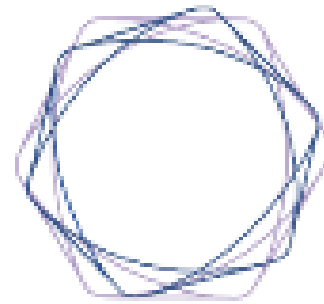


Modélisation des taux de défaillance par composante des départs Basse Tension

B. Ravel – Datastorm

O. Faivre – ENEDIS



DataStorm



ENEDIS
L'ELECTRICITE EN RESEAU

Contexte

- ENEDIS, entreprise de service public, gère la presque totalité des réseaux de distribution d'électricité en France métropolitaine.
- Ceci inclut la maintenance de l'ensemble des équipements du réseau allant des postes sources (liens avec le réseau de Transport géré par RTE) jusqu'aux utilisateurs finaux (consommateurs ou producteurs).
- Les départs Basse Tension constituent la partie du réseau en sortie des postes de transformations et alimentent :
 - Jusqu'à 150 maisons individuelles
 - Jusqu'à 300 logements en collectifs
- La fiabilité de ces départs est en enjeu majeur pour la qualité de service du réseau électrique.

La fiabilité des départs Basse Tension

- Un taux moyen de défaillance annuel de 0,6%
- Des causes de défaillances multiples
 - Travaux
 - Ancienneté des câbles
 - Défaillance de matériel
- Un modèle probabiliste de défaillance sur le départ qui a permis d'optimiser la politique de renouvellement en ciblant les départs à risque.
- Mais un modèle qui ne fournit que des indications au niveau de l'ensemble du départ

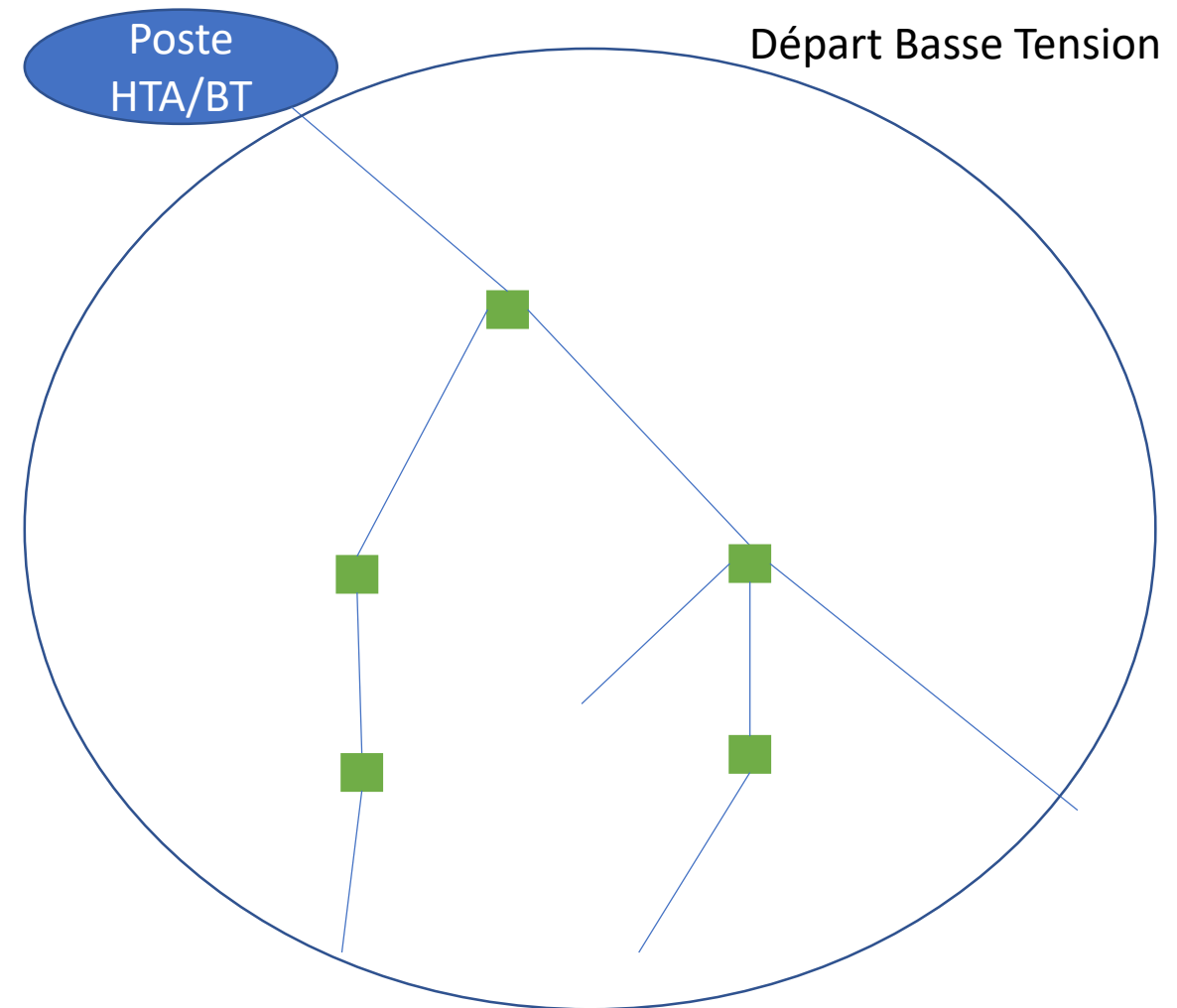
Le départ Basse Tension est un système composite

Un départ est composé de tronçons de câbles électriques et de boîtiers de raccordements

Ces composants ont chacun des caractéristiques propres (taille, technologie, etc.)

La défaillance de l'ensemble du départ peut provenir de n'importe quel composant de cet ensemble

Mais cette information n'est pas collectée dans les bases historiques



La question à résoudre

- Au-delà du ciblage des départs, quels composants faut-il cibler pour le renouvellement au sein d'un départ ?
- Sachant que :
 - Seule la défaillance globale du départ est connue
 - Que la défaillance du départ est équivalente à la défaillance d'au moins un de ces composants (montage en série)
- Peut-on modéliser la fiabilité des composants sans l'avoir observée mais en exploitant la diversité des configurations des départs ?

Les données disponibles

- Une base de données sur 2 millions de départs comprenant :
 - La composition du départ en termes de tronçons et de boitiers,
 - La technologie et la longueur de chaque tronçon
 - L'âge des boitiers de raccordement
- 6 années d'observation de défaillances sur les départs mais avec la seule connaissance de la défaillance au niveau de l'ensemble du départ.
- Parmi toutes les variables disponibles, les expertises métier ont conduit à des recodages.

Les données exploitées

■ Concernant les tronçons :

- Leur longueur, en mètres (l)
- Leur technologie, CPI ou Synthétique (X_1, X_2)

■ Concernant les boitiers :

- leur âge en tranche en 5 classes (1946, Avant 1975, 76 à 86, 86 à 2002 et post 2002) : A_1, A_2, A_3, A_4, A_5
- leur degré en 3 classes : deux tronçons raccordés, 3 tronçons raccordés, 4 ou plus tronçons raccordés. On note ces classes : D_1, D_2, D_3
- son type selon le type de tronçons que le boitier raccorde en 2 classes hétérogène ou homogène : H_1, H_2

Le modèle mis en œuvre (1/2)

- 2 grands partis pour cette modélisation :
 - Se baser sur des modèles à risques proportionnels pour les composants,
 - Considérer des composants sans vieillissement
- Le premier choix est motivé par le fait que les modèles à risques proportionnels permettent de facilement comprendre l'impact des covariables sur les défaillances.
- Le second choix est pragmatique : l'introduction de lois intégrant des facteurs de forme liés au vieillissement aurait considérablement complexifié les calculs pour cette première expérimentation. C'est en plus cohérent avec l'expérience.

$$S(t, \mathbf{X}) = \prod_p e^{-\lambda_{0,p} t} e^{-\theta_p \cdot X_p}$$

Fonction de survie conditionnelle

Produit des fonctions de survie conditionnelles des composants

Modèle à risque proportionnel

Le modèle mis en œuvre (1/2)

- 2 grands partis pour cette modélisation :
 - Se baser sur des modèles à risques proportionnels pour les composants,
 - Considérer des composants sans vieillissement
- Le premier choix est motivé par le fait que les modèles à risques proportionnels permettent de facilement comprendre l'impact des covariables sur les défaillances.
- Le second choix est pragmatique : l'introduction de lois intégrant des facteurs de forme liés au vieillissement aurait considérablement complexifié les calculs pour cette première expérimentation. C'est en plus cohérent avec l'expérience.

$$S(t, \mathbf{X}) = \prod_p e^{-\lambda_{0,p} t} e^{-\theta_p \cdot X_p}$$

Fonction de survie conditionnelle

Produit des fonctions de survie conditionnelles des composants

Modèle à risque proportionnel

Le modèle mis en œuvre (2/2)

- Equation de survie des tronçons :

- $S_{TR,i}(t) = e^{-\lambda_{TR}t} e^{-\theta_1 X_{1,i}}$

- Equation des survie des départs :

- $S_{BT,j}(t) = e^{-\lambda_{BT}t} e^{-[(\alpha_1 A_{1,j} + \alpha_2 A_{2,j} + \alpha_3 A_{3,j} + \alpha_4 A_{4,j}) + (\beta_1 D_{1,j} + \beta_2 D_{2,j}) + (\gamma_1 H_{1,j})]}$

- Equation de survie des départs :

- $S_k(t) = \prod_{\substack{\text{tronçons} \\ i \text{ du départ } k}} e^{-\lambda_{TR}t} e^{-\theta_1 X_{1,i}} \prod_{\substack{\text{boitiers} \\ j \text{ du départ } k}} e^{-\lambda_{BT}t} e^{-[(\alpha_1 A_{1,j} + \alpha_2 A_{2,j} + \alpha_3 A_{3,j} + \alpha_4 A_{4,j}) + (\beta_1 D_{1,j} + \beta_2 D_{2,j}) + (\gamma_1 H_{1,j})]}$

→ 10 paramètres à estimer (2 facteurs de risque Lambda, 1 cofacteur pour les tronçons et 7 pour les boitiers)

Comment estimer ces modèles ?

- Pas de librairie standard pour l'estimation de telles fonctions.
- Estimation par le maximum de Vraisemblance sur les échantillons d'apprentissage :
 - Vraisemblance = Produit
 - Des densités aux départs défaillants
 - Des fonctions de survie aux départs non défaillants (donnée censurée)

▪ La densité étant fournie par la dérivée de la fonction de survie ($f = -\frac{dS}{dt}$)

▪ Le système à résoudre, après passage au log est donc de trouver le maximum de la fonction suivante à l'aide de nos échantillons :

$$\bullet V(\lambda_{TR}, \lambda_{BT}, \theta_1, \alpha_1, \dots, \gamma_1) = \sum_{k \text{ défaillants}} \text{départs} [\log(C_k) - C_k t_k] - \sum_{l \text{ vivants}} \text{départs} [C_l t_k]$$

• Avec

$$\bullet C_k = \lambda_{TR} \sum_{i \text{ du départ } k} \text{tronçons} e^{-\theta_1 X_{1,i}} + \lambda_{BT} \sum_{j \text{ du départ } k} \text{boitiers} e^{-[(\alpha_1 A_{1,j} + \alpha_2 A_{2,j} + \alpha_3 A_{3,j} + \alpha_4 A_{4,j}) + (\beta_1 D_{1,j} + \beta_2 D_{2,j}) + (\gamma_1 H_{1,j})]}$$

L'optimisation

- La recherche du maximum est passée par l'utilisation de NLOPT, package qui fait l'interface vers la librairie libre NLOPT qui offre de nombreuses fonctions d'optimisations.
- La préparation des données pour faciliter le codage de la fonction à évaluer à chaque itération a été faite à l'aide du package DATATABLE
- Le programme d'optimisation a été itéré dans de nombreux cas d'applications différents selon le choix des paramètres activés ou pas dans le modèle exposé.
- Une application Shiny a été mise en œuvre pour exploiter les résultats des estimations (plus de 800 modèles estimés).

Résultats

mono-tronçons all-tronçons Boitiers - 1v Boitiers - 2v Boitiers - 3v **Tr - Boitiers - 1v** Full

Onglet de choix de la famille de modèles

modèles Tr Tr_type_lg Tr_lg

modèles Bt Bt_degree Bt_age

age age_max

Puissance TR FALSE TRUE

Puissance BT FALSE

Filtres pour choisir les modèles

Show 10 entries

models_tr	models_bt	age	puis_tr	puis_bt	data	λ_{tr}	λ_{bt}	X_cpi	X_Synth	H_hete	H_homo	D_3p	D_2	A_0_1946	A_1_pre75	A_2_76_86	A_4_post2003	A_3_87_2002	P_puis_tr	P_puis_bt		
Tr	Bt	age_max	false	false	2008	5.86e-05	1.00e-08															
Tr	Bt	age_max	false	false	2009	6.46e-05	1.00e-08															
Tr	Bt	age_max	true	false	2008	3.33e-05	1.00e-08														-2.84e+00	
Tr	Bt	age_max	true	false	2009	3.63e-05	2.14e-06															-2.83e+00
Tr	Bt_age	age_max	false	false	2008	4.15e-05	3.64e-05							1.59e+01	-5.87e-01	2.50e-01	-7.77e-01	0e+00				
Tr	Bt_age	age_max	false	false	2009	4.88e-05	3.17e-05							1.38e+01	-7.09e-01	4.99e-01	-8.09e-01	0e+00				
Tr	Bt_age	age_max	true	false	2008	1.39e-05	4.32e-05							1.43e+01	-2.04e-01	2.82e-01	-7.05e-01	0e+00				-3.76e+00
Tr	Bt_age	age_max	true	false	2009	1.93e-05	3.75e-05							1.34e+01	-3.16e-01	3.55e-01	-7.86e-01	0e+00				-3.50e+00
Tr	Bt_degree	age_max	false	false	2008	5.45e-05	1.04e-05					1.15e+01	0e+00									
Tr	Bt_degree	age_max	false	false	2009	5.83e-05	1.57e-05					1.20e+01	0e+00									

Showing 1 to 10 of 36 entries

Résultats de l'estimation (au survol, $\frac{1}{\lambda}$ ou e^{-coef} est affiché)

Années

stats

Show 10 entries

Statistique de confusion sur des échantillons d'application

models_tr	models_bt	age	puis_tr	puis_bt	data	0_Q5	0_Q4	0_Q3	0_Q2	0_Q1	1_Q5	1_Q4	1_Q3	1_Q2	1_Q1	AUC	W_AUC
Tr	Bt	age_max	false	false	2008	19.85%	19.88%	19.90%	20.17%	20.20%	33.77%	23.82%	15.03%	14.69%	12.69%	0.61	
Tr	Bt	age_max	false	false	2009	19.85%	19.88%	19.90%	20.17%	20.20%	33.77%	23.82%	15.03%	14.69%	12.69%	0.61	
Tr	Bt	age_max	true	false	2008	19.78%	19.97%	20.04%	20.08%	20.13%	41.39%	22.90%	16.12%	12.04%	7.54%	0.67	
Tr	Bt	age_max	true	false	2009	19.79%	19.97%	20.04%	20.07%	20.13%	41.28%	22.90%	16.15%	12.09%	7.58%	0.67	

Synthèse des conclusions

- Validation par la pratique de la méthode d'estimation de la durée de vie de composants de systèmes montés en série à l'aide des seules observations de la défaillance de l'ensemble du système
- Démonstration de la possibilité de réaliser de telles estimations dans des temps raisonnables sur une base de données très volumineuse (quelques heures au maximum)
- Des résultats chiffrés qui offrent des compléments aux modèles probabilistes de défaillance des départs en permettant de mieux cibler les composants à remplacer sur les départs supposés à risque