



**HAL**  
open science

## Application de méthodes de classification sur des vitesses météorologiques de dégradation de compteurs d'eau.

Frédéric Bertrand, Myriam Maumy

### ► To cite this version:

Frédéric Bertrand, Myriam Maumy. Application de méthodes de classification sur des vitesses météorologiques de dégradation de compteurs d'eau.. XIVe Rencontre de la Société francophone de classification – SFC 2007, Sep 2007, Paris, France. hal-00177005

**HAL Id: hal-00177005**

**<https://hal.science/hal-00177005>**

Submitted on 5 Oct 2007

**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.

# Application de méthodes de classification sur des vitesses métrologiques de dégradation de compteurs d'eau

F. Bertrand<sup>1</sup> et M. Maumy<sup>1</sup>

1. IRMA-ULP, 7, rue René Descartes, 67084 Strasbourg cedex  
(fbertran,mmaumy)@math.u-strasbg.fr

Mots clés : Applications, discrimination.

## Introduction

Les compteurs d'eau, en vieillissant, fournissent une mesure de plus en plus imprécise de la consommation d'eau. Cette dégradation se traduit généralement par un sous-comptage. Ce phénomène est source de problèmes pour les distributeurs d'eau qui ont mis en place des stratégies de gestion des parcs-compteurs ayant comme objectif la réduction des pertes économiques et le respect d'une politique de comptage équitable pour les différents usagers. Toute stratégie nécessite préalablement la compréhension du mécanisme de dégradation et la quantification du sous-comptage. Le vieillissement des compteurs est décrit à travers un modèle dynamique à états discrets, représentant chacun une certaine qualité métrologique. Ce modèle, couplé avec l'observation des erreurs de mesure à l'intérieur de chaque état, permet l'estimation notamment du taux de compteurs défaillants et de l'évolution de la précision de la mesure en fonction de la durée de service du dispositif. L'estimation des paramètres du modèle et la prédiction des valeurs des grandeurs d'intérêt pratique, ont été réalisées dans un cadre bayésien, avec l'utilisation de techniques de simulation MCMC.

Le but de cet article est de classifier en un certain nombre de groupes les compteurs d'eau d'un parc-compteurs d'un distributeur d'eau dans le but d'anticiper sur le changement de ces derniers et par conséquent d'atteindre les deux objectifs préalablement cités. Le papier s'articule en trois parties. La première partie présente le problème : introduction du vocabulaire, liens entre compteurs et qualité de mesure et présentation des données. La deuxième partie concerne la modélisation stochastique du vieillissement des compteurs. Nous introduisons une première ébauche de modèle de dégradation qu'il est nécessaire de compléter en introduisant d'autres facteurs explicatifs. Enfin, la troisième partie développe des techniques de classification pour proposer différents groupes de compteurs d'eau. L'originalité de cet article repose sur l'application de méthodes de classification classiques sur des données issues de modèles bayésiens.

## 1 Description du problème

### 1.1 Définitions et vocabulaire

Comme tous les instruments de mesure, les compteurs d'eau sont susceptibles d'erreur. Sur un branchement, le volume enregistré ( $v_{\text{enr}}$ ), et donc facturé, est généralement

différent du volume effectivement consommé ( $v_{\text{réel}}$ ).

L'étude statistique de ce papier, se limitera exclusivement aux données des compteurs dits volumétriques (largement majoritaires en France) caractérisés par la présence d'un organe de mesure qui se déplace sous l'effet de la poussée hydrodynamique, refoulant un volume déterminé d'eau à chaque tour.

**L'erreur relative de mesure d'un compteur** ( $e$ ) est obtenue en divisant la différence entre le volume enregistré ( $v_{\text{enr}}$ ) et le volume réellement écoulé ( $v_{\text{réel}}$ ) par ce dernier.

**La courbe métrologique** est la représentation graphique de la relation entre l'erreur relative de mesure ( $e$ ) et le débit circulant ( $q$ ).

**Le rendement d'un compteur** ( $r$ ) est le rapport entre le volume enregistré ( $v_{\text{enr}}$ ) et le volume consommé ( $v_{\text{réel}}$ ).

## 1.2 Classement des compteurs selon leur qualité de mesure

Le modèle de vieillissement prévoit le passage par quatre états métrologiques, de qualité décroissante  $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4$ . La définition des quatre états est inspirée par la réglementation actuelle et ses développements [1]. Nous réservons l'état  $\mathcal{E}_1$  aux compteurs qui respectent la conformité aux erreurs maximales tolérées sur toute la gamme de débits. L'état  $\mathcal{E}_2$  est caractéristique des compteurs ayant une métrologie imparfaite mais encore acceptable. Les compteurs qui se trouvent dans l'état  $\mathcal{E}_3$  présentent des rendements très mauvais et sont non-conformes. Enfin dans l'état  $\mathcal{E}_4$  nous avons mis les compteurs bloqués à tout débit, c'est-à-dire ceux qui mesurent une valeur fixe de débit ne variant plus avec le débit réel.

## 1.3 Données

Deux sources différentes de données sont disponibles pour l'étude. D'une part nous utilisons les résultats d'étalonnage de compteurs en service, réalisés dans les laboratoires d'essai d'un distributeur d'eau (base métrologique) et d'autre part nous nous servons des informations issues d'un outil interne de gestion des parcs de compteurs (base des blocages) pour analyser les phénomènes de blocages. Le recours aux deux bases de données est nécessaire puisque l'état absorbant  $\mathcal{E}_4$  est pratiquement inobservé dans la base métrologique.

À partir de la courbe métrologique de chaque compteur, nous pouvons déterminer son état et son rendement. Les individus sont en suite regroupés et dénombrés par état et par âge pour obtenir les  $n_i(t)$ , soit le nombre de compteurs qui se trouvent dans l'état  $\mathcal{E}_i$  à l'âge  $t$ . Leurs rendements sont utilisés pour estimer les paramètres des lois de probabilité des rendements à l'intérieur de chaque état.

# 2 Modélisation stochastique

## 2.1 Premier modèle de dégradation

Un mécanisme markovien homogène de dégradation a été imaginé sur la base de considérations techniques. En effet des tests d'endurance réalisés par des fabricants sur banc d'essai montrent que le « vieillissement naturel » des compteurs est très lent et ne provoque de changement d'état, à partir de l'état  $\mathcal{E}_1$ , qu'après des périodes extrêmement longues. Les changements d'état sont donc essentiellement liés à des accidents de parcours. En outre, il est réaliste de faire l'hypothèse que la dégradation est irréversible et donc la

chaîne a pour état absorbant  $\mathcal{E}_4$ . Le modèle est à temps discret, et l'unité temporelle est une année. Les éléments  $\Theta_{ij}$  de la matrice de transition  $\Theta$  sont les probabilités de passage de l'état  $\mathcal{E}_i$  (âge  $t - 1$ ) à l'état  $\mathcal{E}_j$  (âge  $t$ ).

Le vecteur ligne  $\mathbb{P}(t) = \{P_1(t), P_2(t), P_3(t), P_4(t)\}$  des probabilités des quatre états, à l'âge  $t$ , est exprimé en fonction de  $\mathbb{P}(t-1)$  et de la matrice de transition  $\Theta$  par :  $\mathbb{P}(t) = \mathbb{P}(t-1)\Theta$ . À l'intérieur de chaque état, le rendement suit une loi de probabilité caractéristique de l'état, dont les paramètres sont éventuellement fonction de l'âge. Le couplage entre le modèle de dégradation et les lois de probabilité des rendements permet de modéliser l'évolution de la métrologie d'un ensemble de compteurs par mélange (en proportions variables avec l'âge) d'individus appartenant aux différents états. Pour de plus amples renseignements sur le modèle de dégradation, nous renvoyons à la thèse de Alberto Pasanisi [4].

## 2.2 Amélioration du modèle de dégradation : à la recherche de nouveaux facteurs explicatifs

Le modèle exposé précédemment peut-être amélioré en introduisant des facteurs explicatifs pertinents. En effet, puisqu'un distributeur d'eau souhaite produire de l'information aussi précise que possible, il est souhaitable d'avoir alors recours à un découpage de son réseau suivant plusieurs critères comme les directions régionales, les centres opérationnels, les agences, etc. Mais un découpage plus fin et particulièrement intéressant en pratique peut être réalisé avec les contrats. Les contrats sont des unités géographiques forcément inégales en extension et nombre d'usagers. Cette répartition du territoire a une utilité opérationnelle. Le paramètre le plus significatif, dans le cadre de cette étude, pour décrire la taille d'un contrat est le nombre de compteurs, qui équivaut en pratique au nombre d'abonnés desservis.

Dans la pratique, il est bien reconnu par les experts et les exploitants que la dégradation des compteurs n'est pas la même sur tous les sites. La principale variable explicative typiquement évoquée comme responsable de cette différence de comportement est la « typologie » d'eau distribuée.

Pour résumer, d'une part la méconnaissance de l'effet de chaque facteur sur le mécanisme de dégradation des compteurs et d'autre part la complexité des conditions réelles d'exploitation, difficilement traduisibles en un nombre réduit de paramètres, poussent à aborder le problème différemment. En effet, l'approche choisie consiste à associer à chaque site d'exploitation une notion d'agressivité qui traduit l'effet global de tous les facteurs.

Les avis a priori des experts dans le domaine du vieillissement des compteurs volumétriques convergent sur un nombre de groupes d'agressivité égal à trois. Les méthodes développées dans [4], se déroulant en trois phases, permettent de constituer ces trois groupes de contrats caractérisés par différents intervalles d'agressivité délimités par trois valeurs particulières d'agressivité ( $A_1, A_2, A_3$  en ordre croissant) :

1. Estimation d'un paramètre d'agressivité uniquement sur la base des résultats métrologiques. Ce paramètre est appelé la **vitesse de dégradation métrologique**  $\lambda$ . En s'appuyant sur la valeur de la vitesse  $\lambda$ , les contrats sont découpés en trois groupes d'agressivité croissante.
2. Pour chacun des trois groupes ainsi obtenus, nous calculons les taux de blocage des compteurs de référence, suivant les informations de la base des blocages. Les calculs dans [4] montrent une cohérence entre les deux indicateurs d'agressivité (vitesse de dégradation métrologique et taux de blocage) : les sites où les compteurs

se dégradent plus vite sont aussi ceux où nous observons, en proportion, plus de blocages.

3. Estimation de l'agressivité des contrats pour lesquels uniquement les informations de la base des blocages sont disponibles, grâce à l'utilisation de l'association entre le taux de blocage et l'agressivité. En pratique nous assignons un contrat à un des trois groupes précédemment découpés en fonction de la valeur observée des taux de blocage des compteurs de référence. Si, par exemple, les taux observés sur un site d'agressivité inconnue sont plus proches de ceux qui caractérisent le groupe  $A_1$ , alors nous assignons le contrat examiné au groupe  $A_1$  etc.

### 3 Classification des contrats

L'objectif principal de cette étude est de proposer une classification des contrats à partir du calcul de l'estimation de la vitesse de dégradation métrologique  $\lambda$ .

Un premier problème se pose : savoir si nous devons faire des groupes et le cas échéant, déterminer le nombre de ces groupes. Il est à noter que, jusqu'à présent, à la fois la détermination du nombre de groupes et leur constitution, était réalisée grâce à la connaissance et aux avis des experts du vieillissement des compteurs d'eau. En utilisant ce nombre de groupes fixé a priori, nous avons utilisé des méthodes classiques de classification comme la classification hiérarchique ascendante, la méthode des  $k$ -moyennes et celle des cartes de Kohonen. Puis, nous avons comparé les résultats obtenus avec les groupes constitués par les experts, puisqu'une comparaison directe, c'est-à-dire sans référence extérieure, des résultats de ces trois méthodes n'a pas de sens.

La méthode, la plus compatible avec les avis des experts en la matière et dont nous nous sommes donc finalement servis, est la classification hiérarchique ascendante pour diverses distances associées à la méthode de liaison de Ward. Ce constat nous a naturellement amené à nous demander en combien de groupes il était le plus adéquat de couper le dendrogramme issu des classifications hiérarchiques ascendantes obtenues précédemment. Nous avons déterminé statistiquement le nombre de groupes d'agressivité à considérer en utilisant les règles de Mojena [3] particulièrement adaptées aux méthodes hiérarchiques, voir le livre de Everitt [2] pour de plus amples détails. Il s'est dégagé un nombre optimal de groupes égal à quatre et la répartition des contrats parmi ces derniers a été validée par les experts du domaine. L'exposé présentera cette démarche de manière détaillée et illustrée par des données réelles.

- [1] A. Costes et Y. Pia, "Les compteurs d'eau en France : la réglementation et son évolution", *Techniques, Sciences et Méthodes* 7, 2000, 21–27.
- [2] B. S. Everitt, *Cluster analysis*, 4<sup>th</sup> edition, Arnold Publication, 2001.
- [3] R. Mojena, "Hierarchical grouping methods and stopping rules: An evaluation", *Computer Journal* 20 (4), 1977, 359–363.
- [4] A. Pasanisi, "Aide à la décision dans la gestion des parcs de compteurs d'eau potable", *Thèse présentée pour obtenir le grade de Docteur de l'ENGREF en Sciences de l'Eau (Option Statistique)*, 2004.