Titre: Title:	Algorithme de l'alpiniste pour l'étude de cartes de contrôle en coordonnées parallèles
Auteur: Author:	Sébastien Henwood
Date:	2019
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Henwood, S. (2019). Algorithme de l'alpiniste pour l'étude de cartes de contrôle en coordonnées parallèles [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/3963/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/3963/
Directeurs de recherche: Advisors:	Samuel Bassetto
Programme: Program:	Maîtrise recherche en génie industriel

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Algorithme de l'alpiniste pour l'étude de cartes de contrôle en coordonnées parallèles

SÉBASTIEN HENWOOD
Département de mathématiques et de génie industriel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Génie industriel

Juin 2019

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Ce mémoire intitulé :

Algorithme de l'alpiniste pour l'étude de cartes de contrôle en coordonnées parallèles

présenté par **Sébastien HENWOOD**
en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

François LEDUC-PRIMEAU, président
Samuel Jean BASSETTO, membre et directeur de recherche
Luc ADJENGUE, membre

DÉDICACE

A plusieurs reprises
une chance m'a été donnée
grâce à vous : MERCI

À Audrey pour ta patience fois trois-mille et ma famille parce que

REMERCIEMENTS

Je remercie mes collègues du laboratoire CIMAR LAB de Polytechnique Montréal. Je remercie tout particulièrement Andres, Ben, Milad et Sonia pour notre travail conjoint au sein de l'équipe Data Science du CIMAR, ainsi que Mathilde et Garrick, mes camarade de bureau.

Je remercie par ailleurs l'entreprise SPN Consultants Inc. et l'organisme MITACS pour leur investissement dans ce projet (Projet ANEMONE : Algorithmes d'anticipation des Défauts, de leur Modélisation et Neutralisation Anticipée / Numéro IT : IT11878).

RÉSUMÉ

Ce mémoire propose l'algorithme de l'alpiniste, une méthodologie visant à estimer l'état de santé d'un équipement ou d'un processus multivarié pour prévenir l'apparition d'anomalies. On applique cet algorithme sur une carte de contrôle en coordonnées parallèles, sur laquelle sont projetées les observations prises sur le processus. Une fois ces observations projetées, on construit une approximation de la densité de ces observations à partir de leur historique. L'algorithme de l'alpiniste échantillonne la densité des observations de façon à obtenir un histogramme bidimensionnel. L'algorithme ajuste également l'histogramme pour tirer parti d'informations fournies sur les observations (anomalies ou non). On récolte par la suite pour toute nouvelle observation le chemin emprunté sur ces histogrammes : on se sert des caractéristiques géométriques de ce chemin pour calculer une fonction d'énergie. La fonction d'énergie proposée est paramétrisable et on peut optimiser ces paramètres par un apprentissage supervisé. On peut alors utiliser l'énergie obtenue pour la classification et le diagnostic des anomalies.

L'algorithme proposé est testé sur le jeu de données Tennessee Eastman Process, représentant un processus de production de l'industrie chimique. On réalise ensuite l'apprentissage supervisé des paramètres de la fonction d'énergie par une régression logistique. Sur ce jeu de données, l'algorithme de l'alpiniste obtient une aire sous la courbe sensibilité/spécificité de 94.1%. La précision de l'algorithme de l'alpiniste sous ce modèle de régression logistique pour un choix du seuil de décision équilibré est de 89.67%, avec une sensibilité de 90.54% et une spécificité de 83.87%. Ces résultats démontrent qu'il est possible d'utiliser l'algorithme de l'alpiniste pour la classification et le diagnostic des anomalies.

ABSTRACT

This dissertation introduces the alpinist's algorithm, a methodology to quickly estimate the health factor of an equipment or a multivariate process in order to prevent abnormalities. We apply this algorithm on a control chart in parallel coordinates on which are projected the process samples. Once the samples are projected, we then approximate the density of those samples based on the available data. The algorithm proceed to sample those densities in order to build a two dimensional histogram. The densities are adjusted based on the available samples's labels (abnormalities or not). For any new observation, we can then look at the path on the histogram : we make use of the geometrical features of that path to compute an energy function. This energy function is parametric and it can be optimised by a supervised learning algorithm. Eventually the energy can be used to classify and diagnose the abnormalities.

The alpinist's algorithm is tested on the dataset Tennessee Eastman Process, which represent a production process in the chemical industry. We undergo a supervised learning of the energy function's parameters with a logistic regression. On that dataset, the alpinist's algorithm achieve an area under the curve sensitivity/specificity of 94.1%. The accuracy of the alpinist's algorithm under that logistic regression model is of 89.67% for a balanced decision boundary, with a sensitivity of 90.54% and a specificity of 83.87%. Those results demonstrate the usability of the alpinist's algorithm for the purpose of classification and diagnosis of abnormalities.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES SIGLES ET ABRÉVIATIONS

PAA	Piecewise Aggregate Approximation
SAX	Symbolic Aggregate Approximation
EHF	Equipment Health Factor (Indicateur de Santé de l'Equipement)
MSP/SPC	Maîtrise Statistique des Procédés/Statistical Process Control
CBM	Condition Based Maintenance
PCA	Principal Component Analysis
PLS	Partial Least Square
TEP	Tennessee Eastman Process
CAM	Champs Aléatoires de Markov
MBE	Modèles Basés sur l'Energie

LISTE DES ANNEXES

CHAPITRE 1 INTRODUCTION

Entre $100,000$\$ et $1,000,000$\$: c'est le coût de la destruction d'un lot dans l'industrie des semiconducteurs [1]. Une cause probable : l'état d'un ou de plusieurs équipements défaillants lors du traitement du lot. Et les coûts d'un équipement défaillant ne s'arrêtent pas qu'à la perte de produits : maintenance non prévue, retards dans le processus font eux aussi partie des inconvénients d'un soudain décrochage de l'état de la machine. Mais est-ce si soudain que l'on pourrait le penser ?

Ce caractère soudain semblerait signifier que la défaillance est immédiate et absolue : rien n'aurait pu prévenir de son apparition. Or ne pourrait-on pas, à la manière d'une voiture sur laquelle son conducteur habitué pourrait déceler des variations dans son vrombissement usuel, déceler des changements plus fins et subtils dans l'état de l'équipement ? Quel est l'état de "santé" de notre équipement de production ? Cette question fait écho à la maîtrise statistique des procédés : les détériorations du processus seraient détectables en regardant les mesures que l'on pourrait faire sur l'équipement de production. Les cartes de contrôle constituent un outil permettant de présenter les statistiques dans un graphique avec des règles simples : est-ce qu'une mesure effectuée sur l'équipement reste dans sa zone statistiquement habituelle ?

Historiquement, ces cartes de contrôle ont révolutionné l'industrie : c'est un outil simple pour tenter de réduire les défaillances si coûteuses. Tant et si bien que le nombre de variables mesurées sur l'équipement a augmenté au point d'avoir de trop nombreuses cartes de contrôle. On a ainsi tenté de synthétiser à nouveau plusieurs cartes de contrôle en une seule, pour simplifier la tâche de maîtrise des procédés. Or ces nouvelles cartes de contrôle multivariées ont perdu en pouvoir de diagnostic en venant cacher les variables originelles pour les synthétiser en un indicateur global. Une approche, baptisée Open-Up, pour pallier à cette perte de diagnostic fut proposée par Tilouche [2] pour regagner ce pouvoir de diagnostic. Les variables composant nos mesures sont ainsi projetées dans une carte de contrôle en ne cherchant pas à réduire leur nombre. On observe plutôt la relation entre les variables prises deux à deux et on suppose que c'est cette relation qui pourrait nous renseigner statistiquement sur l'état de l'équipement.

Or, une problématique dans cette carte de contrôle est que les défaillances ont un caractère soudain et absolu : nous ne sommes pas encore au niveau du conducteur chevronné mentionné plus tôt. Comment alors tirer parti de la carte de contrôle Open-Up pour calculer un état de "santé" de l'équipement ?

1.1 Éléments de la problématique

Nous avons donc un environnement mettant l'emphase sur la relation des variables prises deux à deux. Le calcul de l'état de santé devrait donc tirer parti de cette caractéristique : que peut-on déduire de l'état de l'équipement à partir de la géométrie de la carte de contrôle ? Cette géométrie est apprise au travers des protocoles d'entraînements d'Open-Up. C'est donc un élément dynamique reposant sur la richesse des mesures fournies au modèle. Ces mesures peuvent être rares selon le contexte d'utilisation : par exemple un lancement en production ou une faible fréquence de mesures [2]. L'inverse est aussi vrai : les mesures peuvent être abondantes au point où il sera difficile pour la carte de contrôle d'en tirer pleinement parti. Un historique de plusieurs années pourrait induire une augmentation importante des temps de calcul, au point où il serait potentiellement impossible de tirer pleinement parti d'un jeu de données trop volumineux.

Ces mesures, ou observations, peuvent bénéficier d'un potentiel libellé d'origine humaine : on pourra alors associer à chaque observation une étiquette nous renseignant par exemple sur le type d'anomalie observée. Cette étape d'identification des anomalies, bien que nécessaire pour l'entraînement des classificateurs dits "supervisés", peut s'avérer plus difficile qu'on ne pourrait le penser. Différents facteurs peuvent entrer en jeu : des problématiques de coûts, de manque de temps des experts, ou encore de manque de connaissances sur l'équipement (dans le cas de lancements en production par exemple). De nombreuses mesures pourraient donc paradoxalement avoir une très faible proportion de libellés. Dans tous les cas, l'identification d'une anomalie se fera à posteriori de sa détection, avec un temps variable selon les facteurs exposés plus tôt.

La géométrie de la carte de contrôle est un élément faillible. Elle repose sur l'hypothèse que les mesures utilisées pour l'entraînement seront représentatives de l'équipement. Or, deux entraînement des données différentes issues d'un même processus pourraient conduire à des conclusions divergentes de la carte de contrôle. Notre indicateur de santé devra donc tenir compte des zones de la carte de contrôle n'ayant jamais été observées sans immédiatement se détériorer. Ces zones sont des valeurs possibles dans le jeu de données, mais jamais observées. Une fois observées à plusieurs reprises ces zones sans anomalies seraient le témoin d'un fonctionnement normal du système. Cela revient à modifier la géométrie de la carte de contrôle comme si une nouvelle observation était ajoutée dans l'ensemble des mesures utilisées pour l'entraînement. Ainsi, notre méthodologie devrait fonctionner par un entraînement incrémental : la géométrie de la carte de contrôle évoluerait à chaque nouvelle observation influençant ainsi le calcul de l'indicateur de santé.

Ces dynamiques devraient être encouragées : on ne voudrait pas que l'indicateur de santé de l'équipement soit en surapprentissage. Les zones nouvelles devront avoir un poids comparable aux zones habituelles dans le calcul de l'état de santé. Cette approche incrémentale devrait bénéficier à notre indicateur de santé. Les mesures libellées comme anomalies devraient influencer différemment la géométrie. Ainsi l'indicateur de santé obtenu devrait être en mauvais état dans une géométrie d'observations fautives.

Pour résumer, une méthodologie tirant parti de la géométrie de la carte de contrôle devra être proposée pour le calcul de l'indicateur de santé de l'équipement. Cette géometrie sera amenée à être modifiée tant par des données recoltées sur l'équipement que par un étiquetage réalisé par des experts : on parlera alors de géometrie dynamique. La méthodologie devra ensuite être testée pour savoir si elle répond aux attentes que l'on peut s'en faire : l'indicateur obtenu nous renseigne-t-il sur l'évolution de l'état de l'équipement ? Nous donne-t-il accès à une information qualitative ? L'indicateur est-il stable entre deux observations ? Nous permet-il de différencier des observations bonnes des observations fautives ? Cela devra se faire en respectant les partis pris initiaux d'Open-Up : la géométrie de la méthode proposée devra respecter la relation deux à deux établie entre les variables sans hypothèse statistique sur les données.

1.2 Objectifs de recherche

Les objectifs de la recherche proposés en conséquence des problématiques exposées plus tôt :

1. proposer un Indicateur de Santé de l'Equipement (EHF) se basant sur les informations visuelles de la carte de contrôle en coordonnées parallèles Open-Up.
2. qualifier la capacité de l'indicateur proposé à donner une information qualitative quand à l'état intrinsèque du système observé : cohérent entre deux observations subséquentes, permettant de discriminer en toute fiabilité les observations témoignant d'un dysfonctionnement de celles d'un fonctionnement normal.

1.3 Plan du mémoire

Le reste de ce mémoire se divise en 5 chapitres :

Chapitre 2 - Ce chapitre couvre la revue de littérature. Nous nous intéresserons à la maîtrise statistique des procédés, aux techniques de cartes de contrôle et enfin à Open-Up : l'objectif est de comprendre les pratiques et méthodes importantes pour s'assurer que le travail réalisé en respectera les principes clés. Une partie sera dédiée aux méthodologies existantes de calcul de l'EHF : quelles sont les pratiques admises et

peut-on trouver des échos au travail proposé dans ce mémoire pour s'en inspirer. S'en suivront une partie sur les techniques de réduction dimensionnelles et les modèles énergétiques tels que les Champs Aléatoires de Markov qui seront utilisés dans la méthodologie proposée.

Chapitre 3 - Ce chapitre présente la méthodologie proposée : l'algorithme de l'alpiniste. Les concepts mis en oeuvre et leur justification seront abordés ainsi qu'une analyse du modèle. Enfin, un exemple jouet sera proposé pour plus de clarté.

Chapitre 4 - Ce chapitre aborde le processus de test et de validation du modèle mis en oeuvre. Une simulation d'unité de production sera utilisée comme jeu de données sur lequel sera entraîné l'algorithme.

Chapitre 5 - Ce chapitre étudie les résultats des tests et les discussions associées. Nous vérifierons si les objectifs de recherches ont été atteints. La discussion aura pour but de venir critiquer les résultats obtenus.

Chapitre 6 - Ce sixième et dernier chapitre porte sur la conclusion amenée à ce mémoire ainsi que des réflexions et futurs axes de travail.

CHAPITRE 2 REVUE DE LITTÉRATURE

On propose de commencer la revue de littérature par la maîtrise statistique des procédés afin de mieux comprendre quel est le contexte technique dans lequel devra opérer l'indicateur de santé de l'équipement à réaliser.

2.1 Maîtrise statistique des procédés (MSP)

L'objectif de la maîtrise statistique des procédés, *Statistical Process Control*, méthodologie proposée dans les années 30 par Walter Shewart des laboratoires Bell, est de prévenir l'apparition des défauts plutôt que de les détecter à posteriori. Les 3 hypothèses de base de cette démarche sont les suivantes :

1. c'est le processus qui élabore le produit ;
2. le comportement des processus fluctue dans le temps ;
3. les processus ont tendance à se désorganiser et se dégrader dans le temps.

On cherche alors à mesurer les caractéristiques du processus pour prévenir l'apparition de défauts. Cette mesure peut s'effectuer sur les produits à travers des variables physiques ou encore directement sur les outils de production à travers des capteurs. Ainsi, l'outil de prédilection de la MSP est la carte de contrôle selon [3] qui la définit comme :

> "[...] un enregistrement chronologique des données caractéristiques d'un processus sous forme d'un graphique."

Cette représentation graphique nous donne une information sur l'évolution du processus observé : est-il encore sous contrôle ?
Cela ouvre la porte aux trois objectifs de la MSP énoncés par [4] :

— donner aux opérateurs un outil de pilotage de la machine ;

— formaliser la notion de capabilité d'un moyen de production ;

— faire le tri entre les situations ordinaires et extraordinaires qui nécessitent une action.

Une carte de contrôle classique est la carte de contrôle de Shewart. On surveille une variable qui suit une loi normale. On considère que le processus est capable, c'est à dire qu'il peut produire dans les limites des spécifications qui lui sont assignées : 6 écarts types σ de la loi normale tiennent dans l'intervalle de tolérance I du processus. Chez certains constructeurs automobiles, le processus doit être encore plus capable : on doit alors vérifier $6\sigma * 1.67 < I$.
La variable peut se rapporter à une valeur moyenne (carte de contrôle dite $\bar{X}$), à une étendue

(dite R), etc. Pour savoir s'il le processus est sous contrôle, il est d'usage de proposer une limite statistique en dehors de laquelle on pourrait considérer le système observé comme hors contrôle. Par exemple pour notre carte de Shewart, cette valeur est fixée aux $\mu \pm 3\sigma$ où μ est la moyenne des observations et σ est l'écart type d'une loi normale à 1 dimension. Cela signifie que environ 0.3% des valeurs seraient en dehors de cette limite [5].

Les causes de cette perte de contrôle peuvent alors être assignées à deux catégories distinctes : assignables (réglage défectueux par exemple) ou fruit de la variabilité aléatoire (usure par exemple). D'autres méthodologies et heuristiques permettent de qualifier l'état actuel du processus : notamment la détection de tendances à la baisse ou à la hausse sur les 7 dernières observations, des limites de surveillances (moins restrictives que les limites de contrôle vues précédemment), ou encore la trop grande proximité d'une des limites [4–7]. Un exemple graphique de cette carte de contrôle est visible à la Figure 2.1 où les limites de contrôles sont visibles en pointillés, la moyenne est visible en ligne pleine, et où une observation est en dehors des limites de contrôle : il s'agit de l'observation 18.

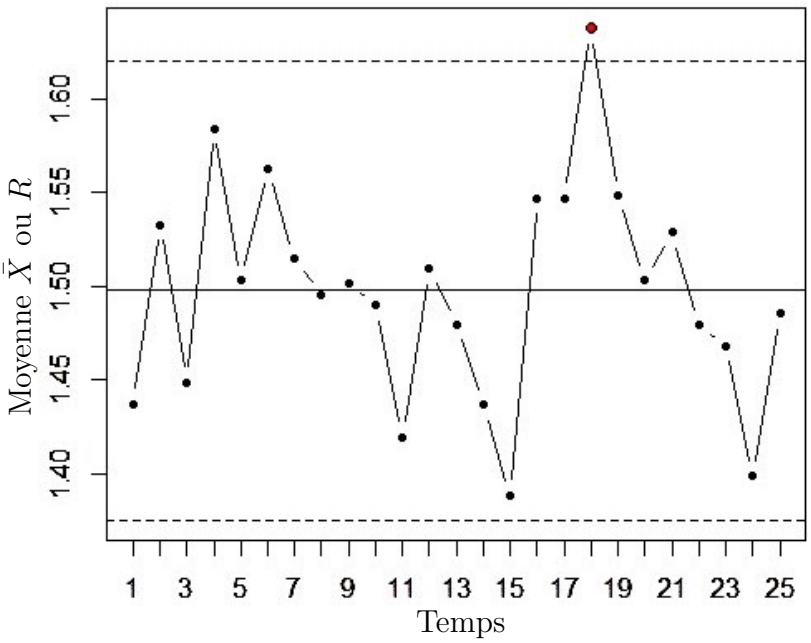


Figure 2.1 Illustration d'une carte de contrôle Shewart

Historiquement, les premières cartes de contrôles proposées furent univariées (surveillance d'une seule variable) : on peut citer entre autres les cartes Shewhart que nous venons de voir, EWMA, ou encore CUSUM [4, 8]. Les cas d'application varient entre les cartes : par exemple la méthodologie pour la carte CUSUM la rend adaptée pour la recherche de dérives faibles et lentes. Chaque type de carte propose des méthodologies de calcul des limites appropriées.

Des variantes multivariées furent par la suite proposées, la plus célèbre étant la carte de Hotelling [9] (cette dernière venant généraliser la carte de Shewart).
La carte de Hotelling utilise la distance T^2 comme indicateur à surveiller. Cette dernière combine l'information de la dispersion et des moyennes de plusieurs variables [5]. Cette valeur pourrait être comprise comme une extension multivariée du test t de Student. Elle est définie à l'Équation 2.1 avec X le vecteur des observations, m le vecteur des moyennes, $(X - m)$ le vecteur des déviations par rapport à la moyenne, S^{-1} l'inverse de la matrice de covariance, c une constante calculée sur la taille de l'échantillon à partir duquel la matrice de covariance fut estimée.

$$T^2 = c(X - m)'S^{-1}(X - m). \tag{2.1}$$

On peut remarquer que les calculs font appel à une inversion de matrice de covariance, ce qui peut rapidement devenir un problème en présence de nombreuses variables. En effet, il s'agit d'un algorithme de complexité $\mathcal{O}(n^3)$ avec n le nombre de variable, dans les implémentations courante [10]. Des travaux ayant pour but d'étendre les cartes CUSUM, EWMA à des données multivariées existent : par exemple le cas CUSUM multivarié [11]. La méthodologie proposée réutilise la méthode CUSUM sur chaque variable individuelle. En cas d'indépendance des variables, une synthèse en un indicateur unique est proposée ; à l'inverse du cas de dépendance qui revient à surveiller autant d'indicateurs/cartes de contrôles qu'il y a de variables observées.

On remarque que ces cartes de contrôle reposent sur une hypothèse restrictive de normalité des données en se fiant à l'hypothèse du théorème central limite [4] :

> "Tout système, soumis à de nombreux facteurs, indépendants les uns des autres,
> et d'un ordre de grandeur équivalent, génère une loi normale."

Par exemple, la construction à $\pm 3\sigma$ vue plus tôt repose sur cette hypothèse. Cependant il existe plusieurs cas de figure où la normalité des données n'est pas une bonne hypothèse [12] : réactions chimiques, nombres de défauts manufacturiers, défauts coaxiaux entre autres sont mieux capturés par des lois de probabilité tierces. Les limites de contrôle basées sur l'hypothèse d'une loi normale peuvent alors induire en erreur. Des alternatives non paramétriques sont proposées pour dépasser le théorème central limite. Par exemple une forme paramétrique non gaussienne est proposée par [13], mais cela exige un travail préliminaire pour déterminer la distribution suivie par les données. Un autre exemple, purement non paramétrique, est présenté par [14] qui propose une carte de contrôle multivariée utilisant la technique des Support Vector Machine (SVM). On se réfère à [2] pour une revue étendue des pratiques non paramétriques. Une problématique commune à tous ces modèles multivariés est que la perte d'information vient diminuer la capacité de diagnostic des anomalies [15].

Ainsi, la question de surveiller le processus au travers d'une carte de contrôle s'étend à comment détecter la présence d'anomalies multivariées sans hypothèse de normalité et les expliquer de manière simple et visuelle à des opérateurs pour guider les actions correctives. Récemment, Open-Up [2] a proposé une approche basée sur les coordonnées parallèles venant répondre à cette problématique.

2.1.1 Open Up

Open-Up est un outil de maîtrise des processus visuel et non paramétrique. Il est visuel dans le sens où il permet de présenter d'un seul tenant de multiples relations entre les variables et non-paramétrique car il ne se repose pas sur les hypothèses de normalité des données observées. Il se base sur la technique des coordonnées parallèles pour projeter l'ensemble des variables sur un espace en deux dimensions.

Les coordonnées parallèles [16, 17] sont une technique de représentation d'une information multivariée : les observations sont projetées en deux dimensions et la dimension x (à l'abscisse) permet d'indexer des axes parallèles représentant chacun une des variables composant les observations. Par exemple, si l'on dispose d'observations contenant n variables et que l'on souhaite en faire une représentation en coordonnées parallèles, la représentation fera appel à un ensemble $x = \{1, \ldots, n\} \in \mathbb{N}$ pour indexer ces variables. La valeur à l'ordonnée (dans la dimension y) sur chaque axe est une projection de l'espace de la variable originale sur cet axe. Les variables sont normalisées pour que l'étendue de chaque chaque variable dans la dimension y soit comparable.

Un exemple serait de considérer les observations du Tableau 2.1. Notre projection en coordonnées parallèles comprendrait donc 3 axes parallèles sur x aux index $\{1, 2, 3\}$. Les valeurs de ces variables seront visibles à l'axe y qui est égale à la valeur normalisée du y_i original. La projection en coordonnées parallèles de cet exemple est visible à la Figure 2.2. Les variables ne sont pas normalisées étant donné que nous sommes dans un exemple où les valeurs ont été choisies dans un intervalle comparable.

Tableau 2.1 Jeu de données exemple

Observation	x_1	x_2	x_3
1	7	3	9
2	5	4	6

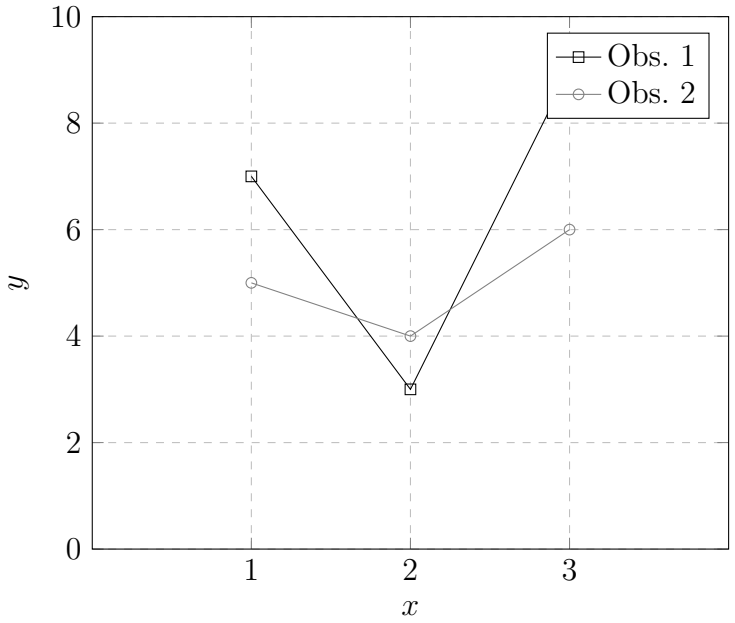


Figure 2.2 La projection en coordonnées parallèles du jeu de données exemple

Ainsi, on dispose d'un environnement visuel permettant la représentation de nombreuses dimensions en un seul plan.

Le protocole d'entraînement d'Open-Up comprend deux tâches séquentielles : premièrement on ordonne les variables et ensuite on utilise l'historique d'observation disponible pour définir les zones de bon fonctionnement connues.

1. L'ordonnancement des variables :
   Il fut proposé l'utilisation du critère d'Information Générale $GI(x_1, x_2, F, H, G)$ pour ordonnancer les variables, fonction prenant en entrée deux mesures de probabilités différentes $F(x_1, x_2)$ et $H(x_1, x_2)$ ainsi qu'une fonction continue $G(.)$ mesurant la "distance" entre les deux mesures de probabilités. Par exemple, F et H sont définies comme la probabilité jointe et la probabilité marginale dans [2], tandis que G est fixée à un critère d'information mutuelle. L'information mutuelle est une quantité mesurant la dépendance statistique entre 2 variables :

   $$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p(x)\, p(y)} \right). \tag{2.2}$$

   Plusieurs choix de F, H et G sont possibles : on peut ainsi tendre à arranger les variables pour maximiser la dépendance statistique entre chaque paire de variables contiguës ou encore séparer au mieux les données pour mieux distinguer les sous-ensembles dans les

observations.

Une fois le calcul du critère d'information générale fait entre chaque paire de variables, on les classe par ordre décroissant de cette valeur. Les variables ayant l'information générale la plus élevée sont les plus importantes au regard de l'information qu'elles contiennent sur l'ensemble des variables observées, en excluant les variables déjà sélectionnées. L'ordonnancement des variables permet de fixer l'ordre des coordonnées parallèles utilisées. Cette étape est commune aux deux approches.

Ensuite les observations sont transformées en une vue par coordonnées parallèles sur un espace cartésien bidimensionnel où l'axe des abscisses x est la dimension de l'index des variables et l'axe des ordonnées y est la dimension de la valeur de la variable. Sur l'axe des abscisse, une interpolation linéaire est réalisée entre chaque paire de variables. Sur l'axe des ordonnées, les valeurs sont normalisées : cela permet d'avoir une étendue comparable de la première à la dernière variable.

Une fois cette transformation effectuée, les données utilisées pour l'entraînement sont projetées dans un but de visualisation. Cette fois-ci, deux méthodologies différentes existent, ce sont les deux approches proposées dans la thèse originale présentant Open-Up, à savoir par clusters de bon fonctionnement et par densité.

2. Apprentissage historique :
   Deux méthodologies ont été proposées. La première, par clusters de fonctionnement et la seconde par densité. L'approche par densité consiste à représenter les probabilités vues lors de l'entraînement d'Open-Up pour guider la réflexion de l'utilisateur. D'autre part, l'approche par cluster de fonctionnement simplifie la réflexion en groupant les observations par un algorithme de partitionnement. Cette approche n'est pas retenue dans la suite de ce mémoire. Ces deux approches ont pour but commun de simplifier les segments de la projection en coordonnées parallèles comme ceux visibles à la Figure 2.2 : un trop grand nombre de segments juxtaposés n'aboutirai qu'à un ensemble incohérent sans signification à nos yeux.

   L'approche par densité utilise un lissage par noyau pour calculer la fonction de densité de l'historique d'observations [18]. Cette méthode généralise l'utilisation d'un histogramme pour obtenir une valeur continue sur l'intervalle de définition. Ces densités sont calculées pour chaque index x de la carte de contrôle, incluant les espaces inter-variables qui ont été échantillonnés pour cette tâche. On applique l'équation 2.3 qui prend en entrée t observations, une taille de fenêtre h, et k un noyau tel que le noyau

gaussien de l'équation 2.4. Un exemple de résultat est visible à la Figure 2.3.

$$\hat{f}(x) = \frac{1}{th} \sum_{i=1}^{t} k\left(\frac{|x - x_i|}{h}\right), \tag{2.3}$$

$$k(z) = \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2} z^2). \tag{2.4}$$

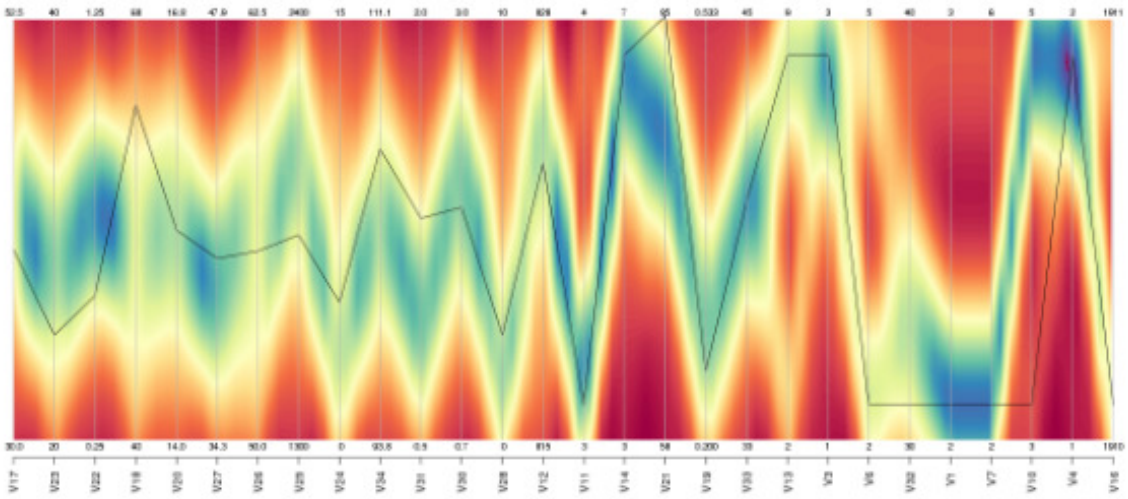


Figure 2.3 La carte de contrôle en coordonnées parallèles Open-Up par densité [2]

Légende

Gradient de couleur :

Bleu sombre : forte densité	Rouge vif : faible densité
Abscisse : variables ordonnées	Ordonnée : valeur de la variable en x
Trait gris : observation unitaire	

Dans les deux cas, Open-Up permet un diagnostic visuel des anomalies par l'interprétation qu'il est possible de faire via l'utilisation des coordonnées parallèles. Toute sortie de cluster pour l'approche par cluster ou transition par zone de faible densité pour l'approche par densité est assignable à une paire de variables dans laquelle la relation observée n'est pas conforme à ce qui a été vu dans le jeu de données d'entraînement.

Cependant, l'auteur de [2] note plusieurs défauts à la méthode Open-Up, dont :

— le besoin d'un large historique pour l'entraînement, mais sans plus de précisions sur le volume,

— le manque d'une étude statistique permettant d'évaluer théoriquement les performances obtenues,

— le nécessaire besoin de connaître avec une bonne précision la classe des données utilisées pour l'entraînement.

Il est aussi noté qu'il serait intéressant d'implémenter une méthodologie automatique de détection de défauts pour les cartes densité, ainsi qu'une prise en compte du temps. On peut remarquer que contrairement aux cartes plus classiques Open-Up ne permet par exemple pas d'exploiter les concepts de capabilité (la capacité d'un processus à respecter des spécifications), ou les erreurs de première et de seconde espèce (probabilité d'accepter un faux négatif et probabilité de rejeter un faux positif).

Pour conclure sur cette revue des pratiques en maîtrise statistique des procédés et sur Open-Up, l'information est souvent binaire : état correct ou incorrect. Dans la prochaine section, une revue des pratique sur les indicateurs de santé de l'équipement sera réalisée : l'objectif est alors d'avoir une information continue plutôt que binaire.

2.2 Indicateur de santé de l'équipement

Le concept de l'indicateur de santé de l'équipement (EHF) part d'une observation : on peut imaginer qu'il serait utile à un décideur de savoir si le système en question va plus ou moins bien. Au lieu d'une information binaire (correct/incorrect), on pourrait différencier plus finement deux états de fonctionnement corrects. On considère alors deux extrêmes : état parfait du système et état défaillant du système. On va tenter de qualifier où se situe le système actuel sur cette échelle. Cette transition d'un indicateur binaire à une valeur continue n'est pas immédiate.

Le calcul de l'EHF relève de la statistique multivariée : connaissant une observation du système composée d'une multitude de variables on calcule une valeur de l'EHF. Cela pourrait être la probabilité conditionnelle $Pr(\text{observation défaillante} \,|\, \text{historique d'observations})$. Il faudrait alors détecter les dérives, potentiellement mono-variables et son inverse, et qualifier le différentiel apporté à l'indicateur. Cela fait écho aux démarches de maintenance basées sur la condition (*Condition Based Maintenance*, CBM) : en exposant l'EHF on donne un outil de décision pour déclencher la maintenance du système. On se réfère à [19] pour une revue des pratiques de CBM.

La CBM est une pratique de maintenance relativement récente utilisant l'état (ou condition) d'un système pour déclencher les activités de maintenance [20]. Cela ferait ainsi partie de la maintenance prédictive. On distingue maintenance curative, après l'apparition d'un défaut, et préventive, cherchant à prévenir l'apparition des défauts. La préventive peut se faire de deux façons : soit à intervalles réguliers pour tenter de maintenir le système dans un état souhaitable (maintenance dite systématique), soit en se conditionnant sur l'état observé du système (maintenance dite prédictive ou encore conditionnelle). L'objectif de la maintenance

conditionnelle est de pouvoir anticiper l'apparition de défauts en détectant les glissements vers un état défaillant. La maintenance conditionnelle fait ainsi l'objet de nombreux efforts de recherche afin de raffiner les pratiques de maintenance existantes [21].

Une méthodologie CBM basée sur la carte de contrôle Hotelling [20] fait écho aux présents travaux. Le travail se limite cependant dans sa définition de système complexe, et ne propose pas la granularité d'Open-Up dans ses résultats. On ne peut ainsi pas tracer quelles variables sont en défaut. De même, Open-Up propose potentiellement des résultats supérieurs à la carte d'Hotelling [2]. Ceci vient renforcer la pertinence de vouloir y adjoindre des métriques similaires dans leur but (ici, l'EHF).

Une méthodologie basée sur une réduction dimensionnelle par Principal Component Analysis (PCA) et Partial Least Squares (PLS) est proposée par [22]. Avec PCA [23] l'objectif est de réduire linéairement les variables initiales d'un jeu de données en un ensemble de variables non corrélées (appelées Composantes). Cela se fait par le biais d'une décomposition de la matrice de covariance dans les directions contenant le maximum d'information (ici, qui expliquent un maximum la variation des données). Il s'agit d'une projection linéaire dans un espace de dimension inférieure minimisant la perte d'inertie du nuage de points.

PLS [24] se base aussi sur l'idée d'existence de quelques facteurs explicatifs sous-jacents au jeu de données que l'on pourrait généralement nommer Facteurs Latents. L'objectif général est d'utiliser ces facteurs latents pour réaliser une prévision de la réponse du système observé. PLS fait ainsi partie des techniques de modélisations indirectes telles que Principal Component Regression (PCR) ou encore Maximum Redundancy Analysis (MRA). La particularité ici est de

> "rechercher les directions dans l'espace latent qui sont associées avec de fortes variations dans la réponse (du système) mais en les biasant dans les directions qui sont prédites avec précision" [24]

Ainsi il s'agirait de surveiller par une carte d'Hotelling les Composantes Principales obtenues en appliquant PLS/PCA comme le propose [22]. Une seconde carte de contrôle est proposée pour surveiller les résidus des estimations, ce qui pourrait être révélateur d'un défaut sur le processus. Une critique possible de ce travail est à nouveau l'incapacité du modèle à supporter le diagnostic. La réduction dimensionnelle fait perdre tout sens aux données originales. De même, le cas d'application étudié dans l'article est assez précis ce qui ne nous renseigne pas sur la capacité de ce modèle à s'étendre à des systèmes de production plus complexes.

Des approches tentent de conserver la puissance explicative des variables initiales. Une méthodologie basée sur l'extraction de composantes sur les variables observées est présentée par [25]. Quatre types de variables sont ainsi proposées avec les composantes d'intérêt asso-

ciées :

1. comportement oscillant ;
   (a) analyse fréquentielle,
   (b) analyse temps-fréquence,
   (c) coefficients d'une modélisation en série temporelle,
2. fonction de pas (*piecewise constant* dans le texte original) ;
   (a) coefficients de régression,
   (b) caractérisation des pulsations,
   (c) intégration,
3. comportement par pics ;
   (a) détection des pics,
   (b) caractérisation des pics,
   (c) intégration des pics,
4. comportement lisse ;
   (a) coefficients de régression,
   (b) analyse des résidus,
   (c) coefficients d'une modélisation en série temporelle.

Une classification automatique des variables est proposée, se basant sur l'utilisation de la méthodologie Support Vector Machine (SVM) ou par arbre de décision. Ensuite, un état de santé de chaque caractéristique est calculé en se basant sur les méthodologies de carte de contrôle exposées plus haut. Enfin, ces valeurs sont consolidées par groupe de variables homogènes comme par exemple la température, la pression, etc. Enfin, elle sont agrégées dans un indice unique d'EHF. Cette méthodologie semble prometteuse dans sa capacité explicative pour le diagnostic et son utilisation des fondements de la SPC. On peut cependant noter que l'ensemble des méthodologies d'extraction de composantes peut rapidement devenir un frein à l'application à grande échelle de cette méthodologie de par le coût inhérent en temps de calcul. De même, il est remarqué qu'un effort d'ingénierie est nécessaire pour implémenter ce système. L'article cite d'une part la nécessité de créer une règle particulière pour le flux de gaz dans le système surveillé, ce qui implique la nécessaire connaissance d'experts ; d'autre part les variables ayant une forte corrélation sont élaguées : on pourrait donc passer à côté de l'information nécessaire à la détection d'une anomalie nonobstant les variables trop fortement corrélées.

Les techniques d'EHF présentées partagent une vision statistique du problème : les données sont considérées comme un ensemble flou aboutissant à l'EHF, par une boite noire s'efforçant de comprendre la relation entre les variables. Or aucune approche n'est faite sur une définition géométrique de l'EHF, ce qui pour rappel est le contexte de la question de recherche à laquelle nous cherchons à répondre. A plusieurs reprises ont été mentionnées des techniques de réduction dimensionnelles pour le calcul de l'EHF : nous allons nous y intéresser plus en avant dans la suite de la revue de littérature. Étant donné que Open-Up par densité utilise une fonction de densité estimée par la méthode de lissage par noyau, les valeurs de probabilités sont alors continues. L'objectif est de pouvoir les discrétiser pour pouvoir y effectuer des opérations élémentaires (addition, soustraction, etc). Cela nous donnera une marge de manoeuvre pour pouvoir raisonner sur les observations qui y seront projetées : pour rappel, un défaut de la méthode Open-Up par densité est qu'une classification automatique n'existe pas encore.

2.3 Réduction dimensionnelle

L'objectif de la réduction dimensionnelle dans notre cas d'application à la carte de contrôle Open-Up par densité est de trouver une surjection d'une densité vers un espace discret. Cela devra conserver la forme des données. Dans la littérature, on observe que les techniques de réduction de dimensionnalité sont particulièrement développées pour les séries temporelles : une série temporelle représentant un phénomène réel peut en effet rapidement croître en taille, dépendamment de la fréquence d'échantillonage des observations dans cette série temporelle et la durée totale d'échantillonage. En l'occurrence, les techniques dites de représentation de séries temporelles sont définies comme suit :

> "Soit x une série temporelle de taille n, alors la représentation de x est un modèle avec une dimensionnalité réduite p (p«n) ressemblant étroitement à x." [26]

Ces techniques sont intéressantes car elles pourraient se transposer sur les coordonnées parallèles : la relation entre les index x et $x + 1$ pourrait être assimilée à une série. En ce sens, les coordonnées parallèles seraient une forme de processus stochastique.

On se réfère à [27] pour lister les plus importantes techniques de représentations de séries temporelles existantes. Il y est aussi suggéré que ces techniques sont utilisées pour :

1. une réduction significative de la dimensionalité des séries temporelles ;
2. mettre l'emphase sur les caractéristiques essentielles des formes de la série temporelle à l'étude ;
3. une manière implicite de tenir compte du bruit présent dans les données ;

4. et enfin, l'impact de simplification subséquent en termes temporels et mémoriels pour les algorithmes utilisant les représentations plutôt que les données brutes.

Par rapport aux objectifs initiaux de simplification de la géométrie, les techniques de représentations de séries temporelles semblent pertinentes : on pourrait ainsi faire ressortir les caractéristiques essentielles de la carte de contrôle pour mieux les utiliser. Il y est par ailleurs proposé quatre grandes familles de techniques de représentations [28] ainsi que les méthodologies affiliées les plus populaires [29] :

1. non adaptatives aux données ;
   (a) PAA : Piecewise Aggregate Approximation
   PAA est une technique de réduction de dimensionalité très simple pour l'extraction de données sur une série temporelle [30, 31]. On réduit un vecteur X de dimension v en un vecteur $\bar{X}$ de dimension w en utilisant w intervalles de tailles égales pris dans v sur lesquels on effectue une moyenne (Eq 2.5).

$$\bar{X}_i = \frac{w}{v} \cdot \sum_{j=\frac{v}{w}(i-1)+1}^{(v/w)\cdot i} x_j \, , i = 1, \ldots, w \, . \tag{2.5}$$

   C'est un algorithme dont plusieurs extensions ont été proposées, dont par exemple [32] : le but étant de venir améliorer ses capacités de représentation. Il existe une mesure de distance d'une représentation PAA toujours inférieure ou égale aux données originales, ce qui garanti théoriquement aucun faux négatif [31]. On se réfère à l'article pour une preuve de cette propriété sur une fonction de mesure de distance.

   (b) DWT : Discrete (Haar) Wavelet Transform
   L'utilisation de l'Ondelette de Haar pour la représentation en série temporelle est plus restrictive que PAA [33]. En effet, une contrainte majeure est que la longueur de la série transformée se doit d'être d'une puissance de 2. PAA et DWT sont des représentations équivalentes selon [31] dans le sens où la meilleure reconstruction des données originales est identique et la distance entre deux séries transformées est identique selon l'une ou l'autre des méthode : DWT semble donc plus contraignant à résultats similaires que PAA.

   (c) DFT : transformée de Fourier discrète
   Permet d'obtenir une représentation spectrale discrète de la série

   (d) DCT : transformée en cosinus discrète
   Se rapproche de la DFT, mais utilise une fonction cosinus pour projeter la série

ce qui abouti à des coefficients réels, contrairement à la DFT qui aboutit à des coefficients complexes.

(e) SMA : Moyenne mobile simple *Simple Moving Average*
Une moyenne mobile simple d'ordre c un entier est une moyenne effectuée sur l'observation actuelle et les $c-1$ observations précédentes (voir Équation 2.6), avec x un élément du vecteur original, k l'élément du vecteur sur lequel on cherche à connaître la valeur SMA). x_k est donc remplacé par $SMA(x_k)$ pour tout k.

$$SMA(x_k) = \begin{cases} \frac{1}{c} \sum_{i=k-p}^{k+p} x_i \text{ si } c \text{ est impair, } c = 2p+1 \\ \frac{1}{c} \sum_{i=k-p}^{k+p-1} x_i \text{ si } c \text{ est pair, } c = 2p \end{cases} . \tag{2.6}$$

(f) PIP : *Perceptually Important Points*
Les points saillants sont ceux qui seraient jugés comme importants pour l'identification de motifs dans la série. Chaque point est identifié itérativement en choisissant le point avec la distance perpendiculaire maximum à une ligne tracée entre chaque paire de points saillant adjacents jusqu'à obtenir le nombre de points saillants désiré.

2. adaptatives aux données ;

(a) SAX : *Symbolic Aggregate Approximation*
En se basant sur PAA, une utilisation de symboles pour réduire à nouveau la complexité de la série temporelle est proposée [34]. L'algorithme bénéficie des performances peu coûteuse en temps et espace de PAA. Un alphabet A de cardinalité a est utilisé pour projeter une représentation PAA d'une série temporelle.
Un fondement majeur de SAX est la qualification par une loi normale des séries temporelles z-normalisées. En effet, l'auteur a remarqué empiriquement que de nombreuses séries temporelles suivent cette observation. Ainsi, SAX propose un découpage des intervalles des valeurs prises par une loi normale pour avoir une aire sous la courbe de la loi normale équivalente pour chaque symbole a utilisé. La conversion d'une représentation PAA en SAX est donc de faible complexité : il suffit de déterminer dans quel intervalle la valeur observée se trouve, puis d'en déduire le symbole a associé.
C'est un algorithme qui a déjà été appliqué pour trouver des anomalies dans des séries temporelles avec succès [35], et une évolution [36] permet d'étendre SAX à des jeux de données colossaux.

(b) PLA : *Piecewise Linear Approximation*

Consiste à approcher une série par des segments linéaires représentant au mieux les formes observées.

3. basées sur un modèle ;
   (a) Profil saisonnier moyen,
   (b) Modèles de représentation saisonniers basés sur des modèles linéaires
4. dictées par les données (données écrêtées) ;
   (a) FeaClip : extraction de caractéristiques de représentations rognées,
   (b) FeaTrend : extraction de caractéristiques de représentations des tendances,
   (c) FeaClipTrend : extraction de caractéristiques de représentations rognées et tendancières

Selon [26, 29], les représentations non adaptatives aux données utilisent des paramètres identiques indépendamment des séries temporelles observées, tandis que les représentations adaptatives changent de paramètres selon les données. D'autre part, les représentations basées sur un modèle ajustent les paramètres d'un modèle théoriques aux données observées. Enfin, les représentations dictées par les données compressent les séries temporelles en utilisant directement les données brutes de ladite série.

Pour conclure sur cette revue des pratiques de représentation, plusieurs catégories de méthodologies ont été proposées. Les méthodes basées sur un modèle ou dictées par les données ne sont pas en phase avec nos objectifs initiaux de simplification de la géométrie. Au contraire des techniques adaptatives et non-adaptatives aux données, notamment SAX qui pourrait s'adapter à la projection en coordonnées parallèles en simplifiant la géométrie par une discrétisation de l'axe des ordonnées selon un fondement statistique.

Il nous manque maintenant un fondement théorique sur lequel nous pourrions discriminer de façon continue deux observations. On propose de regarder de plus près le cadre théorique des modèles basés sur l'énergie qui sont en phase avec cet objectif.

2.4 Modèles basés sur l'énergie

Les modèles basés sur l'énergie (MBE) ont notamment été appliqués pour le débruitage d'image. Cela repose sur une idée : l'état de l'image est imparfait et on peut mesurer cette imperfection. Ainsi, l'objectif dans cette section est de déterminer s'il est possible d'utiliser les modèles basés sur l'énergie afin de mesurer l'imperfection d'une observation sur la carte de contrôle Open-Up.

Les modèles basés sur l'énergie permettent de diagnostiquer la dépendance entre les variables en assignant une valeur d'énergie à chaque configuration des variables observées [37]. L'inférence y est définie comme la fixation de la valeur des variables observées puis la recherche des valeurs pour le reste des variables non observées minimisant l'énergie. Enfin, l'apprentissage sur l'historique d'observations disponible est la recherche d'une fonction d'énergie minimisant l'énergie des configurations observées.

Par rapport aux modèles probabilistes qui requièrent une normalisation pour que l'intervalle des valeurs prises soit $[0; 1]$, les MBE n'ont pas besoin de cette procédure. Les calculs à effectuer s'en retrouvent simplifiés car cette phase de normalisation peut demander l'évaluation d'intégrales insolubles [37]. Cela peut cependant être nécessaire, par exemple si l'on souhaite combiner plusieurs valeurs d'énergie, ou encore si l'on souhaite utiliser la valeur d'énergie pour piloter un système tierce. On a alors besoin que ces valeurs soient sur un pied d'égalité pour ne pas biaiser en cascade l'ensemble des opérations en découlant.

Dans ce cas de figure, il est alors d'usage de normaliser les valeurs d'énergie en utilisant une distribution de Gibbs (Eq. 2.7), où Y est la valeur à prédire, Y_ω l'ensemble des valeurs possibles pour Y, X est l'ensemble des variables d'entrées d'une observation, $P(Y|X)$ signifie probabilité de Y sachant X (conditionnement de Y selon X), $E(Y, X)$ désigne la fonction d'énergie appliquée, β une constante arbitraire positive.

$$P(Y|X) = \frac{e^{-\beta E(Y,X)}}{\int_{y \in Y_\omega} e^{-\beta E(y,X)} \partial y} \,. \tag{2.7}$$

D'autres mesures de probabilités peuvent être utilisées. Comme mentionné précédemment, cela n'est possible que si l'intégrale présente au dénominateur est soluble.

Pour l'entraînement, on considère une fonction de perte et une architecture. Ce duo n'est valable qu'en cas d'entraînement supervisé : on aurait ainsi accès aux libellés pour chaque observation. L'entraînement serait ainsi du sur-mesure pour le jeu de données utilisé. La notion d'architecture demeure valide dans tous les cas et vient qualifier l'objectif du MBE utilisé : régression, classification binaire, multiclasse entre autres.

On s'intéresse plus particulièrement aux MBE tels que les champs aléatoires de Markov (dont le modèle d'Ising est un cas particulier) pour leur capacité à opérer sur des images, qui sont alors une matrice de dimension arbitraire. L'objectif demeure de proposer un EHF de la carte de contrôle en coordonnées parallèles par densité en exploitant l'information géométrique, information pouvant être représentée par une matrice.

2.4.1 Champs aléatoires de Markov, modèle d'Ising

Un champ aléatoire de Markov (CAM) [38,39] est un modèle graphique non orienté. Il possède ainsi un ensemble de noeuds correspondant à une variable ou un ensemble de variables, chaque paire de noeuds pouvant potentiellement être connectée.
On définit la notion d'indépendance conditionnelle : s'il n'existe pas de chemin entre deux ensembles de noeuds A et B ne passant pas par l'ensemble de noeuds C, on considère que A et B sont conditionnellement indépendant sachant C (Eq. 2.8).

$$A \perp\!\!\!\perp B|C\,. \tag{2.8}$$

Cela peut s'exprimer en terme probabilistes en considérant deux noeuds i et j, avec $p(x_i, x_j|x_{\setminus\{i,j\}})$ la probabilité jointe de x_i et x_j sachant le reste des x, $p(x_k|x_{\setminus\{i,j\}})$ la probabilité marginale de x_k sachant les x excluant x_i et x_j (Éq. 2.9).

$$p(x_i, x_j|x_{\setminus\{i,j\}}) = p(x_i|x_{\setminus\{i,j\}})p(x_j|x_{\setminus\{i,j\}})\,. \tag{2.9}$$

On définit les cliques maximales du graphe observé afin de réduire l'analyse à ces ensembles de noeuds. On associe à cela une fonction de potentiel ψ sur les cliques du graphes. Cela nous permet de réécrire l'équation précédente en :

$$p(x) = \frac{1}{Z}\prod_C \psi_C(x_c)\,, \tag{2.10}$$

où Z est une fonction de partition permettant de normaliser la distribution de $p(x)$, C l'ensemble des cliques du graphes, x_c une clique du graphe.
Il est d'usage de conditionner les fonctions de potentiels ψ pour s'assurer de leur positivité stricte [37, 40], par exemple :

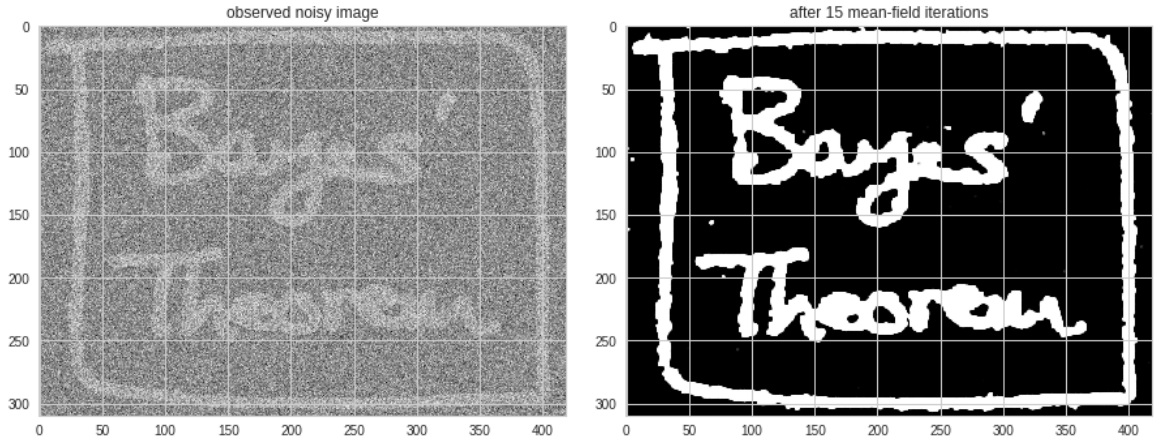
$$\psi_c(x_c) = exp\{-\bar{E}(x_c)\}. \tag{2.11}$$

Où $\bar{E}(x_c)$ est une fonction d'énergie prenant en entrée une clique du graphe. L'énergie totale est donc obtenue en faisant la somme de l'énergie de chacune des cliques avec la fonction de potentiel ψ_c. La fonction de potentiel exprime quelle configuration des variables locales est préférable [38].

On vient de brosser les fondements des CAM. Une des applications possibles est le débruitage d'images : on parle du modèle d'Ising. La fonction d'énergie utilisée est alors :

$$E(x, y) = h\sum_i x_i - \beta\sum_{\{i,j\}} x_i x_j - \eta\sum_i x_i y_i\,. \tag{2.12}$$

où y est l'image observée, h, β, η 3 valeurs positives représentant la force des composantes associées, x_i le pixel considéré de l'image recherchée, y_i le pixel considéré de l'image observée et x_j les pixels de la clique de x_i. La première composante de la fonction d'énergie vient favoriser un signe particulier, la seconde encourage des pixels voisins de même signe et la dernier contribue à ce que le pixel brut et celui obtenu par le modèle soient de même signe. C'est un exemple particulièrement intéressant dans la simplicité dont il est formulé. Un résultat d'application [1] du modèle d'Ising (utilisant le Mean Field Variationnal Inference pour la résolution) est visible Figure 2.4.



(a) Image bruitée (b) Image débruitée

Figure 2.4 Une application du modèle d'Ising

Pour conclure sur les modèles basés sur l'énergie, il existe des modèles relativement simples comme celui d'Ising qui exploitent la topologie d'une image pour y assigner une valeur d'énergie. Bien que le but premier y soit l'optimisation, cela nous apporte un fondement théorique pour notre proposition d'une évaluation continue entre deux observations que nous pourrons utiliser pour le calcul de l'EHF.

2.5 Conclusion de la revue de littérature

Dans cette revue de littérature nous nous sommes intéressés aux concepts de la maîtrise statistique des procédés pour présenter le domaine des cartes de contrôles dont l'outil Open-Up. Traditionnellement, les cartes de contrôle ne sont pas un socle au calcul de l'EHF, et Open-Up ne déroge pas à la règle. A contrario des cartes de contrôles usuelles, Open-Up

1. En suivant le tutoriel disponible à l'adresse suivante : `https://github.com/vsmolyakov/experiments_with_python`

dans son approche multivariée non paramétrique est un outil visuellement plus intéressant sur lequel le calcul de l'EHF pourrait être intéressant.
Une revue des pratiques sur les techniques d'EHF à par ailleurs été menée : utiliser la relation géométrique entre les variables n'a pas encore été exploitée pour le calcul de l'EHF comme nous souhaitons le faire. L'EHF relève d'une boîte noire possiblement multivariée mais sans pouvoir explicatif et difficilement diagnosticable.
Nous avons ensuite vu les techniques de réduction dimensionnelle qui pourraient nous être utiles pour mettre en avant les caractéristiques saillantes de la géométrie d'une carte de contrôle, avant d'explorer les modèles basés sur l'énergie qui pourraient nous apporter un fondement méthodologique sur lequel construire notre EHF.
Dans le prochain chapitre, nous rentrerons dans l'algorithme proposé pour calculer l'EHF sur la carte de contrôle Open-Up.

CHAPITRE 3 ALGORITHME DE L'ALPINISTE

Dans ce chapitre, on détaille la proposition pour obtenir un EHF à partir d'une carte de contrôle en coordonnées parallèles par densité. Dans la section 3.1 est présentée une vue d'ensemble du processus de fonctionnement. La procédure d'initialisation est présentée section 3.2. Dans la section 3.3 la méthodologie d'estimation incrémentale des densités utilisées est exposée, tandis que l'algorithme de l'alpiniste est formulé dans la section 3.4. La dernière section 3.5 offrira un exemple jouet.

3.1 Vue d'ensemble

L'algorithme traite un ensemble d'observation de notre milieu de fonctionnement et évalue à chaque temps t l'état du système observé. L'ensemble d'observation $O^{(t)} = \{\mathbf{o}^{(1)}, \ldots, \mathbf{o}^{(t)}\}$ désigne les observations disponibles à l'instant t. Le vecteur $\mathbf{o}$ est défini sur $\mathbb{R}^n$ où n est le nombre de variables observées et $\mathbf{o}^{(t)}$ la t-ieme observation. La notation $o_x^{(t)}$ représente la valeur de la x-ième variable à la t-ième observation. Ces observations sont supposées mises à disposition par un système au fil du temps selon une méthode d'échantillonage quelconque. Le temps $t \in \mathbb{N}$ est discrétisé sur les entiers positifs indépendamment de la fréquence d'échantillonage.

Les observations $\mathbf{o}^{(t)}$ sont projetées en coordonnées parallèles [16, 17] : un plan cartésien bidimensionnel dans lequel l'abscisse $1 \leq x \leq n \in \mathbb{N}$ est l'espace des indices de variables et l'ordonnée $y \in \mathbb{R}$ est la valeur de la variable. Ce faisant, on réduit la complexité des interactions entre toutes les variables à celle d'index x voisins. On obtient une fonction de densité bidimensionnelle. Dans ce cadre, plutôt que de s'intéresser à la probabilité jointe de toutes les variables, on va plutôt chercher à estimer la probabilité jointe de variables d'index consécutifs.

Pour ce faire on exploite les caractéristiques géométriques de la fonction de densité en l'approchant par un histogramme dont la construction sera expliquée à la Section 3.2. On peut par la suite incrémentalement calculer ces histogrammes pour les adapter aux variations dans la fonctions de densité suivant les nouvelles observations à chaque nouveau t.

L'algorithme proposé requiert une initialisation en trois étapes : ordonnancement des variables (Section 3.2.1), discrétisation et normalisation de l'espace des valeurs de variables (Section 3.2.2) et enfin suréchantillonage de l'espace des indices des variables (Section 3.2.3). L'étape de suréchantillonage sert alors à restituer la relation entre deux indices x d'une

projection en coordonnées parallèles. Cette initialisation suppose disponible un historique d'observations initial pour estimer les paramètres de l'ordonnancement, de discrétisation et de normalisation des valeurs. Ensuite, on ajuste l'estimation de la densité par un mécanisme incrémental (Section 3.3). Enfin, on applique l'algorithme de l'alpiniste sur les densités estimées (Section 3.4). Cet algorithme repose sur une fonction d'énergie exploitant les caractéristiques géometriques de l'histogramme aux points où une observation coïncide avec ce dernier. On peut alors, disposant de l'énergie associée à une observation, effectuer un choix de classification en anomalie ou non.

3.2 Initialisation

3.2.1 Ordonnancement des variables

On attribue un index à chaque variable selon l'ordonnancement proposé par [2] sur l'ensemble d'observations initial tel que décrit à la Section 2.1.1.

3.2.2 Discrétisation et normalisation de l'espace des variables

La projection en coordonnées parallèles requiert la normalisation de l'espace des variables tel que la valeur de $o_n^{(t)}$ est dans l'intervalle $[0, 1]$. Cela permet de comparer des mesures prises sur des espaces initiaux différents. On cherche ensuite à approcher les densités par un histogramme bidimensionnel : on a besoin d'un espace discrétisé représentant l'intervalle $[0, 1]$ par un ensemble de points, et ce pour chaque index x. Plutôt que d'utiliser une discrétisation uniforme de l'intervalle, on propose d'utiliser SAX une méthodologie d'échantillonage non uniforme ayant permis empiriquement d'obtenir de bons résultats dans d'autres domaines [35], que nous avons décrit à la Section 2.3. On appelle ψ_y la taille de l'ensemble résultant de cet échantillonage et $\tilde{y}$ la version discrétisée d'une valeur y. A ce stade, l'histogramme bidimensionnel est de dimension $n \times \psi_y$.

3.2.3 Suréchantillonage de l'espace des indices des variables

On est intéressé à représenter le lien de dépendance entre chaque paire de variables contigües dans la projection en coordonnées parallèles dans notre histogramme bidimensionnel. Pour ce faire, on ajoute des index intermédiaires pour représenter la transition entre chaque x et $x+1$. Ces index intermédiaires ont une interprétation géométrique et représentent les valeurs prises par une interpolation linéaire réalisée sur les coordonnées parallèles entre les paires de variables.

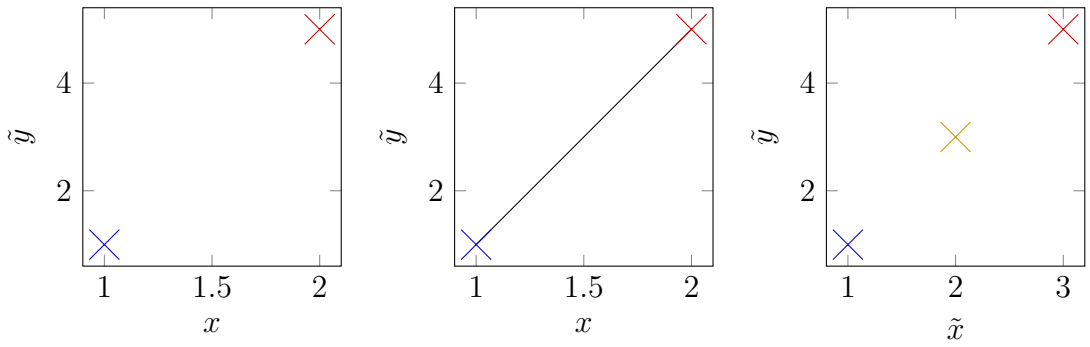


Figure 3.1 Procédure de suréchantillonage

Par exemple, on peut voir Figure 3.1 dans la première sous-figure le cas de départ avec le vecteur observation $\mathbf{o} = \begin{bmatrix} 1 & 5 \end{bmatrix}$. L'interpolation linéaire est visible à la deuxième sous-figure. Enfin pour créer les index intermédiaires on suréchantillonne les indices de variable en utilisant un paramètre ψ_x qui détermine le nombre de suréchantillons prélevés : on prendra $\psi_x - 1$ suréchantillon. C'est le cas sous-figure 3 avec $\psi_x = 2$ ou un index $\tilde{x}$ est utilisé pour designer l'espace des indices de variable suréchantillonné. Après suréchantillonnage, une variable à l'index x devient $\tilde{x} = (\psi_x * (x-1)) + 1$. Les nouveaux index sont définis sur $[0, (\psi_x * (n-1)) + 1]$. A ce stade, l'histogramme bidimensionnel est de dimension $(\psi_x * (n-1) + 1) \times \psi_y$.

3.3 Estimation incrémentale de la densité

On obtient une projection en coordonnée parallèle discrétisée, représentée par une matrice H de taille $a \times b$ définie sur $\mathbb{Z}$ où a est égal à $(\psi_x * (n-1)) + 1$ et b est égal à ψ_y. Le choix de ψ_x et ψ_y sera important : un pas trop petit ne nous permettrait pas d'obtenir un histogramme fidèle aux densités.

On va maintenant incrémenter la matrice H en fonction des observations disponibles dans $O^{(t)}$ afin d'obtenir notre histogramme bidimensionnel. Le pseudo-code pour cette partie est visible à l'Algorithme 1. Cet algorithme fait appel aux paramètres suivants : λ le taux appliqué par la fonction de décroissance exponentielle, $f(H, \lambda)$ une fonction de décroissance exponentielle, $\omega(\mathbf{o}^{(t)})$ un libellé utilisé pour l'apprentissage supervisé. Lorsque $\omega(\mathbf{o}^{(t)})$ est vrai cela signifie que c'est une anomalie, mais une valeur fausse signifie que nous n'avons pas d'information.

On considère que chaque $\mathbf{o}^{(t)}$ de $O^{(t)}$ est discrétisée selon les paramètres d'initialisation retenus. Chaque observation de $O^{(t)}$ est donc une matrice binaire de même dimension que H. Ceci permet de mettre à jour la matrice H, qui représente notre histogramme, en y additionnant ou soustrayant (si $\omega(\mathbf{o}^{(t)})$ est vrai) la nouvelle observation $\mathbf{o}^{(t)}$. On peut le voir à la ligne 5 et

Algorithme 1 : Estimation incrémentale de la densité

Input : H une matrice de dimension $a \times b$
$O^{(t)} = \{\mathbf{o}^{(1)}, \mathbf{o}^{(2)}, \ldots, \mathbf{o}^{(t)}\}$ une séquence d'observations de même support que H
Output : H mise à jour

```
1 for i ← 1 to t do
2     if ω(o^(i)) then
3         H ← H − o^(i)
4     else
5         H ← H + o^(i)
6     H ← f(H, λ)
7 return H
```

ligne 4. Cette estimation de la densité n'est pas normalisée. On propose deux méthodes pour s'adapter aux cas de données non stationnaires et à une potentielle source de supervision.

3.3.1 Fonction de décroissance exponentielle

On cherche à borner les valeurs des éléments de H pour s'assurer que les densités soient normalisées. On propose d'appliquer une fonction de décroissance exponentielle $H \leftarrow f(H, \lambda) = H * e^{-\lambda}$ où λ est le paramètre d'une fonction de décroissance exponentielle. On considère pour simplifier les calculs que la fonction $f(H, \lambda)$ ne sera pas appliquée à chaque nouvelle observation, mais toutes les k observations avec un paramètre λ_k adapté en conséquence : $\lambda_k = \frac{\lambda}{k}$. De plus, H étant définie sur $\mathbb{Z}$, le résultat de la fonction de décroissance exponentielle est arrondi.

3.3.2 Supervision faible

Comme remarqué dans [2], Open-Up tendrait à mieux opérer dans un contexte d'entraînement uniquement basé sur les observations sans anomalies. On émet l'hypothèse que dans un contexte de production, la probabilité d'apparition d'une anomalie sera raisonnablement faible : le système sera plus souvent dans un état fiable que dans un état non fiable. Les anomalies seront donc proportionnellement peu nombreuses. On considère un oracle $\omega(\mathbf{o}^{(t)})$ nous donnant accès à l'information booléenne : vrai si l'observation $\mathbf{o}^{(t)}$ est une anomalie avec certitude, faux sinon. On propose d'adapter ces éléments à l'estimation de la densité par histogrammes par une faible supervision : si $\omega(\mathbf{o}^{(t)})$ est vrai, on diminue de 1 unité la hauteur de l'histogramme pour l'observation $\mathbf{o}^{(t)}$ correspondante.

Ainsi, H est la différence entre deux histogrammes bidimensionnels : $H = H_0 - H_1$, avec

H_0 l'histogramme des observations pour lesquelles $\omega(\mathbf{o}^{(t)})$ est faux, et H_1 l'histogramme des observations pour lesquelles $\omega(\mathbf{o}^{(t)})$ est vrai. Ainsi, la matrice H peut devenir négative sur certains emplacements à cause d'observations pour lesquelles $\omega(\mathbf{o}^{(t)})$ est vrai.

3.4 Calcul de l'EHF : algorithme de l'alpiniste

Le but de l'EHF est de représenter la vraisemblance d'une observation $\mathbf{o}^{(t)}$. Dans notre cas, on considère une fonction d'énergie $E(\mathbf{o}^{(t)}|H)$ définie sur $\mathbb{R}_+$ représentant l'EHF : on considère que plus cette énergie est élevée moins l'EHF sera dans un bon état. Cette fonction est définie pour toute observation $\mathbf{o}^{(t)}$ de même support que H sachant la matrice H dans un état donné. On considère que la fonction d'EHF globale est composée de K fonctions EHF élémentaires de sorte que $E(\mathbf{o}^{(t)}|H) = \sum_{k=1}^{K} E_k(\mathbf{o}^{(t)}|H)$.

La valeur de l'énergie en soit ne permet pas de déterminer si une observation représente une anomalie ou non. Pour ce faire, on cherche un seuil nous permettant de départager les deux classes. Plus l'énergie d'une observation $\mathbf{o}^{(t)}$ sera importante relativement à celles de l'ensemble $O^{(t)}$ plus elle devrait témoigner d'une potentielle anomalie. On va discuter plus amplement du choix du seuil ainsi que de conditions tierces que l'énergie devrait vérifier à la Section 4.3.2.

Algorithme 2 : Algorithme de l'alpiniste

Input : H une matrice de dimensions $a \times b$
$\mathbf{o}^{(t)}$ une observation de même support que H
Output : e une valeur d'énergie pour $\mathbf{o}^{(t)}$
Paramètres : $E(\mathbf{o}^{(t)}|H)$ une fonction d'énergie avec K composantes conditionnées sur H

1 *entier* $e = 0$
2 **foreach** $k \in K$ **do** $e \leftarrow e + E_k(\mathbf{o}^{(t)}|H)$
3 **return** e

Le pseudo-code pour cette partie est visible à l'Algorithme 2. On souhaite donc utiliser la matrice H comme carte de contrôle pour connaître l'EHF d'une observation $\mathbf{o}^{(t)}$. On s'inspire des champs aléatoires de Markov et modèles basés sur l'énergie [37–39] pour proposer une fonction d'énergie associant l'observation $\mathbf{o}^{(t)}$ à la matrice H. Nous n'avons pas accès aux libellés nécessaires pour faire un apprentissage supervisé tel que défini dans [37], nous sommes donc dans un contexte avec une faible supervision. On s'inspire du modèle d'Ising (aussi appelé *pairwise maximum entropy model*) pour proposer heuristiquement une fonction d'énergie par composantes (Eq. 3.1) définies par les équations 3.2 à 3.4, où $|.|$ est la valeur absolue, $\mathbb{1}_p$ la fonction indicateur égale à 1 si p est vrai, 0 sinon, β_1, β_2, β_3 des poids attribués

aux composantes pris dans $\mathbb{R}$, β_0 un biais utilise pour centrer l'énergie.

Chaque composante est déterminée sur chaque indice $\tilde{x}$ pris dans H sur une observation $\mathbf{o}^{(t)}$. On propose $K = 3$ composantes. E_1 (Équation 3.2) est la composante des transitions entre chaque paire de $\tilde{x}$: on veut s'assurer que la densité entre deux index soit stable. Cela nous montre que la transition qui correspond à l'observation actuelle est une transition habituelle. E_2 (Équation 3.3) est la composante de la distance au mode de l'histogramme à l'index $\tilde{x}$ considéré : on veut autant que possible être dans une zone de densité proche de la densité maximum sur l'index $\tilde{x}$. Enfin E_3 (Équation 3.4) est la composante des valeurs négatives : si $\mathbf{o}^{(t)}$ traverse un emplacement de H négatif on sait que c'est typique d'une anomalie et l'énergie doit en conséquence être supérieure. L'équation 3.1 fait appel à la fonction sigmoïde $\sigma(x) \in]0, 1[$ définie comme suit : $\sigma(x) = \frac{1}{1+e^{-x}}$ pour normaliser les résultats, l'ordre étant préservé.

$$E(\mathbf{o}^{(t)}|H) = \sigma(\beta_0 + \beta_1 * E_1(\mathbf{o}^{(t)}) + \beta_2 * E_2(\mathbf{o}^{(t)}) + \beta_3 * E_3(\mathbf{o}^{(t)})) \tag{3.1}$$

$$E_1(\mathbf{o}^{(t)}) = \sum_{\tilde{x}=0}^{a-1} |H(\tilde{x}, o_{\tilde{x}}^{(t)}) - H(\tilde{x}+1, o_{\tilde{x}+1}^{(t)})| \tag{3.2}$$

$$E_2(\mathbf{o}^{(t)}) = \sum_{\tilde{x}=0}^{a} \max_{\bar{y}}(H(\tilde{x}, \bar{y})) - H(\tilde{x}, o_{\tilde{x}}^{(t)}) \tag{3.3}$$

$$E_3(\mathbf{o}^{(t)}) = \sum_{\tilde{x}=0}^{a} \mathbb{1}_{(H(\tilde{x}, o_{\tilde{x}}^{(t)})<0)} |H(\tilde{x}, o_{\tilde{x}}^{(t)})| \,. \tag{3.4}$$

L'algorithme est nommé ainsi car les caractéristiques utilisées pour construire les composantes de la fonction d'énergie peuvent rappeler l'énergie que pourrait engager un alpiniste ($\mathbf{o}^{(t)}$) pour traverser la "montagne" (H).

3.5 Exemple jouet

On considère un système fictif composé de deux capteurs dont on connaît les gammes de mesure : température en C° $\in [-60, 60]$ et humidité en pourcents d'eau dans un mètre cube d'air. On s'intéresse à illustrer l'utilisation d'une carte de contrôle en coordonnées parallèles utilisant l'algorithme de l'alpiniste avec ce système fictif. On suit les étapes d'initialisation, d'estimation incrémentale de la densité et de calcul de l'EHF.

3.5.1 Initialisation

La procédure d'initialisation permet de déterminer les paramètres de normalisation et de discrétisation des observations afin d'obtenir une matrice H qui servira ensuite d'histogramme bidimensionnel. Chaque observation est ainsi projetée en coordonnées parallèles sous forme de matrice binaire de même dimension que H, et la somme des matrices binaires des observations disponibles nous permet d'obtenir l'histogramme.

Dans notre cas de figure nous n'avons que deux variables, on décide donc d'attribuer l'index suivant l'ordre d'énumération ci-dessus : x_1 sera donc la température et x_2 l'humidité. On va normaliser ces variables puis les discrétiser en $\psi_x = 3$ classes 1, 2, 3. Ce choix de ψ_x est arbitraire et n'a pas de signification pour l'exemple.

On calcule donc une valeur normalisée dans l'intervalle $[0, 1]$ pour chaque x suivant la formule

$$x' = (x - x_{min})/(x_{max} - x_{min})\,. \tag{3.5}$$

On discrétise x'_1 et x'_2 suivant SAX. On suppose pour les besoins de cet exemple que la distribution de ces deux variables normalisées suit une loi normale réduite, que l'on peut alors centrer par soustraction d'une valeur constante (0.5). On regarde à l'aire sous la courbe de la loi normale pour décider des limites de chaque intervalle de discrétisation. Comme $\psi_x = 3$, on cherche des intervalles qui totalisent $\frac{1}{3} \approx 33\%$ de l'aire sous la courbe. Une table de la loi normale nous indique que les classes sont séparés aux valeurs $-0.42, 0.42$.

Supposons une observation $\mathbf{o}^{(1)}$ telle que $x'_1 = -0.6$ et $x'_2 = 0.12$. Ainsi, comme x'_1 est dans l'intervalle de la classe 1, alors $\mathbf{o}_1^{(1)} = 1$ et x'_2 est dans l'intervalle de la classe 2, alors $\mathbf{o}_2^{(1)} = 2$. On obtient la matrice de dimension $n \times \psi_y$ (où n est le nombre de variables) suivante

$$\mathbf{o}^{(1)} = \begin{array}{r} \\ \text{Classe 3} \\ \text{Classe 2} \\ \text{Classe 1} \end{array} \begin{array}{c} \begin{array}{cc} x_1 & x_2 \end{array} \\ \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \end{array}. \tag{3.6}$$

Pour terminer l'initialisation, on choisi un paramètre $\psi_x = 2$ pour le suréchantillonage de l'espace des indices de variable. Ce choix est à nouveau arbitraire et n'a aucune signification pour cet exemple. Ceci nous conduit à une matrice H de dimensions $(\psi_x * (n-1)) + 1 \times \psi_y$, soit 3×3. Après interpolation linéaire, on trouve que l'index intermédiaire pour l'observation

$\mathbf{o}^{(1)}$ vaut -0.24, ce qui correspond à la classe 2. La projection de l'observation devient

$$\mathbf{o}^{(1)} = \begin{array}{c} \\ \text{Classe 3} \\ \text{Classe 2} \\ \text{Classe 1} \end{array} \begin{array}{c} \begin{array}{ccc} x_1 & & x_2 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix} \end{array}. \tag{3.7}$$

Nous avons ainsi trouvé les paramètres nous permettant de projeter toute observation de dimension 2 en coordonnées parallèles, sous forme de matrice binaire de même dimension que H. On va utiliser les observations sous cette forme afin de construire l'histogramme bidimensionnel.

3.5.2 Estimation incrémentale de la densité

La matrice $H(3 \times 3)$ qui va nous servir d'histogramme bidimensionnel est initialisée à 0 partout. L'opération d'estimation incrémentale consiste à projeter toute nouvelle observation dans une matrice binaire de même dimension que H puis de faire la somme de H et de cette nouvelle observation projetée. On obtient ainsi un histogramme bidimensionnel. Nous avons déjà projeté $\mathbf{o}^{(1)}$, on peut donc mettre à jour H

$$H' = H + \mathbf{o}^{(1)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}. \tag{3.8}$$

On observe neuf observations supplémentaires afin d'obtenir un histogramme plus fidèle aux densités réelles (qui nous sont inconnues). Par l'opération de projection et discrétisation, certaines observations ont la même matrice, auquel cas ont le signifie en multipliant cette matrice par le nombre d'observations identiques.

$$H' = H + \sum_{i=2}^{10} \mathbf{o}^{(i)} = \tag{3.9}$$

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix} + 4 \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} + 2 \begin{bmatrix} 10 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} + 2 \begin{bmatrix} 10 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 10 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} =$$

$$\begin{bmatrix} 0 & 0 & 4 \\ 2 & 7 & 4 \\ 8 & 3 & 2 \end{bmatrix}$$

On note que dans le cadre de cet exemple avec peu d'observations, on n'applique pas de fonction de décroissance exponentielle. Si cela avait été le cas, cela aurait pris la forme d'une multiplication de H par un coefficient $\lambda < 1$ toutes les k observations.

3.5.3 Calcul de l'EHF

On souhaite maintenant utiliser la matrice H obtenue pour calculer l'énergie d'une nouvelle observation $\mathbf{o}^{(11)}$, qui projetée est :

$$\mathbf{o}^{(11)} = \begin{array}{c} \\ \text{Classe 3} \\ \text{Classe 2} \\ \text{Classe 1} \end{array} \begin{array}{c} \begin{array}{ccc} x_1 & & x_2 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \end{array}. \tag{3.10}$$

Avec pour rappel, une matrice H calculée précédemment telle que

$$H = \begin{array}{c} \\ \text{Classe 3} \\ \text{Classe 2} \\ \text{Classe 1} \end{array} \begin{array}{c} \begin{array}{ccc} x_1 & & x_2 \end{array} \\ \begin{pmatrix} 0 & 0 & 4 \\ 2 & 7 & 4 \\ 8 & 3 & 2 \end{pmatrix} \end{array}. \tag{3.11}$$

Pour rappel, le calcul de l'énergie fait appel à trois composantes que l'on peut indépendamment calculer. Chaque composante est une somme pour chaque index $\tilde{x}$ de H :

- de la différence entre l'index $\tilde{x}$ et l'index $\tilde{x} + 1$ de H suivant là où la matrice de l'observation vaut 1
- pour l'index $\tilde{x}$ considéré, la différence entre la valeur maximum de H et la valeur de H là où l'observation vaut 1
- enfin, si la valeur de H est négative là où l'observation vaut 1, on prend la valeur absolue de cette valeur négative (dans le cas où la supervision faible à été appliquée)

Pour effectuer le calcul, on peut calculer deux matrices intermédiaires : le produit de Hadamard $\odot$ entre H et $\mathbf{o}^{(11)}$

$$H \odot \mathbf{o}^{(11)} = \begin{bmatrix} 0 & 0 & 4 \\ 2 & 7 & 4 \\ 8 & 3 & 2 \end{bmatrix} \odot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 7 & 4 \\ 0 & 0 & 0 \end{bmatrix}, \tag{3.12}$$

que l'on peut simplifier en prenant la valeur absolue maximum par indice de $\tilde{x}$

$$\begin{bmatrix} 0 & 7 & 4 \end{bmatrix}, \tag{3.13}$$

et le maximum de H pour chaque indice $\tilde{x}$

$$\begin{bmatrix} 8 & 7 & 4 \end{bmatrix}. \tag{3.14}$$

On prend alors pour le calcul de la première composante la différence entre chaque valeur du produit de Hadamard simplifié, soit

$$E_1 = |0 - 7| + |7 - 4| = 10\,, \tag{3.15}$$

puis pour le calcul de la seconde la somme de la différence entre la valeur du produit de Hadamard simplifié et la maximum de H pour chaque indice de $\tilde{x}$

$$E_2 = 8 - 0 + 7 - 7 + 4 - 4 = 8\,, \tag{3.16}$$

et enfin prendre la somme sur toutes les valeurs négatives du produit de Hadamard simplifié, soit 0 dans notre cas de figure.

Nous disposons ainsi de la valeur de toutes les composantes de la fonction d'énergie pour l'observation $\mathbf{o}^{(11)}$. Sachant les valeurs $\boldsymbol{\beta}$, on pourrait alors calculer une valeur normalisée dans l'intervalle $[0, 1]$ pour l'observation $\mathbf{o}^{(11)}$ et décider ou non de sa classification en tant qu'anomalie.

Nous venons de voir une mise en pratique de l'algorithme de l'alpiniste au travers d'un exemple jouet. Nous allons maintenant nous intéresser à appliquer l'algorithme pour la détection d'anomalies dans un système de production simulé.

CHAPITRE 4 PROTOCOLES DE TEST

Dans ce chapitre, on présente la méthodologie utilisée pour tester l'algorithme de l'alpiniste couplée à l'estimation incrémentale de la densité. La Section 4.1 présentera le jeu de données utilisé, tandis que la paramétrisation des algorithmes sera abordée Section 4.2, et la Section 4.3 la méthode de classification utilisée et les métriques attendues. Enfin, la Section 4.4 abordera les attentes sur l'estimation de la densité, et la Section 4.5 sur la capacité d'aide au diagnostic du modèle.

4.1 Jeu de données

On propose d'utiliser le jeu de données Tennessee Eastman Process (TEP) [41, 42] pour évaluer l'algorithme de l'alpiniste. Il est intéressant car il reflète un processus de production avec une dépendance temporelle entre chaque observation. Originellement proposé par J. Downs et al. [43] à but de référence pour la Companie Eastman Chemical, TEP est une **simulation** d'un processus de production industriel à l'échelle d'une usine. Chaque journée y est indépendante, et une journée est échantillonnée toutes les 3 minutes. La simulation couvre cinq unités de production pour cinq intrants et trois extrants. Au total, ce sont 41 variables qui sont mesurées pour 12 variables de contrôle : par exemple la pression, température ou encore les niveaux de liquide. Les données sont bruitées, les interactions entre les variables sont complexes, les anomalies peuvent se propager et empirer : c'est un environnement intéressant pour nos tests. L'objectif sera l'observation des dynamiques de l'énergie de l'algorithme de l'alpiniste.

4.1.1 Description du jeu de données

Le jeu données proposé par [41] contient 55 colonnes :

— la première contient un entier entre 0 et 20, 0 indiquant des conditions normales de fonctionnement : il y a donc 20 types d'anomalies différentes. Comme on ne cherche pas à différencier les types d'anomalies ces derniers sont choisis au hasard.

— la seconde colonne contient le numéro de la simulation (cela représente le noyau utilisé pour le simulateur de nombres aléatoires). Chaque numéro de simulation est une journée η **indépendante** des autres. Chaque journée est donc un départ avec des machines dans un état souhaitable.

— la troisième est l'index de l'échantillon prélevé, sachant une fréquence d'échantillonage toutes les **3 minutes**. Les données d'entraînement furent échantillonnées pendant 25 heures, contre 48 heures. Les états fautifs commencent respectivement à 1 et 8h après le début de la simulation fautive. Une journée fautive aura donc 20 observations pour lesquelles $\omega(\mathbf{o}^{(t)}) = 0$ et 480 observations pour lesquelles $\omega(\mathbf{o}^{(t)}) = 1$. On note O^{η} les observations de la journée η.

— les colonnes 4 à 55 contiennent les variables du processus TEP

4.1.2 Utilisation du jeu de données

Un échantillonage aléatoire sera fait parmi les journées disponibles : chaque journée échantillonnée est donc constituée de 500 observations. On observera une séquence de 91 jours tirés aléatoirement parmi les ensembles fautifs et non fautifs.
Le tirage aléatoire est effectue avec 80% de chance de tirer une journée normale. Une table présentant les échantillons tirés aléatoirement est visible en Annexe 1 A. On fait la prédiction de l'énergie d'abord et on considère que le libellé peut être utilisé immédiatement pour faire la mise à jour de H.

4.2 Paramétrisation

Pour la valeur des paramètres de discrétisation et d'estimation incrémentale de la densité, on propose d'utiliser un choix non informé pour mieux refléter un manque de connaissances à priori pour les utilisateurs du modèle. On considère arbitrairement un paramètre extrême pour $\psi_x = 3$ et un paramètre plus large pour $\psi_y = 15$. On suppose que ψ_y a besoin d'un plus grand ensemble que ψ_x pour que les résultats soient corrects : cette supposition mériterait d'être étayée dans un travail futur. Pour $f(H, \lambda_k)$ les paramètres seront $\lambda_k = 0.9$ et $k = 100$. Malgré ce départ à froid, l'ordonnancement des coordonnées parallèles et la normalisation dans $y \in [0, 1]$ seront fait avec les statistiques de l'ensemble du jeu de données pour soustraire un degré de variabilité supplémentaire. Les coefficients β_0, β_1, β_2 et β_3 seront choisis tel qu'expliqué Section 4.3.1.

4.3 Performance de classification

On cherche à savoir dans cette section si l'énergie obtenue est pertinente pour l'utilisation que nous souhaitons en faire.

4.3.1 Modèle de régression logistique

On propose d'optimiser les coefficients β_1, β_2 et β_3 et le biais β_0 par un algorithme supervisé de régression logistique. L'objectif sera alors d'obtenir une frontière de décision linéaire pour séparer les groupes d'observations normales et fautives. Dans ce cas de figure, on ne tient pas compte des mille premières observations pour éviter le biais du départ à froid sur l'estimation de H car $E(\mathbf{o}^{(t)}|H)$ n'a alors pas la même amplitude qu'une observation prise après ces mille premières observations (étant donné les valeurs maximales de la matrice H). On souhaite observer des h proches de leur borne auquel cas $E(\mathbf{o}^{(t)}|H)$ pourrait être maximal. On propose d'utiliser une régression logistique en considérant la formulation de l'Équation 4.1, qui trouve des similitudes avec la fonction d'énergie définie à l'Équation 3.1. On obtient chaque coefficient β_1, β_2 et β_3 et le biais β_0 par les valeurs $\hat{\beta}_{0,1,2,3}$ estimées par la régression logistique. On note $\boldsymbol{\beta}$ le vecteur des coefficients. Soit A_t la variable binaire qui vaut 1 si l'observation $\mathbf{o}^{(t)}$ est une anomalie et 0 sinon. On considère le modèle de régression logistique de la forme

$$Pr(A_t = 1|\mathbf{o}^{(t)}, \boldsymbol{\beta}) = \sigma(\beta_0 + \beta_1 * E_1(\mathbf{o}^{(t)}) + \beta_2 * E_2(\mathbf{o}^{(t)}) + \beta_3 * E_3(\mathbf{o}^{(t)})) = \sigma_\beta(\mathbf{o}^{(t)})\,. \quad (4.1)$$

Il vient $Pr(A_t = 1|\mathbf{o}^{(t)}, \boldsymbol{\beta}) + Pr(A_t = 0|\mathbf{o}^{(t)}, \boldsymbol{\beta}) = 1$. On adopte la notation $\sigma_\beta(\mathbf{o}^{(t)}) = Pr(A = 1|\mathbf{o}^{(t)}, \boldsymbol{\beta})$.

Par indépendance, la vraisemblance pour l'ensemble des observations est

$$L(\boldsymbol{\beta}|\mathbf{o}^n) = \prod_t \sigma_\beta(\mathbf{o}^{(t)})^{A_t}(1 - \sigma_\beta(\mathbf{o}^{(t)}))^{1-A_t}\,. \quad (4.2)$$

Les coefficients β_j $j = 0, 1, 2, 3$ sont estimés en maximisant le log de la vraisemblance, ou de manière équivalente en minimisant le négatif du log de la vraisemblance, c'est-à-dire

$$-\log L(\boldsymbol{\beta}|\mathbf{o}^n) = -\sum_t (A_t \log(\sigma_\beta(\mathbf{o}^{(t)})) + (1 - A_t)\log(1 - \sigma_\beta(\mathbf{o}^{(t)})))\,. \quad (4.3)$$

Pour ce faire, on modifie itérativement le vecteur $\boldsymbol{\beta}$ selon par exemple une descente de gradient, ou encore en se basant sur le score de Fisher, nous nous référons aux manuels des logiciels correspondants pour plus de détails. La classification qui en résulte étant une variable binaire, on cherche une frontière de décision sur $\sigma_\beta(\mathbf{o}^{(t)})$ pour déterminer si $\mathbf{o}^{(t)}$ est une anomalie ou non. Généralement, si $\sigma_\beta(\mathbf{o}^{(t)}) > 0.5$ on considère que l'on a détecté une anomalie. Pour savoir si le modèle de régression logistique est pertinent, on s'intéressera à la signifiance des composantes de la fonction d'énergie, les résultats de classification bruts seront analysés, et on utilisera la courbe sensibilité/spécificité pour connaître une frontière de décision optimale dans notre cas de figure.

4.3.2 Discussions sur l'énergie comme EHF

On cherche à savoir si la fonction d'énergie vérifie les conditions suivantes inspirées de Y. LeCun et al. [37] sur l'apprentissage par énergie reflétant notre but d'obtenir un EHF. On discutera des résultats obtenus dans le contexte où les coefficients ont été optimisés.

Conditions sur l'énergie

Condition 1 - Pour les observation d'index $t = a$ vérifiant $\omega(\mathbf{o}^{(a)}) = 1$ et l'ensemble $O^{(t)\backslash(a)}$ (sans anomalies) l'algorithme donnera une réponse correcte si :

$$E(\mathbf{o}^{(a)}|H) > E(\mathbf{o}^{(t)}|H) \, \forall \mathbf{o}^{(t)} \in O^{(t)\backslash(a)} \, . \tag{4.4}$$

La réponse correcte est atteinte quand l'énergie des anomalies est systématiquement supérieure à l'énergie des observations normales.

Condition 2 - Pour tout couple d'observations contiguës d'un processus avec mémoire vérifiant $\omega(\mathbf{o}^{(t)}) = 1$ et $\omega(\mathbf{o}^{(t+1)}) = 0$ (une anomalie suivie d'une observation normale) l'algorithme donnera une réponse correcte si :

$$E(\mathbf{o}^{(t+1)}|H) << E(\mathbf{o}^{(t)}|H) \, . \tag{4.5}$$

La réponse correcte est atteinte quand l'énergie en sortie d'une anomalie est grandement inférieure à celle de l'anomalie. Par exemple, une maintenance a été effectuée et le système est considéré réparé. Dans notre cas de figure c'est correct car en sortie d'anomalie une observation normale (presque remise à neuf puisque les journées sont indépendantes) est systématiquement tirée.

Condition 3 - Pour tout η l'algorithme donnera une réponse correcte si :

$$E(\mathbf{o}^{(t)}|H) \leq E(o^{(t+1)}|H) \, \forall \mathbf{o}^{(t)} \in O^{\eta} \, . \tag{4.6}$$

La réponse correcte est atteinte quand l'écoulement du temps dans un processus avec mémoire augmente graduellement l'énergie (par exemple, usure). Étant soumis à un bruit local, on accepte que l'énergie puisse varier contrairement à cette condition tant que le résultat sur la journée η est globalement vérifié.

La régression logistique ne met pas de contrainte sur la réalisation des *condition 2* et *condition 3* mais uniquement à la *condition 1*. On pourra tenter d'observer ces deux conditions tierces

sur l'évolution des valeurs prises $\sigma_\beta(\mathbf{o}^{(t)})$ au fil du temps.

Ces observations seront soumises à des incertitudes apportées par les coordonnées parallèles et les dynamiques d'estimation incrémentale de la densité pour la simulation effectuée, ce qui pourrait alors fausser la réalisation de ces conditions. Leur atteinte même partielle nous renseignerait tout de même sur la capacité de l'algorithme de l'alpiniste à obtenir un EHF remplissant ces conditions. Des tests dans un environnement mieux contrôlé que les coordonnées parallèles et d'estimation incrémentale de la densité pourraient pallier à ces incertitudes, tout en nous renseignant sur les paramètres optimaux à retenir.

4.4 Estimation de la densité

Un point d'intérêt pour la performance de l'algorithme de l'alpiniste est la représentation des densités sur lesquelles la fonction d'énergie est appliquée. Or l'estimation incrémentale des densités est conçue pour s'adapter à un contexte non stationnaire de la distribution des données. On s'intéresse à la matrice H à différents points d'intérêts arbitrairement échantillonnés :

— après 10 jours de données, soit 5000 observations

— après 30 jours de données, soit 15000 observations

— après 90 jours de données, soit 45000 observations

L'objectif est d'observer si la matrice H n'est pas soumise à des variations trop importantes. Un cas trop dynamique nous apprendrait que l'estimation serait encline à trop rapidement s'adapter aux variations dans les données, ce qui pourrait sous-entendre un sur-ajustement : on souffrirait alors d'un oubli catastrophique des événements trop anciens. Inversement, un cas trop stationnaire signifierait que les densités apprises ne s'adapteraient pas à un environnement dynamique : l'apport de $\mathbf{o}^{(t)}$ pour un t arbitrairement grand serait insignifiant. Le cas échéant, cela nous donnerait des pistes d'améliorations pour le calcul de H : le choix de λ, ψ_x et ψ_y notamment.

4.5 Aide au diagnostic

On envisage d'utiliser l'algorithme de l'alpiniste dans le contexte des cartes de contrôle tel que présenté à la Section 2.1. Un objectif annoncé de ces méthodes est d'utiliser la carte de contrôle comme un outil de pilotage de la machine : on peut alors détecter des situations anormales et agir dessus. Pour pouvoir agir dessus, les cartes de contrôle univariées exposent explicitement la variable incriminée. Avec l'algorithme de l'alpiniste notre carte de contrôle

est multivariée mais on peut toujours vouloir diagnostiquer les variables incriminées en sus de l'EHF, qui est une métrique globale.

Pour ce faire, on regarde à chaque $\tilde{x}$ le valeur prise par les composantes $E_{1,2,3}$ de l'Équation 4.1. On propose qu'une aide au diagnostic est possible quand la distribution des composantes sur l'axe $\tilde{x}$ s'éloigne d'une distribution uniforme pour laquelle l'entropie est maximale. On cherche donc à savoir si dans un échantillon limité de $\mathbf{o}^{(t)}$ on observe des irrégularité sur $E_{1,2,3}(\mathbf{o}^{(t)}|H)$.

C'est un travail en entonnoir : la valeur d'énergie d'une observation nous dit si cette dernière est à considérer comme une anomalie, puis la distribution $E_{1,2,3}(\mathbf{o}_{\tilde{x}}^{(t)}|H)$ nous indique les $\tilde{x}$ incriminés. Dans notre cas de figure, on ne connaît pas le processus TEP, on ne sait donc pas si les variables incriminées sont les causes réelles de l'anomalie. Notre analyse se borne donc à la distribution de $E_{1,2,3}(\mathbf{o}_{\tilde{x}}^{(t)}|H)$. Dans le cas où cette information serait disponible on propose une métrique adaptée : quelle est la distance minimale sur $\tilde{x}$ entre l'une des variable en cause dans l'anomalie et le $\tilde{x}$ pour lequel $E(\mathbf{o}^{(t)}|H)$ est maximal ? Cela nous donnerait de bonnes informations sur la capacité de diagnostic de l'algorithme.

Pour conclure ces expérimentations, un récapitulatif des résultats sera proposé. L'objectif sera d'analyser les potentielles limites de l'algorithme de l'alpiniste et de proposer des pistes de résolution.

CHAPITRE 5 APPLICATION ET DISCUSSION

Dans ce chapitre, la méthodologie de test présentée au précédent chapitre sera appliquée. La Section 5.1 présentera la performance de classification selon la régression logistique utilisée : la Section 5.1.1 sera dédiée à l'analyse du modèle et des résultats obtenus, tandis que la Section 5.1.2 analysera les limites de l'énergie comme potentiel EHF. La Section 5.2 étudiera la stabilité de la représentation H en fonction du temps, tandis que la Section 5.3 explorera les possibilités d'aide au diagnostic avec le modèle.

5.1 Performance de classification

On procède au calcul de la matrice H selon l'échantillonnage aléatoire présenté au chapitre précédent. A chaque nouvelle observation $\mathbf{o}^{(t)}$ on mesure $E(\mathbf{o}^{(t)}|H)$. Puis on applique le mécanisme de faible supervision pour modifier H en conséquence. On obtient $E^{(t)}_{\{1,2,3\}}$ les composantes d'énergie pour chaque t observé. On applique ensuite une régression logistique pour optimiser $\boldsymbol{\beta}$ et réaliser une classification.

5.1.1 Modèle de régression logistique

Entraînement du modèle

On exclut les 1000 premières observations pour éviter le biais du départ à froid, mais l'ensemble des observations restantes est considéré : on sépare alors les observations entre un ensemble d'entraînement et un ensemble de test. L'entraînement est réalisé avec le paquet *caret* (langage de programmation R).

Analyse du modèle

Le modèle de régression logistique optimise l'Équation 4.3 mentionnée au chapitre précédent : on cherche les coefficients $\beta_i \, i \in \{0, 1, 2, 3\}$. β_0 peut être interprété comme un biais. Les coefficients estimés sont visibles au Tableau 5.1.

On observe que β_0 et β_1 influencent l'énergie négativement. La composante E_1 est donc contre-intuitive : plus elle est forte plus l'énergie est basse. Dans notre cas de régression logistique, la colonne écart type est la racine carrée des diagonales de la matrice de covariance des estimations $\hat{\beta}_j \, , j = 0, 1, 2, 3$. Chaque coefficient obtient une valeur $Pr(> |z|)$ (test de Wald) très faible : on peut donc raisonnablement écarter l'hypothèse nulle qu'il n'y a pas de lien

Tableau 5.1 Coefficients estimés par la régression logistique

Coefficient	Estimation	Écart type	Valeur z	$Pr(>\|z\|)$
β_0	-3.632e-01	5.811e-02	-6.25	4.1e-10
β_1	-5.160e-04	8.900e-06	-57.97	$<$ 2e-16
β_2	4.611e-04	8.967e-06	51.43	$<$ 2e-16
β_3	3.782e-04	3.511e-05	10.77	$<$ 2e-16

entre les composantes de $E(\mathbf{o}^{(t)}|H)$ et l'état normal/anormal d'une observation. En ce sens les composantes proposées sont signifiantes. On remarque que la valeur z (rapport de l'estimation sur l'écart-type) de β_0 et β_3 est environ 5x plus faible que ceux de β_1 et β_2 : cela pourrait nous indiquer que la composante E_3 est moins fiable. En d'autres termes, on pourrait utiliser ces résultats pour raffiner les paramètres de l'entraînement ou modifier les équations des composantes.

On utilise la déviance pour mesurer la vraisemblance du modèle : une déviance plus haute indiquera un modèle moins adapté aux données. On compare la déviance au nombre de degrés de liberté (nombre d'observation moins le nombre de variables explicatives utilisées) : un fort nombre de degrés de liberté est souhaitable pour éviter un biais de surraprentissage. La déviance sous l'hypothèse nulle est de 34923 avec 44499 degrés de libertés (44500 observations moins le paramètre unique de l'hypothèse nulle). L'ajout des 3 composantes à réduit de plus de moitié la déviance : cette dernière passe à 16880 sur 44496 degrés de liberté (nombre d'observations moins les quatre paramètres de la régression logistique). On a ainsi diminué de moitié la déviance avec une perte de 3 degrés de liberté ce que l'on peut raisonnablement considérer comme un bon résultat.

On peut donc conclure que le modèle par composantes semble au moins partiellement adapté à la problématique de classification binaire des observations, connaissant les coefficients optimaux.

Analyse des résultats

Un graphique présentant l'énergie au fil des observations utilisant les coefficients optimaux est visible Figure 5.2. Un échantillonage toutes les 100 observations à été effectué pour rendre le graphique plus lisible. La figure sera analysée en détail à la sous-section suivante. La classification avec comme point de décision classique $\sigma_\beta(\mathbf{o}^{(t)}) > 0.5$ nous donne la matrice de confusion du Tableau 5.2. La précision est alors de $\frac{38143+3333}{44500} = 93.2\%$, tandis que la sensibilité est de 98.46% et la spécificité de 57.86%.

La sensibilité étant le rapport du nombre de vrais positifs sur la somme du nombre de vrais

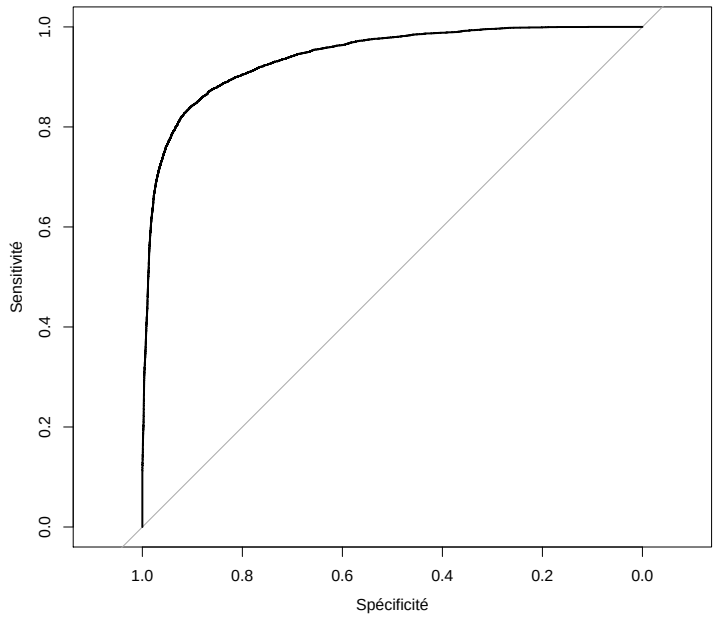


Figure 5.1 Courbe sensibilité/spécificité

positifs et de faux négatifs. La spécificité est le rapport du nombre de vrais négatifs sur la somme du nombre de vrais négatifs et de faux positifs. Généralement, on peut vouloir optimiser le points de décision pour maximiser ces deux métriques. On trace donc la courbe sensibilité/spécificité (aussi appelée courbe ROC (*Receiver Operating Characteristic*)) qui représente l'évolution de l'un sachant l'autre selon le choix du point de décision.

Premièrement, le calcul de l'aire sous la courbe (AUC) ROC visible Figure 5.1 nous donne un indice sur la qualité de notre modèle : on obtient un AUC de 94.1%, ce qui est une valeur relativement proche du classificateur idéal (pour lequel AUC = 100%). Deuxièmement, on sélectionne le point de décision sur lequel sensibilité et spécificité sont maximaux (le point sur la courbe ROC le plus proche du coin en haut à gauche) auquel cas $\sigma_\beta(\mathbf{o}^{(t)}) > 0.166$ est utilisé. La matrice de confusion pour ce cas est visible Tableau 5.3. La précision devient 89.67% mais la sensibilité (90.54%) et la spécificité (83.87%) sont maximales.

Enfin, on pourrait préférer d'autres mesures selon notre cas de figure, notamment l'optimisation du point de décision selon une fonction de coût (si cette dernière est connue), ou encore l'utilisation de la courbe précision/rappel [1] qui serait adaptée à un jeu de données déséquilibré (beaucoup d'observations normales contre peu d'observations fautives). A titre illustratif l'AUC pour la courbe précision/rappel est dans notre cas de 78.7%. Ces métriques font écho

1. Rappel et sensibilité sont deux appellations de la même métrique

à la discussion de la prochaine sous-section.

Tableau 5.2 Matrice de confusion sur les 44500 observations considérées - $\sigma_\beta(\mathbf{o}^{(t)}) > 0.5$

	Référence	
Prédiction	0	1
0	38143 ($\approx 86\%$)	2427 ($\approx 5.5\%$)
1	597 ($\approx 1\%$)	3333 ($\approx 7.5\%$)

Tableau 5.3 Matrice de confusion sur les 44500 observations considérées - $\sigma_\beta(\mathbf{o}^{(t)}) > 0.166$

	Référence	
Prédiction	0	1
0	35074 ($\approx 79\%$)	929 ($\approx 2\%$)
1	3666 ($\approx 8\%$)	4831 ($\approx 11\%$)

5.1.2 Discussions sur l'énergie comme EHF

On se réfère à la Figure 5.2 pour guider notre discussion sur les conditions énoncées à la Section 4.3.2. On rappelle que les observations sont groupées dans des journées soit normales, pour lesquelles toutes les observations vérifient $A_t = 0$ (dans la légende "Obs. normale"), soit fautives auquel cas les vingt premières observations vérifient $A_t = 0$ (dans la légende "Obs. normale"), et le reste $A_t = 1$ (dans la légende "Obs. fautive"). H est mise à jour à chaque nouvelle observation et on s'attend à observer une évolution de H qui dépende des caractéristiques de la journée. On remarque tout d'abord que les conditions 1 à 3 ne sont que partiellement remplies :

— dans la plupart des cas, la condition 1 est observée quand la matrice H à été mise à jour à plusieurs reprises sur un type d'anomalie. Cela se confirme par la croissance importante de l'énergie préalablement à un état stable où les anomalies sont groupées proche de $\sigma_\beta(\mathbf{o}^{(t)}) \approx 1$. L'énergie initiale de ces journées d'anomalies commence avec $\sigma_\beta(\mathbf{o}^{(t)}) \approx 0$. C'est de là que proviennent la plupart des faux négatifs.

— la condition 2 est aussi invalidée au début de la journée suivant une journée d'anomalies. On observe systématiquement un surplus d'énergie : c'est de là que viennent la plupart des faux positifs. Cela pourrait être amélioré en prenant en compte l'information sur la réparation qui intervient naturellement dans le jeu de données (les journées sont indépendantes).

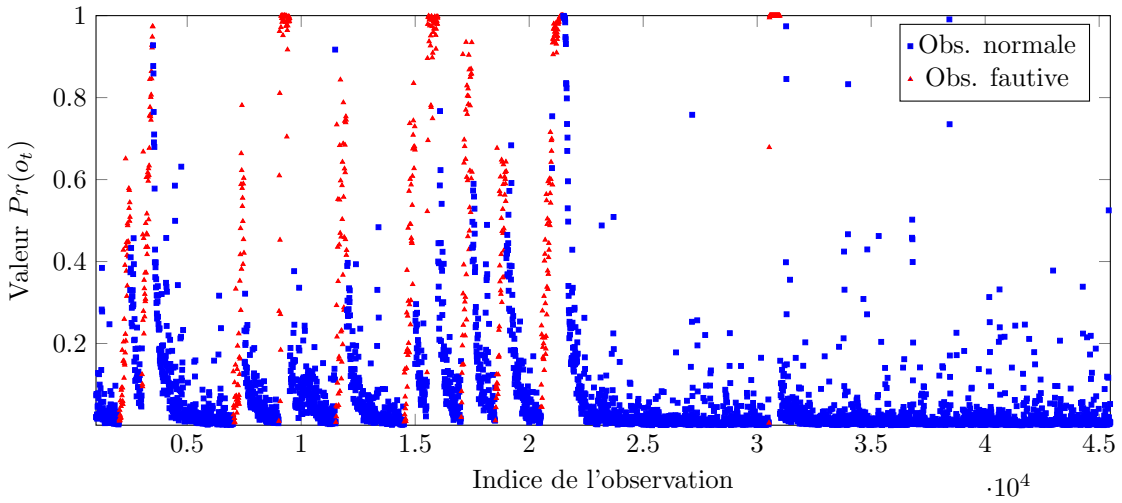


Figure 5.2 Évolution de la fonction d'énergie $E(\mathbf{o}^{(t)}|H)$ (échantillonnage uniforme), vrais libellés

— enfin, aucun mouvement dans les données ne permet de conclure sur la condition 3. Il n'y a pas de dégradation au sein d'une même journée qui nous ferait tendre vers un processus avec mémoire. Un autre jeu de données serait nécessaire.

Pour les conditions 1 et 2 l'estimation incrémentale de la densité aurait un rôle à jouer sur les dynamiques observées. Il faut l'observation répétée d'un état anormal pour que la matrice H nous donne une énergie adéquate pour une nouvelle observation anormale. On peut suggérer que les paramètres utilisés par l'algorithme d'estimation incrémentale fussent mal choisis. De même, concernant la condition 2 on peut suggérer que le pas de discrétisation fut choisi trop petit, ce faisant donnant la même projection dans H à des observations normales et fautives.

5.2 Estimation de la densité

On souhaite à observer si H est relativement similaire sur plusieurs échelles de temps : on s'attend à ce que la mesure de densité demeure globalement stable (le processus observé ne change pas de fonction de distribution chaque jour), mais qu'il puisse localement dévier au fil du temps (on observe un processus dynamique, de légères modifications sont envisageables). Pour ce faire, on regarde à la fin de la 10ème, 30ème et 90ème journée les statistiques sur la matrice H tels que présentés à la Table 5.4.

On peut remarquer les fluctuations sur le minimum et le maximum sont importantes. Le maximum de 900 est presque avoisiné pour la 90ème journée, mais le minimum de -900 ne semble pas approché. La moyenne n'est pas épargnée et fluctue largement. On impute ces écarts à deux éléments : la méthodologie de simulation et les paramètres de l'estimation in-

Tableau 5.4 Statistiques sur H à divers points d'entraînement

	Minimum	1er Quartile	Médiane	Moyenne	3ème Quartile	Maximum
A 10 jours	-95	14	51	83.61	88	698
A 30 jours	-356	-4	1	11.15	14	274
A 90 jours	-4	-4	4	58.63	75	896

crémentale de la densité. Premièrement, le tirage aléatoire de jours pour la simulation laisse les 48 derniers jours sans anomalie. Il y a ainsi une forte concentration d'anomalies dans les 43 premières journées, ce qui pourrait expliquer les variations observées. Secondement, l'estimation incrémentale pourrait être impactée par le paramétrage non optimisé : observation normales et anormales tendraient alors à avoir des similitudes avec une discretisation selon ψ_x, ψ_y utilises. Le mécanisme d'ajout/retrait d'observations serait alors inefficace.

Une meilleure compréhension des coordonnées parallèles et de la typologie des anomalies sachant les paramètres ψ_x, ψ_y nous permettrait de déterminer $\psi_x^\star$ et $\psi_y^\star$ tel que le support de H serait optimal pour la détection de motifs anormaux. Le cas échéant, on s'attendrait à ce que les fluctuations sur H soit maîtrisées par les mécanismes de l'Algorithme 1 d'estimation incrémentale de la densité. Auquel cas le choix de λ n'est pas anodin : dans notre cas de figure il fut choisi indépendamment de la fréquence d'échantillonage des observations et d'apparition des anomalies. Il serait intéressant de proposer des méthodes tirant parti de ces deux éléments pour rendre indépendante l'estimation de H des caractéristiques du jeu de données.

5.3 Aide au diagnostic

Le but de l'aide au diagnostic est d'aider à identifier la/les source(s) de l'anomalie, les variables hors de l'ordinaire sur lesquelles il est possible d'agir pour rectifier la situation. Ces variables sur lesquelles il est possible d'agir sont celles du vecteur dans $\mathbb{R}^n$ qui constituait originellement les observations avant projection en coordonnées parallèles et discrétisation. On peut les retrouver dans la dimension $\tilde{x}$. Or dans notre cas de figure le manque de connaissance sur le processus TEP ne nous permet pas de savoir quelles sont les variables fautives quand une anomalie est observée.

L'objectif sera alors d'estimer si l'entropie de l'énergie sur $\tilde{x}$ est suffisamment faible pour envisager une aide à la décision : on ne sait pas si les variables désignées sont les bonnes, mais il faut que l'on puisse en désigner. Le résultat sera considéré concluant si on peut discerner des $\tilde{x}$ dont l'énergie est suffisamment haute pour la faire sortir du lot. Cela signifierait que l'on pourrait attribuer l'anomalie à un indice pris dans $\tilde{x}$ et donc en déduire le couple ou la

variable dans $\mathbb{R}^n$ incriminée.

Un échantillonage est réalisé pour obtenir chacun des cas de figure de classification possible : 21ème (nommée $\mathbf{o}^{(1)}$) et 42ème (nommée $\mathbf{o}^{(2)}$) observation du 5ème jour, 1ère et 2ème ($\mathbf{o}^{(3)}$ et $\mathbf{o}^{(4)}$) observation du 6ème jour. $\mathbf{o}^{(1)}$ est un faux négatif, c'est aussi la première anomalie observée, tandis que $\mathbf{o}^{(2)}$ est un vrai positif : $\sigma_\beta(\mathbf{o}^{(2)})$ devient suffisamment grand. A l'inverse $\mathbf{o}^{(3)}$ est un faux positif et $\mathbf{o}^{(4)}$ est un vrai négatif. Sans savoir le résultat de la classification, on s'intéresse à la répartition de l'énergie (échelle logarithmique) sur $\tilde{x}$ pour chaque composante pour chaque $\mathbf{o}^{(t)}$.

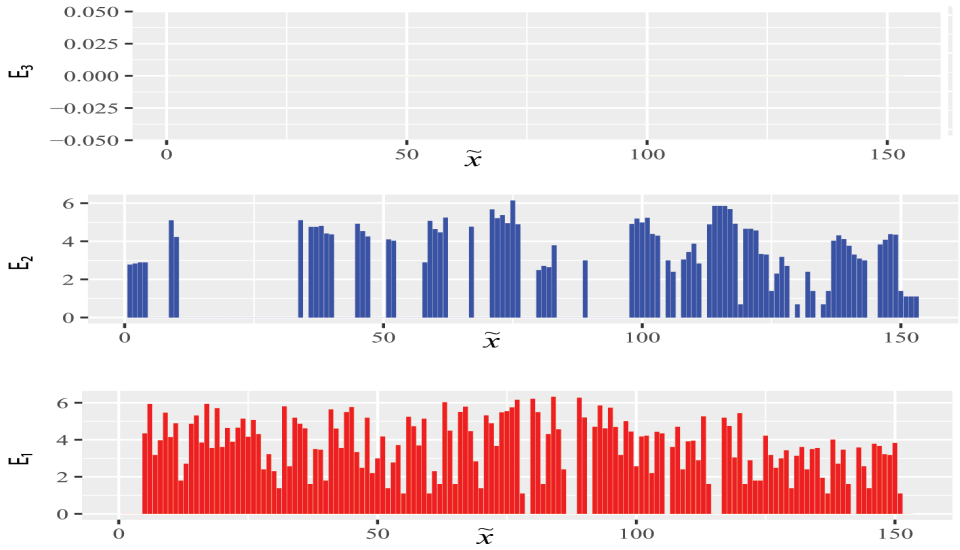


Figure 5.3 $E_{1,2,3}(\mathbf{o}^{(1)}_{\tilde{x}}|H)\,\forall\tilde{x}\in[0,a-1]$

Pour $\mathbf{o}^{(1)}$ (Figure 5.3), on observe que E_3 est nul : c'est un cas de figure normal car la supervision faible n'ayant jamais eu lieu à cet instant il ne peut pas y avoir de valeurs négatives dans la matrice H. E_1 est plus riche d'informations mais on peut néanmoins observer une énergie moyenne sur l'ensemble $\tilde{x}$ qui abouti à un bruit visuel n'aidant pas à la décision. Quelques valeurs extrêmes sont observables mais elle sont trop nombreuses pour conclure. Enfin, E_2 est de loin la composante la plus intéressante : des pics locaux sont observables, et il y a peu de bruit entre ces pics. On peut néanmoins critiquer leur nombre élevé. Cette critique est globalement valable pour $\mathbf{o}^{(2)}$ (Figure 5.4), $\mathbf{o}^{(3)}$ (Figure 5.5) et $\mathbf{o}^{(4)}$ (Figure 5.6). Sur ces observations, on remarque que E_3 est fortement localisé : on peut donc facilement discerner les $\tilde{x}$ d'intérêt pour le diagnostic. De la même manière, E_1 sur $\mathbf{o}^{(3)}$ et $\mathbf{o}^{(4)}$ semble moins bruité : on voit se démarquer des pics quand $\tilde{x}\approx 120$. E_2 semble fluctuer autour d'un bruit acceptable pour le diagnostic. On remarque enfin que les trois composantes se

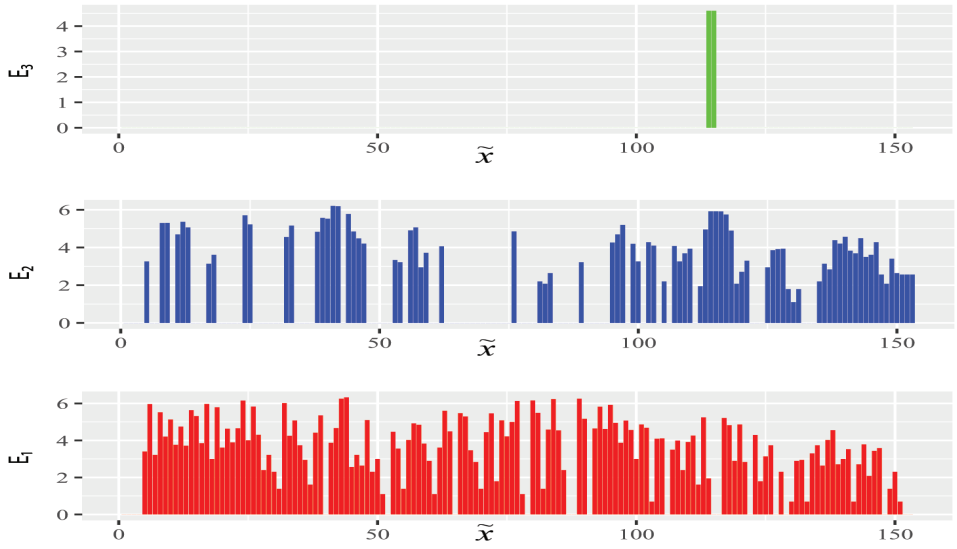


Figure 5.4 $E_{1,2,3}(\mathbf{o}_{\tilde{x}}^{(2)}|H)\,\forall\tilde{x}$

corrèlent visuellement : par exemple pour $\mathbf{o}^{(3)}$ le pic à $\tilde{x} \approx 120$ est partagé par E_1, E_2 et E_3. Pour $\mathbf{o}^{(4)}$ les résultats sont similaires quand $\tilde{x} \approx 0$: on remarque cependant que E_1 est de valeur nulle tandis que E_2 et E_3 sont relativement hauts. Cela corrobore les coefficients estimés par la régression linéaire où β_1 est un coefficient négatif. C'est un résultat contre intuitif mais logique : E_1 caractérisant la transition entre $\tilde{x}_i$ et $\tilde{x}_{i+1}$ pour $\mathbf{o}^{(t)}$, si $\tilde{y}_i \equiv \tilde{y}_{i+1}$ alors $E_1(\mathbf{o}_{\tilde{x}}^{(t)}) = 0$, la première composante sera nulle.

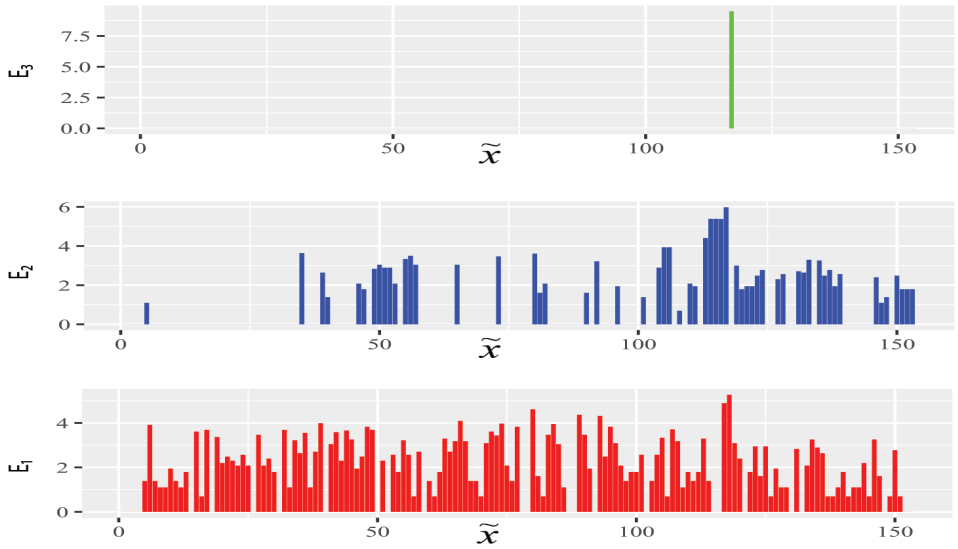


Figure 5.5 $E_{1,2,3}(\mathbf{o}_{\tilde{x}}^{(3)}|H)\,\forall\tilde{x}$

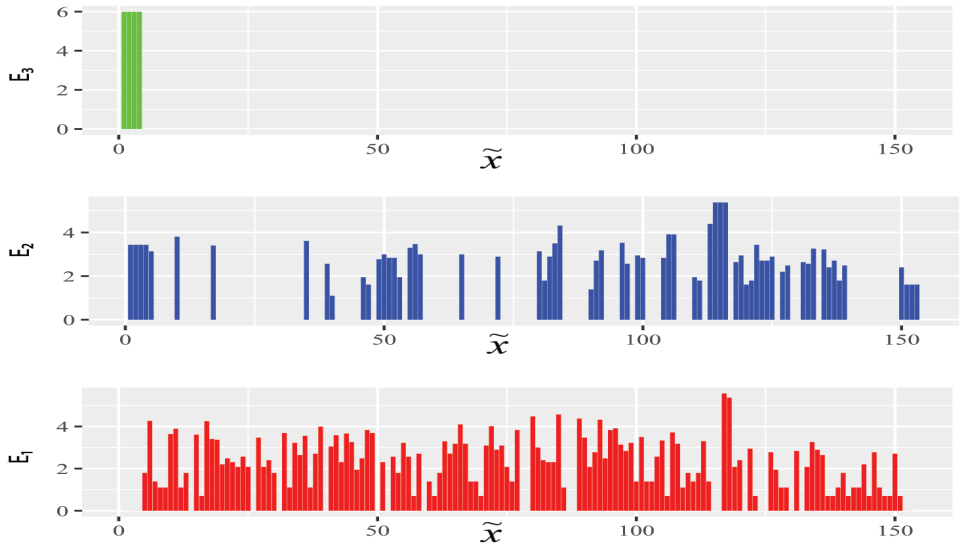


Figure 5.6 $E_{1,2,3}(\mathbf{o}_{\tilde{x}}^{(4)}|H)\,\forall\tilde{x}$

CHAPITRE 6 CONCLUSION ET RECOMMANDATIONS

6.1 Synthèse des travaux

Ce mémoire propose l'algorithme de l'alpiniste pour l'évaluation d'une carte de contrôle en coordonnées parallèles par densité. Cette évaluation permet de déduire une métrique proportionnelle à la probabilité d'observation d'un état du processus, à la manière d'un indicateur de santé de l'équipement (EHF). Le fonctionnement présenté succinctement consiste à utiliser une représentation par un histogramme bidimensionnel de la carte de contrôle en coordonnées parallèles, puis à calculer la force appliquée par l'observation considérée sur les densités discrétisées. L'algorithme est appliqué sur une projection en coordonnées parallèles des variables caractérisant un processus avec mémoire.

Dans les expérimentations, on effectue une estimation incrémentale de la densité discrétisée, et donc non normalisée. On autorise un mécanisme de faible supervision pour tirer parti des libellés potentiel dans l'estimation de la densité, ainsi qu'une décroissance exponentielle pour borner les densités non normalisées. On observe ensuite grâce à l'algorithme de l'alpiniste l'EHF de l'observation qui nous est donnée par une valeur d'énergie. Cette énergie est une somme de trois composantes E_1, E_2 et E_3. Chacune propose une interaction avec l'histogramme bidimensionnel suivant l'observation actuelle : distance au mode de l'abscisse, distance aux valeurs probables de l'abscisse et enfin distance entre deux index de l'abscisse. Un coefficient permet ensuite de moduler l'apport de chacune des composante à l'énergie globale.

Une optimisation sur ces coefficients est conduite au travers d'une régression logistique pour maximiser la détection d'anomalies. Ceci nous permet de jauger de la valeur de l'algorithme proposé connaissant des coefficients idéaux. On observe que les trois composantes sont des caractéristiques suffisantes pour permettre une bonne classification.

6.2 Limites de la solution proposée

En observant de plus près l'énergie obtenue comme une interprétation d'EHF, on se rend compte que l'estimation des densités joue un rôle clé : une densité reflétant des conditions de fonctionnements normales permet une bonne classification. Or, l'estimation incrémentale et le mécanisme de faible supervision, avec un retard dans le temps, semble ne pas aider l'algorithme. Il faut en effet attendre que les anomalies soient détectées pendant une période suffisamment importante pour qu'elles soient classifiées comme telle. Les degrés de libertés

que sont les paramètres utilisés pour la projection en coordonnées parallèle, le choix du pas de discrétisation des densités, les paramètres de la simulation et les mécanismes d'estimation incrémentale s'additionnent : au final il est difficile de discerner lequel de ces élément serait inadéquat, ce faisant rendant floue la critique de l'algorithme de l'alpiniste.

De plus, les résultats de classification obtenus furent corrects mais les paramètres de l'expérimentation ainsi que l'optimisation informée ne reflètent pas des conditions de fonctionnement hypothétiques. On ne peut alors pas conclure sur la validité dans un cadre général des coefficients obtenus : d'autres expérimentation seraient nécessaires. En particulier, le jeu de données simulé Tennessee Eastman Process qui fut utilisé est critiquable : il devait nous rapprocher d'un processus avec mémoire, cas de figure où l'EHF fait du sens. En l'occurrence, les anomalies du jeu de données sont introduites de manière abrupte ce qui tendrait à caractériser un processus sans mémoire. Enfin, on ne tire pas avantage de l'information sur les réparations (en sortie de journée fautive, le processus est comme remis à neuf) : il faudrait déterminer une procédure de "réinitialisation" des densités afin de fixer des zones dans H où une anomalie n'est pas attendue. Dans notre cas de figure, cela nous aurait épargné nombre de faux positifs.

6.3 Améliorations futures

On propose de distinguer deux axes de travail majeurs : la projection en coordonnées parallèles couplée à l'utilisation de l'historique d'observation d'une part, et d'autre part la caractérisation de l'algorithme de l'alpiniste dans un environnement contrôlé.

Comme mentionné plus tôt, les défis et les degrés de libertés introduits par la projection en coordonnées parallèles et le fonctionnement d'un processus de production ne permettent pas de caractériser efficacement l'algorithme de l'alpiniste. La projection en coordonnées parallèles fait appel à un critère d'ordonnancement finalement peu étudié, de même que l'application à des problématiques de processus de production. Le principal travail sur ce critère d'ordonnancement [2] est conscient de ces limites : l'approfondissement de la recherche à ce sujet pourrait donc constituer un travail intéressant. Sur cet axe de travail, il faudrait alors caractériser les anomalies possibles, et comprendre de quelle manière les anomalies apparaissent dans un jeu de données (fréquence, vitesse d'apparition, etc). On pourrait alors déterminer des paramètres robustes pour l'algorithme de l'alpiniste.

On remarque que les résultats obtenus, notamment par la régression logistique, pourront aider a fixer des paramètres initiaux pour de futures utilisations de l'algorithme de l'alpiniste. On peut néanmoins difficilement affirmer que ces paramètres auront une valeur générale sans

études tierces. On peut alors suggérer de regarder au travers de plusieurs jeux de données si les paramètres obtenus permettent à l'algorithme de généraliser à des jeux de données tiers. Plus largement pour l'algorithme de l'alpiniste, on peut suggérer d'avoir une approche plus fondamentale dans la caractérisation de l'algorithme. Puisque nous cherchons à caractériser le fonctionnement d'un processus stochastique on pourrait complètement abandonner la projection en coordonnées parallèles. Des expérimentations avec un processus simulé (un processus Gaussien serait un bon début, puis l'on pourrait généraliser à un processus quelconque) dont les paramètres seraient connus nous permettrait d'observer avec plus de finesse la dynamique de l'énergie obtenue. On éviterait alors la problématique d'estimer la densité d'un processus en coordonnées parallèles (chose qui serait faite dans l'axe de travail 1). Chaque composante pourrait alors être analysée et améliorée en conséquence. En cas de succès, l'algorithme de l'alpiniste pourrait être adapté sur des problématiques avec de grands espaces de recherche pour lesquels des solutions exactes ne sont pas pertinentes.

RÉFÉRENCES

[1] S. R. Montgomery, "Higher Profits from Intelligent Semiconductor-Equipment Maintenance," Rapport technique, 2000. [En ligne]. Disponible : https://pdfs.semanticscholar.org/2599/f6d4d9ee567258045518572c5ba8f64142a8.pdf

[2] S. Tilouche, "Nouvelle approche de maîtrise de processus intégrant les cartes de contrôle multidimensionnelles et les graphes en coordonnées parallèles," Thèse de doctorat, Polytechnique Montréal, 2017. [En ligne]. Disponible : https://publications.polymtl.ca/2953/1/2017_ShaimaTilouche.pdf

[3] G. Javel, *ORGANISATION ET GESTION DE LA PRODUCTION Cours avec exercices corrigés*, 2010.

[4] M. Pillet, *Appliquer la maitrise statistique des processus (MSP/SPC)*. Paris : Editions d'Organisation, 2005.

[5] C. Croarkin, P. Tobias et and others, *NIST/SEMATECH e-Handbook of Statistical Methods*, 2012. [En ligne]. Disponible : http://www.itl.nist.gov/div898/handbook/

[6] *Statistical quality control handbook*. Indianapolis : Western Electric Co., 1956. [En ligne]. Disponible : https://www.worldcat.org/title/statistical-quality-control-handbook/oclc/33858387

[7] L. S. Nelson, "The Shewhart Control Chart—Tests for Special Causes," *Journal of Quality Technology*, vol. 16, n°. 4, p. 237–239, 10 1984. [En ligne]. Disponible : https://www.tandfonline.com/doi/full/10.1080/00224065.1984.11978921

[8] E. S. Page, "CONTINUOUS INSPECTION SCHEMES," *Biometrika*, vol. 41, n°. 1-2, p. 100–115, 6 1954. [En ligne]. Disponible : https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/41.1-2.100

[9] H. Hotelling, "Multivariate Quality Control," *Techniques of Statistical Analysis*, 1947. [En ligne]. Disponible : http://ci.nii.ac.jp/naid/10021322508/en/

[10] F. Le Gall et François, "Powers of tensors and fast matrix multiplication," dans *Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation - ISSAC '14*. New York, New York, USA : ACM Press, 2014, p. 296–303. [En ligne]. Disponible : http://dl.acm.org/citation.cfm?doid=2608628.2608664

[11] H. Li, M. Kazim Khan, A. Tonge, R. Jin et R. S. Meindl, "MULTIVARIATE EXTENSIONS OF CUSUM PROCEDURE," Rapport technique, 1996. [En ligne]. Disponible : https://etd.ohiolink.edu/rws_etd/document/get/kent1185558637/inline

[12] F. Babus, “Contrôle de processus industriels complexes et instables par le biais des techniques statistiques et automatiques,” Thèse de doctorat, 2008. [En ligne]. Disponible : https://tel.archives-ouvertes.fr/tel-00535668

[13] P. Qiu et D. Hawkins, “A Nonparametric Multivariate Cumulative Sum Procedure for Detecting Shifts in All Directions,” p. 151–164, 2003. [En ligne]. Disponible : https://www.jstor.org/stable/4128242

[14] R. Sun et F. Tsung, “A kernel-distance-based multivariate control chart using support vector methods,” *International Journal of Production Research*, vol. 41, n°. 13, p. 2975–2989, 1 2003. [En ligne]. Disponible : https://www.tandfonline.com/doi/full/10.1080/1352816031000075224

[15] S. Bersimis, J. Panaretos et S. Psarakis, “Multivariate Statistical Process Control Charts and the Problem of Interpretation : A Short Overview and Some Applications in Industry,” Rapport technique, 2005. [En ligne]. Disponible : https://arxiv.org/pdf/0901.2880.pdf

[16] M. D’Ocagne, “Coordonnées parallèles et axiales : méthode de transformation géométrique et procédé nouveau de calcul graphique déduits de la considération des coordonnées parallèles,” 1885. [En ligne]. Disponible : https://cds.cern.ch/record/471296/export/hx

[17] A. Inselberg, *Parallel coordinates : visual multidimensional geometry and its applications*. Springer, 2009. [En ligne]. Disponible : https://dl.acm.org/citation.cfm?id=1051759

[18] E. Parzen, “On Estimation of a Probability Density Function and Mode,” *The Annals of Mathematical Statistics*, vol. 33, n°. 3, p. 1065–1076, 9 1962. [En ligne]. Disponible : http://projecteuclid.org/euclid.aoms/1177704472

[19] S. Alaswad et Y. Xiang, “A review on condition-based maintenance optimization models for stochastically deteriorating system,” *Reliability Engineering & System Safety*, vol. 157, p. 54–63, 1 2017. [En ligne]. Disponible : https://www.sciencedirect.com/science/article/pii/S0951832016303714

[20] D. A. Ogden, T. L. Arnold et W. D. Downing, “A multivariate statistical approach for anomaly detection and condition based maintenance in complex systems,” dans *2017 IEEE AUTOTESTCON*, 2017.

[21] S. Fossier et P.-O. Robic, “Maintenance of complex systems — From preventive to predictive,” dans *2017 12th International Conference on Live Maintenance (ICOLIM)*. IEEE, 4 2017, p. 1–6. [En ligne]. Disponible : http://ieeexplore.ieee.org/document/7964123/

[22] E. Zio, P. Baraldi, I. Marton, A. I. Sánchez, S. Carlos et S. Martorell, "Application of Data Driven Methods for Condition Monitoring Maintenance," vol. 33, 2013. [En ligne]. Disponible : www.aidic.it/cet

[23] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, nº. 11, p. 559–572, 11 1901. [En ligne]. Disponible : https://doi.org/10.1080/14786440109462720

[24] R. D. Tobias, "An Introduction to Partial Least Squares Regression," Rapport technique, 2016. [En ligne]. Disponible : https://stats.idre.ucla.edu/wp-content/uploads/2016/02/pls.pdf

[25] C. Krauel et L. Weishäupl, "Multivariate Approach for Equipment Health Monitoring," *IFAC-PapersOnLine*, vol. 49, nº. 12, p. 716–720, 2016.

[26] P. Esling et C. Agon, "Time-Series data mining," *ACM Comput. Surv. 45, 1, Article*, vol. 12, 2012. [En ligne]. Disponible : http://doi.acm.org/10.1145/2379776.2379788

[27] P. Laurinec, "TSrepr R package : Time Series Representations," *Article in The Journal of Open Source Software*, 2018. [En ligne]. Disponible : https://doi.org/10.21105/joss.00577

[28] C. Ratanamahatana, E. Keogh, A. J. Bagnall et S. Lonardi, "A Novel Bit Level Time Series Representation with Implication of Similarity Search and Clustering," dans *Advances in Knowledge Discovery and Data Mining, 9th Pacific-Asia Conference, PAKDD 2005, Hanoi, Vietnam, May 18-20, 2005*, 2005, p. 771–777. [En ligne]. Disponible : http://link.springer.com/10.1007/11430919_90

[29] P. Laurinec, M. Lucka et M. Lucká, *Comparison of Representations of Time Series for Clustering Smart Meter Data Multiple Data Streams Clustering View project Comparison of Representations of Time Series for Clustering Smart Meter Data*, 2016. [En ligne]. Disponible : http://www.ucd.ie/issda/data/commissionforenergyregulationcer/

[30] B.-K. Yi et C. Faloutsos, "Fast Time Sequence Indexing for Arbitrary Lp Norms," dans *Proceedings of the 26th International Conference on Very Large Data Bases*, ser. VLDB '00. San Francisco, CA, USA : Morgan Kaufmann Publishers Inc., 2000, p. 385–394. [En ligne]. Disponible : http://dl.acm.org/citation.cfm?id=645926.671689

[31] E. Keogh, K. Chakrabarti, M. Pazzani et S. Mehrotra, "Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases," *Knowledge and Information Systems*, vol. 3, nº. 3, p. 263–286, 8 2001. [En ligne]. Disponible : http://link.springer.com/10.1007/PL00011669

[32] C. Guo, H. Li et D. Pan, "An Improved Piecewise Aggregate Approximation Based on Statistical Features for Time Series Mining." Springer, Berlin, Heidelberg, 2010, p. 234–244. [En ligne]. Disponible : http://link.springer.com/10.1007/978-3-642-15280-1_23

[33] K.-P. Chan et A. Wai-Chee Fu, "Efficient Time Series Matching by Wavelets," dans *Proceedings of the 15th IEEE international conference on data engineering*, 1999. [En ligne]. Disponible : https://infolab.usc.edu/csci599/Fall2003/Time%20Series/Efficient%20Time%20Series%20Matching%20by%20Wavelets.pdf

[34] J. Lin, E. Keogh et S. Lonardi, "Experiencing SAX : a Novel Symbolic Representation of Time Series," dans *Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, 2003. [En ligne]. Disponible : https://cs.gmu.edu/~jessica/SAX_DAMI_preprint.pdf

[35] E. Keogh, J. Lin et A. Fu, "HOT SAX Finding the Most Unusual Time Series Subsequence : Algorithms and Applications," dans *Fifth IEEE International Conference on Data Mining (ICDM'05)*, 2005, p. 8. [En ligne]. Disponible : http://www.cs.ucr.edu/~eamonn/HOT%20SAX%20%20long-ver.pdf

[36] J. Shieh et E. Keogh, "iSAX Indexing and Mining Terabyte Sized Time Series," dans *14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2008, p. 623–631. [En ligne]. Disponible : http://www.cs.ucr.edu/~eamonn/iSAX.pdf

[37] Y. LeCun, S. Chopra, R. Hadsell, A. Ranzato et F. Jie Huang, "Predicting Structured Data," Rapport technique, 2006. [En ligne]. Disponible : http://yann.lecun.com

[38] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg : Springer-Verlag, 2006.

[39] D. Koller et N. Friedman, *Probabilistic Graphical Models : Principles and Techniques - Adaptive Computation and Machine Learning*. The MIT Press, 2009.

[40] K. P. Murphy, *Machine Learning : A Probabilistic Perspective*. The MIT Press, 2012.

[41] C. A. Rieth, B. D. Amsel, R. Tran et M. B. Cook, "Additional Tennessee Eastman Process Simulation Data for Anomaly Detection Evaluation," 2017. [En ligne]. Disponible : https://doi.org/10.7910/DVN/6C3JR1

[42] ——, "Issues and Advances in Anomaly Detection Evaluation for Joint Human-Automated Systems." Springer, Cham, 2018, p. 52–63. [En ligne]. Disponible : http://link.springer.com/10.1007/978-3-319-60384-1_6

[43] J. Downs et E. Vogel, "A plant-wide industrial process control problem," *Computers & Chemical Engineering*, vol. 17, n°. 3, p. 245–255, 3 1993. [En ligne]. Disponible : https://www.sciencedirect.com/science/article/pii/009813549380018I

ANNEXE A ÉCHANTILLONS UTILISÉS

L'échantillonnage à été réalisé sous R3.5.1 à l'aide du code suivant :

```
faultfree <- rbinom(91, 1, 0.2)
faulty <- faultfree[faultfree == 1]
faultfree <- faultfree[faultfree == 0]

faulty <- sample(seq(1, 20, 1), length(faulty), replace=TRUE)

sim_run <- sample(seq(1, 500, 1), 91, replace=TRUE)
faults <- c(faultfree,faulty)

samples <- data.frame(faults, sim_run)
```

Listing A.1 Code R pour la sélection aléatoires des échantillons

91 échantillons suivant une loi de Bernouilli sont donc générés. Les échantillons *faulty* (ceux différent de 0) sont pris uniformément dans l'ensemble $[\![1, 20]\!]$. Le noyau du générateur de nombre aléatoire est aussi choisi aléatoirement uniformément.

Tableau A.1 Échantillons utilisés

ID sample	Fault	Noyau RNG
1	0	200
2	0	6
3	0	62
4	0	328
5	9	195
6	0	374
7	19	349
8	0	121
9	0	316
10	0	436
11	0	347
12	0	193
13	0	182

Tableau A.1 – *Échantillons utilisés (suite)*

ID sample	**Fault**	**Noyau RNG**
14	0	356
15	4	69
16	0	471
17	0	127
18	0	89
19	12	355
20	0	269
21	0	356
22	0	367
23	0	157
24	17	275
25	0	69
26	0	438
27	0	368
28	0	277
29	0	416
30	16	115
31	0	238
32	12	86
33	0	74
34	0	49
35	17	436
36	0	456
37	0	304
38	15	436
39	0	31
40	0	240
41	0	108
42	3	126
43	7	74
44	0	125
45	0	266
46	0	387

Tableau A.1 – *Échantillons utilisés (suite)*

ID sample	**Fault**	**Noyau RNG**
47	0	417
48	0	392
49	0	434
50	0	214
51	0	274
52	0	360
53	0	412
54	0	381
55	0	426
56	0	230
57	0	283
58	0	105
59	0	149
60	0	452
61	0	344
62	6	333
63	0	84
64	0	165
65	0	410
66	0	415
67	0	389
68	0	136
69	0	89
70	0	403
71	0	323
72	0	379
73	0	497
74	0	473
75	0	79
76	0	279
77	0	332
78	0	499
79	0	137

Tableau A.1 – *Échantillons utilisés (suite et fin)*

ID sample	**Fault**	**Noyau RNG**
80	0	166
81	0	210
82	0	385
83	0	73
84	0	148
85	0	449
86	0	132
87	0	273
88	0	315
89	0	375
90	0	385
91	0	201