



HAL
open science

Étude factuelle du pronostic à l'aide d'une base de données de maintenance existante

Jean-François Boudreau, Normand Amyot, Stéphane Plourde, Mario Céré

► **To cite this version:**

Jean-François Boudreau, Normand Amyot, Stéphane Plourde, Mario Céré. Étude factuelle du pronostic à l'aide d'une base de données de maintenance existante. 2011. hal-00616517

HAL Id: hal-00616517

<https://hal.science/hal-00616517>

Submitted on 22 Aug 2011

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.

Étude factuelle du pronostic à l'aide d'une base de données de maintenance existante

Jean-François Boudreau¹, Normand Amyot¹, Stéphane Plourde² et Mario Céré³

1. *Institut de recherche d'Hydro-Québec, Varennes (Québec) J3X 1S1, Canada*
boudreau.jean-francois@ireq.ca
amyot.normand@ireq.ca

2. *Hydro-Québec TransÉnergie, Montréal (Québec) H5B 1H7, Canada*
plourde.stephane@hydro.qc.ca

3. *Hydro-Québec TransÉnergie, Saint-Hubert (Québec) J3Z 1G3, Canada*
cere.mario@hydro.qc.ca

RÉSUMÉ

Une méthode éprouvée d'estimation de la date de mise au rebut d'un appareil électrique au moyen d'une base de données de maintenance est présentée. Trois types de renseignements contenus dans une telle base de données ont été utilisés : les bons de travail, l'historique des déplacements et la date de la dernière mise à jour de certains des champs de la base. La date de mise au rebut des différents appareils d'une famille peut par la suite être mise à profit dans l'étude de leur fiabilité.

Les auteurs ont choisi les inductances à noyau d'air pour illustrer le processus d'estimation de la date de mise au rebut, ainsi que l'ajustement de différents modèles paramétriques et non paramétriques. Il a ainsi été établi que la fiabilité des inductances à noyau d'air fabriquées à partir de 1982 est supérieure à celle des appareils fabriqués antérieurement.

Cette méthode d'estimation de la date de mise au rebut pourrait être appliquée aux équipements majeurs. Dans ces cas, les diagnostics disponibles pourraient être utilisés comme prédicteurs dans les modèles de fiabilité.

En plus du modèle de temps de défaillance accéléré présenté, l'estimation des dates de mise au rebut permet d'utiliser différents modèles de fiabilité. Le modèle de Cox, ou modèle des risques proportionnels, pourrait par exemple être utilisé pour étudier les liens entre différents

prédicteurs et la fiabilité sans avoir à avancer d'hypothèse sur la distribution du temps avant un événement donné.

1. INTRODUCTION

Hydro-Québec TransÉnergie utilise des projections des coûts de remplacement et de différents facteurs de risque pour planifier ses investissements en matière de grands équipements électriques. Le principal facteur de risque utilisé est le risque moyen d'un groupe d'appareils, où le risque couru par un appareil est égal au produit de sa probabilité de mise au rebut et de son impact sur le réseau. Des courbes de fiabilité, ou les courbes du taux de risque correspondantes, permettent de quantifier la probabilité de mise au rebut en fonction du temps écoulé depuis la mise en service. Ces courbes sont établies pour les différentes familles d'appareils à l'aide des données historiques contenues dans la base de données de maintenance.

Dans un deuxième temps, une analyse détaillée est effectuée afin de choisir les appareils à remplacer. Une matrice de risque construite à l'aide de l'indice d'état des appareils et leur impact sur le réseau est utilisée à cette fin. L'indice d'état est une moyenne pondérée de différents facteurs de dégradation : âge de l'appareil, nombre de manœuvres, charge effective, teneur en gaz dissous, etc.

La méthode actuelle d'estimation des courbes de fiabilité de la première étape ne nécessite pas la connaissance des dates des événements passés, ce qui simplifie la tâche mais induit un biais important. En effet, la fiabilité correspondant à une certaine durée écoulée depuis la mise en service est estimée par la proportion d'appareils encore en service au moment de l'analyse parmi ceux de la cohorte concernée. L'inconvénient d'une telle méthode est qu'elle utilise toute la trajectoire de fiabilité d'une cohorte d'appareils pour estimer la fiabilité à un seul moment.

L'estimation des dates de mise au rebut – ou à tout le moins d'intervalles où se situent ces dates – permettrait d'obtenir des courbes de fiabilité moins biaisées. De plus, la disponibilité des données dans un format conventionnel en fiabilité donnerait la possibilité de moduler la probabilité de mise au rebut à l'aide de l'indice d'état ou d'autres prédicteurs, ce qui reviendrait à combiner les deux étapes ci-dessus. Enfin, cette démarche ouvrirait la voie à l'utilisation des développements passés et futurs des techniques d'analyse de fiabilité et de survie.

Les auteurs exposeront tout d'abord les renseignements pertinents des bases de données de maintenance et d'inventaire d'Hydro-Québec TransÉnergie et la façon dont cette information peut être utilisée pour estimer les dates de mise au rebut. Les grandes lignes de la théorie de la fiabilité ainsi que quelques modèles seront par la suite décrits. Une étude de cas portant sur des inductances à noyau d'air sera présentée à titre d'exemple. En conclusion, certaines nouvelles applications et extensions possibles des travaux présentés seront discutées.

2. PROSPECTION DES DONNÉES

2.1 Données disponibles

La base de données d'inventaire d'Hydro-Québec TransÉnergie comprend, pour tous les appareils, une description, la date de mise en service et les codes d'utilisation. Les deux principaux codes d'utilisation sont « en service » et « rebut ».

Parmi les données de la base de données de maintenance, trois des types sont particulièrement utiles pour l'estimation de la date de mise au rebut, soit : 1) les bons de travail, qui comprennent une description des activités de maintenance accomplies, le motif de création du bon ainsi que le temps, le coût et la date d'exécution ; 2) l'historique des déplacements, qui indique le lieu d'origine, la destination et la date de chaque déplacement ; 3) la date de la dernière mise à jour de différents champs de la base générées automatiquement.

2.2 Utilisation des données disponibles

Le code d'utilisation courant indique si un appareil donné est en service ou s'il a été mis au rebut. Le suivi des appareils commence à la date de mise en service. Si ce champ ne contient que l'année, la date de mise en service est assimilée au 30 juin dans le cas d'un appareil encore en service, et à la période entre le 1^{er} janvier et le 31 décembre dans le cas d'un appareil mis au rebut. Dans le second cas, la date de l'événement est donc censurée par intervalle.

Si un appareil est toujours en service au moment de l'extraction des données, le suivi prend fin à cette date. Le cas échéant, la date d'événement est censurée à droite. Pour chaque appareil mis au rebut, le suivi prend fin à la date de la mise au rebut, celle-ci étant estimée à l'aide des trois types de données énumérés à la section 2.1 et de l'algorithme suivant.

Le dernier bon de travail est tout d'abord examiné afin de déterminer si ce bon concerne la mise au rebut. Dans l'affirmative, la date de mise au rebut est estimée par la date de ce bon de travail.

Si la base de données ne contient pas un tel bon de travail, l'historique des déplacements est examiné. Si celui-ci contient un déplacement d'un lieu « non rebut » à un lieu « rebut », la date de mise au rebut est estimée par la date la plus récente entre celle de ce déplacement et celle de la dernière modification du code d'utilisation.

Enfin, si un tel déplacement n'est pas documenté dans la base de données, un intervalle faisant appel à la date de la dernière modification du code d'utilisation est établi. Étant donné que cette information est générée automatiquement lorsqu'une personne autorisée modifie ce code dans la base de données, elle est utilisée comme borne supérieure de la date de mise au rebut. Quant à la borne inférieure de l'intervalle, elle correspond à la date la plus récente entre celle du dernier bon de travail qui ne porte pas sur la mise au rebut et celle du dernier déplacement entre deux lieux « non rebut ».

Il est à noter que ce travail de prospection de données se fait manuellement puisqu'un automate ne saurait analyser les situations ambiguës et faire des choix éclairés.

3. MODÉLISATION DU PRONOSTIC

Désignons par T la variable aléatoire décrivant la date de mise au rebut et $f(t)$ sa densité de probabilité. La fonction de fiabilité s'exprime comme suit :

$$R(t) = \int_t^{\infty} f(u)du$$

alors que la fonction du taux de défaillance est la suivante :

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr[t \leq T < t + \Delta t | T \geq t]}{\Delta t}$$

Enfin, le taux de défaillance cumulatif est :

$$H(t) = \int_0^t h(u)du$$

Ces quatre fonctions de base contiennent la même information et peuvent être dérivées les unes des autres, par exemple :

$$f(t) = -dR(t)/dt$$

$$h(t) = f(t)/R(t)$$

$$R(t) = \exp(-H(t))$$

Dans un contexte classique de censure à droite et de dates d'événement exactes, l'estimateur non paramétrique Kaplan-Meier de la fonction de fiabilité est le suivant (Kaplan et Meier, 1958 ; Kalbfleisch et Prentice, 2002) :

$$\hat{R}(t) = \prod_{j|t_j \leq t} \frac{n_j - d_j}{n_j}$$

où $t_1 < t_2 < \dots < t_k$ représentent les dates de défaillance observées, n_j , le nombre d'éléments à risque juste avant t_j , et d_j , le nombre d'événements à la date t_j .

En présence de dates d'événements censurées par intervalle, les observations sont notées $t_i \in (l_i, u_i], i = 1, \dots, n$. Les intervalles $(L_j, U_j], j = 1, \dots, m$ sont définis où $L_j \in \{l_1, \dots, l_n\}$, $U_j \in \{u_1, \dots, u_n\}$ et L_j et U_j sont adjacents. L'estimateur non paramétrique de vraisemblance maximale de $R(t)$ est obtenu en maximisant (Kalbfleisch et Prentice, 2002)

$$\prod_{i=1}^n \left\{ \sum_{j=1}^m \delta_{ij} [R(U_j) - R(L_j)] \right\}$$

où $\delta_{ij} = \mathbf{1}\{(L_j, U_j] \subset (l_i, u_i]\}$, $i = 1, \dots, n$ et $j = 1, \dots, m$.

Les estimateurs non paramétriques n'exigent pas d'hypothèse forte. Par contre, l'extrapolation est difficile alors que la fonction du taux de défaillance correspondante n'est pas obtenue facilement. Ce n'est pas le cas lorsqu'un modèle paramétrique est utilisé. Par exemple, les fonctions de base de la loi de Weibull sont les suivantes :

$$f(t) = \alpha \lambda t^{\alpha-1} \exp(-\lambda t^\alpha)$$

$$R(t) = \exp(-\lambda t^\alpha)$$

$$h(t) = \alpha \lambda t^{\alpha-1}$$

$$H(t) = \lambda t^\alpha$$

où $\lambda > 0$ et $\alpha > 0$.

Deux fonctions sont étudiées afin d'évaluer l'adéquation des modèles paramétriques, soit une fonction de la fiabilité et une fonction du temps. Ces fonctions sont choisies de telle façon que leur relation soit linéaire si le modèle paramétrique posé est vrai. Par exemple, dans le cas de la loi de Weibull, un graphique de

$$\ln\{-\ln[\hat{R}(t)]\} \text{ vs. } \ln(t)$$

devrait prendre la forme d'une droite (Klein et Moeschberger, 1997 ; Kalbfleisch et Prentice, 2002).

Jusqu'ici, l'intérêt a porté sur des populations homogènes. Des techniques d'analyse de régression sont utilisées pour l'étude de populations hétérogènes ou du lien entre la fiabilité et différents prédicteurs. Par exemple, la transformation de T à l'aide du logarithme naturel produit une variable dont le support est l'ensemble des nombres réels. Un modèle linéaire est assumé pour $Y = \ln(T)$:

$$Y = \mu + \boldsymbol{\gamma}^T \mathbf{Z} + \sigma W$$

où $\boldsymbol{\gamma}^T$ est le vecteur de coefficients de régression, \mathbf{Z} est le vecteur des variables explicatives, et W est la distribution de l'erreur. Ce modèle s'appelle le modèle de temps de défaillance accéléré (*accelerated failure time model*). Un choix populaire pour W est la loi des valeurs extrêmes, ce qui correspond à un modèle de régression de Weibull pour Y . Lorsque des dates censurées par intervalle sont utilisées, les paramètres de ce modèle peuvent facilement être estimés à l'aide de différents logiciels de statistique courants.

4. ÉTUDE DE CAS : LES INDUCTANCES À NOYAU D'AIR

4.1 Données et suivi

De 1967 à 2010, Hydro-Québec TransÉnergie a mis en service 662 inductances à noyau d'air de tension nominale égale ou supérieure à 120 kV. De ce nombre, 56 inductances avaient été mises au rebut avant le 31 janvier 2011.

La date de mise au rebut a été estimée à l'aide du dernier bon de travail consacré à la mise au rebut dans le cas de huit appareils, et à l'aide d'un déplacement vers un lieu « rebut » dans 41 des cas restants. Quant aux sept derniers cas, la date de mise au rebut a été estimée par un intervalle tel que décrit à la section 2.2.

La date de la mise en service ou de la mise au rebut ayant été estimée sous forme d'un intervalle pour les 56 appareils mis au rebut, elle est censurée par intervalle pour tous ces appareils. En supposant que l'événement est survenu au milieu de l'intervalle, les 56 inductances totalisent 974 machines-années (suivi compris entre 0,4 et 27,1 machines-années). Les 606 appareils encore en service totalisent, quant à eux, 11 596 machines-années (suivi compris entre 0,3 et 38,6 machines-années).

Pour évaluer l'impact des changements apportés aux matériaux utilisés et aux méthodes de fabrication, nous comparerons les appareils fabriqués avant 1982 aux plus récents. L'année choisie pour la coupure est telle que le groupe le plus ancien comprend 51 appareils et 15 événements.

4.2 Modélisation du pronostic

La manipulation des données et les analyses ont été effectuées à l'aide du logiciel R 2.10.1 (R Foundation for Statistical Computing, 2009). Les analyses de fiabilité ont été réalisées à l'aide des fonctions *survfit* et *survreg* de la bibliothèque *survival* 2.35-7.

La figure 1 illustre la fiabilité des inductances à noyau d'air obtenue de l'estimateur Kaplan-Meier. La figure 2 utilise le même estimateur pour les deux groupes d'inductances à noyau d'air définis à l'aide de l'année de fabrication. Le second graphique montre que les appareils plus récents semblent offrir une fiabilité supérieure.

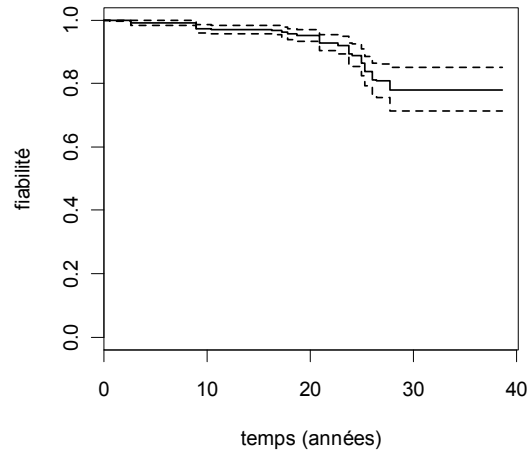


Figure 1 – Estimation Kaplan-Meier de la fiabilité des inductances à noyau d'air et intervalle de confiance à 95 %

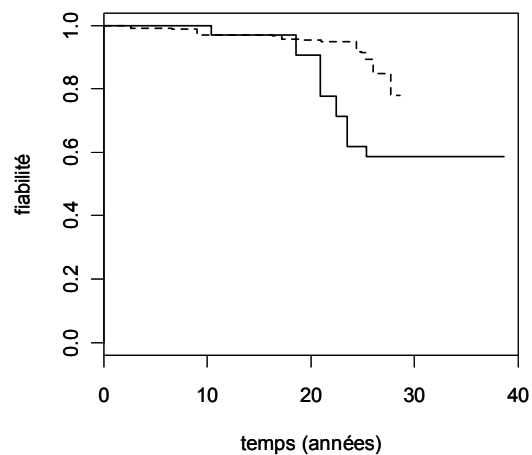


Figure 2 – Estimation Kaplan-Meier de la fiabilité des inductances à noyau d'air selon l'année de fabrication (trait continu : avant 1982 ; trait discontinu : 1982 ou plus récente)

Pour étudier formellement la relation entre la date de fabrication et la fiabilité, un modèle de temps de défaillance accéléré a été utilisé. Pour choisir la distribution appropriée, les graphiques diagnostiques des modèles de régression log-logistique, log-normale, exponentielle et de Weibull ont été étudiés. La linéarité entre les fonctions appropriées de la fiabilité et du temps a été jugée adéquate dans le cas des modèles de régression log-normale et de Weibull; le second est illustré à la figure 3.

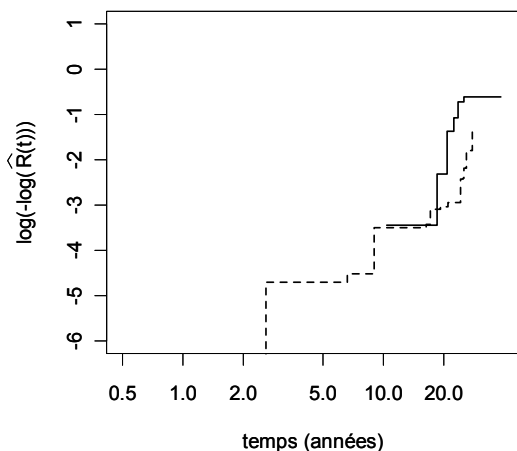


Figure 3 – Graphique diagnostique du modèle de Weibull selon l’année de fabrication (trait continu : avant 1982 ; trait discontinu : 1982 ou plus récente).

Les résultats du modèle de temps de défaillance accéléré de Weibull montrent que l’époque où l’appareil a été fabriqué est un prédicteur statistiquement significatif de la fiabilité ($p < 0,001$). Le risque relatif d’une inductance à noyau d’air fabriquée avant 1982 est 3,58 fois celui d’un appareil plus récent. De façon équivalente, le facteur d’accélération est égal à 1,92, signifiant que le temps médian avant la mise au rebut des inductances à noyau d’air fabriquées en 1982 ou plus récemment est 1,92 fois celui des appareils fabriqués avant 1982.

Enfin, les fonctions du taux de défaillance des deux groupes modélisés sont illustrées à la figure 4, où la différence entre les deux groupes est très nette.

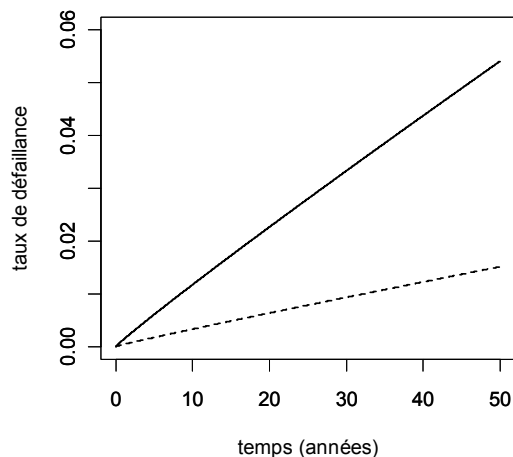


Figure 4 – Estimation paramétrique du taux de défaillance des inductances à noyau d’air basée sur le modèle de temps de défaillance accéléré de Weibull, selon l’année de fabrication (trait continu : avant 1982 ; trait discontinu : 1982 ou plus récente)

5. CONCLUSION

Les auteurs ont montré qu’il est possible d’utiliser une base de données de maintenance comme celle utilisée chez Hydro-Québec TransÉnergie pour estimer la date de mise au rebut d’appareils électriques. Cette information peut à son tour être exploitée pour la modélisation de la fiabilité dans un cadre conventionnel, tel qu’illustré avec dans le cas des inductances à noyau d’air de 120 kV et plus.

Cette étude permet d’envisager une vaste gamme d’analyses. À l’avenir, lorsque la date de mise au rebut de grands équipements tels que les transformateurs sera estimée, les diagnostics sur lesquels est fondé leur indice d’état pourraient être pris en compte dans la modélisation de leur fiabilité. Ces nouveaux modèles auront certainement une influence importante sur la sélection des appareils à remplacer.

Outre les nouveaux appareils, nous nous pencherons prochainement sur la modélisation de l’effet de la maintenance, ainsi que sur la modélisation d’événements récurrents tels que les pannes non fatales.

REMERCIEMENTS

Les auteurs aimeraient remercier Hélène Dumont pour son assistance technique avec les bases de données d’inventaire et de maintenance, ainsi que Claude Rajotte pour les renseignements qu’il a fournis au sujet de la stratégie de gestion des actifs chez Hydro-Québec TransÉnergie.

BIBLIOGRAPHIE

- KALBFLEISCH, J.D., et PRENTICE R.L. (2002). *The Statistical analysis of Failure Time Data*, 2^e éd. Hoboken, New Jersey : John Wiley & Sons, Inc.
- KAPLAN, E.L., et MEIER P. (1958). Nonparametric Estimation from Incomplete Observations, *Journal of the American Statistical Association*, vol. 53, no. 282, p. 457-481.
- KLEIN, J.P., et MOESCHBERGER M.L. (1997). *Survival Analysis Techniques for Censored and Truncated Data*, 1^e éd. New York : Springer-Verlag New York, Inc.
- R FOUNDATION FOR STATISTICAL COMPUTING (2009). R: A Language and Environment for Statistical Computing (Version 2.10.1). R Foundation for Statistical Computing, Vienne, Autriche