régression est constituée d'un ensemble d'arbres simples de prévision, chacun étant capable de produire une réponse numérique lorsqu'on lui présente un sous-ensemble de variables explicatives ou prédicteurs. L'ensemble de prédicteurs est choisi de façon aléatoire à partir de la même distribution et pour tous les arbres. Pour faire la prévision, on fait la moyenne de toutes les prévisions des arbres composant la forêt et qui n'ont pas été pris dans le Bootstrap.

Plus précisément, fixons une observation (X_i, Y_i) de l'échantillon d'apprentissage. On considère maintenant, l'ensemble des arbres construits sur les échantillons Bootstrap ne

contenant pas cette observation, c'est-à-dire pour lesquels cette observation est « Out-Of-Bag ». Nous agrégeons alors uniquement les prévisions de ces arbres pour évaluer la prévision \hat{y}_k de y_k . L'erreur commise lors de cette prévision est appelée l'erreur Out Of Bag (erreur OOB). Elle se calcule de la manière suivante :

$$errOOB_s = \frac{1}{n} \sum_{k=1}^n (y_k - \hat{y}_k)^2, \text{ où } n \text{ est le nombre d'arbres.}$$

Cette démarche est donnée par l'algorithme proposé par (Liaw et Wiener, 2002) :

1. Construire $ntree$ (nombre d'arbres) échantillons Bootstrap à partir des données ;
2. Pour chaque échantillon, construire l'arbre non élagué de la manière suivante : à chaque nœud, au niveau des prédicteurs, le choix de la variable à partager se fera sur un échantillon aléatoire $mtry$ (le nombre de variables sélectionnées pour chaque nœud) et on choisit parmi celles-ci la meilleure coupure ;
3. Pour la prévision sur un nouveau jeu de données, on agrège les prévisions des $ntree$ arbres.

3.6.4 Le traitement des données catégorielles

On peut utiliser plusieurs variables catégorielles et plusieurs variables continues dans une même analyse des FA. Une variable catégorielle, possédant N modalités sera ventilée dans un tableau disjonctif (avec $N-1$ variables) selon un schéma de codification 0-1. Ainsi, une variable catégorielle à N modalités peut être considérée comme un ensemble de $N-1$ variables, dont une seule va prendre la valeur 1 pour une observation donnée. En fait, la possibilité d'effectuer des prévisions sur un sous-ensemble aléatoire de variables prédictives est l'un des atouts du module *Forêts Aléatoires*, qui le rend particulièrement bien adapté au traitement de jeux de données comportant un nombre extrêmement élevé de variables prédictives.

3.6.5 Le traitement des valeurs manquantes

L'algorithme des forêts aléatoires, implémenté dans **R**, est en mesure de gérer les valeurs manquantes dans les variables prédictives. Lors de la construction du modèle, lorsqu'une observation particulière comporte des valeurs manquantes, la prévision réalisée pour cette observation se base sur l'avant dernier nœud (non-terminal) de l'arbre respectif. Ainsi, par exemple, si à un moment donné de la séquence des arbres, une variable prédictive est sélectionnée au nœud racine (ou un autre nœud non-terminal) et que certaines observations n'ont pas de données valides, la prévision de ces observations sera simplement basée sur la moyenne globale du nœud racine (ou du nœud non-terminal). Il n'est donc pas nécessaire d'éliminer ces observations de l'analyse si elles possèdent des valeurs manquantes pour certains prédicteurs, pas plus qu'il n'est nécessaire de calculer des statistiques de division pour le remplacement des valeurs manquantes.

3.6.6 Mise en œuvre sous R

Sous **R**, on utilise la fonction **randomForest** pour générer les forêts aléatoires et qui est incluse dans le package possédant le même nom. Il y a deux paramètres dans cette fonction à préciser, **ntree** représentant le nombre d'arbres composant la forêt et le paramètre **mtry** représentant le nombre de variables d'entrées choisis de manière aléatoire (Bootstrap) à chaque division.

Le tirage, à chaque nœud, des m variables, se fait sans remise et uniformément parmi toutes les p variables explicatives (chaque variable a une probabilité $1/p$ d'être choisie). Le nombre m ($m \leq p$) est fixé au début de la construction de la forêt et est identique pour tous les arbres. C'est un paramètre très important de la méthode.

L'instruction de base est : **randomForest(formula,data,ntree,mtry)**. Par défaut le nombre **ntree** est fixé à 500 et **mtry** vaut $p/3$, c'est ce que suggère Breiman et ce qui est utilisé dans la procédure de (Liaw et Wiener, 2002), implémentée dans ce package.

En sortie, après avoir utilisé la fonction **randomForest()**, on obtient la moyenne des carrées des résidus (MSE), qui représente l'erreur OOB et le pourcentage de la variance expliquée « *percent variance explained* », qui est calculé par la formule (3.21) et qui permet de juger de la qualité du modèle (Liaw et Wiener, 2002).

$$MSE_{OOB} = \frac{\sum_{i=1}^n \{y_i - \hat{y}_i^{OOB}\}^2}{n} \quad (3.1)$$

$$\text{Percent Variance Explained} = 1 - \frac{MSE_{OOB}}{\hat{\sigma}_y^2} \quad (3.2)$$

\hat{y}_i^{OOB} est la moyenne des prédictions OOB de la i ème observation. Le n dans (3.1) est le nombre d'observations.

$\hat{\sigma}_y^2$ est la variance de la variable Y , calculée comme suit :

$$\hat{\sigma}_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}, \text{ avec } \bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

Enfin, il est possible de prédire sur un nouveau jeu de données le modèle de forêt aléatoire par la fonction **predict()**, **predict(modelrf,newdata)**, tel que **modelrf** est le modèle de forêt aléatoire établi et **newdata** est le nouveau jeu de données.

À titre d'exemple d'application des forêts aléatoires, on considère le même exemple traité précédemment par l'arbre de régression. On utilise le même échantillon d'apprentissage, le même échantillon de validation et le même échantillon test que dans le cas précédent.

Sous **R**, on construit la forêt aléatoire en introduisant l'instruction :

```
Exemple.RF = randomForest(TrMW.CE ~ ., data, ntree = 500, mtry = 3, importance = TRUE)
```

L'exécution de cette instruction nous procure l'affichage montré par la Figure 3.10, à savoir : la variance du modèle qui est de 99,09 % et l'ordre d'importance des variables explicatives, selon le critère de la moyenne des erreurs au carré (MSE).

```
Summary of the Forêt aléatoire Model
=====
Number of observations used to build the model: 12264
Missing value imputation is active.

Call:
randomForest(formula = TrMW.Sud ~ .,
              data = crs$dataset[crs$sample, c(crs$input, crs$target)],
              ntree = 500, mtry = 3, importance = TRUE, replace = FALSE, na.action = na.roughfix)

      Type of random forest: regression
                  Number of trees: 500
No. of variables tried at each split: 8

      Mean of squared residuals: 100215.8
          % Var explained: 99.09

Importance des variables
=====

      %IncMSE IncNodePurity
Deph.702 133.06    394860071
Deph.719 126.74    690509436
Deph.708 111.07    374987172
Deph.703  61.91    883708475
Deph.706  52.09    7712467465
Deph.717  41.45    43424205661
Deph.751  29.76    7636061573
Deph.750  28.88    8746339716
Deph.722  26.34    4575613643
Deph.704  21.10    10395529280
```

Figure 3.10 La sortie du modèle de forêt aléatoire

Le grand nombre d'arbres, n=500, a contribué dans la diminution de la variance du modèle, comme le montre la Figure 3.11. Tous ces arbres sont accessibles et compréhensibles. Nous pouvons visualiser les règles de tout arbre construit, en tapant l'instruction :

`printRandomForests(RF, n),`

avec **n** le *nième arbre*. Pour le premier arbre, n=1, quelques règles apparaissent sur la Figure 3.12.

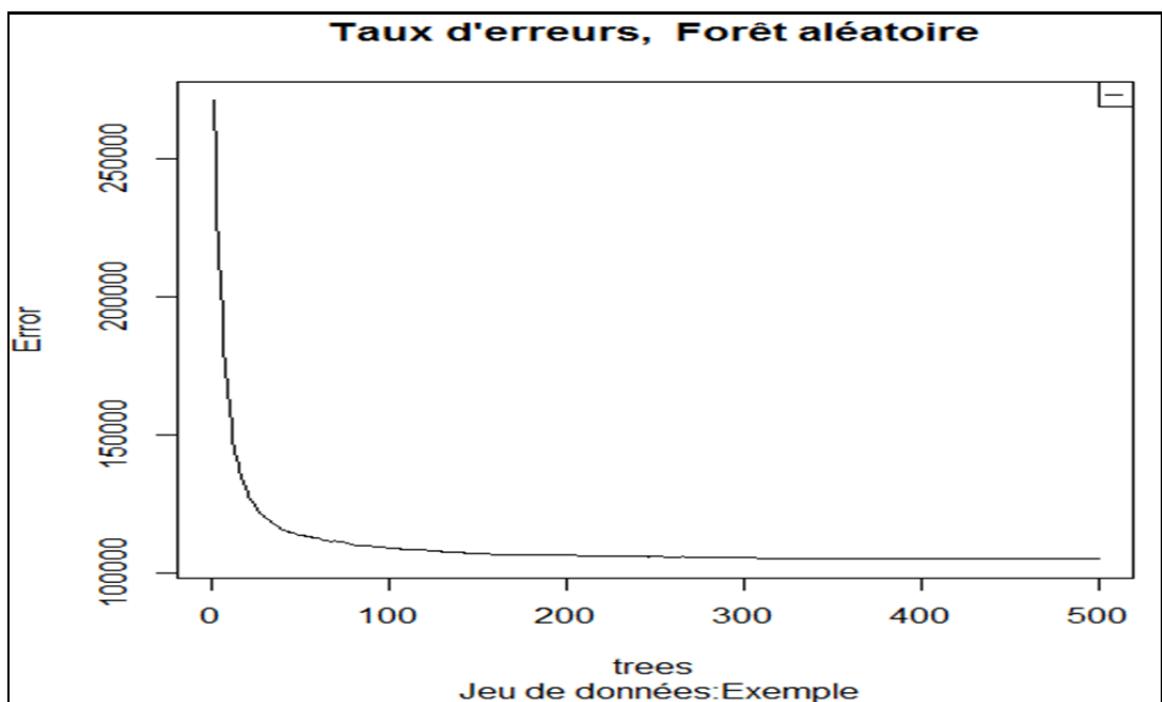


Figure 3.11 Taux d'erreurs en fonction du nombre d'arbres

Pour évaluer la qualité de prévision ou de généralisation du modèle sur l'échantillon **test**, on exécute l'instruction :

`Exemple.pre = predict(Exemple.RF, newdata)`

La qualité du modèle de prévision est évaluée par le biais du coefficient R. Il est de 99,52 %, comme le montre la Figure 3.13.

```
3: Deph.703 > 11.7018113136
4: Deph.706 > 36.92259025575
5: Deph.703 > 13.23468780515
6: Deph.703 > 13.88664531705
7: Deph.708 > -0.964088201523
8: Deph.751 <= 65.39822387695
9: Deph.751 <= 62.53561210635
-----
Tree 1 Rule 2534 Node 1404 Regression (to do - extract predicted value)

1: Deph.706 > 18.48793697355
2: Deph.750 > 51.36391448975
3: Deph.703 > 11.7018113136
4: Deph.706 > 36.92259025575
5: Deph.703 > 13.23468780515
6: Deph.703 > 13.88664531705
7: Deph.708 > -0.964088201523
8: Deph.751 <= 65.39822387695
9: Deph.751 > 62.53561210635
10: Deph.751 <= 63.8485622406
-----
Tree 1 Rule 2535 Node 1405 Regression (to do - extract predicted value)

1: Deph.706 > 18.48793697355
2: Deph.750 > 51.36391448975
3: Deph.703 > 11.7018113136
4: Deph.706 > 36.92259025575
5: Deph.703 > 13.23468780515
6: Deph.703 > 13.88664531705
7: Deph.708 > -0.964088201523
8: Deph.751 <= 65.39822387695
9: Deph.751 > 62.53561210635
10: Deph.751 > 63.8485622406
-----
Tree 1 Rule 2536 Node 489 Regression (to do - extract predicted value)

1: Deph.706 > 18.48793697355
2: Deph.750 > 51.36391448975
3: Deph.703 > 11.7018113136
4: Deph.706 > 36.92259025575
5: Deph.703 > 13.23468780515
6: Deph.703 > 13.88664531705
7: Deph.708 > -0.964088201523
8: Deph.751 > 65.39822387695
-----
Number of rules in Tree 1: 2536
```

Figure 3.12 Les règles du premier arbre de la forêt aléatoire

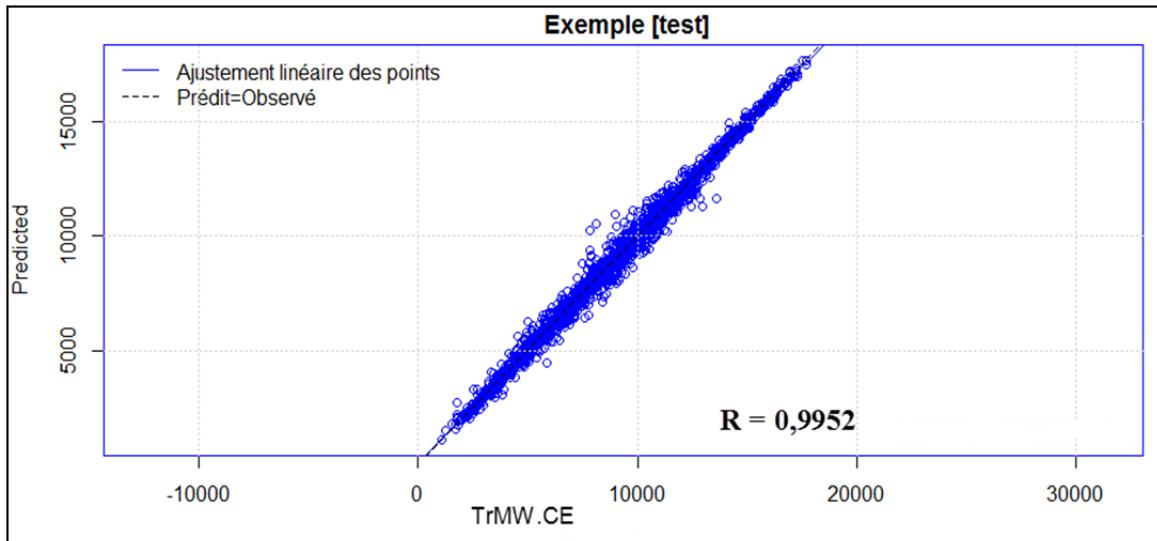


Figure 3.13 L'évaluation du modèle de prévision par la FA -**Exemple-**

3.7 Les réseaux de neurones

3.7.1 Principe

Cette méthode repose sur la notion de neurone formel. Un neurone formel, voir Figure 3.14, est un modèle, caractérisé par des signaux d'entrée (les variables explicatives), une fonction d'activation f , de forme :

$$f \left(\alpha_0 + \sum_i \alpha_i \times x_i \right)$$

f peut être linéaire, à seuil stochastique, et le plus souvent sigmoïde. Le calcul des paramètres se fait par apprentissage.

Les neurones sont ensuite associés en couche, tel que montré sur la Figure 3.15. Une couche d'entrée lit les signaux entrant, un neurone par entrée x^j , une couche en sortie fournit la réponse du système. Une ou plusieurs couches cachées participent au transfert. Un neurone d'une couche cachée est connecté en entrée à chacun des neurones de la couche précédente et en sortie à chaque neurone de la couche suivante. En régression (la variable cible est quantitative), la dernière couche est constituée d'un seul neurone muni de la fonction

d'activation identité tandis que les autres neurones (couche cachée) sont munis de la fonction sigmoïde <http://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-app-rn.pdfs>

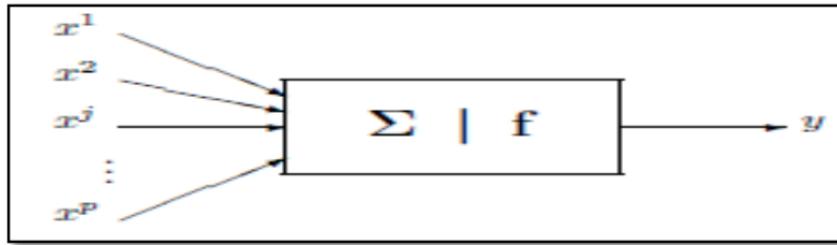


Figure 3.14 Représentation d'un neurone formel

Lors d'apprentissage (ajustement), on minimise une fonction objective $Q(\alpha)$ (perte quadratique). À partir du gradient de cette fonction, on utilise un algorithme d'optimisation. Le modèle dépend de plusieurs paramètres :

- l'architecture du réseau : nombre de couches cachées (une ou deux en général) et le nombre de neurones par couche,
- le nombre d'itérations, l'erreur maximale tolérée et un terme de régularisation (decay).

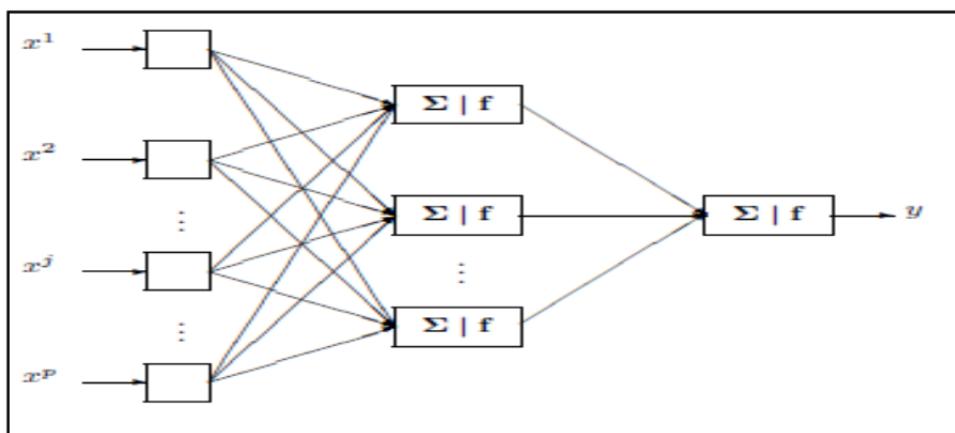


Figure 3.15 Exemple de perceptron multicouche élémentaire

3.7.2 Mise en œuvre sous R

Pour expliciter la mise en œuvre de la méthode NNET sous le logiciel **R**, on traite le même exemple traité dans les deux sections précédentes.

D'abord, on charge le package dédié aux réseaux de neurones artificielles **library(nnet)**. Ensuite on lance l'estimation du modèle tout en précisant le nombre de couches masquées **size**.

```
Exemple.nnet= nnet(TrMW.CE ~ .,data,size=10, linout=TRUE, skip=TRUE,
MaxNWts=10000, trace=FALSE, maxit=100)
```

Puis, pour évaluer la qualité de la prévision, on applique le modèle sur l'échantillon **test** avec l'instruction :

```
Exemple.pre = predict(Exemple.nnet, newdata)
```

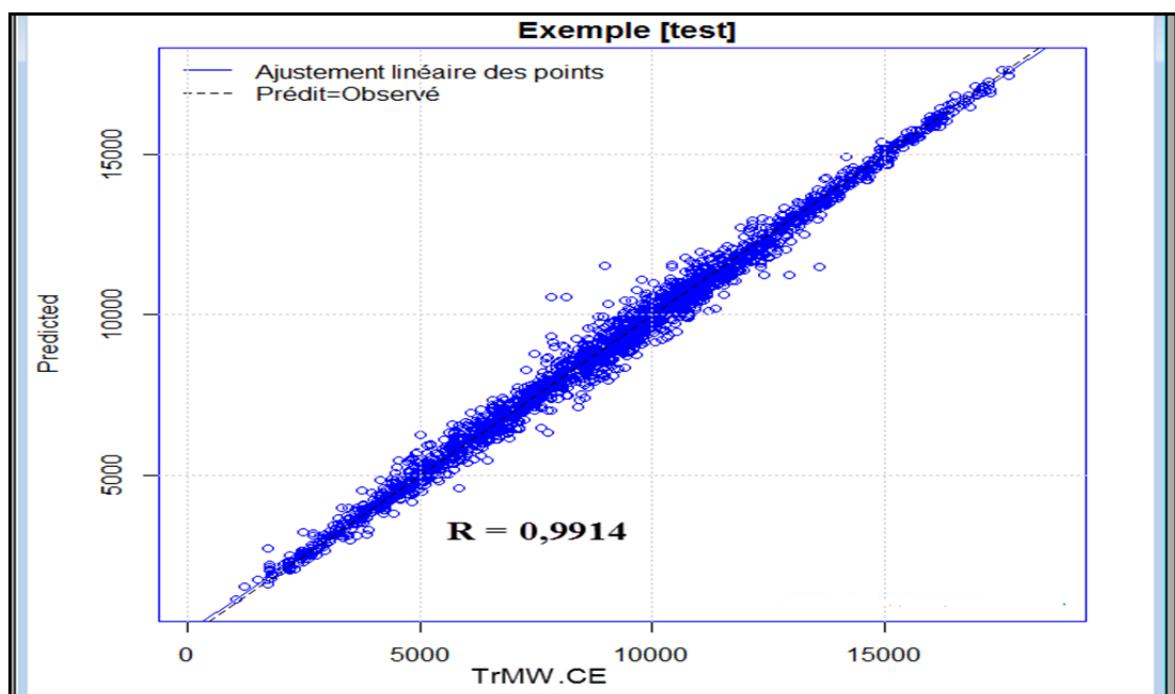


Figure 3.16 L'évaluation du modèle de prévision par le réseau de neurones -Exemple-

La qualité de la prévision de la méthode NNET est donnée par la Figure 3.16 où on peut lire la valeur de R qui est de 99,14 %.

3.8 Les modèles linéaires généralisés

3.8.1 Théorie

Le modèle linéaire généralisé, comme son nom l'indique, est une généralisation du modèle linéaire qui s'écrit sous la forme générale :

$$E(y) = a_1 x_1 + a_2 x_2 + \dots + b + \varepsilon \quad (3.1)$$

Où $E(y)$ représente l'espérance de la variable à expliquer y ; x_1, x_2, \dots sont les variables explicatives; ε représente le terme d'erreur (le bruit) qui doit impérativement suivre une distribution normale et de même variance. C'est-à-dire que la variance de ce terme d'erreur doit être indépendante de la valeur des différentes variables x_1, x_2, \dots .

On cherche à estimer les paramètres a_1, a_2, \dots, b de telle sorte que ce modèle soit très proche de la réalité. Le Tableau 3.1 donne les différents types de modèle linéaire général que l'on peut rencontrer.

Tableau 3.1 Les modèles linéaires générales

Type et nombre de variables explicatives	Type de modèle
Une variable explicative quantitative	Une régression linéaire simple
Plusieurs variables et toutes quantitatives	Une régression linéaire multiple
Une variable explicative qualitative	Une ANOVA à un facteur
Plusieurs variables et toutes qualitatives	Une ANOVA à plusieurs facteurs
Une combinaison de variables quantitatives et qualitatives	Une analyse de covariance (ANACOV ou ANCOVA)

La régression linéaire multiple est un cas particulier du modèle linéaire général. Elle consiste à expliquer une variable quantitative par plusieurs variables quantitatives. ANOVA est aussi un cas particulier du modèle linéaire général. Elle consiste à expliquer une variable qualitative par une ou plusieurs variables qualitative(s).

Le modèle linéaire général repose sur une hypothèse forte qui est la normalité des erreurs et l'égalité de leurs variances. Cependant, nous avons parfois si ce n'est pas souvent, le besoin d'expliquer des variables (et donc leurs erreurs) qui ne suivent pas ce prérequis (Baum, 2006). Pour tenir compte de ces cas, la solution la plus répandue consiste à trouver une transformation mathématique de la variable à expliquer pour la rendre normale (et son erreur avec) et pour en stabiliser les variances. C'est dans ce cadre de transformation que se développe le modèle linéaire généralisé (Generalized Linear Model; GLM). L'idée reste d'utiliser une transformation mathématique sur la variable à expliquer y mais en tenant compte cette fois-ci de la véritable distribution des erreurs. La fonction mathématique utilisée pour transformer la variable à expliquer est appelée *fonction de lien*. Différentes fonctions peuvent être utilisées, selon la distribution réelle de la variable étudiée (et de son erreur). Le modèle à ajuster ou le modèle linéaire généralisé s'écrit :

$$f(E(y)) = a_1 x_1 + a_2 x_2 + \dots + b + \varepsilon \quad (3.2)$$

Où $f(\dots)$ est la fonction de lien.

Les paramètres de ce modèle sont estimés par la méthode du maximum de vraisemblances (Ouellet, Belley-F. et Leblond, 2005). L'estimateur du maximum de vraisemblance a pour but de choisir le vecteur de paramètres pour lequel les données observées sont les plus probables.

Le modèle linéaire généralisé est un outil qui peut être utilisé dans de nombreuses situations, afin d'analyser des variables quelle que soit leurs distributions, même si la variable est de distribution normale. Le GLM est utilisé aussi lorsque la variable à expliquer n'est pas défini

sur $]-\infty, \infty]$. C'est le cas par exemple de données de fréquence ($[0,1]$) ou de comptage (nombres entiers positifs). On utilise alors une fonction qui permet de transformer cette variable en une variable définie sur l'intervalle $]-\infty, \infty]$. Le Tableau 3.2 donne les principales situations rencontrées.

Tableau 3.2 Les familles des GLM

Famille	Type de données	Fonction de lien	Formule
Normale	Variable suivant une loi normale	Identité	$f(y) = y$
Poisson	Comptage	Log	$f(y) = \log(y)$
Binomiale	Pourcentage	Logit	$f(y) = \log(y / (1-y))$
Gamma	Durée	Inverse	$f(y) = 1/y$

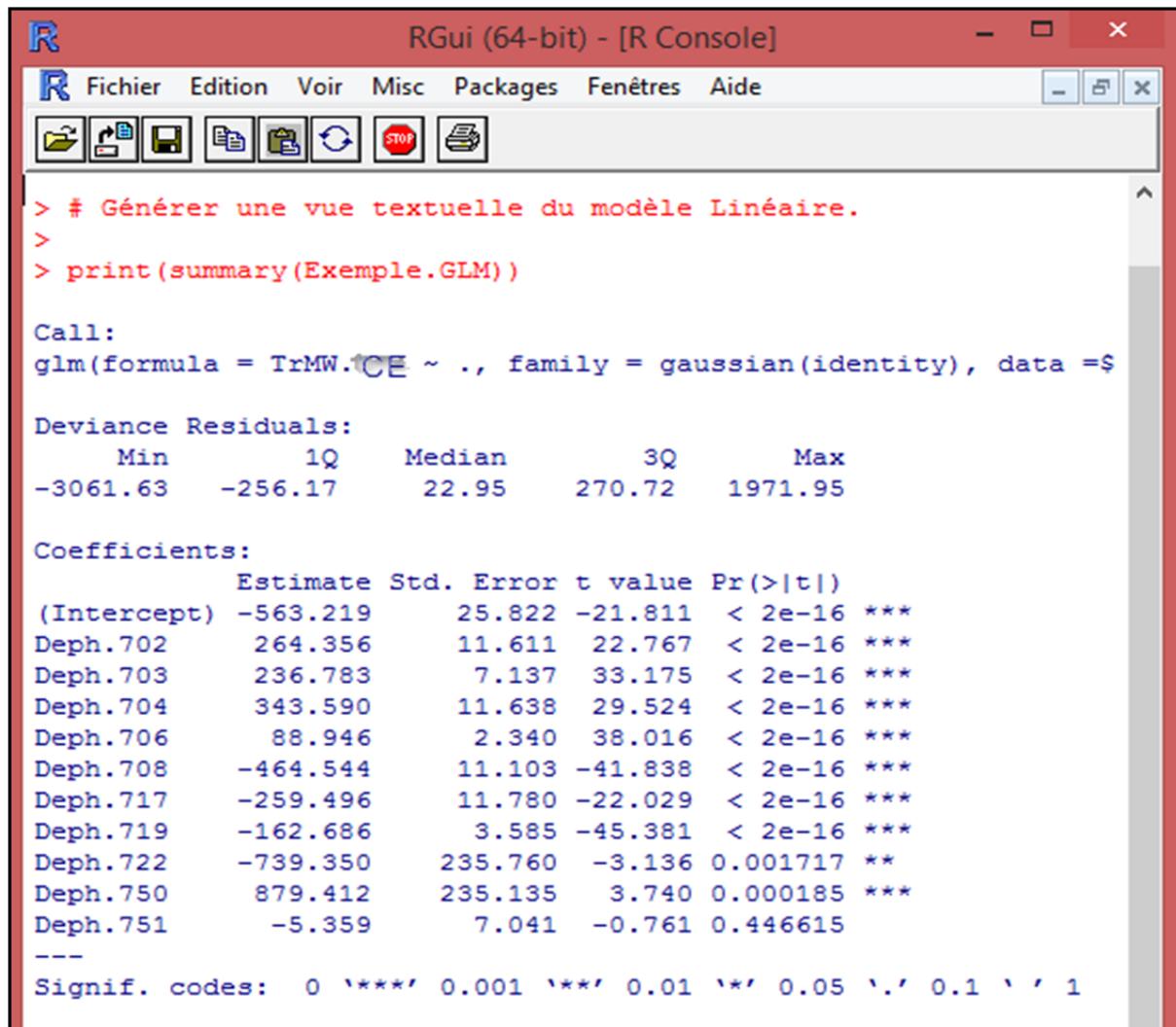
3.8.2 Mise en œuvre sous R

À titre d'exemple d'application du modèle GLM, on considère toujours l'exemple traité précédemment par les trois autres algorithmes : arbre de régression (rpart), forêts aléatoires (randomforest) et les réseaux de neurones (NNET).

Sous **R**, on construit le modèle GLM en introduisant l'instruction :

Exemple.GLM= **glm**(formula = TrMW.CE ~ ., data)

En exécutant l'instruction citée, le résultat s'affiche, comme le montre la Figure 3.17.



```

R Gui (64-bit) - [R Console]
R Fichier Edition Voir Misc Packages Fenêtres Aide
[Icons: File, Edit, View, Help, etc.]

> # Générer une vue textuelle du modèle Linéaire.
>
> print(summary(Exemple.GLM))

Call:
glm(formula = TrMW.1CE ~ ., family = gaussian(identity), data = $)

Deviance Residuals:
    Min      1Q   Median      3Q     Max 
-3061.63 -256.17    22.95   270.72  1971.95 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -563.219   25.822 -21.811 < 2e-16 ***
Deph.702     264.356   11.611  22.767 < 2e-16 ***
Deph.703     236.783    7.137  33.175 < 2e-16 ***
Deph.704     343.590   11.638  29.524 < 2e-16 ***
Deph.706      88.946    2.340  38.016 < 2e-16 ***
Deph.708    -464.544   11.103 -41.838 < 2e-16 ***
Deph.717    -259.496   11.780 -22.029 < 2e-16 ***
Deph.719    -162.686    3.585 -45.381 < 2e-16 ***
Deph.722    -739.350   235.760 -3.136 0.001717 ** 
Deph.750     879.412   235.135   3.740 0.000185 *** 
Deph.751     -5.359     7.041  -0.761 0.446615  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figure 3.17 Les paramètres du modèle GLM -Exemple-

Pour évaluer la qualité de la prévision ou la généralisation du modèle, on applique le modèle sur l'échantillon test avec l'instruction :

`Exemple.pre = predict(Exemple.GLM, newdata).`

La qualité de la prévision est évaluée par un R et il est de 99,14 %, comme le montre la Figure 3.18.

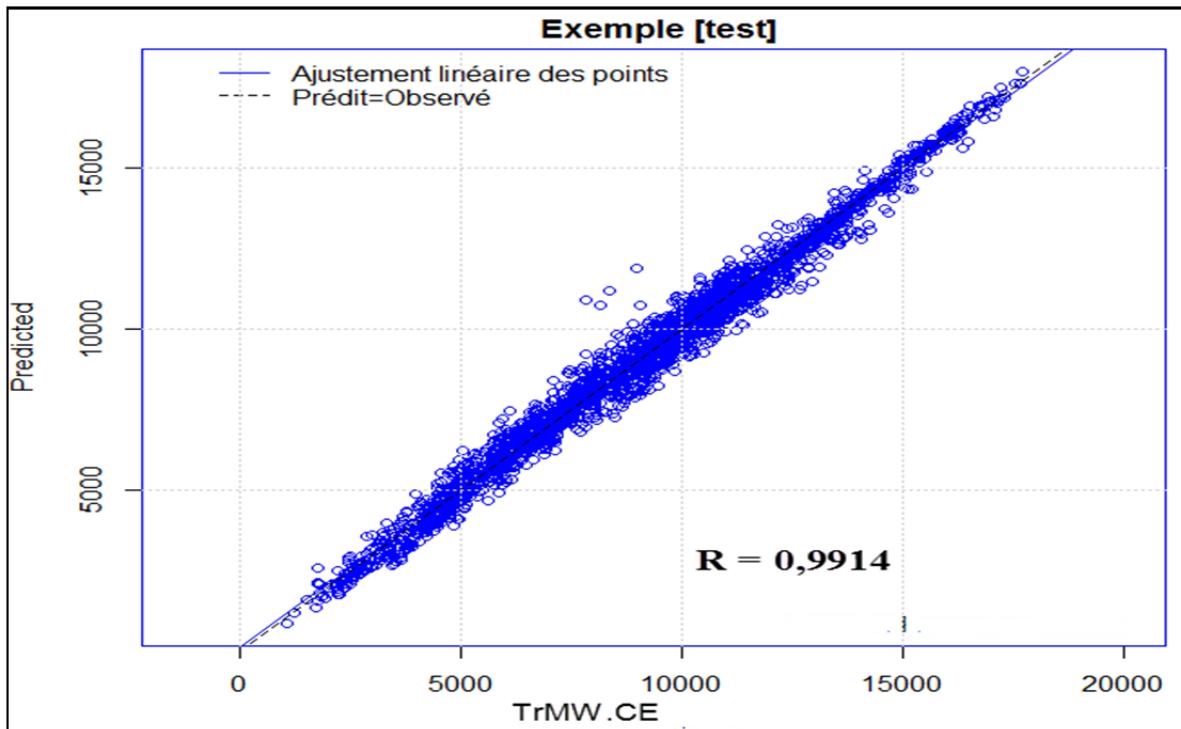


Figure 3.18 L'évaluation du modèle de prévision par GLM

3.9 Discussion

L'exemple traité par les différentes méthodes d'apprentissage statistique est de même nature que notre problème de recherche. Cependant, nous avons utilisé une base de données d'un nombre restreint de variables et d'observations. Le jeu de données est constitué de 11 variables et 17 520 observations. Soulignons que 12 264 de ces observations, équivalent à 70% de l'ensemble des données, sont utilisées pour former le modèle. Les 2 628 autres observations, équivalent à 15%, sont utilisées pour valider le modèle, et le reste des observations, 2 628 autres, sont utilisées pour tester le modèle. Pour les quatre modèles, nous avons utilisé le même échantillon de formation, puis le même échantillon de validation et enfin le même échantillon test. Les résultats de cette étude sont récapitulés dans les tableaux et les figures ci-dessous.

Tableau 3.3 Comparaison des différentes méthodes

Méthode	R de validation	R de test
L'arbre de régression	95,31 %	95,16 %
GLM	99,08 %	99,14 %
Le réseau de neurone	99,08 %	99,14 %
La forêt aléatoire	99,54 %	99,52 %

À partir du Tableau 3.3, la méthode des FA à base d'arbres de régression affiche sans surprise des performances significativement supérieures à celles de l'arbre de régression. Ceci est dû au fait que la méthode des FA est basée sur l'agrégation des résultats de plusieurs arbres de régression.

Par rapport à la méthode de GLM, tout en se référant à la référence (Gareth et al., 2013), ce résultat est attendu. Si la relation entre les variables explicatives et la variable cible est assez bien approximée par un modèle linéaire, alors une approche telle que le GLM surpassera probablement la méthode des FA qui n'exploite pas cette structure linéaire. En revanche, devant une relation non linéaire et complexe entre les caractéristiques et la réponse, la méthode des FA est de loin supérieure à la méthode GLM.

Concernant la comparaison de la méthode des forêts aléatoires avec la méthode NNET, les auteurs de la référence (Kamwa, Samantaray et Joos, 2012) ont effectué de vastes évaluations comparatives de plusieurs méthodes d'apprentissage statistique, entre-autres les méthodes NNET et FA. L'étude de classification était basée sur une base de données à 90% des données réelles du réseau de transport d'HQ. Les modèles élaborés avaient pour but d'anticiper la perception des catastrophes dans les réseaux électriques. La comparaison de la performance prédictive des quatre méthodes étudiées se basait sur la mesure des critères de précision, de fiabilité et de sécurité. La méthode des FA à base d'arbres de classification montrait des performances significativement supérieures à la méthode NNET. Mieux encore, c'est seulement avec les FA que la performance de 99,0% sur les trois critères a été obtenue, répondant ainsi à la norme "99.9%" de fiabilité (Energy, 2008).

La comparaison de l'efficacité des quatre méthodes, appliquées à notre exemple, montre la qualité supérieure de la performance de la méthode FA. Cette comparaison est explicitée à travers la Figure 3.19. Ce graphique évoque les séries chronologiques des erreurs de prévision de chaque modèle. L'erreur de prévision tracée est la différence entre les valeurs réelles et les valeurs prédites. L'amplitude de la différence entre les erreurs de prévision des quatre méthodes est nettement discernable.

Pour mieux cerner cette différence, nous exprimons les erreurs de prévisions en termes statistiques. Le Tableau 3.4 illustre en fonction des quartiles les différentes erreurs de prévision. Dans la première ligne du Tableau 3.3 : « méth » la méthode d'apprentissage, « mean » la moyenne de l'erreur de prévision, « sd » l'écart-type qui représente la dispersion des données de l'erreur, « IQR » l'écart interquartile, « 0% » représente la valeur minimale des erreurs, « 25% », « 50% » et « 75% » les différents quartiles et enfin la valeur maximale de l'erreur de prévision.

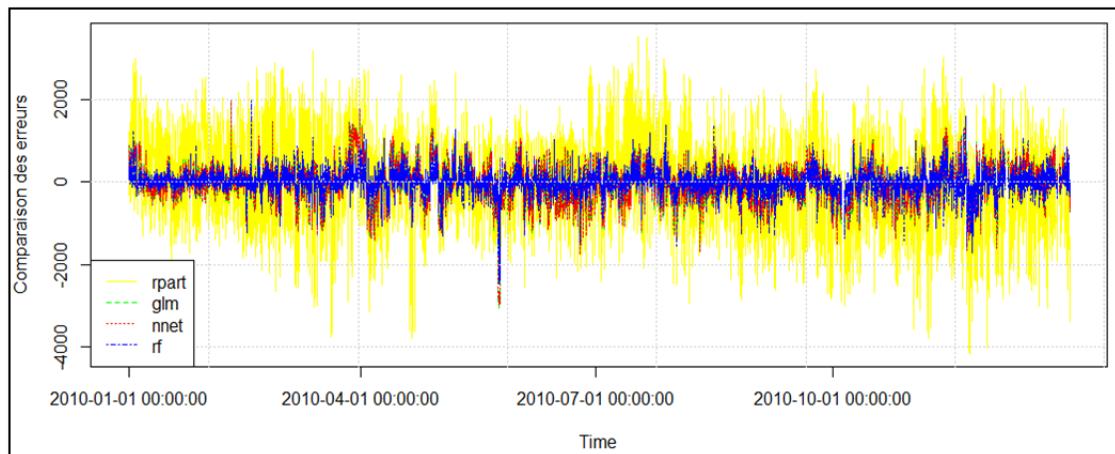


Figure 3.19 Les séries chronologiques des erreurs de prévisions

Tableau 3.4 Résumé des caractéristiques statistiques des erreurs de prévisions

méth	mean	sd	IQR	0%	25%	50%	75%	100%
rpart	6,83	1 006,76	1 341,30	-4 161,21	-649,30	34,130	691,99	3 538,21
glm	2,00	441,76	526,45	-3 061,62	-253,80	23,45	272,64	1 971,94
nnet	2,20	440,68	528,21	-2 994,32	-254,37	22,16	273,83	1 993,61
Rf	0,61	215,48	177,12	-2 472,69	-84,72	5,64	92,40	1 978,07

Avec la méthode des FA, toutes les valeurs des caractéristiques statistiques de l'erreur de prévision sont inférieures à celles des autres méthodes. Par exemple, avec la méthode des FA, la moyenne de l'erreur est à 0,61 alors qu'elle est de 2,20 avec NNET. Avec FA, 75% des données représentant l'erreur de prévision sont inférieures ou égales à 92,40, tandis qu'avec la méthode NNET, 75% de données sont inférieures ou égales à 273,83. Ainsi, l'erreur de prévision avec les FA est plus faible comparativement aux erreurs de prévision obtenues via les autres méthodes. Par conséquent, la méthode des FA à base d'arbres de régression est plus performante pour la prévision, ou la généralisation, dans ce type de problème.

La précision des résultats avec la méthode des FA s'améliore nettement bien en introduisant toutes les variables et toutes les observations qui décrivent le problème. Toutefois, nous nous sommes limités à un nombre restreint de données car le but principal du traitement de cet exemple est de montrer la souplesse indéniable du logiciel **R** face au traitement des problèmes si complexes. Effectivement, nous avions déjà comme hypothèse que la méthode des forêts aléatoires se démarquerait des autres, en raison de la nature non linéaire des données des réseaux électriques et les fortes interactions entre les variables.

Un autre moyen d'améliorer la précision du modèle des FA, en plus d'utiliser un nombre d'arbres assez élevé, est de choisir adéquatement le paramètre m , le nombre de variables explicatives tirées au hasard au niveau de chaque nœud. En effet, comme on peut le voir sur la Figure 3.20, la valeur de m donnée par défaut par l'algorithme « randomforest » ne nous mènent pas nécessairement au meilleur résultat. Dans cet exemple, le résultat optimal est

obtenu pour m égal à huit. Des tests de comparaison étaient effectués pour toutes les valeurs possibles de m et dont les résultats n'apparaissent pas ici dans ce document.

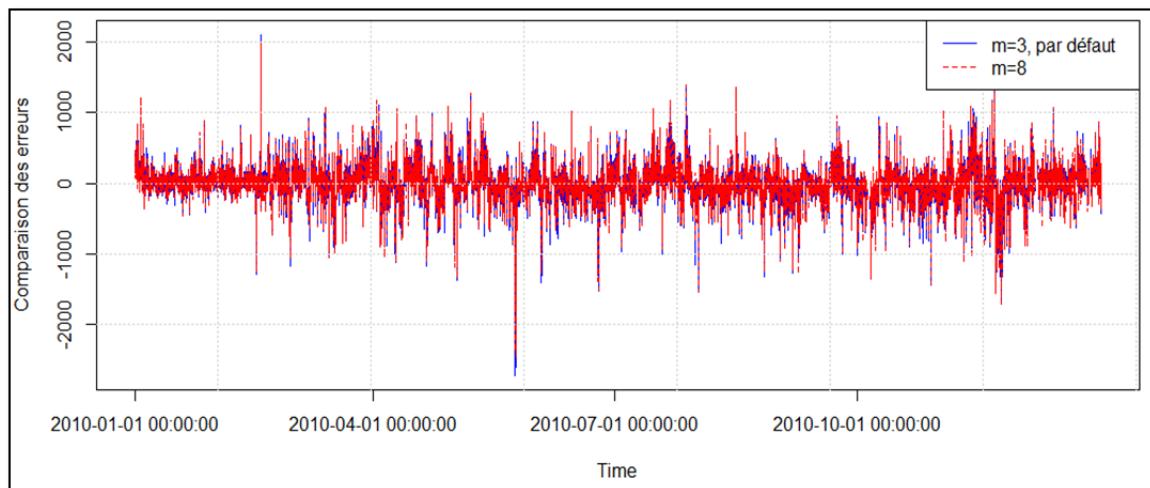


Figure 3.20 Les erreurs de prévisions avec les F.A en fonction du paramètres "m"

3.10 Conclusion

Sous le logiciel **R**, qu'il soit pour les problèmes de régression ou de classification, plusieurs algorithmes d'apprentissage statistique sont offerts (Tufféry, 2015). Nous avons présenté brièvement quelques méthodes que nous aborderons ensuite dans les chapitres ultérieurs et ceci pour différents buts.

À travers ce chapitre, nous avons présenté des éléments de l'apprentissage statistique. Il consiste à définir un modèle à partir d'une très grande base de données dont l'objectif est la prévision. Ainsi, un bon modèle est caractérisé par sa capacité de généralisation sur d'autres nouvelles observations. Conséquemment, l'évaluation de la qualité du modèle de prévision repose sur la mesure de sa précision et sa robustesse. Différemment des autres méthodes, les forêts aléatoires présentent un avantage significatif en terme d'erreur de prévision, et ce, à partir des effectifs d'apprentissage faibles (Gareth et al., 2013).

CHAPITRE 4

L'APPROCHE STATISTIQUE DE L'ÉVALUATION DE LA SÉCURITÉ DYNAMIQUE DU RÉSEAU ÉLECTRIQUE DE TRANSPORT

4.1 Introduction

Ce présent chapitre décrit l'approche statistique adoptée pour mettre en lumière la relation forte et directe du transit actif ainsi que la marge de sécurité associée, des corridors du réseau de transport, avec les angles de phase de tension des barres à 735kV.

On débutera par un bref aperçu sur les limites de capacités de charges et le lien théorique et pratique entre le transit actif et les angles de phase de tension. Puis, nous mettons en évidence l'existence d'une relation forte entre les écarts angulaires de phases de tensions d'une part et d'une liaison, aussi forte qu'on ne peut ignorer, entre ces angles et les marges de sécurité ou les transits de puissance active des couloirs du réseau de transport d'une autre part. Puis, par une investigation détaillée de tous les prédicteurs, en considérant les saisons, les mois, les jours de semaine, les heures de pointe et hors pointe, nous caractérisons les signatures d'un comportement typique et nous identifions les événements atypiques. Enfin, nous concluons ce chapitre par la calibration statistique (*baselining*) des prédicteurs pour un fonctionnement normal du réseau électrique, et de ce fait la potentialité de ces indicateurs de contribuer à la surveillance de la sécurité du réseau de transport.

Tout au long du reste de ce document, afin de simplifier notre exposé, nous utilisons l'expression de l'angle de phase de tension pour désigner l'écart angulaire de cette tension par rapport à celle de Boucherville (le poste D01).

4.2 Fondements théoriques

Le fonctionnement normal d'un réseau électrique est assujetti à des conditions initiales données et fixées en avance, à savoir les tensions, les moyens de compensation réactive, les sources de production et la topologie des lignes. En pratique, avant et après tout évènement,

il faut toujours maintenir les tensions à des valeurs précises. De plus, on cherche continuellement à maximiser le transit de puissance (P), quelles que soient les contraintes de stabilité. La stabilité d'un tel réseau, dont l'état initial est donné, est interprétée en limite de transit de puissance. Plus c'est stable, mieux on est capable de transiter plus de puissance. De même, une augmentation (ou diminution) des limites de transit indique une augmentation (ou diminution) de la stabilité. Dès que l'on change ces conditions de fonctionnement, les limites de transit sont modifiées et elles sont directement indiquées à l'opérateur. Ainsi, l'indicateur de la stabilité est traduit directement par la limite de transit. Dans un tel cas réel, pour augmenter le transit de puissance, on augmente les angles de tension. La stabilité est interprétée par la limite de transit de puissance et cette limite de transit de puissance correspond à des angles limites des tensions précises.

À Hydro-Québec, supposer une relation entre les angles et la stabilité est tout à fait réaliste car, la stabilité est traduite sous la forme de limite de transit de puissance et cette limite de transit de puissance correspond directement à des angles limites des tensions précises, pour des tensions et des réactances de lignes fixées (l'état initial de réseau).

À travers cette section, nous justifions de façon théorique l'utilisation des angles de tension comme prédicteurs des limites, sachant que dans cette étude, les limites sont toujours connues. Nous présentons les relations entre les marges de transit de puissance et les angles de tension. Nous verrons que près des limites, les angles deviennent directement proportionnels au transit de puissance.

4.2.1 Relation entre les limites de puissance et les phaseurs

Considérons la différence des angles de phase entre deux postes à 735 kV. Le premier est le poste d'observation, appelé le poste B et le deuxième est le poste de référence, appelé poste Réf, qui est le poste Boucherville dans notre étude. Le réseau ayant de très grandes valeurs de réactances, l'impédance est représentée par une réactance équivalente X_{Th} . Nous pouvons représenter le réseau entre ces deux postes par le modèle illustré par la Figure 4.1.

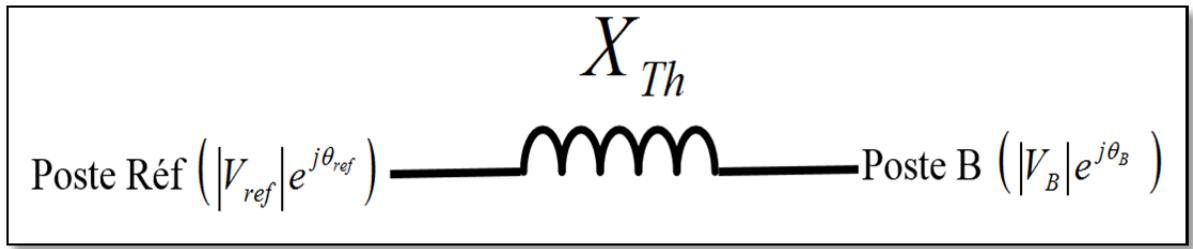


Figure 4.1 Le modèle simplifié du réseau de transport entre deux postes

Le transit de puissance entre ces deux postes est donné par :

$$P_{ref-B} = \frac{|V_{ref}| |V_B|}{X_{Th}} \sin (\theta_B - \theta_{ref}) \quad (4.1)$$

4.2.1.1 Modèles linéaires (faible écart d'angle de tension)

Lorsque la différence des angles est très faible, ce qui est possible lors d'un faible transit ou dans des réseaux très maillés, l'équation (4.1) se simplifie en :

$$P_{ref-B} \approx \frac{|V_{ref}| |V_B|}{X_{Th}} (\theta_B - \theta_{ref})$$

$$P_{ref-B} \cong F(V_i) \Delta\theta_{ref-B}$$

Ainsi, la relation entre puissance et différence d'angles devient alors directement linéaire. Dans cette situation, l'angle de la tension est alors directement relié à la puissance transisée. On peut noter que lorsque l'amplitude de la tension change peu, ce qui est le cas dans la plupart des réseaux, plus le transit augmente plus l'angle de la tension augmente. Cette relation directe de proportionnalité explique le fort potentiel d'utiliser les angles de tension, ceux-ci reflétant directement la puissance. C'est un indice fort qu'il peut exister une relation linéaire entre les deux.

Considérons une condition de réseau quelconque. Il existe alors une limite de puissance, pas nécessairement connue, qui peut se traduire directement en termes de différence d'angles limite :

$$\Delta\theta_{ref-Lim} \equiv \left(\frac{X_{Th}}{|V_{ref}| |V_{Lim}|} \right) P_{ref-Lim}$$

ou autrement,

$$\Delta\theta_{ref-Lim} \equiv \left(\frac{1}{F(V_i)_{Lim}} \right) P_{ref-Lim}$$

Comparons l'angle de tension limite avec l'angle de tension actuel, du même réseau, mais pour un transit de puissance plus faible. En considérant que les amplitudes de tension sont très similaires dans les deux situations, nous avons alors

$$\Delta\theta_{ref-B} - \Delta\theta_{ref-Lim} \equiv \left(\frac{1}{F(V_i)} \right) (P_{ref-B} - P_{ref-Lim})$$

Sachant que,

$$\begin{aligned}\Delta\theta_{ref-B} &= \theta_B - \theta_{ref} \\ \Delta\theta_{ref-Lim} &= \theta_{Lim} - \theta_{ref} \\ \Delta\theta_{ref-B} - \Delta\theta_{ref-Lim} &= \theta_B - \theta_{Lim} \\ \Delta\theta_{ref-B} - \Delta\theta_{ref-Lim} &= \Delta\theta_{Lim-B}\end{aligned}$$

Alors,

$$\Delta\theta_{Lim-B} \equiv \left(\frac{1}{F(V_i)} \right) \Delta P_{Lim-B} \quad (4.2)$$

$$\Delta\theta_{Lim-B} \cong \left(\frac{-1}{F(V_i)} \right) Mr \quad (4.2)$$

Où Mr est la marge de transit,

$$Mr = P_{ref-Lim} - P_{ref-B},$$

Selon (4.2), on peut observer que plus la marge de transit diminue plus l'angle augmente et vice-versa.

Ainsi, dans un réseau où les angles de tension varient peu, il existe toujours une relation directe entre les transits et les angles de tension. Les valeurs varient selon la configuration du réseau (la réactance équivalente Thévenin) et les niveaux de tension. Deux variables qui habituellement changent peu dans un réseau (car la tension est maintenue dans un intervalle strict).

4.2.1.2 Modèle non-linéaire (indépendamment des écarts angulaires de tension)

Considérons des écarts d'angle de tension plus importants et regardons la différence entre le transit et la limite de transit dans un réseau pour une condition d'opération quelconque. Nous avons :

$$\Delta P_{Lim-B} = P_{ref-B} - P_{ref-Lim} \quad (4.3)$$

En utilisant l'expression du transit de puissance en fonction de l'angle de tension (4.1) et en considérant que près des limites les tensions sont similaires, l'expression (4.3) devient

$$\Delta P_{Lim-B} = \left(\frac{|V_{ref}| |V_{Lim}|}{X_{Th}} \right) (\sin(\Delta\theta_{ref-B}) - \sin(\Delta\theta_{ref-Lim}))$$

$$\Delta P_{Lim-B} = F(V_i) (\sin(\Delta\theta_{ref-B}) - \sin(\Delta\theta_{ref-Lim})) \quad (4.4)$$

Sous forme générale,

$$\sin(\theta_1) - \sin(\theta_2) = 2 \cos\left(\frac{\theta_1 + \theta_2}{2}\right) \sin\left(\frac{\theta_1 - \theta_2}{2}\right),$$

Ainsi l'expression (4.4) s'écrit,

$$\Delta P_{Lim-B} = 2 F(V_i) \cos\left(\frac{\Delta\theta_{ref-B} + \Delta\theta_{ref-Lim}}{2}\right) \sin\left(\frac{\Delta\theta_{Lim-B}}{2}\right), \quad (4.5)$$

Nous pouvons observer que la relation entre la différence des transits et l'angle de tension est en général plus complexe. Loin des limites de transit, cette relation n'est plus linéaire, c'est pourquoi l'utilité des angles de tension peut paraître moins évidente ou plus difficile. Toutefois, elle se simplifie grandement près de la limite de transit. Lorsque le transit s'approche de la limite, l'angle de tension s'approche de sa valeur limite :

$$\Delta\theta_{ref-B} \equiv \Delta\theta_{ref-Lim}$$

et leur différence se réduit à :

$$\Delta\theta_{Lim-B} = \theta_{ref-B} - \theta_{ref-Lim} \equiv 0$$

Donc,

$$\sin\left(\frac{\Delta\theta_{Lim-B}}{2}\right) \equiv \frac{\Delta\theta_{Lim-B}}{2}$$

et par suite l'expression (4.5) s'écrit,

$$\Delta P_{Lim-B} \equiv F(V_i) \cos(\Delta\theta_{ref-Lim}) \Delta\theta_{Lim-B}$$

D'où, on obtient :

$$\Delta\theta_{Lim-B} \equiv \frac{\Delta P_{Lim-B}}{F(V_i) \cos(\Delta\theta_{ref-Lim})}$$

$$\Delta\theta_{Lim-B} \equiv \frac{-Mr}{F(V_i) \cos(\Delta\theta_{ref-Lim})} \quad (4.6)$$

On retrouve bien une relation linéaire entre la marge de transit et l'angle de tension. Même dans le modèle non-linéaire, il existe bel et bien une zone de linéarité. On peut comparer l'expression (4.6) avec l'expression (4.2) du modèle linéaire, du paragraphe précédent. On peut noter qu'elles sont identiques lorsque l'angle limite de tension $\Delta\theta_{ref-Lim}$ est faible.

En effet, lorsque

$$\Delta\theta_{ref-Lim} \equiv 0$$

$$\cos(\Delta\theta_{ref-Lim}) \equiv 1$$

Ainsi, utiliser la simple relation linéaire entre la puissance et l'angle de tension est acceptable lorsque la valeur limite de l'angle de tension est faible. Ce n'est pas nécessairement le cas pour le réseau Hydro-Québec, un réseau radial qui comporte de longue distance et d'important transit. Dans ce cas, près des limites, en termes de marge de transit :

$$\Delta\theta_{ref-B} \equiv \Delta\theta_{ref-Lim} - \frac{Mr}{F(V_i) \cos(\Delta\theta_{ref-Lim})} \quad (4.7)$$

D'après l'expression (4.7), nous pouvons constater que près des limites de transit, même dans le cas non linéaire, il existe bel et bien une proportionnalité entre les angles et les transits. Les coefficients ne sont plus seulement dépendants des tensions et des réactances mais aussi du cosinus de l'angle de tension limite (une variable inconnue initialement).

Ainsi dans un réseau en général, il y existe un lien non linéaire entre les angles de tension et les transits de puissance. Ce lien redevient linéaire (proportionnel) lorsqu'on s'approche des limites de transit.

En conclusion, prendre comme hypothèse, que les angles de tension sont directement reliés au transit de puissance est d'autant plus vrai qu'on s'approche des limites de transit ou que la marge de transit diminue. Autrement dit plus la marge de transit diminue plus l'angle de tension se comporte comme les transits de puissance.

4.2.2 L'évaluation de la capacité de charge des lignes de transport à haute tension et/ou des corridors de transport

L'étude analytique de la capacité de charge d'une ligne de transport à haute tension ou d'un corridor de transport a été réalisée par (Gutman, 1988; Gutman, Marchenko et Dunlop, 1979). Le fondement théorique des auteurs se basait sur une représentation simplifiée du réseau électrique, en représentant la ligne et les paramètres du réseau par deux équivalents de Thevenin aux extrémités de cette dernière (Voir la Figure 4.2).

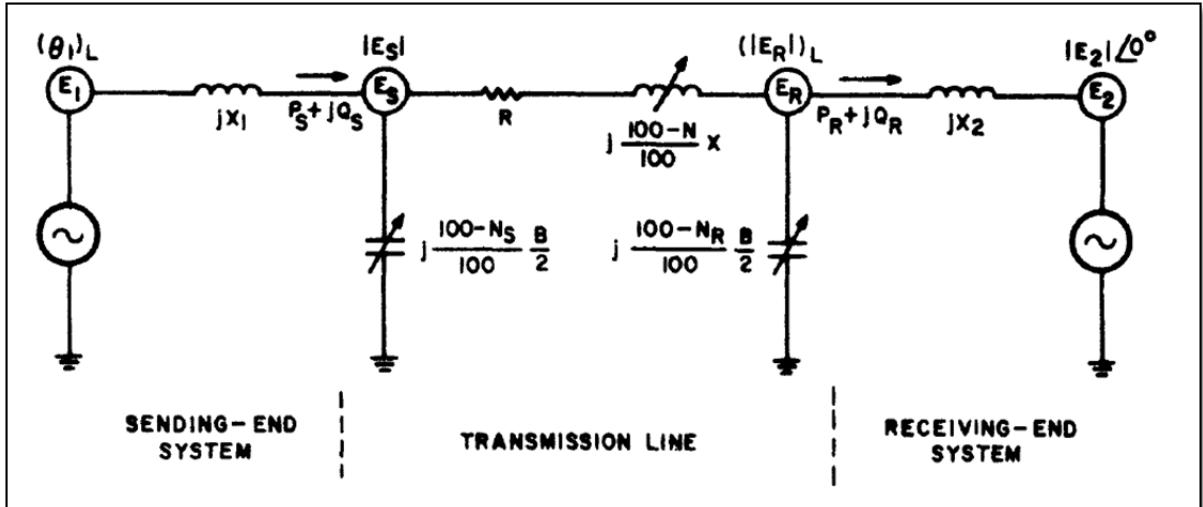


Figure 4.2 La ligne de transport avec les deux équivalents de Thevenin
Tirée de Gutman, Marchenko et Dunlop (1979, p. 2)

Cette simple équivalence a ensuite été utilisée pour évaluer la capacité de charge avec les critères de puissance thermique, de chute de tension et de la marge de stabilité en régime permanent (basé sur les angles aux bouts du système).

Ce modèle analytique permet la reproduction des courbes St. Clair existantes (CLAIR, 1953), ainsi que leur extension à d'autres niveaux de tension, longueurs de ligne et d'autres paramètres du réseau (Gutman, 1988). La conclusion tirée par cette étude analytique est que les lignes sont thermiquement limitées quand ils sont moins de 50 miles de long. De plus, entre 50 et 190 miles, la charge est limitée en raison de la chute de tension aux bornes de la ligne. Enfin, pour les lignes longues de 190 miles, le chargement est limité en raison de la marge de la stabilité à l'état stable dans tout le système. Les auteurs de la référence (Kay et al., 1982) explicitent encore ces conclusions.

Par la Figure 4.3 et la Figure 4.4, on reprend un résultat du travail (Gutman, 1988) accompagné du commentaire des auteurs par rapport aux angles des extrémités de tout le système.

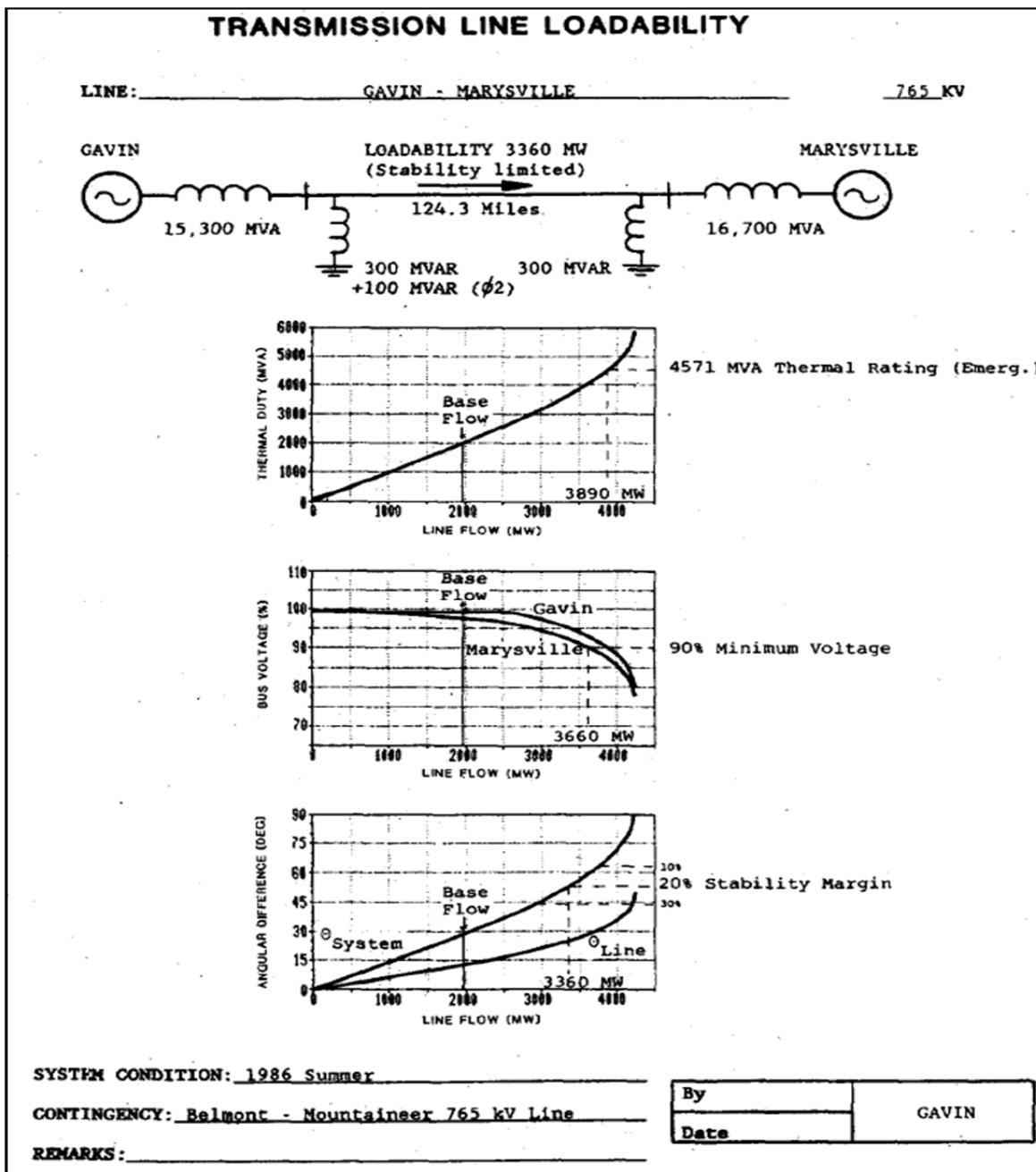


Figure 4.3 Les caractéristiques de capacité de charge d'une ligne de transport
Tirée de Gutman (1988, p. 4)

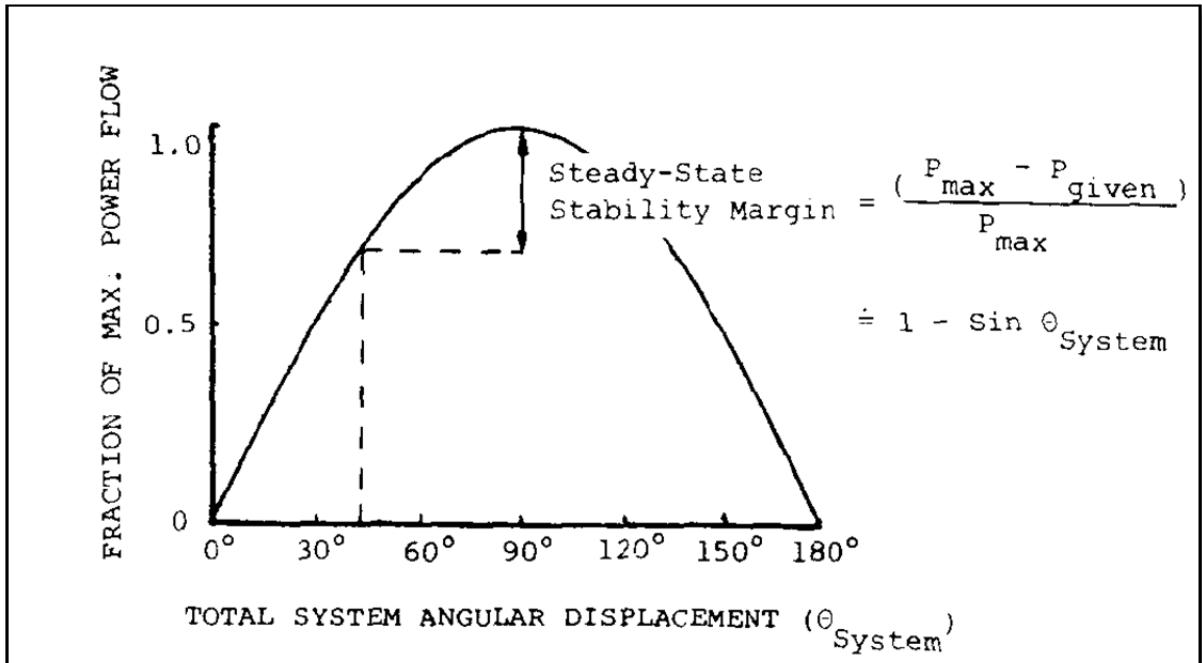


Figure 4.4 La marge de stabilité à l'état stable
Tirée de Gutman (1988, p. 5)

The angular displacement curve shows the phase angle difference across the line terminals (θ_{Line}) and across the complete system model (θ_{System}) as a function of the line flow. The (θ_{System}) and (θ_{Line}) curves are similar only when the system reactance is negligible relative to the line reactance (e.g., when a long transmission line connects two strong systems). In general, the (θ_{System}) is significantly greater than (θ_{Line}) and, therefore, only (θ_{System}) should be used in assessing the system steady-state stability. System stability is usually measured in terms of the available stability margin, which is defined as the margin between the maximum possible power flow (P_{\max}) and a given power flow (P_{given}) expressed in percent of P_{\max} . (Gutman, 1988, p. 1430).

En se basant sur le même concept montré par la Figure 4.2, d'autres travaux de recherche ont été effectués pour évaluer les marges de stabilité en utilisant les données des synchrophaseurs (Sauer, 2011; Sauer et Dominguez-Garcia, 2014). Les auteurs ont proposé une conjecture à base de la limite de stabilité de 90 degrés aux bouts des deux équivalents Thevenin. Si un système s'approche de la capacité de charge maximale (écroulement de tension), au moins

une ligne affichera cette condition par un calcul approprié de l'équivalent de Thevenin aux bouts du système de la ligne de transport. Par conséquent, l'examen de l'angle entre les deux équivalents de Thevenin pour chaque ligne offrira un degré de proximité à la stabilité statique (steady-state stability).

4.3 Les données de l'EMS-SCADA

Nous utilisons dans notre investigation les angles de phase de tensions des 37 postes identifiés sur la Figure 4.5. Le Tableau 4.1 et le Tableau 4.2 récapitulent toutes les variables utilisées dans notre travail de recherche. Nous visualisons les séries chronologiques de 2010 des angles de phase de tension de quelques postes de transports sur la Figure 4.6.a et les séries chronologiques des marges de sécurité des cinq couloirs de réseau de transport ainsi que la série chronologique de la demande sur la Figure 4.6.b.

La marge de sécurité d'un corridor, décrite dans le Tableau 4.2, est la différence entre la limite de transit de ce corridor et le transit de puissance active le long de ce dernier, normalisée par le besoin global ou la charge globale.

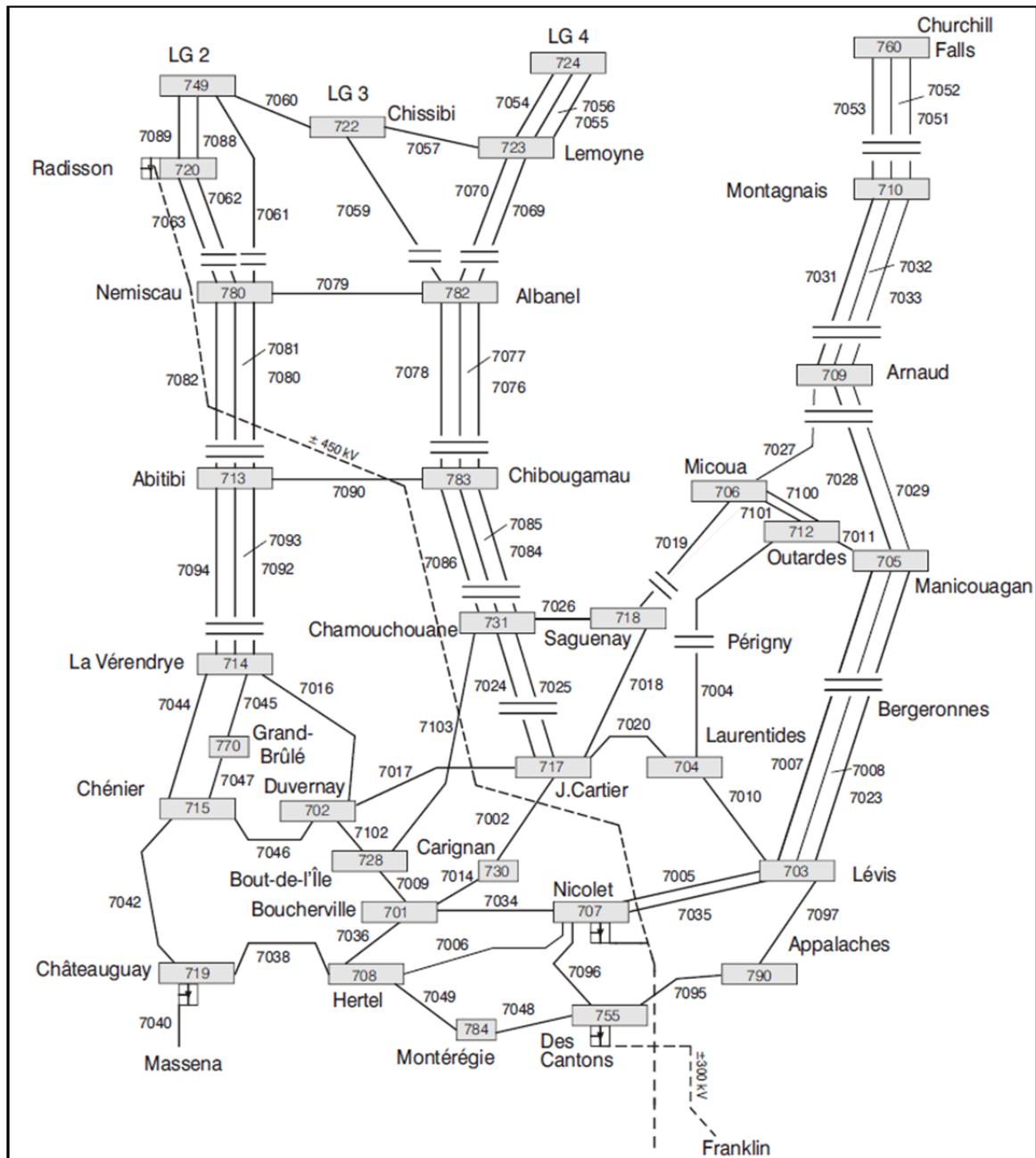


Figure 4.5 Les postes du réseau de transport

Tableau 4.1 Description des variables d'entrée

N	Variables d'entrée	Description des variables
1	D01	Angle de tension (degrés) du poste Boucherville
2	D02	Angle de tension (degrés) du poste Duvernay
3	D03	Angle de tension (degrés) du poste Lévis
4	D04	Angle de tension (degrés) du poste Laurentides
5	D05	Angle de tension (degrés) du poste Manicouagan
6	D06	Angle de tension (degrés) du poste Micoua
7	D07	Angle de tension (degrés) du poste Nicolet
8	D08	Angle de tension (degrés) du poste Hertel
9	D09	Angle de tension (degrés) du poste Arnaud
10	D10	Angle de tension (degrés) du poste Montagnais
11	D13	Angle de tension (degrés) du poste Abitibi
12	D14	Angle de tension (degrés) du poste Laverendrye
13	D15	Angle de tension (degrés) du poste Chénier
14	D17	Angle de tension (degrés) du poste Jacques-Cartier
15	D18	Angle de tension (degrés) du poste Saguenay
16	D19	Angle de tension (degrés) du poste Chateauguay
17	D20	Angle de tension (degrés) du poste Radisson
18	D22	Angle de tension (degrés) du poste Chissibi
19	D23	Angle de tension (degrés) du poste Lemoyne
20	D24	Angle de tension (degrés) du poste Tilly
21	D30	Angle de tension (degrés) du poste Carignan
22	D31	Angle de tension (degrés) du poste Chamouchouane
23	D49	Angle de tension (degrés) du poste LG2
24	D50	Angle de tension (degrés) du poste LG3
25	D51	Angle de tension (degrés) du poste LG3
26	D52	Angle de tension (degrés) du poste LG3
27	D55	Angle de tension (degrés) du poste Des Cantons

N	Variables d'entrée	Description des variables
28	D60	Angle de tension (degrés) du poste Churchill Falls
29	D64	Angle de tension (degrés) du poste LG4
30	D65	Angle de tension (degrés) du poste LG4
31	D66	Angle de tension (degrés) du poste LG4
32	D70	Angle de tension (degrés) du poste Grand Brûlé
33	D80	Angle de tension (degrés) du poste Némiscau
34	D82	Angle de tension (degrés) du poste Albanel
35	D83	Angle de tension (degrés) du poste Chibougamau
36	D84	Angle de tension (degrés) du poste Montérégie
37	D90	Angle de tension (degrés) du poste Appalaches

Tableau 4.2 Description des variables de sortie

N	Variables de sortie	Description des variables
1	TransitMW_BJN	Transit de puissance active pour l'axe Baie-James Nord
2	TransitMW_BJS	Transit de puissance active pour l'axe Baie-James Sud
3	TransitMW_CHU	Transit de puissance active pour l'axe Churchill Falls
4	TransitMW_MANQ	Transit de puissance active pour l'axe Manic-Québec
5	TransitMW_Sud	Transit de puissance active pour l'axe Sud
6	LimiteMW_BJN	Limite de transit sur l'axe Baie James Nord
7	LimiteMW_BJS	Limite de transit sur l'axe Baie James Sud
8	LimiteMW_CHU	Limite de transit sur l'axe Churchill Falls
9	LimiteMW_MQ	Limite de transit sur l'axe Manic Québec
10	LimiteMW_Sud	Limite de transit sur l'axe Sud
11	Mr_BJN	Marge de sécurité de l'axe Baie-James Nord
12	Mr_BJS	Marge de sécurité de l'axe Baie-James Sud
13	Mr_CHU	Marge de sécurité de l'axe Churchill Falls
14	Mr_MANQ	Marge de sécurité de l'axe Manic-Québec
15	Mr_Sud	Marge de sécurité de l'axe Sud

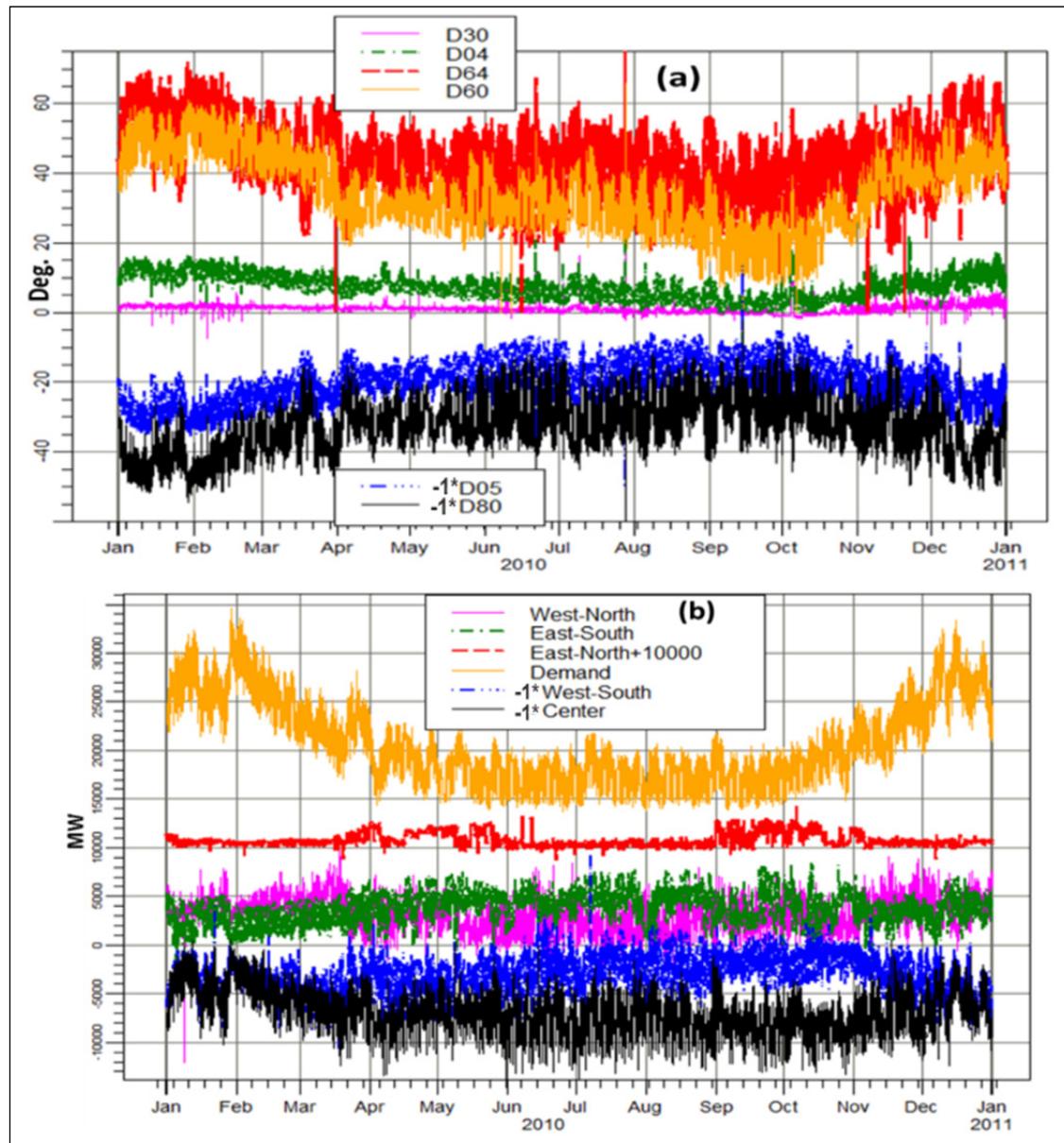


Figure 4.6 Séries chronologiques à partir des archives CILEX et EMS

- (a) Les écarts angulaires des tensions de quelques barres
- (b) Les transits actifs à travers les cinq interfaces

Les limites de transit varient selon les équipements et la configuration des lignes et fluctuent au gré des conditions du réseau. La charge varie selon l'année : elle est forte de décembre à février puis atteint un creux autour des mois de juin-juillet. Le transit de puissance et la marge de sécurité ne sont ainsi pas constants et varient selon l'année.

Sur la Figure 4.6, on note les points suivants :

- les valeurs des angles fluctuent et reflètent les sollicitations quotidiennes de la demande et des transits du réseau ;
- les angles sont plus élevées durant la période d'hiver et diminuent durant la période estivale tel que reflété par le niveau de charge globale ;
- les changements d'angle jusqu'à 40 degrés sont observés. Ils indiquent les changements de distances électriques : moins il y a de lignes présentes plus l'impédance (la distance électrique) augmente ;
- les postes situés près de la charge ont tendance à avoir moins de variation d'angle que ceux éloignées ;
- la demande augmente durant les périodes d'hiver et diminuent durant les périodes d'été;
- les transits suivent en général le profil de la demande ;
- des fluctuations périodiques de 1 000 MW de la demande sont observées ;
- des variations de transit de puissance sont très variées selon leur localisation ;
- on peut observer des écarts variant entre quelques centaines de MW jusqu'à près de 1 000 MW ;
- certains axes sont plus sollicités et plus variables que d'autres ;
- le transit de l'interface Sud est le plus élevé et le plus fluctuant tandis que l'interface Churchill (CHU) est le plus faible et le plus constant.

4.4 Les variations des angles de phase de tension des postes à 735kV

4.4.1 Diagramme en boite (*Box plot*)

Les diagrammes en boite, Figure 4.7, sont idéaux pour comparer une même caractéristique sur plusieurs séries statistiques. Ils font apparaître immédiatement la distribution et la variabilité des données, et permettent d'identifier les observations aberrantes.

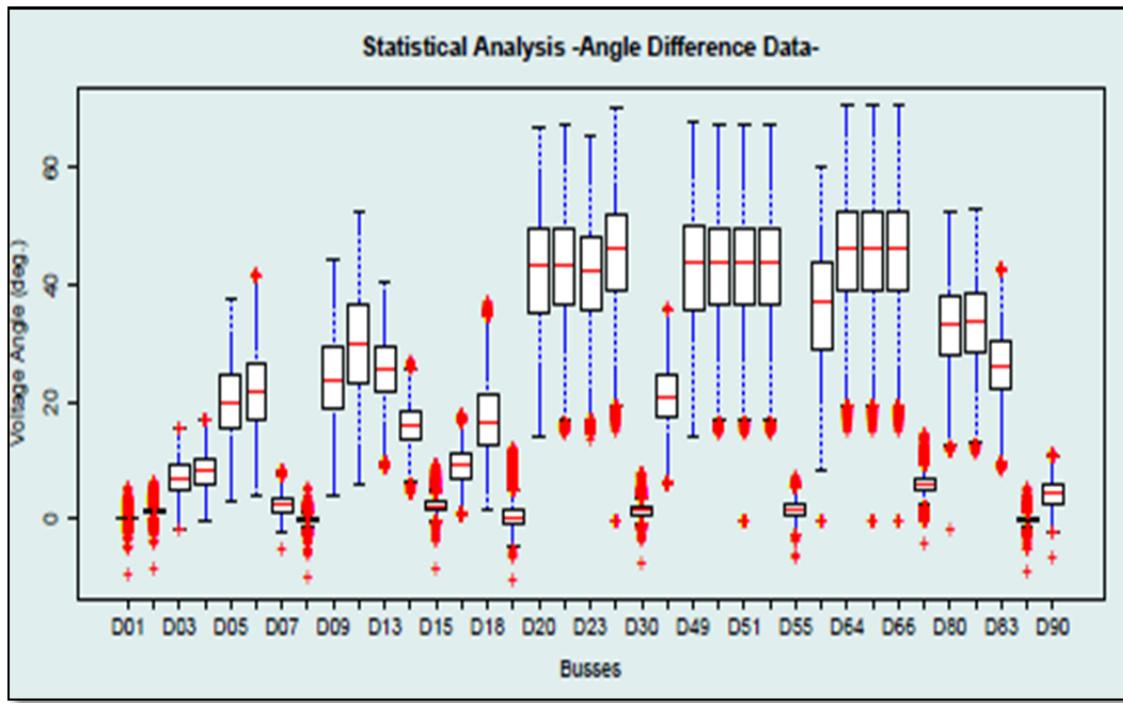


Figure 4.7 Box plot des angles de phase de tension des barres à 735kV
(175 000 échantillons)

Les caractéristiques du diagramme de boite sont :

Le centre (la médiane) : désigné par le trait gras à l'intérieur du rectangle sur la Figure 4.7. Il indique si la distribution est symétrique ou non. Dans le cas où le centre est au milieu du rectangle, la distribution est alors symétrique ;

La largeur de la boite : elle informe sur la dispersion. Les extrémités du rectangle sont les quartiles supérieur et inférieur. Le rectangle couvre donc l'écart interquartile, autrement dit la

distribution de 50% des valeurs de l'ensemble des données. Ce concept de la largeur élimine l'influence des valeurs aberrantes ;

Les deux « moustaches » inférieure et supérieure : dites valeurs adjacentes. Elles sont représentées par les points en bleu, sur la Figure 4.7. Elles sont basées sur 1,5 fois la hauteur de la boîte (Le Guen, 2001) ;

Les valeurs extrêmes : sont appelées aussi valeurs atypiques, exceptionnelles, ou *outliers*. Elles sont situées au-delà des valeurs adjacentes. Elles sont représentées par des marqueurs en rouge.

4.4.2 Interprétation

Selon les largeurs des boîtes, il est clair que, plus on s'éloigne des zones de charges plus les largeurs de boîtes se démarquent. Ces largeurs traduites par les interquartiles interprètent la dispersion des données : la variabilité des angles de tension.

En se référant aux Figure 4.5 et Figure 4.7, on peut facilement distinguer au moins deux zones spécifiques. L'une est caractérisée par des variations d'angle de phase très étendues allant de 10 degrés à 15 degrés. Cette zone renferme les postes qui se situent plus au nord, proche des zones de production.

La deuxième zone est caractérisée par des boîtes très étroites, indiquant ainsi une dispersion très serrée. Cette zone comprend les postes situés près de la charge. Leurs interquartiles sont moins de 5 degrés.

Un autre point c'est que dans la zone où les variations sont importantes, ces dernières s'accentuent encore lorsque les liens électriques sont moins forts. Comme par exemple, les variations d'angle de tension de LG2 (D49) sont plus étendues que les variations d'angle de tension de Churchill (D60).

Ainsi, la variation d'angle de tension des postes est intimement liée à la position géographique par rapport à la charge et au lien électrique.

Indépendamment de la connaissance exacte ou non du réseau, les boîtes fournissent rapidement les postes ayant des caractéristiques similaires. Nous pouvons les catégoriser selon la valeur de l'angle de phase qui leur est associé. Cette partition en catégories permet de classer les postes selon leur variabilité et leur proximité géographique relative à la charge. Ainsi, cette catégorisation est une caractérisation des postes selon leur degré de sollicitation.

En voici quelques exemples :

1. Les postes D80 (Némiscau) et D82 (Albanel) ont des valeurs et des variations parfaitement identiques ce qui signifie qu'ils sont voisins et voient ainsi le même environnement électrique ;
2. Le poste D10, situé à Montagnais, à un niveau de proximité équivalent à celui du poste D82 (Albanel) mais à des écarts plus importants équivalents à celui du poste D49 (LG2) centrale éloignée. Le poste Montagnais est ainsi proche de la charge comme le poste Albanel, même si celui-ci n'est pas dans le même axe, mais électriquement, sa caractéristique est typique d'un poste éloigné proche d'une centrale. Ce qui est tout à fait vérifié dans la réalité car le poste Montagnais est directement dépendant d'une centrale éloignée ;
3. Les postes D15 (Chénier), D30 (Carignan), D55 (Des Cantons) et D70 (Grand-Brûlé) ont tous une variabilité similaire sans qu'ils soient nécessairement situés à la même distance électrique (près de la charge de Montréal). Cela signifie qu'ils sont sollicités de la même façon, ce qui est bien le cas car ils sont situés dans la même région où se trouve un corridor important de transport ;

4. Les postes D19 (Chateauguay), D07 (Nicolet) et D90 (Appalaches) sont près de la charge avec des écarts similaires. Ils sont associés à des zones où on retrouve des échanges avec les réseaux voisins ;
5. Les postes Lévis (D03), Jacques-Cartier (D17) et Laurentides (D04) sont situés dans la même région de Québec. Au niveau de l'angle de tension, l'ordre est plutôt Lévis, suivi de celui des Laurentides et enfin Jacques-Cartier. C'est le reflet direct du nombre de liens selon le corridor Est de transport, où le poste Jacques-Cartier a le moins de lien avec cet axe. Au niveau des écarts, ils sont identiques entre eux car ils répondent directement au niveau de charge de la région de Québec.

Nous pouvons aussi regarder les résultats selon les valeurs d'angle de tension de la médiane (trait rouge) sur la figure 4.7. Nous présentons, ci-dessous, quelques exemples :

1. Entre 0 et 20 degrés : le poste situé le plus au Nord est le poste D05 (Manicouagan) et celui le plus au Sud est le poste D19 (Chateauguay). Les postes ayant les boites les plus larges sont les postes D05 (Manicouagan) et D18 (Saguenay) ;
2. Entre 20 et 40 degrés : les postes situés le plus au Nord sont le poste D60 (Churchill Falls) suivi de D82 (Albanel) et D80 (Némiscau). Les plus au Sud sont les postes D31 (Chamouchouane) et D06 (Micoua). Les postes ayant les boites les plus larges sont les postes D60 (Churchill Falls) et D10 (Montagnais) suivi du poste D09 (Arnaud). Les plus faibles sont les postes D13 (Abitibi) et D63 (Chibougamau) ;
3. Entre 40 et 60 degrés : les postes situés le plus au Nord sont le poste D23 (Le Moyne) suivi du poste LG4 (D64-66). Les plus au Sud sont le poste D23 (Le Moyne) suivi de D20 (Radisson) et D22 (Chissibi). Les postes ayant les boites les plus larges sont les postes D20 (Radisson) et D49 (LG2). Les plus faibles valeurs sont similaires sur un grand nombre de poste (ex LG4 (D64-66), LG3 (D50-52) D22 (Chissibi) D24 (Tilly)).

Mettons en évidence, selon l'ordre croissant de l'angle de tension, les constatations suivantes :

1. Près de la charge : Le poste Chateauguay suivi de D84 (Montérégie), D08 (Hertel), Boucherville, Carignan, Des Cantons, Duvernay, Chénier, Nicolet, Appalaches, Grand-Brûlé, Lévis, Laurentides, Jacques-Cartier, La Vérendrye, Saguenay, *Manicouagan*, Chamouchouane, *Micoua*, *Arnaud*, Abitibi, Chibougamau, *Montagnais*, Némiscau-Albanel, *Chuchill*, Le Moyne, Radisson-Chissibi, LG2-LG3, et LG4-Tilly ;
2. Au-delà des régions de charges, le réseau d'Hydro-Québec comporte deux axes en V. À l'est c'est l'axe Manic-Québec, à l'ouest c'est l'axe Baie-James (nous avons indiqué en souligné l'axe Baie-James Ouest).

La position géographique de l'axe Manic-Québec n'est pas reflétée au niveau de la distance électrique. Chaque poste de ce corridor se trouve décalé au sud par rapport aux postes de l'axe Baie-James. Par exemple, le poste Churchill est situé à une latitude entre le poste Le Moyne au Sud et le poste Radisson au nord mais électriquement parlant, le poste Churchill se décale plus au Sud et est plutôt situé entre le poste Albanel au Sud et le Poste Radisson au Nord.

Dans l'axe Baie-James, les axes situés à l'Ouest et à l'Est sont aux mêmes latitudes ce n'est pas le cas au niveau des distances électriques. L'axe Baie-James Ouest est situé plus près de la charge de Montréal que l'axe Baie-James Est qui est affecté par la charge de Québec. Par exemple, géographiquement le poste Abitibi, à l'Ouest de l'axe est situé entre les postes, à l'Est, de Chibougamau et d'Albanel mais sa distance électrique est située au Sud du Poste Albanel. Géographiquement, le poste La Vérandrye est situé entre le poste Laurentides et le poste Saguenay et électriquement il est situé au Sud de ce dernier poste.

4.5 La dépendance mutuelle des postes

Un réseau électrique étant un ensemble de lignes et de postes où transite de la puissance, le comportement à différents postes peut être similaire pour des raisons de proximité électrique.

Il est utile de déterminer leur influence entre eux afin de capturer rapidement les caractéristiques électriques et dynamiques d'un réseau, ce qui permet de réduire le nombre de variables à considérer. Pour connaître le degré d'interdépendance des postes entre eux, nous utilisons la notion de corrélation. Plus le facteur de corrélation entre deux variables d'entrée est grand, plus forte est la relation linéaire entre elles. Avec le logiciel **R**, nous mesurons ce lien. Nous résumons les résultats de corrélation obtenus par le biais de la Figure 4.8.

4.5.1 Matrice de corrélation des angles de phase de tension des postes à 735kV

Les résultats de corrélation dans la Figure 4.8 sont donnés par la forme et la couleur des éléments graphiques :

- un cercle parfait indique qu'il n'y pas (ou très peu) de corrélation entre les variables ;
- plus les ellipses sont aplatis, plus la corrélation entre deux variables est grande ;
- les couleurs utilisées donnent un autre indice de la force de corrélation. L'intensité de la couleur est maximale (noir) pour une corrélation parfaite, et minimale (blanc) s'il n'y a aucune corrélation. Des nuances de rouge sont utilisées pour des corrélations négatives et de bleu pour des corrélations positives ;
- les lignes droites sur la diagonale de la parcelle indiquent que toute variable est parfaitement corrélée avec elle-même.

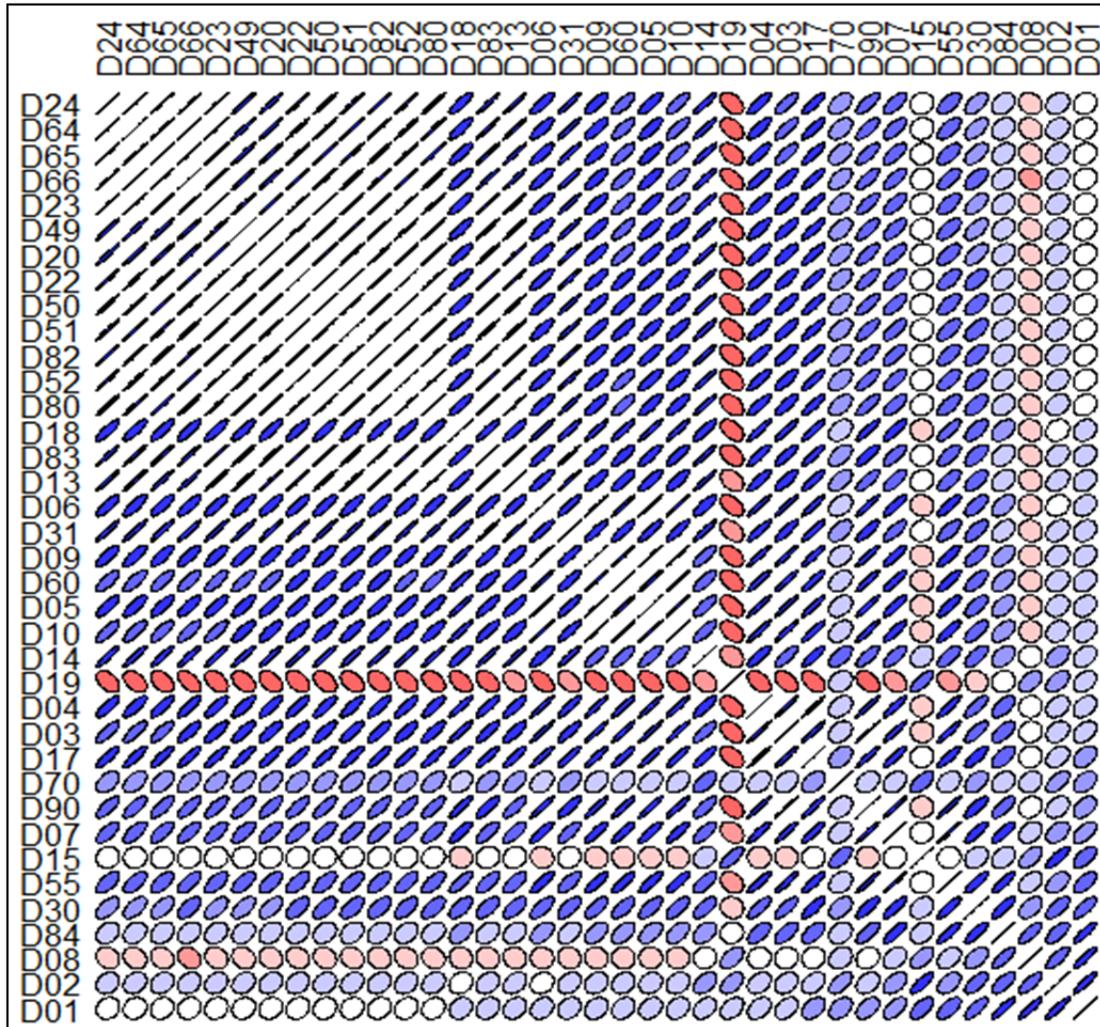


Figure 4.8 Matrice de corrélation des angles de phase de tension des 37 postes

4.5.2 Interprétation

L'analyse des résultats de corrélation fournit un portrait des influences électriques entre les différents postes électriques. On constate bien que :

1. Tout poste est corrélé avec ses postes voisins et le niveau de corrélation dépend de sa source principale d'alimentation. Par exemple, le poste Des Cantons (D55) suit ses postes adjacents à savoir, le poste Montérégie avec un facteur de 0,87, le poste Nicolet avec un facteur de corrélation de 0,98 et le poste Appalaches avec un facteur de 0,97. La

corrélation augmente dans la direction de l'Est, là où les liens sont les plus forts et d'où provient le transit de puissance ;

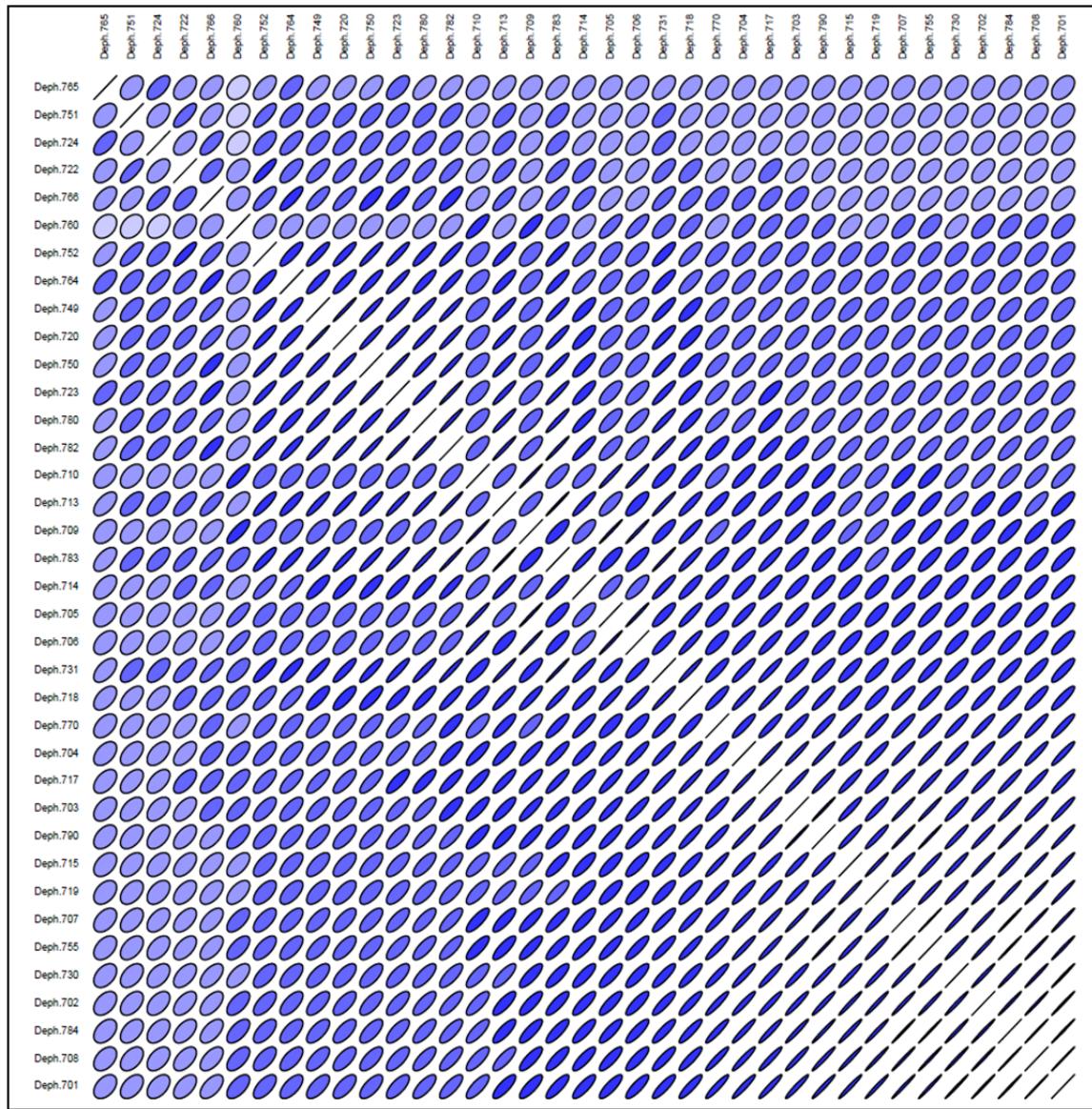
2. Le poste Chénier est le moins corrélé avec tous les autres postes, sauf avec le poste Duvernay et Chateauguay. Son facteur de corrélation, en valeur absolue, avec les autres postes est entre 0,03 et 0,4 et avec le poste Duvernay et Chateauguay est de 0,79 et 0,73 respectivement. C'est un indice de son importance et de sa position particulière dans le réseau. Il est situé dans une zone où se trouve des transits de puissance important ayant différentes origines le rendant moins dépendant de postes particuliers ;
3. Pour le poste Boucherville, dont la phase est la référence, son facteur de corrélation avec tous les autres postes passe de 0,13 à 0,85, comme le résume le Tableau 4.3. Les deux postes Chénier et Boucherville sont ceux qui ont le niveau de charge le plus élevé. Le tableau indique qu'il est le moins relié avec les postes dans l'axe Baie-James Nord et est le plus relié avec les postes Duvernay, Hertel, Montérégie et Carignan. Notons que le poste Nicolet, pourtant adjacent n'est pas dans la liste tandis que le poste Montérégie est dans la liste alors qu'il n'est pas adjacent. Les corrélations les plus fortes respectent les directions habituelles des transits de puissance ;
4. Les postes Chateauguay et Hertel sont ceux qui réagissent en opposition avec l'ensemble ou un sous-ensemble du réseau. Si on note qu'ils sont situés au-delà des charges principales et du poste de référence (Boucherville), l'opposition n'est que le reflet de leur emplacement ;
5. Le poste Grand-Brûlé est le poste baromètre, universel, du réseau. C'est le poste qui suit similairement toutes les variations du réseau. Étant situé dans un corridor avec le transit le plus élevé, alimentant des charges élevées, il subit aussitôt des changements lors des perturbations sur les autres axes de transit ;

6. Les deux postes Chamouchouane et Saguenay sont géographiquement très proches mais sont influencés par des axes de transit différents. Le poste Saguenay suit plus solidement l'axe Manic-Québec, au Nord, ainsi que la zone de charge de la région de Québec situé au Sud. Le poste Chamouchouane suit plus solidement les postes dans l'axe Baie-James au Nord, que les postes situés au Sud (La Vérandrye et Saguenay). De plus, il est légèrement plus corrélé avec le poste La Vérandray situé géographiquement beaucoup plus loin que le poste Saguenay.

Tableau 4.3 La corrélation du poste Boucherville avec le reste des postes à 735kV

Deph	D66	D24	D64	D65	D23	D52	D49	D20
Cor	0,13	0,15	0,15	0,16	0,17	0,17	0,18	0,18
Deph	D22	D50	D51	D82	D80	D18	D83	D13
Cor	0,18	0,18	0,18	0,19	0,20	0,22	0,22	0,24
Deph	D31	D06	D09	D05	D19	D60	D10	D14
Cor	0,25	0,26	0,29	0,29	0,32	0,32	0,32	0,32
Deph	D04	D03	D70	D17	D90	D07	D15	D55
Cor	0,37	0,38	0,38	0,38	0,43	0,53	0,55	0,57
Deph	D30	D84	D08	D02	D01			
Cor	0,76	0,84	0,86	0,87	1,00			

Des blocs communs de postes similaires peuvent être regroupés. Le premier bloc comprend les postes situés dans la région de la Baie-James (les postes Nemiscau et Albanel et ceux situés au nord). Un deuxième bloc comprend les postes Micoua et Manic ainsi que ceux situés dans l'axe vers Churchill (Montagnais, Arnaud et Churchill). Un dernier bloc est celui des postes situés dans la région de Québec (Laurentides, JCA et Levis).



Ces blocs indiquent les zones influençant le poste Boucherville, soit la zone de charge principale de Montréal et moindrement la zone de charge de Québec. On peut déduire qu'électriquement des défauts dans les postes avec de forte corrélation seront vus au poste Boucherville tandis qu'un défaut situé dans la région de Baie-James ne sera pas perçu à moins d'actions effectuées dans la zone de charge.

La non-corrélation identifiée entre quelques postes dans la Figure 4.8, ne signifie pas nécessairement qu'il n'y a pas de relations entre ces postes. En effet, il s'agit juste de l'absence de lien linéaire. Cependant une relation non linéaire entre eux n'est pas exclue. Effectivement, pour les 37 variables d'entrée, nous avons calculé la différence d'angles aux instants (t) et ($t+1$) et nous avons évalué la corrélation entre ces 37 nouvelles séries chronologiques. Les résultats de cette étape sont illustrés dans la Figure 4.9. On constate bien qu'il y a de liens entre les postes de la boucle du sud : D15, D02, D01, D30, D07, D03, D90, D55, D84, D08 et D19.

4.6 Les angles de phase de tension versus les transferts et les marges de puissance

Après l'exploration par corrélation des relations entre les angles de phase des tensions, nous exploitons le même outil pour découvrir le niveau de lien entre ces différents angles et les transits, ainsi que les marges de puissance à travers les couloirs du réseau de transport. Les résultats de la

Figure 4.10 montrent que de nombreux angles sont corrélés positivement à plus de 95% avec les transits de puissance des interfaces. Le peu d'angle de phase corrélés négativement sont ceux qui ont des facteurs de corrélation négatifs avec les autres angles de phases, tels que D08, D19. Concernant l'étude des marges de sécurité pour les interfaces, la plupart des facteurs de corrélation sont négatifs avec les angles de phase, mais ils sont supérieurs à 90%, à l'exception de l'interface Churchill (CHU) où la corrélation est limitée à 50%.

Nous explorons les mêmes séries mais cette fois-ci, elles sont étendues sur 3 ans par minute, de 2010 à 2012. Les résultats sont aussi bons que pour l'étude d'une année. La Figure 4.11 illustre la corrélation maximale entre les transits, les limites et les marges de puissance des couloirs du réseau de transport avec les angles de tension. Les facteurs de corrélation, concernant l'évaluation du transit de puissance sont supérieurs à 95% pour tous les couloirs sauf pour le couloir CHU qui est inférieur à 90%. Cependant, cette interface fonctionne sous un contrat fixe de long terme, ce qui rend son comportement atypique mais prévisible. Les facteurs de corrélation examinant les marges de sécurité au niveau de toutes les interfaces

sont plus faibles et négatifs. Ils interprètent les relations complexes entre les angles de phases des barres à 735kV et les limites de transferts. Ces dernières sont de nature complexes étant donné que ce sont des fonctions du lieu de contingence, de la topologie du réseau, de la compensation réactive, des modèles de système, des modèles de charge, des modèles de système de contrôle et de plan de puissance, etc.

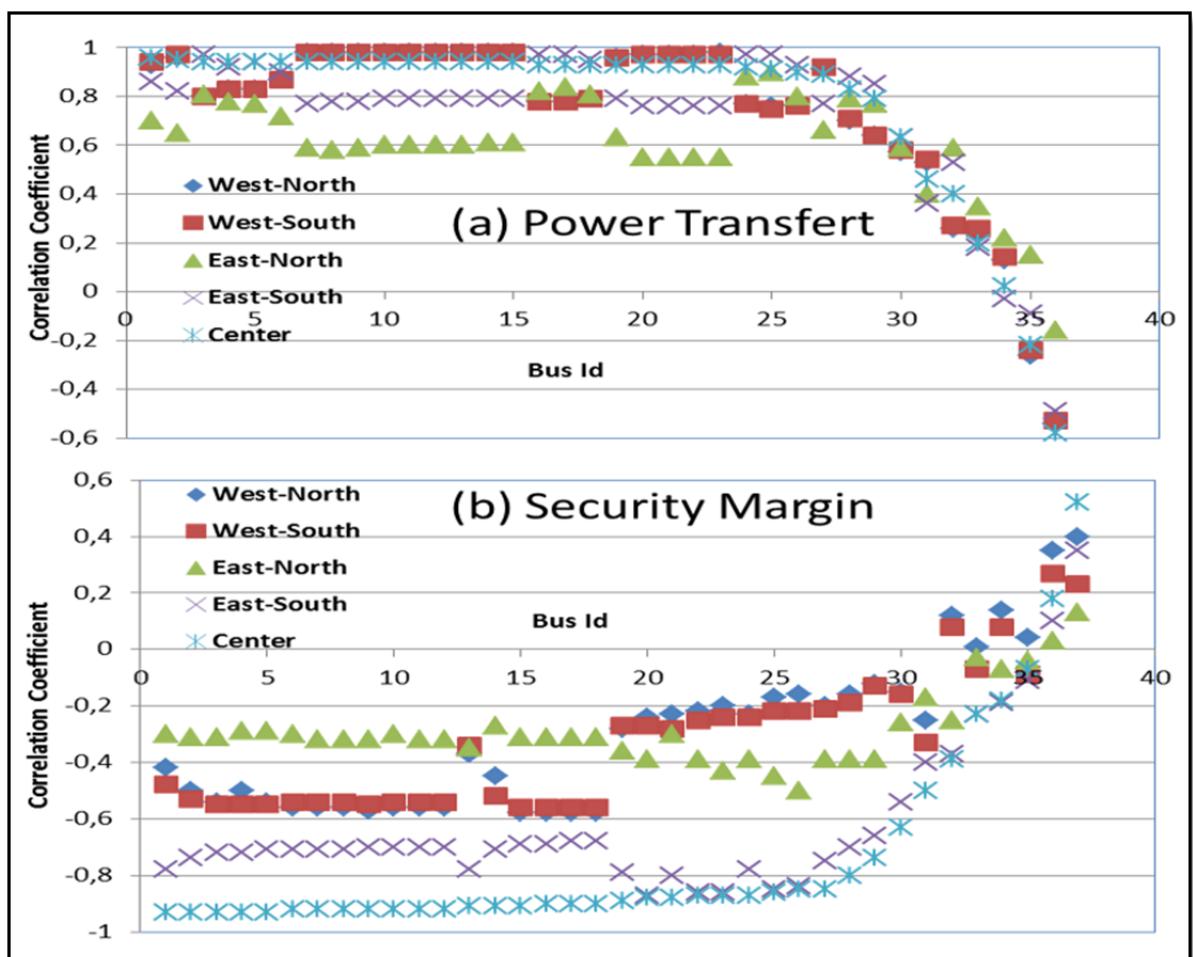


Figure 4.10 La corrélation des angles de phase des barres avec : (a) Les transferts de puissance, (b) Les marges de sécurité

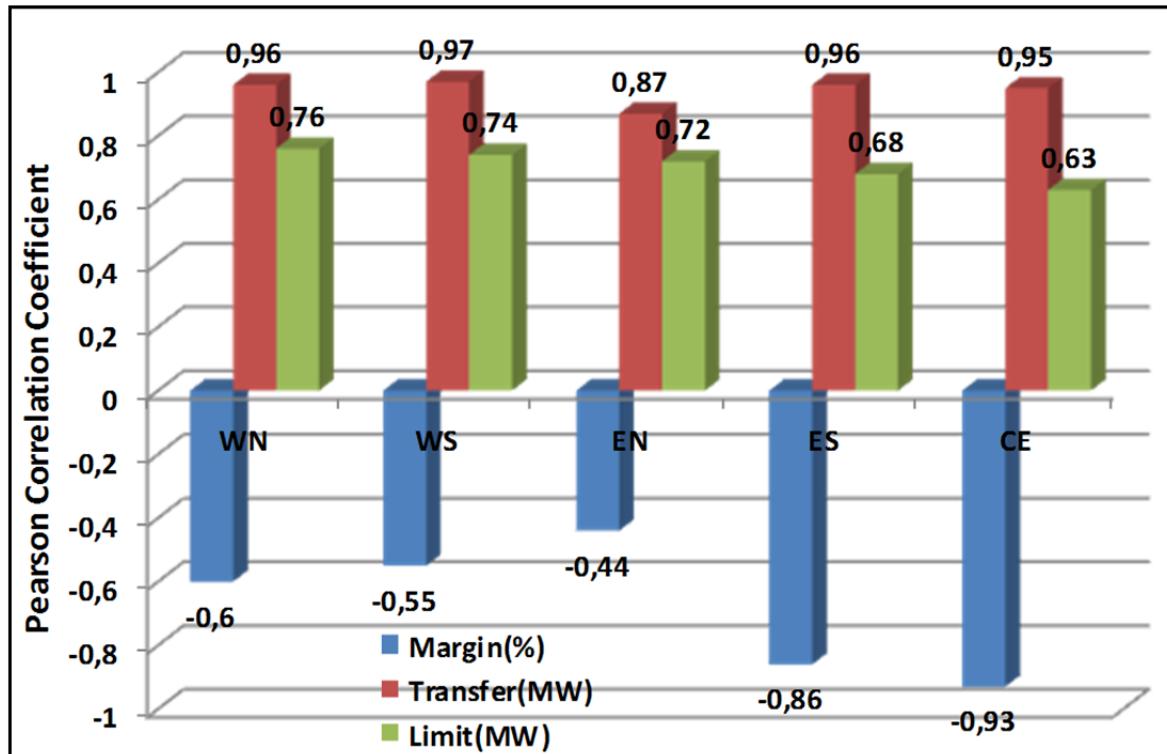


Figure 4.11 Le niveau maximum de corrélation des angles avec les transferts, les marges et les limites de puissance sur 3 ans par minutes

4.7 Les prédicteurs de la stabilité dynamique

À partir des résultats de l'analyse des corrélations entre les angles de tensions des postes entre eux-mêmes et puis avec les transits de puissance et les marges de sécurité des différents couloirs du réseau de transport, nous percevons l'empreinte de la caractéristique dynamique du réseau. Effectivement, les Figure 4.8-

Figure 4.10 mettent en évidence les interrelations dynamiques des angles de phase de tension des différentes barres. Ils reflètent l'ensemble des variations des conditions de fonctionnement du réseau de transport. Empiriquement, les angles constituent la variable qui varie instantanément avec la puissance P , Q et la topologie du réseau.

En plus de l'investigation statistique réalisée, afin de confirmer la légitimation des écarts angulaires comme des prédicteurs de la stabilité dynamique du réseau de transport, nous avons mis côté à côté plusieurs profils de tous les 37 angles de phase de tension avec les 5 transits de puissance active et les 5 marges de puissance active pour comparer leurs aspects. Un cas parmi les autres, le nomogramme illustré par la Figure 4.12 qui interprète les tendances saisonnières de la marge de sécurité à l'interface Sud et l'interface Manic-Québec (MQ) face aux angles de phase de tension du poste Churchill Falls (D60) et du poste LG4 (D64). Nous voyons que, lorsque les angles de phase de ces postes au nord sont élevés, les marges de sécurité des deux interfaces sont faibles et vice-versa.

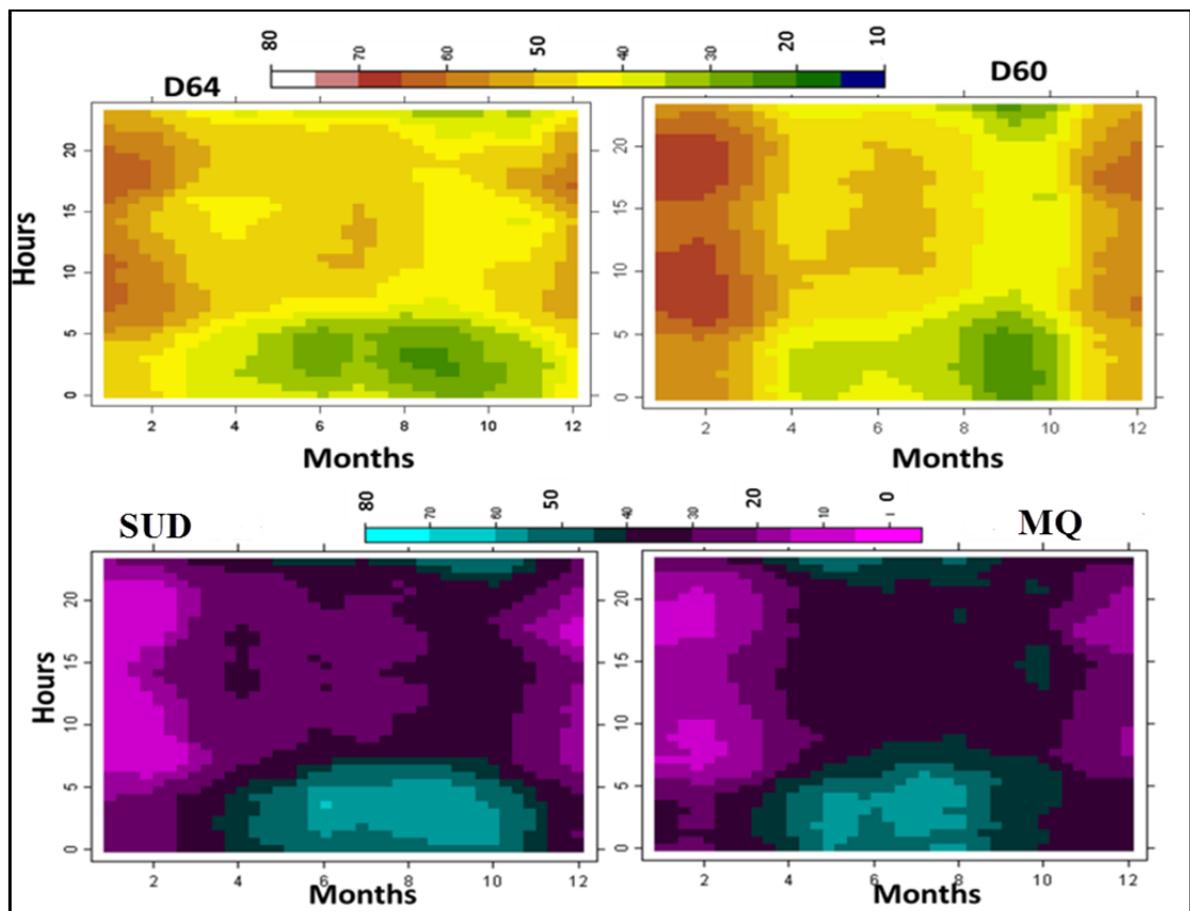


Figure 4.12 Les tendances mensuelles et horaires en 2010 des marges de transferts vs. les angles de phase de tension

Un autre cas où nous visualisons les profils des transferts de puissance et des marges de sécurité en face des angles de phase de tension, le nomogramme sur la Figure 4.13. L'échelle en couleur est l'échelle de transfert de puissance. L'échelle verticale est celle de l'angle D60 et l'échelle horizontale est celle de D64.

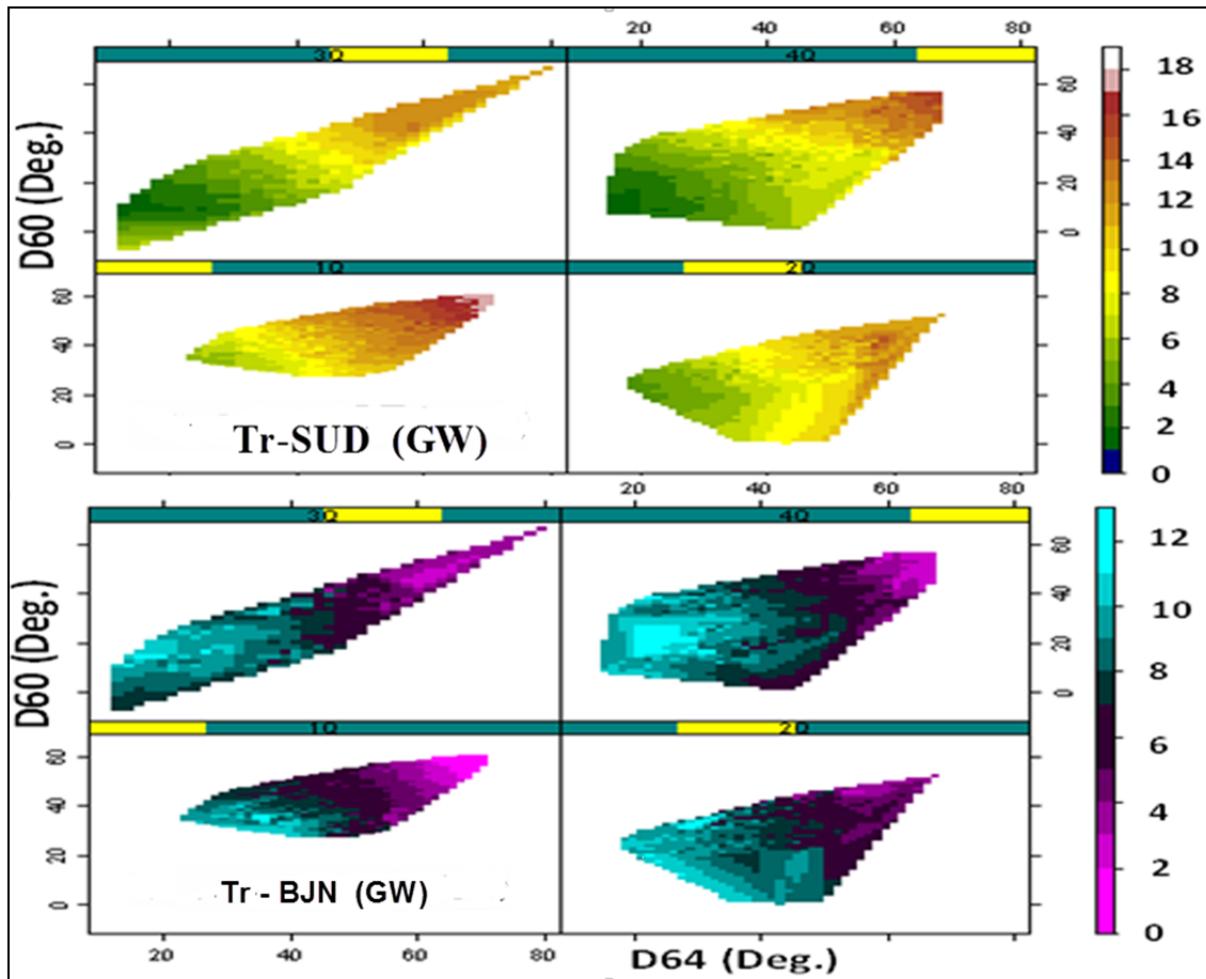


Figure 4.13 Les nomogrammes de transits (en haut) à travers l'interface SUD et (en bas) à travers l'interface BZN à base des phases de 2010

Toutes les allures obtenues et les études de corrélations faites montrent une cohérence générale ou globale entre les transferts et les angles et de même entre les marges de sécurité et les angles. Ce qui nous laisse supposer que les angles de phase de tension peuvent relater ou informer sur les transits de puissance active et aussi sur les marges de sécurité. Ainsi, une

représentation de ces relations complexes avec plus de précision et moins de nuances, nous permet d'intégrer les angles de phase de tension des barres à 735kV dans le processus de surveillance de la sécurité dynamique du réseau.

Pour atteindre le but de rendre les représentations plus claires et plus compréhensibles, nous avons manipulé les données brutes d'une année par minutes, de plusieurs façons. Nous les avons explorées en données par saison, par mois, par jour, par période de pointe et hors pointe. La période de pointe est la période où le réseau d'Hydro-Québec est plus sollicité, soit entre 6h et 9 h du matin et en fin de journée entre 16 h et 20 h, les jours de semaine. C'est pendant ces périodes qu'une pointe, c'est-à-dire une demande exceptionnellement élevée, peut se produire. Par la Figure 4.14, nous pouvons voir les variations des valeurs maximales et minimales quotidiennes tout au long de l'année 2010. L'écart entre les trois grandeurs présentées le long de toute l'année est considérable, ce qui nous amène à utiliser la variation quotidienne pour surveiller la sécurité du réseau. Avec le graphe en boite, nous pouvons mieux voir ces variations du fait que la variabilité des valeurs extrêmes sera mieux distinguée et repérée.

Par la Figure 4.15, nous présentons les comportements de tous les angles de phase, en séparant les deux saisons hiver et été pour une période de novembre 2008 jusqu'à décembre 2012. À partir d'une vue globale, nous constatons une grande différence, de presque 10 degrés, entre les valeurs extrêmes des deux saisons. Sur la Figure 4.16, on note le même décalage entre le mois de janvier 2010 (saison d'hiver) et le mois d'août 2010 (saison d'été), pour quelques angles de phase de tension.

En gardant dans l'esprit cette caractéristique de variabilité des angles de phase entre les mois d'une année, nous avons étudié le degré d'influence des heures de pointe sur cette caractéristique. La Figure 4.17 récapitule la variabilité des angles de phase pendant tous les mois en fonction des heures de pointe et hors pointe. Cette variabilité de l'angle est accompagnée aussi des changements de la marge de sécurité vécus par le réseau en 2010. On note, que les valeurs maximales extrêmes de l'angle pour tous les mois, pour les deux périodes en pointe et hors pointe sont très proches sauf pour les mois de juin et juillet. Ainsi,

lors de ces deux périodes, les valeurs minimales extrêmes de l'angle changent largement. Pendant la période de pointe, et ceci le long de tous les mois, face à la diminution de la marge de sécurité, les médianes de l'angle se décalent considérablement et les variabilités diminuent par rapport aux périodes hors pointe.

Ces deux derniers points peuvent faire l'objet d'une surveillance pour distinguer un fonctionnement normal et alerter dans le cas d'un fonctionnement à risque.

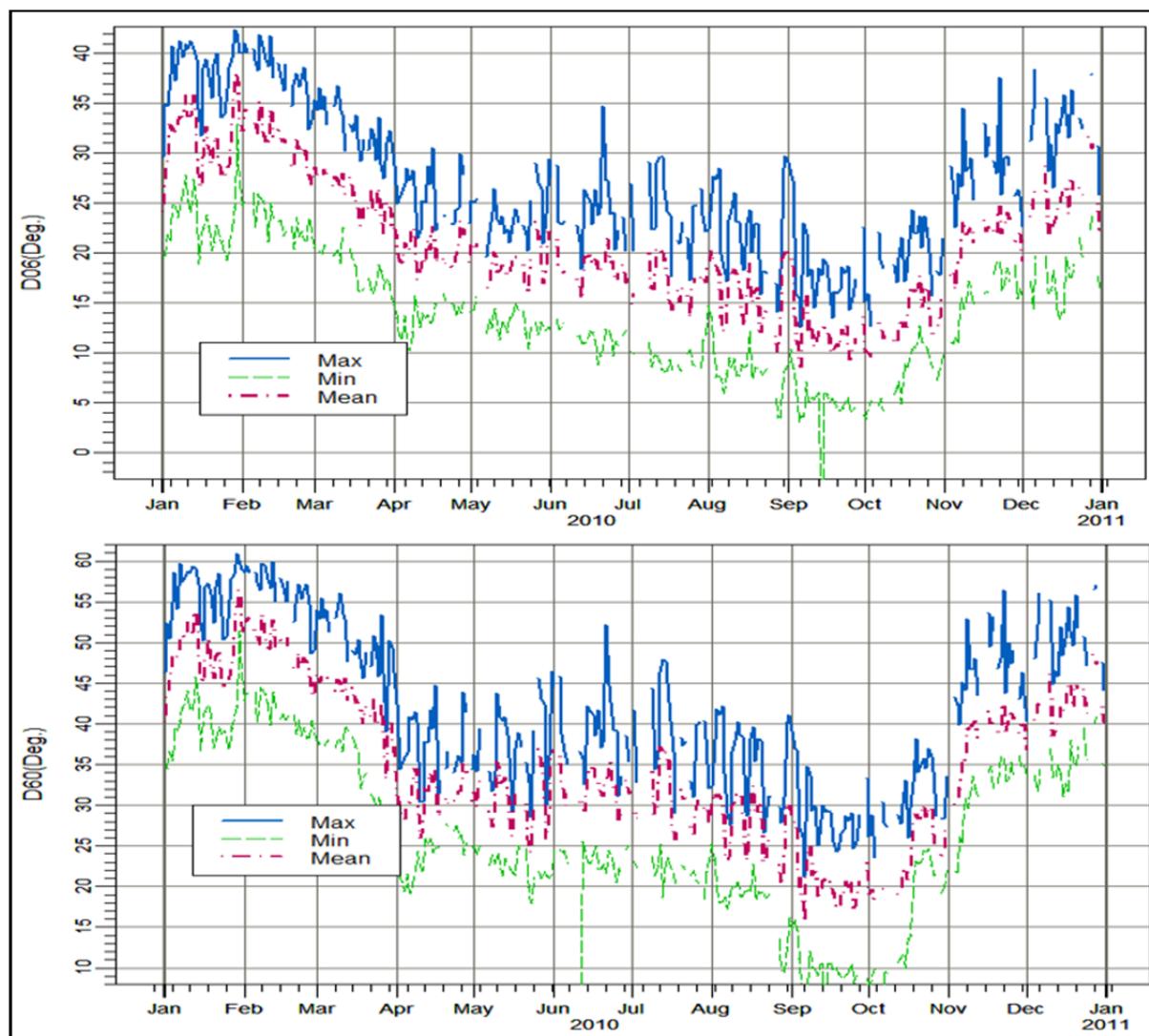


Figure 4.14 Les variations quotidiennes des angles de phase de tension en 2010

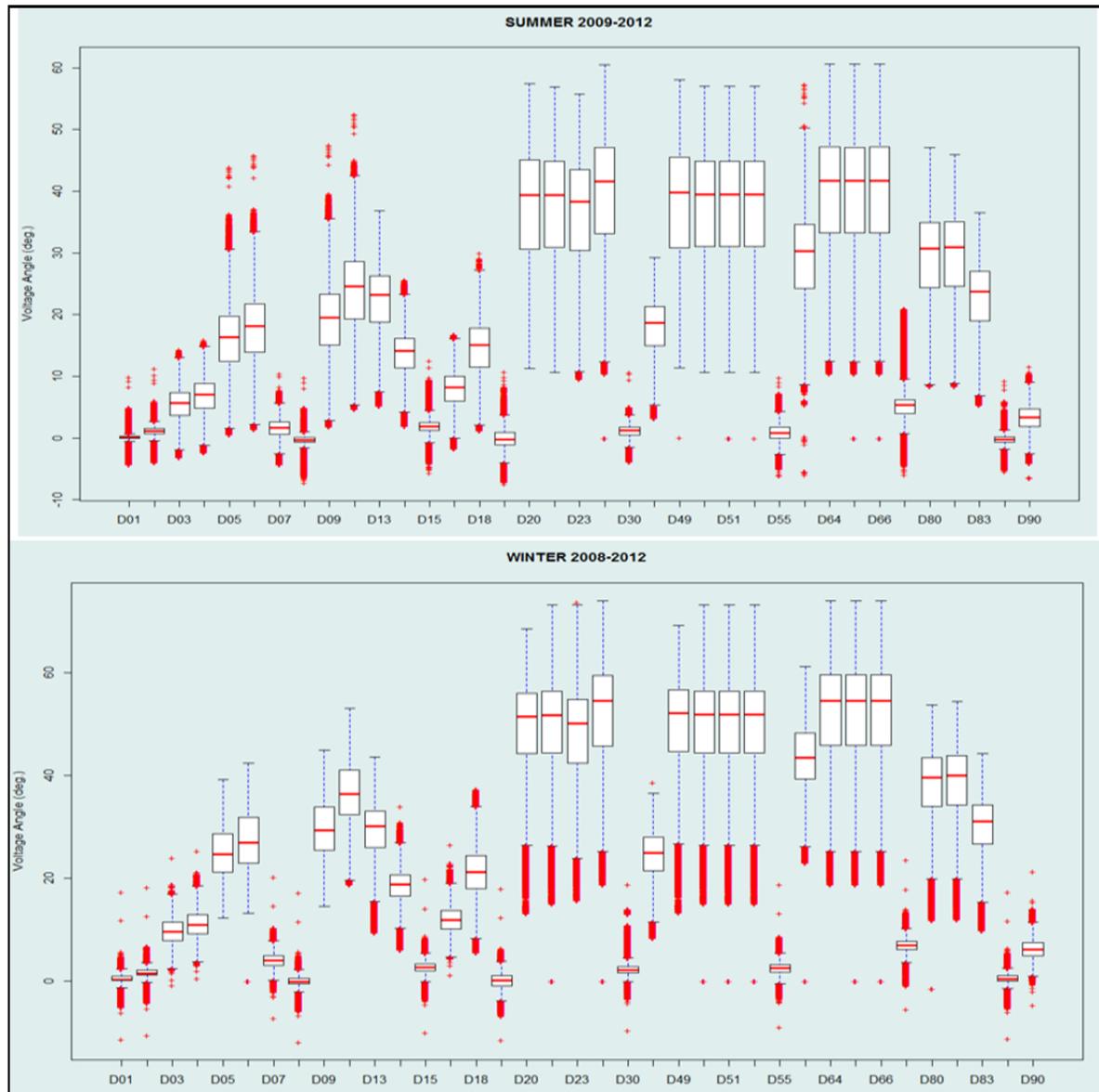


Figure 4.15 Profilage saisonnier des angles de phase de tension

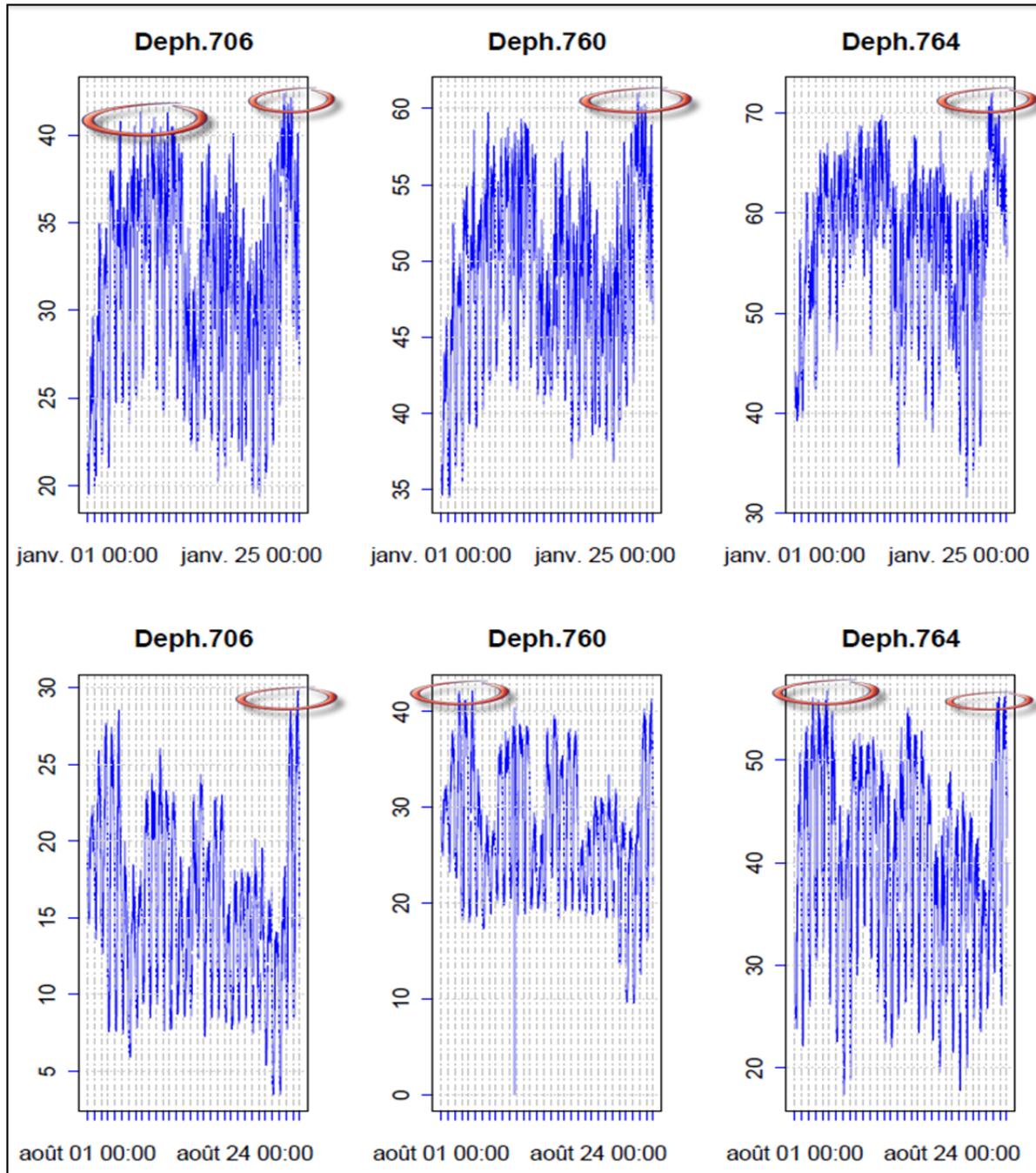


Figure 4.16 Comparaison des profils des angles de phase de tension pendant les mois de janvier et d'août de 2010

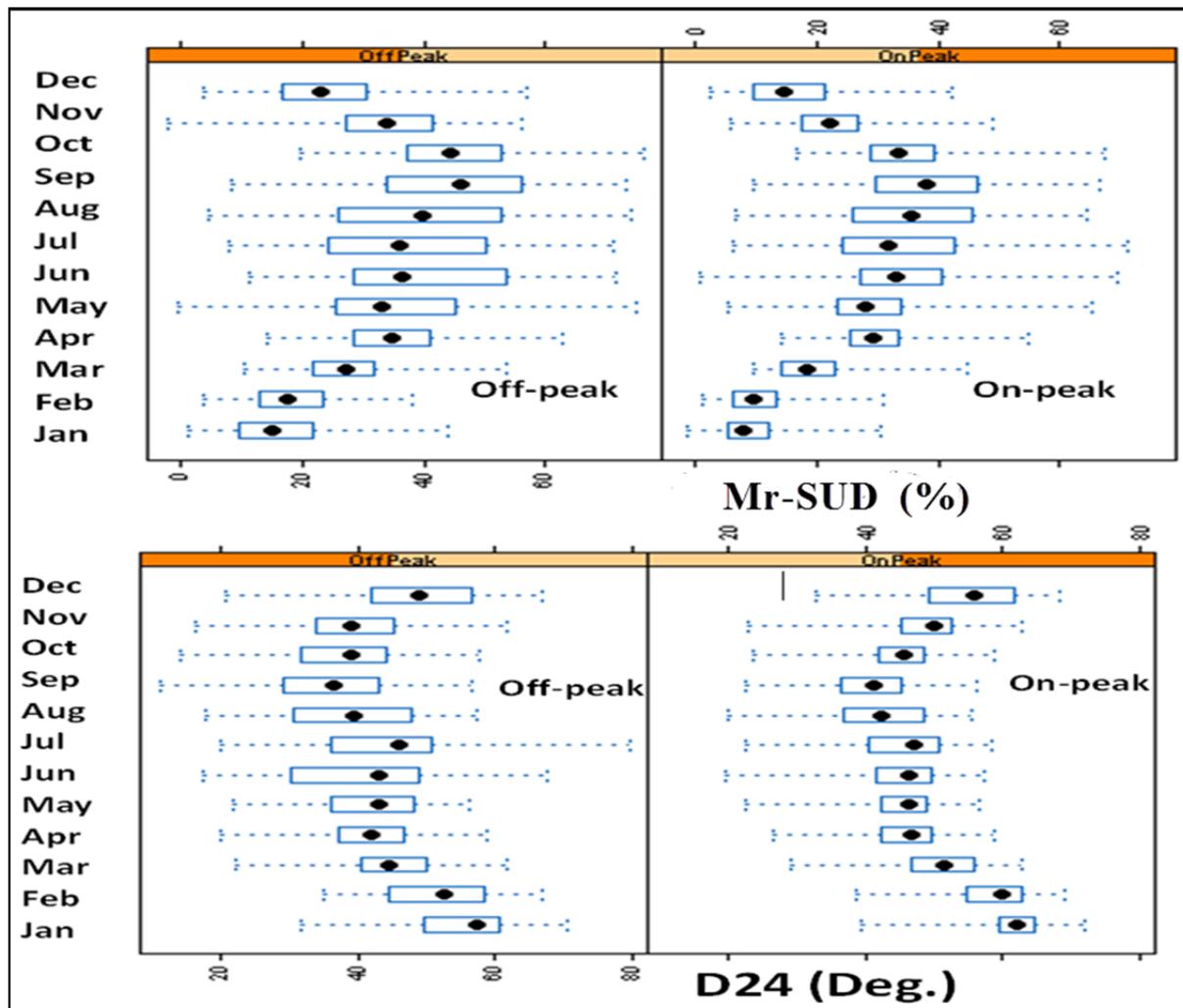


Figure 4.17 Les tendances mensuelles -la marge vis-à-vis l'angle de phase-

4.7.1 L'identification des angles de phase clefs par PAM

Tous au long de notre travail, nous avons analysé le comportement de tous les angles et sous différents profils. Cependant, pour présenter nos résultats, nous avons jugé plus intéressant et plus instructif de présenter seulement quelques-uns. En se basant sur les observations de tous les résultats comme par exemple ceux présenté par les Figure 4.7-Figure 4.9, nous pouvons distinguer des regroupements des angles dont les éléments sont très liés. Ainsi, au lieu de présenter les résultats de la calibration statistique de tous les prédicteurs, nous nous limitons à la présentation des prédicteurs « médoids » ou clefs qui devraient véhiculer les mêmes

informations que l'ensemble des 37 angles, sur la sécurité du réseau. Pour cela, nous utilisons l'instruction PAM (*Partitioning Around Medoids*) du logiciel **R** et qui est disponible aussi dans d'autres logiciels de statistique (Sabine et Mia, 2005). Cet algorithme regroupe les angles autours d'un certain nombre de médoides. Cependant, c'est à l'utilisateur de fixer le nombre de clusters ou de regroupements. Pour estimer le nombre de cluster, nous utilisons l'algorithme iVAT (*Improved Visual Assessment of Cluster Tendency*). Le résultat d'iVAT (Havens et Bezdek, May 2012) présenté par la Figure 4.18 montre deux représentations. La première est la reproduction de la disposition des 37 angles introduits. La deuxième représentation est le regroupement issu où les angles de phases de tension des barres similaires sont contigus. Nous reconnaissions facilement trois groupes dominants. En fixant le nombre de clusters à trois (3) dans l'algorithme PAM, nous obtenons la répartition montrée par la Figure 4.19. Les résultats de cette représentation sont assemblés dans le Tableau 4.4. Cependant, Cette répartition n'est pas unique. On peut aussi distinguer, sur Figure 4.18, en se concentrant sur la couleur bleu, d'autres regroupements autours de la diagonale, allant jusqu'à 10 clusters. Par conséquent, nous pouvons raisonnablement prendre le nombre de regroupement égal à 8.

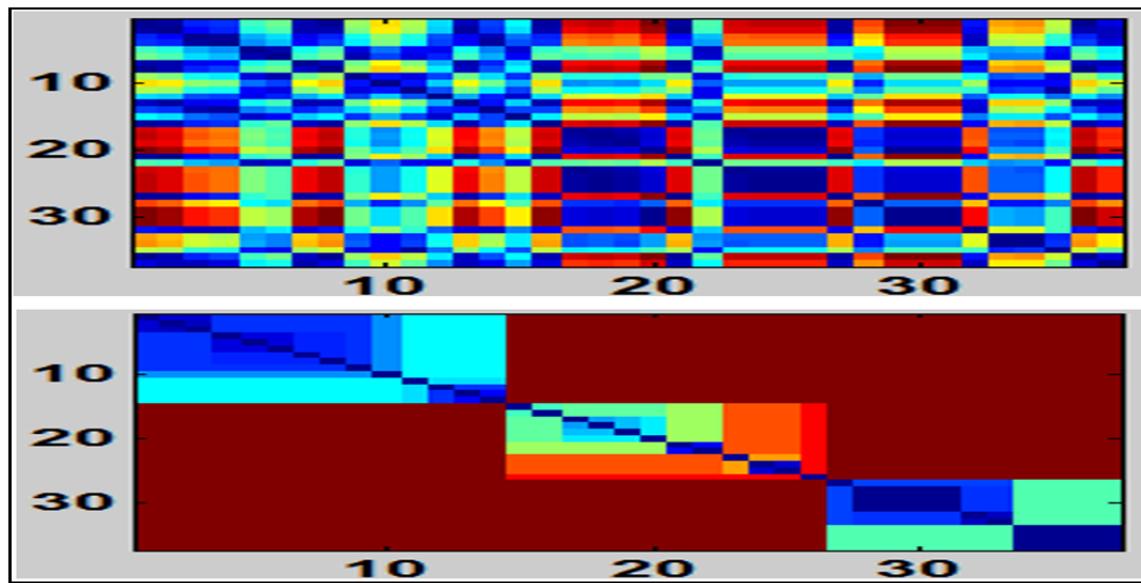


Figure 4.18 La visualisation de la distance de Manhattan utilisée par la matrice de dissimilarité

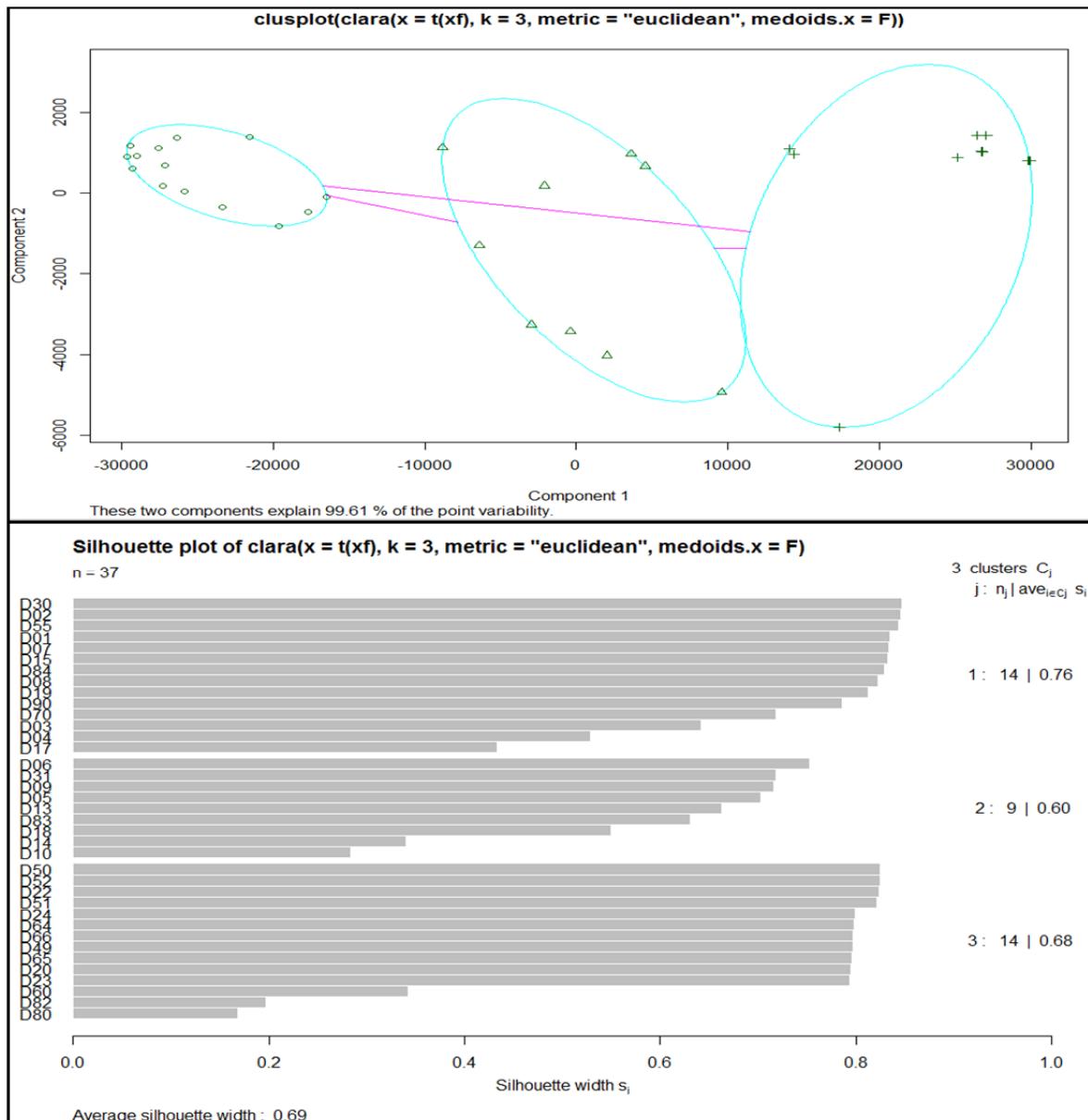


Figure 4.19 La répartition des angles de phases de tension la plus optimale

Tableau 4.4 Les regroupements suite à l'algorithme PAM

Le premier regroupement	Le deuxième regroupement	Le troisième regroupement
Le poste médoid Carignan (D30)	Le poste médoid Micoua (D06)	Le poste médoid LG3 (D50)
Boucherville (D01), Duvernay (D02), Levis (D03), Laurentides (D04), Nicolet (D07), Hertel (D08), Chenier (D15), Jacques- Cartier (D17), Chateauguay (D19), Des Cantons (D55), Grand Brûlé (D70), Montérégie (D84), Appalaches (D90).	Manicouagan (D05), Arnaud (D09), Montagnais (D10), Abitibi (D13), Laverendrye (D14), Saguenay (D18), Chamouchouane (D31), Chibougamau (D83).	Radisson (D20), Chissibi (D22), Lemoyne (D23), Tilly (D24), LG2 (D49), LG3(D51), LG3 (D52), Churchill Falls (D60), LG4 (D64), LG4 (D65), LG4 (D66), Nemiscau (D80), Albanal (D82).

4.7.2 Calibration statistique des angles de phase clefs

En se référant aux résultats obtenus suite à l'investigation détaillée de l'ensemble des prédicteurs, nous retenons le fait que la variabilité saisonnière prime sur la variabilité mensuelle. En outre, par rapport au profit quotidien, la spécificité des périodes de pointe et hors pointe n'est pas insignifiante. Sur la Figure 4.20, les écarts des valeurs des angles de phases entre les saisons d'hiver et d'été sont très significatifs, de même pour les valeurs extrêmes maximales. L'écart des extrémités des courbes de l'hiver et l'été est illustré et il est de 10 à 15 degrés. En période de pointe et hors pointe, nous pouvons surveiller les valeurs minimales extrêmes. Nous pouvons visualiser l'information apportée suite à la séparation de ces deux périodes, en pointe et hors pointe sur les illustrations, de la Figure 4.21 à la Figure 4.26, référant aux quelques barres médiods identifiées par les boîtes vertes sur la Figure 2.1

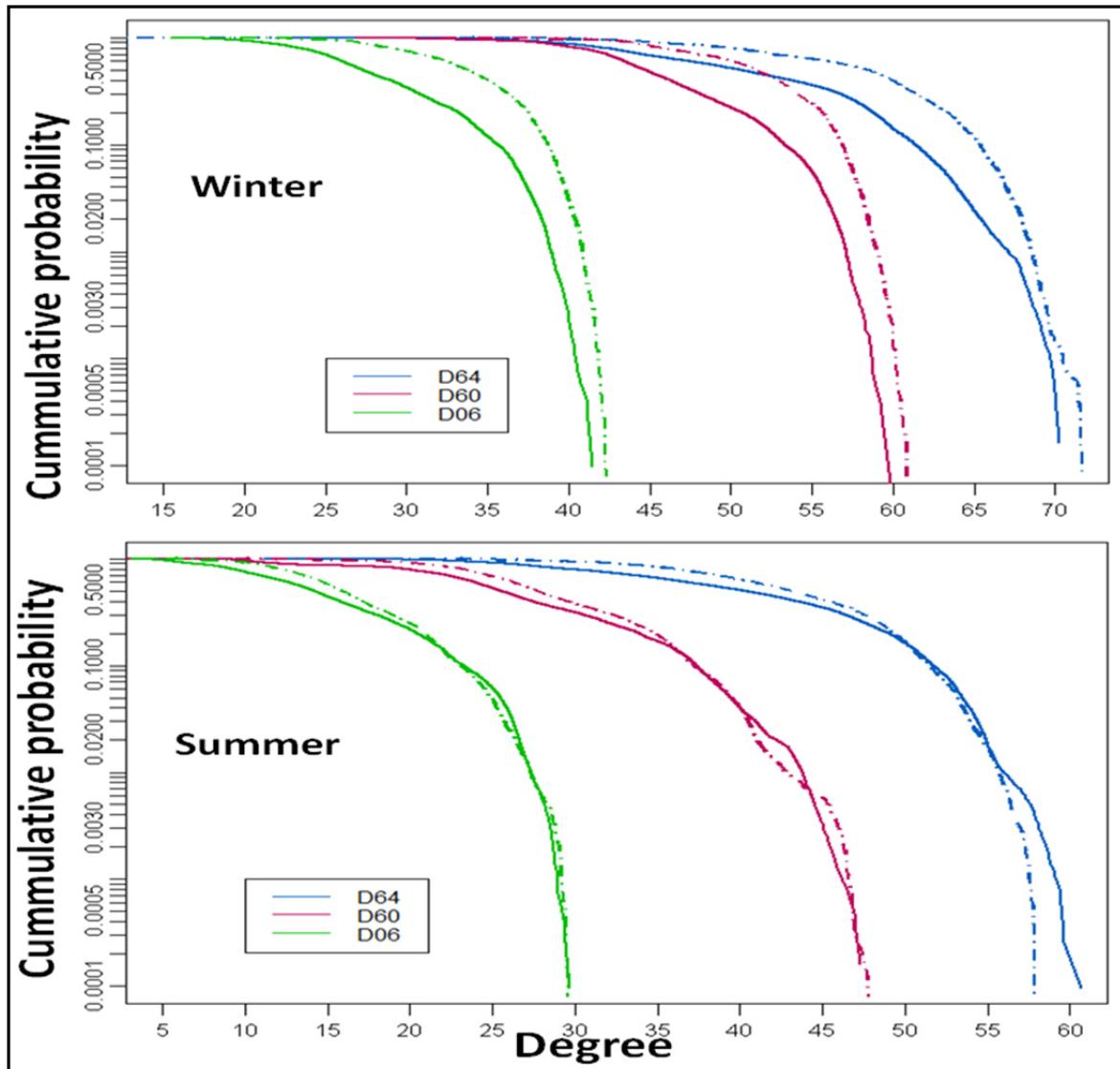


Figure 4.20 Les courbes des dépassements saisonniers en périodes de pointe (en pointillé) et hors pointe (en continu)

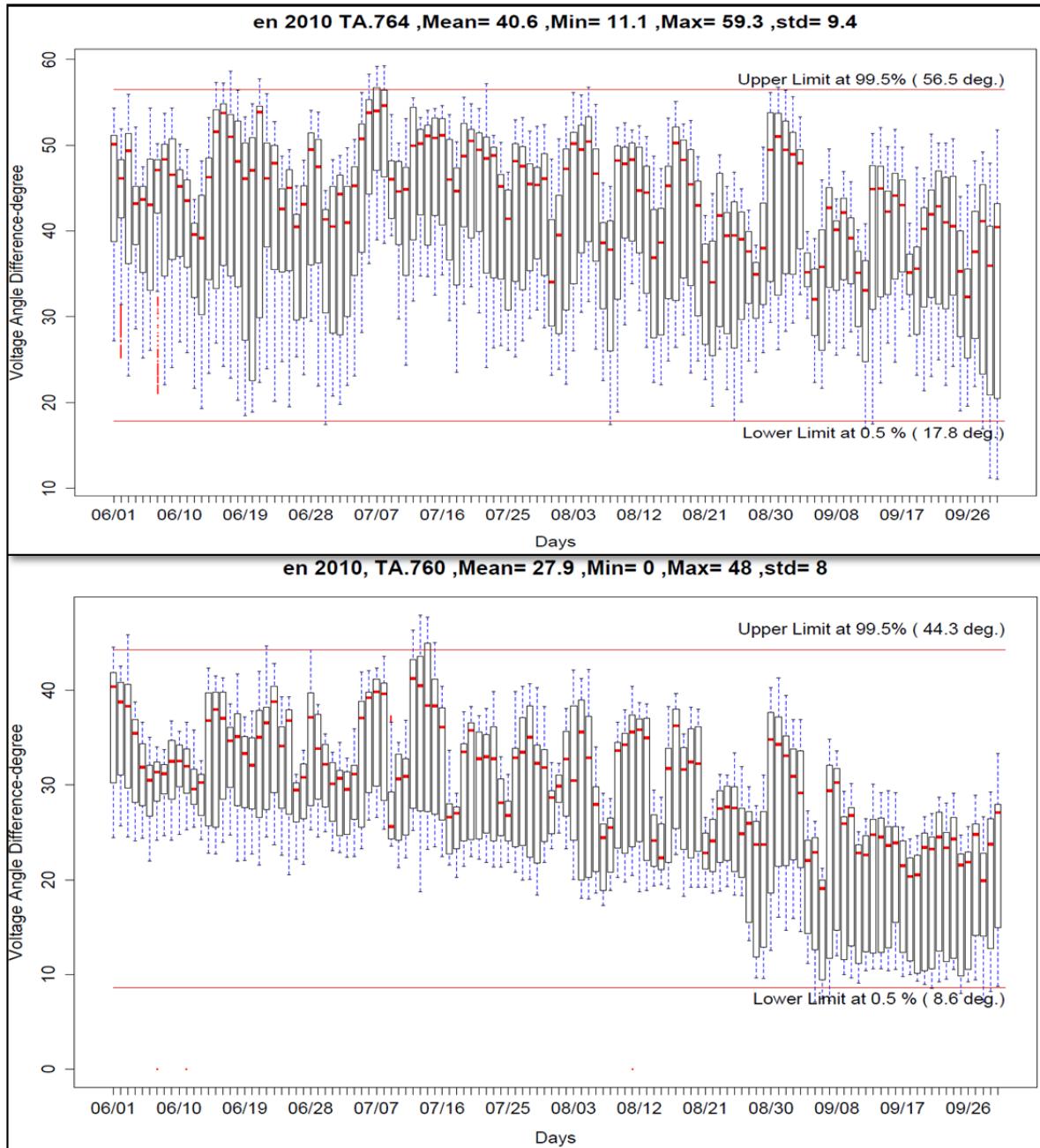


Figure 4.21 La variation quotidienne des angles de phase de tension en été 2010

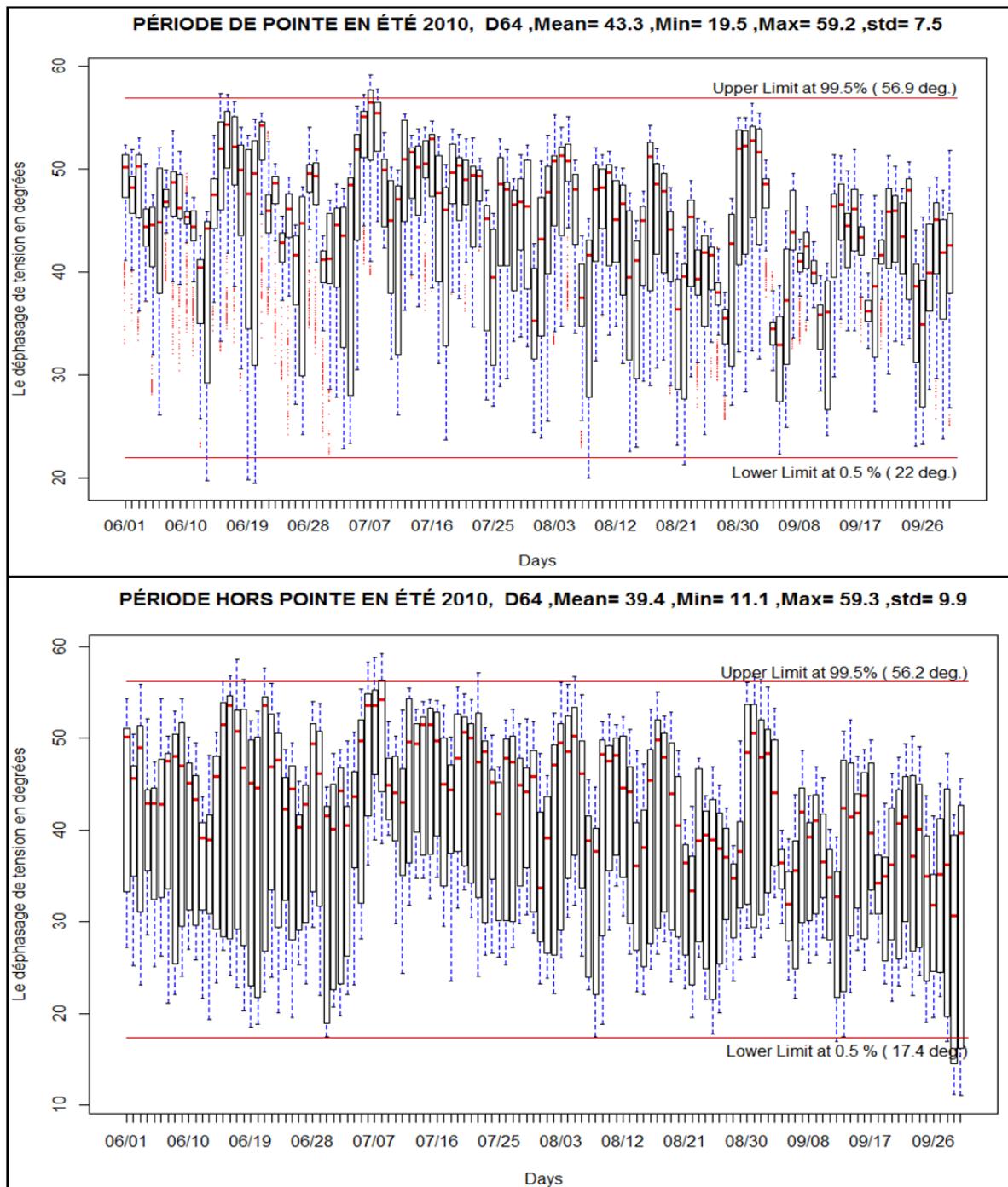


Figure 4.22 La variation quotidienne de l'angle de phase de la tension D64 en périodes de pointe et hors pointe de l'été 2010

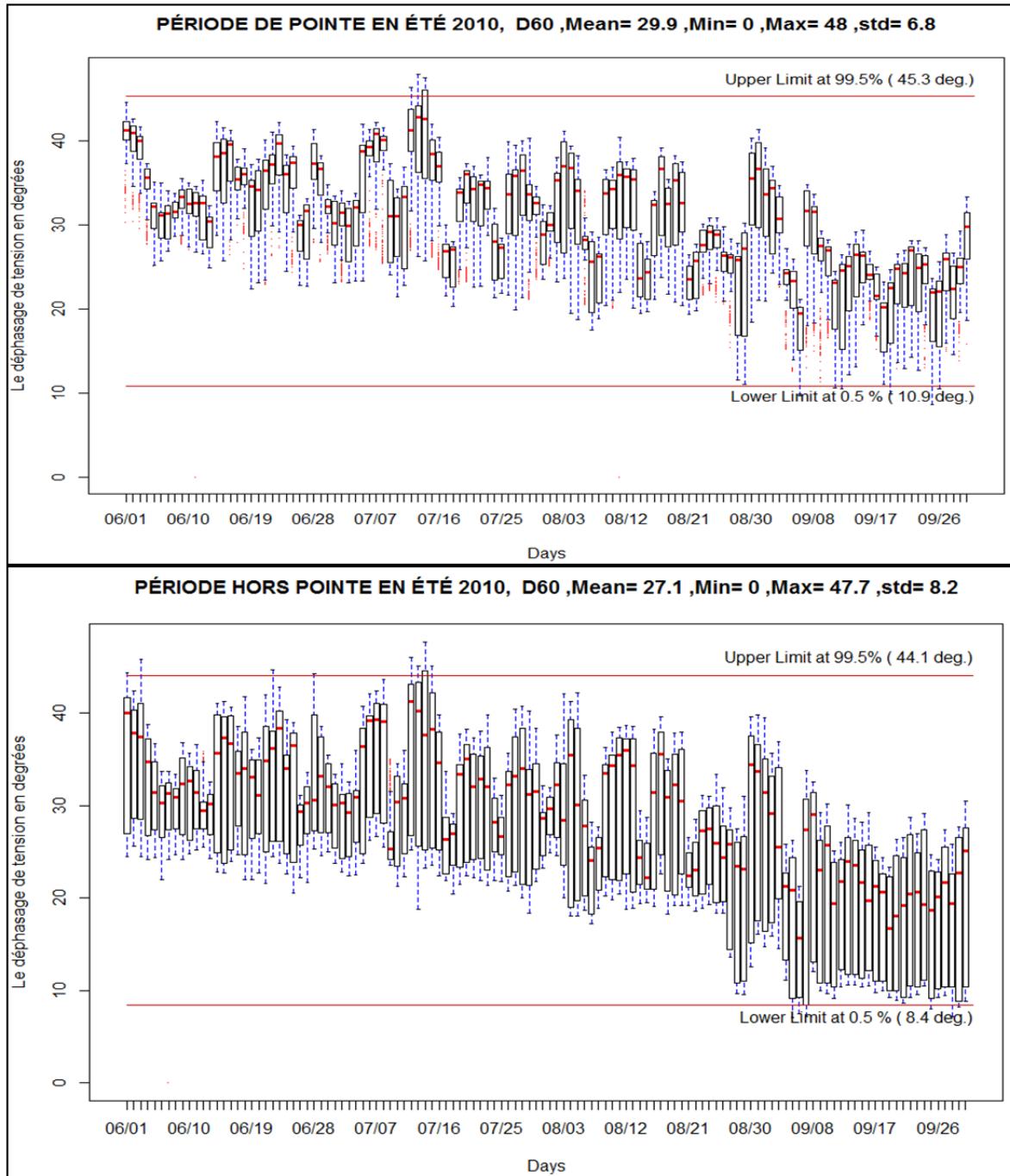


Figure 4.23 La variation quotidienne de l'angle de phase de la tension D60 en périodes de pointe et hors pointe de l'été 2010

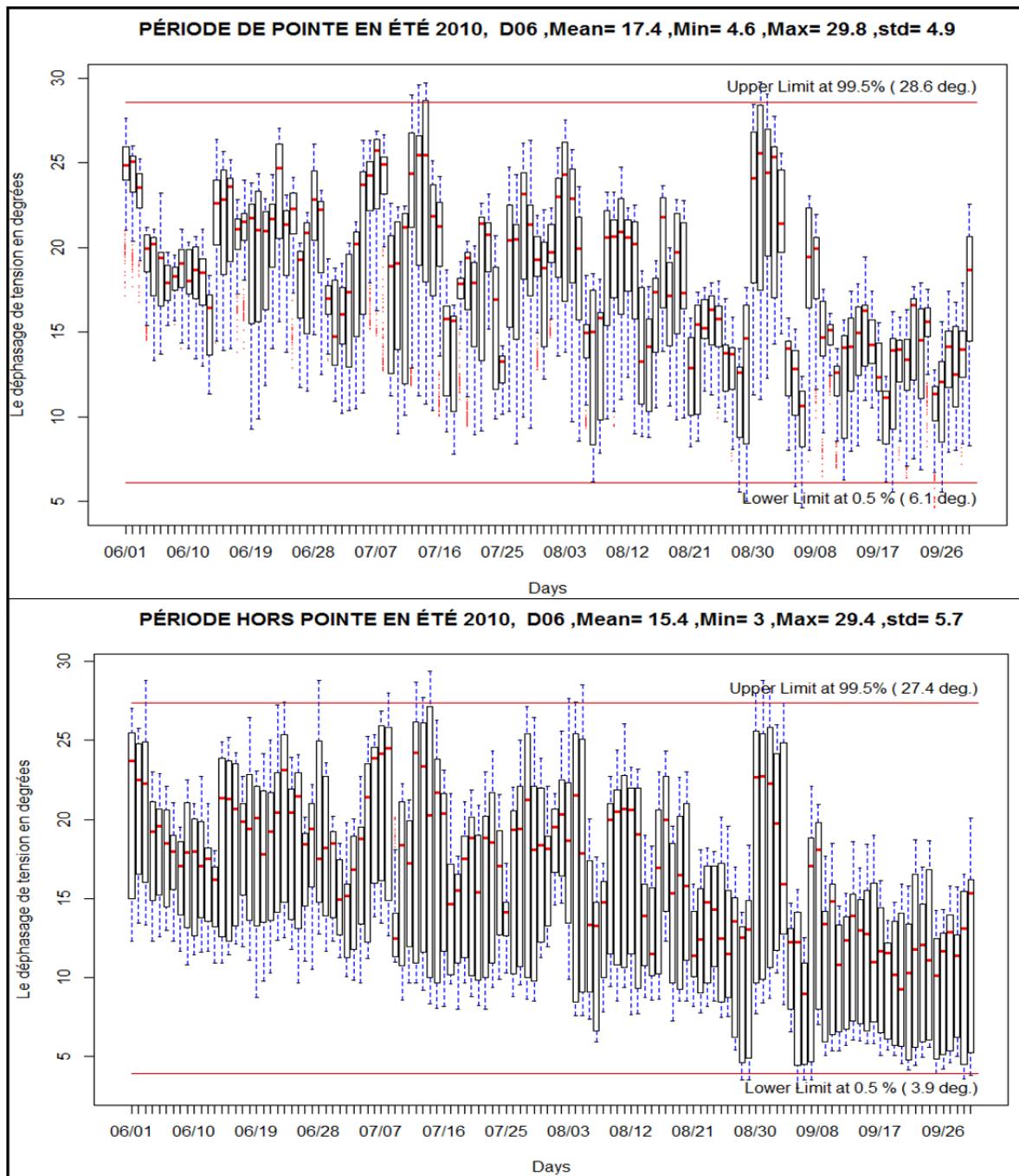


Figure 4.24 La variation quotidienne de l'angle de phase de la tension D06 en périodes de pointe et hors pointe de l'été 2010

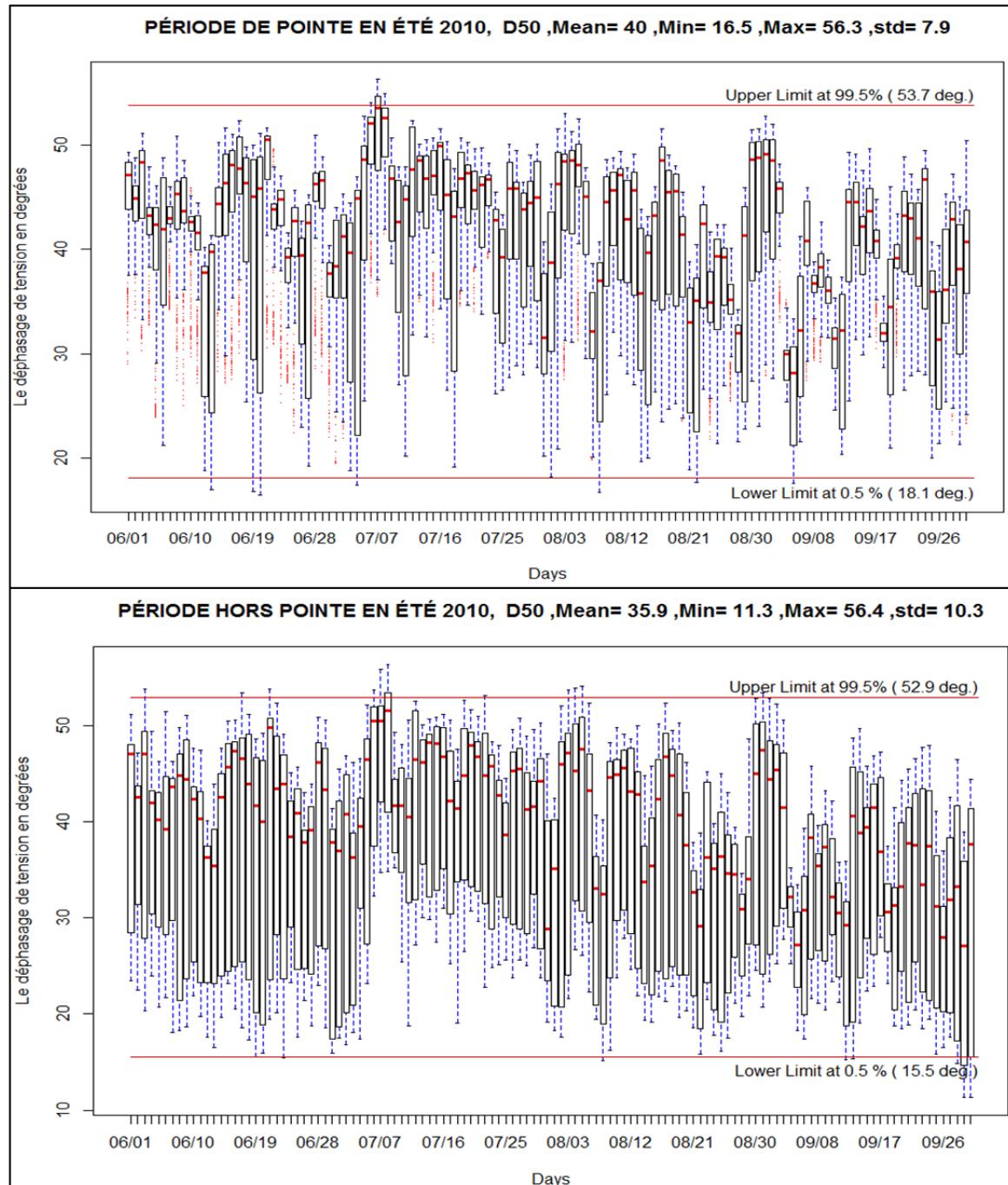


Figure 4.25 La variation quotidienne de l'angle de phase de la tension D50
en périodes de pointe et hors pointe de l'été 2010

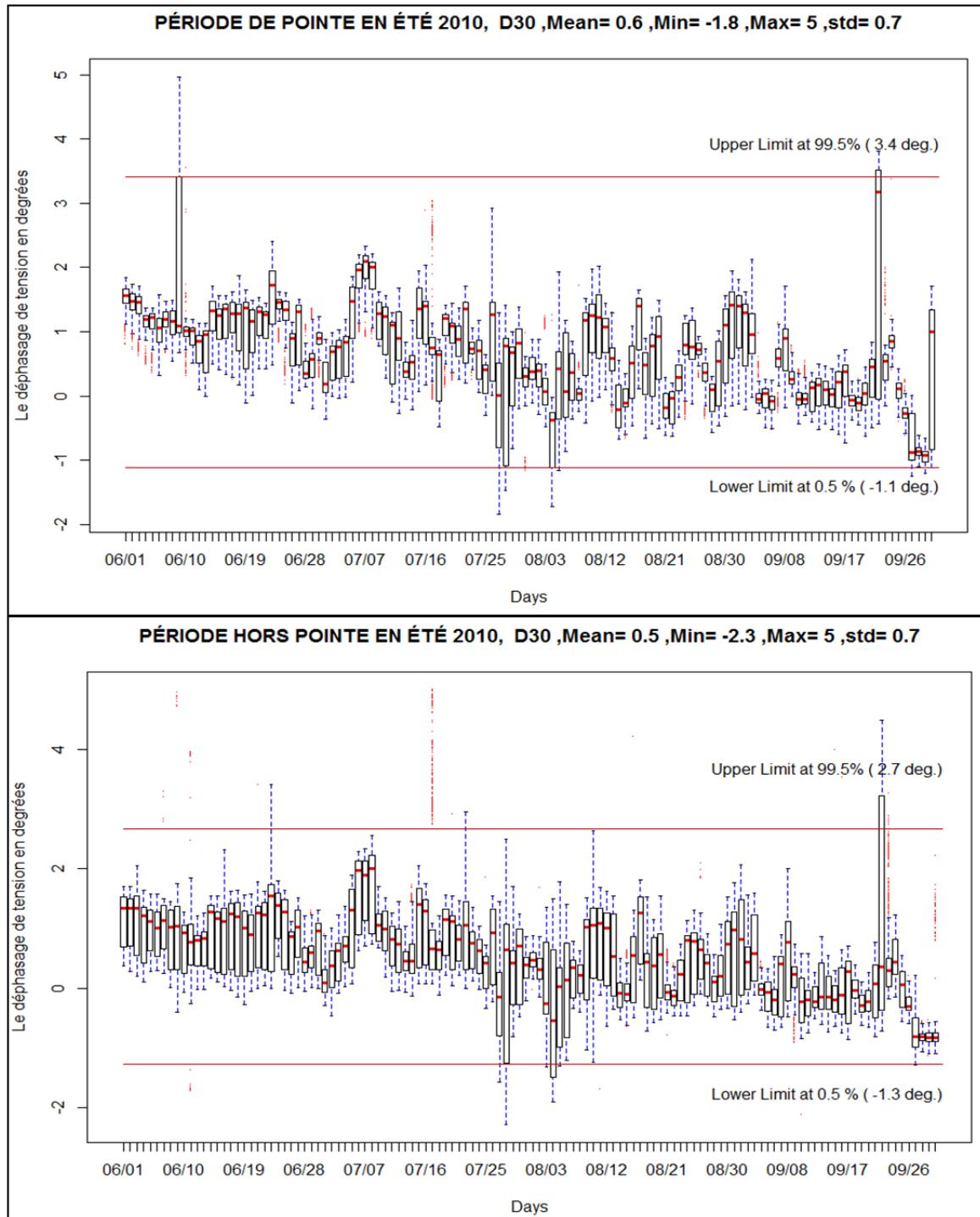


Figure 4.26 La variation quotidienne de l'angle de phase de la tension D30 en périodes de pointe et hors pointe de l'été 2010

Pour pouvoir tirer profits de ces constations, une étude de calibration statistique à base des prédicteurs nous semble nécessaire. Cette étude a pour finalité d'établir un cas de référence pour pouvoir distinguer d'une manière opérationnelle un comportement atypique d'un comportement normal des angles de tension des barres du réseau de transport.

Suite aux résultats obtenus de l'analyse des séries chronologiques de tous les angles de tension, nous proposons une limite standard pour tous les prédicteurs. Avec une limite de 99,95% pour chaque angle de phase, nous arrivons à reconnaître les comportements à surveiller, comme ceux identifiés sur la Figure 4.16, dans le souci d'assurer la sécurité du réseau. Nous rapportons quelques résultats dans le Tableau 4.5.

Tableau 4.5 Caractéristiques statistiques des angles de phase de tension de quelques postes pendant les saisons typiques

	Médiane		Quantile 99,5%	
Hiver	Hors	Pointe	Hors	Pointe
Barre 64	50,55	58,48	68,18	69,04
Barre 60	44,70	51,56	57,57	59,46
Barre 06	27,28	33,79	39,54	41,18
Été	Hors	Pointe	Hors	Pointe
Barre 64	40,24	43,12	57,60	56,43
Barre 60	25,47	27,71	45,36	45,36
Barre 06	14,16	15,91	28,15	28,27

Afin de mieux évaluer la faisabilité de la surveillance du réseau à base de dépassement du seuil de 99,95% sur les caractéristiques des angles de phase, nous avons analysé les séries chronologiques en minute des angles de tension des huit médiods sur une année complète pour détecter les événements atypiques connus dans les cinq corridors sur la même période. La Figure 4.27 illustre un ensemble des instances de l'angle de phase du médiods D06 dépassant le seuil de 99,95%. Ils ont eu lieu au cours de la saison d'hiver. À ces

moments, il est montré que les événements s'approchent de la faible marge de la sécurité dynamique qui correspond à 0,1%.

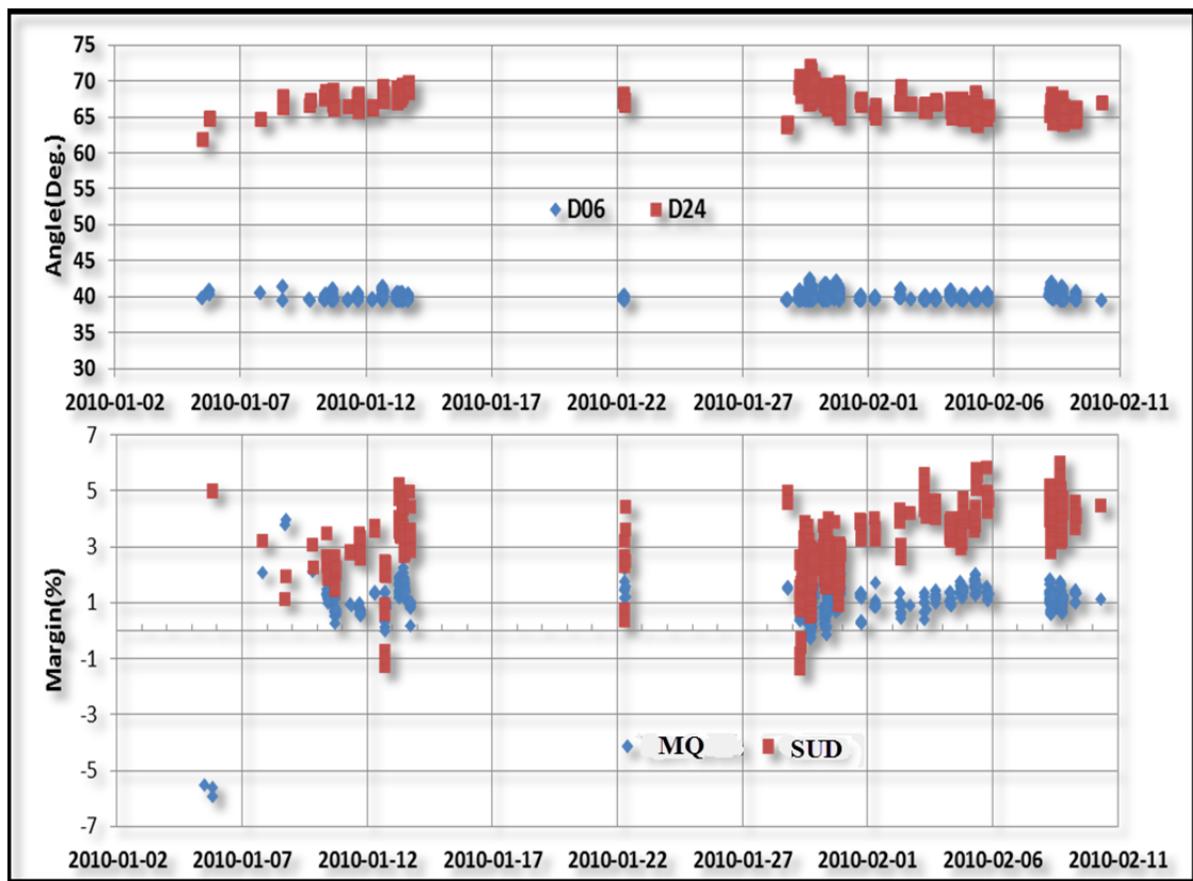


Figure 4.27 Le dépassement du seuil des angles

Pour plus de clarté, deux événements spécifiques consécutifs le même jour sont par ailleurs représentés sur la Figure 4.28. Dans le premier incident, la violation a été brève, avec des conditions de stress apparaissant dans le corridor MQ où la marge a chuté au-dessous du seuil critique de 1%, avant que la situation soit corrigée rapidement. Cette reprise se refléta correctement reflétée dans le comportement de l'angle de phase médoids qui était de retour à la normale. Un incident s'est produit alors de nouveau à 18h, où l'angle de phase dépassa le seuil de 99,95%, avec une plus longue durée et nous avons observé une plus grande variation de la marge du corridor MQ. Dans ce cas, la marge du corridor Churchill (CHU) a également diminué en dessous du seuil critique de 1%, mais, quand toutes les marges ont été restaurées,

l'angle était de retour dans la plage normale de 99,95%. En comparant ces deux événements, d'une durée d'environ 12 minutes, avec l'archive enregistré des événements du réseau, nous avons pu confirmer que leur détection à partir des écarts angulaires coïncident en effet avec des vrais phénomènes sur le réseau, comme la perte d'une ligne à cause du givrage lorsque la charge est élevée.

D'autres événements atypiques confirmés ont été également détectés à partir des angles de phase des médoïds dépassant le niveau de 99,95%. En effet, après une inspection globale des résultats sur un an, nous avons constaté que chaque dépassement de l'angle de phase médoïd de cette limite correspond au moins à un corridor dans un état d'insécurité, c'est-à-dire fonctionnant à proximité ou au-dessous des limites normales.

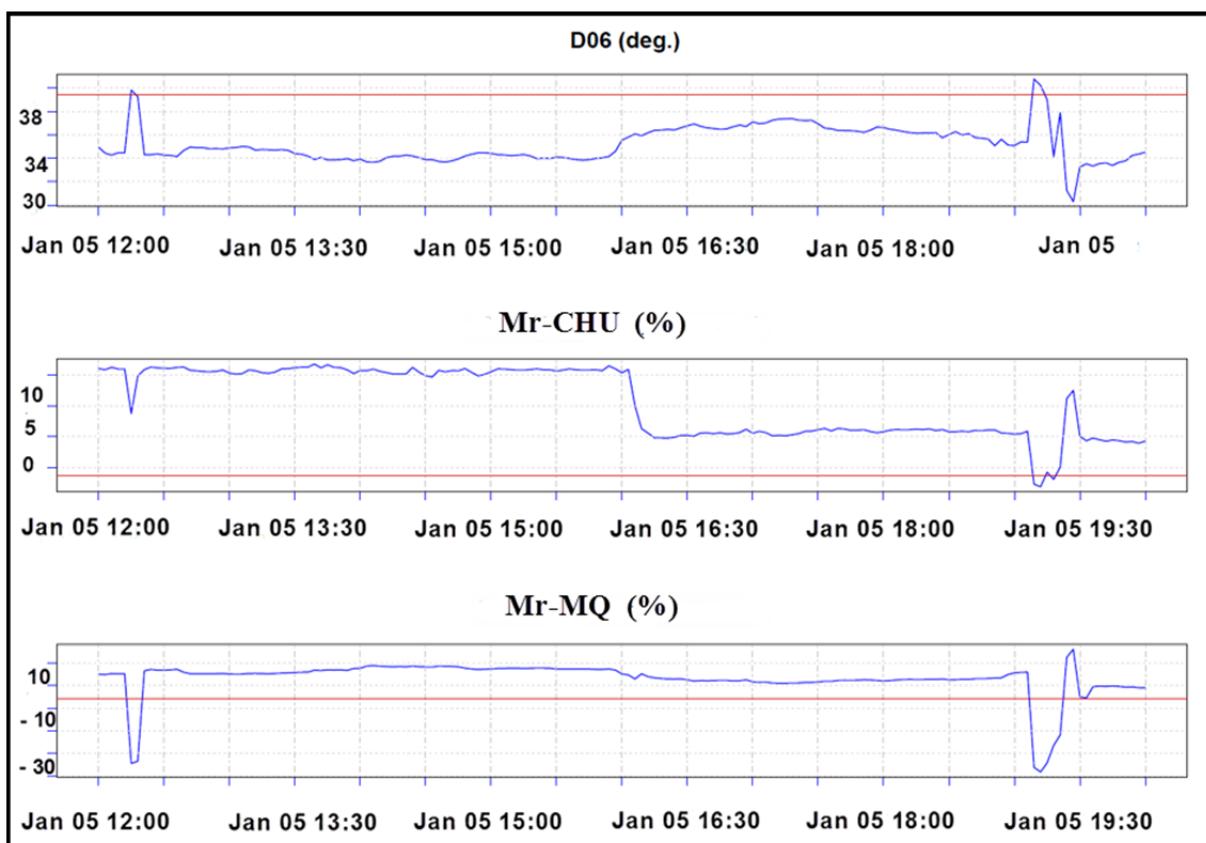


Figure 4.28 Détection des évènements à base des dépassements des angles

4.8 Conclusion

Par une approche typiquement statistique, nous avons présenté les données des angles de phase de tension extraites de la base de données l'EMS/SCADA sous formes de diagrammes en boîtes, de nomogrammes, des courbes de dépassement et des matrices de corrélation. Ces figures simples, représentant plus de 2 102 400 valeurs pour chaque variable, sont très significatives et interprétables par les connaisseurs des réseaux. À cet effet, des commentaires associés aux figures de ce chapitre ont été établis par des gestionnaires du réseau de transport d'Hydro-Québec.

Avec une approche similaire, nous avons représenté les profils des angles de tension versus les profils des transits actifs et des marges de sécurité des corridors principaux. La concordance manifeste de ces comportements nous a permis d'effectuer une étude de calibration statistique des angles de phase de tension. Comme corollaire de cette dernière analyse, nous avons constaté que chaque dépassement d'angle de la limite établie correspond au moins à un corridor dans un état d'insécurité, fonctionnant à proximité ou au-dessous des limites normales.

CHAPITRE 5

PRÉVISION DES LIMITES DE SÉCURITÉ DYNAMIQUE PAR LES FORÊTS ALÉATOIRES

5.1 Introduction

Après avoir mis en évidence dans le chapitre quatre l'existence d'une relation non linéaire entre les angles des phases mesurées et les transferts de puissance et/ou les marges de sécurité, nous nous intéresserons dans ce chapitre à la « modélisation » des transferts actifs et des marges de sécurité dynamique du réseau de transport d'HQ. Cependant, le but ultime de cet apprentissage statistique ou « modélisation » est la prévision de ces variables à base des angles de phases de tension des barres à 735 kV. Face à ce genre de problèmes où les variables cibles sont de nature si complexe, nous utilisons l'algorithme des forêts aléatoires (FA) vu leur résistance aux variances et aux biais. Nous modélisons indépendamment chaque transfert actif et chaque marge de sécurité, relatifs aux cinq principales interfaces du réseau en fonction de l'ensemble des 37 angles. Ensuite, nous répondrons à la question « est-t-il suffisant d'utiliser l'ensemble des angles médiods afin de garantir la sécurité du réseau ? ». Pour cette raison, nous recommençons la modélisation des mêmes variables, mais cette fois-ci, avec seulement les angles de phases des barres médiods. Puis, nous synthétisons les résultats de notre étude par une comparaison de performances des différents modèles. Nous terminerons ce chapitre par la vérification de la cohérence de l'approche adoptée pour valider la conjecture énoncée.

5.2 Prévision des transits et des marges de sécurité par les forêts aléatoires

Avec la méthode des forêts aléatoires, nous modélisons les transferts de puissance et les marges de sécurité relatifs aux cinq principales interfaces du réseau. Nous fixons notre travail sur la base de données de 2010 filtrée par un intervalle de temps de 3 minutes. Ainsi, pour modéliser chaque transfert actif et chaque marge de sécurité, nous utilisons, en premier lieu, les 37 prédicteurs chronologiques dont chacun contient 175 000 observations. Avec l'algorithme des forêts aléatoires, nous subdivisons aléatoirement cet ensemble

d'observations en trois sous-ensembles. Le premier sous-ensemble est constitué de 70% des données et nous l'utilisons pour former les modèles, c'est ce qu'on appelle l'ensemble d'apprentissage. Le deuxième sous-ensemble est constitué de 15% des données et nous l'utilisons pour évaluer et valider nos modèles. Le troisième sous-ensemble est formé de 15% des données et nous l'utilisons pour tester nos modèles de prévision. La Figure 5.1 schématise ce principe d'application des modèles des forêts aléatoires sur l'ensemble test.

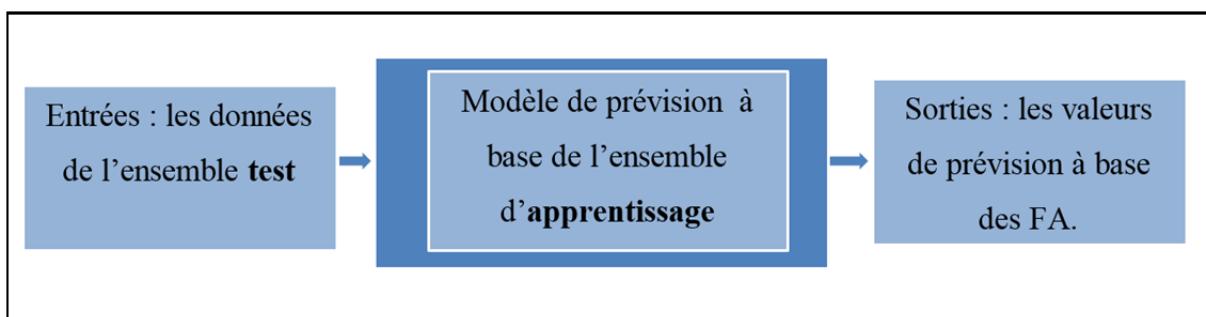


Figure 5.1 Le principe de prévision des transits actifs et des marges de sécurité à base des FA

Pour cette démarche d'apprentissage, nous avons utilisé 100 arbres de régression. Ce choix est largement justifié, en se référant à l'erreur OOB qui se stabilise à partir de cette valeur comme on le voit dans la Figure 5.2. En second lieu, nous utilisons la même méthode avec les mêmes paramètres, de l'algorithme des FA (les mêmes observations, le même nombre d'arbres, la même répartition des sous-ensembles), mais avec moins de variables d'entrée. Nous utilisons comme données d'entrée les observations des séries chronologiques des huit angles de phases des barres médoides pour prédire les transits actifs et les marges de transits des cinq corridors séparément. Les angles des huit barres médoides sont ceux identifiés dans le chapitre 4 et sont : D22, D24, D82, D83, D14, D03, D04 et D06.

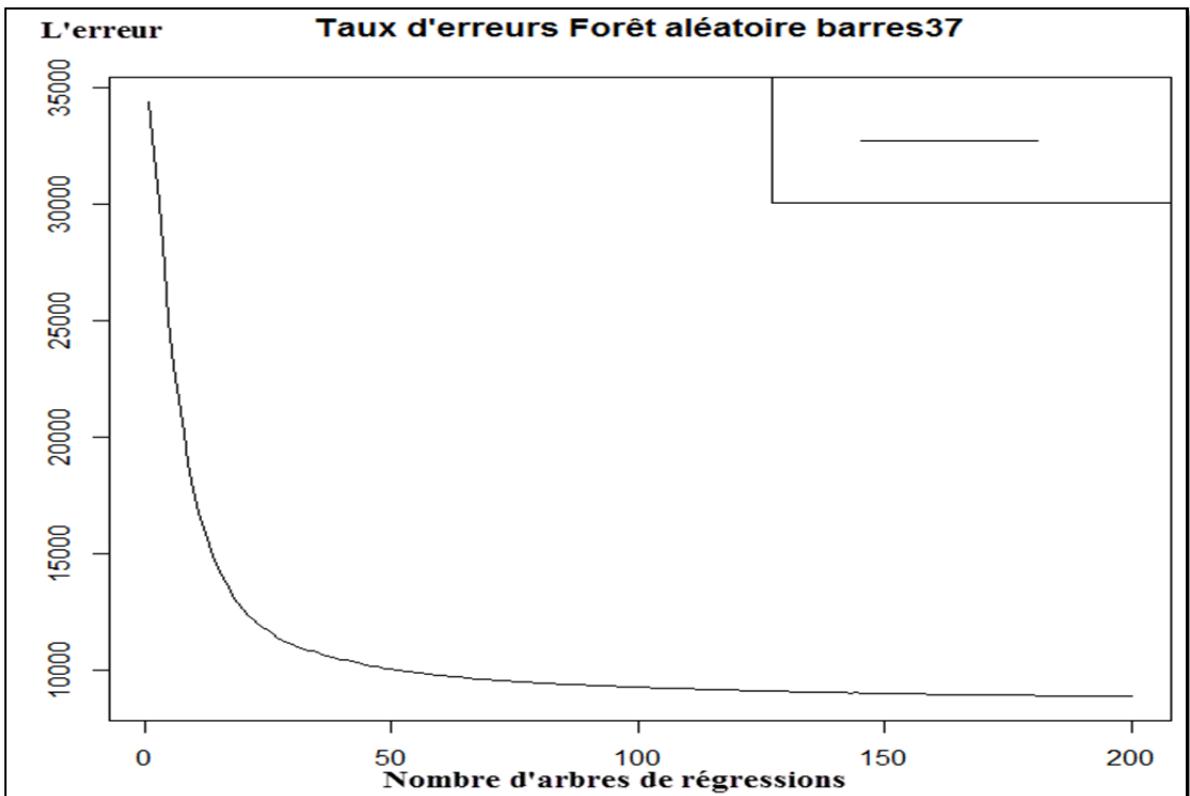


Figure 5.2 L'erreur de prévision en fonction de nombres d'arbres de régression

5.3 Résultats et commentaires

Relativement à l'erreur MSE, le Tableau 5.1 classe en ordre décroissant le degré d'importance des prédicteurs dans la formation des modèles de prévision. Dans tous les modèles de prévision que nous avons construits, la majorité des prédicteurs classés en premiers appartiennent à la boucle sud identifiée dans le chapitre 4. Ces prédicteurs sont indiqués par la couleur jaune dans le Tableau 5.1.

Tableau 5.1 L'importance relative des prédicteurs lors de la formation des modèles de prévision des transits actifs

SUD (MSE %)	MQ (MSE %)	CHU (MSE %)	BJS (MSE %)	BJN (MSE%)
D14 49,05	D15 40,99	D15 31,15	D15 57,47	D15 59,55
D15 44,88	D02 37,78	D19 28,34	D19 43,71	D02 49,93
D70 40,76	D70 34,22	D02 22,44	D30 41,77	D70 46,11
D19 37,67	D13 32,89	D14 20,66	D55 41,17	D55 44,01
D07 36,36	D01 31,25	D70 19,61	D02 40,37	D18 42,33
D02 35,27	D08 28,20	D08 19,54	D09 37,36	D10 41,48
D84 31,81	D84 27,77	D66 18,94	D05 36,87	D60 40,96
D08 28,58	D14 26,50	D13 17,74	D08 36,30	D30 39,85
D01 27,22	D66 26,40	D01 16,83	D07 35,90	D01 38,93
D30 24,35	D19 26,32	D64 16,15	D84 35,84	D09 38,09
D55 21,99	D24 25,58	D10 15,69	D60 35,76	D08 37,37
D66 17,20	D20 25,31	D60 15,23	D01 35,15	D05 37,22
D13 14,86	D49 24,50	D80 13,23	D90 34,13	D07 37,11
D24 10,84	D23 24,40	D24 12,95	D06 30,32	D03 35,95
D65 10,14	D64 22,10	D20 12,71	D17 29,94	D84 35,82
D51 8,06	D65 19,63	D82 12,68	D04 29,81	D04 35,20
D50 7,91	D80 15,82	D84 12,52	D03 27,55	D90 34,15
D22 7,52	D51 15,57	D22 12,39	D10 27,08	D19 33,35
D04 7,42	D82 11,70	D06 12,38	D70 21,57	D17 32,22
D64 7,33	D52 9,20	D51 12,28	D18 19,89	D13 20,66
D52 7,18	D60 9,17	D65 12,03	D14 9,17	D14 9,52
D20 6,83	D10 8,86	D50 11,95	D52 7,31	D06 8,42
D03 6,62	D03 8,41	D09 11,54	D51 7,21	D50 8,01
D60 6,60	D06 8,27	D05 11,23	D49 7,15	D51 7,33
D06 6,49	D30 8,14	D23 10,93	D20 6,47	D22 7,06
D17 6,37	D07 8,12	D49 10,88	D80 6,41	D52 6,38
D82 6,17	D50 7,95	D04 10,04	D22 6,28	D20 5,88
D18 6,00	D90 7,27	D83 9,76	D50 5,79	D49 5,53
D10 5,96	D18 7,13	D03 9,57	D31 5,52	D66 5,25
D49 5,94	D31 7,10	D31 9,49	D13 5,38	D80 4,52
D09 5,89	D05 6,81	D30 8,28	D66 4,42	D23 4,47
D05 5,81	D09 6,79	D18 7,84	D23 4,39	D31 4,28
D90 5,56	D55 6,56	D07 7,75	D82 4,39	D64 4,21
D31 5,49	D22 6,04	D52 7,17	D83 3,88	D82 4,17
D23 5,19	D04 5,99	D55 6,48	D64 3,82	D24 4,15
D80 4,58	D83 5,70	D17 6,33	D65 3,73	D65 3,77
D83 4,57	D17 5,52	D90 6,31	D24 3,53	D83 3,62

La Figure 5.3 illustre quelques résultats de prévision sur le sous-ensemble test qui n'a pas été pris en considération lors de la construction du modèle. Cette figure présente le lissage linéaire de la prévision en fonction des valeurs réelles des marges (en % relativement à la demande correspondante) et des transferts actifs (en MW) à travers certaines interfaces. L'axe des abscisses représente les valeurs réelles des points appartenant au sous-ensemble test et l'axe des ordonnées représente les valeurs prédictes à partir du modèle déjà construit par l'ensemble d'apprentissage.

Pour la prévision de la marge de sécurité des corridors SUD et MQ, en fonction de huit angles, nous obtenons un R élevé. Cependant, nous remarquons que ces modèles ne sont guère aussi robustes que les modèles de prévision des transits actifs. Plus les points se condensent au voisinage de la diagonale, plus le modèle est robuste. Cette constatation est fortement appuyée par les variances exprimées dans le Tableau 5.2.

Tableau 5.2 L'évaluation de la performance des modèles de FA
(Base de données : l'année 2010 par 3 minutes)

	Avec les 37 barres de à 735 kV			Avec les huit (8) principales barres		
	Variance exp. (%)	R	MAE ; 3Q (%)	Variance exp. (%)	R	MAE ; 3Q (%)
Mr. SUD	98,69	99,78	0,49 ; 0,52	97,77	99,61	0,70 ; 0,78
Mr. MQ	97,70	99,62	0,31 ; 0,29	95,88	98,05	0,46 ; 0,45
Mr. CHU	96,81	99,51	0,14 ; 0,13	94,12	97,3	0,20 ; 0,20
Mr. BJS	94,72	99,12	0,47 ; 0,48	90,34	95,22	0,72 ; 0,80
Mr. BJJ	94,56	99,10	0,46 ; 0,46	91,04	95,75	0,66 ; 0,70
			MW			MW
Tr. SUD	99,87	99,93	43, 50	99,70	99,85	64, 73
Tr. MQ	99,79	99,89	33, 37	99,27	99,60	61, 67
Tr. CHU	99,62	99,81	19, 21	96,63	98,46	58, 64
Tr. BJS	99,81	99,91	32, 37	99,59	99,80	46, 50
Tr. BJJ	99,81	99,91	33, 39	99,60	99,80	48, 53

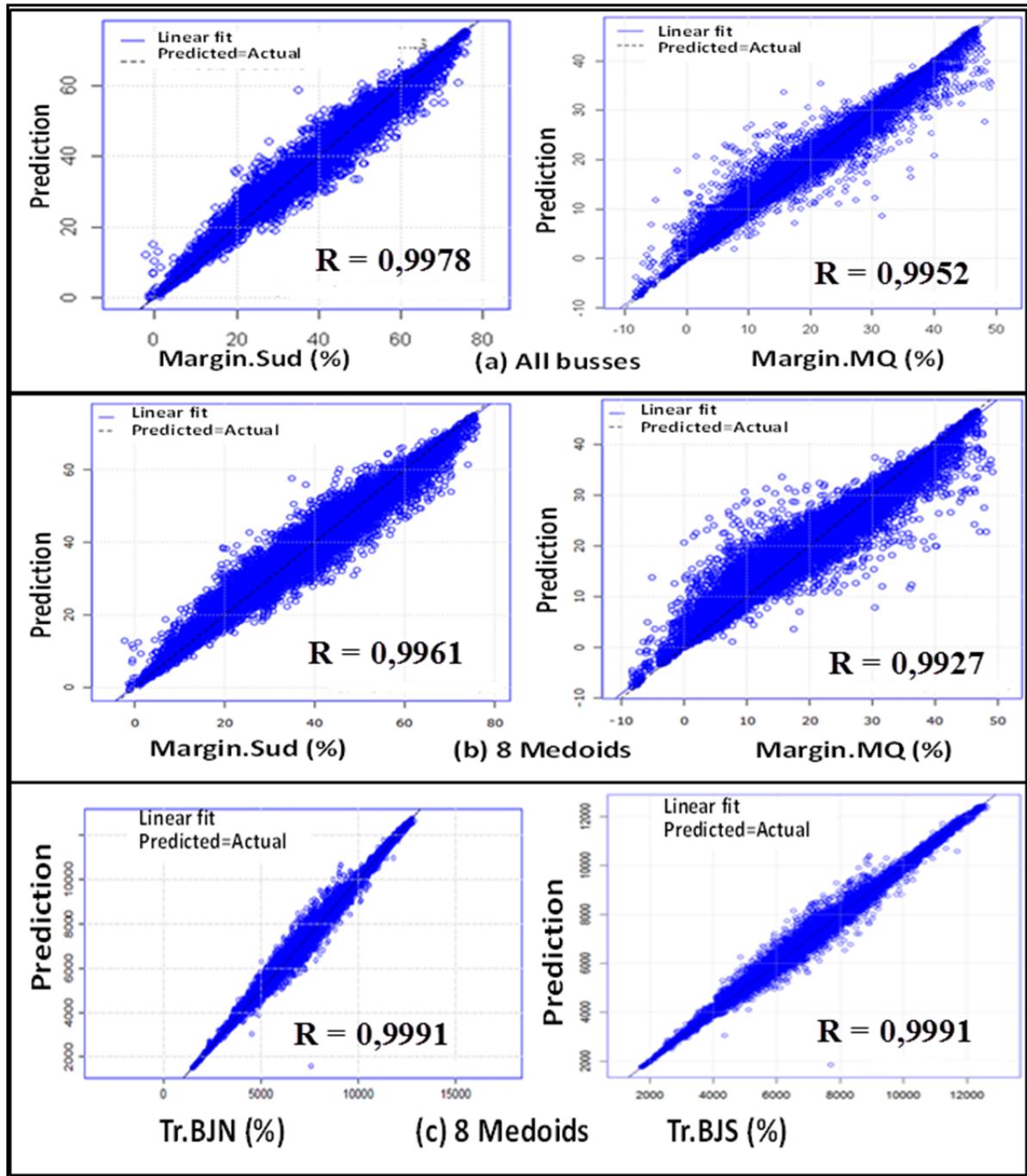


Figure 5.3 L'évaluation du modèle de prévision avec la méthode des FA

Globalement, le Tableau 5.2 présente une évaluation complète de la précision de tous les modèles de prévisions par les FA, d'une part avec tous les prédicteurs et d'une autre part avec juste les prédicteurs correspondants aux barres médiods. En comparant les modèles de prévisions des marges de sécurité avec ceux de prévisions des transits actifs, nous avons établi que ces derniers ont une meilleure précision que les premiers. Or concernant les modèles de prévisions des marges de sécurité, il importe de préciser que la valeur de R soit supérieure à 99% dans certains cas, ce qui montre la qualité de ces derniers. En comparant les modèles de prévisions en fonction des huit prédicteurs avec ceux des 37 prédicteurs, nous pouvons conclure que lorsque le modèle avec 37 prédicteurs est précis (R élevé) et robuste (Variance expl. élevée), alors cette même qualité de prévision peut être obtenue avec un nombre restreint de prédicteurs. C'est le cas de prévisions des transits actifs des interfaces SUD, BJJN et BJS.

Les cinq figures, Figure 5.4-Figure 5.8, illustrent les séries chronologiques des marges de sécurité de chacune des cinq interfaces. Nous représentons les séries de la marge réelle, de la marge prédite par les FA et de l'erreur brute entre les deux marges. Chaque figure regroupe deux graphes résultant d'une étude à base des angles des barres médiods d'une part et, à base de tous les angles de barres à 735kV d'autre part.

À partir de ces graphes, on peut remarquer que les erreurs de prévisions des marges des interfaces ne sont pas biaisées et elles ont tendance à accroître lorsque les marges sont élevées. Pour les interfaces CHU et MQ, l'erreur de prévisions à base de tous les angles est nettement inférieure à l'erreur de prévisions avec les angles des barres médiods. Alors que pour les autres interfaces telles que BJJN, BJS et SUD, l'erreur est modérément moindre.

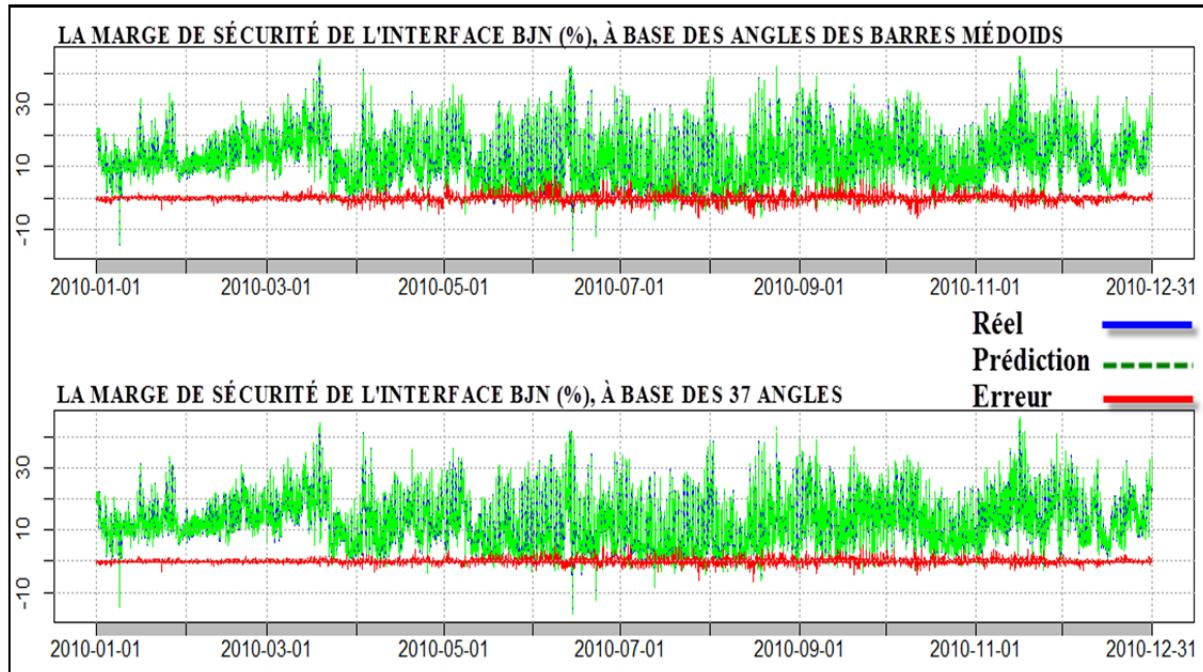


Figure 5.4 Les séries chronologiques des erreurs de prévisions avec les modèles de FA de la marge de sécurité du corridor BJT

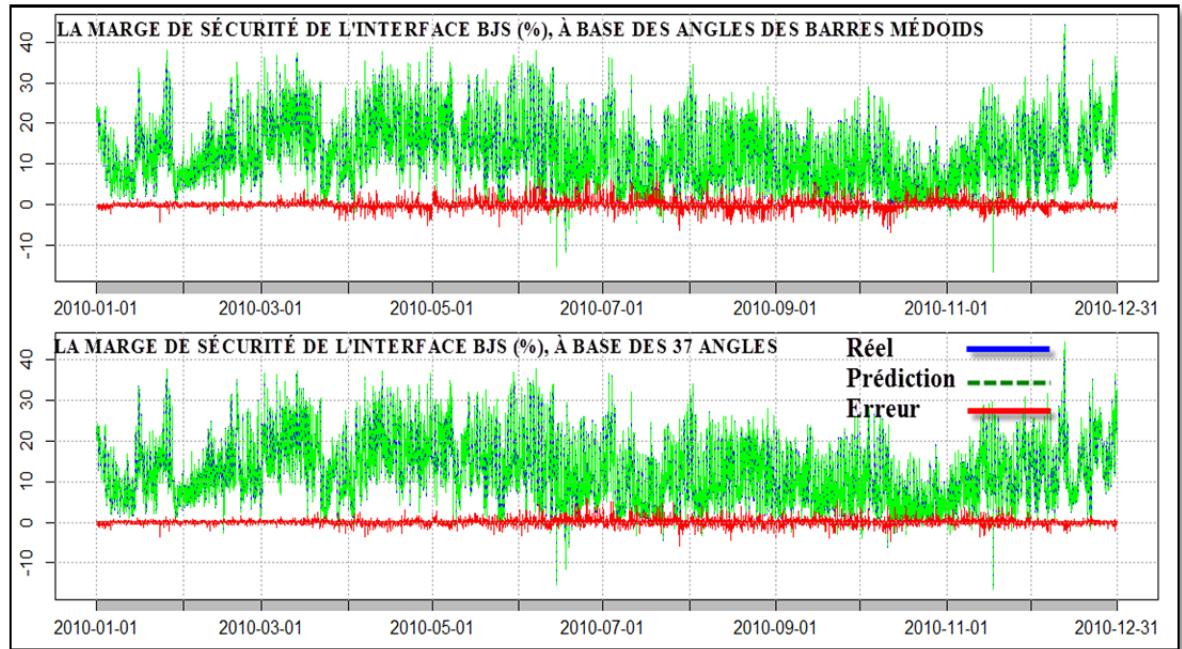


Figure 5.5 Les séries chronologiques des erreurs de prévisions avec les modèles de FA de la marge de sécurité du corridor BJS

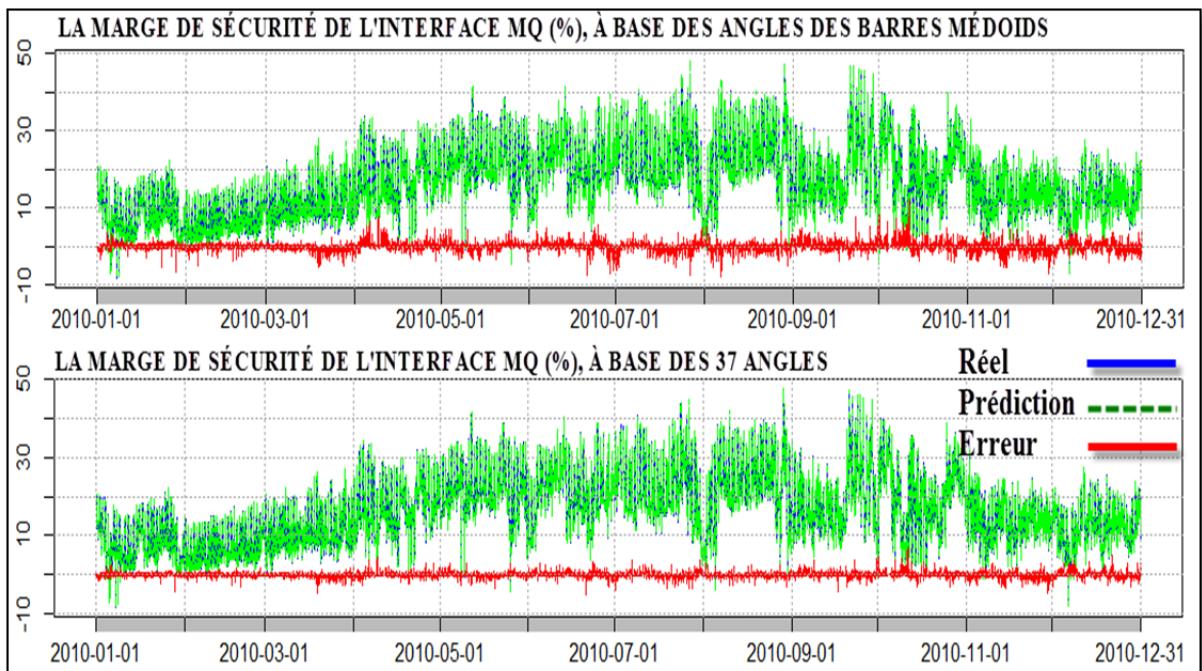


Figure 5.6 Les séries chronologiques des erreurs de prévisions avec les modèles de FA de la marge de sécurité du corridor MQ

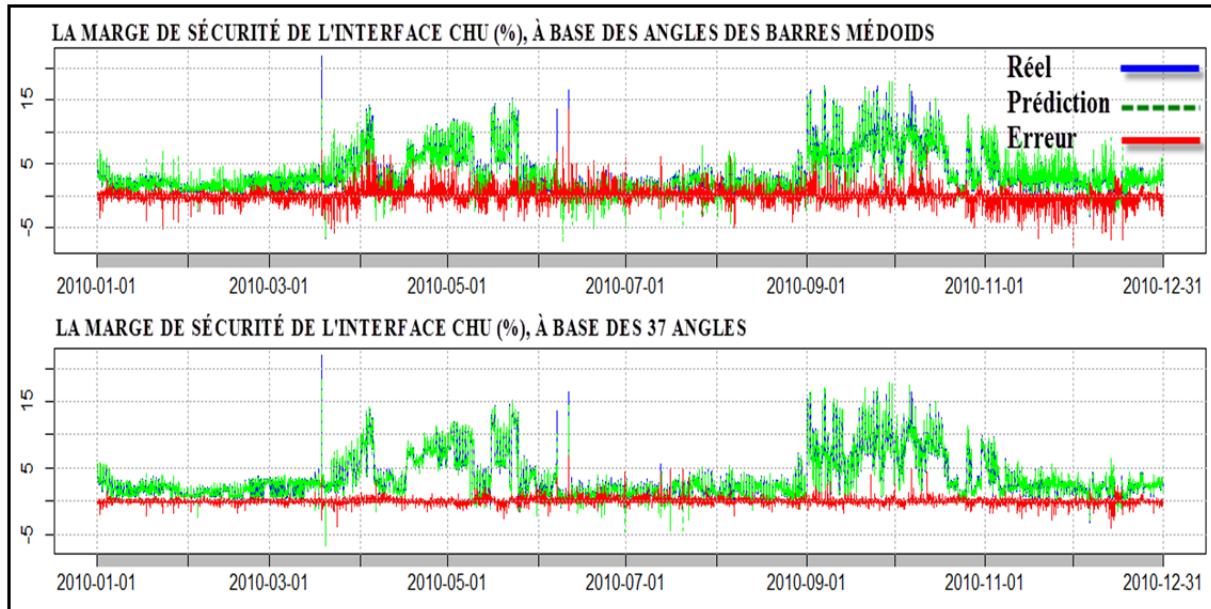


Figure 5.7 Les séries chronologiques des erreurs de prévisions avec les modèles de FA de la marge de sécurité du corridor CHU

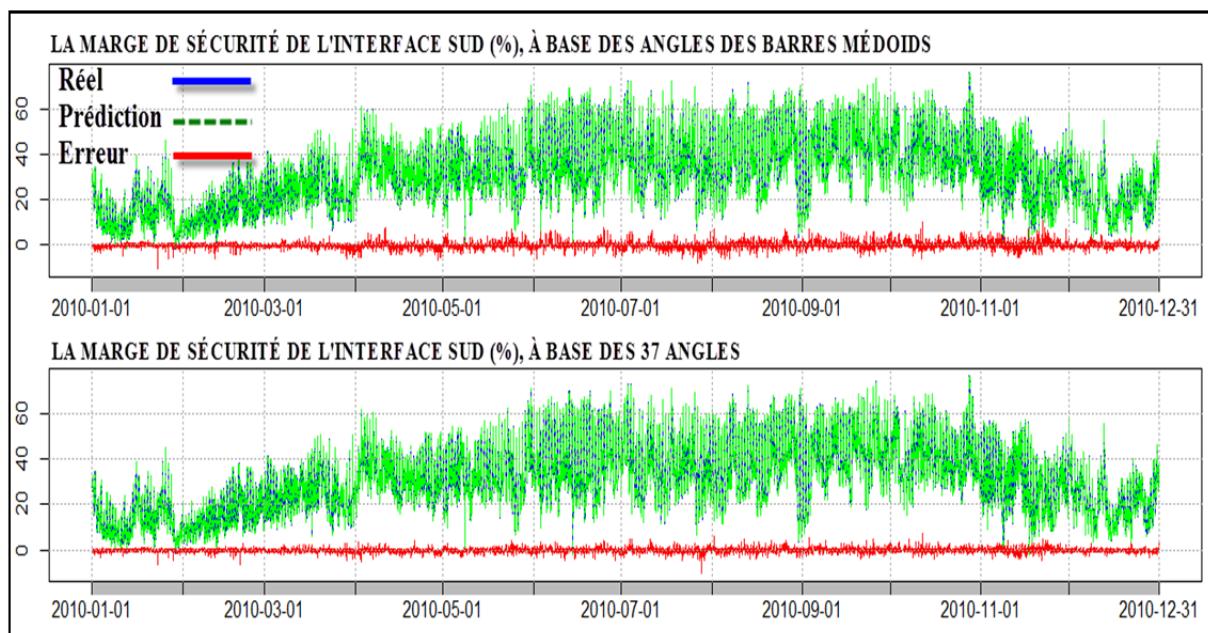


Figure 5.8 Les séries chronologiques des erreurs de prévisions avec les modèles de FA de la marge de sécurité du corridor SUD

Les deux figures, Figure 5.9 et Figure 5.10, illustrent les séries chronologiques des transits actifs de chacune des cinq interfaces. Nous représentons les séries du transit actif réel, du transit actif prédict à base des angles des barres médoids par les FA et de l'erreur brute entre les deux transits. De ces graphes, mise à part l'interface CHU, toutes les erreurs de prévisions des transits des interfaces du réseau de transport ne sont pas biaisées.

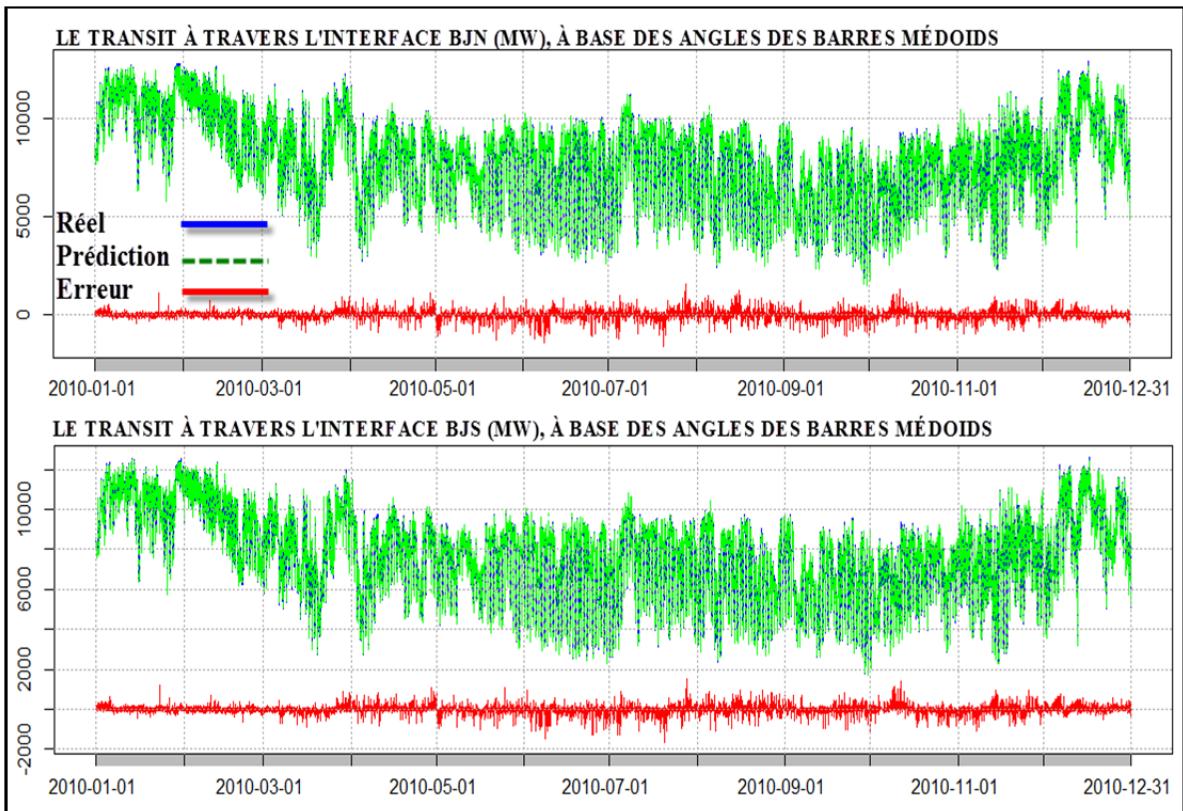


Figure 5.9 Les séries chronologiques des erreurs de prévisions par les FA des transferts de puissance de l'année 2010 par 3 minutes à travers les interfaces BZN et BJS

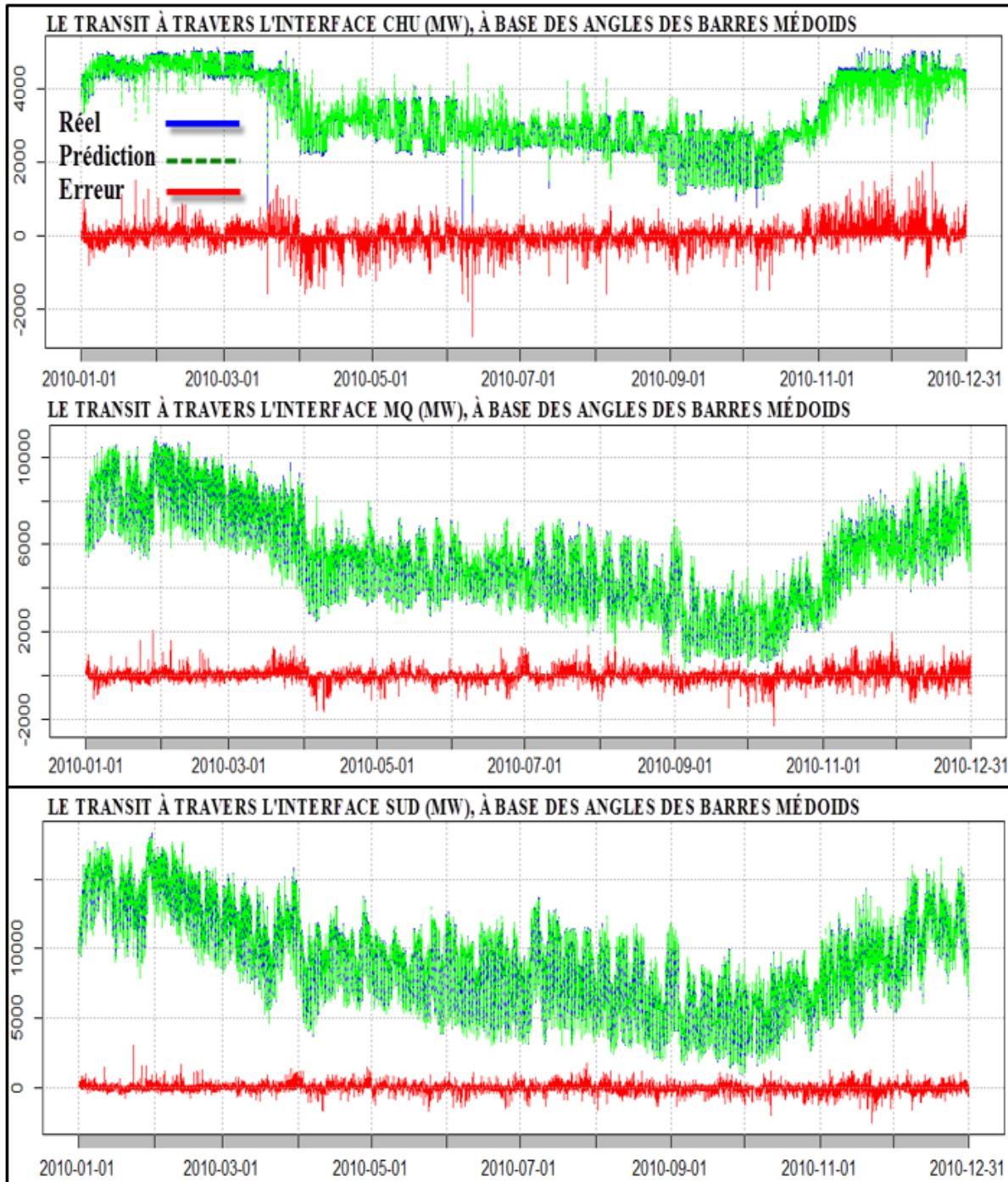


Figure 5.10 Les séries chronologiques des erreurs de prévisions par les FA des transferts de puissance de l'année 2010 par 3 minutes à travers les interfaces CHU, MQ et SUD

Les quatre figures, Figure 5.11-Figure 5.14, présentent les densités des erreurs de prévisions des différents modèles à base des FA. Tous les graphes montrent explicitement que les erreurs sont centrées presque à l'origine et ont tendance à suivre une loi normale. Ceci indique clairement que les erreurs sont sans-biais. Dans ce cas, d'après le théorème central limite, plus on rajoute des informations ou des observations (à des fractions d'une minute), plus l'erreur devient clairement gaussienne et centrée à l'origine. Autrement dit, nous avons des processus asymptotiquement sans-biais donc les erreurs sont des bruits blancs. Ainsi, toutes les informations sont concentrées dans la partie principale du modèle. Par conséquent, plus on rajoute des informations, nos modèles deviennent plus stables et plus cohérents et nous permettent d'avoir une précision de qualité supérieure. L'ensemble des résultats, du Tableau 4.3 au Tableau 5.7, récapitulent les caractéristiques statistiques des modèles de prévisions des transits actifs.

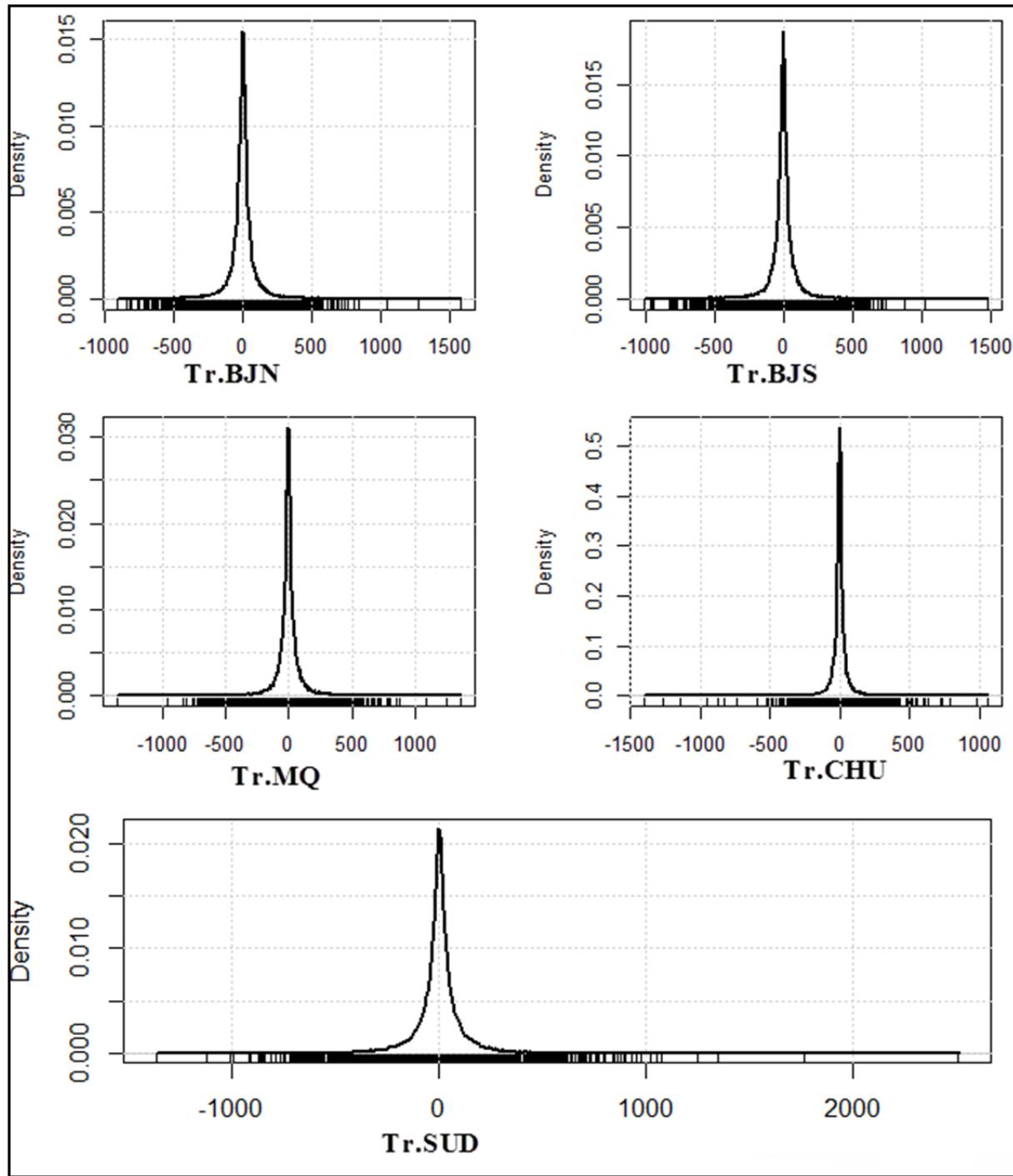


Figure 5.11 Les densités des erreurs de prévisions des transits actifs du sous-ensemble test en fonction de tous les angles par les modèles de FA établis sur l'ensemble d'apprentissage de données de l'année 2010 à 3 minutes

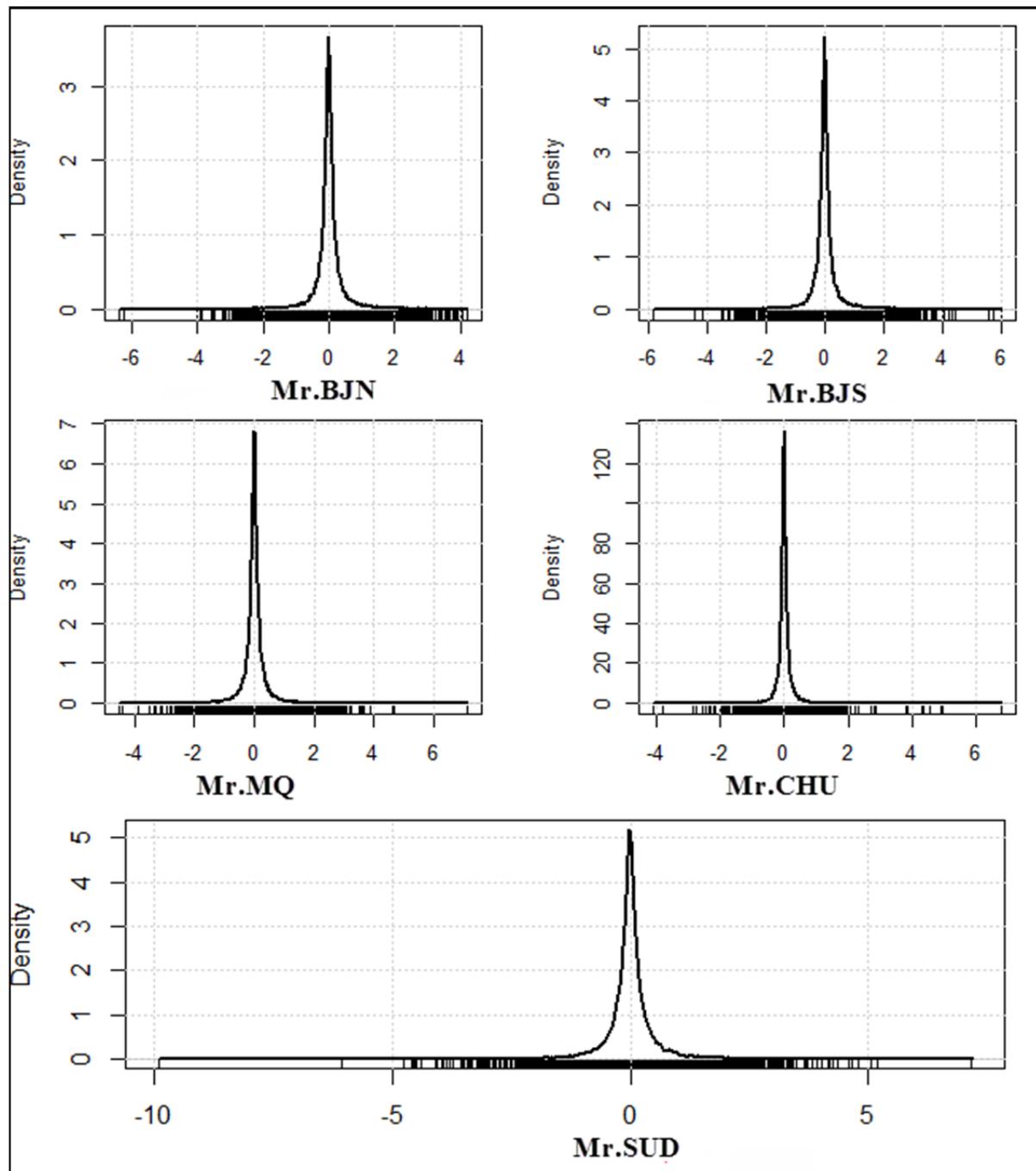


Figure 5.12 Les densités des erreurs de prévisions des marges de sécurité du sous-ensemble test en fonction de tous les angles par les modèles de FA établis sur l'ensemble d'apprentissage de données de l'année 2010 à 3 minutes

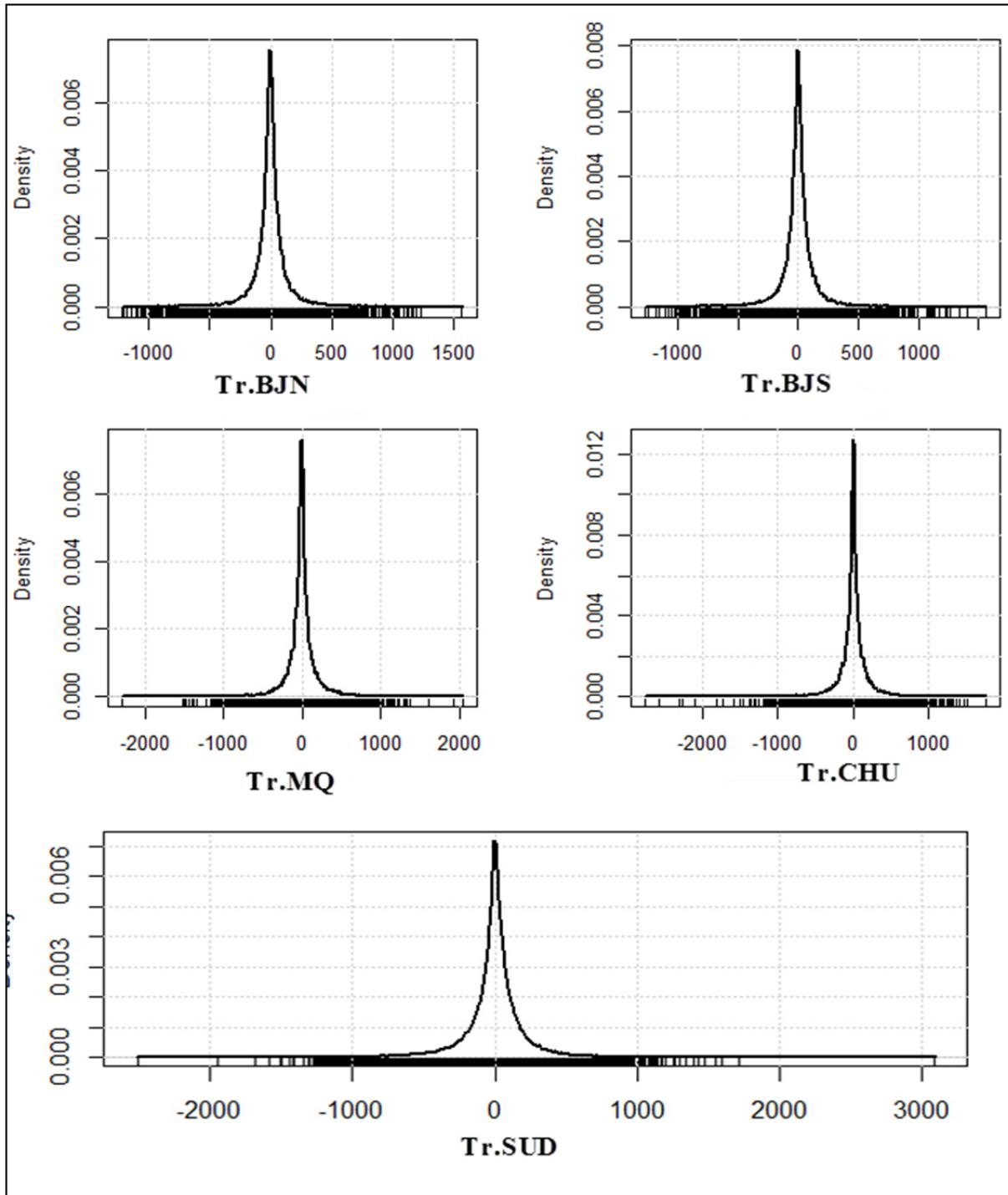


Figure 5.13 Les densités des erreurs de prévisions des transits actifs du sous-ensemble test en fonction des angles des barres médoides par les modèles de FA établis sur l'ensemble d'apprentissage de données de l'année 2010 à 3 minutes

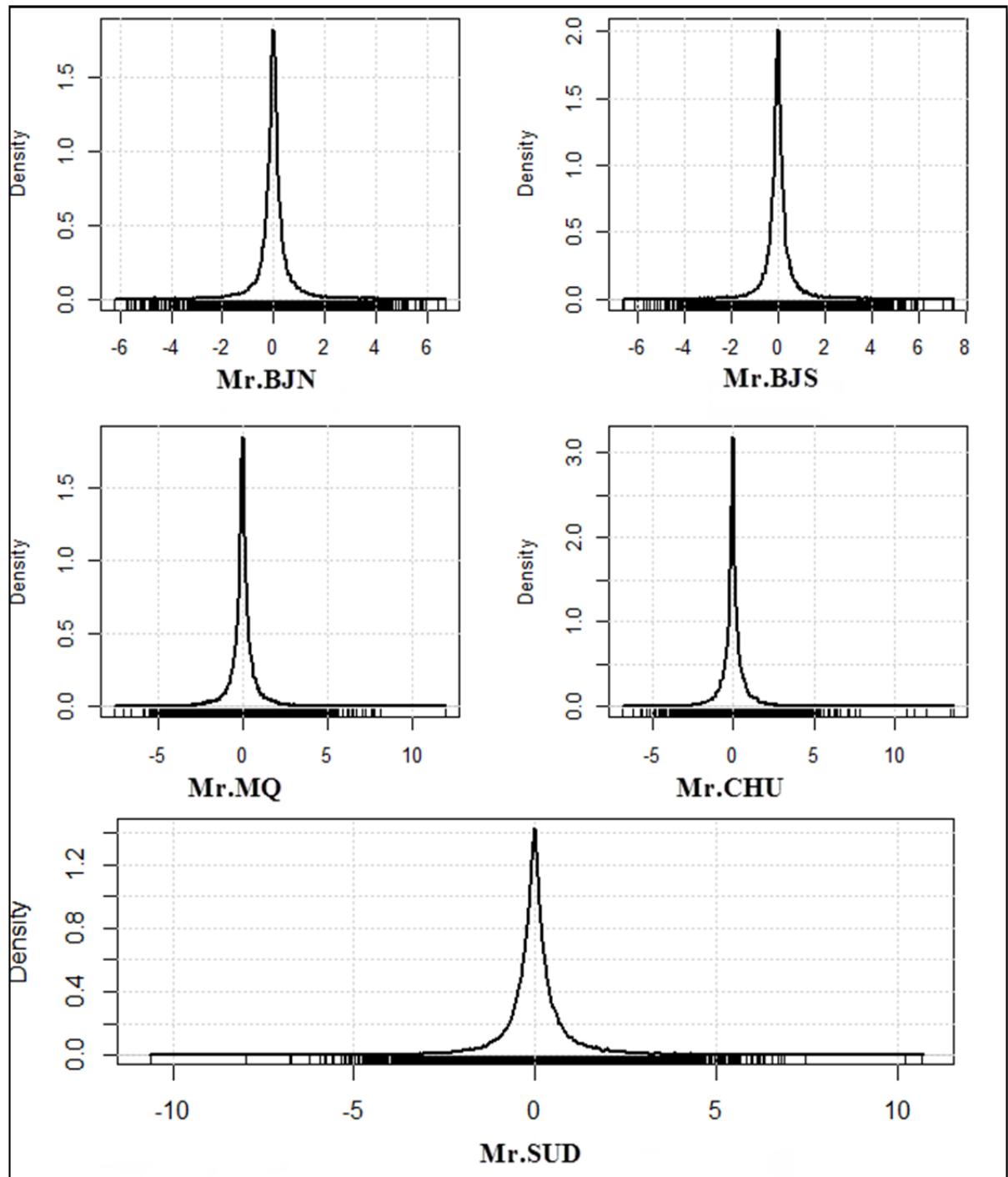


Figure 5.14 Les densités des erreurs de prévisions des marges de sécurité du sous-ensemble test en fonction des angles des barres médoides par les modèles de FA établis sur l'ensemble d'apprentissage de données de l'année 2010 à 3 minutes

Tableau 5.3 Caractéristiques statistiques de l'ensemble test extrait de la base de données de 2010 à 3 minutes

Transits actifs (MW)	BJN	BJS	MQ	CHU	SUD
Minimum	1 484	1 747	293,5	0	929,5
1 ^{er} Quartile	6 355	6 267	3 717,7	2 734	6 682,1
Médiane	8 129	7 966	5 422,8	3 190	9 322,9
Moyenne	7 855	7 703	5 391,4	3 438	9 187,2
3 ^{ème} quartile	9 358	9 138	6 698,3	4 443	11 326,5
Maximum	12 747	12 514	10 951,6	5 112	17 987,3

Tableau 5.4 Caractéristiques statistiques de l'ensemble prédit par les FA en fonction de tous les angles de la base de données de 2010 à 3 minutes

Prévision par les FA	BJN	BJS	MQ	CHU	SUD
Minimum	1 492	1 755	329,7	à 735,7	991,3
1 ^{er} Quartile	6 365	6 285	3 715,7	2 731,6	6 689,2
Médiane	8 139	7 971	5 428,4	3 183,8	9 319,7
Moyenne	7 857	7 705	5 391,0	3 438,9	9 189,0
3 ^{ème} quartile	9 356	9 129	6 686,6	4 440,0	11 313,4
Maximum	12 726	1 2417	10 859,4	5 079,5	17 954,3

Tableau 5.5 Caractéristiques statistiques des erreurs de prévisions en fonction de tous les angles par les modèles de FA établis sur l'ensemble d'apprentissage de données de l'année 2010 à 3 minutes

Erreur brute de prévision	BJN	BJS	MQ	CHU	SUD
Minimum	-914,2516	1 009,0245	1 355,0436	1 402,2560	1 364,6801
1 ^{er} Quartile	-29,5283	26,5633	26,4809	14,8623	36,3297
Médiane	-0,5422	0,1918	-1,0015	0,1202	-0,7143
Moyenne	-2,0801	-1,9562	0,4097	-0,4654	-1,7743
3 ^{ème} quartile	26,9631	26,5795	24,7739	15,0342	34,7867
Maximum	1 578,8518	1 477,0958	1 357,8778	1 059,4198	2 505,7217
Écart type	93,50934	89,98169	93,82656	58,73351	117,39826

Tableau 5.6 Caractéristiques statistiques de l'ensemble prédit par les FA en fonction des angles de phase des barres médiods de la base de données de 2010 à 3 minutes

Prévision par les FA	BJN	BJS	MQ	CHU	SUD
Minimum	1 494	1 754	320,6	1 075	994,4
1 ^{er} Quartile	6 372	6 288	3 722,6	2 733	6 691,5
Médiane	8 132	7 969	5 439,8	3 204	9 308,6
Moyenne	7 856	7 704	5 391,9	3 439	9 188,6
3 ^{ème} quartile	9 349	9 117	6 688,4	4 416	11 305,3
Maximum	12 719	12 392	10 854,2	5 047	17 877,8

Tableau 5.7 Caractéristiques statistiques des erreurs de prévisions en fonction des angles des barres médiods par les modèles de FA établis sur l'ensemble d'apprentissage de données de l'année 2010 à 3 minutes

Erreur brute de prévision	BJN	BJS	MQ	CHU	SUD
Minimum	-1 217,9983	-1 272,813	2 285,5910	2762,4340	2 514,929
1 ^{er} Quartile	-44,8800	-42,011	58,9608	51,4874	64,343
Médiane	-0,4013	1,451	-3,1356	1,3025	3,137
Moyenne	-1,6114	-1,085	-0,4684	-0,8135	-1,311
3 ^{ème} quartile	45,1625	46,139	53,1702	53,7784	70,610
Maximum	1 574,4141	1 558,254	2 040,7569	1 751,8465	3 090,073
Écart type	145,5888	143,1278	185,7168	181,3590	204,6377

5.4 La consistance de l'approche de prévision de la sécurité dynamique par les angles de phase de tension des barres à 735kV avec la méthode des FA

Après les études et les analyses élaborées sur les observations d'une base de données de 2010, provenant de l'EMS SCADA, nous avons pu obtenir les données des autres années à savoir celles des années 2009, 2011 et 2012. L'acquisition de cette base de données, nous a permis de vérifier la consistance de l'approche adoptée pour démontrer la validité de la conjecture énoncée.

À l'aide de la méthode des FA, nous élaborons des modèles de prévisions des transits actifs et les marges de sécurité à partir d'une base de données s'échelonnant sur quatre ans, de 2009 à 2012. Pendant toute cette période, le réseau de transport d'HQ a connu des conditions de

fonctionnement différentes d'une année à l'autre. En effet, les exportations nettes d'Hydro-Québec ont changé énormément au cours de cette période: 18,5, 12,6, 20,8 et 30,2 TWh en 2009, 2010, 2011 et 2012 respectivement. Cela implique une énorme variation dans les modèles d'écoulement et les conditions du réseau par saison et par heure de la journée, entraînée en grande partie par l'activité en temps réel des marchés de l'électricité des réseaux voisins. Face à ces différentes variations et contraintes, nous étions amenés à étudier la consistance de notre approche dans la construction de nouveaux modèles plus généraux et plus complexes.

Il advient que lors de l'établissement de ces nouveaux modèles, seulement 70% des données de cette base de données de 4 ans sont utilisées dans la formation de ces modèles. Ainsi, chaque modèle n'utilise pas 30 % des conditions disponibles du réseau. Nous résumons les résultats obtenus dans le Tableau 5.8.

Tableau 5.8 L'évaluation des modèles de prévisions par les FA des transits actifs et des marges de sécurité à base de données des années 2009 à 2012

Les interfaces	Les transits actifs (MW)			Les marges de sécurité (%)		
	SUD	MQ	BJN	SUD	MQ	BJN
La variance expliquée à (%)	99,22	98,7	99,22	98,69	97,70	94,56
R , test	99,83	99,73	99,83	99,78	99,62	99,10
MAE	109,02	37,962	80,19	0,49	0,31	0,46

Pour les trois interfaces du réseau de transport d'HQ, l'interface SUD, MQ et BJR, que ce soit pour la prévision des marges de sécurité ou la prévision des transferts actifs, les modèles sont assez précis pour prédire 15 % de données correspondants à des situations et conditions de réseau non vues par le modèle des FA lors de l'apprentissage. À titre illustratif,

considérant le transfert de puissance le long du corridor SUD, le coefficient R mesuré sur les données test de 4 ans est de 99,83 % par rapport à 99,93 % mesuré sur les données test de l'année 2010. Les R de la marge de stabilité dynamique de l'interface SUD entre la validation sur une période de 4 ans et sur une période seulement d'une année 2010 sont similaires. Encore une fois, cette performance (sur plusieurs années) d'un modèle unique se compare favorablement à la performance du modèle élaboré à partir des données de 2010, confirmant ainsi la bonne caractéristique de généralisation des angles de phase de tension comme facteurs prédictifs des transferts dynamiques et des marges de sécurité du réseau de transport réel.

Les représentations, de la Figure 5.15 à la Figure 5.19, illustrent les séries chronologiques des transits actifs et des marges de sécurité sur une période de 3 ans (de 2010 à 2012 avec un pas de 9 minutes). Cette réduction de la fréquence d'étude est intimement liée à la contrainte concernant la vitesse d'exécution de l'ordinateur. Chaque graphe de ces figures, rassemble la série chronologique réelle, la série chronologique prédite et la série chronologique de l'erreur de prévisions. Nous notons clairement que les erreurs ne sont pas biaisées. L'ensemble de résultats, du Tableau 5.9 au Tableau 5.11, résume quelques caractéristiques statistiques des transits actifs, à savoir la série réelle, la série prédite et la série de l'erreur de prévision de chaque variable cible de chacune des interfaces du réseau de transport. Les illustrations, de la Figure 5.20 à la Figure 5.23, regroupant les erreurs de prévisions de tous les modèles sur l'ensemble test, appuient ce constat. Tous les graphiques associés montrent explicitement que les erreurs suivent une loi gaussienne centrée à l'origine. Nous arrivons à la même conclusion que celle de l'étude précédente basée sur les données d'une année où toutes les erreurs sont des bruits blancs.

Suite à l'ensemble des résultats obtenus dans ce chapitre, nous venons de valider la conjecture qu'on s'est proposé comme l'un des objectifs faisant l'objet de ce travail de thèse. Par ailleurs, il est indéniable que l'utilisation de l'algorithme des FA était d'un apport considérable pour la qualité des erreurs obtenues. Ces performances peuvent être justifiées par le fait que cet algorithme repose sur une démarche optimale de Bootstrap.

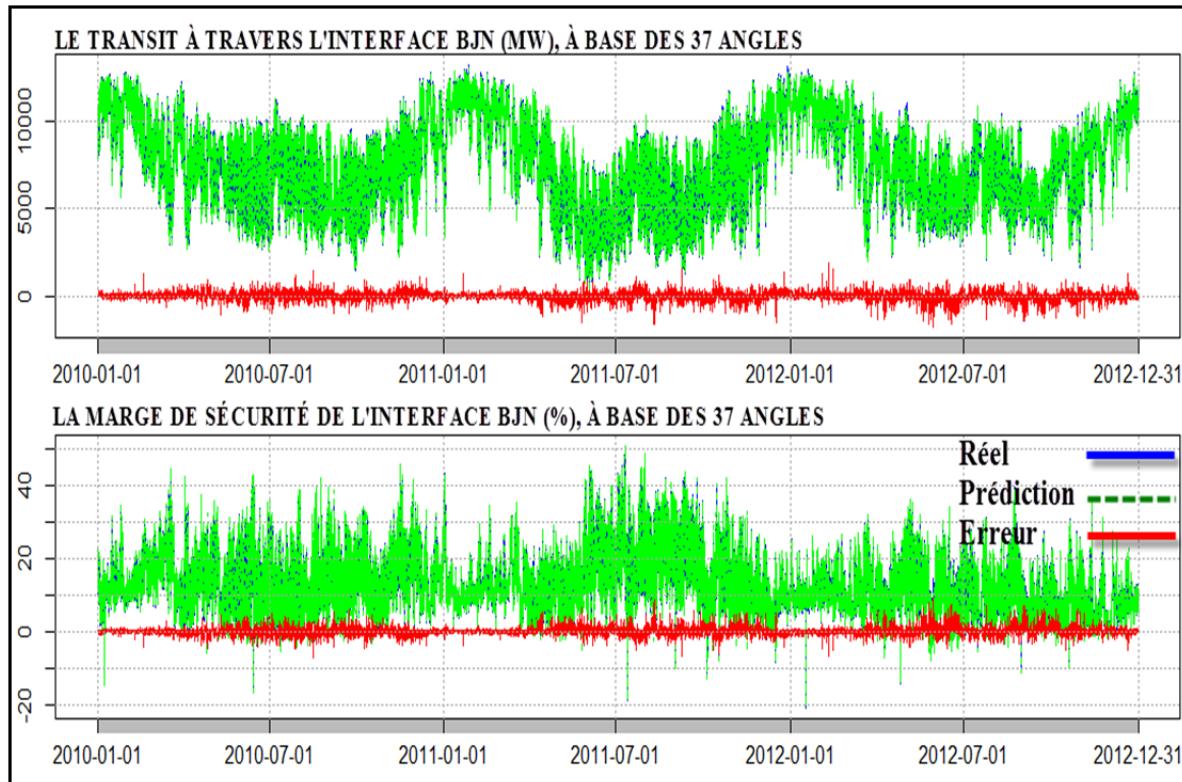


Figure 5.15 Les erreurs de prévisions par les FA du transit actif à travers l'interface BJT et de la marge de sécurité associée sur une période de 3 ans par 9 minutes

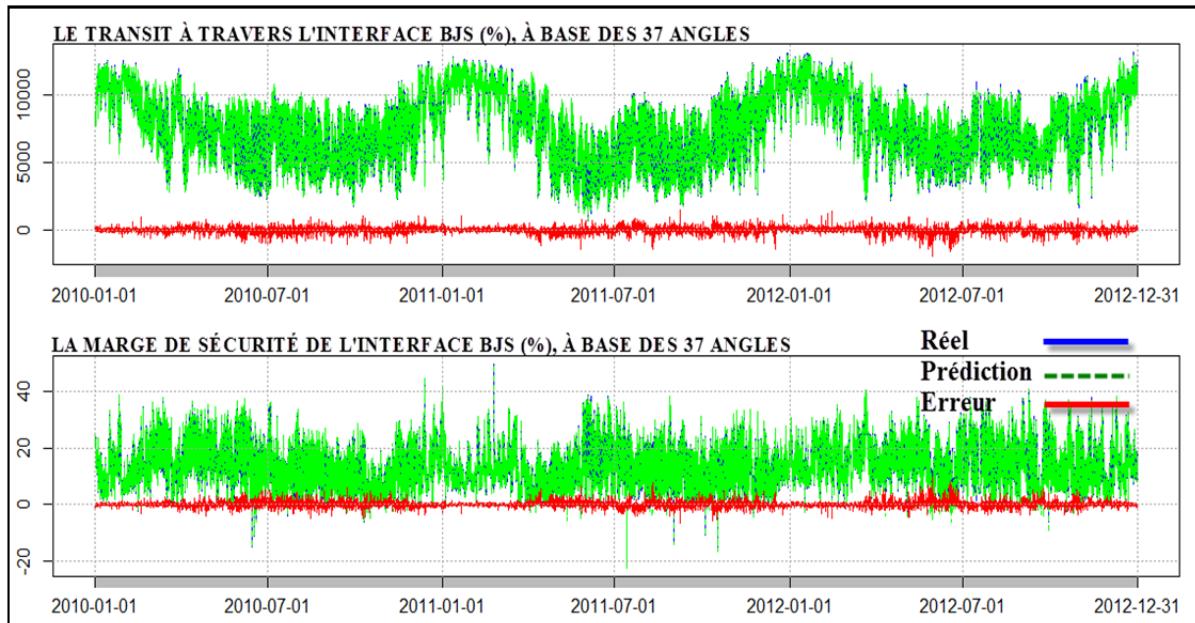


Figure 5.16 Les erreurs de prévisions par les FA du transit actif à travers l'interface BJS et de la marge de sécurité associée sur une période de 3 ans par 9 minutes

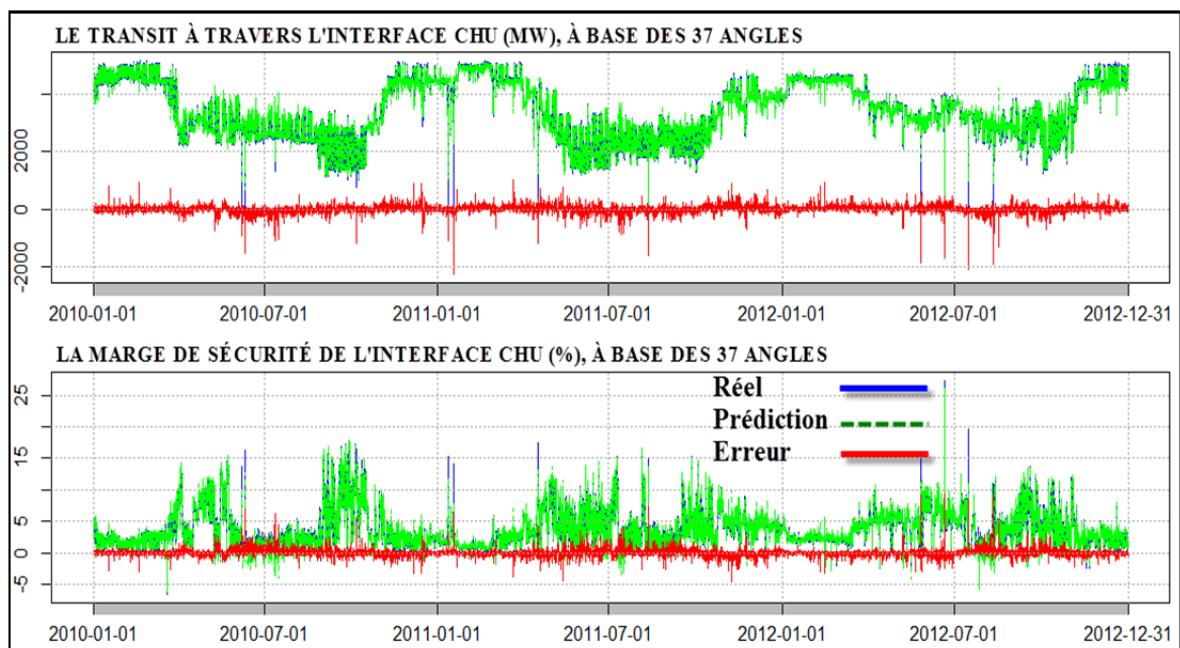


Figure 5.17 Les erreurs de prévisions par les FA du transit actif à travers l'interface CHU et de la marge de sécurité associée sur une période de 3 ans par 9 minutes

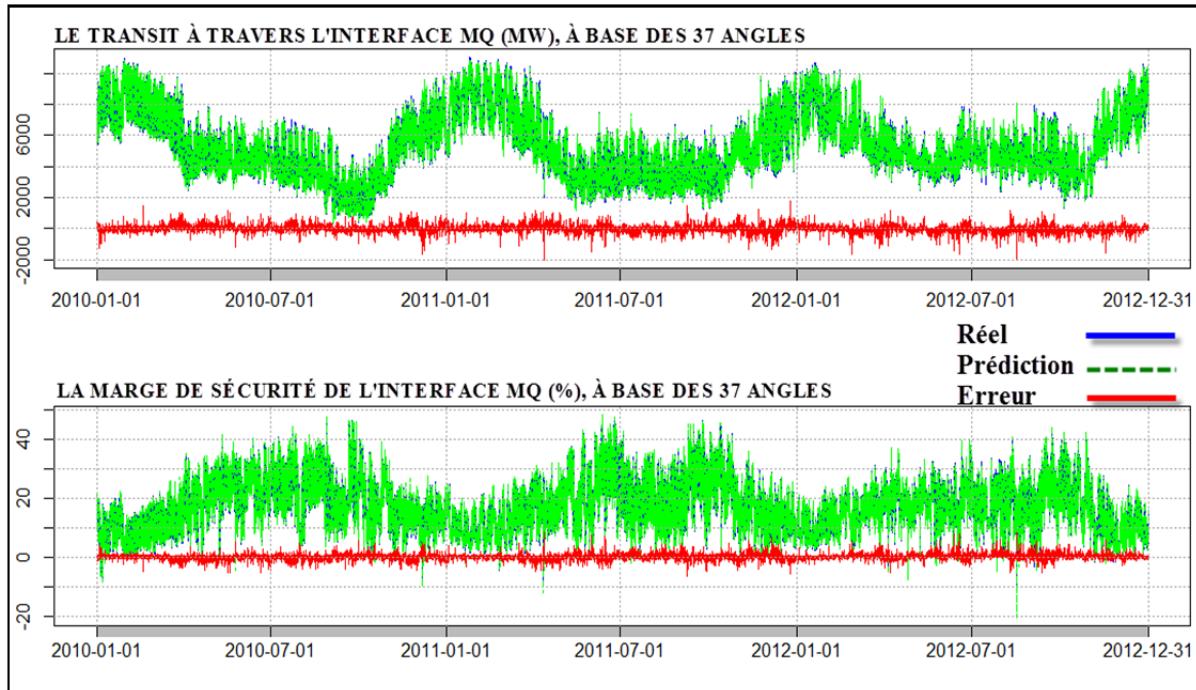


Figure 5.18 Les erreurs de prévisions par les FA du transit actif à travers l'interface MQ et de la marge de sécurité associée sur une période de 3 ans par 9 minutes

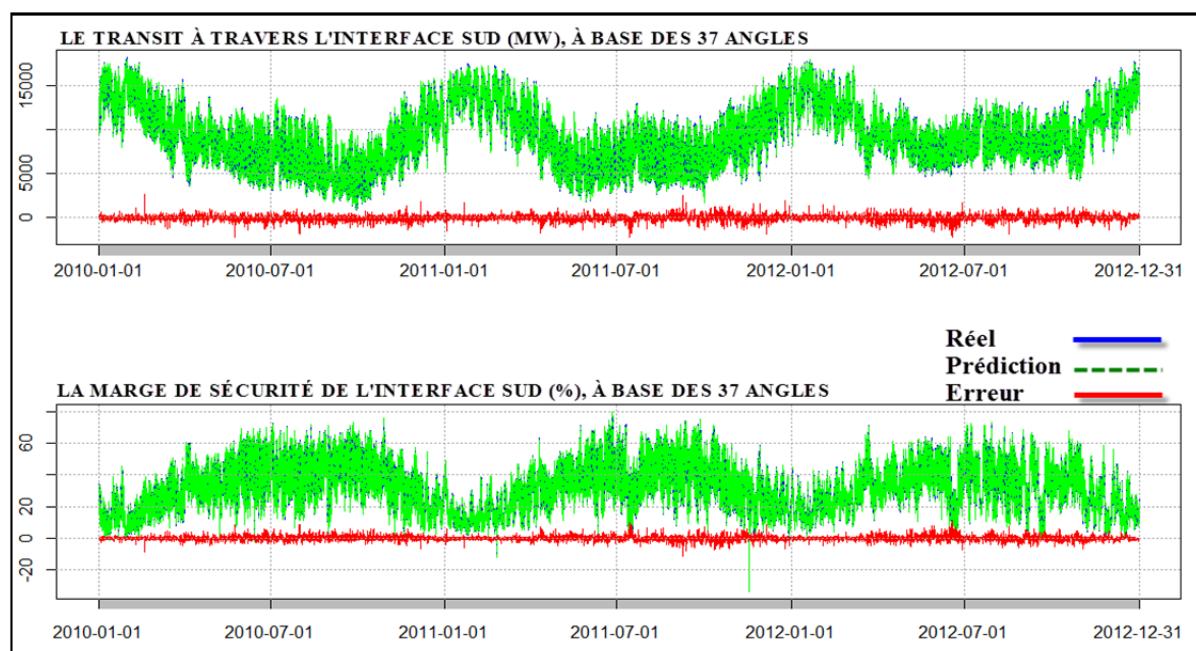


Figure 5.19 Les erreurs de prévisions par les FA du transit actif à travers l'interface SUD et de la marge de sécurité associée sur une période de 3 ans par 9 minutes

Tableau 5.9 Caractéristiques statistiques de l'ensemble test extrait de la base de données de 3 ans à 9 minutes

Transits actifs (MW)	BJN	BJS	MQ	CHU	SUD
Minimum	326,2	882,9	426,8	-56,63	930,7
1 ^{er} Quartile	6 115,5	6 252,8	4 006,4	2 798,20	7 624,6
Médiane	8 007,9	8 057,4	5 265,7	3 427,87	9 872,9
Moyenne	7 802,5	7 852,2	5 456,6	3 487,01	9 827,0
3 ^{ème} quartile	9 594,4	9 577,5	6 661,8	4 417,39	11 963,8
Maximum	13 130,2	13 040,0	10 918,2	5 125,45	17 995,5

Tableau 5.10 Caractéristiques statistiques de l'ensemble prédit par les FA en fonction de tous les angles à base de données de 3 ans à 9 minutes

Prévision par les FA	BJN	BJS	MQ	CHU	SUD
Minimum	327	887,6	505,2	195,3	1 037
1 ^{er} Quartile	6 163	6 281,8	4 040,5	2 805,7	7 677
Médiane	7 998	8 050,2	5 257,4	3 417,8	9 877
Moyenne	7 807	7 856,2	5 455,8	3 487,4	9 832
3 ^{ème} quartile	9 572	9 567,2	6 625,5	4 401,0	11 929
Maximum	12 933	12 884,5	10 732,9	5 076,1	17 922

Tableau 5.11 Caractéristiques statistiques des erreurs de prévisions en fonction de tous les angles par les modèles de FA établis sur l'ensemble d'apprentissage de données de 2010 à 2012 à 9 minutes

Erreur brute de prévision	BJN	BJS	MQ	CHU	SUD
Minimum	1 790,258	1 969,884	1 634,0200	2 257,985	2 229,428
1 ^{er} Quartile	62,828	56,740	62,7056	26,260	82,032
Médiane	3,067	3,915	0,4212	4,286	4,341
Moyenne	4,625	3,917	0,8218	0,397	5,286
3 ^{ème} quartile	69,999	63,187	67,3998	34,761	88,953
Maximum	1 846,932	1 472,204	1 760,3410	995,090	2 618,440
Écart type	176,18997	160,90341	1 76,02026	96,17156	224,18306

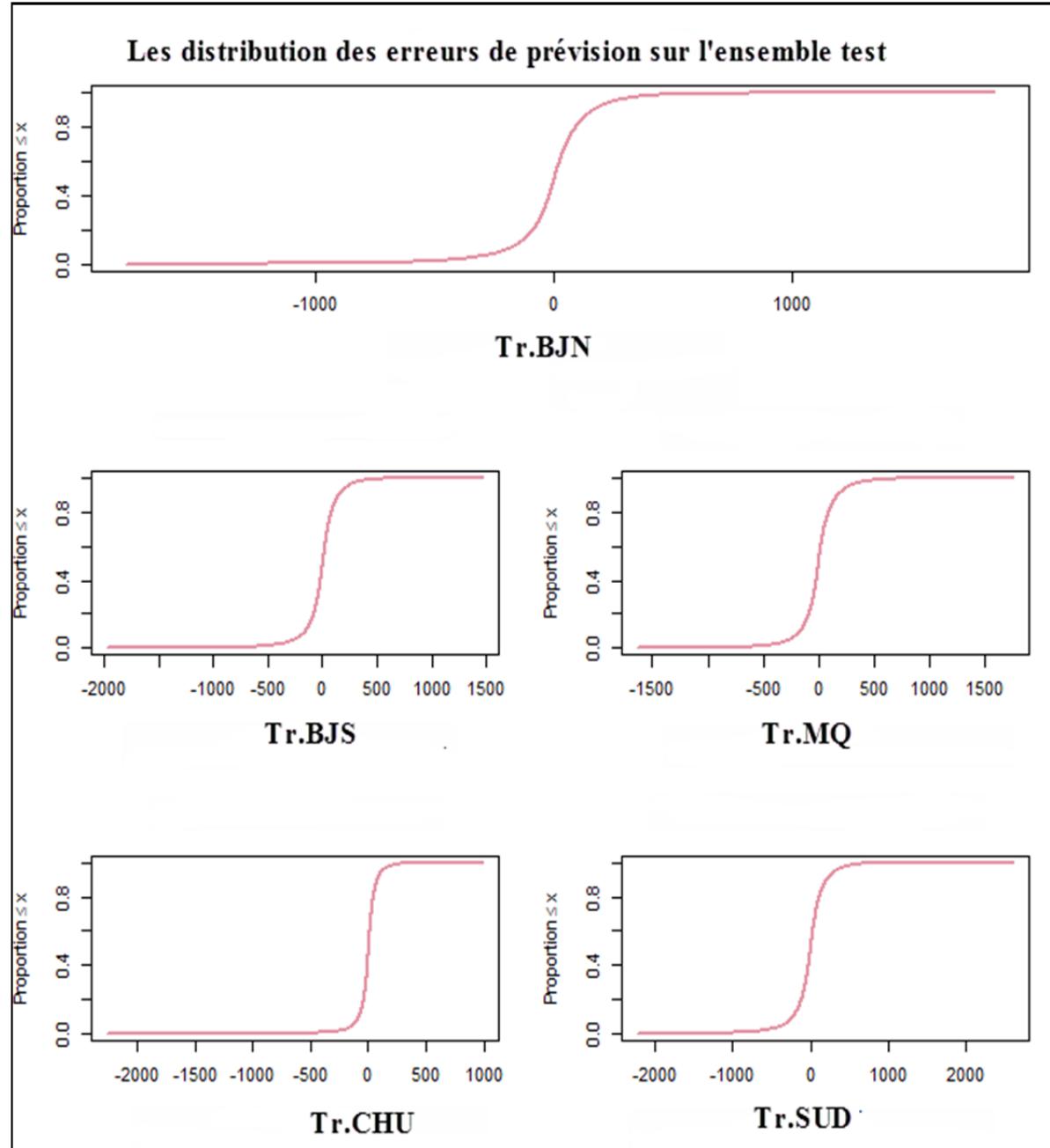


Figure 5.20 Les distributions des erreurs de prévisions des transits actifs du sous-ensemble test en fonction de tous les angles par les modèles de FA établis sur l'ensemble d'apprentissage de données de trois ans à 9 minutes

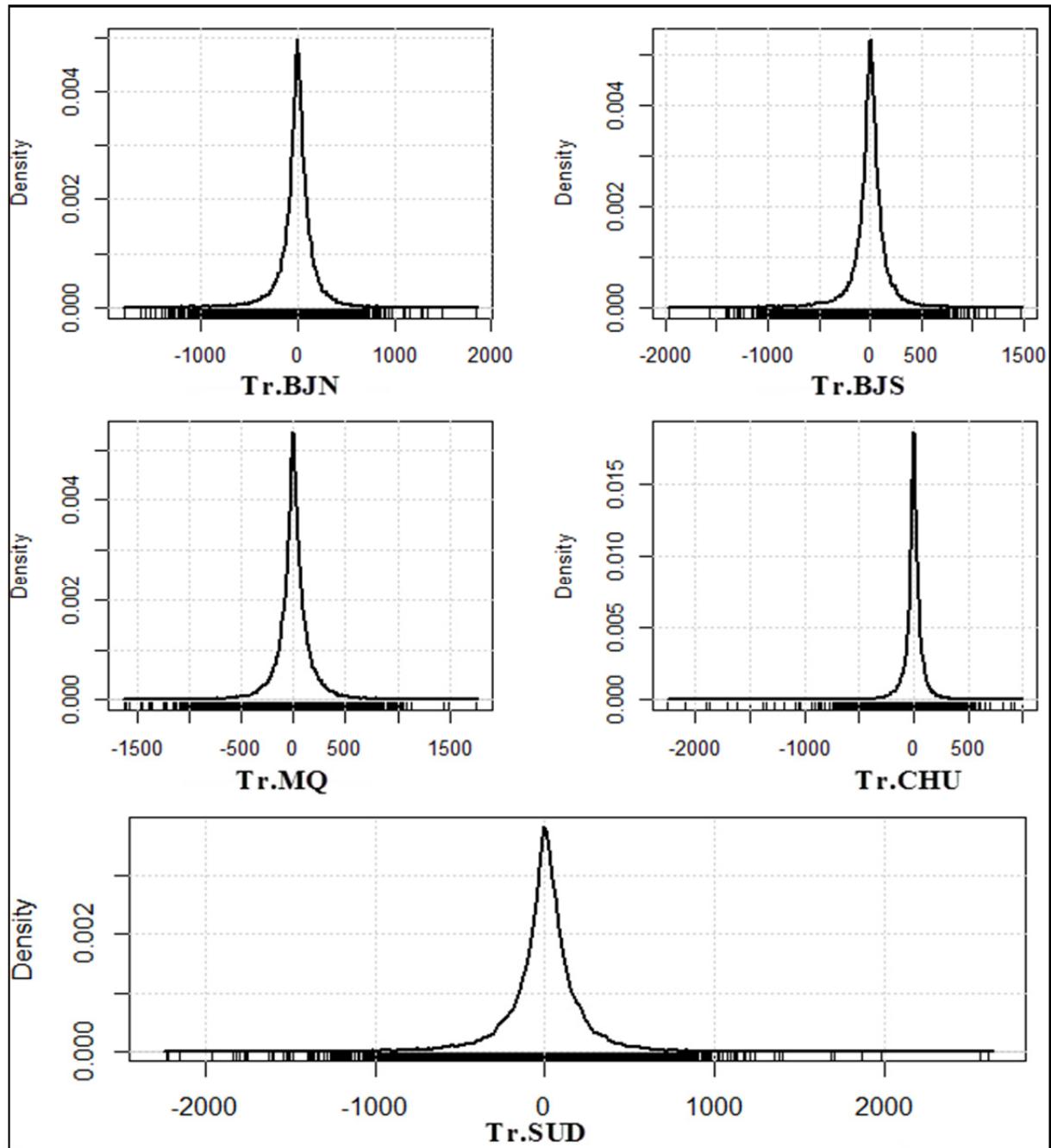


Figure 5.21 Les densités des erreurs de prévisions des transits actifs du sous-ensemble test en fonction de tous les angles par les modèles de FA établis sur l'ensemble d'apprentissage de données de trois ans à 9 minutes

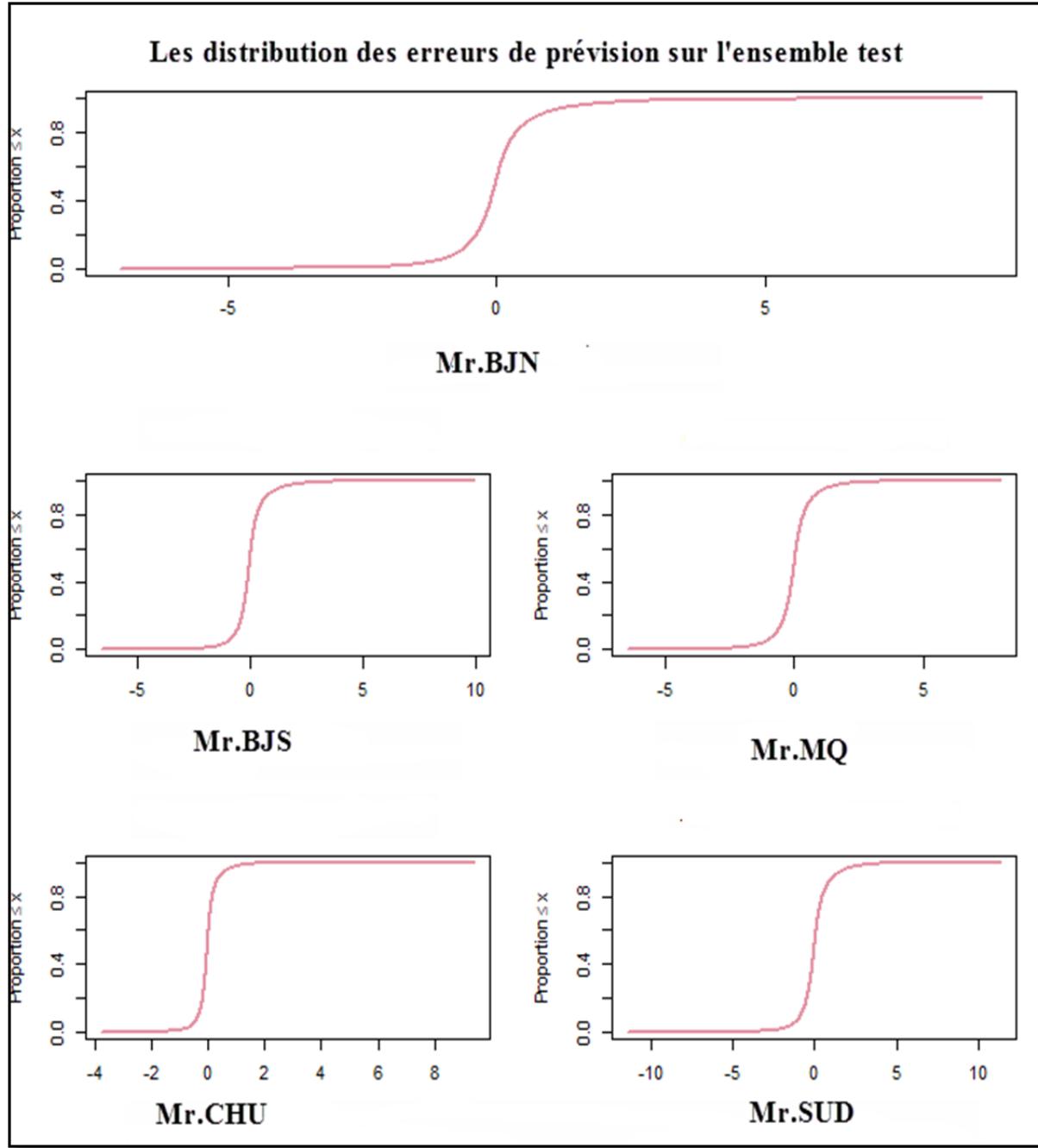


Figure 5.22 Les distributions des erreurs de prévisions des marges de sécurité du sous-ensemble test en fonction de tous les angles par les modèles de FA établis sur l'ensemble d'apprentissage de données de trois ans à 9 minutes

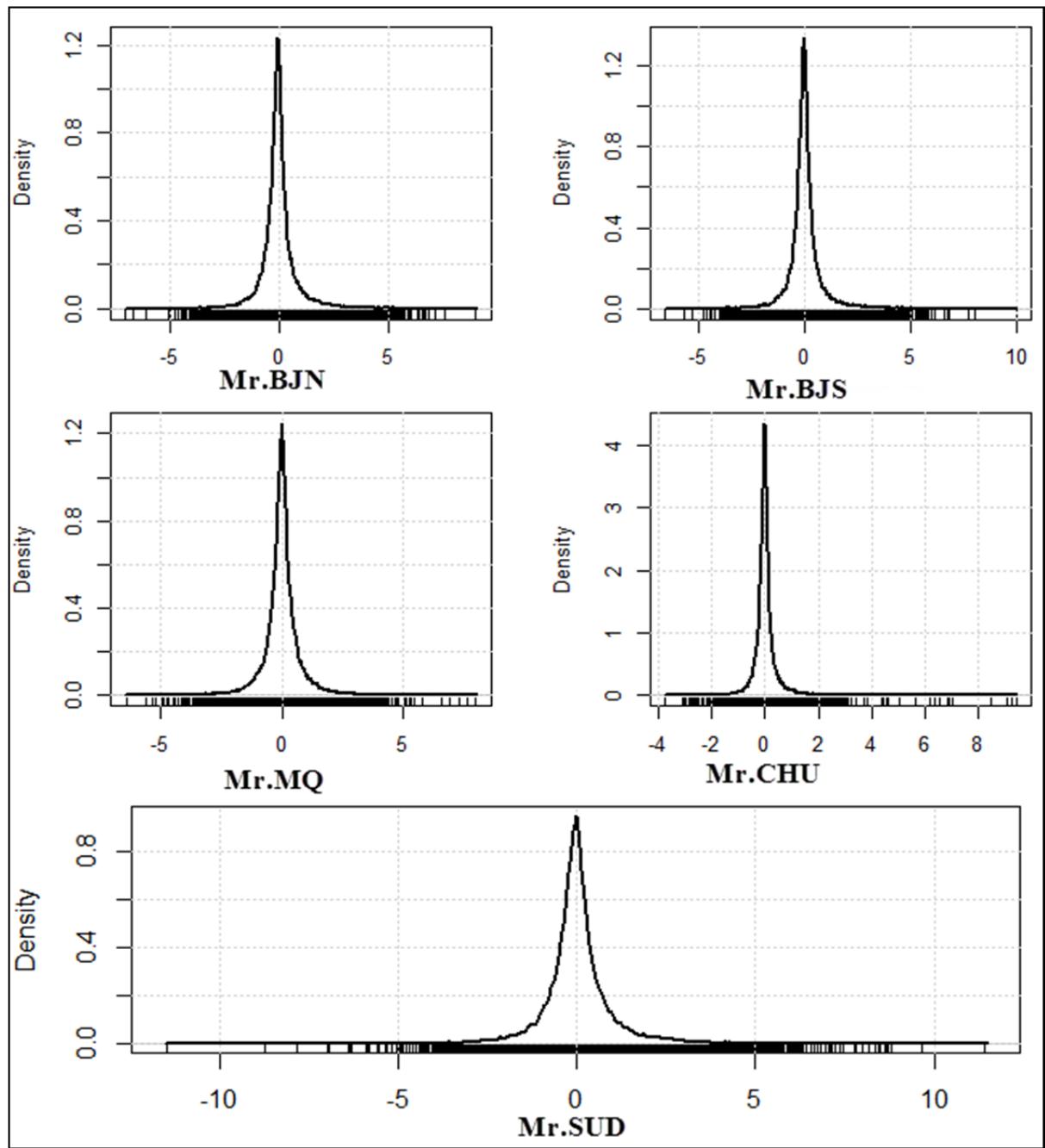


Figure 5.23 Les densités des erreurs de prévisions des marges de sécurité du sous-ensemble test en fonction de tous les angles par les modèles de FA établis sur l'ensemble d'apprentissage de données de trois ans à 9 minutes

5.5 Conclusion

Dans ce chapitre, nous avons établi que les modèles de prévisions des transits et des marges de sécurité de toutes les interfaces, en fonction de tous les angles, ont un R dépassant 99,5%. Ce qui signifie que nos modèles statistiques de prévisions des marges de sécurité et des transits actifs des corridors principaux du réseau de transport, en fonction des angles de phase de tension des barres à 735kV, sont cohérents relativement aux variables explicatives, pertinents pour la prévision de la sécurité et robuste relativement à la généralisation pour de nouvelles données et sur de nouvelles périodes. D'où la validité de notre hypothèse de départ que les angles de phase des barres à 735 kV sont des bons prédicteurs de la stabilité du réseau.

Par souci d'avoir des modèles de prévisions parcimonieux, nous avons générés des modèles de prévisions à base des angles des barres médiods. D'après les courbes des densités des erreurs de prévisions, Figure 5.13 et Figure 5.14, montrant un comportement asymptotique des bruits blancs gaussien, nous laisse estimer que les modèles élaborés sont d'une bonne qualité. Cependant, en les comparant aux modèles de prévisions sur la même période, à base de tous les angles et qui correspondent à la Figure 5.11 et la Figure 5.12, nous jugeons que ces modèles simplifiés nécessitent une amélioration. La qualité relative de ces modèles apparaît dans l'importance de l'effet de bord. L'amélioration la plus plausible à apporter est de revoir la méthode de choix des variables explicatives.

Toutes les prévisions en MW, concernant les données de 2010, relatives aux cinq interfaces semblent très précises. Les R obtenus sont supérieurs à 99,8%. Le R le plus élevé est supérieur à 99,9% et il correspond aux interfaces SUD, BJR et BJS. Selon (Energy, 2008), ce pourcentage de précision est très favorable pour la surveillance de la sécurité. Cependant, pour les précisions en %, les R sont de moindre qualité. L'amélioration de la précision en % peut se jouer sur la définition elle-même de la marge. Dans notre travail, la marge, telle qu'elle est définie, est très complexe. Elle est en fonction des variables complexes à savoir le transit, la limite de transit calculée par LIMSEL et la charge globale. Cependant, nous avons amorcé l'analyse de la marge définie comme le rapport de la différence entre la limite du

transit et le transit par rapport à la limite et nous avons obtenus des résultats de prévision préliminaires beaucoup plus précis pour la plupart des interfaces.

Avec la méthode des FA à base d'arbres de régressions, nous avons atteint un niveau significatif de précision et de robustesse dans la prévision des limites de sécurité dynamique, en fonction des angles de phase de tension des barres du réseau de transport. En conséquence, nous avons démontré la consistance de l'approche adoptée pour légitimer les angles de phase comme des prédicteurs de la stabilité dynamique. En plus, en se fiant à la qualité des erreurs sans-biais obtenue, nous jugeons que l'ajout des données plus précises comme celles obtenues par les synchrophaseurs ne sera que bénéfique pour l'amélioration de la précision de nos modèles.

Les résultats décrits dans ce chapitre ont fait l'objet de l'article (Kaci et al., 2014b) où, les auteurs ont mis l'accent sur la potentiel de cette méthode à rajouter un plus dans l'évaluation de la sécurité du réseau en temps réel. En effet, la lecture chronologique des phaseurs d'estimateur d'état sur une année complète, effectuée pour vérifier les niveaux de dépassement d'angle de phase et les marges de sécurité confirme que l'approche prononcée est tout à fait réaliste et pourrait être facilement étendue aux données des synchrophaseurs actuelles de PDC d'Hydro Québec (Kamwa et al., 2006; Kamwa et al., 2009) pour une migration potentielle à la salle de contrôle.

CHAPITRE 6

PRÉVISIONS ANTICIPÉES DES ÉCARTS ANGULAIRES DES TENSIONS ET LEURS INTÉGRATION DANS LA CONDUITE DU RÉSEAU

6.1 Introduction

Dans les opérations, il est souvent avantageux de combiner les prévisions de charges et de transactions de puissance avec les plans de maintenance pour se préparer à l'avenir (Grenier et al., 2006; Hong et Shahidehpour, 2015; HQD, 2010; Slaven, Brian et Hongming, 2015). Pouvons-nous entreprendre le même procédé avec les angles de phase de tension des barres du réseau de transport ?

Pour répondre à cette question, nous subdivisons notre démarche de travail en deux étapes. Nous débutons par proposer différents modèles de prévisions des angles de phase de tension des barres du réseau de transport. Sur la base des résidus des modèles proposés, nous effectuons une étude de comparaison entre les algorithmes d'apprentissage statistique, rpart, GLM et NNET, dans le but de choisir le mieux adapté à notre contexte. Nous poursuivons notre étude avec l'algorithme adopté GLM. Nous évaluons les performances de prévisions par les mesures des erreurs et nous concluons cette étape par une discussion des résultats.

En seconde étape, nous utilisons les modèles de prévisions, développés auparavant par le biais des FA, mais en considérant comme entrées les nouveaux angles prévus par GLM. Avec la mesure des erreurs de prévisions des transferts actifs et des marges de sécurité des interfaces clefs du réseau de transport, nous discutons la possibilité d'intégrer ce procédé dans la salle de conduite.

6.2 Modélisation des angles de phase de tension des barres du réseau de transport

À partir des constations et résultats obtenus, dans les chapitres précédents, sur les angles de phase de tension en tant que prédicteurs de la stabilité du réseau de transport, nous proposons de modéliser leurs comportements dans un objectif de prévision. Cette modélisation se fera dans le même sens que celle habituellement utilisée dans d'autres travaux antérieurs

concernant les charges (HQD, 2010; Slaven, Brian et Hongming, 2015). En s'inspirant du modèle à base de temps présenté par (Ferryman et Amidan, 2012), nous avons construit nos modèles basés sur celui-ci, et ce en rajoutant des termes qui ont mené à une amélioration visible de ce dernier.

Les modèles élaborés pour les prévisions des angles de phase de tension sont en fonction des variables catégorielles et réelles, énumérées dans le Tableau 6.1. Les variables catégorielles font référence à une base de temps basée sur les mois d'une année, les jours d'une semaine et huit intervalles horaires d'un jour, de trois heures chacun. Elles sont de différents mode ou niveau. La variable *Mois* identifie le mois de l'année et elle est de 12 modes. La variable *Jour* identifie le jour de la semaine et elle est de 7 modes et la variable *S* identifie un intervalle de temps de 3 heures sur un jour et elle est de 8 modes. Les variables catégorielles sont définies de sorte :

Mois : Janvier, Fevrier, Mars,..., Decembre

Jour : Lundi, Mardi,..., Dimanche

S : S_1, S_2, \dots, S_8

Tel que,

$$S_1 = [00:00\ 03:00], \quad S_2 = [03:00\ 06:00], \dots, \quad S_8 = [21:00\ 00:00] \quad (6.1)$$

Les variables réelles sont les variables à prévoir et les variables explicatives extraites de la base de données EMS/SCADA. Elles sont définies comme suit :

BG : le besoin global;

Pred.Charge : la valeur de la charge prévue la veille ;

Pred24H.Charge : la valeur de la charge prévue 24 H à l'avance ;

θ_t : la valeur de l'angle de phase de la tension à l'instant t (en minutes) ;

$\theta_{t-\Delta}$: la valeur de l'angle de phase de la tension, Δ en avance par rapport à t , tel que

$\Delta \in \{24\ H, 4\ H, 1\ H\}$;

e_t : l'erreur de prévision à l'instant t .

Tableau 6.1 Les variables des modèles des angles de phase de tension des barres à 735kV

Les variables catégorielles		Les variables numériques
Nom	Mode	
<i>Mois</i>	12 modes : Janvier, Février, Mars, Avril, Mai, Juin, Juillet, Août, Septembre, Octobre, Novembre, Décembre	BG $Pred.Charge$ $Pred24H.Charge$ θ_t
<i>Jour</i>	7 modes : Lundi, Mardi, Mercredi, Jeudi, Vendredi, Samedi, Dimanche	θ_{t-1H}
<i>S</i>	8 modes : S1, S2, S3, S4, S5, S6, S7, S8	θ_{t-4H} θ_{t-24H} e_t

Ainsi, la structure générale des modèles proposés de prévisions des angles de phase de tension des barres à 735kV est donnée par (6.2). Celle-ci prend en considération les effets saisonniers, les effets diurnes, la charge globale du réseau et l'état passé de l'angle.

$$\theta_t = \theta_0 + \theta_1 Mois + \theta_2 Jour + \theta_3 S + \theta_4 BG + \theta_5 Pred.Charg e + \\ \theta_6 Pred24H.Charg e + \theta_7 \theta_{t-\Delta} + e_t \quad (6.2)$$

6.3 Prévisions des angles de phase de tension des postes à 735kV du réseau de transport

À base de l'apprentissage statistique fondé sur les données à une minute de 2010 à 2012 des angles de phase de tension des postes à 735kV du réseau de transport, nous formons les modèles de prévisions sur le même principe expliqué dans les chapitres antérieurs. Nous ajustons, ainsi, les modèles de la forme (6.2) sur les 75% seulement des données de

l'ensemble d'observations et nous testons la qualité de la généralisation de ces derniers sur un autre ensemble qui contient les données restantes.

À base des variables explicatives données dans le Tableau 6.1, nous modélisons tous les angles de phase de tension. Pour chaque angle de phase, nous élaborons trois modèles de prévisions selon la valeur de Δ ; un modèle de prévision vingt-quatre heures en avance, un modèle de prévision quatre heures en avance, un modèle de prévision une heure en avance.

À titre d'illustration, pour la prévision de l'angle de phase de tension D06, nous présentons les modèles de prévisions définis par le biais des expressions (6.3), (6.4) et (6.5) :

1. La modélisation une journée en avance de l'angle de phase de tension D06 par le modèle G06.24H :

$$G06.24H = \beta_0 + \beta_1 Mois + \beta_2 Jour + \beta_3 S + \beta_4 BG + \beta_5 Pred.Charge + \beta_6 Pred24H.Charge + \beta_7 D06_{t-24*60} + e_t \quad (6.3)$$

2. La modélisation quatre heures en avance de l'angle de phase de tension D06 par le modèle G06.4H :

$$G06.4H = \theta_0 + \theta_1 Mois + \theta_2 Jour + \theta_3 S + \theta_4 BG + \theta_5 Pred.Charge + \theta_6 Pred24H.Charge + \theta_7 D06_{t-4*60} + r_t \quad (6.4)$$

3. La modélisation une heure en avance de l'angle de phase de tension D06 par le modèle G06.1H :

$$G06.1H = \alpha_0 + \alpha_1 Mois + \alpha_2 Jour + \alpha_3 S + \alpha_4 BG + \alpha_5 Pred.Charge + \alpha_6 Pred24H.Charge + \alpha_7 D06_{t-60} + s_t \quad (6.5)$$

où e_t , r_t et s_t sont respectivement les résidus des modèles $G06.24H$, $G06.4H$ et $G06.1H$.

6.4 Les critères d'évaluation de la qualité des modèles de prévisions

Pour mieux cerner l'efficacité des modèles proposés, nous utilisons une multitude de critères comme moyen d'évaluation de la précision. Les critères considérés sont la moyenne de l'erreur absolue MAE (*Mean absolute error*), la moyenne des erreurs au carré MSE (*Mean square error*), le facteur de corrélation R et le coefficient de détermination R-carré ou Rsq. Ces différentes mesures sont calculées à partir des équations (6.6) à (6.9) respectivement. Dans ces expressions, D représente la valeur réelle de l'angle de phase de tension et G représente la valeur prévue de cet angle par le modèle de prévision.

$$MAE_{Dm} = \frac{\sum_{k=1}^n |G_k - D_k|}{n} \quad (6.6)$$

$$MSE_{Dm} = \frac{\sum_{k=1}^n (G_k - D_k)^2}{n} \quad (6.7)$$

$$R_{Dm} = cor(D, G) \quad (6.8)$$

$$Rsq = 1 - \frac{MSE_G}{\text{var}(G)} \quad (6.9)$$

6.5 Les performances des différents algorithmes de régression

La construction des modèles de prévisions de type (6.2) revient à traiter un problème d'apprentissage statistique de type régression. La variable à expliquer est numérique et les variables explicatives sont de type catégoriel et numérique. Les modèles de même nature que (6.3), (6.4) et (6.5) peuvent être traité aisément par les algorithmes d'arbre de régression (rpart), de modèles linéaire généralisé (GLM) ou par l'algorithme des réseaux de neurones

(NNET), sous l'environnement **R**. Afin de voir lequel de ces algorithmes est le mieux adapté à notre cas, nous avons modélisé selon (6.5), pour une prévision anticipée d'une heure en avance, tous les angles de phase de tension par les trois algorithmes. Puis, par les mesures (6.6), (6.8) et (6.9), nous avons évalué la qualité de ces prévisions de chaque angle par ces trois algorithmes.

Lors de la modélisation de tous les angles de phase de tension avec les algorithmes de l'arbre de décision (rpart) et les réseaux de neurones (NNET), nous avons choisi les paramètres de manière à avoir les meilleurs résultats. Pour l'algorithme rpart, nous avons forcé le paramètre de complexité à zéro pour entraîner la construction de l'arbre à sa profondeur maximale. Pour l'algorithme NNET, nous avons pris le nombre de couches égal à 10. L'ensemble des résultats de cette étude est mentionné dans le tableau-A III-1 de l'annexe III. Il est clair que pour la prévision de la même variable, si l'erreur de prévision MAE diminue, les facteurs R et Rsq augmentent. Autrement dit, autant que les erreurs de prévisions de la même variable diminuent, les précisions de ces prévisions s'améliorent. Les deux représentations, Figure 6.1 et Figure 6.2, montrent les variations des caractéristiques d'évaluation de prévisions MAE et R selon la méthode de prévision utilisée. Les résultats obtenus montrent que les méthodes de GLM et NNET se distinguent de la méthode de l'arbre de décision (rpart), concernant leur qualité de prévision. La modélisation avec les deux méthodes GLM et NNET mène à la même qualité de prévision et ceci pour tous les angles de phase. En revanche, la construction des modèles avec l'algorithme GLM nécessite moins de temps qu'avec l'algorithme NNET.

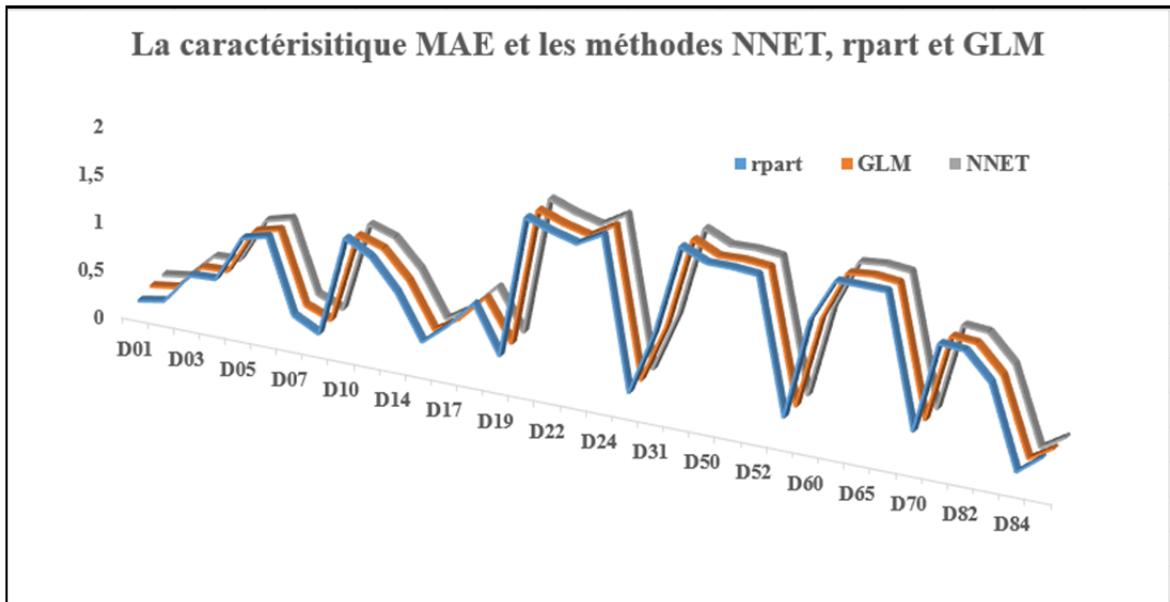


Figure 6.1 Les méthodes NNET, rpart et GLM et la caractéristique MAE

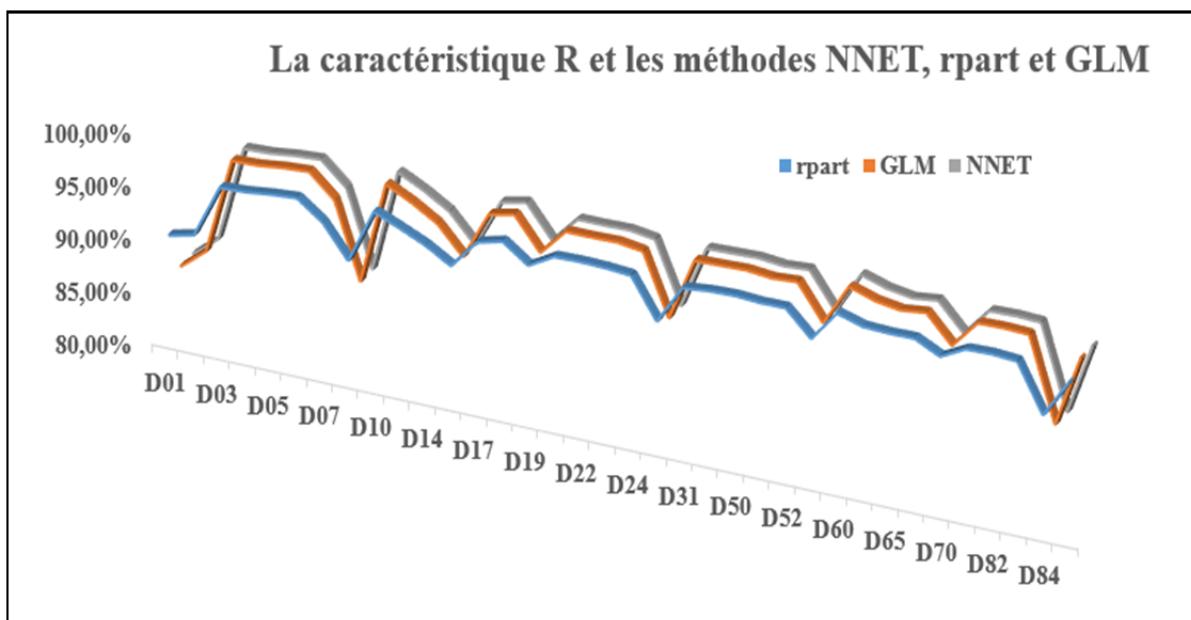


Figure 6.2 Les méthodes NNET, rpart et GLM et la caractéristique R

Sur la base de ces résultats, nous concluons que dans notre contexte, la méthode GLM se veut la plus appropriée pour la prévision des angles à base des modèles de type (6.5). Ainsi, dans la partie qui reste de notre travail, nous utilisons l'algorithme GLM pour anticiper la prévision de tous les angles de phase de tension du réseau de transport, quatre heures et vingt-quatre heures en avance.

Un tel constat concernant la performance d'une méthode simple comme GLM devant des méthodes cotées dans le domaine des réseaux électriques n'est pas une exception dans la littérature. En effet, comme montré dans (Hong, Wang et Willis, 2011) les modèles de régression polynomiales surpassent, dans certains cas, des approches plus sophistiquées comme les réseaux de neurones.

6.6 Prévisions anticipées des angles de phase de tension par GLM

Une fois qu'on s'est fixé sur le choix de la méthode de GLM, nous modélisons chacun des 37 angles de phase de tension, suivant les trois modèles (6.3), (6.4) et (6.5). Les résultats de prévisions anticipées d'une/ (des) heure(s) en avance de chaque angle de phase de tension, par le biais de la méthode de GLM, regroupent des informations sur la distribution du résidu du modèle, les paramètres (coefficients) estimés de chaque variable et les tests de significativité des coefficients (leurs intervalles de confiance, la valeur de la statistique et la P-value du test).

Le Tableau 6.2 Un aperçu sur le modèle G06.1H avec la méthode GLM donne un aperçu sur le modèle de prévision, une heure en avance, de l'angle de phase de tension de la barre D06. Les caractéristiques des modèles de prévisions, quatre heures en avance et vingt-quatre heures en avance, du même angle de phase de tension D06 sont mentionnées dans l'annexe III.

Tableau 6.2 Un aperçu sur le modèle G06.1H avec la méthode GLM

	Estimate	Std. Error	t value	Pr(> t)	(95%)	CI	code
(Intercept)	-2,4077	1,400e-02	-171,944	< 2e-16	-2,4347	-2,3807	***
Moisavril	-0,4247	7,004e-03	-60,633	< 2e-16	-0,4382	-0,4112	***
Moisdécembre	-0,4812	9,704e-03	-49,585	< 2e-16	-0,4999	-0,4625	***
Moisfévrier	-0,3275	1,009e-02	-32,466	< 2e-16	-0,3469	-0,3081	***
Moisjanvier	-0,4853	1,056e-02	-45,958	< 2e-16	-0,5057	-0,4650	***
Moisjuillet	-0,1090	6,812e-03	-16,001	< 2e-16	-0,1221	-0,0959	***
Moisjuin	0,1730	6,786e-03	25,497	< 2e-16	0,1599	0,1861	***
Moismai	-0,2563	6,685e-03	-38,348	< 2e-16	-0,2692	-0,2435	***
Moismars	-0,3195	7,995e-03	-39,963	< 2e-16	-0,3349	-0,3041	***
Moisnovembre	-0,0485	7,932e-03	-6,110	9,99e-10	-0,0637	-0,0332	***
Moisoctobre	-0,3620	7,144e-03	-50,671	< 2e-16	-0,3758	-0,3482	***
Moisseptembre	0,1106	6,878e-03	16,077	< 2e-16	0,0973	0,1238	***
Joursjeudi	-0,1436	5,355e-03	-26,807	< 2e-16	-0,1539	-0,1332	***
Jourslundi	-0,1116	5,325e-03	-20,948	< 2e-16	-0,1218	-0,1013	***
Joursmardi	-0,1362	5,366e-03	-25,384	< 2e-16	-0,1465	-0,1259	***
Joursmercredi	-0,0925	5,369e-03	-17,231	< 2e-16	-0,1029	-0,0822	***
Jourssamedi	-0,0415	5,166e-03	-8,040	9,01e-16	-0,0515	-0,0316	***
Joursvendredi	-0,1624	5,294e-03	-30,675	< 2e-16	-0,1726	-0,1522	***
Sample2H	1,1036	5,657e-03	195,083	< 2e-16	1,0927	1,1145	***
Sample3H	1,4090	6,499e-03	216,820	< 2e-16	1,3965	1,4216	***
Sample4H	0,0473	6,699e-03	7,057	1,70e-12	0,0344	0,0602	***
Sample5H	-0,2704	6,482e-03	-41,718	< 2e-16	-0,2829	-0,2579	***
Sample6H	0,7289	6,709e-03	108,639	< 2e-16	0,7160	0,7418	***
Sample7H	-0,4045	6,794e-03	-59,537	< 2e-16	-0,4176	-0,3914	***
Sample8H	-1,3098	5,863e-03	-223,387	< 2e-16	-1,3211	-1,2985	***
Deph.706Anc1H	0,7895	4,983e-04	1 584,278	< 2e-16	0,7885	0,7905	***
BQ.PR	0,0001	3,452e-06	29,205	< 2e-16	0,0001	0,0001	***
BQ.PR24H	-0,0003	3,193e-06	-104,246	< 2e-16	-0,0003	-0,0003	***
BG	0,0005	1,751e-06	280,162	< 2e-16	0,0005	0,0005	***

Niveau de signification de code : **0 '***'** **0,001 '**'** **0,01 '*'** **0,05 '!** **0,1 ''** **1**

Déviance Résiduelle:

Min	1Q	Médiane	3Q	Max
-43,689	-0,972	-0,042	0,901	39,948

L'erreur résiduelle standard est : 1,414 sur 1 025 779 degrés de liberté.

Multiple R-squared: 0,955 9. Adjusted R-squared: 0,955 9

F-statistic: 7,947e+05 on 28 and 1 025 779 DF, p-value: < 2,2e-16

La lecture de ces tableaux et l'interprétation de ces résultats sont basées sur la référence (Ouellet, Belley-F. et Leblond, 2005; Rakotomalala, 2011). Le Rsq est le coefficient de détermination. Il exprime la manière dont les données réelles s'ajustent au modèle. C'est le premier critère que l'on regarde dans une régression (Rakotomalala, 2011). Il exprime la part de la variance expliquée par le modèle ou la proportion de la variation de la variable dépendante (à expliquer) qui est expliquée par les variables indépendantes (les variables explicatives). Le Rsq qui semble petit ne veut pas dire nécessairement que le modèle est rejeté. Un faible Rsq donne plutôt l'indice qu'il manquerait des variables explicatives pertinentes au modèle. Dans le cas de la régression, des erreurs d'interprétation peuvent survenir. Par exemple, l'objectif d'interprétation n'est pas de comparer les coefficients de deux variables indépendantes du modèle afin de vérifier laquelle a une influence plus grande sur la variable dépendante (Ouellet, Belley-F. et Leblond, 2005).

Dans notre cas, on s'intéresse à la relation entre une variable explicative en particulier et la variable à expliquer. Donc, les autres représentent des variables de contrôle que nous devrons par ailleurs interpréter. Ainsi, nous voulons souligner que le lien de causalité prévaut entre la variable explicative et la variable à expliquer d'intérêt et non pas entre les variables explicatives.

Dans la première colonne du Tableau 6.2, on retrouve nos variables explicatives. Les coefficients sont dans la seconde colonne (les coefficients sont toujours exprimés dans les

unités de la variable dépendante). Dans la troisième colonne, on retrouve les écarts types estimés. La quatrième colonne donne la statistique t qui est interprétée à l'aide de la table de Student. Elle est essentielle afin de déterminer si les coefficients sont significatifs. La colonne « *code* » pour exprimer le niveau de la significativité des coefficients. Finalement, les dernières colonnes donnent l'intervalle de confiance à un niveau de 95%. Avec le test de Student, $F=t^2$, on prend un risque $\alpha = 5\%$ d'accepter une variable. Les variables explicatives des modèles proposés dans notre travail présentent un t^2 élevé. Elles sont très significatives, en tous les cas avec un p-value largement en-deçà du niveau que l'on s'est fixé (5%). Pour les variables non significatives, le p-value est $> 5\%$.

Le Tableau 6.3 récapitule le résultat du test ANOVA sur le modèle de prévision anticipée une heure en avance de l'angle de phase de tension D06. Il rapporte l'effet des variables explicatives sur le modèle de la variable continue D06. Ce test permet, en ajoutant séquentiellement les variables explicatives, de tester la significativité de chacune par la diminution de la déviance. Les variables sont rajoutées de la première à la dernière selon leur introduction dans le modèle. Cependant, nous avons retracé ce tableau de façon à classer les variables explicatives de la plus influente à la moindre. Ça apparaît que la variable explicative Deph.706Anc1H est la plus influente. Ce classement n'est pas le même pour les autres modèles de prévisions.

Tableau 6.3 Le test d'ANOVA du modèle G06.1H avec la méthode GLM

Les variables explicatives	Df	Deviance	Resid. Df	Resid.DV
NULL			1 011 286	45 528 022
Deph.706Anc1H	1	36 867 528	1 011 285	8 660 495
Sample	7	2 700 510	1 011 284	5 959 984
BG	1	507 200	1 011 273	5 452 785
BQ.PR24H	1	167 468	1 011 267	5 285 317
Mois	11	151 183	1 011 266	5 134 134
Jours	6	34 558	1 011 259	5 099 576
BQ.PR	1	1 128	1 011 258	5 098 448

Conditionnellement à la connaissance des valeurs des variables prédictives, les paramètres inconnus du modèle (θ_i et σ^2) sont estimés par maximisation de la vraisemblance en supposant l'hypothèse de la normalité de la variable d'erreur (Julian, 2002). Les 111 modèles réalisés (37 angles * 3 modèles de prévision anticipée) sont statistiquement significatifs avec des P-values largement inférieures à 10^{-3} .

En fonction du critère d'évaluation R, la Figure 6.3 montre les résultats de prévisions anticipées de l'angle D06 ainsi que les deux autres angles médoids D30 et D50. L'ensemble des résultats des prévision anticipées, des trente-sept angles de phase de tension des postes à 735kVdu réseau de transport avec les trois modèles (24H, 4H et 1H en avance), sur le même ensemble test, sont reportés dans le tableau-A III-2 de l'annexe III. La Figure 6.4 retrace l'allure de l'efficacité de ces modèles sur l'ensemble des angles.

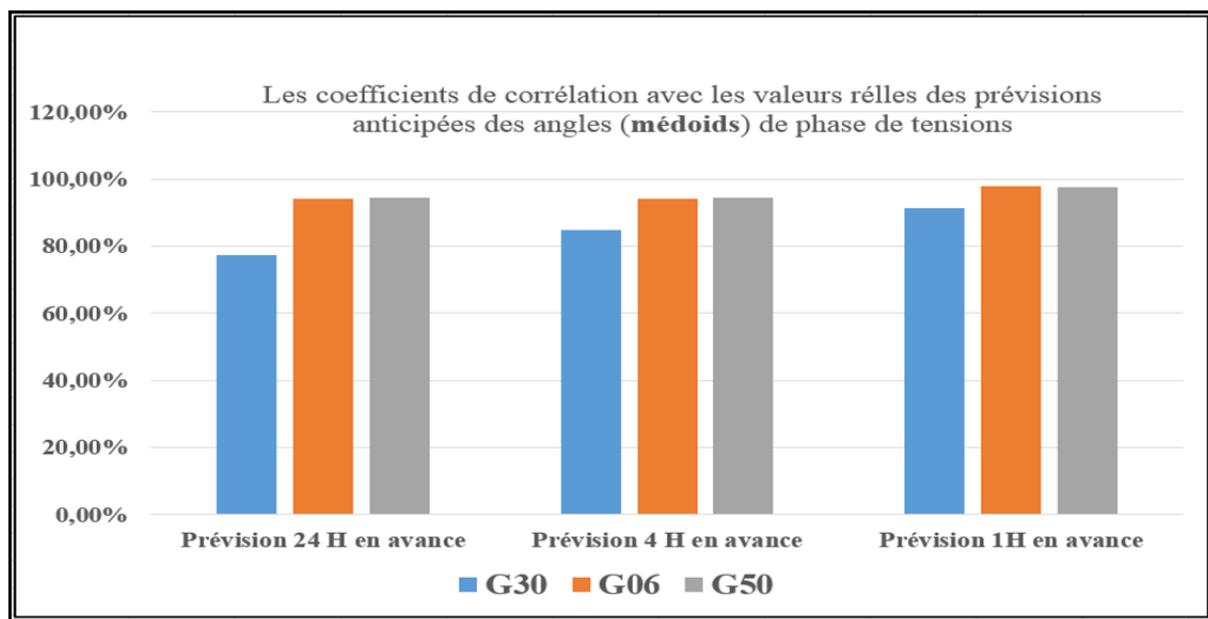


Figure 6.3 Les prévisions anticipées des angles médoids

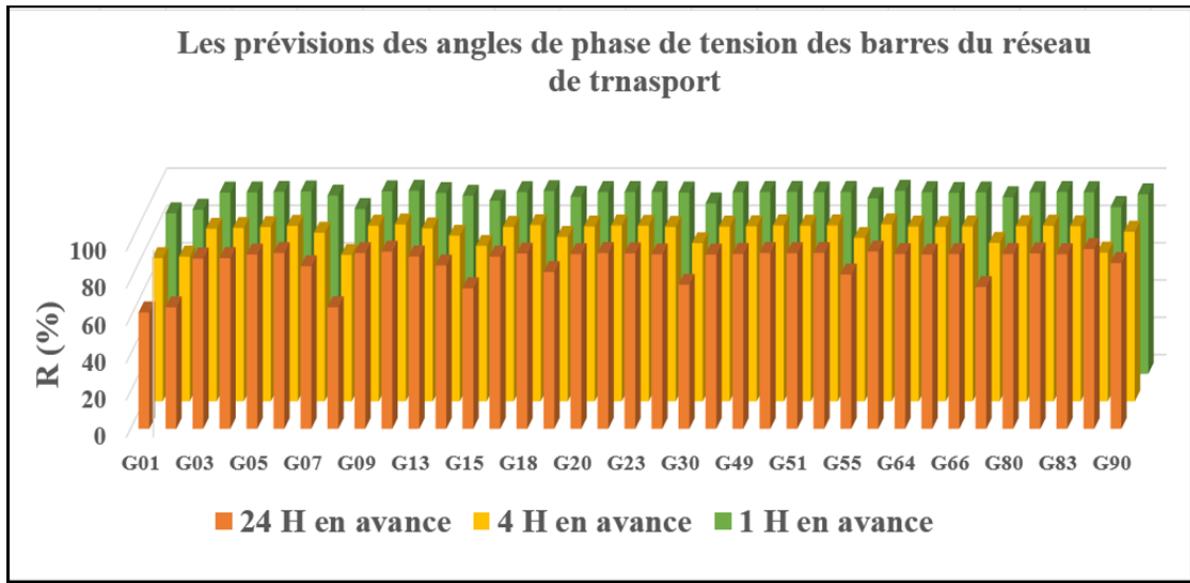


Figure 6.4 Prévisions anticipées des angles de tension

La prévision anticipée une heure en avance rapporte des coefficients de corrélation plus précis que les deux autres prévisions. Ceci était prévisible car l'horizon fini d'une heure est le plus proche dans le temps au présent. Le facteur le plus élevé est au niveau de 98% et la majorité des coefficients sont supérieurs à 95%.

En effet, tous les angles de tension du deuxième et du troisième groupe dont les médiods sont respectivement les postes LG3 et Micoua (D06) sont prévus, pour le cas d'une heure en avance, avec une précision d'au moins 97%, à l'exception de l'angle D14 de la Vérendrye ayant un coefficient de 95%. Les précisions de prévisions anticipées de vingt-quatre heures ou de quatre heures en avance ne semblent pas différentes pour ces deux ensembles. Pour tous les angles, un écart constant de 4% apparaît entre la précision d'une prévision anticipée d'une heure et la précision d'une prévision anticipée de quatre heures ou vingt-quatre heures.

Pour les angles du premier groupe dont, le médoid est le poste Carignan (D30), la précision de la prévision anticipée d'une heure en avance s'améliore visiblement par rapport à celles de

prévisions anticipées de vingt-quatre heures et quatre heures. Néanmoins selon les angles, elle varie entre 97,30% et 86,08%.

Une analyse plus approfondie de l'efficacité des modèles de prévisions anticipées d'une heure en avance en fonction des différents critères d'évaluation (6.6-6.9) a été effectuée. Nous avons reporté les résultats de cette analyse dans le tableau-A III-3 de l'annexe. La Figure 6.5 retrace la performance de ces modèles de prévisions de tous les angles de phase de tension du réseau de transport.

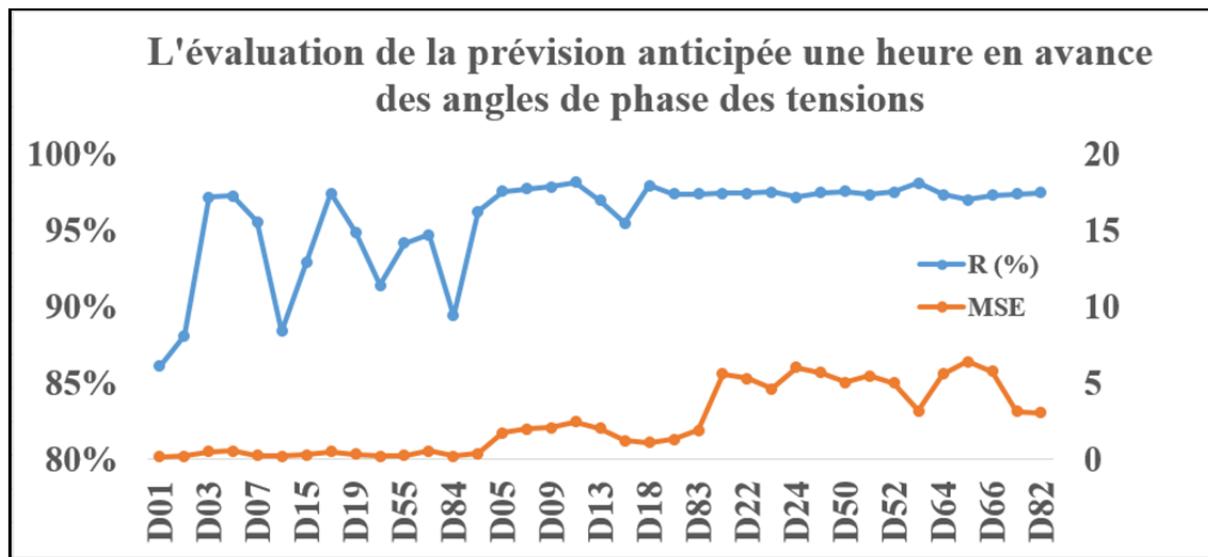


Figure 6.5 Les performances des modèles de prévisions une heure en avance

Avec les nuages de points, en fonction des valeurs réelles de l'angle de phase de la tension D06, nous représentons les points prédits avec les trois modèles de prévisions. La Figure 6.6 montre les résultats de prévisions anticipées vingt-quatre heures en avance, 4 heures en avance et une heure en avance versus les valeurs réelles. La Figure 6.7 rapporte la vue de l'ensemble de points prédits par le modèle de prévision anticipée une heure en avance face aux valeurs réelles de l'angle de phase de la tension D06, en formes de nuage de points et de séries chronologiques.

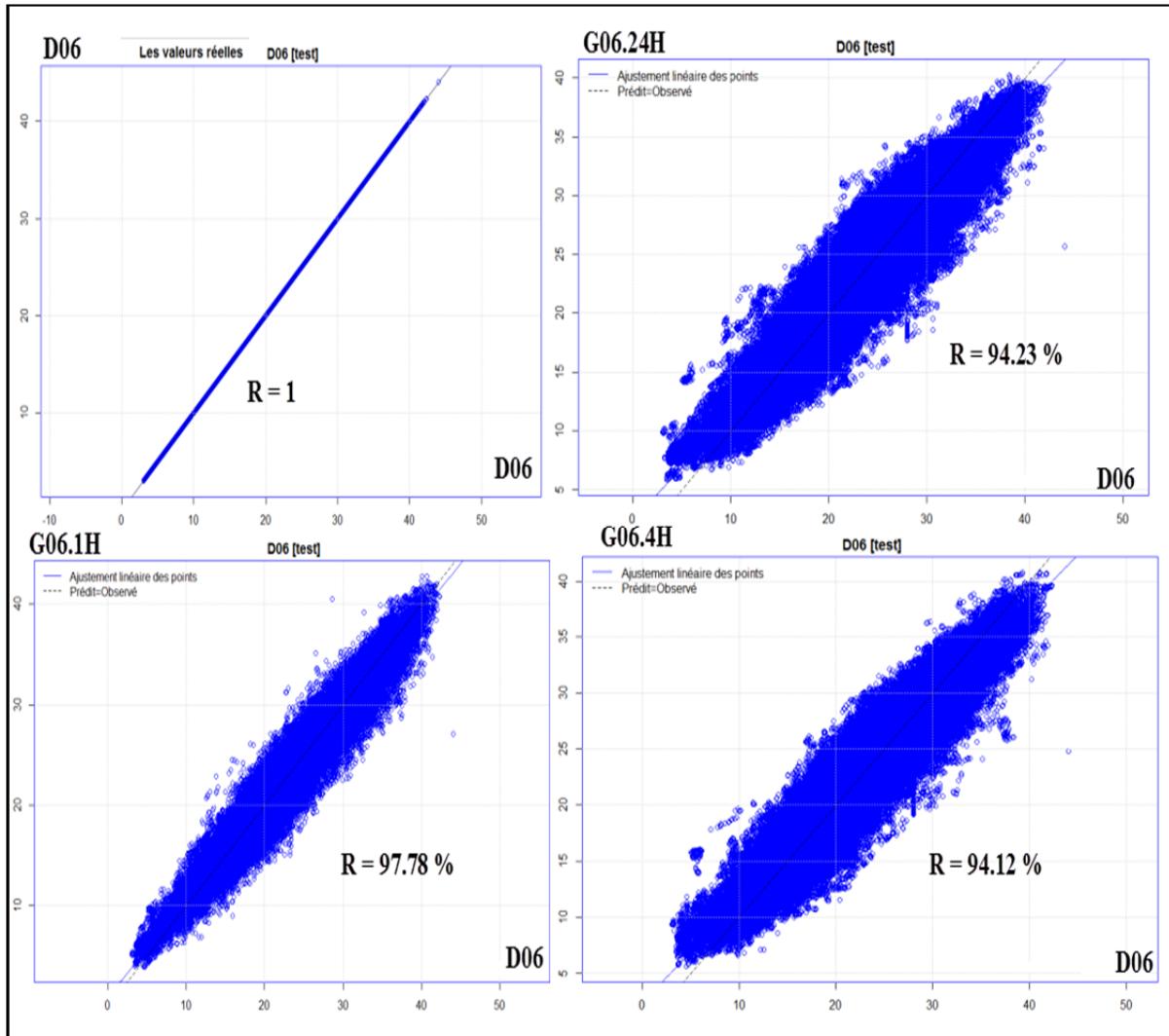


Figure 6.6 Les corrélations entre les ensembles prévus et l'ensemble de mesures réelles de l'angle de phase de tension du poste Micoua sur le même ensemble test

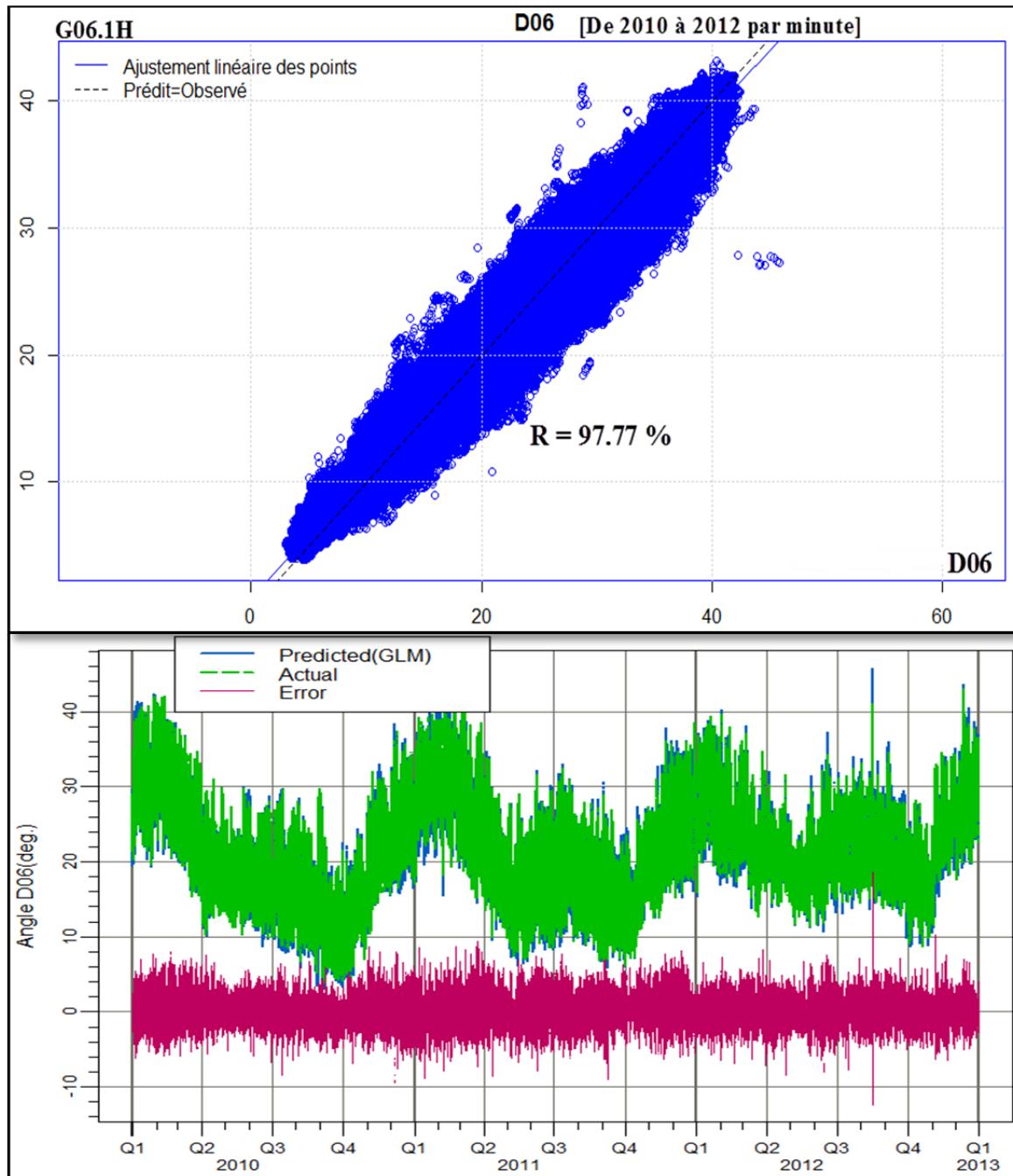


Figure 6.7 Évaluation de la performance du modèle GLM avec le critère R (haut) et les résidus (bas)

La Figure 6.8 expose graphiquement, selon les colonnes, les distributions des angles de phase de tension des barres médoids :

Première colonne : les distributions des valeurs réelles des angles de l'ensemble test;

Autres colonnes : les distributions des valeurs prédictes des angles, une heure, quatre heures et vingt-quatre en avance, respectivement.

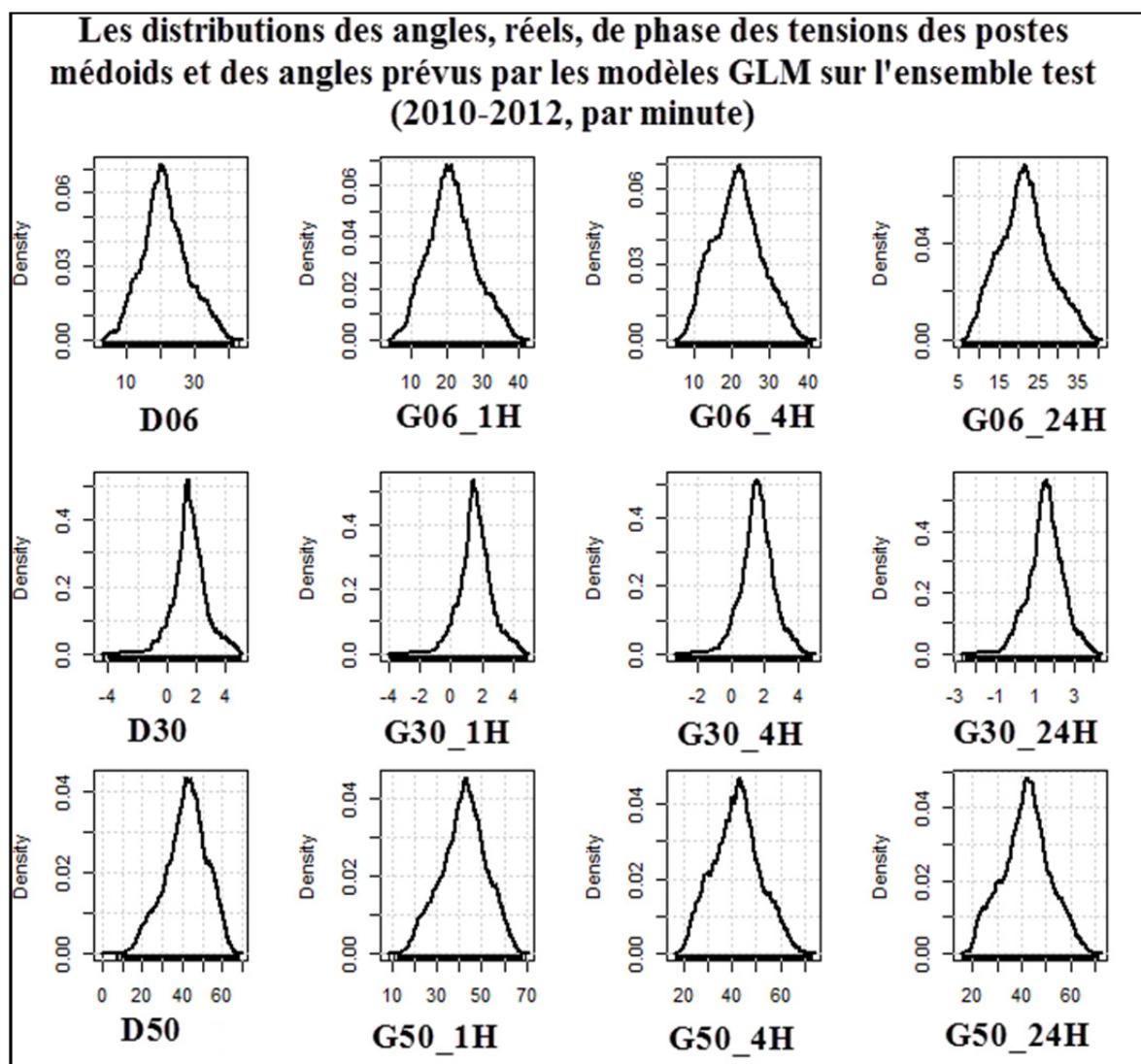


Figure 6.8 Les prévisions anticipées des angles médoids

La Figure 6.9 met en exergue les distributions des erreurs brutes de prévisions des mêmes angles de phase de tension.

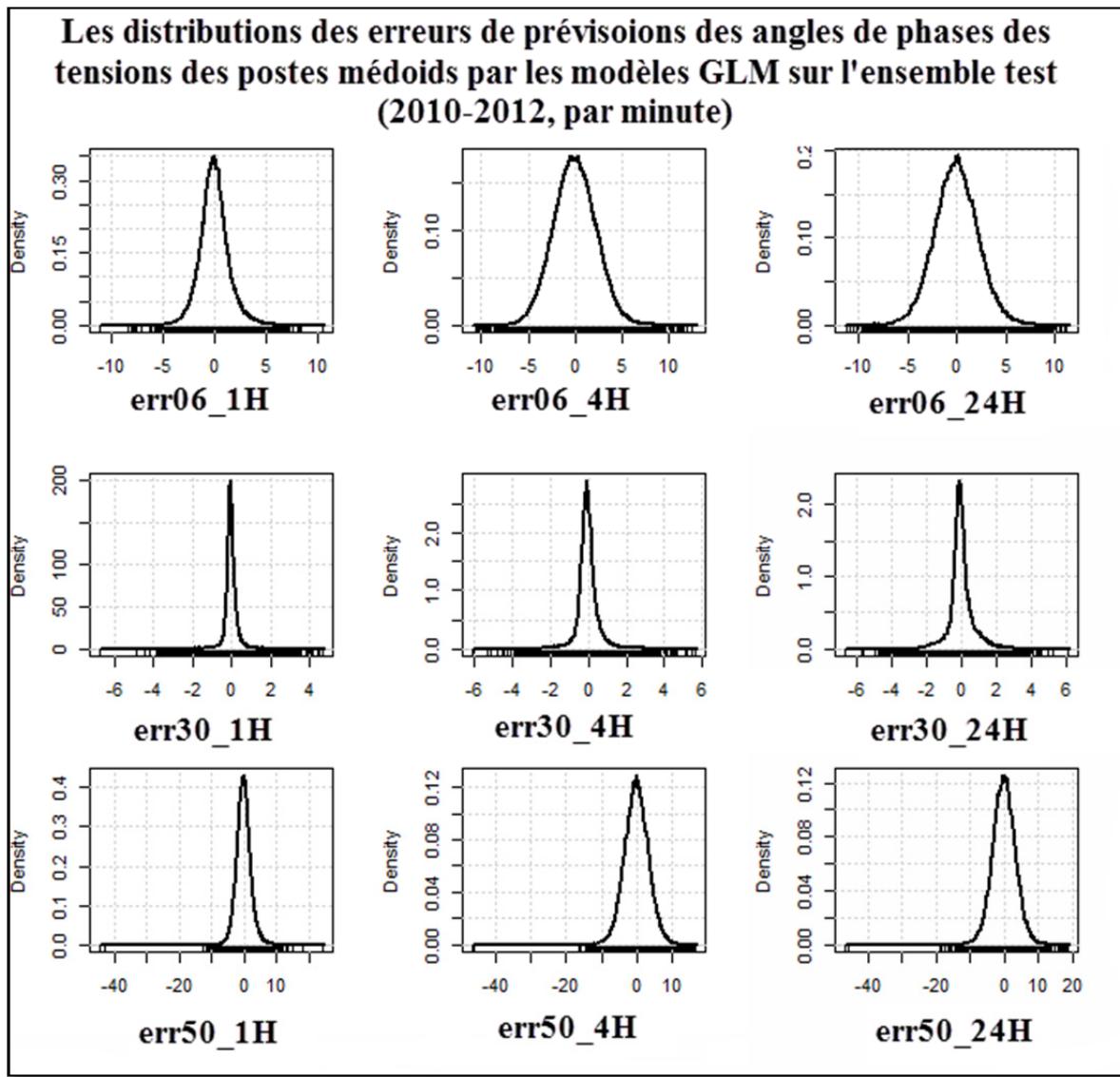


Figure 6.9 Les erreurs brutes de prévisions anticipées des angles médoides

En guise de conclusion, concernant cette première étape, nous soulignons que malgré que dans la pratique les valeurs des coefficients de corrélation R et d'explication de variance Rsq ne sont pas viables, il convient de noter que d'un point de vue théorique, elles sont de qualité largement supérieure. De plus, les distributions des erreurs brutes montrent que la structure

du modèle n'est pas dénuée de sens. En effet, les graphiques associés à ces erreurs présentés dans la Figure 6.9, montrent que ces dernières suivent une loi gaussienne centrée avec une symétrie assez prononcée. De plus, le domaine $[-3\sigma, 3\sigma]$ [constitue un intervalle de confiance incluant plus de 95% de données. En outre, les paramètres estimés de tous les modèles sont très significatifs, selon les résultats des tests de validité présentés dans la colonne « code » Tableau 6.2.

6.7 Le couplage entre les approches à base de GLM et les modèles à base de FA

Pour introduire ces informations obtenues à partir des différents modèles de GLM dans la conduite du réseau, nous suggérons d'utiliser le mappage de ces angles sur les modèles de prévisions des transits et des marges déjà établis et ceci afin de prévoir des transits et des marges de sécurité une/des heure(s) en avance.

À partir des modèles développés dans les sections précédentes, nous établissons trois bases de données constituées des séries chronologiques des angles prévus par la méthode de GLM. Une base de données composée des séries chronologiques par minute des angles de phase de tension des postes à 735kV du réseau de transport prévus une heure en avance par le modèle de forme (6.5). Une autre base de données formées de 37 séries chronologiques par minute des angles prévus quatre heures en avance. Enfin, une dernière constituées de 37 séries chronologiques par minute des angles prévus vingt-quatre heures en avance par le modèle à base de GLM, de forme (6.3). Pour chaque interface, du réseau de transport, nous appliquons aux modèles préétablis par les forêts aléatoires, une de ces bases de données comme entrée (input) afin de générer d'autres valeurs de transits actifs et de marges de sécurité. Ce principe de procéder est illustré par la Figure 6.10 et la Figure 6.11.

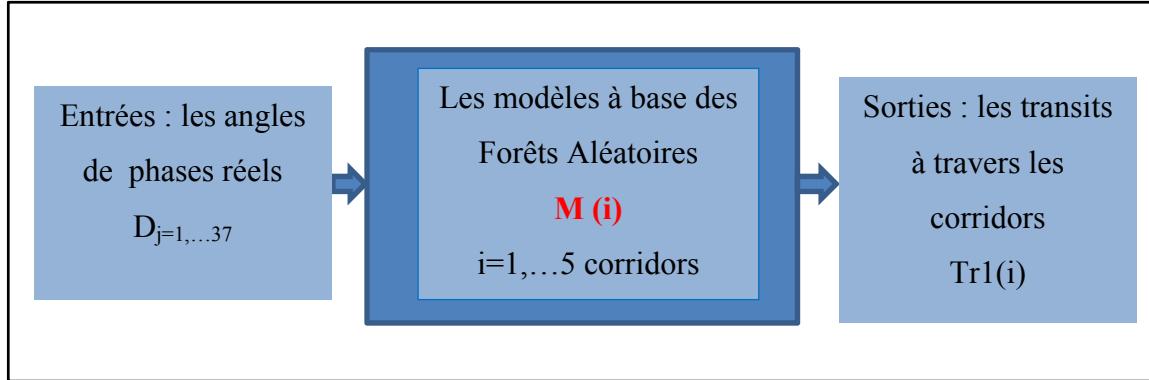


Figure 6.10 Construction des modèles de prévisions des transits actifs et des marges de sécurité à base des angles de phase réels par la méthode de FA

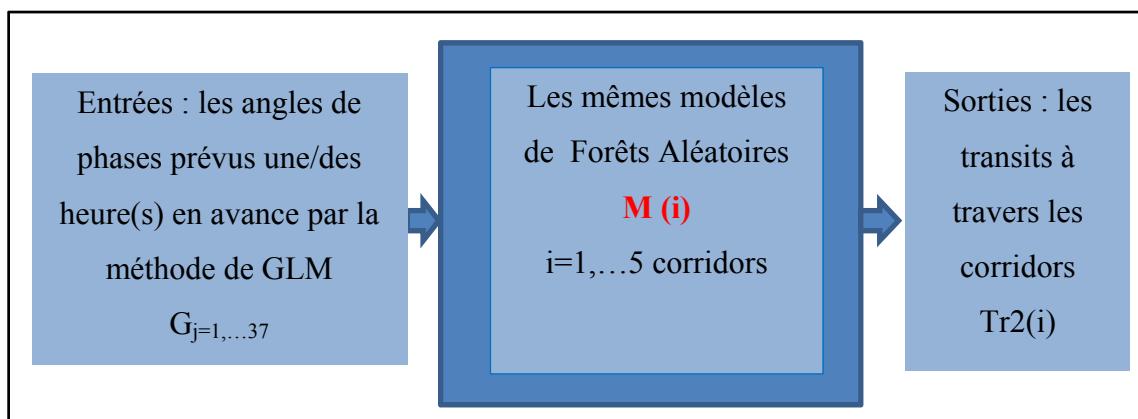


Figure 6.11 Le mappage de l'ensemble des angles prévus par GLM sur les modèles de prévisions déjà établis des transits actifs et des marges de sécurité

Les modèles **M(i)**, dans la Figure 6.10, sont obtenus par la méthode des forêts aléatoires et ils sont construits à partir d'une base de données réelle contenant les angles de phase de tensions des barres à 735kV comme entrées et les transits actifs via les différentes interfaces clef du réseau de transport.

Dans la Figure 6.11, on garde ces mêmes modèles **M(i)** élaborés et on introduit de nouvelles entrées (G_j). Ces dernières sont les angles prévus une/des heure(s) en avance, estimés par les modèles GLM. À la phase finale de cette procédure, nous obtenons ($Tr2$) les valeurs de transfert actifs une /des heure(s) en avance.

6.8 Évaluation de prévisions anticipées des transferts actifs et des marges de sécurité

Nous réalisons le mappage des différentes bases de données sur les modèles des transits et des marges de sécurité. Le Tableau 6.4 englobe les mesures d'évaluation de prévisions des modèles **M(i)**, dégagées du principe décrit dans la Figure 6.10 et les résultats de prévisions par mappage de l'ensemble des angles G sur les modèles à base des forêts aléatoires **M(i)**, décrits dans la Figure 6.11.

Tableau 6.4 Les mesures d'évaluation des modèles de prévisions des transits de puissance et des marges de sécurité

	Modèles	MAE.TR (MW)	R.TR (%)	Rsq.TR (%)	MAE. Mr (%)	R. Mr (%)	Rsq.Mr (%)
1	SUD	48,6529	99,96	99,91	0,201571	99,96	99,91
2	SUD.1H	365,082	98,8	97,56	1,49389	98,91	97,79
3	SUD.4H	557,987	97,42	94,64	2,30983	97,53	94,91
4	SUD.24H	546,885	97,48	94,66	2,23394	97,71	95,25
5	BJN	40.1668	99,95	99,89	0,17026	99,88	99,76
6	BJN.1H	359,124	97,89	95,61	1,49169	95,99	92,03
7	BJN.4H	609,568	94,33	87,73	2,5545	89,55	79,73
8	BJN.24H	607,118	94,15	87,24	2,53696	89,68	80,11
9	CHU	19,1846	99,91	99,81	0,079293	99,77	99,53
10	CHU.1H	107,528	98,37	95,87	0,414593	95,97	92,1
11	CHU.4H	223,216	94,14	85,36	0,915208	86,47	74,38
12	CHU.24H	212,002	94,5	86,25	0,859921	87,56	76,33
13	MQ	38,6605	99,93	99,86	0,156214	99,92	99,85
14	MQ.1H	271,6	98,24	96,35	1,08245	98,25	96,47
15	MQ.4H	489,285	95,11	89,39	1,9745	94,93	89,66
16	MQ.24H	457,181	95,63	90,4	1,84264	95,55	90,77

Les lignes 1, 5, 9 et 13, du Tableau 6.4, représentent les mesures d'évaluation de prévisions des transits et des marges de sécurité à base des modèles des FA. Les MAE.TR et MAE.Mr sont les erreurs moyennes absolues des transits actifs en Mw et des marges de sécurité en % définies dans le chapitre 4. Les R et Rsq sont les coefficients de linéarité et les coefficients carrés en %. Plus ils sont proches de 100%, plus ils sont meilleurs. Les autres lignes du Tableau 6.4 sont les mesures d'évaluation de prévisions des transits et des marges de sécurité, par mappage des angles de phases de tension prévus une/des heure(s) en avance par les différents modèles de GLM.

La Figure 6.12 reprend la mesure de la relation de linéarité entre les valeurs réelles des transits et les valeurs des transits prévus par les forêts aléatoires à travers quatre interfaces du réseau de transport. La qualité de précision de ces prévisions est justifiée par les facteurs de linéarité qui dépassent 99,9%.

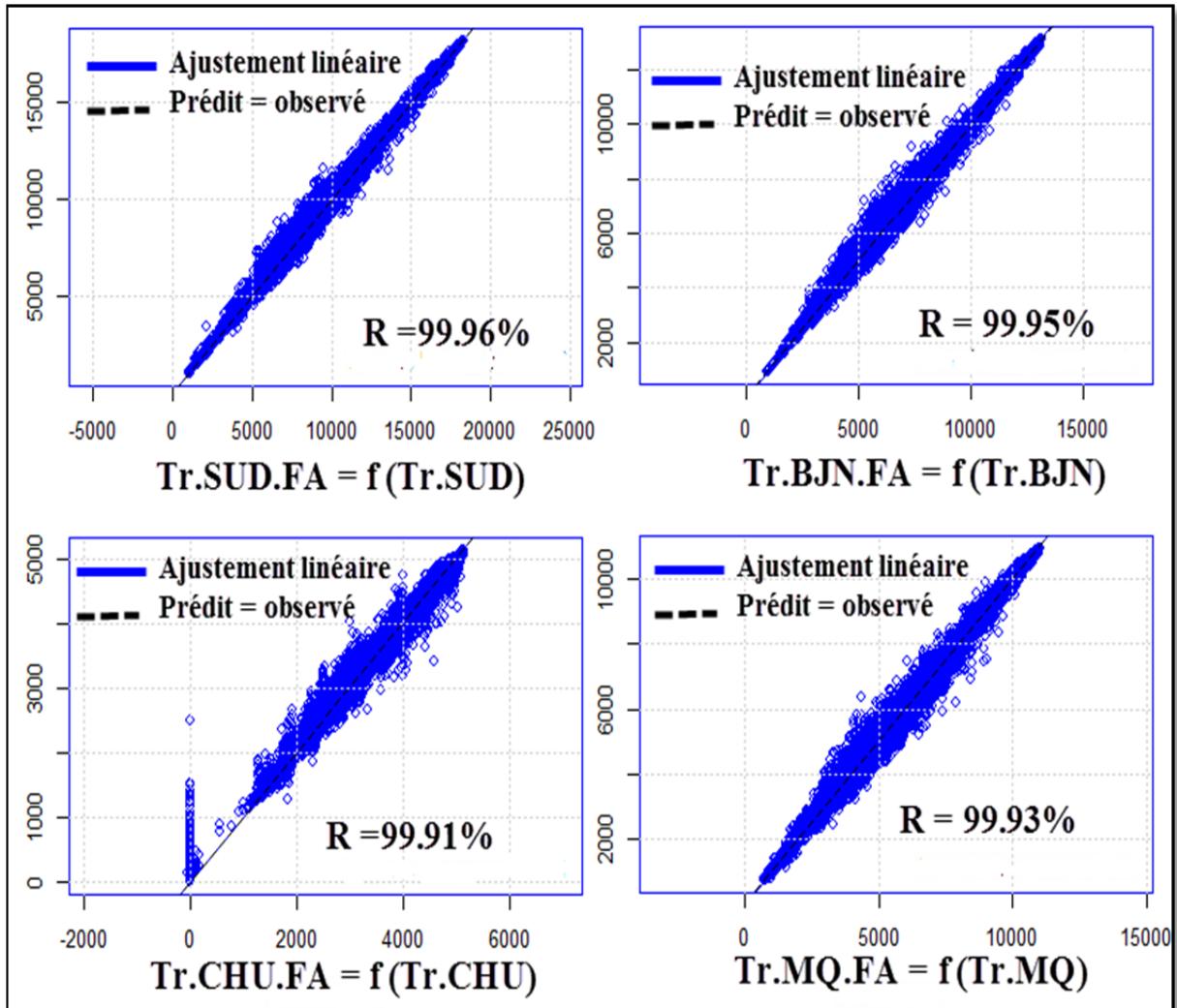


Figure 6.12 Les modèles des FA des transits actifs à travers les différentes interfaces à base des angles de phase réels

Les résultats illustrés par les cinq figures, la Figure 6.13 à Figure 6.17, représentent les produits de prévisions une/des heure(s) en avance, par les Forêts aléatoires, des transits actifs le long des corridors principaux et la marge de sécurité de l'interface SUD.

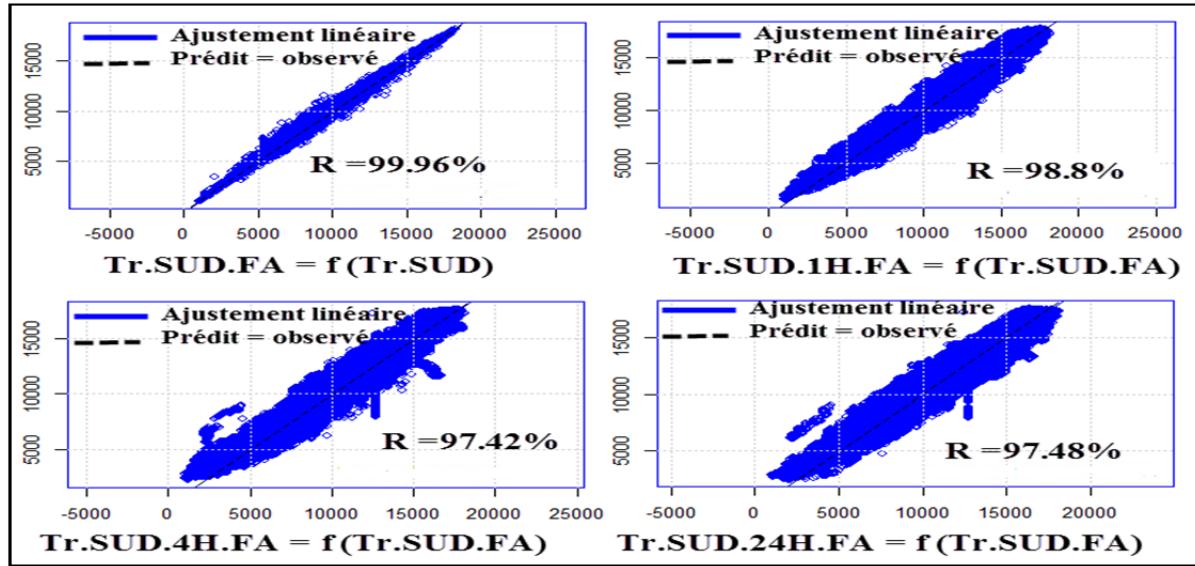


Figure 6.13 Les prévisions anticipées des transits en MW à travers le corridor SUD

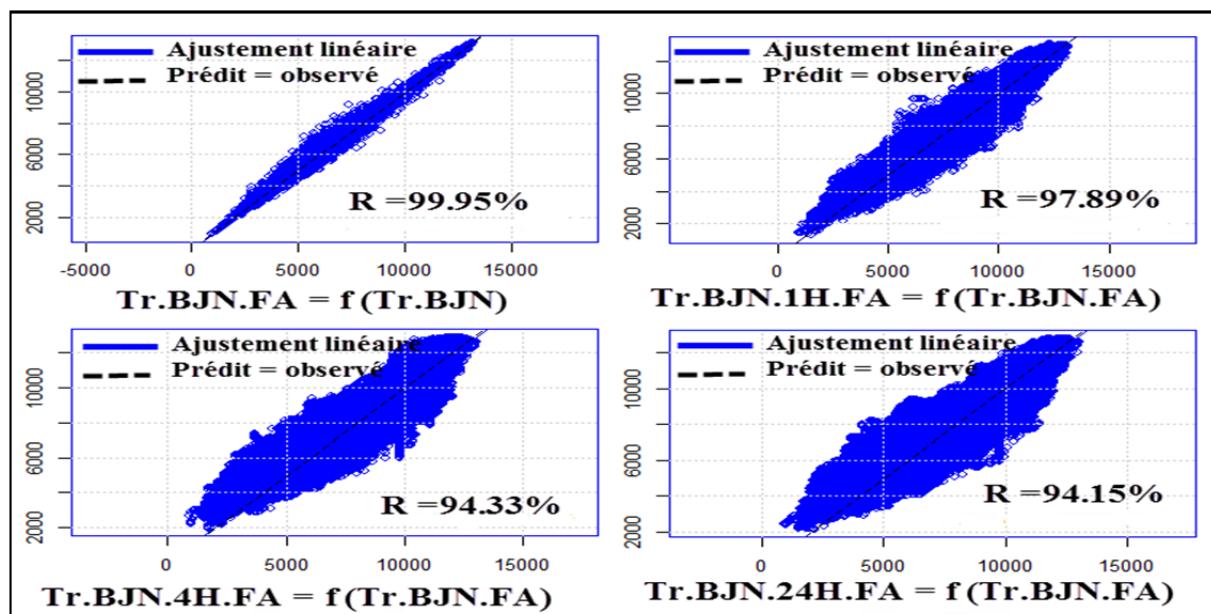


Figure 6.14 Les prévisions anticipées des transits en MW à travers le corridor BZN

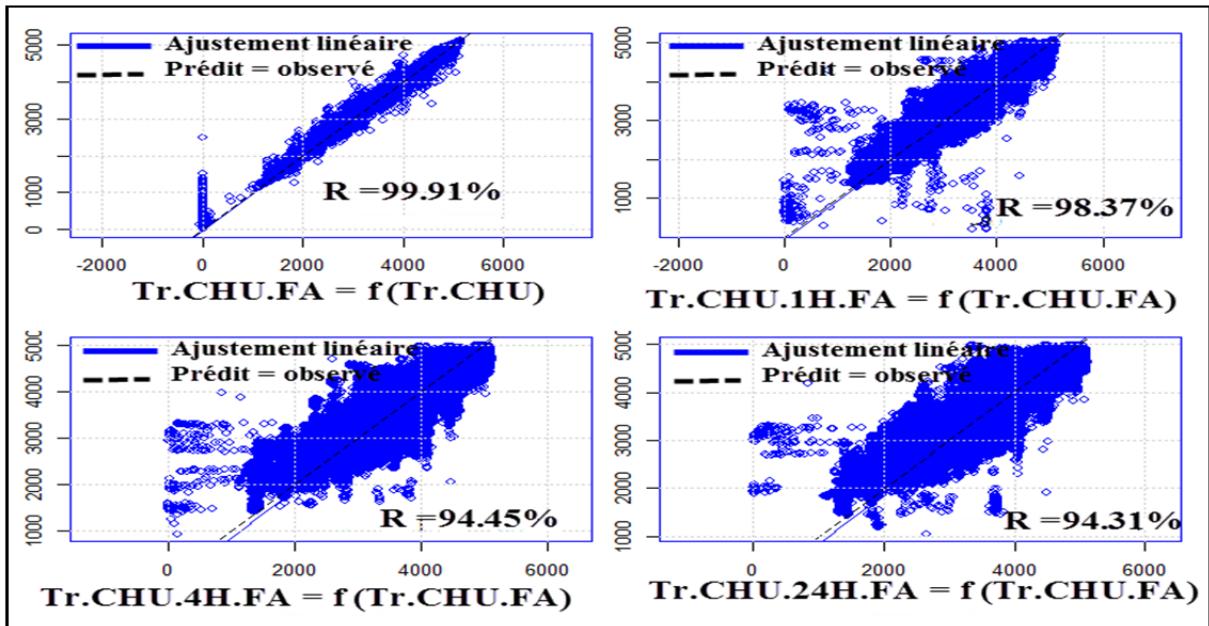


Figure 6.15 Les prévisions anticipées des transits en MW à travers le corridor CHU

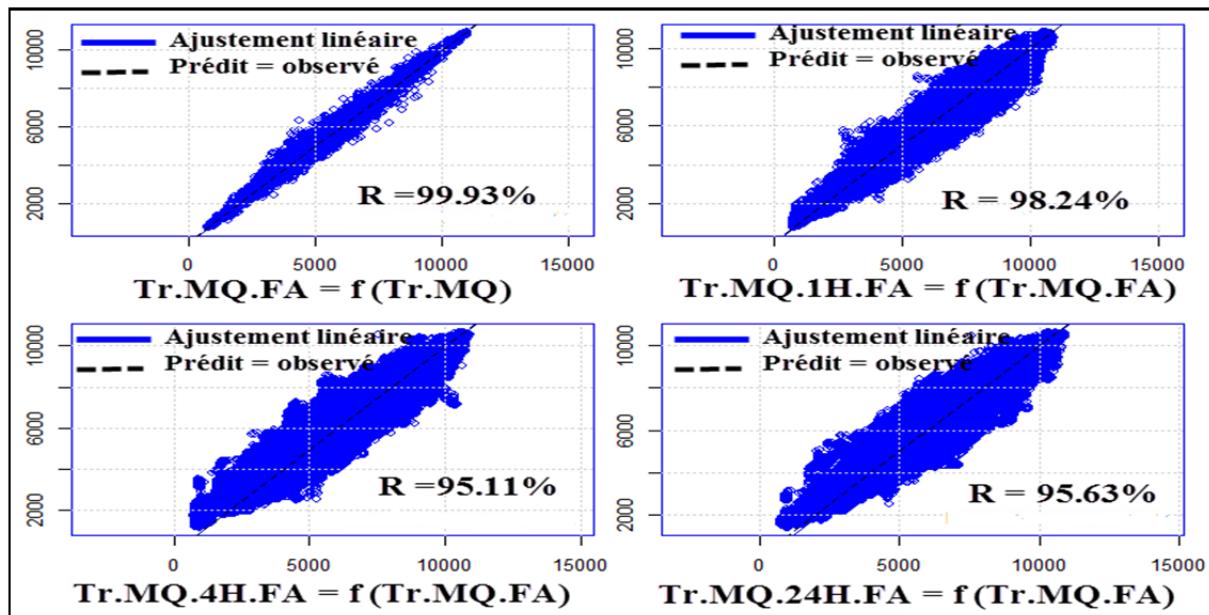


Figure 6.16 Les prévisions anticipées des transits en MW à travers le corridor MQ

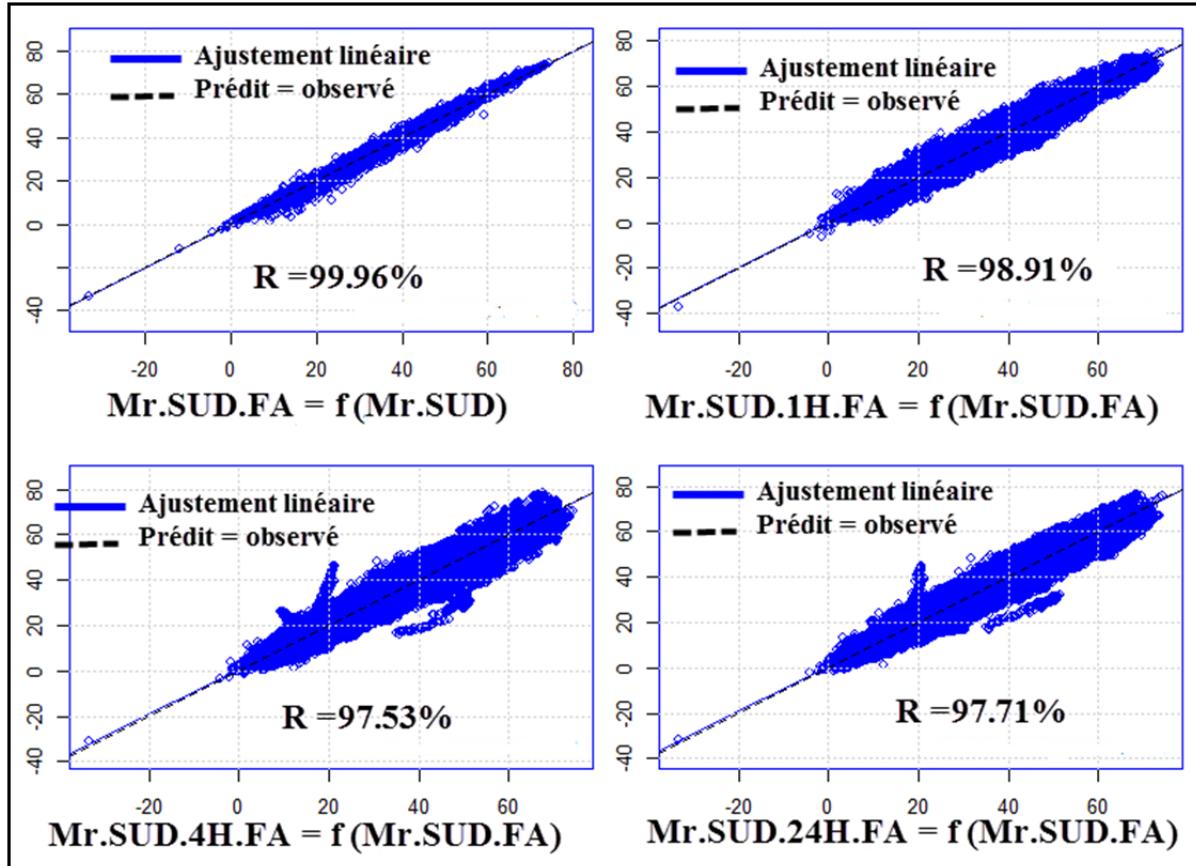


Figure 6.17 Les prévisions anticipées des marges de sécurité de l'interface SUD

Les prévisions des transits et des marges basées sur le mappage des angles de phases prévus une heure en avance semblent de bonne précision. Effectivement la mesure de linéarité R, entre les valeurs prévues et les valeurs réelles des transits actifs des quatre interfaces, varie de 97,89% à 98,8%. Concernant les marges de sécurité des mêmes corridors, cette mesure est dans la fourchette de 95,97% à 98,91%. Ceci nous laisse prétendre que les prévisions anticipées d'une heure des marges de sécurité et des transits actifs le long des corridors, à base de cette démarche de couplage de deux algorithmes FA et GLM, peut être utile dans la pratique.

Une autre évaluation de la qualité de prévisions des transits et des marges par le mappage des angles de phase prévus une/des heure(s) en avance, sur le modèle d'origine, consiste à

comparer ces prévisions par rapport aux valeurs réelles. Le Tableau 6.5 récapitule les résultats de cette comparaison. La Figure 6.18 illustre l'ajustement linéaire de la marge une heure en avance à base de GLM versus la marge prévue par les angles réels de l'interface SUD.

Tableau 6.5 Les prévisions par le mappage et les valeurs réelles

Interface	Le transit actif Tr(D) vs. Tr(GLM.D)		La marge de sécurité Mr.(D) vs. Mr. (GLM.D)	
	R (%)	MAE (MW)	R (%)	MAE (%)
BJN	97,75	380,38	96	1,64
CHU	97,89	123,36	95,6	1,7
SUD	98,6	397,07	98,71	0,54

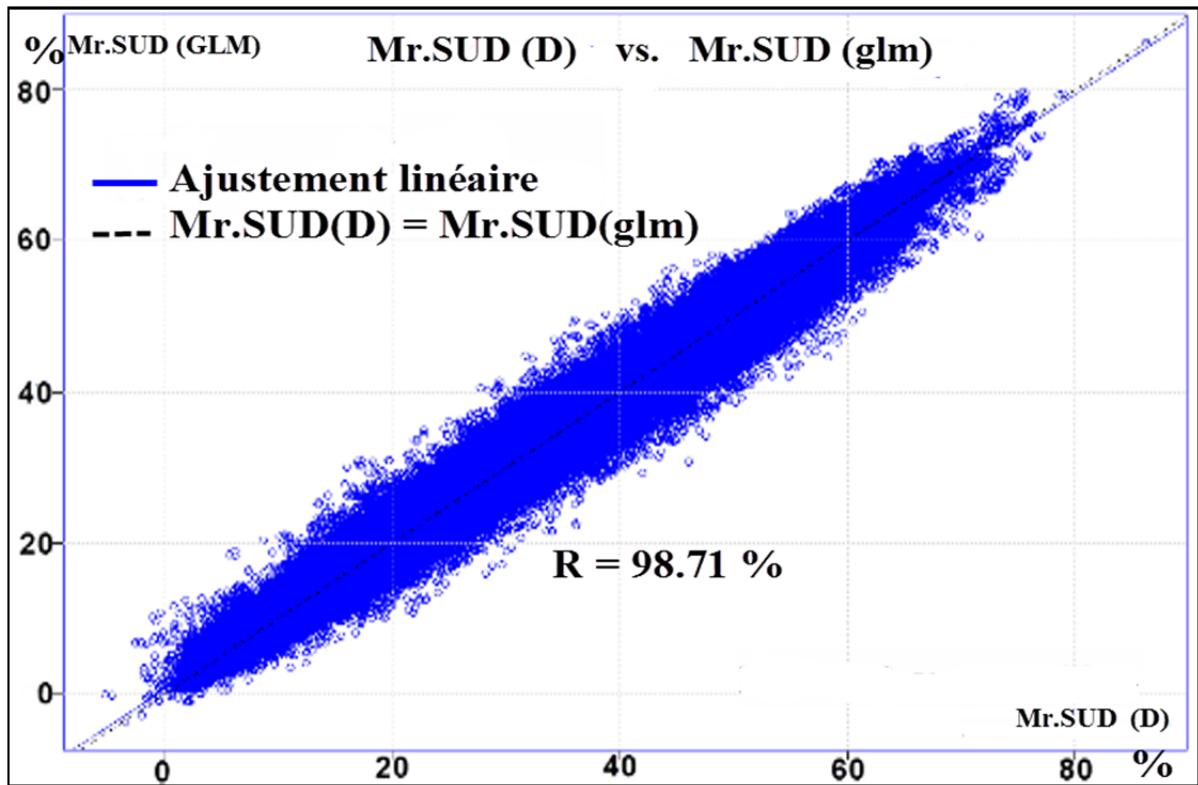


Figure 6.18 La marge prévue et la marge réelle du corridor SUD

6.9 Conclusion

Au début, nous avons élaboré trois modèles de prévision anticipée pour chaque angle de phase de tension à 735 kV du réseau de transport, en utilisant l'algorithme GLM. Les modèles élaborés, basés sur les données de l'ensemble de formation, ont servi à prévoir les angles de phase de tension appartenant au même ensemble test utilisé dans le chapitre précédent. Nous avons estimé la qualité de prévisions de tous ces modèles par le biais de différents critères d'évaluation. Les résultats obtenus nous laissent supposer que les modèles sont assez consistants. Toutefois, la précision des modèles à base de GLM peut s'améliorer en analysant au préalable les données d'étude pour distinguer les valeurs extrêmes des valeurs standards.

Nous avons réalisé le mappage des angles prévus via les modèles à base de GLM sur les modèles des forêts aléatoires. Ces derniers sont les modèles de prévisions des marges de sécurité et des transits actifs le long des corridors clefs du réseau de transport, développés dans le chapitre 5.

En conséquence, nous avons générée de nouvelles valeurs des transits et des marges qui correspondent aux valeurs « futur » ou prévues des angles de phase. Ce procédé de prévision anticipée des transits et des marges se voit aussi robuste que précis pour une utilisation pratique dans l'évaluation de la sécurité dynamique du réseau de transport.

Ainsi dans ce chapitre, nous avons effectué le couplage des modèles à base des forêts aléatoires avec les modèles à base de GLM et ceci, dans le cadre d'apprentissage statistiques utilisant les données de l'EMS-SCADA de l'année 2010 jusqu'à l'année 2012.

CONCLUSION

En guise de conclusion, nous synthétisons notre élaboration et nos contributions en points suivants :

1. Le sujet principal de cette thèse concerne la manière d'utiliser le grand volume des données du réseau électrique, obtenues à partir des mesures des synchrophaseurs, pour surveiller le fonctionnement du réseau de transport et assurer sa sécurité. Ces mesures qualifiées de qualité supérieure telle que la précision, la latence et la visibilité concernent entre autre les courants, les tensions, les fréquences et les phaseurs. Ainsi, la problématique fondamentale est comment présenter ce grand volume de données au gestionnaire de réseau afin d'évaluer la sécurité dynamique du réseau de transport ?
2. Pour amener une réponse à cette problématique, nous nous sommes basés sur deux points essentiels. Le premier point concerne la base de données utilisée. À défaut d'une base de données à base des mesures de synchrophaseurs, nous avons utilisé une base de données extraite de l'EMS/SCADA. Le deuxième point concerne l'utilisation exclusive des angles de phase de tension pour surveiller le fonctionnement et assurer la sécurité du réseau de transport ;
3. En se basant sur des rapports d'études des grands blackouts connus et sur les constats des experts des réseaux de transport, nous avons proposé les angles de phase de tension, comme des prédicteurs de la stabilité des réseaux de transport ;
4. Par une approche statistique, en utilisant des boxplots, nomogrammes et autres, nous avons réuni et présenté le grand volume des données. Puis, par une étude de calibration statistique (baselining), nous avons proposé des angles de phase limites, où le réseau semble stable;

5. Pour la gestion des réseaux de transport, nous avons élaboré des modèles de prévisions des marges de sécurité et des transits actifs sur les couloirs principaux du réseau de transport, en fonction des angles de phase de tension des barres. Ces modèles de prévisions sont établis par la méthode de forêts aléatoires dans le contexte d'apprentissage statistique ;
6. Sur le principe que les angles de phase de tension sont des prédicteurs de la stabilité, nous avons construit des modèles pour anticiper leur prévision. Ces modèles de prévision des angles de phase sont établis par l'algorithme de GLM;
7. En s'inspirant de la procédure standard suivie dans l'exploitation des réseaux à base de la prévision des charges (Hong et Shahidehpour, 2015; Slaven, Brian et Hongming, 2015), nous avons proposé une démarche basée sur la combinaison des résultats obtenus en utilisant les FA et ceux obtenus par le GLM. Ce couplage consiste à mapper les angles prévus sur les modèles préétablis des marges de sécurité et des transits actifs par les FA;
8. Les auteurs des articles (Kaci et al., 2014a; 2014b) suggèrent la possibilité d'intégrer ces procédures dans la conduite des réseaux.

Les résultats obtenus dans le cadre de cette thèse mettent en évidence deux aspects essentiels concernant une éventuelle politique future du développement des réseaux de transport d'Hydro-Québec.

En premier lieu, les modèles établis dans ce travail à base des angles de tension, permettent de mettre en place une procédure de protection, assez efficace, des limites de stabilité des réseaux. Celle-ci facilite, par conséquent, la gestion de ce type de situations. À cet effet, et à titre illustratif, nous avons mis en évidence, que chaque dépassement d'angle de la limite établie correspond au moins à un corridor dans un état d'insécurité, fonctionnant à proximité ou au-dessous des limites normales.

Le deuxième aspect concerne la pertinence de la généralisation de l'utilisation des synchrophaseurs dans les réseaux de transport d'Hydro-Québec. En effet, il apparaît d'un point de vue statistique, que l'utilisation des données mieux étoffées et plus précises, permettrait d'obtenir de meilleures estimations concernant les coefficients de nos modèles (coefficients estimés entachés d'erreurs d'intensités moindres). Ceci permettrait d'améliorer le degré d'efficacité de nos modèles de prévision. Ce qui met en évidence, d'un point de vue d'une gestion saine et efficace, le caractère primordial de l'installation et de l'utilisation des synchrophaseurs dans les réseaux de transport d'Hydro-Québec.

RECOMMANDATIONS

1. Nous avons entrepris toutes ces analyses en se basant sur les données EMS/SCADA du réseau de transport d'Hydro-Québec. Cependant, réaliser la même étude avec une base de données plus enrichie par des mesures précises des synchrophaseurs et avec un calculateur plus puissant, améliorera la qualité des informations émises. En effet, on sera capable de capter les événements rapides qu'on n'a pas pu voir durant cette étude, vu qu'ils se passent en quelques millisecondes, et du coup, on peut renforcer le postulat avancé : les écarts angulaires sont des prédicteurs de la stabilité dynamique ;
2. L'analyse élaborée dans le cadre de cette thèse concerne typiquement le réseau de transport d'Hydro-Québec, qui est de type radial. Par conséquent, cette approche se voit également applicable à d'autres réseaux de même type tels que les réseaux de Chine, Brésil, WECC, etc. Par ailleurs, la justification théorique donne à croire que l'approche est générale et peut s'appliquer à tout réseau maillé ou non ;
3. L'utilisation des forêts aléatoires dans notre étude s'est montrée très performante. Ainsi pour améliorer l'approche proposée, particulièrement dans les réseaux extrêmement maillés, nous suggérons d'introduire, en plus des angles de phase comme entrées dans le modèle des FA des transferts actifs, d'autres fonctionnalités telles que : les réserves réactives, les injections de puissance active et réactive, les barres critiques, les grandeurs des barres, etc. ;
4. Nous suggérons une étude similaire à base des écarts angulaires par rapport aux angles différents du réseau de transport pour mener à l'identification d'autres signatures de défauts;
5. Pour améliorer la prévision des angles de phase de tension, l'utilisation des forêts aléatoires avec la logique floue nous semble faisable et prometteuse.

ANNEXE I

PROGRAMMER AVEC R

Cette annexe présente les différentes lignes de code de la programmation avec le logiciel **R** permettant de construire les modèles GLM des angles de phase de tension des barres à 735 kV, présentés dans le chapitre 6.

Programme principal :

```
# Charger un jeu de données R
crs$jeudedonnees <- D06

# Afficher la structure du jeu de données.
Str(crs$jeudedonnees)

# Construire les jeux de données de formation/validation/test.

Crv$seed <- 42
set.seed(crv$seed) # Une valeur prédéfinie est utilisée pour réinitialiser la racine aléatoire de
façon à ce que les résultats puissent se répéter.

Crs$nobs <- nrow(crs$jeudedonnees)
crs$train <- sample(nrow(crs$jeudedonnees), 0.7*crs$nobs)
crs$sample <- crs$train
crs$validate <- sample(setdiff(seq_len(nrow(crs$jeudedonnees)), crs$train),
0.15*crs$nobs)
crs$test <- setdiff(setdiff(seq_len(nrow(crs$jeudedonnees)), crs$train), crs$validate)

# Définir les variables d'entrées et la variable de sortie

crs$input <- c("Mois", "Jours", "Sample", "Deph.706Anc24H",
" BQ.PR", " BQ.PR24H", "BG")
crs$numeric <- c("Deph.706Anc24H", "BQ.PR", "BQ.PR24H", "BG")
crs$categoric <- c("Mois", "Jours", "Sample")
crs$target <- "Deph.706"
crs$ident <- "DateHeure"
crs$ignore <- c("Deph.706Anc4H", "Deph.706Anc1H")

# Construction du modèle de prévision anticipée de 24 H en avance de l'angle de phase
Deph.706 par le GLM
```

```
crs$glm <- glm(Deph.706 ~ ., data=crs$jeudedonnees[crs$train,c(crs$input, crs$target)], family=gaussian(identity))
```

Générer une vue textuelle du modèle Linéaire.

```
Print(summary(crs$glm))
print(anova(crs$glm))
```

Évaluer la performance du modèle sur l'ensemble de validation

Générer un graphique : Prédit vs. Observé

```
crs$pr <- predict(crs$glm, type= »response », newdata=crs$jeudedonnees[crs$validate, c(crs$input, crs$target)])
```

Obtenir la sortie observée pour le jeu de données.

```
Obs <- subset(crs$jeudedonnees[crs$validate, c(crs$input, crs$target)], select=crs$target)
```

Gérer si une cible catégorique est traitée comme étant numérique.

```
Obs.rownames <- rownames(obs)
obs <- as.numeric(obs[[1]])
obs <- data.frame(Deph.706=obs)
rownames(obs) <- obs.rownames
```

Regrouper les valeurs observées avec celle prédites

```
fitpoints <- na.omit(cbind(obs, Prédit=crs$pr))
```

Obtenir la corrélation R

```
fitcorr <- format(cor(fitpoints[,1], fitpoints[,2]), digits=4)
cat('===== Validation fitcorr G06 24H ===== ')
print(fitcorr)
```

Tracer les paramètres pour les points réels et le meilleur ajustement

```
op <- par(c(lty="solid", col="blue"))
```

Afficher les points de (X) observé en fonction du (Y) prédit

```
plot(fitpoints[[1]], fitpoints[[2]], asp=1, xlab="Deph.706", ylab="Predicted")
```

Générer un ajustement linéaire simple entre les valeurs prédites et observées.

```

Prline <- lm(fitpoints[,2] ~ fitpoints[,1])

# Ajouter l'ajustement linéaire au graphique

abline(prline)

# Ajouter une diagonale représentant une corrélation parfaite

par(c(lty="dashed", col="black"))
abline(0, 1)

# Inclure le facteur de corrélation R au graphique

legend("bottomright", print("R=%s", fitcorr), bty="n")

# Ajouter un titre et une grille au graphique.

Title(main= »Prédit vs. Observé
Le modèle linéaire G06_24H vs.D06 sur l'ensemble de validation « )
grid()

# Évaluer la performance du modèle sur l'ensemble test

# Générer un graphique Prédit vs. Observé

crs$pr <- predict(crs$glm, type= »response », newdata=crs$jeudedonnees[crs$test,
c(crs$input, crs$target)])

# Obtenir la sortie observée pour le jeu de données

obs <- subset(crs$jeudedonnees[crs$test, c(crs$input, crs$target)], select=crs$target)

# Regrouper les valeurs observées avec celle prédites

fitpoints <- na.omit(cbind(obs, Prédit=crs$pr))

# Obtenir la corrélation R

fitcorr <- format(cor(fitpoints[,1], fitpoints[,2]), digits=4)
cat('===== Test fitcorr G06 24H ===== ')
print(fitcorr)

# Tracer les paramètres pour les points réels et le meilleur ajustement

op <- par(c(lty="solid", col="blue"))

```

```

# Afficher les points de (X) observé en fonction du (Y) prédit
plot(fitpoints[[1]], fitpoints[[2]], asp=1, xlab="Deph.706", ylab="Predicted")

# Générer un ajustement linéaire simple entre les valeurs prédictes et observées.

Prline <- lm(fitpoints[,2] ~ fitpoints[,1])

# Ajouter l'ajustement linéaire au graphique
abline(prline)

# Ajouter une diagonale représentant une corrélation parfaite
par(c(lty="dashed", col="black"))
abline(0, 1)

# Inclure le facteur de corrélation R au graphique
legend("bottomright", print(" R= %s ", fitcorr), bty="n")

# Ajouter un titre et une grille au graphique
Title(main= »Prédit vs. Observé
Le modèle linéaire G06_24H vs.D06 sur l'ensemble test »)
grid()

```

Revoir l'ensemble de données :

Il est souvent important de visualiser les données. Les instructions les plus usuelles sont :

```

dim(crs$jeudedonnees)
names(crs$jeudedonnees)
head(crs$jeudedonnees)
tail(crs$jeudedonnees)
str(crs$jeudedonnees)
summary(crs$jeudedonnees)

```

ANNEXE II

LES PARAMÈTRES DES MODÈLES G06_24H ET G06_4H

Cette présente annexe englobe les résultats des deux modèles de prévision anticipées vingt-quatre heures et quatre heures en avance de l'angle de phase D06. Les paramètres de chacun des modèles sont inscrits dans les tableaux-A II-1 et II-3 respectivement.

Les tests ANOVA permettent de rapporter l'effet d'une variable catégorielle sur une variable continue par le rapport de vraisemblance ou la différence de déviance. Ce rapport consiste à évaluer l'apport des variables explicatives. Pour ces tests, les termes sont rajoutés séquentiellement de la première variable à la dernière, dans l'ajustement du modèle. Les tableaux-A II-2 et II-4, récapitulent les résultats des tests ANOVA.

Le tableau A II-5 résume les tendances des résidus des trois modèles de prévisions de D06.

Tableau- A II-1 L'estimation des coefficients des paramètres de **G06.24H**

Coefficients	Estimate	Std. Error	t value	Pr(> t)	Code
(Intercept)	-5,741519401	0,022007479	-260,89	<2e-16	***
Moisavril	-1,195461370	0,011118607	-107,52	<2e-16	***
Moisdécembre	-1,493605488	0,015624651	-95,59	<2e-16	***
Moisfévrier	-1,237115865	0,016194409	-76,39	<2e-16	***
Moisjanvier	-1,396214091	0,016900578	-82,61	<2e-16	***
Moisjuillet	-0,248533087	0,010877808	-22,85	<2e-16	***
Moisjuin	0,486156754	0,010804544	45,00	<2e-16	***
Moismai	-0,610178586	0,010633575	-57,38	<2e-16	***
Moismars	-1,028136436	0,012742429	-80,69	<2e-16	***
Moisnovembre	-0,21à 7352421	0,012666798	-17,16	<2e-16	***
Moisoctobre	-1,032216825	0,011381151	-90,69	<2e-16	***
Moisseptembre	0,175857426	0,010970185	16,03	<2e-16	***
Joursjeudi	-0,436190215	0,008532125	-51,12	<2e-16	***
Jourslundi	0,468108779	0,008559345	54,69	<2e-16	***
Joursmardi	-0,423467035	0,008556748	-49,49	<2e-16	***
Joursmercredi	-0,229860122	0,008563169	-26,84	<2e-16	***
Jourssamedi	-0,831813771	0,008291773	-100,32	<2e-16	***
Joursvendredi	-0,594078405	0,008461559	-70,21	<2e-16	***
Sample2H	-0,146629112	0,008966712	-16,35	<2e-16	***
Sample3H	-0,523218758	0,010349130	-50,56	<2e-16	***
Sample4H	-0,597758708	0,010751957	-55,59	<2e-16	***
Sample5H	-0,699384451	0,010372490	-67,43	<2e-16	***
Sample6H	-0,709085153	0,010719928	-66,15	<2e-16	***
Sample7H	-0,765537140	0,010876127	-70,39	<2e-16	***
Sample8H	-0,334843213	0,009299014	-36,01	<2e-16	***
Deph.706Anc24H	0,433429879	0,000706723	613,29	<2e-16	***
BQ.PR	-0,000083257	0,000005565	-14,96	<2e-16	***

Coefficients	Estimate	Std. Error	t value	Pr(> t)	Code
BQ.PR24H	-0,000425024	0,000005136	-82,75	<2e-16	***
BG	0,001231564	0,000002658	463,41	<2e-16	***

Tableau- A II-2 La table de la déviance ANOVA suite à la modélisation de D06 vingt-quatre heures en avance

Les paramètres	Df	Deviance	Resid. Df	Resid. Dev
NULL			1011286	45528022
Mois	11	20324574	1011275	25203449
Jours	6	731811	1011269	24471638
Sample	7	8739124	1011262	15732514
Deph.706Anc24H	1	8173327	1011261	7559188
BQ.PR	1	1349254	1011260	6209934
BQ.PR24H	1	28809	1011259	6181125
BG	1	1082676	1011258	5098448

Tableau- A II-3 L'estimation des coefficients des paramètres de **G06.4H**

Les paramètres	Estimate	Std. Error	t value	Pr(> t)	Code
(Intercept)	-8,881175922	0,022246860	-399,21	<2e-16	***
Moisavril	-1,330352562	0,011240029	-118,36	<2e-16	***
Moisdécembre	-1,972256051	0,015761181	-125,13	<2e-16	***
Moisfévrier	-1,681814236	0,016427503	-102,38	<2e-16	***
Moisjanvier	-2,045142597	0,017202711	-118,89	<2e-16	***
Moisjuillet	-0,325847560	0,010989612	-29,65	<2e-16	***
Moisjuin	0,580637665	0,010927465	53,14	<2e-16	***
Moismai	-0,674122559	0,010755984	-62,67	<2e-16	***
Moismars	-1,316892465	0,012927362	-101,87	<2e-16	***
Moisnovembre	-0,425605999	0,012816012	-33,21	<2e-16	***
Moisoctobre	-1,085310643	0,011488598	-94,47	<2e-16	***
Moisseptembre	0,306837444	0,011092678	27,66	<2e-16	***
Joursjeudi	-0,525843610	0,008646480	-60,82	<2e-16	***
Jourslundi	-0,389565483	0,008596550	-45,32	<2e-16	***
Joursmardi	-0,531930292	0,008660933	-61,42	<2e-16	***
Joursmercredi	-0,445082806	0,008667259	-51,35	<2e-16	***
Jourssamedi	-0,236553519	0,008330706	-28,39	<2e-16	***
Joursvendredi	-0,641312056	0,008539995	-75,09	<2e-16	***
Sample2H	2,003796763	0,009710854	206,35	<2e-16	***
Sample3H	3,710554965	0,012124376	306,04	<2e-16	***
Sample4H	1,869483126	0,011304835	165,37	<2e-16	***
Sample5H	0,371553685	0,010564558	35,17	<2e-16	***
Sample6H	1,211909107	0,011057921	109,60	<2e-16	***
Sample7H	1,164762815	0,011249146	103,54	<2e-16	***
Sample8H	-0,668413531	0,009431898	-70,87	<2e-16	***
Deph.706Anc4H	0,382348508	0,000658212	580,89	<2e-16	***
BQ.PR	-0,000374681	0,000005553	-67,48	<2e-16	***

Les paramètres	Estimate	Std. Error	t value	Pr(> t)	Code
BQ.PR24H	-0,000215139	0,000005165	-41,66	<2e-16	***
BG	0,001417565	0,000002571	551,27	<2e-16	***
Signification du code	0 ***!	0,001 **!	0,01 *!	0,05 !!	0,1 ''
					1

Tableau- A II-4 La table de la déviance ANOVA suite à la modélisation de quatre heures en avance de D06

Column1	Df	Deviance	Resid. Df	Resid. Dev
NULL			1016848	46133570
Mois	11	20689313	1016837	25444256
Jours	6	à 735266	1016831	24708991
Sample	7	8894643	1016824	15814347
Deph.706Anc4H	1	6938471	1016823	8875876
BQ.PR	1	2048514	1016822	6827362
BQ.PR24H	1	1267	1016821	6826095
BG	1	1570701	1016820	5255394

Tableau- A II-5 Les déviances des résidus des modèles de prévision de D06

Les déviances des résidus G06	Min	1Q	Median	3Q	Max
Prévision 24H en avance	-19,3524	-1,4675	-0,0013	1,4548	20,1 à 735
Prévision 4H en avance	-19,5985	-1,5231	-0,0231	1,4955	21,0732
Prévision 1H en avance	-12,3529	-0,8408	-0,0476	0,7751	18,5689

ANNEXE III

RÉSULTATS DE PRÉVISIONS

Cette annexe recueillie les résultats de prévisions des angles de phase de tension.

Nous reprenons la répartition de l'ensemble des angles de phases de tension autour des angles medoids, obtenue préalablement dans le chapitre 3. Ainsi nous distinguons dans ces tableaux par trois couleurs différentes, les trois groupes des angles de phases de tension.

Le tableau A III-1 récapitule les performances des différents algorithmes utilisés pour la modélisation des angles de phase de tension pour les prévisions d'une heure en avance.

Tableau- A III-1 Les performances des algorithmes : rpart, GLM et NNET

An gle	MAE. rpart	MAE. GLM	MAE. NNET	R. rpart	R. GLM	R. NNET	Rsq. rpart	Rsq. GLM	Rsq. NNET
D01	0,165	0,2005	0,2005	0,9021	0,8611	0,8611	0,7729	0,6613	0,6613
D02	0,2144	0,2407	0,2407	0,9079	0,8811	0,8811	0,7869	0,7189	0,7189
D03	0,5344	0,5001	0,5001	0,9579	0,9719	0,9719	0,9093	0,9419	0,9419
D04	0,5493	0,5162	0,5162	0,9593	0,9729	0,9729	0,912	0,944	0,944
D05	1,03	0,9841	0,9841	0,9623	0,976	0,976	0,9193	0,9507	0,9507
D06	1,101	1,058	1,058	0,9643	0,9777	0,9777	0,9239	0,9544	0,9544
D07	0,3331	0,322	0,322	0,9465	0,956	0,956	0,8826	0,907	0,907
D08	0,1942	0,2202	0,2202	0,9156	0,8843	0,8843	0,8066	0,7265	0,7265
D09	1,139	1,076	1,076	0,9647	0,9788	0,9788	0,9247	0,9567	0,9567
D10	1,251	1,161	1,161	0,9673	0,9818	0,9818	0,9304	0,9629	0,9629
D13	1,102	1,077	1,077	0,9571	0,9698	0,9698	0,9074	0,9374	0,9374
D14	0,7914	0,7943	0,7943	0,9464	0,9553	0,9553	0,8821	0,9053	0,9053
D15	0,326	0,3357	0,3357	0,9321	0,9291	0,9291	0,8471	0,8433	0,8433
D17	0,55	0,5077	0,5077	0,9599	0,9747	0,9747	0,9136	0,9479	0,9479

An	MAE.	MAE.	MAE.	R.	R.	R.	Rsq.	Rsq.	Rsq.
gle	rpart	GLM	NNET	rpart	GLM	NNET	rpart	GLM	NNET
D19	0,3425	0,3551	0,3551	0,9487	0,9492	0,9492	0,8867	0,8907	0,8907
D22	1,746	1,708	1,708	0,9632	0,9747	0,9747	0,9211	0,9479	0,9479
D23	1,676	1,643	1,643	0,9627	0,9754	0,9754	0,9201	0,9494	0,9494
D24	1,827	1,806	1,806	0,9608	0,972	0,972	0,9158	0,9421	0,9421
D30	0,2287	0,2431	0,2431	0,9227	0,9142	0,9142	0,8251	0,8078	0,8078
D31	0,9105	0,8694	0,8694	0,9598	0,9741	0,9741	0,9136	0,9468	0,9468
D49	1,865	1,822	1,822	0,9625	0,975	0,975	0,9196	0,9484	0,9484
D50	1,747	1,701	1,701	0,9635	0,976	0,976	0,9219	0,9507	0,9507
D51	1,748	1,712	1,712	0,9621	0,9737	0,9737	0,9164	0,9458	0,9458
D52	1,729	1,691	1,691	0,9625	0,9755	0,9755	0,9093	0,9497	0,9497
D55	0,2958	0,2961	0,2961	0,9386	0,9418	0,9418	0,8638	0,8746	0,8746
D60	1,366	1,271	1,271	0,9681	0,9816	0,9816	0,9322	0,9626	0,9626
D64	1,828	1,796	1,796	0,961	0,9739	0,9739	0,9159	0,9462	0,9462
D65	1,83	1,824	1,824	0,9601	0,9703	0,9703	0,9133	0,9385	0,9385
D66	1,828	1,8	1,8	0,9606	0,9733	0,9733	0,9026	0,9449	0,9449
D70	0,4221	0,423	0,423	0,9488	0,9473	0,9473	0,8875	0,8866	0,8866
D80	1,394	1,353	1,353	0,9612	0,9742	0,9742	0,9165	0,9468	0,9468
D82	1,375	1,335	1,335	0,962	0,975	0,975	0,9184	0,9485	0,9485
D83	1,094	1,055	1,055	0,9604	0,9741	0,9741	0,9148	0,9466	0,9466
D84	0,216	0,2381	0,2381	0,9148	0,8945	0,8945	0,8052	0,7555	0,7555
D90	0,4221	0,4036	0,4036	0,9512	0,9629	0,9629	0,8938	0,9223	0,9223
MED.	1,094	1,055	1,055	0,9604	0,9737	0,9737	0,9136	0,9458	0,9458
MOY.	1,02306	0,99799	0,99799	0,95209	0,95821	0,95821	0,89404	0,90858	0,90858

Le tableau A III-2 récapitule les résultats d'évaluation, avec le critère du coefficient de corrélation R, de la performance des modèles de prévision des angles de phase de tension une heure, quatre heures et vingt-quatre heures en avance, avec l'algorithme GLM.

Tableau-A III-2 Les performances des 111 modèles de prévisions des angles de phase

Phase G	R	R	R
	24 H en avance	4 H en avance	1 H en avance
G01	62,41	76,80	86,08
G02	65,16	77,61	88,05
G03	91,36	92,54	97,20
G04	91,58	92,97	97,29
G05	93,46	93,45	97,6
G06	94,23	94,12	97,78
G07	87,20	90,39	95,59
G08	65,21	78,49	88,38
G09	94,25	94,17	97,88
G10	95,03	94,85	98,18
G13	92,42	92,73	96,99
G14	87,59	88,81	95,53
G15	75,40	83,35	92,85
G17	92,30	93,56	97,46
G18	94,02	94,45	97,99
G19	84,10	88,15	94,89
G20	93,70	93,77	97,48
G22	94,27	94,28	97,47
G23	94,11	94,15	97,55
G24	93,53	93,56	97,24
G30	77,29	84,83	91,40
G31	93,45	93,7	97,42

Phase G	R 24 H en avance	R 4 H en avance	R 1 H en avance
G49	93,81	93,86	97,51
G50	94,33	94,32	97,59
G51	94,19	94,20	97,35
G52	94,30	94,35	97,54
G55	82,75	87,60	94,18
G60	95,07	94,89	98,16
G64	93,70	93,71	97,40
G65	93,48	93,53	97,07
G66	93,69	93,78	97,30
G70	76,05	84,95	94,71
G80	93,74	93,94	97,42
G82	94,04	94,19	97,51
G83	93,67	93,90	97,41
G84	96,41	79,76	89,44
G90	88,87	90,93	96,30

Le tableau A III-3 récapitule les résultats d'évaluation, avec des différents critères, la performance du modèle de prévision des angles de phase de tension une heure en avance.

Tableau- A III-3 Les performances des modèles de prévisions une heure en avance

Angle	MAE (degrés)	R (%)	MSE	Rsq (%)
D01	0,2005	0,8611	0,1919	0,6613
D02	0,2407	0,8811	0,2215	0,7189
D03	0,5001	0,9719	0,5264	0,9419
D04	0,5162	0,9729	0,5679	0,944
D05	0,9841	0,976	1,759	0,9507
D06	1,058	0,9777	1,998	0,9544
D07	0,322	0,956	0,286	0,907
D08	0,2202	0,8843	0,2109	0,7265
D09	1,076	0,9788	2,091	0,9567
D10	1,161	0,9818	2,461	0,9629
D13	1,077	0,9698	2,034	0,9374
D14	0,7943	0,9553	1,256	0,9053
D15	0,3357	0,9291	0,327	0,8433
D17	0,5077	0,9747	0,5286	0,9479
D18	0,7819	0,9799	1,127	0,9589
D19	0,3551	0,9492	0,3411	0,8907
D20	1,805	0,9747	5,599	0,948
D22	1,708	0,9747	5,322	0,9479
D23	1,643	0,9754	4,641	0,9494
D24	1,806	0,972	6,024	0,9421
D30	0,2431	0,9142	0,2145	0,8078
D31	0,8694	0,9741	1,343	0,9468
D49	1,822	0,975	5,706	0,9484
D50	1,701	0,976	5,055	0,9507

Angle	MAE (degrés)	R (%)	MSE	Rsq (%)
D51	1,712	0,9737	5,471	0,9458
D52	1,691	0,9755	5,008	0,9497
D55	0,2961	0,9418	0,2658	0,8746
D60	1,271	0,9816	3,213	0,9626
D64	1,796	0,9739	5,621	0,9462
D65	1,824	0,9703	6,432	0,9385
D66	1,8	0,9733	5,779	0,9449
D70	0,423	0,9473	0,575	0,8866
D80	1,353	0,9742	3,151	0,9468
D82	1,335	0,975	3,052	0,9485
D83	1,055	0,9741	1,935	0,9466
D84	0,2381	0,8945	0,2214	0,7555
D90	0,4036	0,9629	0,3846	0,9223

LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES

- Abed, A. M. 1999. « WSCC voltage stability criteria, undervoltage load shedding strategy, and reactive power reserve monitoring methodology ». In *Power Engineering Society Summer Meeting, 1999. IEEE.* (18-22 Jul 1999) Vol. 1, p. 191-197 vol.1.
- Ackerman, P. 1930. « Abridgment of the problem of service security in large transmission systems ». *A.I.E.E., Journal of the*, vol. 49, n° 9, p. 767-770.
- Al-Masri, A. N., M. Z. A. Ab Kadir, H. Hizam et N. Mariun. 2013. « A Novel Implementation for Generator Rotor Angle Stability Prediction Using an Adaptive Artificial Neural Network Application for Dynamic Security Assessment ». *Power Systems, IEEE Transactions on*, vol. 28, n° 3, p. 2516-2525.
- Allen, A. J., S. Santoso et W. M. Grady. 2010. « Voltage phase angle variation in relation to wind power ». In *Power and Energy Society General Meeting, 2010 IEEE.* (25-29 July 2010), p. 1-7.
- Allen, A., S. Santoso et E. Muljadi. 2013 *Algorithm for Screening Phasor Measurement Unit Data for Power System Events and Categories and Common Characteristics for Events Seen in Phasor Measurement Unit Relative Phase-Angle Differences and Frequency Signals.* En ligne. Coll. « Technical Report NREL », TP-5500-58611. U.S.Departement of Energy, 79 p. <<http://www.nrel.gov/docs/fy13osti/58611.pdf>>. Consulté le 11 octobre 2015.
- Alves da Silva, A. P. 2005. « Overcoming limitations of NNs for on-line DSA ». In *Power Engineering Society General Meeting, 2005. IEEE.* (12-16 June 2005), p. 2653-2660 Vol. 3.
- Aminifar, FARROKH, MAHMUD Fotuhi-Firuzabad, A Davoudi, A Safdarian et MOHAMMAD Shahidehpour. 2015. « Synchrophasor Measurement Technology in Power Systems: Panorama and State-of-the-Art ». *The journal for rapid open access publishing*, vol. 2, n° 2014, p. 1607-1628.
- Anderson, Paul M., et A. Bose. 1983. « A Probabilistic Approach to Power System Stability Analysis ». *Power Apparatus and Systems, IEEE Transactions on*, vol. PAS-102, n° 8, p. 2430-2439.
- Baldick, R., B. Chowdhury, I. Dobson, Dong Zhaoyang, Gou Bei, D. Hawkins, H. Huang, Joung Manho, D. Kirschen, Li Fangxing, Li Juan, Li Zuyi, Liu Chen-Ching, L. Mili, S. Miller, R. Podmore, K. Schneider, Sun Kai, D. Wang, Wu Zhigang, Zhang Pei, Zhang Wenjie et Zhang Xiaoping. 2008. « Initial review of methods for cascading failure analysis in electric power transmission systems IEEE PES CAMS task force on understanding, prediction, mitigation and restoration of cascading failures ». In

Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE. (20-24 July 2008), p. 1-8.

Baldick, R., B. Chowdhury, I. Dobson, Dong Zhaoyang, Gou Bei, D. Hawkins, Huang Zhenyu, Joung Manho, Kim Janghoon, D. Kirschen, S. Lee, Li Fangxing, Li Juan, Li Zuyi, Liu Chen-Ching, Luo Xiaochuan, L. Mili, S. Miller, M. Nakayama, M. Papic, R. Podmore, J. Rossmaier, K. Schneider, Sun Hongbin, Sun Kai, D. Wang, Wu Zhigang, Yao Liangzhong, Zhang Pei, Zhang Wenjie et Zhang Xiaoping. 2009. « Vulnerability assessment for cascading failures in electric power systems ». In *Power Systems Conference and Exposition, 2009. PSCE '09. IEEE/PES.* (15-18 March 2009), p. 1-9.

Basu, C., A. Agrawal, J. Hazra, A. Kumar, D. P. Seetharam, J. Béland, S. Guillon, I. Kamwa et C. Lafond. 2014. « Understanding events for wide-area situational awareness ». In *Innovative Smart Grid Technologies Conference (ISGT), 2014 IEEE PES.* (19-22 Feb. 2014), p. 1-5.

Baum, Christopher F. 2006. *An introduction to modern econometrics using Stata*. United States of America: Stata press, 341 p.

Besse, Philippe. 2009. *Apprentissage statistique & data mining*. Toulouse: INSA, 124 p.

Besse, Philippe, Aurélien Garivier et Jean Michel Loubes. 2014. « Big Data Analytics-Retour vers le Futur 3 : De Statisticien à Data Scientist ». *arXiv*, p. 1-14.

Bose, A. 1984. « Application of Direct Methods to Transient Stability Analysis of Power Systems ». *Power Apparatus and Systems, IEEE Transactions on*, vol. PAS-103, n° 7, p. 1629-1636.

Breiman, Léo. 2001. « Random Forests ». *Machine Learning*, vol. 45, n° 1, p. 5-32.

Breiman, Leo, Jerome Friedman, Charles J Stone et Richard A Olshen. 1984. *Classification and regression trees*, 1st ed. Coll. « Wadsworth Statistics/Probability series ». Boca Raton: Chapman and Hall/ CRC, 368 p.

Cepeda, J. C., D. G. Colome et N. J. Castrillon. 2011. « Dynamic vulnerability assessment due to transient instability based on data mining analysis for Smart Grid applications ». In *Innovative Smart Grid Technologies (ISGT Latin America), 2011 IEEE PES Conference on.* (19-21 Oct. 2011), p. 1-7.

Chen-Ching, Liu, et K. Tomsovic. 1986. « An Expert System Assisting Decision-Making of Reactive Power/Voltage Control ». *Power Systems, IEEE Transactions on*, vol. 1, n° 3, p. 195-201.

- Chengxi, Liu, C. L. Bak, Chen Zhe et P. Lund. 2014. « Dynamic security assessment of western Danish power system based on ensemble decision trees ». In *Developments in Power System Protection (DPSP 2014), 12th IET International Conference on.* (March 31 2014-April 3 2014), p. 1-6.
- CLAIR, H.P.ST. 1953. « Practical Concepts in Capability and Performance of Transmission Lines ». *Power Apparatus and Systems, Part III. Transactions of the American Institute of Electrical Engineers*, vol. 72, n° 2, p. 1151-1157.
- De Montigny, M., A. Heniche, I. Kamwa, L. Cauchon, R. Mailhot, S. Lebeau, D. Lefebvre et L. Bernier. 2013. « Multiagent Stochastic Simulation of Minute-to-Minute Grid Operations and Control to Integrate Wind Generation Under AC Power Flow Constraints ». *Sustainable Energy, IEEE Transactions on*, vol. 4, n° 3, p. 619-629.
- DOE. 2011. *Department of Energy Quadrennial Technology Review Framing Document*. En ligne, 40 p. <<http://energy.gov/QTR>>. Consulté le 2 octobre 2015.
- DOE. 2012. *QTR : REPORT ON THE FIRST QUADRENNIAL TECHNOLOGY REVIEW*. Coll. « TECHNOLOGY ASSESSMENTS ». Washington, DC, 292 p.
- DOE. 2013. *Synchrophasor Technologies and Their Deployment in the Recovery Act Smart Grid Smart Programs*. Coll. « Amerriican Recovery and Reinvestment Act of 2009 ». US: Electricity Delivery and Energy Reliability, 17 p.
- DOE. 2014. *Factors Affecting PMU Installation Costs : Smart Grid InvestmentGrant Program* Coll. « American Recovery and Reinvestment Act of 2009 ». US: Electricity Delivery and Energy Reliability, 44 p.
- Dommel, H. W., et N. Sato. 1972. « Fast Transient Stability Soultions ». *Power Apparatus and Systems, IEEE Transactions on*, vol. PAS-91, n° 4, p. 1643-1650.
- Doyle, Patrick, R. Mailhot, H. Horisberger, E. Aboumrad, J. Gauthier et G. Versailles. 2004. *LIMSEL: An efficient software for power system reliability monitoring and control*. C2-204. Paris: CIGRE, 8 p.
- El-Keib, A. A., et X. Ma. 1995. « Application of artificial neural networks in voltage stability assessment ». *Power Systems, IEEE Transactions on*, vol. 10, n° 4, p. 1890-1896.
- Energy, U.S. Dept. 2008. *Smart grid: Enabler of the new energy economy*. <<http://www.oe.energy.gov/DocumentsandMedia/final-smart-grid-report.pdf>>.
- Fayyad, Usama M, Gregory Piatetsky-Shapiro et Padhraic Smyth. 1996. « Knowledge Discovery and Data Mining: Towards a Unifying Framework ». In *KDD*. Vol. 96, p. 82-88.

FERC. 2009a. *Proposed Smart grid policy Policy Statement and Action Plan*. En ligne, 1 p. < www.ferc.gov/EventCalendar/Files/20090319121942-E-22-fact-sheet.pdf >. Consulté le 11 octobre 2015.

FERC. 2009b. *smart Grid Policy Statement*. en ligne, 105 p. < <http://www.ferc.gov/whats-new/comm-meet/2009/071609/E-3.pdf> >. Consulté le 09 octobre 2015.

Ferryman, T. A. , et B. G. Amidan. 2012. « Investigation of Phase Angle Differences Using Statistical Analysis of Real World State Estimator Data ». *Proc. 45th Hawaii Int. Conf. on System Sciences*, n° January, p. 2062-2071.

Fouad, A. A., F. Aboytes, V. F. Carvalho, S. L. Corey, K. J. Dhir et R. Vierra. 1988. « Dynamic security assessment practices in North America ». *Power Systems, IEEE Transactions on*, vol. 3, n° 3, p. 1310-1321.

Fouad, A. A., Zhou Qin et V. Vittal. 1994. « System vulnerability as a concept to assess power system dynamic security ». *Power Systems, IEEE Transactions on*, vol. 9, n° 2, p. 1009-1015.

Fouad, A. A., S. Vekataraman et J. A. Davis. 1991. « An expert system for security trend analysis of a stability-limited power system ». *Power Systems, IEEE Transactions on*, vol. 6, n° 3, p. 1077-1084.

Gareth, James, Daniela Witten, Trevor Hastie et Robert Tibshirani. 2013. *An Introduction to statistical learning: with Applications in R* Coll. « Texts in Statistics ». New York: Springer 441 p.

Giri, Jay, Manu Parashar, Jos Trehern et Vahid Madani. 2012. « The situation room: Control center analytics for enhanced situational awareness ». *Power and Energy Magazine, IEEE*, vol. 10, n° 5, p. 24-39.

Grenier, M., J. Gagnon, C. Mercier et J. Richard. 2006. « Short-term load forecasting at Hydro-Quebec TransEnergie ». In *Power Engineering Society General Meeting*. (Montréal, Québec), p. 1-5. IEEE.

Guillon, S. 2005. *Présentation, bilan et synthèse d'une activité au coeur d'une exploitation sécuritaire et fiable du réseau de TransÉnergie : détermination des limites d'exploitation et des stratégies d'exploitation associées*. IREQ-2005-021.

Gutman, R. 1988. « Application of line loadability concepts to operating studies ». *Power Systems, IEEE Transactions on*, vol. 3, n° 4, p. 1426-1433.

Gutman, R., P. P. Marchenko et R. D. Dunlop. 1979. « Analytical Development of Loadability Characteristics for EHV and UHV Transmission Lines ». *Power Apparatus and Systems, IEEE Transactions on*, vol. PAS-98, n° 2, p. 606-617.

- Hakim, Hossein. 1992. « Application of Pattern Recognition in Transient Security Assessment ». *Electric Machines & Power Systems*, vol. 20, n° 1, p. 1-15.
- Hatziaargyriou, N. D. , S. A. Papathanassiou et M. P. Papadopoulos. 1995. « Decision trees for fast security assessment of autonomous power systems with a large penetration from renewables ». *Energy Conversion, IEEE Transactions on*, vol. 10, n° 2, p. 315-325.
- Havens, T. C., et J. C. Bezdek. May 2012. « An efficient formulation of the improved visual assessment of cluster tendency (iVAT) algorithm ». *IEEE Trans. Knowl. Data Eng.* , vol. 24, n° 5, p. 813-822.
- Heniche, A., I. Kamwa et M. Dobrescu. 2013. « Hydro-Québec's defense plan: Present and future ». In *Power and Energy Society General Meeting (PES), 2013 IEEE*. (21-25 July 2013), p. 1-5.
- Heqing, Huang, Zhao Feng et Ye Min. 2010. « Estimate the influential level of vulnerability instance based on hybrid ranking for dynamic network attacking scenarios ». In *Information Sciences Signal Processing and their Applications (ISSPA), 2010 10th International Conference on*. (10-13 May 2010), p. 586-589.
- Hong, Tao, et Mohammad Shahidehpour. 2015. *Load Forecasting : Case Study*. DE-OE0000316. EISPC and NARUC: U.S. Department of Energy, 171 p. <http://www.naruc.org/Grants/Documents/Load%20Forecasting%20Case%20Study%20_%20UNCC.pdf>. Consulté le 09 octobre 2015.
- Hong, Tao, Pu Wang et H. L. Willis. 2011. « A Naïve multiple linear regression benchmark for short term load forecasting ». In *Power and Energy Society General Meeting, 2011 IEEE*. (24-29 July 2011), p. 1-6.
- HQD. 2010. *Simulateur SIRE: Nouvelle approche par simulation pour évaluer les impacts de l'intégration de l'aproduction éolienne sur le service de régulation de la fréquence*. Coll. « Hydro-Québec Distribution », Demande R-3748-2010. En ligne, 30 p. <http://publicsde.regie-energie.qc.ca/projets/12/DocPrj/R-3748-2010-B-0025-DDR-REPDDR-2011_03_15.pdf>. Consulté le le 10 octobre 2015.
- Huang, J. A., L. Loud, G. Vanier, B. Lambert et S. Guillon. 2012. « Experiences and challenges in contingency analysis at Hydro-Quebec ». In *Power and Energy Society General Meeting, 2012 IEEE*. (22-26 July 2012), p. 1-9.
- Huang, J. A., A. Valette, M. Beaudoin, K. Morison, A. Moshref, M. Provencher et J. Sun. 2002. « An intelligent system for advanced dynamic security assessment ». In *Power System Technology, 2002*. (13-17 Oct 2002) Vol. 1, p. 220-224.

- Husson, François, Lê Sébastien et Pagès Jérôme. 2009. *Analyse de données avec R*. Coll. « Didact Statistique ». Presses universitaires de Rennes, 224 p.
- IEEE. 2006. *IEEE Standard for Synchrophasors for Power Systems*. New York: Power System Relaying Committee, 65 p.
- Jihun, Kim, Lee Byongjun, Han Sangwook, Shin Jeong-Hoon, Kim Taekyun, Kim Sangtae et Moon Younghwan. 2013. « Study of the Effectiveness of a Korean Smart Transmission Grid Based on Synchro-Phasor Data of K-WAMS ». *Smart Grid, IEEE Transactions on*, vol. 4, n° 1, p. 411-418.
- Julian, Faraway. 2002. *Practical regression and ANOVA using R*. University of Bath, 213 p.
- Jun, Zhu, E. Zhuang, C. Ivanov et Yao Ziwen. 2011. « A Data-Driven Approach to Interactive Visualization of Power Systems ». *Power Systems, IEEE Transactions on*, vol. 26, n° 4, p. 2539-2546.
- Kaci, A., I. Kamwa, L. A. Dessaint et S. Guillon. 2014a. « Phase angles as predictors of network dynamic security limits and further implications ». In *PES General Meeting. Conference & Exposition, 2014 IEEE*. (27-31 July 2014), p. 1-6. <<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6939281>>. Consulté le 11 octobre 2015.
- Kaci, A., I. Kamwa, L. A. Dessaint et S. Guillon. 2014b. « Synchrophasor Data Baselining and Mining for Online Monitoring of Dynamic Security Limits ». *Power Systems, IEEE Transactions on*, vol. 29, n° 6, p. 2681- 2695.
- Kai, Sun, S. Likhate, V. Vittal, V. S. Kolluri et S. Mandal. 2007. « An Online Dynamic Security Assessment Scheme Using Phasor Measurements and Decision Trees ». *Power Systems, IEEE Transactions on*, vol. 22, n° 4, p. 1935-1943.
- Kalyani, S., et K. S. Swarup. 2011. « Classification and Assessment of Power System Security Using Multiclass SVM ». *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 41, n° 5, p. 753-758.
- Kamwa, I., J. Beland et D. McNabb. 2006. « PMU-Based Vulnerability Assessment Using Wide-Area Severity Indices and Tracking Modal Analysis ». In *Power Systems Conference and Exposition, 2006. PSCE '06. 2006 IEEE PES*. (Oct. 29 2006-Nov. 1 2006), p. 139-149.
- Kamwa, I., J. Beland, G. Trudel, R. Grondin, C. Lafond et D. McNabb. 2006. « Wide-area monitoring and control at Hydro-Québec: past, present and future ». In *Power Engineering Society General Meeting, 2006*. (Montréal, Québec), p. 1-12. IEEE.

- Kamwa, I., et R. Grondin. 2002. « PMU configuration for system dynamic performance measurement in large, multiarea power systems ». *Power Systems, IEEE Transactions on*, vol. 17, n° 2, p. 385-394.
- Kamwa, I., A. Heniche, M. De Montigny, R. Mailhot, S. Lebeau, L. Bernier et A. Robitaille. 2012. « Long-Term Statistical Assessment of Frequency Regulation Reserves Policies in the Québec Interconnection ». *Sustainable Energy, IEEE Transactions on*, vol. 3, n° 4, p. 868-879.
- Kamwa, I., A. K. Pradhan et G. Joos. 2007. « Automatic Segmentation of Large Power Systems Into Fuzzy Coherent Areas for Dynamic Vulnerability Assessment ». *Power Systems, IEEE Transactions on*, vol. 22, n° 4, p. 1974-1985.
- Kamwa, I., A. K. Pradhan, G. Joos et S. R. Samantaray. 2009. « Fuzzy Partitioning of a Real Power System for Dynamic Vulnerability Assessment ». *Power Systems, IEEE Transactions on*, vol. 24, n° 3, p. 1356-1365.
- Kamwa, I., S. R. Samantaray et G. Joos. 2009. « Development of Rule-Based Classifiers for Rapid Stability Assessment of Wide-Area Post-Disturbance Records ». *Power Systems, IEEE Transactions on*, vol. 24, n° 1, p. 258-270.
- Kamwa, I., S. R. Samantaray et G. Joos. 2010. « Catastrophe Predictors From Ensemble Decision-Tree Learning of Wide-Area Severity Indices ». *Smart Grid, IEEE Transactions on*, vol. 1, n° 2, p. 144-158.
- Kamwa, I., S. R. Samantaray et G. Joos. 2012. « On the Accuracy Versus Transparency Trade-Off of Data-Mining Models for Fast-Response PMU-Based Catastrophe Predictors ». *Smart Grid, IEEE Transactions on*, vol. 3, n° 1, p. 152-161.
- Kassabalisidis, I. N., M. A. El-Sharkawi et R. J. Marks, II. 2002. « Border identification for power system security assessment using neural network inversion: an overview ». In *Evolutionary Computation, 2002. CEC '02. Proceedings of the 2002 Congress on*. (2002) Vol. 2, p. 1075-1079.
- Kay, T. W., P. W. Sauer, R. D. Shultz et R. A. Smith. 1982. « EHV and UHV Line Loadability Dependence on Var Supply Vapability ». *Power Apparatus and Systems, IEEE Transactions on*, vol. PAS-101, n° 9, p. 3568-3575.
- Kezunovic, M., Xie Le et S. Grijalva. 2013. « The role of big data in improving power system operation and protection ». In *Bulk Power System Dynamics and Control - IX Optimization, Security and Control of the Emerging Power Grid (IREP), 2013 IREP Symposium*. (25-30 Aug. 2013), p. 1-9.
- Khan, Asfar Ali. 2008. « A Simple Method for Tracing PV Curve of a Radial Transmission Line ». In *Proceedings of World Academy of Science, Engineering and Technology*,

Vol 27, sous la dir. de Ardin, C. Vol. 27, p. 347-350. Coll. « Proceedings of World Academy of Science Engineering and Technology ». Canakkale: World Acad Sci, Eng & Tech-Waset.

Khatib, A. R., R. F. Nuqui, M. R. Ingram et A. G. Phadke. 2004. « Real-time estimation of security from voltage collapse using synchronized phasor measurements ». In *Power Engineering Society General Meeting, 2004. IEEE*. (6-10 June 2004), p. 582-588 Vol.1.

Klump, R., R. E. Wilson et K. E. Martin. 2005. « Visualizing Real-Time Security Threats Using Hybrid SCADA / PMU Measurement Displays ». In *System Sciences, 2005. HICSS '05. Proceedings of the 38th Annual Hawaii International Conference on*. (03-06 Jan. 2005), p. 55c-55c.

Kumar, A. B. R., A. Ipakchi, V. Brandwajn, M. El-Sharkawi et G. Cauley. 1991. « Neural networks for dynamic security assessment of large-scale power systems: requirements overview ». In *Neural Networks to Power Systems, 1991., Proceedings of the First International Forum on Applications of*. (23-26 Jul 1991), p. 65-71.

Lafaye de Micheaux, Pierre , Rémy Drouilhet et Benoit Liquet. 2014. *The R Software : fundamentals of Programming and Statistical Analysis*, 40. Statistics and Computing. New York: Springer Science & Business, 628 p.

Lafaye de Micheaux, Pierre, Rémy Drouilhet et Benoît Liquet. 2011. *Le logiciel R : Maîtriser le langage-Effectuer des analyses statistiques*, 1ère éd. Coll. « Statistique et probabilités appliquées ». Paris: Springer Science & Business Media, 490 p.

Le Guen, Monique. 2001. « La boîte à moustaches de TUKEY, un outil pour initier à la statistique ». *Statistiquement Votre-SFDS*, n° 4, p. 1-3.

Lei, Wang, Tong Jianzhong et Li Haoen. 2012. « An application example for on-line DSA systems ». In *Power System Technology (POWERCON), 2012 IEEE International Conference on*. (Oct. 30 2012-Nov. 2 2012), p. 1-5.

Liacco, T. E. D. 1967. « The Adaptive Reliability Control System ». *Power Apparatus and Systems, IEEE Transactions on*, vol. PAS-86, n° 5, p. 517-531.

Liaw, Andy, et Matthew Wiener. 2002. « Classification and Regression by randomForest ». *R news*, vol. 2, p. 18-22.

Liqiang, Chen, K. Tomsovic, A. Bose et R. Stuart. 2000. « Estimating reactive margin for determining transfer limits ». In *Power Engineering Society Summer Meeting, 2000. IEEE*. (2000) Vol. 1, p. 490-495 vol. 1.

- Loud, L., S. Guillon, G. Vanier, J. A. Huang, L. Riverin, D. Lefebvre et J. C. Rizzi. 2010. « Hydro-Québec's challenges and experiences in on-line DSA applications ». In *Power and Energy Society General Meeting, 2010 IEEE*. (25-29 July 2010), p. 1-8.
- Lu, Wei. 2009. « Le délestage optimal pour la prévention des grandes pannes d'électricité ». Institut National Polytechnique de Grenoble-INPG.
- Mansour, Y., A. Y. Chang, J. Tamby, E. Vaahedi, B. R. Corns et M. A. El-Sharkawi. 1997. « Large scale dynamic security screening and ranking using neural networks ». *Power Systems, IEEE Transactions on*, vol. 12, n° 2, p. 954-960.
- McGillis, D., K. El-Arroudi, R. Brearley et G. Joos. 2006. « The Process of System Collapse Based on Areas of Vulnerability ». In *Power Engineering, 2006 Large Engineering Systems Conference on*. (26-28 July 2006), p. 35-40.
- Ming, Zhou, V. A. Centeno, J. S. Thorp et A. G. Phadke. 2006. « An Alternative for Including Phasor Measurements in State Estimators ». *Power Systems, IEEE Transactions on*, vol. 21, n° 4, p. 1930-1937.
- Morison, K., L. Wang et P. Kundur. 2004. « Power system security assessment ». *Power and Energy Magazine, IEEE*, vol. 2, n° 5, p. 30-39.
- NASPI. 2009. *Update and Technology Roadmap*. En ligne. 10 p. < <https://www.naspi.org/documents> >. Consulté le 11 octobre 2015.
- NERC. 2004. *Technical Analysis of the August 14, 2003, Blackout: What Happened, Why, and What Did We Learn?* Coll. « Report to the NERC Board of Trustees by the NERC Steering Group ». En ligne 124 p. < http://www.nerc.com/docs/docs/blackout/NERC_Final_Blackout_Report_07_13_04.pdf >.
- NERC. 2014. *Improving Coordinated Operations Across The Electric Reliability Organization (ERO) Enterprise*. En ligne. 16 p. < http://www.nerc.com/AboutNERC/Documents/ERO_Enterprise_Operating_Model_February_2014.pdf >. Consulté le 02 octobre 2015.
- Ouellet, Estelle, Isabelle Belley-F. et Simon Leblond. 2005. *Guide économétrique pour STATA Pour ECN 3950 et FAS 3900*. Université de Montréal, 69 p.
- Pao, Y. H., et D. J. Sobajic. 1992. « Combined use of unsupervised and supervised learning for dynamic security assessment ». *Power Systems, IEEE Transactions on*, vol. 7, n° 2, p. 878-884.

- Parashar, M., et Mo Jianzhong. 2009. « Real Time Dynamics Monitoring System (RTDMS): Phasor Applications for the Control Room ». In *System Sciences, 2009. HICSS '09. 42nd Hawaii International Conference on.* (5-8 Jan. 2009), p. 1-11.
- Parashar, Manu, et Jianzhong Mo. 2009. « Real time dynamics monitoring system (RTDMS): phasor applications for the control room ». In *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on.* p. 1-11. IEEE.
- Patel, M. . 2011. *Eastern Interconnection Wide Area SynchroPhasor Angles Baseline Study*. En ligne. <<https://www.naspi.org/documents>>. Consulté le 11 octobre 2015.
- Patel, M., et al. 2010. *Real-Time Application of Synchrophasors for Improving Reliability*. En ligne. Coll. « NERC Special rep., Oct. 18, 2010 », 77 p. <<http://www.nerc.com/docs/oc/rapirtf/RAPIR%20final%20101710.pdf>>. Consulté le 11 octobre 2015.
- Patton, A. D. 1974. « Assessment of the security of operating electric power systems using probability methods ». *Proceedings of the IEEE*, vol. 62, n° 7, p. 892-901.
- Phadke, A. G. 2002. « Synchronized phasor measurements-a historical overview ». In *Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. IEEE/PES.* (6-10 Oct. 2002) Vol. 1, p. 476-479 vol.1.
- Phadke, A. G., J. S. Thorp et M. G. Adamiak. 1983. « A New Measurement Technique for Tracking Voltage Phasors, Local System Frequency, and Rate of Change of Frequency ». *Power Apparatus and Systems, IEEE Transactions on*, vol. PAS-102, n° 5, p. 1025-1038.
- Powertech-Labs. « DSA TOOLS: Dynamic Security Assessment Software ». En ligne. <<http://www.dsatools.com>>. Consulté le 10 octobre 2015.
- Québec, Régie de l'énergie. « Documentation administrative publique ». En ligne. <<http://www.regie-energie.qc.ca>>. Consulté le 05 octobre 2015.
- R. 2001. « The R Project for Statistical Computing ». En ligne. <<http://cran.r-project.org>>. Consulté le 05 octobre 2015.
- Rakotomalala, Ricco. 2011. « Pratique de la régression logistique ». *Régression Logistique Binaire et Polytomique*, Université Lumière Lyon, vol. 2.
- Reinhard, Karl E., Peter W. Sauer et Alejandro D. Dominguez-Garcia. 2013. « On Computing Power System Steady-State Stability Using Synchrophasor Data ». In *System Sciences (HICSS), 2013 46th Hawaii International Conference on.* (7-10 Jan. 2013), p. 2312-2318.

- Ribbens-Pavella, M., et F. J. Evans. 1985. « Direct methods for studying dynamics of large-scale electric power systems—A survey ». *Automatica*, vol. 21, n° 1, p. 1-21.
- Rovnyak, S., S. Kretsinger, J. Thorp et D. Brown. 1994. « Decision trees for real-time transient stability prediction ». *Power Systems, IEEE Transactions on*, vol. 9, n° 3, p. 1417-1426.
- Sabine, V. , et H. Mia. 2005. « LIBRA: A MATLAB library for robust analysis ». *Chemometr. Intell. Lab. Syst*, vol. 75, n° 2, p. 127-136.
- Samantaray, S. R., K. El-Arroudi, G. Joos et I. Kamwa. 2010. « A Fuzzy Rule-Based Approach for Islanding Detection in Distributed Generation ». *Power Delivery, IEEE Transactions on*, vol. 25, n° 3, p. 1427-1433.
- Samantaray, S. R., I. Kamwa et G. Joos. 2010. « Ensemble decision trees for phasor measurement unit-based wide-area security assessment in the operations time frame ». *Generation, Transmission & Distribution, IET*, vol. 4, n° 12, p. 1334-1348.
- Saporta, Gilbert. 2011. *Probabilités, analyse des données et statistique*, 3e éd. Paris: Editions Technip, 622 p.
- Sauer, P. W. 2011. « Computation of margins to powersystem loadability limits using phasor measurement unit data ». *Proc. 2011 IEEE PES General Meeting, Presentation at the DSAWG*.
- Sauer, P. W., et A. D. Dominguez-Garcia. 2014. « Data driven dynamic security assessment ». In *PES General Meeting, Conference & Exposition, 2014 IEEE*. (27-31 July 2014), p. 1-2.
- Sauer, P. W., K. Tomsovic, J. Dagle, S. Widergren, T. Nguyen et L. Schienbein. 2004. *Integrated Security Analysis, CERTS* 72 p.
- Sawhney, H., et B. Jeyasurya. 2004. « On-line transient stability assessment using artificial neural network ». In *Power Engineering, 2004. LESCOPE-04. 2004 Large Engineering systems Conference on*. (28-30 July 2004), p. 76-80.
- Shrestha, P. 2012. *Case Study: Disturbance event comparing PMU and SE voltage angle difference in ERCOT*. Power Point Presentation. En ligne, 16 p. < <http://www.ercot.com/search/results?q=Case+Study%3A+Disturbance+event+comparing+PMU+and+SE+voltage+angle+difference+in+ERCOT> >. Consulté le 11 octobre 2015.
- Silverstein, A., et J. E. Dagle. 2012. « Successes and Challenges for Synchrophasor Technology: An Update from the North American SynchroPhasor Initiative ». In

System Science (HICSS), 2012 45th Hawaii International Conference on. (4-7 Jan. 2012), p. 2091-2095.

Slaven, K., T. Brian et Z. Hongming. 2015. « Bridgen the Gap between Operation and Planning in WECC - a Step toward Dynamic Assessment of System Limits- ». *IEEE PES* p. 25.

Sobajic, D. J., et Y. H. Pao. 1989. « Artificial neural-net based dynamic security assessment for electric power systems ». *Power Systems, IEEE Transactions on*, vol. 4, n° 1, p. 220-228.

Stagg, Glen W. , et Ahmed H. El-Abiad. 1968. *Computer Methods in Power Systems Analysis*, 1st ed. Coll. « Electronic Systems ». Tokyo: McGraw Hill Kogakusha Ltd, 438 p.

Stedman, H. G. A. 1931. « Tees-side Sub-Centre: Chairman's address. Security And Reliability of Electricity supply ». *Electrical Engineers, Journal of the Institution of*, vol. 69, n° 410, p. 263-269.

Taylor, Carson W. 1994. *Power system voltage stability*. Université du Michigan, N.Y.: McGraw-Hill Ryerson, 273 p.

Thorp, J. S., A. G. Phadke et K. J. Karimi. 1985. « Real Time Voltage-Phasor Measurement For Static State Estimation ». *Power Apparatus and Systems, IEEE Transactions on*, vol. PAS-104, n° 11, p. 3098-3106.

Tiako, R., D. Jayaweera et S. Islam. 2011. « A case-based reasoning approach for dynamic security assessment of power systems with large penetration of wind power ». In *Universities Power Engineering Conference (AUPEC), 2011 21st Australasian*. (25-28 Sept. 2011), p. 1-6.

Trudel, G., Jean-Pierre Gingras et J. R. Pierre. 2005. « Designing a Reliable Power System: Hydro-Quebec's Integrated Approach ». *Proceedings of the IEEE*, vol. 93, n° 5, p. 907-917.

Tufféry, Stéphane. 2005. *Data mining et statistique décisionnelle: l'intelligence dans les bases de données*, 1ère ed. Paris: Editions Technip, 379 p.

Tufféry, Stéphane. 2010. *Data mining et statistique décisionnelle: l'intelligence des données*, 3ème ed. Paris: Editions Technip, 705 p.

Tufféry, Stéphane. 2015. *Modélisation prédictive et apprentissage statistique avec R*. Paris: Editions TECHNIP 420 p.

- Vaiman, M., M. Vaiman, S. Maslennikov, E. Litvinov et Luo Xiaochuan. 2010. « Calculation and Visualization of Power System Stability Margin Based on PMU Measurements ». In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*. (4-6 Oct. 2010), p. 31-36.
- Valette, A., J. A. Huang, S. Guillon, L. Loud, G. Vanier, F. Levesque, L. Riverin, J. C. Rizzi et F. Guillemette. 2008. « An integrated approach for optimizing dynamic transfer limits at Hydro-Québec ». In *Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*. (20-24 July 2008), p. 1-8.
- Valette, Alain, Huang Jin An, S. Guillon, L. Loud, G. Vanier, F. Levesque, L. Riverin, J. C. Rizzi et F. Guillemette. 2009. « An Integrated Approach for Optimizing Dynamic Transfer Limits at Hydro-Quebec ». *Power Systems, IEEE Transactions on*, vol. 24, n° 3, p. 1310-1317.
- Wang, L. 2012. *On-Line Dynamic Security Assessment: Its Role and Challenges for Smart Control Centers*. En ligne. Powertech Labs Inc., 14 p. <http://mydocs.epri.com/docs/PublicMeetingMaterials/1202/MVNQQ2YVLLT/Lei_Wang.pdf>. Consulté le 11 octobre 2015.
- Wang, L., et K. Morison. 2006. « Implementation of online security assessment ». *Power and Energy Magazine, IEEE*, vol. 4, n° 5, p. 46-59.
- Wehenkel, L., Th Van Cutsem et M. Ribbens-Pavella. 1989. « An Artificial Intelligence Framework for On-Line Transient Stability Assessment of Power Systems ». *Power Engineering Review, IEEE*, vol. 9, n° 5, p. 77-78.
- Weiying, Jiang, V. Vittal et G. T. Heydt. 2007. « A Distributed State Estimator Utilizing Synchronized Phasor Measurements ». *Power Systems, IEEE Transactions on*, vol. 22, n° 2, p. 563-571.
- Williams, Graham. 2011. *Data Mining with Rattle and R: The Art of Excavating Data for knowledge Discovery*. New york: Springer Science & Business Media, 374 p.
- Zhang, G., P. Hirsch et S. Lee. 2007. « Wide Area Frequency Visualization using Smart Client Technology ». In *Power Engineering Society General Meeting, 2007. IEEE*. (24-28 June 2007), p. 1-8.