



HAL
open science

Analyse statistique des pannes du réseau HTA

Patrick Garcia

► **To cite this version:**

Patrick Garcia. Analyse statistique des pannes du réseau HTA. Méthodologie [stat.ME]. 2012. dumas-00728956

HAL Id: dumas-00728956

<https://dumas.ccsd.cnrs.fr/dumas-00728956v1>

Submitted on 7 Sep 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Patrick GARCIA
Master de Mathématiques 2ème année,
Statistique et Applications.
Année scolaire 2011/2012.

RAPPORT DE STAGE : **Analyse statistique des pannes du réseau HTA.**



Responsables :
M. Daniel WAGNER, Groupe Électricité de Strasbourg.
M. Nicolas POULIN, Université de Strasbourg.
Du 1^{er} Février 2012 au 31 Juillet 2012.

Remerciements.

Je remercie tout d'abord le Groupe Électricité de Strasbourg pour m'avoir accueilli au sein de l'entreprise lors de mon stage de fin d'études.

Je remercie particulièrement mon maître de stage, Daniel WAGNER, pour son accueil, pour le projet qu'il m'a fourni, pour le temps qu'il m'a consacré et pour la confiance qu'il m'a accordée dès mon arrivée au sein de l'entreprise.

Je tiens à remercier également mon tuteur universitaire, M. Nicolas POULIN, pour m'avoir accueilli et conseillé lors du mi-parcours.

Merci également à Mme Myriam MAUMY-BERTRAND pour ses conseils sur les processus de Poisson et l'écriture du rapport.

Un grand merci à mes collègues de bureau, Éric RUHLMANN, Taha AFERHANE, Sébastien MULLER et Frédéric Krüger, pour m'avoir si bien intégré dans l'équipe, pour leur sympathie et leur coopération professionnelle tout au long de ce stage.

Merci à Bernard CROUX pour ses informations concernant la topologie du réseau de l'ÉS et merci d'avoir répondu à toutes mes questions.

Enfin je remercie l'ensemble du personnel du groupe Électricité de Strasbourg et les stagiaires et cadres d'OSI et ESSIDIS que j'ai rencontrés pour leur accueil, ainsi que toutes les personnes qui m'ont permis de passer un stage agréable et intéressant.

Table des matières

Remerciements.	1
Introduction.	5
1 L'entreprise Électricité de Strasbourg.	5
1.1 Historique de l'entreprise.	5
1.2 Organisation générale de l'entreprise.	6
1.3 Les réseaux électriques de l'Électricité de Strasbourg.	6
1.4 Les groupes d'exploitation.	7
2 Présentation du stage.	8
3 Diagramme de Gantt.	9
Analyse des incidents.	10
1 Les incidents du réseau électrique.	10
1.1 Trois types d'incidents.	10
1.2 L'application « Incidents ».	11
1.3 Le critère B	11
2 Statistiques descriptives sur les incidents HTA et HTB.	12
2.1 Les variables.	12
2.2 Représentations graphiques.	12
3 Modélisation des incidents HTA et HTB sur les quatre dernières années.	18
3.1 Les processus aléatoires.	18
3.2 Les processus de Poisson.	19
3.2.1 Processus de Poisson homogène.	19
3.2.2 Processus de Poisson non homogène.	20
3.2.3 Processus de Poisson composé.	20
3.3 Les tests d'adéquation.	20
3.4 Modélisation par processus de Poisson.	22
4 Simulation de pannes souterraines pour chaque départ HTA.	32
4.1 « Espérance » de panne.	32
4.1.1 Calcul des longueurs des câbles.	32
4.1.2 Taux d'incident au kilomètre.	33
4.1.3 Calcul des espérances de panne E_{panne} pour chaque cellule.	34
4.2 Critère B attendu par an.	34
4.2.1 Durées des incidents.	34
4.2.2 Nombre de clients par départ.	37
4.2.3 Calcul des contributions au critère B attendu.	37
4.3 Simulation de pannes souterraines.	37

4.4 Quelques représentations graphiques.	39
Conclusion et perspectives.	44
Annexes.	45
1 Interface de l'application Incidents.	45
2 Fiche d'incident réseau HTA.	46
3 Statistiques descriptives – Résultats de la console R.	48
4 Seuil critique $c_n = \sqrt{n} * d_n$ du test de Kolmogorov-Smirnov.	49
5 Programme R utilisé pour obtenir la figure 17, page 25.	49
6 Programme R utilisé pour obtenir les simulations des pannes souterraines.	50
7 Analyse Factorielle de Données Mixtes.	50
Références.	55

Table des figures

1	Organisation générale de l'ÉS	6
2	Les différents réseaux électriques	7
3	Groupes d'exploitation	7
4	Diagramme de Gantt	9
5	Diagramme circulaire : groupes d'exploitation	12
6	Diagramme circulaire : défauts à l'origine des incidents	13
7	Diagramme circulaire : causes des incidents	13
8	Diagramme simplifié des causes des incidents	14
9	Boîtes à moustaches du nombre de postes coupés	14
10	Figure 9 recentrée sur l'intérieur des boîtes à moustaches	14
11	Boîtes à moustaches du nombre de clients coupés	15
12	Boîtes à moustaches pour le critère B de l'incident	16
13	Matrice de nuages de points	17
14	Cumul des incidents de 2008 à 2011	22
15	Q-Q Plot des temps inter-arrivées	23
16	Ajustement à une loi exponentielle	24
17	Seuils critiques c_n par la méthode du bootstrap paramétrique	25
18	Histogramme de la distribution simulée de D	29
19	Histogramme de la distribution simulée de W^2	30
20	Simulations des incidents 2008 à 2011	31
21	Exemple du départ n° 15 du poste de Reichstett	33
22	Schéma illustrant la formule de d_1	36
23	Boîtes à moustaches du critère B sur 100 ans	40
24	Boîtes à moustaches du nombre d'incidents sur 100 ans	40
25	Distribution du critère B sur 100 ans	40
26	Distribution du nombre d'incidents sur 100 ans	40

Introduction.

1 L'entreprise Électricité de Strasbourg¹.

Électricité de Strasbourg SA est une société anonyme au capital de 71 543 860 €, cotée à la Bourse sur Euronext Paris qui a dégagée un chiffre d'affaires de 562,1 M€ et un résultat net consolidé de 60,1 M€. Le groupe Électricité de Strasbourg est composé de plus de 1 000 salariés répartis dans les différentes sociétés et filiales.

L'actionnaire majoritaire d'Électricité de Strasbourg est EDF Développement et Environnement (EDEV) qui détient 88,82% du capital. Le groupe ÉS est composé d'Électricité de Strasbourg (maison mère), d'ÉS Réseaux (Gestionnaire de réseaux) et de plusieurs filiales (ÉS Énergies Strasbourg, Écotral...).

1.1 Historique de l'entreprise.

Électricité de Strasbourg (ÉS) a été créée le 14 décembre 1899 sous le nom de « Elektrizitäts-werk Strassburg », à l'initiative de la ville de Strasbourg et grâce aux capitaux de l'entreprise allemande AEG et du Crédit Suisse, à travers sa filiale Elektrobank. AEG fabriquait alors des équipements électrotechniques, tandis que Elektrobank finançait la construction et l'exploitation d'entreprises d'électricité.

En 1908, la ville de Strasbourg devient actionnaire majoritaire. En ce début de siècle, Électricité de Strasbourg s'engage dans les grands programmes d'électrification des campagnes et des villes.

Après une courte période de transition à l'époque de la première guerre mondiale, l'entreprise reprend son expansion. Elle crée une société immobilière qui, en 1925, deviendra Sofidal, sa plus ancienne filiale.

En 1954, la ville de Strasbourg cède ses actions à Électricité de France (EDF) qui devient ainsi l'actionnaire majoritaire de l'entreprise.

Dans les années 70, années de crises pétrolières, l'heure est aux économies d'énergie en France. Électricité de Strasbourg crée une filiale qui deviendra « Écotral » chargée de promouvoir des installations de chauffage électrique économiques, avec garantie de résultat.

1^{er} juillet 2004 : cette date marque une étape décisive dans le processus d'ouverture du marché de l'énergie avec l'éligibilité de tous les clients, ménages exceptés. À partir du 1^{er} octobre 2005, ÉS élargit son activité et devient également fournisseur de gaz. En 2007, c'est l'ouverture totale du marché de l'électricité et du gaz en France.

En 2009, Électricité de Strasbourg, en lien avec les obligations juridiques qui encadrent l'ouverture des marchés européens de l'énergie, a procédé à la filialisation de son commercialisateur « ÉS Énergies Strasbourg ». Elle crée également la marque du groupe « Groupe ÉS » et la marque du distributeur « ESR » pour ÉS Réseaux.

En avril 2012, Énerest, fournisseur historique de gaz de Strasbourg et ex-filiale de Réseau GDS, devient une filiale du groupe ÉS.

1. Source : <http://www.es-groupe.fr/>

1.2 Organisation générale de l'entreprise.

Le groupe ÉS est organisé autour de trois activités principales :

- la gestion du réseau de distribution électrique assurée par *ÉS Réseaux*.
- la fourniture d'énergies et de services liés assurée par *ÉS Énergies Strasbourg*.
- le conseil, l'assistance et le service pour tout projet de construction ou de rénovation assurés par *Écotral*.



FIGURE 1 – Organisation générale de l'ÉS

1.3 Les réseaux électriques de l'Électricité de Strasbourg.

Le réseau de l'Électricité de Strasbourg est divisé en quatre parties :

- le réseau *EDF-RTE* : c'est le réseau électrique appartenant à EDF. C'est un réseau maillé qui transporte du courant de 400 000 volts et 225 000 volts. Actuellement, le réseau HTB (voir plus bas) de l'ÉS est alimenté en cinq points par le réseau RTE à 225 000 volts.
- le réseau *HTB* : c'est également un réseau maillé dont les tensions électriques valent plus de 50 000 volts (en général, 63 000 volts). Le réseau est composé de postes HTB (ou postes source) qui sont divisés en plusieurs cellules (ou départs HTA). Chaque cellule alimente une partie du réseau « haute tension ». Le réseau électrique de l'ÉS est composé de 40 postes HTB.
- le réseau *HTA* : c'est la haute tension de la distribution publique appelée haute tension de type A (HTA) ou moyenne tension (MT). Cette tension vaut généralement 20 000 volts. Il s'agit ici d'un réseau arborescent qui se délimite par des postes de transformation HTA/BT ou des postes de transformation HTA/BT client. Le réseau de l'ÉS est composé de 7 000 postes HTA/BT.

- le réseau *BT* : c'est le réseau basse tension (230 et 400 volts) qui est aussi un réseau arborescent. La racine de l'arbre étant, comme il a été dit précédemment, soit un poste de transformation HTA/BT, soit un poste de transformation HTA/BT client. Dans le premier cas, le poste alimente des particuliers (également appelés clients basse tension). Les particuliers sont reliés au réseau BT par des coffrets. En moyenne, un poste HTA/BT alimente 100 clients basse tension. Dans le second cas, le poste n'est relié qu'à un seul client qui est appelé un client HTA. Les clients HTA sont généralement des entreprises.

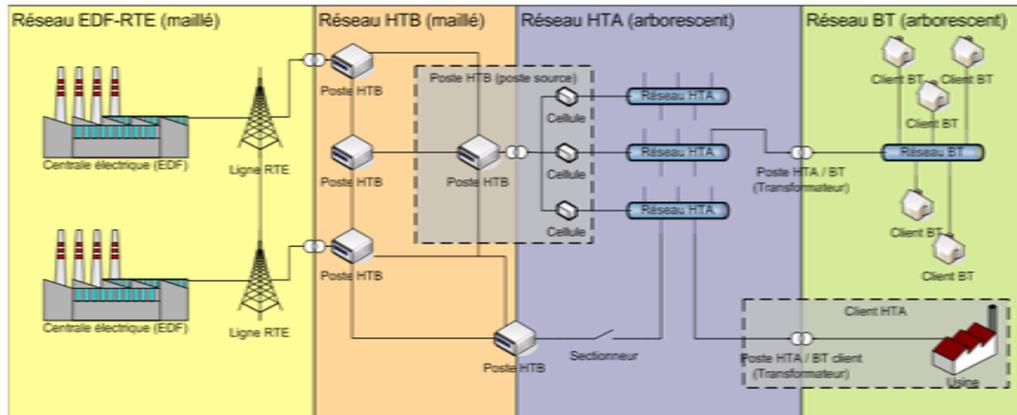


FIGURE 2 – Les différents réseaux électriques

En HTB, les réseaux sont dits bouclés, c'est-à-dire que s'il y a un incident sur une partie de la ligne, le reste du réseau n'est pas perturbé. En revanche, en BT, les réseaux sont en antennes, c'est-à-dire qu'un problème sur une ligne entraîne des perturbations pour tous les clients qui sont en aval. En HTA, les réseaux sont en antennes dites bouclables.

1.4 Les groupes d'exploitation.

Électricité de Strasbourg Réseaux alimente la majeure partie du Bas-Rhin, et son réseau est divisé en trois groupes d'exploitation : le GEC, Groupe d'Exploitation Centre ; le GEN, Groupe d'Exploitation Nord ; et le GES, Groupe d'Exploitation Sud. La répartition est la suivante :



FIGURE 3 – Groupes d'exploitation

2 Présentation du stage.

Mon stage s'est déroulé au sein du département Expertise et Soutien au Système d'Information du Distributeur (ESSIDIS), entité de l'ESR, qui se situe au Centre Opérationnel de Mundolsheim (COM).

Dans un premier temps, nous avons analysé les interventions effectuées par l'ÉS depuis 2008. Une intervention peut être une intervention technique (relevé spécial, vérification, impayé/coupure...) ou une intervention contrat (cessation, souscription, modification de contrat...). Cette partie, que nous ne présentons pas, a duré un mois et m'a permis d'améliorer mes compétences sur le logiciel de statistique R, notamment en utilisant de gros jeux de données ainsi qu'en réalisant des statistiques descriptives et des représentations graphiques.

Dans un second temps, nous avons analysé les incidents du réseau de l'ÉS. L'objectif principal de ce stage a consisté à analyser les incidents HTA et HTB.

Nous présentons en premier lieu certaines statistiques descriptives sur les incidents de 2008 à 2011. 609 incidents HTA ou HTB sont survenus ces quatre dernières années, il n'est donc pas aisé de déceler si des choses particulières se dégagent des données. C'est en ce sens que les statistiques descriptives sont pertinentes.

Ensuite, nous nous sommes intéressés aux occurrences des pannes sur le réseau électrique qui sont considérées comme des événements aléatoires car nous ne pouvons pas les prévoir. Nous avons réalisé des tests statistiques afin de savoir s'il est possible de modéliser les incidents des quatre dernières années en utilisant des processus aléatoires et plus particulièrement des processus de Poisson car ces processus sont souvent bien adaptés pour expliquer des processus « d'arrivées ».

Puis, nous nous sommes plus particulièrement intéressés aux incidents sur les câbles souterrains. À partir de la topologie du réseau, nous avons calculé les longueurs des câbles souterrains et nous avons pu estimer une espérance des pannes pour chaque départ HTA ².

Ensuite, nous avons cherché les distances que doivent parcourir les exploitants pour isoler le tronçon en panne lorsqu'un incident a lieu sur telle ou telle cellule d'un poste source, et de ce fait nous avons évalué les durées des incidents pour chacune de ces cellules.

Et ainsi nous avons pu calculer une espérance du critère B qui correspond au temps moyen de coupure d'un client par an.

Puis, après avoir calculé le critère B attendu par an, nous avons cherché à simuler des pannes pour chaque départ HTA pendant 100 ans. De cette façon nous avons pu calculer le critère B pour 100 ans d'affilée, observer la dispersion, puis comparer la moyenne au critère B attendu que nous avons calculé.

Enfin nous avons réalisé une Analyse Factorielle des Données Mixtes (AFDM) qui permet l'analyse factorielle simultanée de variables quantitatives et qualitatives sur les incidents. Cette analyse est présentée en annexe 7, page 50.

Tout au long de mon stage, j'ai utilisé le tableur Excel et le logiciel de statistique R.

2. Un départ HTA (ou cellule) alimente le réseau HTA à partir d'un poste source HTB (cf. section 1.3).

3 Diagramme de Gantt.

Le diagramme de Gantt est un outil qui permet de visualiser dans le temps les différentes tâches qui composent un projet. C'est un outil très utilisé en gestion de projet car il permet de représenter graphiquement l'avancement d'un projet.

Voici le diagramme de Gantt qui correspond au déroulement de mon stage :

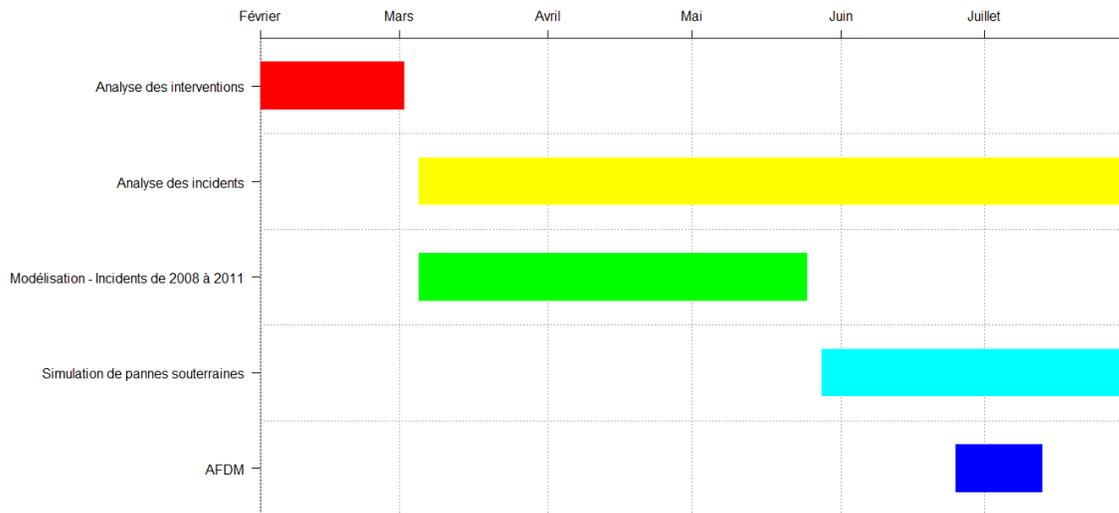


FIGURE 4 – Diagramme de Gantt

Analyse des incidents.

1 Les incidents du réseau électrique.

Rappelons brièvement que le réseau électrique se compose de l'ensemble des câbles électriques connectés entre eux. Il existe différents types de réseaux :

- le réseau Haute Tension (ou réseau HT) qui est utilisé pour le transport et la répartition de l'électricité. Il se décompose en deux sous-réseaux :
 - le réseau Haute Tension B (ou réseau HTB) : tensions supérieures à 50 kV (Kilovolts). Il sert à l'alimentation générale, du niveau national au niveau régional.
 - le réseau Haute Tension A (ou réseau HTA) : tensions variant de 1 000 V à 50 kV. C'est le réseau de distribution local, dans un rayon de 10 à 20 kilomètres autour du poste source.
- le réseau Basse Tension (ou réseau BT) : tensions inférieures à 1 000 V. Il est utilisé pour desservir les quartiers et les communes.

1.1 Trois types d'incidents.

Il existe trois types d'incidents différents qui peuvent survenir sur le réseau électrique :

- un *incident BT* ou incident « basse tension » est un incident dans une petite zone géographique. Il touche généralement quelques rues, ce qui représente quelques dizaines jusqu'à une centaine de clients. Un incident BT peut être causé par la coupure d'un poste HTA/BT ou être lié à un problème dans l'arborescence du réseau BT (arrachage de la ligne électrique par la chute d'une branche par exemple).
- un *incident HTA* est un incident de grande ampleur. Il s'agit en général de la coupure d'un départ HTA (une cellule). Ce départ étant la racine d'un réseau HTA, cette coupure entraîne la coupure de plusieurs postes HTA/BT et donc une coupure chez les clients BT raccordés à chacun des postes. Un tel incident affecte quelques milliers de clients.
- un *incident HTB* est le pire des incidents pour l'ÉS. En effet, un tel incident entraîne la coupure de plusieurs cellules, ce qui entraîne la coupure de plusieurs dizaines de postes HTA/BT. Il s'agit donc d'un incident de très grande ampleur. Le réseau HTA étant protégé par des disjoncteurs dans le poste source (les cellules), en cas de défaut, c'est tout le départ qui est coupé soit plusieurs dizaines de milliers de clients.

Dans la suite, le travail portera sur les incidents HTA et HTB.

1.2 L'application « Incidents »³.

L'application Incidents a été développée par le département Exploitation du groupe ÉS et permet la gestion des incidents électriques. Le but principal est d'avoir un historique sur les incidents qui se sont produits sur le réseau de l'ÉS.

Il est possible également d'obtenir des statistiques et des détails sur chaque incident. Tous les utilisateurs de l'application ont ainsi accès à toutes les informations possibles sur un incident. L'historique des incidents est construit soit grâce au journal de bord du système de télé-conduite utilisé par l'ÉS pour contrôler à distance le réseau Haute Tension, soit grâce à des fiches saisies par les agents de l'ÉS. Une fiche d'incident HTA est présentée en annexe 2, page 46.

Extraction

L'application Incidents permet directement l'extraction des données sur les incidents dans un fichier CSV.

1.3 Le critère B

Le critère B est un marqueur très important pour l'ESR. Il représente la durée moyenne annuelle de coupure par client desservi.

Grâce à l'application Incidents, le calcul du critère B sera plus précis du fait de la précision des données. C'est une métrique universelle de la qualité des réseaux électriques. À l'ESR, le critère B varie autour de 10 minutes par année alors qu'en Europe, selon les réseaux, le critère B peut être de plusieurs heures.

Il se calcule de la façon suivante :

$$\text{Critère B} = \sum_{\text{Incidents } i \text{ de l'année}} \left(\frac{\text{Durée de l'incident } i * \text{Nb de clients coupés dans l'incident } i}{\text{Nombre total de client ÉS}} \right)$$

3. Une interface de l'application Incidents est donnée en annexe 1, page 45.

2 Statistiques descriptives sur les incidents HTA et HTB.

2.1 Les variables.

Les variables concernant les incidents sont nombreuses, nous avons décidé d'en retenir 10 :

- 6 variables *qualitatives* : type de l'incident, secteur où a eu lieu la coupure, nombre de clients de la commune coupée, nom du poste source, défaut de l'incident et cause de l'incident.
- 4 variables *quantitatives* : nombre de postes coupés, nombre de clients coupés, critère B de l'incident et puissance coupée de l'incident.

2.2 Représentations graphiques.

Il y a 609 incidents HTA ou HTB qui sont survenus de 2008 à 2011. Les statistiques descriptives vont permettre de nous renseigner sur chacune des variables étudiées. Les résultats de la console R sont donnés en annexe 3, page 48. Nous présentons maintenant les représentations graphiques pour quelques variables.

Les variables qualitatives.

Les représentations graphiques liées aux variables qualitatives sont la représentation en diagramme circulaire et la représentation en bâtons. Dans ce manuscrit, nous représentons les variables qualitatives sous forme de diagrammes circulaires car ils sont bien adaptés pour des vues d'ensemble.

Secteur où a eu lieu la coupure.

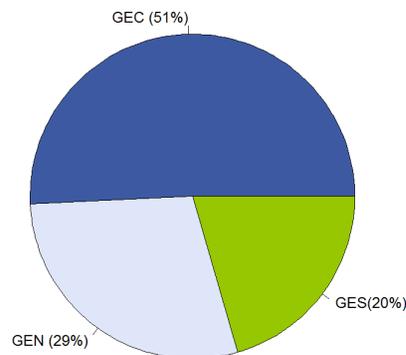


FIGURE 5 – Diagramme circulaire : groupes d'exploitation

Plus de la moitié des incidents ont eu lieu dans le GEC. C'est dans ce secteur que se trouve la ville de Strasbourg, qui regroupe près de 20% des postes de l'ESR mais aussi 50% des clients.

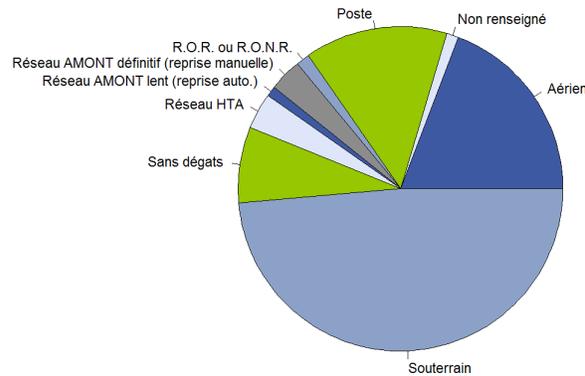
Défaut de l'incident.

FIGURE 6 – Diagramme circulaire : défauts à l'origine des incidents

Quasiment la moitié des défauts sont souterrains. Environ $1/5^e$ sont des défauts aériens. Les défauts sur câbles sont bien plus fréquents que les défauts sur postes.

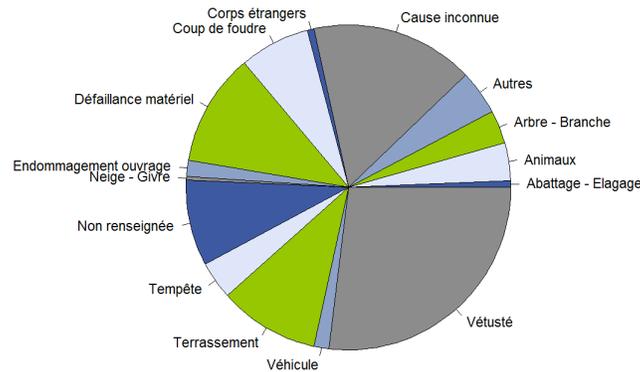
Cause de l'incident.

FIGURE 7 – Diagramme circulaire : causes des incidents

La principale cause des incidents est la vétusté des câbles. Nous pouvons remarquer également que les causes naturelles (animaux, coup de foudre, tempête. . .) sont responsables d'un grand nombre d'incidents. Ces causes sont des causes inévitables pour l'ESR.

Représentation simplifiée.

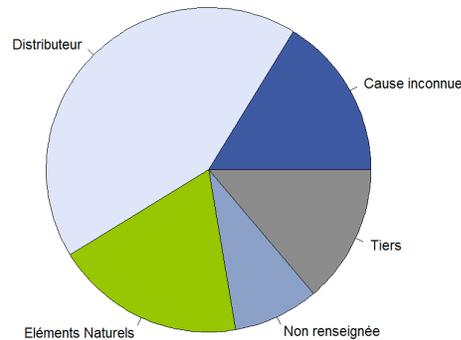


FIGURE 8 – Diagramme simplifié des causes des incidents

La représentation simplifiée, où la variable cause de l'incident à cinq modalités, nous montre que le distributeur a une part de responsabilité dans plus de 40% des incidents. Nous remarquons également que près de 20% des incidents sont dus à des causes naturelles et environ 15% sont de causes inconnues.

Les variables quantitatives.

Une des représentations adéquates pour les variables quantitatives est l'histogramme. Une autre représentation intéressante est la boîte à moustaches. C'est cette représentation que nous avons retenue car elle permet de comparer facilement une même variable en fonction d'un paramètre. Par exemple, nous pouvons comparer les variables en fonction des différentes classes de communes. Nous avons donc en abscisse la taille des communes concernées par l'incident.

Nombre de postes coupés.

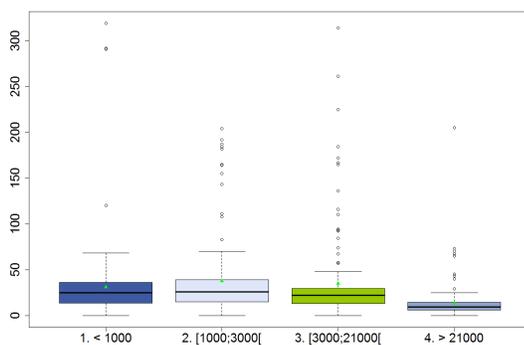


FIGURE 9 – Boîtes à moustaches du nombre de postes coupés

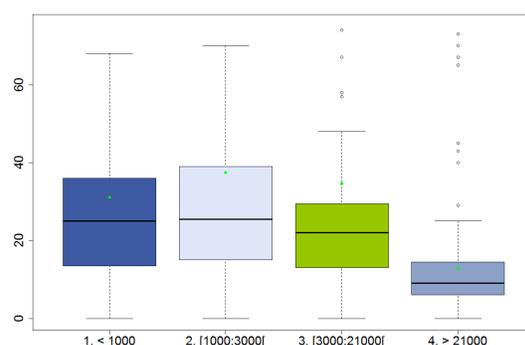


FIGURE 10 – Figure 9 recentrée sur l'intérieur des boîtes à moustaches

Nous distinguons un certain nombre de valeurs extrêmes pour chaque classe sur la figure 9, page 14. Ainsi, dans les quatre groupes de communes, il s'est produit des incidents qui ont entraîné la coupure de nombreux postes HTA/BT. Ceci est dû aux incidents de type HTB.

L'échelle de la figure 9 n'est pas très agréable car les boîtes à moustaches sont concentrées vers le bas. Nous allons donc recentrer le graphique sur l'intérieur des boîtes à moustaches pour obtenir un graphique que nous pourrions mieux étudier. Nous obtenons ainsi la figure 10, page 14.

Sur cette figure 10, nous constatons que les médianes sont plus faibles pour les grandes communes que pour les petites communes. Ceci peut s'expliquer par le fait que les petites communes sont alimentées par des départs très longs avec beaucoup de postes alors qu'à l'inverse les grandes communes ont moins de postes par départ car l'habitat est plus dense et donc il y a plus de clients par poste. La moyenne de chaque groupe est assez éloignée de la médiane, ceci s'explique par le fait que les valeurs extrêmes tirent vers le haut la moyenne qui est moins robuste que la médiane.

Nombre de clients coupés.

Comme précédemment, les boîtes à moustaches sont concentrées vers le bas avec des valeurs extrêmes pour chaque groupe. Il y a donc eu dans chaque classe de communes des incidents qui ont entraîné la coupure de nombreux clients. Cela est dû aux incidents de type HTB. Nous présentons directement le graphique recentré sur les boîtes à moustaches à la figure 11.

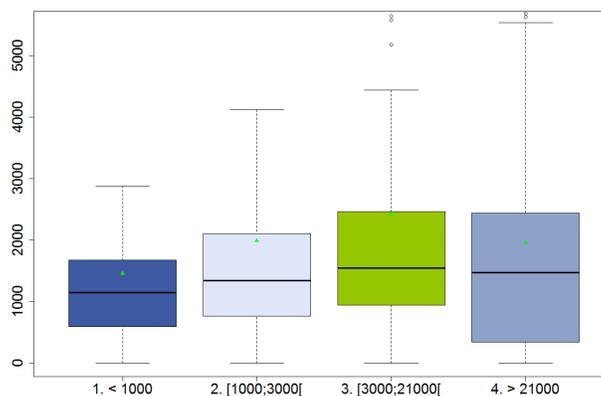


FIGURE 11 – Boîtes à moustaches du nombre de clients coupés

Nous remarquons que le nombre de clients coupés est plus élevé dans les grandes communes. Il semble donc qu'il y a moins de postes coupés mais plus de clients coupés dans les grandes communes. Comme nous l'avons dit plus haut, il y a moins de clients sur les départs des petites communes, donc un incident va affecter moins de clients dans les petites communes que dans les grandes communes.

Critère B de l'incident.

Le critère B d'un incident est la contribution en secondes qu'apporte un incident sur le critère B à l'année. Nous présentons, là encore, directement le graphique centré sur les boîtes à moustaches à la figure 12.

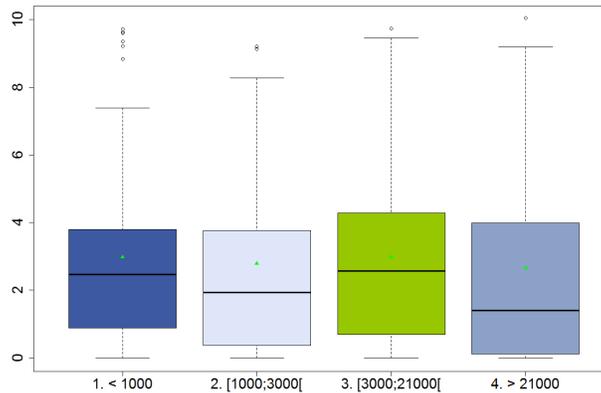


FIGURE 12 – Boîtes à moustaches pour le critère B de l'incident

Nous remarquons que le critère B moyen d'un incident est le même pour chaque groupe de communes et il est environ égal à trois secondes. Nous observons également que la dispersion autour de la moyenne pour chacun des groupes est assez identique.

Il n'y a donc pas de différence au niveau de ce critère entre les classes de communes contrairement aux deux variables étudiées précédemment (nombre de postes coupés et nombre de clients coupés) où nous avons observé des différences entre les petites et les grandes communes.

Liens entre les variables quantitatives.

Nous pouvons établir graphiquement des liens entre les variables quantitatives en utilisant une matrice de nuages de points. La fonction *scatterplotMatrix* du logiciel R permet de réaliser ces nuages de points. De plus, nous pouvons y ajouter les droites des moindres carrés (en vert), les courbes de lissage (en rouge) et également les histogrammes sur la diagonale.

Nous observons, figure 13, page 17, une relation entre le nombre de postes coupés et le nombre de clients coupés, selon deux groupes. De manière générale, un nombre important de clients coupés correspond à un grand nombre de postes coupés. Une explication serait que les deux groupes que nous observons diffèrent suivant la taille des communes. Comme nous l'avons vu dans l'étude des boîtes à moustaches, dans les grandes communes il y a plus de clients par postes. Pour un même nombre de postes coupés, le nombre de clients coupés est plus important dans les grandes communes que dans les petites communes.

Il semble également qu'il y ait une liaison entre la puissance coupée et le nombre de postes coupés et aussi une liaison entre la puissance coupée et le nombre de clients coupés. C'est logique, plus il y a de postes coupés ou de clients coupés, plus la puissance coupée de l'incident est élevée.

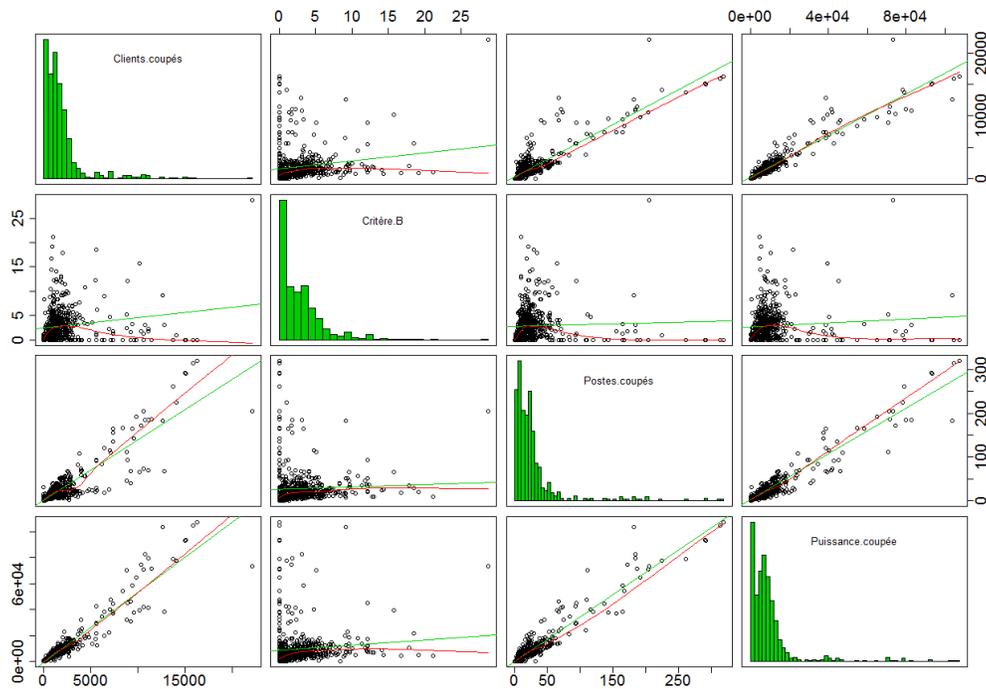


FIGURE 13 – Matrice de nuages de points

Nous pouvons afficher également la matrice des corrélations qui donne les coefficients de corrélation de Bravais-Pearson entre les variables quantitatives grâce à la fonction *cor* de R :

$$\begin{array}{l}
 \textit{clients coupés} \\
 \textit{critère B} \\
 \textit{postes coupés} \\
 \textit{puissance coupée}
 \end{array}
 \begin{pmatrix}
 \textit{clients coupés} & 1.0000000 & 0.1627597 & 0.8715437 & 0.9416155 \\
 \textit{critère B} & 0.1627597 & 1.0000000 & 0.0448003 & 0.0892804 \\
 \textit{postes coupés} & 0.8715437 & 0.0448003 & 1.0000000 & 0.9449067 \\
 \textit{puissance coupée} & 0.9416155 & 0.0892804 & 0.9449067 & 1.0000000
 \end{pmatrix}$$

Nous obtenons les mêmes conclusions qu'avec la matrice de nuages de points : il existe une liaison linéaire entre le nombre de postes coupés et le nombre de clients coupés, entre le nombre postes coupés et la puissance coupée et également entre le nombre de clients coupés et la puissance coupée.

Néanmoins, l'étude graphique nous a permis en plus de distinguer deux groupes dans la relation linéaire entre le nombre de postes coupés et le nombre de clients coupés.

3 Modélisation⁴ des incidents HTA et HTB sur les quatre dernières années.

Les occurrences des pannes sur le réseau électrique sont considérées comme des événements aléatoires car nous ne pouvons pas les prévoir. Nous supposons alors que nous pouvons modéliser les incidents HTA/HTB par un processus aléatoire. Dans la suite, nous réaliserons des tests d'adéquation, notamment les tests de Kolmogorov-Smirnov et de Cramer-Von Mises, afin de savoir s'il est possible de modéliser les incidents du réseau de l'ÉS par des processus de Poisson qui sont souvent utilisés pour expliquer des processus « d'arrivées ».

3.1 Les processus aléatoires.

Les processus aléatoires (ou stochastiques) sont des processus qui décrivent l'évolution d'une variable aléatoire en fonction du temps. Introduisons des processus qui vont nous servir dans la suite de l'étude :

Soit $(T_n)_{n \geq 1}$ le *processus des temps d'occurrences* qui représente les instants d'occurrences des incidents. Par convention, $T_0 = 0$. À partir de ce processus $(T_n)_{n \geq 1}$, il est possible de définir le *processus de comptage*, noté $N(t)$, qui représente le nombre d'incidents qui se sont produits entre 0 et t :

$$N(t) = \sum_{i=1}^n \mathbf{1}(T_i \leq t), \forall t \geq 0$$

Soit $(X_n)_{n \geq 1}$ le *processus des temps inter-incidents* qui correspond au temps d'attente entre deux incidents :

$$X_n = T_n - T_{n-1}, \forall n \geq 1$$

Introduisons quelques définitions :

Définition 1 *Un processus est dit à accroissements indépendants si $\forall 0 < t_1 < t_2 < \dots < t_n$:*

$$N(t_1), N(t_2) - N(t_1), \dots, N(t_n) - N(t_{n-1}) \text{ sont indépendants.}$$

Définition 2 *Un processus est dit à accroissements stationnaires si $\forall 0 \leq s < t$:*

$$N(t) - N(s) \text{ et } N(t - s) \text{ sont identiquement distribués.}$$

Définition 3 *L'intensité $\lambda(t)$ d'un processus $N(t)$ se définit de la façon suivante :*

$$\lambda(t) = \lim_{h \rightarrow 0} \frac{\mathbb{P}[N(t+h) - N(t) \geq 1]}{h}$$

L'intensité d'un processus peut s'interpréter comme le taux d'occurrence d'un incident.

Définition 4 *L'intensité cumulée $\Lambda(t)$ d'un processus $N(t)$ est le nombre moyen d'incidents survenus entre 0 et t :*

$$\Lambda(t) = \int_0^t \lambda(v) dv = \mathbb{E}[N(t)]$$

4. Pour cette analyse, nous nous sommes appuyés sur [1] et [2] référencés page 55.

3.2 Les processus de Poisson.

Les processus de Poisson sont bien adaptés pour expliquer des processus « d'arrivées ». Des exemples de ces processus sont larges : appels téléphoniques à un standard, arrivée d'un client à un guichet, sinistres subis par une compagnie d'assurance, panne sur une machine. . .

Les processus de Poisson que j'ai étudiés sont des processus « temporels », mais il est bon de savoir qu'il existe des processus de Poisson dans d'autres espaces. Les processus de Poisson temporels se divisent en trois types : les processus de Poisson homogènes, les processus de Poisson non homogènes et les processus composés.

3.2.1 Processus de Poisson homogène.

Définition 5 *Un processus de comptage $N(t)$ est un processus de Poisson homogène d'intensité $\lambda > 0$ si :*

1. $N(0) = 0$;
2. *Le processus est à accroissements indépendants ;*
3. *Le processus est à accroissements stationnaires ;*
4. $\forall 0 \leq s < t$, $N(t) - N(s)$ *suit une loi de Poisson de paramètre $\lambda(t - s)$:*

$$\forall k \in \mathbb{N}, \mathbb{P}[N(t) - N(s) = k] = e^{-\lambda(t-s)} \frac{(\lambda(t-s))^k}{k!}$$

Une définition équivalente et qui est plus répandue est la suivante :

Définition 6 *Un processus de comptage $N(t)$ est un processus de Poisson homogène d'intensité $\lambda > 0$ si :*

1. $N(0) = 0$;
2. *Le processus est à accroissements indépendants ;*
3. *Le processus est à accroissements stationnaires ;*
- 4.

$$\mathbb{P}[N(t+h) - N(t) = k] = \mathbb{P}[N(h) = k] = \begin{cases} 1 - \lambda h + o(h) & \text{si } k = 0 \\ \lambda h + o(h) & \text{si } k = 1 \\ o(h) & \text{si } k \geq 2 \end{cases}$$

Cette définition nous montre que plusieurs occurrences ne peuvent avoir lieu en même temps, ce qui peut poser problème dans certains cas. Dans la suite, nous pourrons lever cette hypothèse grâce aux processus composés.

La proposition suivante est très utile pour modéliser un processus de Poisson homogène :

Proposition 1 *Un processus de comptage est un processus de Poisson homogène si et seulement si les $X_n = T_n - T_{n-1}$ sont indépendants et identiquement distribués selon une loi exponentielle de paramètre $\lambda > 0$, c'est-à-dire si les X_n ont pour densité de probabilité :*

$$f(x) = \begin{cases} 0 & \forall x < 0 \\ \lambda e^{-\lambda x} & \forall x \geq 0 \end{cases}$$

3.2.2 Processus de Poisson non homogène.

Un processus de Poisson non homogène diffère d'un processus de Poisson homogène par le fait que ses accroissements ne sont plus stationnaires. L'intensité d'un tel processus n'est donc plus constante mais dépend du temps.

Définition 7 *Un processus de comptage $N(t)$ est un processus de Poisson non homogène d'intensité $\lambda(t)$ s'il satisfait les conditions suivantes :*

1. $N(0) = 0$;
2. Le processus est à accroissements indépendants ;
3. $N(t)$ suit une loi de Poisson de paramètre $\Lambda(t)$:

$$\forall k \in \mathbb{N}, \mathbb{P}[N(t) = k] = e^{-\Lambda(t)} \frac{(\Lambda(t))^k}{k!}$$

où $\Lambda(t)$ est une fonction déterministe.

Dans ce cas, les temps inter-incidents X_n ne sont plus identiquement distribués.

3.2.3 Processus de Poisson composé.

Les processus de Poisson composés permettent de lever l'hypothèse selon laquelle plusieurs événements ne peuvent se produire en même temps.

Définition 8 *Un processus de comptage $Z(t)$ est un processus de Poisson composé si :*

$\forall t \geq 0, Z(t) = \sum_{n=0}^{N(t)} Y_n$ où $N(t)$ est un processus de Poisson sous-jacent, $(Y_n)_{n \geq 1}$ une famille de variables aléatoires indépendantes et identiquement distribuées et $Y_0 = 0$. De plus, $N(t)$ et $(Y_n)_{n \geq 1}$ sont supposés indépendants.

3.3 Les tests d'adéquation ⁵.

Les tests d'adéquation sont des tests d'hypothèse statistique qui permettent de déterminer s'il y a adéquation ou non entre un échantillon de données et une distribution connue.

Le test du Chi-2 est certainement un des tests d'adéquation les plus connus. Par nature, ce test s'applique plus généralement à des variables discrètes. Mais les variables continues peuvent être catégorisées et donc le test peut également s'appliquer aux distributions continues. Néanmoins, la statistique de test va dépendre de la répartition choisie.

Il y a également les tests d'adéquation basés sur la fonction de répartition empirique. La fonction de répartition empirique est basée sur l'échantillon X_1, \dots, X_n et est définie par :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i \leq x), \forall x \in \mathbb{R}$$

5. Source : http://www.aiaccess.net/French/Glossaires/GlosMod/f_gm_adequation.htm

Ces tests vérifient que la fonction de répartition empirique $F_n(x)$ est une bonne approximation de la fonction de répartition théorique $F(x)$. Ils testent alors :

$$H_0 : F_n = F \text{ contre } H_1 : F_n \neq F$$

où F_n est la fonction de répartition empirique de l'échantillon et F est la fonction de répartition de la distribution de référence.

Parmi ces tests, nous pouvons citer le test de Kolmogorov-Smirnov et celui de Cramér-Von Mises dont nous présentons les statistiques de test.

Test de Kolmogorov-Smirnov.

Le test de Kolmogorov-Smirnov s'applique aux distributions continues et dont la loi théorique est entièrement spécifiée, c'est-à-dire sans paramètres inconnus. Si les paramètres sont estimés à partir des données, la région critique de la statistique de test n'est pas valable et doit être déterminée par simulation.

La statistique de test D_n mesure l'écart entre les valeurs observées (échantillon) et les valeurs théoriques (distribution de référence) en prenant la plus grande valeur absolue de la différence entre la fonction de répartition empirique et la fonction de répartition théorique :

$$D_n = \sup_{x \in \mathbb{R}} | F_n(x) - F(x) |$$

La distribution de D_n sous H_0 n'est pas connue mais elle ne dépend pas de $F(x)$. Elle a été abondamment simulée et tabulée.

Test de Cramer-Von Mises.

Si le test de Kolmogorov-Smirnov est raisonnablement puissant, il peut paraître surprenant qu'il soit conçu uniquement sur l'écart entre une observation de l'échantillon et une valeur de la fonction de répartition théorique.

Le test de Cramér-Von Mises va plutôt mesurer la différence entre les deux fonctions de répartition en les comparant sur l'ensemble de leur domaine. La statistique de test est la suivante :

$$W^2 = \frac{1}{12n} + \sum_{i=1}^n \left(\frac{2i-1}{2n} - F(x_{(i)}) \right)^2$$

où $x_{(i)}$ est la statistique d'ordre de rang i de l'échantillon.

La distribution de W^2 est inconnue mais elle est identique pour toutes les fonctions de répartition $F(x)$. Elle a été abondamment simulée et tabulée.

3.4 Modélisation par processus de Poisson.

Nous avons à notre disposition les dates d'arrivées des incidents HTA/HTB depuis le 1^{er} janvier 2008 jusqu'au 31 décembre 2011. Nous pouvons donc définir la fonction de comptage du processus du cumul des incidents par :

$$N(t) = \sum_{i=1}^n \mathbf{1}(T_i \leq t), \forall t \geq 0$$

où les T_i correspondent aux instants d'occurrences des incidents.

La figure 14 représente la trajectoire observée du cumul des incidents entre 2008 et 2011 :

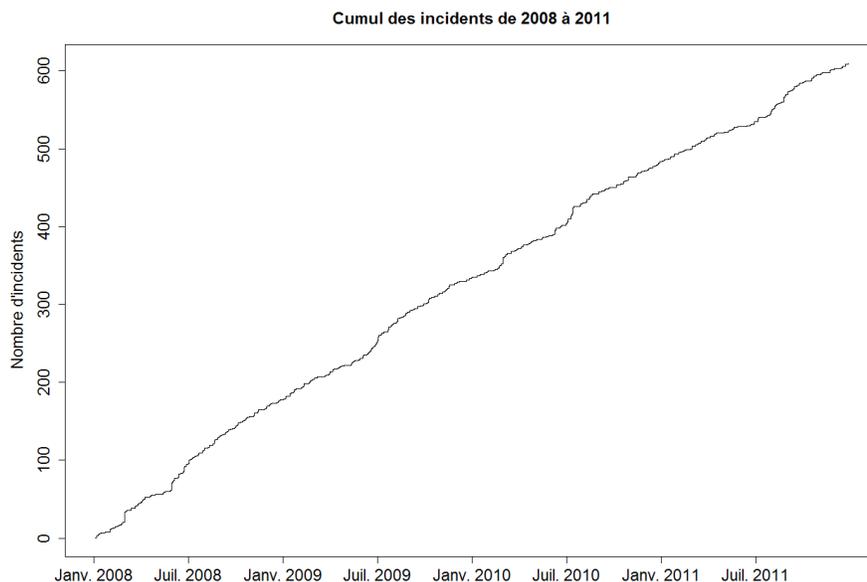


FIGURE 14 – Cumul des incidents de 2008 à 2011

Nous pouvons observer qu'il y a eu 609 incidents HTA/HTB lors des quatre dernières années et nous constatons une légère accélération du cumul lors de chaque été.

Nous remarquons également que plusieurs incidents ont eu lieu le même jour puisque certaines marches de la trajectoire observée sont plus hautes. Nous nous retrouvons alors avec plusieurs temps d'occurrences égaux et cela n'est pas possible dans le cadre d'un processus de Poisson car il ne peut y avoir plusieurs instants d'occurrences en même temps.

Nous allons donc travailler avec un processus de Poisson composé $\sum_{n=0}^{N(t)} Y_n$, où $N(t)$ est obtenu en ne considérant qu'une seule fois chaque temps d'occurrences, ce qui correspond au processus sous-jacent, et Y_n correspond au nombre d'incidents ayant eu lieu à la date T_n . Par convention, $T_0 = 0$ et $Y_0 = 0$. Si nous avions codé les instants d'arrivées à la seconde, nous n'aurions pas eu d'occurrences simultanées, mais il est plus commode de travailler avec les dates.

Dès lors que nous considérons le processus sous-jacent, nous n'avons plus que 438 instants d'occurrences et nous définissons le processus des temps inter-arrivées $X_n = T_n - T_{n-1}, \forall n \geq 1$, qui correspond au temps d'attente entre deux occurrences. Les X_n ne peuvent être nulles puisque nous travaillons avec le processus $N(t)$ sous-jacent.

Le processus sous-jacent est-il un processus de Poisson homogène ?

Nous pouvons désormais tester si le processus sous-jacent $N(t)$ correspond à un processus de Poisson homogène. Pour cela, nous allons utiliser la proposition 1, page 19, selon laquelle un processus de comptage est un processus de Poisson homogène si et seulement si les $X_n = T_n - T_{n-1}$ sont indépendants et identiquement distribués selon une loi exponentielle de paramètre $\lambda > 0$.

Le Q-Q Plot des temps inter-arrivées permet d'obtenir de l'information sur le caractère exponentiel ou non des X_n . Il suit l'idée des tests d'adéquation basés sur la fonction de répartition empirique :

$$\begin{aligned} F_n(x) &\approx F(x), \forall x \in \mathbb{R} \\ \Leftrightarrow \frac{i}{n} &\approx 1 - e^{-\lambda x_{(i)}} \\ \Leftrightarrow x_{(i)} &\approx \frac{-1}{\lambda} \ln \left(1 - \frac{i}{n} \right) \end{aligned}$$

$\Leftrightarrow \left(x_{(i)}, -\ln \left(1 - \frac{i-\frac{1}{2}}{n} \right) \right)$ forme un nuage de points qui s'ajuste à une droite de pente λ .

Si les points sont bien alignés, nous pouvons supposer que le processus est un processus de Poisson homogène d'intensité λ que nous pouvons estimer grâce à la pente de la droite.

Voici le Q-Q Plot des temps inter-arrivées qui correspond à nos données :

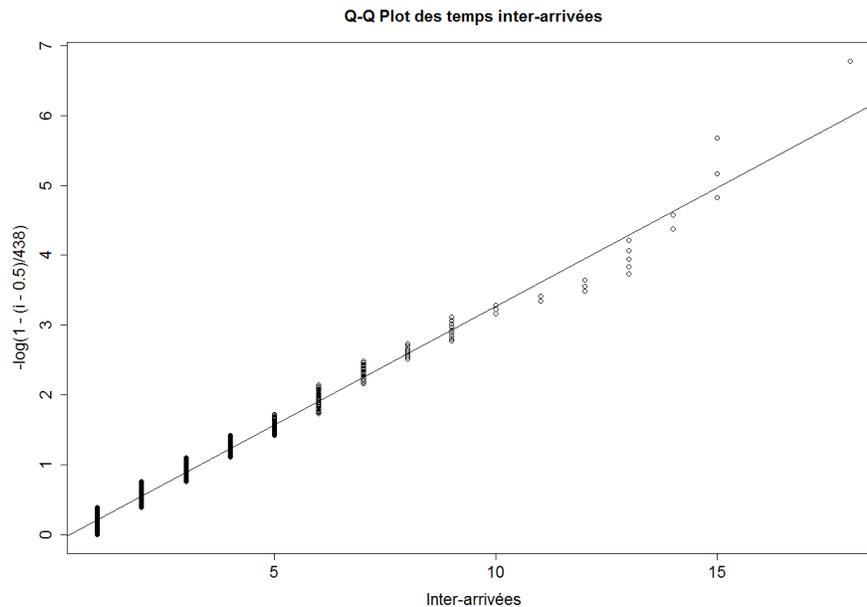


FIGURE 15 – Q-Q Plot des temps inter-arrivées

La figure 15, page 23 montre plusieurs ensembles de points qui sont caractérisés par une même valeur du temps inter-arrivées alors qu'en théorie il n'y a pas plusieurs valeurs égales pour une même variable continue. En fait, il ne se passe pas exactement un nombre entier de jours entre deux incidents mais les données sont arrondies à l'entier le plus proche. La méthode graphique nous montre qu'un ajustement linéaire n'est pas recommandé car les points ne sont pas alignés et donc il ne serait pas adéquat de modéliser le processus sous-jacent $N(t)$ par un processus de Poisson homogène.

Nous allons tout de même tester si l'ajustement exponentiel des temps inter-arrivées X_n est acceptable ou non selon le test de Kolmogorov-Smirnov. Pour cela, nous pouvons utiliser la fonction *ks.test* de R qui teste⁶ l'adéquation d'un échantillon de données à une distribution connue.

Voici une représentation graphique qui montre l'écart entre la fonction de répartition empirique des X_n et la fonction de répartition de la loi exponentielle de paramètre λ estimé :

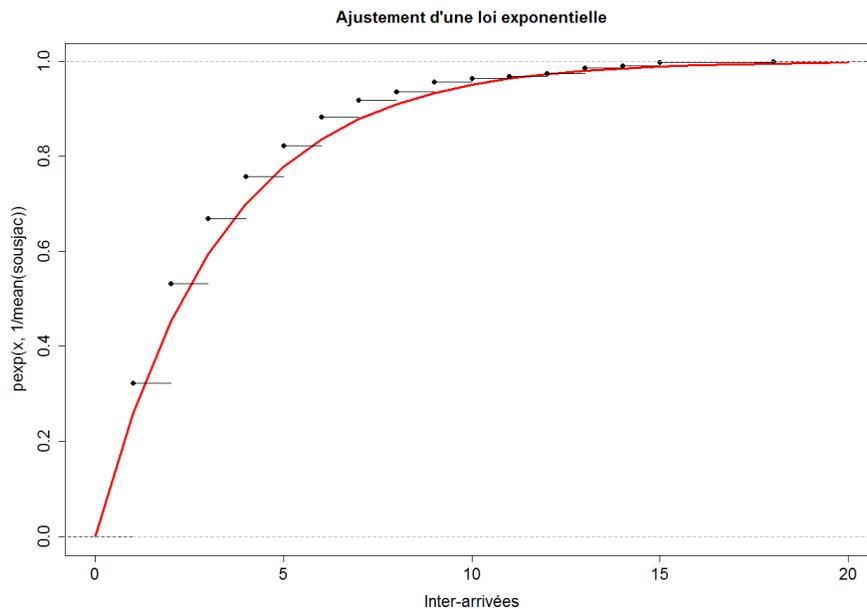


FIGURE 16 – Ajustement à une loi exponentielle

La statistique de test de Kolmogorov-Smirnov que nous obtenons est : $D_{438} = 0,2596$. Nous avons vu précédemment que pour le test de Kolmogorov-Smirnov, la distribution théorique doit être entièrement spécifiée. Ainsi, nous ne devons pas tenir compte de la p-valeur donnée par le logiciel R car le paramètre λ est inconnu et a été estimé sur l'échantillon à tester.

Cependant, J. DURBIN s'est intéressé au test de Kolmogorov-Smirnov en cas de paramètres inconnus dans l'article [3], publié dans la revue sur les statistiques *Biometrika*.

6. *ks.test* peut également tester si deux échantillons suivent la même loi.

Cet article présente une technique qui permet d'obtenir la distribution exacte de la statistique de test lorsque les paramètres sont estimés. Puis cette technique est appliquée au cas d'une distribution exponentielle. Une table des seuils critiques c_n tels que $\mathbb{P}[\sqrt{n}D_n > c_n] = \alpha$ est donnée dans l'article pour plusieurs valeurs de α (1%, 5%, 10%...).

Nous donnons la table pour $\alpha = 5\%$ en annexe 4, page 49. Une interpolation linéaire peut être réalisée pour les valeurs ne figurant pas dans la table. Pour les valeurs de $n > 100$, la méthode du bootstrap paramétrique⁷ a permis d'obtenir une estimation de la valeur du seuil critique. La figure 17 montre que c_n peut être estimé à $c_n = \sqrt{nd_n} \approx 1,08$ pour $n > 100$:

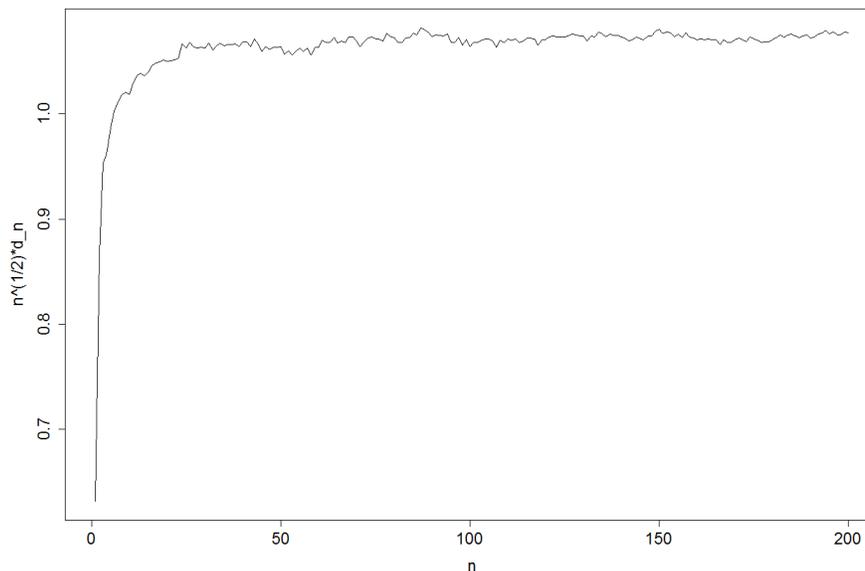


FIGURE 17 – Seuils critiques c_n par la méthode du bootstrap paramétrique

Nous pouvons ainsi obtenir les valeurs critiques d_n tel que $d_n = \frac{c_n}{\sqrt{n}}$. Nous obtenons d_n pour $n = 438$ qui vaut : $d_{438} = \frac{1,08}{\sqrt{438}} = 0,0516044$. Nous décidons donc, avec un risque de première espèce $\alpha = 5\%$, de rejeter l'hypothèse nulle selon laquelle les temps inter-arrivées sont distribués exponentiellement car $D_{438} = 0,2596 > d_{438} = 0,0516044$.

Un autre moyen pour tester si les X_n sont distribués exponentiellement est d'utiliser le package de R « *fitdistrplus* » et ses fonctions *fitdist* et *gofstat*.

La fonction *fitdist* ajuste⁸ une distribution théorique aux données et la fonction *gofstat* calcule les statistiques des tests d'adéquation suivants :

- Test du Chi-2 pour les distributions discrètes
- Tests de Kolmogorov-Smirnov, de Cramer-Von Mises et d'Anderson-Darling pour les distributions continues.

N.B. Le test d'Anderson-Darling est de la même famille que le test de Kolmogorov-Smirnov et de Cramer-Von Mises, il est basé sur la fonction de répartition empirique. La différence est que le test d'Anderson-Darling donne plus de poids aux queues de distribution.

7. Le code R pour réaliser la figure 17 a été repris du rapport [2] et est donné en annexe 5, page 49.

8. Par défaut, c'est la méthode du maximum de vraisemblance qui est utilisée.

Étant donné que nous testons l'adéquation des temps inter-arrivées à une distribution exponentielle, nous obtenons les statistiques suivantes :

- Statistique de Kolmogorov-Smirnov : $D_{438} = 0,2596$
- Statistique de Cramer-Von Mises : $W^2 = 3,1603$
- Statistique d'Anderson-Darling : $A^2 = 20,7110$

Les trois tests rejettent l'hypothèse nulle. Mais pour la statistique de Kolmogorov-Smirnov, comme précédemment, nous ne devons pas tenir compte de la décision affichée puisque la distribution testée n'est pas entièrement spécifiée, et nous devons comparer D_{438} avec la valeur critique d_{438} obtenue par simulation. Le résultat est le même puisque nous obtenons $D_{438} = 0,2596 > d_{438} = 0,0516044$. Nous concluons au rejet de l'hypothèse nulle avec un risque de première espèce $\alpha = 5\%$.

Le processus sous-jacent est-il un processus de Poisson non homogène⁹ ?

Le processus sous-jacent $N(t)$ ne correspond pas à un processus de Poisson homogène, nous allons alors tester s'il correspond mieux à un processus de Poisson non homogène.

Comme nous l'avons vu précédemment, les accroissements d'un processus de Poisson non homogène ne sont plus stationnaires et donc les temps inter-arrivées ne sont plus identiquement distribués. L'intensité d'un tel processus dépend alors du temps. Cela est peut être plus réaliste dans notre cas puisque nous avons remarqué qu'il y avait une petite accélération du cumul des incidents chaque été.

Les processus de Poisson non homogènes sont souvent utilisés en fiabilité et notamment pour les modèles sur les défaillances de systèmes réparables. Un des modèles les plus connus est sans doute le « Power Law Model ». Le processus Power Law a une fonction d'intensité donnée par :

$$\lambda(t) = abt^{b-1}; \quad a, b > 0, \quad t > 0$$

Nous allons donc ajuster nos données à un processus de Poisson non homogène d'intensité $\lambda(t) = abt^{b-1}$. Dans un premier temps, nous devons estimer les paramètres de l'intensité $\lambda(t)$, puis ensuite nous testerons si cet ajustement est acceptable ou non.

Estimation paramétrique de l'intensité.

Pour estimer les paramètres a et b de $\lambda(t)$, nous utilisons la *méthode d'estimation par maximum de vraisemblance*. La log-vraisemblance du processus $N(t)$ d'intensité $\lambda(t)$ s'écrit :

$$L(T) = \sum_{i=1}^{N(t)} \log \lambda(T_i) - \int_0^t \lambda(v) dv$$

Ainsi, dans le cas d'une intensité d'un processus Power Law, nous obtenons :

$$L(a, b) = N(t)(\log a + \log b) + (b - 1) \sum_{i=1}^{N(t)} \log T_i - at^b$$

En annulant les dérivées partielles, nous pouvons calculer les estimateurs \hat{a} et \hat{b} :

9. Cette partie s'appuie sur l'article [4].

- Estimateur \hat{a} de a :

$$\begin{aligned}\frac{dL(a, b)}{da} = 0 &\Leftrightarrow \frac{N(t)}{a} - t^b = 0 \\ &\Leftrightarrow \frac{N(t)}{a} = t^b \\ &\Leftrightarrow a = \frac{N(t)}{t^b}\end{aligned}$$

Nous obtenons l'estimateur suivant :

$$\hat{a} = \frac{N(t)}{t^{\hat{b}}}$$

où \hat{b} est l'estimateur par maximum de vraisemblance de b .

Rappelons que :

$$\frac{du^x}{dx} = u^x \ln u$$

- Estimateur \hat{b} de b :

$$\begin{aligned}\frac{dL(a, b)}{db} = 0 &\Leftrightarrow \frac{N(t)}{b} + \sum_{i=1}^{N(t)} \log T_i - at^b \log t = 0 \\ &\Leftrightarrow \frac{N(t)}{b} + \sum_{i=1}^{N(t)} \log T_i - \frac{N(t)}{t^b} t^b \log t = 0 \\ &\Leftrightarrow \frac{N(t)}{b} = N(t) \log t - \sum_{i=1}^{N(t)} \log T_i \\ &\Leftrightarrow b = \frac{N(t)}{N(t) \log t - \sum_{i=1}^{N(t)} \log T_i}\end{aligned}$$

Nous obtenons l'estimateur suivant :

$$\hat{b} = \frac{N(t)}{N(t) \log t - \sum_{i=1}^{N(t)} \log T_i}$$

La période d'étude est de 4 ans, du 1^{er} janvier 2008 au 31 décembre 2011, donc $t = 1461$ jours. $N(t) = 438$ puisque nous travaillons avec le processus sous-jacent. Nous pouvons ainsi calculer les estimateurs \hat{a} et \hat{b} avec le logiciel R :

$$\hat{b} \simeq 0,9618 \text{ et } \hat{a} \simeq 0,3958$$

N.B. Une approche numérique est possible lorsqu'une solution analytique n'existe pas ou n'est pas facilement réalisable pour trouver les estimateurs par maximum de vraisemblance. La fonction *nlm* de R permet de minimiser une fonction à plusieurs variables. Il nous suffit alors de minimiser l'opposée de la log-vraisemblance.

Tests d'ajustement à un processus de Poisson non homogène.

Nous allons maintenant mettre en place deux tests d'adéquation à un processus de Poisson non homogène dans le cas d'une paramétrisation Power Law pour tester si nos données s'ajustent bien à un tel processus : les tests modifiés de Kolmogorov-Smirnov et de Cramer-von Mises qui sont décrits dans l'article [4].

Ces tests statistiques considèrent une hypothèse nulle de la forme :

$$H_0 : T \sim \text{processus de Poisson non homogène}$$

contre

$$H_1 : T \not\sim \text{processus de Poisson non homogène}$$

Statistique de test « modifiée » de Kolmogorov-Smirnov.

$$D = \max(D^+, D^-)$$

avec

$$D^+ = \max_{1 \leq i \leq n} \left(\frac{i}{n} - \widehat{V}_i \right), \quad D^- = \max_{1 \leq i \leq n} \left(\widehat{V}_i - \frac{i-1}{n} \right)$$

où

$$\widehat{V}_i = \left(\frac{T_i}{t} \right)^{\hat{b}} \text{ et } n = N(t) = 438.$$

Le calcul de la statistique de test observée donne $D_{obs} \simeq 0,0370$.

Puis comme cela est mentionné dans l'article, nous pouvons simuler la distribution de D . Nous réalisons 100 000 échantillons de (\widehat{V}_i) où :

$$\widehat{V}_i = U_{(i)}^{-n / \sum_{i=1}^n \log U_i} \quad \forall i = 1, \dots, n \text{ et } U_i \sim U[0, 1].$$

Nous pouvons ainsi calculer la valeur critique du test, pour $\alpha = 5\%$, en prenant le quantile à 95% de la distribution simulée D_{sim} .

La fonction *quantile* de R nous permet d'obtenir $D_{crit} \simeq 0,0518$.

Ainsi, nous avons :

$$D_{obs} \simeq 0,0370 \leq D_{crit} \simeq 0,0518$$

La p-valeur vaut :

$$\mathbb{P}[D_{sim} \geq D_{obs}] \simeq 0,3493$$

Ainsi, nous décidons de ne pas rejeter l'hypothèse nulle et donc le processus sous-jacent $N(t)$ s'ajuste bien à un processus de Poisson non homogène avec une fonction d'intensité $\lambda(t) = abt^{b-1}$.

Voici l'histogramme qui représente la distribution simulée D_{sim} avec la valeur de la statistique de test observée D_{obs} et la valeur critique D_{crit} :

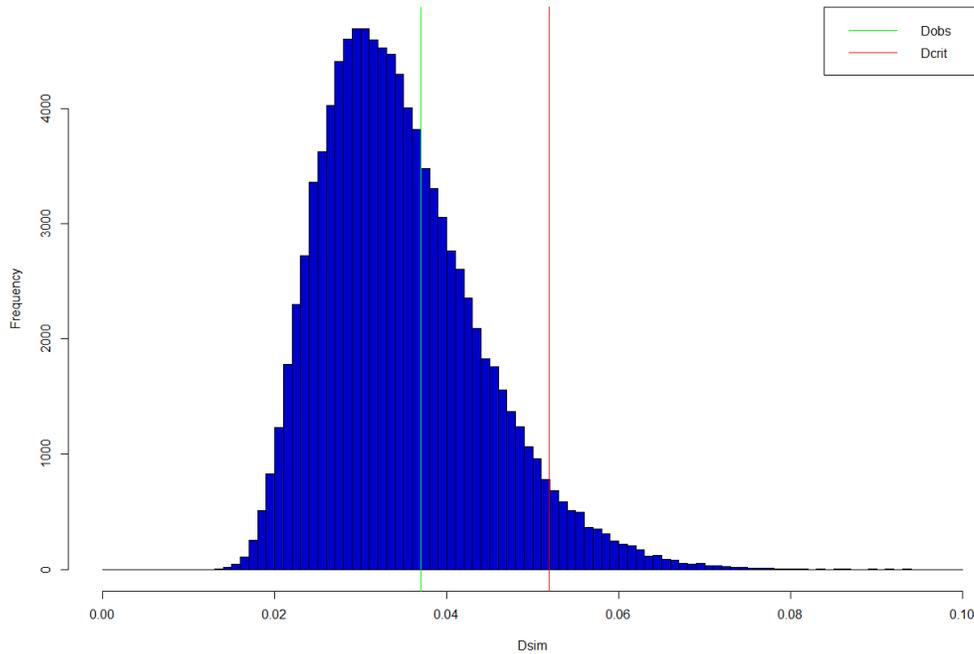


FIGURE 18 – Histogramme de la distribution simulée de D

Statistique de test « modifiée » de Cramer-Von Mises.

$$W^2 = \sum_{i=1}^n n \left(\widehat{V}_i - \frac{2i-1}{2n} \right)^2 + \frac{1}{12n}$$

où

$$\widehat{V}_i = \left(\frac{T_i}{t} \right)^{\hat{b}} \text{ et } n = N(t) = 438.$$

Le calcul de la statistique de test observée donne $W_{obs}^2 \simeq 0,1591$.

Puis comme pour le test de Kolmogorov-Smirnov, nous pouvons simuler la distribution de W^2 . Nous réalisons 100 000 échantillons de (\widehat{V}_i) où :

$$\widehat{V}_i = U_{(i)}^{-n / \sum_{i=1}^n \log U_i} \quad \forall i = 1, \dots, n \text{ et } U_i \sim U[0, 1].$$

Nous pouvons ainsi calculer la valeur critique du test, pour $\alpha = 5\%$, en prenant le quantile à 95% de la distribution simulée W_{sim}^2 .

La fonction *quantile* de R nous permet d'obtenir $W_{crit}^2 \simeq 0,2232$.

Ainsi, nous avons :

$$W_{obs}^2 \simeq 0,1591 \leq W_{crit}^2 \simeq 0,2232$$

La p-valeur vaut :

$$\mathbb{P}[W_{sim}^2 \geq W_{obs}^2] \simeq 0,1280$$

Ainsi, nous décidons de ne pas rejeter l'hypothèse nulle et donc le processus sous-jacent $N(t)$ s'ajuste bien à un processus de Poisson non homogène avec une fonction d'intensité $\lambda(t) = abt^{b-1}$.

Voici l'histogramme qui représente la distribution simulée W_{sim}^2 avec la valeur de la statistique de test observée W_{obs}^2 et la valeur critique W_{crit}^2 :

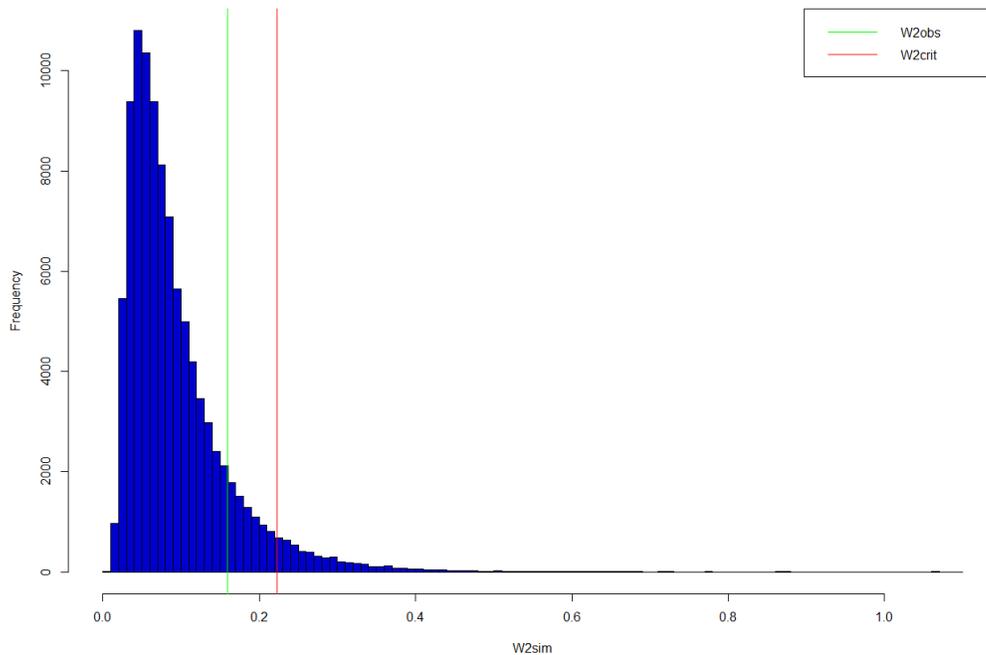


FIGURE 19 – Histogramme de la distribution simulée de W^2

N.B. Un algorithme pour simuler un processus de Poisson non homogène avec une fonction d'intensité Power Law est également donné dans l'article [4].

Pour pouvoir conclure et nous assurer que les incidents HTA/HTB peuvent être modélisés par un processus composé, il reste à tester l'indépendance entre Y_n , qui est le nombre d'incidents ayant lieu à chaque date T_n , et le processus sous-jacent $N(t)$. Pour cela, nous pouvons utiliser le test d'indépendance du Chi-2 via la commande `chisq.test` de R.

Le test d'indépendance du Chi-2, entre un échantillon aléatoire issu de la distribution des Y_n et un échantillon aléatoire issu de la distribution de Poisson de paramètre $\Lambda(t) = at^b$ pour t quelconque, ne rejette pas l'hypothèse nulle d'indépendance.

Conclusion :

Les incidents HTA/HTB du réseau de l'ÉS entre 2008 et 2011 peuvent être modélisés par un processus de Poisson composé $\sum_{n=0}^{N(t)} Y_n$, où $N(t)$ est un processus de Poisson non homogène d'intensité $\lambda(t) = abt^{b-1}$, et Y_n correspond au nombre d'incidents ayant eu lieu à la date T_n . Par convention, $T_0 = 0$ et $Y_0 = 0$.

Voici 1000 simulations du processus d'arrivées des incidents de l'ÉS de 2008 à 2011 :

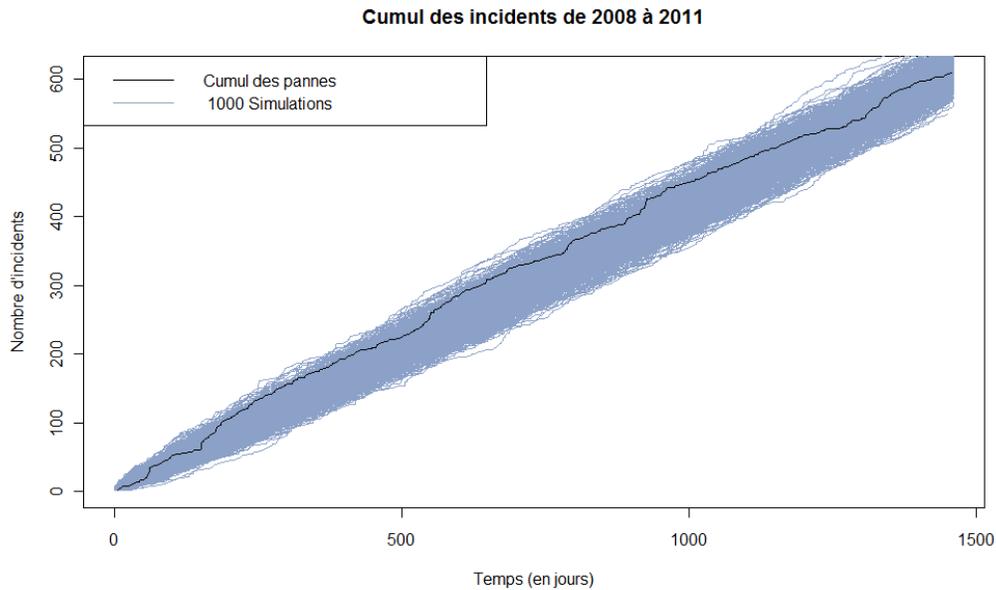


FIGURE 20 – Simulations des incidents 2008 à 2011

Tout ce chapitre 3 a également été fait en tenant compte uniquement des incidents souterrains d'une part, et aériens d'autre part. Ces deux analyses, que nous ne présentons pas ici, ont fourni les mêmes conclusions aux tests que pour l'analyse ci-dessus.

4 Simulation de pannes souterraines pour chaque départ HTA.

Nous nous intéressons maintenant aux incidents souterrains. Il y a deux sous-type d'incidents souterrains :

- les incidents souterrains sur câbles *papiers* (câbles à isolation par du papier imprégné d'huile diélectrique).
- les incidents souterrains sur câbles *synthétiques* (câbles à isolation en polyéthylène).

Les incidents sur câble papier sont plus fréquents que ceux sur câble synthétique, alors que la longueur totale de câbles papiers est largement plus faible que la longueur totale de câbles synthétiques (cf. section 4.1.2, page 33). Pour preuve, sur les quatre dernières années, nous avons comptabilisé 221 incidents sur câble papier et 73 sur câble synthétique.

Les longueurs des câbles vont également jouer un rôle car il y a plus de chance de voir apparaître une panne sur un départ long que sur un départ court.

4.1 « Espérance » de panne.

Le premier objectif ici est d'obtenir une « espérance » de panne pour chaque départ HTA par an. Comme nous l'avons expliqué en introduction, le réseau HTA est composé de plusieurs départs HTA (ou cellules) à partir des postes sources dits postes HTB. Le réseau comporte 40 postes HTB et chacun alimente 10 départs HTA en moyenne.

Nous allons donc dans un premier temps calculer les longueurs des câbles papiers et synthétiques pour chaque départ HTA. Puis, en se basant sur les données observées de 2008 à 2011, nous pourrions calculer un taux d'incident au kilomètre pour les câbles papiers et synthétiques pour chacun des départs HTA. Enfin, à partir de ces deux taux et des longueurs de câbles de chacune des cellules, nous pourrions calculer une espérance de panne pour chaque départ HTA par an.

4.1.1 Calcul des longueurs des câbles.

Les longueurs des câbles papiers et synthétiques sont disponibles dans une base de données dans laquelle figure la topologie du réseau.

La figure 21, page 33, montre un exemple de la topologie du réseau concernant le départ n° 15 du poste source de Reichstett. Nous distinguons clairement la structure arborescente du départ HTA.

Un départ est composé de plusieurs circuits, ici le départ n° 15 en compte neuf. Les extractions réalisées et qui ont été mises à notre disposition ont été faites pour chaque circuit de chaque départ HTA pour chaque poste source.

Ainsi pour calculer la longueur totale de câbles papiers et synthétiques de chaque départ, il a fallu être vigilant sur les doublons, c'est-à-dire ne pas comptabiliser plusieurs fois une longueur de câble donnée qui a déjà été prise en compte.

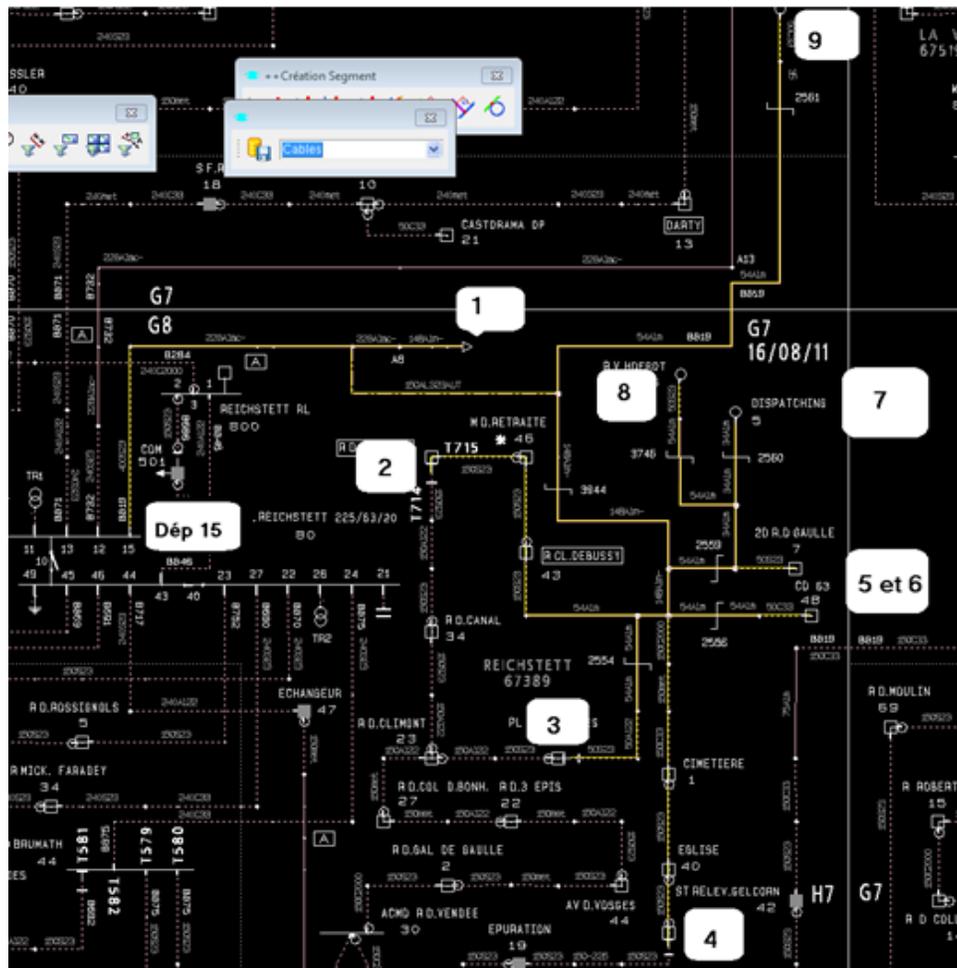


FIGURE 21 – Exemple du départ n° 15 du poste de Reichstett

4.1.2 Taux d'incident au kilomètre.

À partir des résultats précédents, nous pouvons calculer la longueur totale de câbles papiers et de câbles synthétiques :

- longueur totale de câbles papiers $\simeq 493$ kilomètres.
- longueur totale de câbles synthétiques $\simeq 2920$ kilomètres.

Puisque de 2008 à 2011 il y a eu 221 incidents sur câble papier et 73 sur câble synthétique, nous obtenons les deux taux d'incident suivants sur 4 ans :

- taux d'incident papier $\simeq 0,448139$ au kilomètre.
- taux d'incident synthétique $\simeq 0.024995$ au kilomètre.

Ce qui donne sur 1 an :

- taux d'incident papier : $tx_{papier} \simeq 0,112035$ au kilomètre.
- taux d'incident synthétique : $tx_{synth} \simeq 0.006249$ au kilomètre.

4.1.3 Calcul des espérances de panne E_{panne} pour chaque cellule.

Nous allons maintenant pouvoir calculer le nombre attendu d'incidents par an pour chacun des deux types de câbles.

Pour cela, nous multiplions les taux tx_{papier} et tx_{synth} par les longueurs de câbles l_{papier} et l_{synth} respectives de chaque départ et nous obtenons un nombre attendu d'incidents par an pour les deux types de câbles. Et ainsi en faisant la somme des deux, nous obtenons le résultat souhaité :

$$E_{panne} = tx_{papier} * l_{papier} + tx_{synth} * l_{synth}$$

N.B. En faisant la somme de toutes les espérances de panne E_{panne} obtenues pour chacun des départs HTA, nous retrouvons le nombre d'incidents survenus en 1 an, c'est-à-dire 73,5 incidents, qui n'est rien d'autre que la moyenne du nombre d'incidents souterrains qui se sont produits de 2008 à 2011.

4.2 Critère B attendu par an.

Pour calculer le critère B attendu par an, nous avons besoin de la durée des incidents pour chaque départ, du nombre de clients par départ et des espérances de panne E_{panne} calculées précédemment :

$$\text{Critère B attendu} = \sum_{\text{Départ } k} \left(\frac{\text{Durée incident} * \text{Nb de clients} * E_{panne}}{\text{Nombre total de client ÉS}} \right)$$

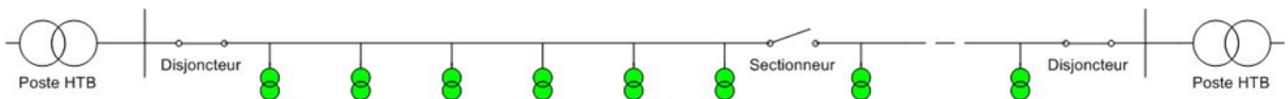
4.2.1 Durées des incidents.

Dans cette partie, nous tentons de modéliser du mieux possible la durée d'un incident qui correspond à la durée nécessaire à la localisation du défaut, de l'isolement du tronçon en défaut et à la réalimentation des tronçons sains.

Nous présentons¹⁰ ici plus en détail la séquence de réalimentation des clients touchés par un incident HTA compte tenu de la topologie et de la connaissance de la position des organes de manœuvres télécommandés ou non. Les clients sont tous coupés en une fois lors du court-circuit mais sont réalimentés en plusieurs étapes, car il faut réalimenter le réseau en gardant hors tension le segment en défaut en réorganisant la topologie du réseau.

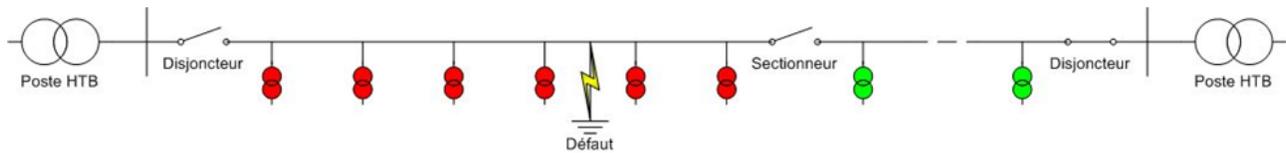
Dans les cas les plus simples, la recherche et l'élimination d'un défaut nécessitent au minimum 3 étapes. Dans le cas que nous allons présenter, nous diviserons l'élimination du défaut en 5 étapes pour bien comprendre le principe :

0 : Le réseau fonctionne normalement avant l'apparition du défaut. Ici, nous utilisons la couleur vert pour un poste HTA/BT alimenté et rouge pour un poste HTA/BT coupé.

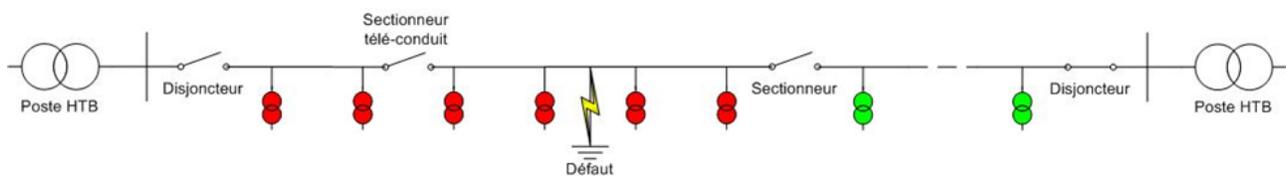


10. La présentation de la séquence de réalimentation des clients a été entièrement reprise de [5].

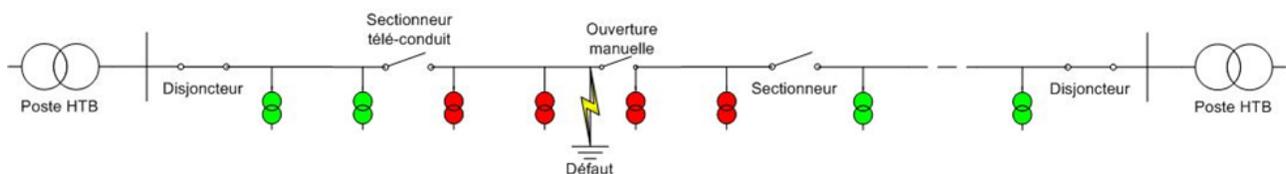
1 : A l'apparition d'un défaut, le disjoncteur de la ligne où a lieu le défaut, s'ouvre. La ligne n'étant plus alimentée, tous les postes reliés à cette ligne sont alors coupés, soit 100% des clients alimentés par la ligne.



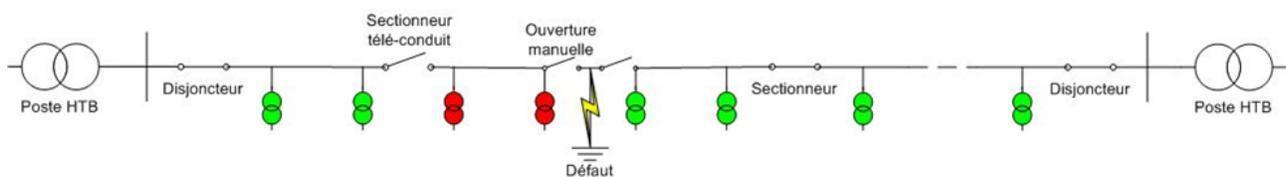
2 : C'est alors que commence le travail de localisation du défaut. Le Bureau Central de Conduite (BCC) a à sa disposition des sectionneurs télé-conduits placés à des endroits stratégiques sur certaines lignes pour lui permettre d'ouvrir ces lignes à un endroit précis.



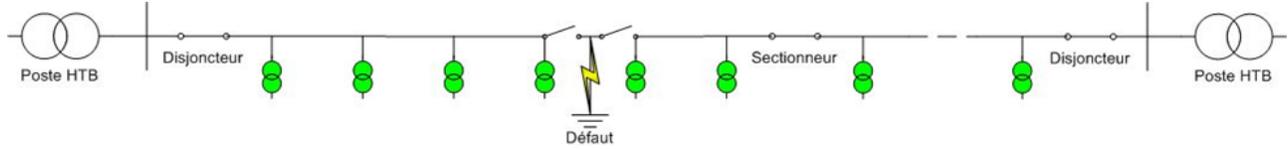
3 : Le sectionneur télé-conduit ouvert, le courant pourra à nouveau alimenter les postes HTA/BT situés entre ce sectionneur et le poste HTB. A noter qu'il s'agit d'un cas précis et simplifié pour ne pas troubler le lecteur. N'ayant plus de sectionneur télé-conduit entre le défaut et le dernier sectionneur utilisé, il faut aller sur le terrain pour ouvrir différents sectionneurs manuellement. A ce stade, environ 50% des clients ont déjà été réalimentés grâce aux sectionneurs télé-conduits. La réalimentation par la télé-conduite prend quelques minutes.



4 : Le réseau HTA est relié de part et d'autre à un poste HTB. Après une ouverture manuelle correcte d'un sectionneur, les parties de lignes entre le poste HTB et le sectionneur ouvert manuellement peuvent à nouveau être alimentées.



5 : Les exploitants (personnes réalisant les travaux sur les sites touchés) procèdent ainsi par dichotomie pour localiser le défaut et l'isoler jusqu'à obtenir un nombre de clients coupés minimum. Les contrôles pour la dichotomie sont réalisés grâce aux DLD (Dispositifs Lumineux de Défaut) qui indiquent si le défaut se trouve en amont ou en aval du poste examiné (visible uniquement sur place).



Les clients sont donc réalimentés en 3 étapes (étape 3, 4 et 5). Il s'agit ici d'un cas d'école où la réalimentation des clients se déroule sans problème.

Notre but est de calculer au plus proche de la réalité, la durée d'un incident pour chaque départ HTA. Pour cela, nous procédons de la façon suivante :

- Nous avons à notre disposition la liste des postes HTA/BT de chacun des départs HTA, et également les coordonnées en « Lambert » (X, Y) de ces postes.

Nous décidons de prendre la distance entre (X_{min}, Y_{min}) et (X_{max}, Y_{max}) de manière à prendre la distance d_1 la plus large possible dans la zone dans laquelle se trouvent les postes HTA/BT d'un départ donné :

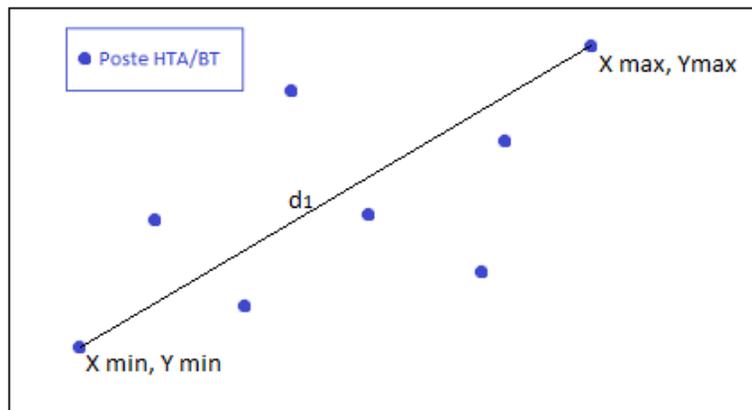


FIGURE 22 – Schéma illustrant la formule de d_1

$$d_1 = \sqrt{(X_{min} - X_{max})^2 + (Y_{min} - Y_{max})^2}$$

- Puis, nous multiplions d_1 par $\sqrt{2}$ pour modéliser le fait que les routes ne sont pas toujours en ligne droite, et nous divisons le tout par 1000 afin d'obtenir une distance d_2 en kilomètres :

$$d_2 = \frac{d_1 * \sqrt{2}}{1000}$$

- Enfin, pour obtenir une durée, nous supposons que les exploitants se déplacent à une vitesse de 30 km/h. Pour être plus réaliste, nous ajoutons 1 minute par poste parcouru pour un départ donné et nous supposons que les exploitants mettent environ 10 minutes pour se rendre au premier poste HTA/BT du départ concerné par l'incident. Nous obtenons ainsi le temps t_{inc} (en minutes) de durée d'un incident :

$$\begin{aligned} t_{inc} &= \frac{60 * d_2}{30} + nb_{postes} + 10 \\ &= 2 * d_2 + nb_{postes} + 10 \end{aligned}$$

4.2.2 Nombre de clients par départ.

Nous avons à notre disposition le nombre de clients rattachés à chaque poste HTA/BT. Nous avons également la liste des postes HTA/BT de chaque départ HTA.

Ainsi, à l'aide d'un tableau croisé dynamique, il nous est possible d'obtenir directement le nombre de clients pour chaque départ HTA, noté $nb_{clients\ coupés}$.

La somme des clients départ par départ donne un nombre total de clients ÉS égal à 432 038.

4.2.3 Calcul des contributions au critère B attendu.

Nous avons désormais toutes les données pour calculer les contributions au critère B de chaque départ HTA. Pour un départ donné, la contribution se calcule de la façon suivante :

$$\begin{aligned} \text{Contribution} &= \left(\frac{t_{inc} * nb_{clients\ coupés} * E_{panne}}{\text{Nombre total de client ÉS}} \right) \\ &= \left(\frac{t_{inc} * nb_{clients\ coupés} * E_{panne}}{432\,038} \right) \end{aligned}$$

En additionnant les contributions de chaque départ HTA, nous obtenons le critère B attendu à l'année qui vaut 10,17 minutes.

4.3 Simulation de pannes souterraines.

L'idée maintenant est de simuler des pannes souterraines pour chaque départ HTA et ainsi nous pourrions calculer le critère B pour plusieurs années d'affilée et les comparer au critère B que nous venons de calculer ci-dessus.

Pour cela, il faut savoir quelle loi serait convenable pour réaliser ces simulations. Pour trouver cette loi, nous nous basons sur nos données de 2008 à 2011.

Nous avons trouvé précédemment que les incidents de l'ÉS pouvaient être modélisés par un processus composé avec un processus sous-jacent modélisé par un processus de Poisson non homogène de fonction d'intensité Power Law. Mais cette modélisation correspond aux incidents toutes causes confondues (réseau aérien, poste HTA/BT, câble souterrain).

Ici, nous souhaitons modéliser les incidents uniquement souterrains et pris cellule par cellule. Les temps inter-incidents sont donc plus espacés et il est très rare qu'il y ait deux ou plusieurs incidents la même journée pour un même départ HTA. Cela laisse supposer que nous pouvons tenter de modéliser ces incidents par un processus de Poisson homogène cette fois-ci.

Nous allons donc tester sur nos données de 2008 à 2011, s'il est possible de modéliser les incidents souterrains pris cellule par cellule par un processus de Poisson homogène. Il nous faut alors tester si les temps inter-incidents suivent une loi exponentielle de paramètre $\lambda > 0$.

Comme les données que nous avons ne portent que sur quatre années, il n'est pas évident d'avoir un échantillon assez large pour un départ donné. Nous allons donc réaliser nos tests sur le départ qui a subi le plus d'incidents souterrains durant ces quatre dernières années. Il s'agit du départ n° 14 du poste source Port du Rhin. Il s'est produit 13 incidents souterrains de 2008 à 2011 sur ce départ avec les temps d'arrivées T_n et les temps inter-arrivées X_n suivants :

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}	T_{13}
164	356	578	603	626	647	741	866	874	956	1049	1057	1173
X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}
164	192	222	25	23	21	94	125	8	82	93	8	116

Nous testons si les X_n sont distribuées suivant une loi exponentielle avec la fonction *ks.test* de R. Nous obtenons la statistique de test de Kolmogorov-Smirnov : $D_{13} = 0,2124$.

Nous comparons D_{13} à la valeur critique $d_{13} = \frac{c_{13}}{\sqrt{13}} = \frac{1,0354}{\sqrt{13}} = 0,2871683$. $D_{13} \leq d_{13}$ et donc nous décidons de ne pas rejeter l'hypothèse nulle.

Utilisons également la librairie « *fitdistrplus* » et ses fonctions *fitdist* et *gofstat*. Nous obtenons :

- Statistique de Kolmogorov-Smirnov : $D_{13} = 0,2124$
- Statistique de Cramer-Von Mises : $W^2 = 0,0866$
- Statistique d'Anderson-Darling : $A^2 = 0,4697$

La décision du test de Kolmogorov-Smirnov n'est pas calculée. Dans l'aide de la fonction *gofstat*, il est dit que le résultat du test de Kolmogorov-Smirnov n'est donné seulement si le jeu de données est supérieur à 30 observations, or ce n'est pas le cas de notre échantillon. Ce test n'est donc pas recommandé pour notre échantillon même si nous pouvons comparer D_{13} à d_{13} comme nous l'avons fait précédemment.

Les deux autres tests donnent leur décision pour des jeux de données supérieurs à cinq observations. Notre jeu de données est composé de 13 observations donc les décisions des tests sont données et ils ne rejettent pas l'hypothèse nulle.

Nous pouvons ainsi conclure que les temps inter-incidents souterrains pour le départ n° 14 du poste source du Port du Rhin peuvent être modéliser par un processus de Poisson homogène d'intensité $\lambda > 0$.

Il n'est pas possible de tester sur tous les autres départ HTA des autres postes sources car nous n'aurions à chaque fois que des échantillons avec un nombre d'observations très faible, mais dans la suite nous supposons que nous pouvons modéliser les incidents souterrains pour chaque cellule par un processus de Poisson homogène.

Simulations des pannes.

Si pour un départ donné, les pannes peuvent se modéliser par un processus de Poisson homogène $\lambda > 0$, nous pouvons simuler les temps inter-incidents par la loi exponentielle de paramètre λ d'après la proposition 1, page 19.

Nous allons donc simuler des temps inter-incidents grâce à la loi exponentielle pour chacun des départs HTA. Le paramètre λ de la loi exponentielle sera pour chaque départ l'espérance E_{panne} que nous avons calculé plus haut. Puis, nous cumulons les temps inter-incidents simulés pour chaque départ HTA pour obtenir les dates d'arrivées des incidents souterrains de chaque départ. Le programme R de ces simulations est donnée en annexe 6, page 50.

Ensuite, nous décidons d'observer ce qu'il se passe sur 100 années. Nous comptabilisons ainsi le nombre d'incidents qui se trouvent dans l'intervalle $[0;365]$, $]365;730]$, \dots , $]36\ 135;36\ 500]$ pour chaque cellule de chaque poste source.

Nous obtenons donc un grand tableau avec beaucoup de 0, de 1 et de 2, qui correspondent au nombre d'incidents qui ont eu lieu telle ou telle année pour un départ donné. Il y a également quelques 3 et 4 qui se trouvent dans le tableau. Cela arrive très rarement et s'explique par le fait que certains départs HTA sont très long et donc nous ne pouvons pas exclure que sur un même départ, il arrive parfois trois ou quatre incidents dans la même année.

Une fois ce travail effectué, nous avons toutes les données à disposition pour calculer le critère B pour 100 ans d'affilée par la formule que nous rappelons :

$$\text{Critère B} = \sum_{\text{Incidents } i \text{ de l'année}} \left(\frac{\text{Durée de l'incident } i * \text{Nb de clients coupés dans l'incident } i}{\text{Nombre total de client ÉS}} \right)$$

Nous pouvons alors maintenant comparer le moyenne des critères B obtenus avec le critère B que nous avons calculé avec les espérance de panne :

- Critère B attendu par an = 10,17 minutes.
- Moyenne des critères B sur 100 années de simulations de pannes = 9,56 minutes.

Les deux résultats sont très proches.

4.4 Quelques représentations graphiques.

La figure 23, page 40, est très intéressante car elle nous permet de remarquer que les quatre dernières années ont été atypiques au niveau du critère B, et cela est plus particulièrement vrai pour l'année 2010 et 2011 avec un critère B de 3 minutes. Nous observons également la dispersion qu'il peut y avoir d'une année à l'autre, le critère B pouvant être de 6 à 14 minutes suivant les années.

La figure 24 nous montre aussi une grande dispersion au niveau du nombre d'incidents. D'une année à l'autre nous pouvons avoir de 50 à 100 incidents. 2008 et 2009 ont été des années avec un nombre d'incidents plutôt élevé au contraire de 2010 et 2011 qui sont en dessous de la moyenne sur les 100 années de simulations.

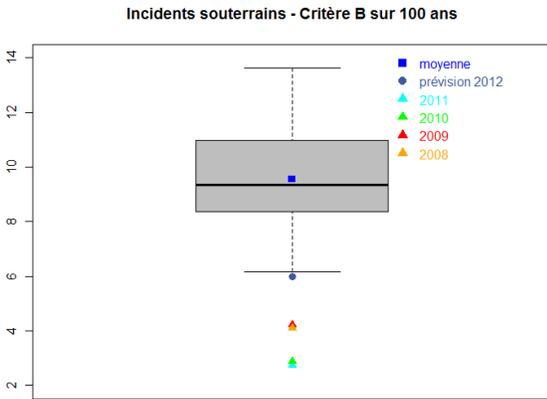


FIGURE 23 – Boîtes à moustaches du critère B sur 100 ans

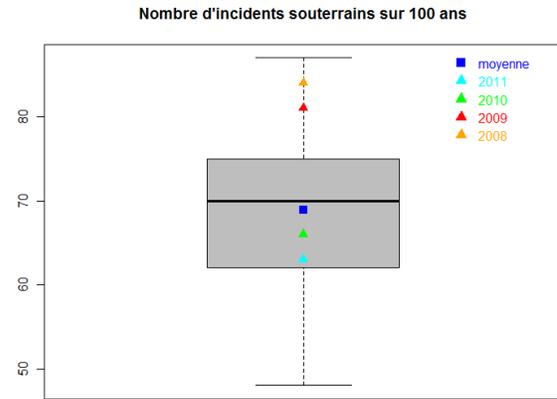


FIGURE 24 – Boîtes à moustaches du nombre d'incidents sur 100 ans

Nous pouvons également observer les distributions associées au critère B et au nombre d'incidents :

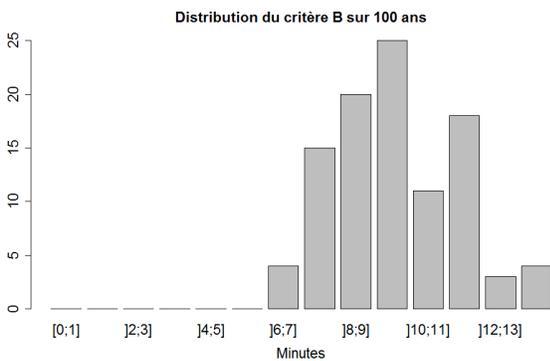


FIGURE 25 – Distribution du critère B sur 100 ans

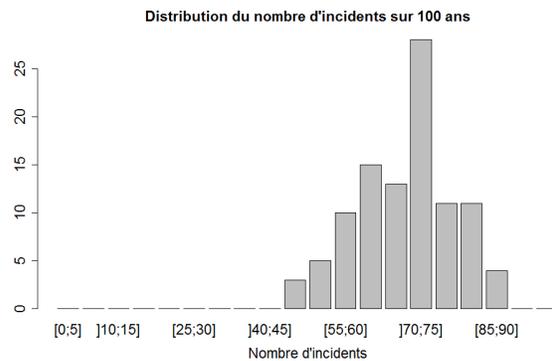
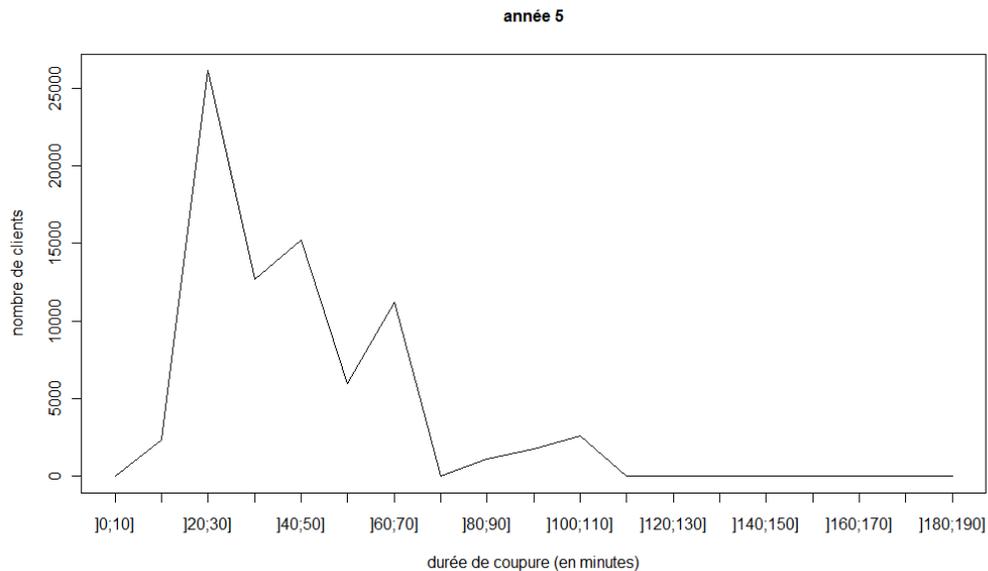


FIGURE 26 – Distribution du nombre d'incidents sur 100 ans

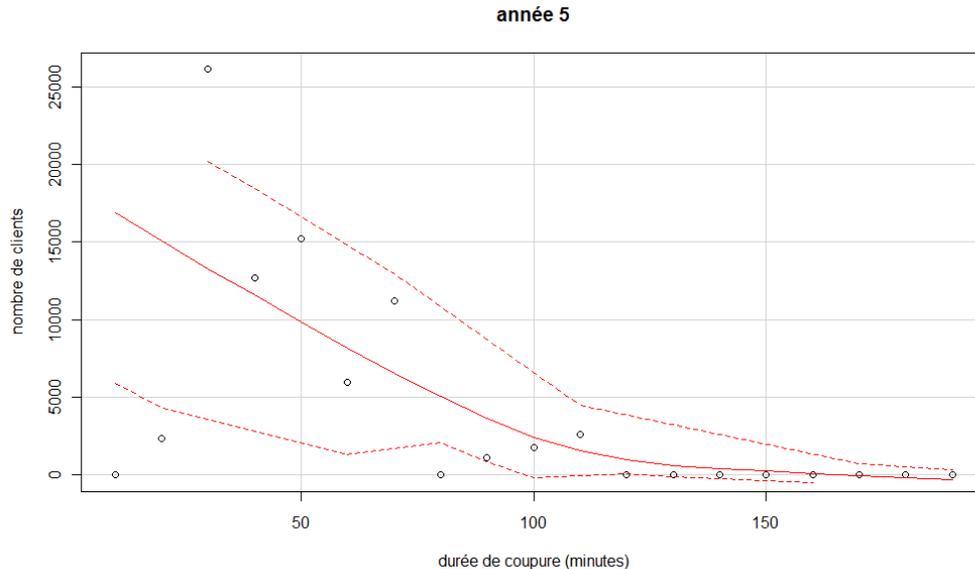
Nous observons à la figure 25 qu'un quart des critères B se situent entre 9 et 10 minutes, et quasiment la moitié entre 8 et 10 minutes.

Concernant le nombre d'incidents, la figure 26 montre que plus d'un quart des nombres d'incidents sont entre 70 et 75 incidents, et près de 60% des nombres d'incidents sont entre 60 et 75 incidents.

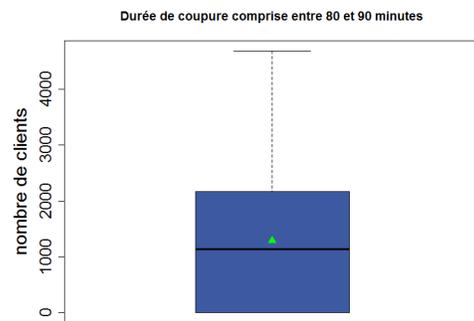
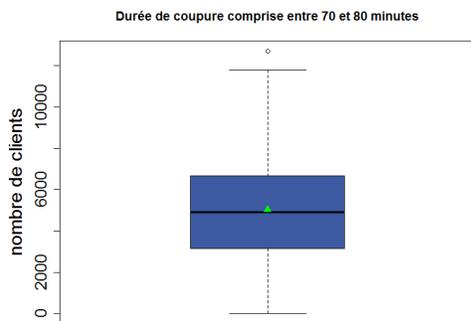
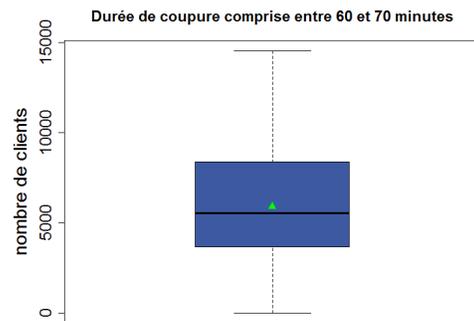
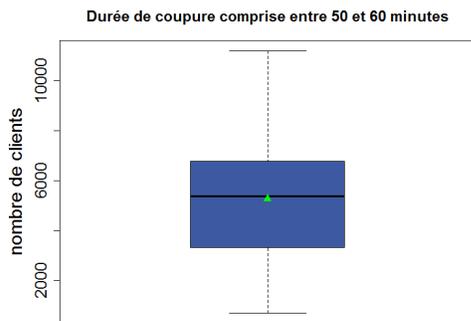
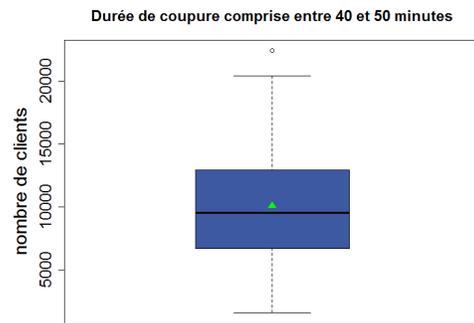
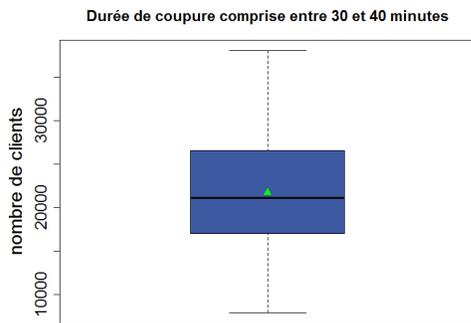
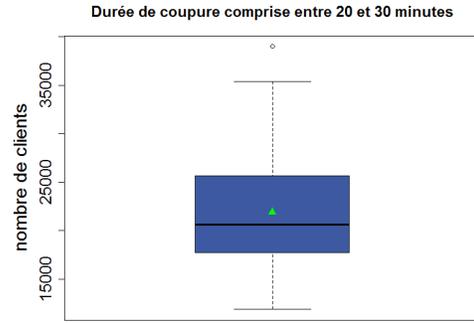
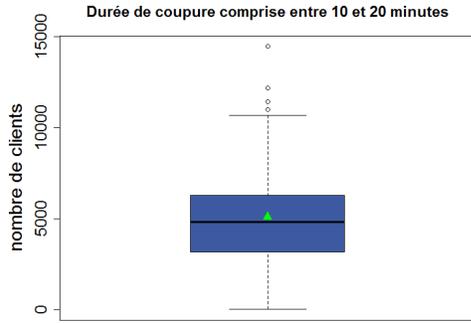
Nombre de clients en fonction de la durée de coupure.



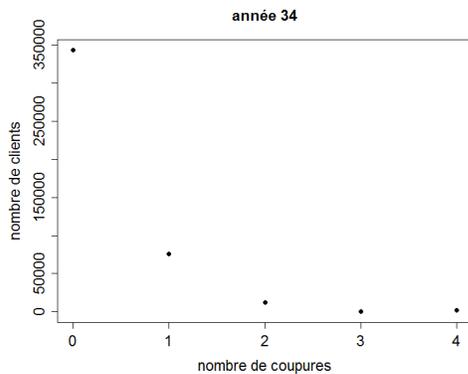
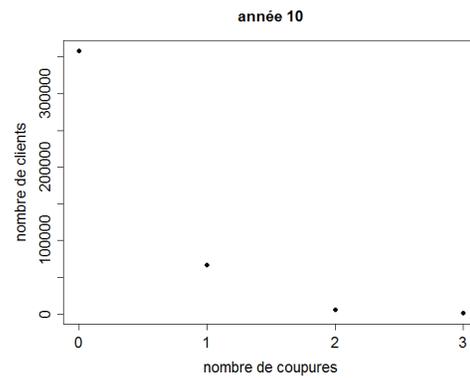
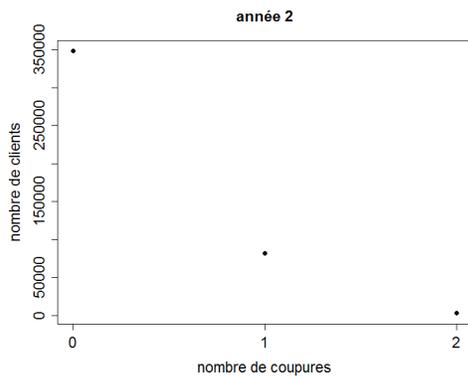
Nous pouvons afficher une fonction de lissage « loess » basée sur une régression non paramétrique grâce à la fonction *scatterplot* de R.



Nous avons représenté ici la 5^e année des simulations, mais nous pourrions le faire pour les 100 années. L'allure de la courbe est souvent la même pour toute les années mais il pourrait être intéressant de voir la dispersion du nombre de clients pour les différentes classes de durées sur les 100 ans. C'est ce que nous affichons à la page suivante.



Nous pouvons afficher également le nombre de clients en fonction du nombre de coupures :



L'allure des graphes est toujours le même suivant les années : environ 350 000 clients qui n'ont aucune coupure à l'année, entre 50 000 et 100 000 clients qui ont une coupure à l'année et entre 5 000 et 15 000 clients qui ont deux coupures la même année.

Comme nous l'avons déjà dit précédemment, il est rare que des clients subissent trois à quatre pannes dans la même année, mais cela peut arriver parfois. Cela arrive sur des départs HTA avec une grande longueur de câbles et concerne généralement moins de 10 000 clients.

Conclusion – Perspectives.

Ce stage en entreprise au sein du Groupe Électricité de Strasbourg a été très intéressant. Ce stage m'a permis de découvrir le monde de l'entreprise et le mode de fonctionnement d'un grand groupe comme celui d'Électricité de Strasbourg.

Ce stage m'a permis également d'approfondir mes compétences au niveau du logiciel R et du tableur Excel. Il m'a permis de faire de nombreuses recherches sur différents points mathématiques et statistiques et ainsi approfondir mes connaissances, notamment au sujet des processus de Poisson.

Je me suis également rendu compte des difficultés auxquelles un statisticien peut être confrontés, et notamment le recueil des données et la mise en équation d'un problème réel.

Ce stage a été un grand enrichissement personnel que ce soit au niveau humain ou au niveau des compétences développées et je remercie encore une fois toutes les personnes qui par leurs accueils, leurs aides et leurs conseils ont participé au bon déroulement de mon stage.

Au niveau des perspectives, les résultats du chapitre 4 sont très intéressants car ils permettent de situer le critère B d'une année dans la distribution obtenue par les simulations et ainsi de vérifier si une dérive « technique » liée à un vieillissement accéléré des câbles papiers devait survenir.

Ces résultats portent sur les incidents souterrains, il serait peut être intéressant également de porter cette analyse au cas des incidents aériens.

2 Fiche d'incident réseau HTA.



es
électricité de strasbourg

FICHE D'INCIDENT RESEAU HTA

Groupes d'exploitation

Centre
 Nord
 Sud

RAPPORT SUCCINCT

N°

Chargé d'expl. :

1^{ère} saisie :

Dern. Valid. :

LOCALISATION

Déclenchement
 Ouverture
 kV
 Heure de 1^{re} réalimentation

Jour
 Date
 Heure de réalimentation générale

Commune
 Puissance installé coupée (hors postes clients) MW

Poste source
 N° cellule
 Nom du départ perturbé

Ligne principale N° ligne
 OU
 Dérivation Nom
 Supports N° de à

Défaut entre et

Fonctionnement normal :

Fusion fusibles :
 Extérieurs
 Intérieurs
 DLD OUI
 NON
 poste

Nombre
 Calibre A
 Compte passage OUI
 NON
 poste

Réparation :
 immédiate
 décalée

TYPE DE DEFAUT

Défaut Aérien
 Défaut Souterrain
 Défaut Poste
 Sans dégâts

DEFAUT SOUTERRAIN

Câble :
 Section mm²
 Boîte :
 Jonction
 Fonte
 Rétractable à froid

CPI à ceinture
 Cu
 OU
 Dérivation
 Injectée

CPI métallisé
 Al
 Extrémité :
 Int.
 Prémoulée
 Poinçonnage

Câble synthétique
 Ext.
 Thermorétractable
 Serrage méc.

CAUSES

Causes:

Coup de foudre
 Neige / Givre
 Tempête
 Arbre / Branche
 Oiseaux
 Autres animaux
 Corps étrangers
 Cause inconnue
 Erreur identifiée

Endommagement d'ouvrage

N° R:

Tiers:

OUI
 NON

Identifié: OUI
 NON

Installations client
 Véhicule
 Trav. abattage-élag.
 Malveillance
 Trav. terrassement
 Autres causes

Defaillance du matériel

Marque
 N° lot/Série

due à:

Vétusté
 Mise en oeuvre
 Défaut de fabrication
 Defaillance de protection
 Cause non déterminée
 Autres
 Rech. en cours

OBSERVATIONS/DETAILS

Dégâts Réseau/Matériel OUI
 NON

DETAILS DES POSTES COUPES 1/1

Incident n°: 82837
 Agence d'exploitation : GEC
 Lieu : MEINAU
 Cellule concernée : --C.27--

Début : 22/11/11 22:55:17
 Fin : 22/11/11 23:18:04
 Durée de l'incident : 00:22:47

Postes coupés : 10
 Clients coupés : 1176
 Puissance coupée : 4520

Critère B : 2.2277498
 Critère B 2 : 2.3191855

état : Acquité par Yannick Fischer (23/11/11 09:29:28)
 commentaire :

Début: 22/11/2011 22:55:17
 Fin: 22/11/2011 22:56:35
 Durée de la coupure : 00:01:18
 Postes coupés: 3
 Clients coupés: 328
 Puissance coupée: 1260
 Critère B: 0.048069738
 Critère B 2: 0.050079376

Insee	Commune	Code poste	Description	Nb clients
67482	STRASBOURG	N 001	12 C DU HEYRITZ	0
67482	STRASBOURG	N 093	C D HEYRITZ COLL PAS	158
67482	STRASBOURG	PN 076	R.D.IMPRIMEURS	158

Début: 22/11/2011 22:55:17
 Fin: 22/11/2011 23:18:04
 Durée de la coupure : 00:22:47
 Postes coupés: 7
 Clients coupés: 848
 Puissance coupée: 3260
 Critère B: 2.17968
 Critère B 2: 2.2691061

Insee	Commune	Code poste	Description	Nb clients
67482	STRASBOURG	PN 027	1,R.D.DOUBS	158
67482	STRASBOURG	PN 077	R DOUBS	250
67482	STRASBOURG	PS 047	SEITA 12,R.DOUBS	0
67482	STRASBOURG	PS 048	ORSA 12 R D DOUBS	0
67482	STRASBOURG	PS 055	GDS 10 R DU DOUBS	0
67482	STRASBOURG	PS 056	R.DU.DOUBS AUTOTR. S	0
67482	STRASBOURG	PS 094	8 R D DOUBS	158

3 Statistiques descriptives – Résultats de la console R.

```
> summary(statdescript)
```

Type	Secteur	Classe.commune	Poste.source	Postes.coupés
HTA:548	GEC:309	1. < 1000 :144	PORT DU RHIN: 48	Min. : 0.00
HTB: 61	GEN:175	2. [1000;3000[:138	MEINAU : 32	1st Qu.: 9.00
	GES:125	3. [3000;21000[:147	BRUMATH : 28	Median : 18.00
		4. > 21000 :180	GRAFFENST : 28	Mean : 27.97
			OBERNAI : 26	3rd Qu.: 29.00
			HOLZMATT : 25	Max. :319.00
			(Other) :422	

Clients.coupés	Critère.B
Min. : 0	Min. : 0.000
1st Qu.: 691	1st Qu.: 0.390
Median : 1365	Median : 2.070
Mean : 1952	Mean : 2.837
3rd Qu.: 2145	3rd Qu.: 3.980
Max. :22164	Max. :28.710

	Défaut	Cause
Souterrain	:296	Vétusté :164
Aérien	:117	Cause inconnue : 99
Poste	: 87	Défaillance matériel: 68
Sans dégats	: 46	Terrassement : 61
Réseau HTA	: 22	Non renseignée : 52
Réseau AMONT définitif (reprise manuelle):	20	Coup de foudre : 43
(Other)	: 21	(Other) :122

	Cause.2	Puissance.coupée
Cause inconnue	: 99	Min. : 0
Distributeur	:259	1st Qu.: 3360
Éléments Naturels:	115	Median : 7100
Non renseignée	: 52	Mean : 10155
Tiers	: 84	3rd Qu.: 11100
		Max. :107525

4 Seuil critique $c_n = \sqrt{n} * d_n$ du test de Kolmogorov-Smirnov pour une distribution exponentielle de moyenne inconnue.

n	c_n	n	c_n
2	0,8673	22	1,0511
3	0,9540	24	1,0531
4	0,9687	26	1,0549
5	0,9884	28	1,0565
6	1,0007	30	1,0580
7	1,0084	35	1,0609
8	1,0153	40	1,0633
9	1,0212	45	1,0652
10	1,0258	50	1,0668
12	1,0327	60	1,0694
14	1,0381	70	1,0714
16	1,0424	80	1,0729
18	1,0458	90	1,0742
20	1,0486	100	1,0753

5 Programme R utilisé pour obtenir la figure 17, page 25.

```
> bootstrap <- function(i, I) {
+   taille <- 1:i
+   quant <- 1:i
+   par <- runif(1, 0.1, 10)
+   M_sim <- NULL
+   for (n in 1:i) {
+     N_res <- NULL
+     x_sim <- 1:I
+     x_sim <- rexp(I, par)
+     M_sim <- rbind(M_sim, x_sim)
+     for (j in 1:I) {
+       stat <- sqrt(n) * ks.test(M_sim[, j], "pexp", 1/mean(M_sim[,
+       j]))$statistic
+       N_res <- c(N_res, stat)
+     }
+     quant[n] <- quantile(N_res, 0.95)
+   }
+   plot(taille, quant, xlab = "n", ylab = "n^(1/2)*d_n", type = "l")
+ }
> bootstrap(200, 10000)
```

6 Programme R utilisé pour obtenir les simulations des pannes souterraines.

```
> lambda <- read.csv("lambda.csv", sep = ";", dec = ",", header = TRUE)
> delta_t <- matrix(rep(NA, 25000), ncol = 40)
> delta_j <- matrix(rep(NA, 25000), ncol = 40)
> d_j <- matrix(rep(NA, 25000), ncol = 40)
> for (i in 1:625) {
+   delta_t[i, ] <- rexp(40, lambda[i, ])
+   delta_j[i, ] <- delta_t[i, ] * 365
+   d_j[i, ] <- cumsum(delta_j[i, ])
+ }
```

7 Analyse Factorielle de Données Mixtes.

Lorsque nous souhaitons réaliser une analyse factorielle sur des données qualitatives, nous utilisons l'Analyse des Correspondances Multiples (ACM). Pour une analyse factorielle sur des données quantitatives, nous avons à notre disposition l'Analyse en Composantes Principales (ACP). Mais ici nous avons à la fois des variables qualitatives et quantitatives. Nous devons donc utiliser une méthodologie factorielle qui permet d'inclure les deux types de variables en tant qu'éléments actifs : l'Analyse Factorielle de Données Mixtes (AFDM).

L'AFDM est un problème fréquent. À partir d'anciens travaux d'Escofier (1979) et Saporta (1990), proposant d'inclure des données mixtes dans le cadre de l'ACM (pour Escofier) et dans le cadre de l'ACP (pour Saporta), Jérôme PAGÈS a proposé une méthode qui prend en compte les variables quantitatives comme une ACP normée et les variables qualitatives comme une ACM. Elle fournit une représentation simultanée des deux types de variables en plus des représentations usuelles de l'ACP et de l'ACM. Pour plus de détails, nous nous reporterons à l'article [6].

L'AFDM est implémentée dans le package « Factominer » du logiciel R et c'est ce que nous utilisons dans la suite.

Les variables sur lesquelles nous réalisons l'AFDM sont les mêmes que celles de la section 2.1, page 12. Nous présentons ici l'AFDM réalisée avec comme seules variables actives : le nombre de postes coupés, le nombre de clients coupés, le critère B, la puissance coupée (4 variables quantitatives) et le nombre de clients de la commune coupée (variable qualitative nommée « Classe.commune »).

L'AFDM avec les 10 variables regroupées n'est pas intéressante car elle n'explique que 12% de l'inertie totale et n'est pas visuellement interprétable car il y a beaucoup de modalités pour chaque variable.

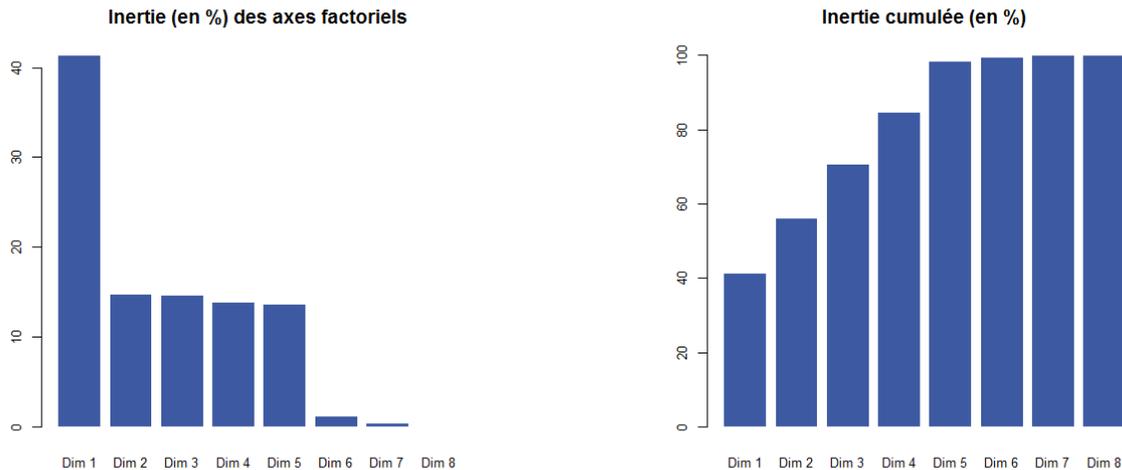
Valeurs propres.

Les valeurs propres, pourcentages de variance expliquée et cumul des pourcentages de variance expliquée sont donnés ci-dessous :

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	2.895764e+00	4.136805e+01	41.36805
comp 2	1.032452e+00	1.474931e+01	56.11736
comp 3	1.025579e+00	1.465113e+01	70.76849
comp 4	9.755366e-01	1.393624e+01	84.70473
comp 5	9.578743e-01	1.368392e+01	98.38865
comp 6	8.138952e-02	1.162707e+00	99.55135
comp 7	3.140531e-02	4.486473e-01	100.00000
comp 8	2.871112e-31	4.101588e-30	100.00000

Nous constatons que le plan factoriel (1,2) explique quasiment 60% de l'inertie totale.

Nous pouvons représenter par un histogramme la décroissance de l'inertie expliquée et également le cumul de l'inertie expliquée.



Nous pouvons sélectionner un nombre d'axes à garder suivant le *critère de Kaiser* ou le *critère du « coude »* :

- Critère de Kaiser : retenir les axes dont l'inertie est supérieure à l'inertie moyenne.

Valeur propre moyenne = $\frac{7}{8} = 0,875$. Nous retenons donc 5 axes d'après les valeurs propres de la page précédente.

- Critère du « coude » : déceler un changement de pente sur le diagramme en bâtons de la décroissance de l'inertie des axes. Nous sélectionnons les axes avant le « coude » : 5 axes.

Néanmoins, nous ne donnerons les représentations graphiques uniquement pour les axes 1 et 2 pour ne pas surcharger l'analyse.

Résumé.

Nous donnons dans l'ordre suivant, Lg, RV, coordonnées, contributions, qualités de représentation et corrélations de l'analyse.

- *Lg* : Les coefficients Lg de liaison permettent de mesurer à quel point les variables sont liées deux à deux.

	Postes.coupés	Clients.coupés	Critère.B	Puissance.coupée	Classe.commune	MFA
Postes.coupés	1.000000000	0.75958849	0.002007065	0.892848666	0.061996282	0.9380739
Clients.coupés	0.759588492	1.000000000	0.026490722	0.886639691	0.016857395	0.9287968
Critère.B	0.002007065	0.02649072	1.000000000	0.007970993	0.001827294	0.3585569
Puissance.coupés	0.892848666	0.88663969	0.007970993	1.000000000	0.033784268	0.9742658
Classe.commune	0.061996282	0.01685740	0.001827294	0.033784268	3.000000000	1.0755246
MFA	0.938073919	0.92879685	0.358556893	0.974265791	1.075524609	1.4763697

- *RV* : Les coefficients RV de liaison correspondent à une normalisation des coefficients Lg. La valeur des RV est donc comprise entre 0 et 1, ce qui en facilite l'interprétation. Nous pouvons observer ici que les variables les plus liées sont les postes coupés, les clients coupés et la puissance coupée.

	Postes.coupés	Clients.coupés	Critère.B	Puissance.coupée	Classe.commune	MFA
Postes.coupés	1.000000000	0.759588492	0.002007065	0.892848666	0.035793570	0.7720395
Clients.coupés	0.759588492	1.000000000	0.026490722	0.886639691	0.009732622	0.7644044
Critère.B	0.002007065	0.026490722	1.000000000	0.007970993	0.001054989	0.2950941
Puissance.coupée	0.892848666	0.886639691	0.007970993	1.000000000	0.019505356	0.8018255
Classe.commune	0.035793570	0.009732622	0.001054989	0.019505356	1.000000000	0.5110485
MFA	0.772039452	0.764404378	0.295094088	0.801825541	0.511048546	1.0000000

- *Coordonnées* : Les coordonnées des variables sur un axe sont des mesures de liaison entre l'axe et les variables. Ces coordonnées sont utilisées pour créer le graphe des variables.

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Postes.coupés	0.92726064	0.009109487	0.008352648	0.0006917636	0.009674932
Clients.coupés	0.91214371	0.022889019	0.004318770	0.0104161392	0.007067467
Critère.B	0.02385803	0.028765834	0.422323018	0.2314397875	0.292885086
Puissance.coupée	0.97132257	0.001216555	0.000927981	0.0000638312	0.005851357
Classe.commune	0.06117886	0.970470717	0.589656431	0.7329250686	0.642395470

- *Contributions* : Décrit l'importance que prend une variable dans la construction d'un axe.

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Postes.coupés	32.0212800	0.8823161	0.81443255	0.070911087	1.0100419
Clients.coupés	31.4992441	2.2169580	0.42110560	1.067734343	0.7378283
Critère.B	0.8238943	2.7861677	41.17899066	23.724357433	30.5765675
Puissance.coupée	33.5428796	0.1178316	0.09048364	0.006543189	0.6108690
Classe.commune	2.1127021	93.9967265	57.49498756	75.130453948	67.0646933

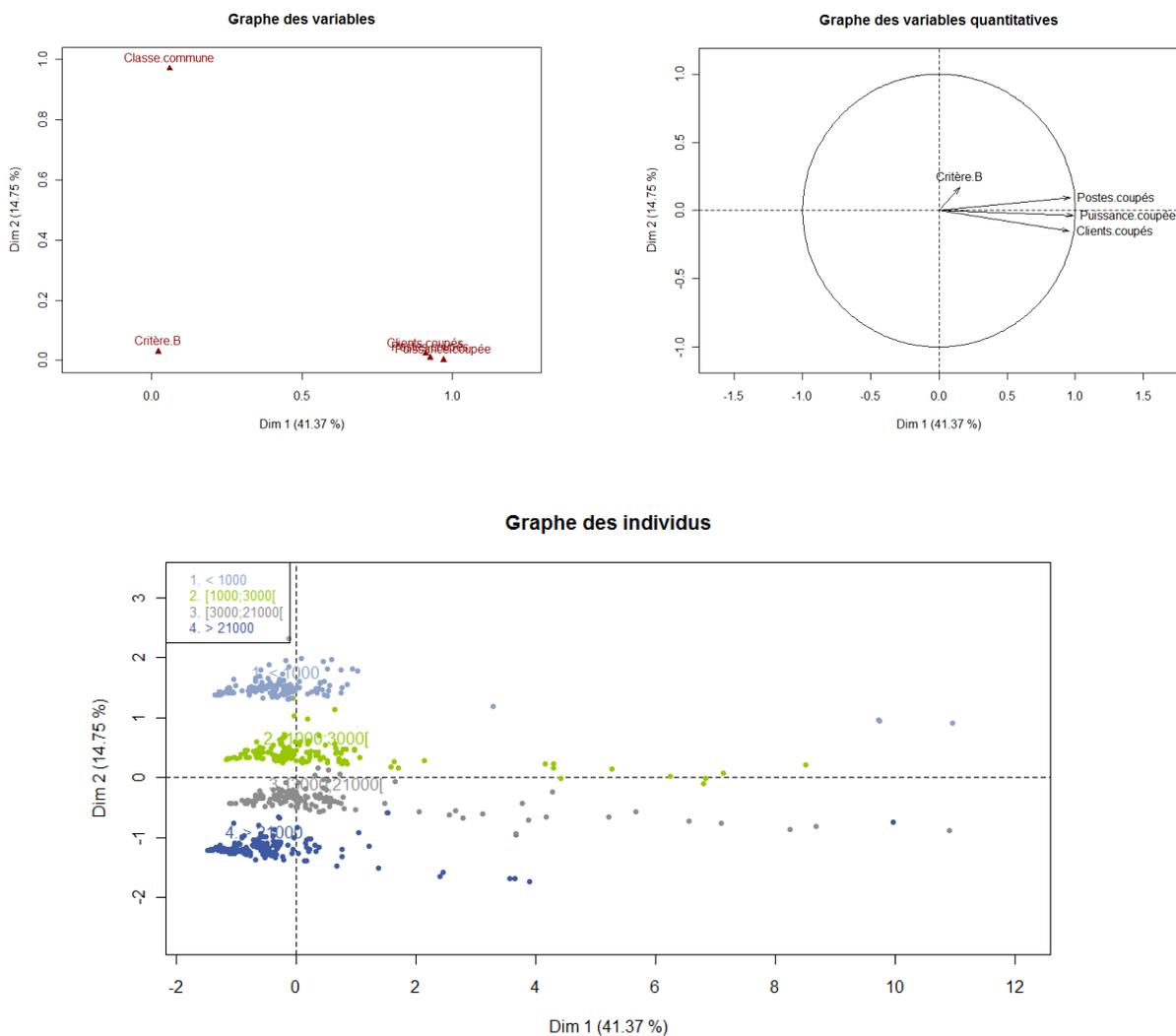
- *Qualités de représentation* : La qualité de représentation est mesuré par le « \cos^2 ». C'est une mesure de proximité entre un point et le plan (mesure de l'angle formé avec le plan).

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Postes.coupés	0.8598122886	8.298275e-05	6.976673e-05	4.785369e-07	9.360431e-05
Clients.coupés	0.8320061503	5.239072e-04	1.865177e-05	1.084960e-04	4.994909e-05
Critère.B	0.0005692057	8.274732e-04	1.783567e-01	5.356438e-02	8.578167e-02
Puissance.coupée	0.9434675313	1.480005e-06	8.611488e-07	4.074423e-09	3.423838e-05
Classe.commune	0.0012476177	3.139378e-01	1.158982e-01	1.790597e-01	1.375573e-01

- *Corrélations* : Pour les variables quantitatives, cela correspond aux corrélations avec un axe. Concernant les variables qualitatives, il s'agit du rapport de corrélation qui est une mesure de liaison entre un axe et une variable. Si le rapport de corrélation est élevé, cela veut dire que les modalités de la variable forment autant de sous-population homogènes et bien séparées le long de l'axe. C'est le cas, ici, nous nous attendons à ce que la variable « Classe.commune » ait des modalités bien séparées le long de l'axe 2.

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Postes.coupés	0.9629437	0.09544363	0.09139282	0.026301399	0.09836123
Clients.coupés	0.9550622	0.15129117	0.06571735	0.102059489	0.08406823
Critère.B	0.1544605	0.16960493	0.64986385	0.481081893	0.54118859
Puissance.coupée	0.9855570	0.03487915	0.03046278	0.007989443	0.07649416
Classe.commune	0.2473436	0.98512472	0.76789090	0.856110430	0.80149577

Représentations graphiques.



Nous pouvons voir sur le graphe des variables que le premier axe factoriel est fortement lié aux variables « Postes.coupés », « Clients.coupés » et « Puissance.coupée ». Quant au second axe factoriel, il est très lié à la variable « Classe.commune ».

Le graphe des variables quantitatives nous confirme que les trois variables « Postes.coupés », « Clients.coupés » et « Puissance.coupée » sont liées entre elles et nous informe sur le fait que la variable « Critère.B » n'est pas bien représentée car elle n'est pas à proximité du bord du cercle des corrélations.

Sur le graphe des individus, les incidents survenus dans les communes de plus de 3 000 clients sont dans la partie négative du second axe, alors que les incidents dans les plus petites communes sont dans la partie positive du second axe.

De plus, nous observons que la variable « Classe.commune » a des modalités qui forment bien des sous-populations homogènes puisqu'elles sont bien séparées le long du deuxième axe factoriel.

Références.

- [1] « *Analyses statistiques des évènements de coupures sur les réseaux électriques* », Adrien BARRET.
Rapport de stage, ENSAI Rennes, 2011.
- [2] « *Processus de Poisson* », Christel RUWET.
Mémoire, Université de Liège, 2007.
- [3] « *Kolmogorov-Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests on spacing* », J. DURBIN.
Biometrika, Vol 62, No. 1 (Apr., 1975), pp. 5-22.
- [4] « *Monte Carlo Exact Goodness-of-fit Tests for Nonhomogeneous Poisson Processes* », BO H. LINDQVIST et BJARTE RANNESTAD.
Norwegian University of Science and Technology, Department of Mathematical Sciences.
- [5] « *Méthode de détermination du nombre de clients d'Électricité de Strasbourg mal alimentés en tension.* », Mélanie CHOLET.
Rapport de Stage, Université de Strasbourg, 2009.
- [6] « *Analyse Factorielle de Données Mixtes.* », J. PAGÈS.
Revue de statistique appliquée, tome 52, n° 4 (2004), p. 93-111.