



## OFFRE D'ALLOCATION DE THESE / PhD GRANT

ÉCOLE DOCTORALE SCIENCES EXACTES ET LEURS  
APPLICATIONS - ED 211 / NATURAL SCIENCES DOCTORAL SCHOOL  
Avenue de l'université BP 1155 64 013 PAU Cedex – France

### SUJET DE THESE / PhD SUBJECT

#### TITRE / TITLE:

**Cartes de contrôle multivariées non paramétriques pour la détection de changement en ligne -  
Application à la surveillance de données environnementales**

/

**Nonparametric multivariate control charts for online change detection - Application to the  
monitoring of environmental data**

#### RESUME :

Cette proposition de thèse concerne le domaine de la statistique mathématique.

L'objectif est de développer une méthodologie statistique pour la détection, en ligne, d'un changement dans une série de données. A l'origine, ces approches, appelées cartes de contrôle, étaient destinées à la supervision de procédés industriels. Aujourd'hui, elles sont également employées dans un contexte de surveillance environnementale (surveillance de la qualité de l'eau, surveillance des ressources aquatiques, etc.).

L'accent sera mis sur le développement de méthodes à la fois non paramétriques et multivariées. Une étude théorique des propriétés de ces cartes de contrôle sera menée et des simulations de Monte Carlo permettront d'évaluer les performances des cartes à échantillon fini.

Enfin, une partie de la thèse concernera l'application de ces approches sur des données environnementales (surveillance de la qualité de l'eau, par exemple).

#### ABSTRACT:

This PhD proposal pertains to the area of mathematical statistics.

The aim of the proposed research project is to develop statistical methodology for online change detection in multivariate data. Originally, these approaches, called control charts, were intended to the supervision of industrial processes. Today, they are used for the monitoring of environmental data (water quality monitoring, monitoring of aquatic resources, etc.).

The focus will be on the development of nonparametric and multivariate methods. From a theoretical perspective, the asymptotic validity of the proposed methods will be mathematically demonstrated and Monte Carlo experiments will be conducted in order to investigate the finite-sample performance of the developed methods.

The project will be concluded with at least one application to environmental data (water quality monitoring for example).

**Mots clés : Statistique mathématique, détection en ligne de changement, cartes de contrôle non paramétriques, données environnementales.**

**Keyword: Mathematical statistics, Online change detection, nonparametric control charts, environmental data.**

### CONDITIONS D'EXERCICE / WORKING CONDITIONS

**Laboratoire : Laboratoire de Mathématiques et de leurs Applications de Pau (LMAP)**

**Site web : <https://lma-umr5142.univ-pau.fr/fr/index.html>**

**Directeur de thèse (PhD Director): Ghislain VERDIER**

<b>Lieu (Place) : Université de Pau et des Pays de l'Adour, site de Pau</b>	
<b>Date début (start): 1<sup>er</sup> Janvier 2021 (au plus tard)</b>	<b>Durée (duration): 3 ans (years)</b>
<b>Employeur (employer): Université de Pau et des Pays de l'Adour (UPPA)</b>	
<b>Salaire mensuel brut (monthly salary before taxes): 1768 €</b>	

## SAVOIR-FAIRE DU LABORATOIRE / HOST LABORATORY PROFILE

Un des axes de recherche de l'équipe Probabilités et Statistique du LMAP concerne la fiabilité et la sûreté de fonctionnement. Plusieurs chercheurs de l'équipe disposent donc d'un savoir-faire en matière de cartes de contrôle, ou plus généralement dans le domaine de la détection de rupture de modèle, en ligne ou hors ligne (Paoissin et al. (2016), Verdier (2013) par exemple). Enfin, un des domaines d'application des méthodologies statistiques étudiées dans l'équipe Probabilités et Statistique concerne l'environnement et la biostatistique.

*Reliability theory and surveillance models are a core activity for the Probability and Statistics team of the LMAP. Several researchers of the team therefore have expertise in control charts, or more generally in the field of model change detection, online or offline (Paoissin et al. (2016), Verdier (2013) for example). Finally, one of the fields of application of the statistical methodologies studied in the Probability and Statistics team concerns environment and biostatistics.*

## MISSION - ACTIVITES PRINCIPALES / MISSION – PRINCIPAL ACTIVITIES

### I. Le contexte scientifique / *Scientific Context*

Les premières cartes de contrôle ont été mises au point par Shewhart en 1931. Depuis, de nombreuses cartes ont été proposées pour des données univariées ou multivariées (voir Montgomery (2007) pour un aperçu). Le but de ces méthodes est de détecter un changement, quand il a lieu, dans la distribution de probabilité d'une série d'observations, le plus rapidement possible. Pour les méthodes les plus classiques, il est souvent supposé que les données « sous contrôle » sont des réalisations d'une loi normale. Evidemment, il s'agit là d'une hypothèse très restrictive, notamment lorsque la surveillance concerne plusieurs paramètres mesurés en même temps. En effet, une hypothèse de normalité dans le cas multivarié est beaucoup plus restrictive que dans le cas univarié.

Au cours des 25 dernières années, de nombreux travaux ont été menés pour proposer des cartes de contrôle non paramétriques, c'est-à-dire des cartes qui puissent être utilisées sans qu'aucune hypothèse paramétrique ne soit faite sur la loi des observations (voir Chakraborti and Graham (2019) et Qiu (2018)). La plupart de ces approches sont basées sur les rangs des observations (Ross et Adams (2012) par exemple). Cependant, une minorité d'entre elles permettent de détecter tous types de changement dans la distribution.

*The first control chart was proposed by Shewhart in 1931. Since then, many control charts have been introduced for univariate or multivariate data generating processes; see Montgomery (2007) for an overview. The aim of these methods is to detect a change (when it occurs) in the probability distribution of the produced series of observations as soon as possible (the sooner the better). In most cases, when "under control", the data generating process distribution is assumed to be normal. Obviously, this is a very restrictive assumption, notably when the monitoring consists of the supervision of several parameters at the same time. Indeed, in the multivariate case, the normality assumption is even less likely to hold than in the univariate case.*

*Over the last 25 years, many studies were carried out to propose nonparametric control charts, that is, control charts that can be used when no parametric assumptions on the observations (such as normality) are made (see Chakraborti and Graham (2019) and Qiu (2018) for an overview). Most existing nonparametric control charts are based on the ranks of the observations; see, for example, Ross and Adams (2012). However, only a minority of approaches are designed to be able to detect any type of change in the underlying distribution.*

## **II. Les objectifs / Objectives**

L'objectif de ce projet de thèse est de développer de nouvelles cartes de contrôle non paramétriques, c'est-à-dire sans qu'aucune hypothèse sur la distribution des observations ne soit formulée, mais également multivariées. Par ailleurs, alors que beaucoup de cartes de contrôle se focalisent sur des types de changement particuliers (un changement de moyenne ou de variance), les méthodes développées dans cette thèse pourront concerner des changements plus généraux, comme un changement de distribution, ou un changement dans la structure de dépendance des observations. Ces cartes de contrôle pourraient être construites, par exemple, à partir de la fonction caractéristique empirique.

Des résultats théoriques sont attendus. Plus précisément, une étude asymptotique des propriétés de ces cartes de contrôle sera menée. Elle consiste à étudier ce qui se passe lorsque la taille de l'échantillon d'apprentissage est infiniment grand. Parallèlement, des simulations de Monte Carlo permettront d'évaluer les performances des cartes pour des échantillons d'apprentissage de taille finie. Ce dernier point est complémentaire à l'étude théorique : savoir qu'une méthode fonctionne asymptotiquement est rassurant, mais vérifier que ses performances restent satisfaisantes lorsque la quantité de données disponibles est « raisonnable » ne l'est pas moins.

Enfin, une partie de la thèse concernera l'application de ces approches sur des données environnementales collectées à l'occasion de projets dans lesquels des chercheurs de l'équipe de Probabilités et Statistique du LMAP sont, ou ont été, impliqués.

*The objective of the PhD is to derive multivariate control charts for detecting online changes that are nonparametric. Moreover, while many control charts focus on specific types of change (a change in mean or variance), the methods developed in this thesis will be able to detect more general changes, such as changes in the distribution, or changes in the dependency structure of the observations. These control charts could be based, for example, on the empirical characteristic function.*

*An important objective of this project is to state mathematical conditions under which the proposed control charts are (asymptotically) valid. It consists in studying what happens when the amount of data increases "to infinity". In addition to the aforementioned theoretical investigations, the finite-sample performance of the proposed procedures will also be investigated through simulation experiments. The latter is complementary to the theoretical results: knowing that a method works asymptotically is of course reassuring, but verifying that its performance is good when the amount of data is "reasonable" is not less important.*

*Finally, the PhD project will end with at least one application to environmental data collected within other projects in which researchers from the Probability and Statistics team of the LMAP are involved.*

## **III. Plan de travail / Work plan**

La première étape de cette thèse consistera à effectuer une étude bibliographique sur les cartes de contrôle, tout en acquérant les techniques probabilistes et statistiques en vue d'obtenir les résultats théoriques évoqués précédemment. Une première carte de contrôle basée sur la fonction caractéristique empirique pour des données iid (indépendantes et identiquement distribuées) pourra être considérée (premiers tests numériques).

La possibilité d'étendre cette approche à des données dépendantes pourra être étudiée. Parallèlement, un travail initial sur une application sera mené. Enfin, la dernière partie de la thèse sera consacrée à la finalisation de l'application et à l'écriture du manuscrit.

*The first step of this PhD will be the study of the existing literature on control charts, and the learning of the probabilistic and statistical techniques for the planned theoretical derivations. Then a first control chart based on the empirical characteristic function in an iid (independent and identically distributed) context could be considered (first implementations and tests).*

*The possibility of extending the proposed control chart to dependent observations could then be studied. In addition, initial work on an application will be done.*

*Finally, the last year of the PhD will be devoted to the finalization of the application and the writing of the PhD manuscript.*

#### **IV. Références bibliographiques (Literature References)**

- Montgomery, D. C. (2007). Introduction to statistical quality control. John Wiley & Sons.
- Paroissin, C., Penalva, L., Pétrau, A., & Verdier, G. (2016). New control chart for monitoring and classification of environmental data. *Environmetrics*, 27(3), 182-193.
- Qiu, P. (2018). Some perspectives on nonparametric statistical process control. *Journal of Quality Technology*, 50(1), 49-65.
- Ross, G. J., & Adams, N. M. (2012). Two nonparametric control charts for detecting arbitrary distribution changes. *Journal of Quality Technology*, 44(2), 102-116.
- Shewhart, W. A. (1931). Economic control of quality of manufactured product. ASQ Quality Press.
- Verdier, G. (2013). Application of Copulas to Multivariate Control Charts, *Journal of Statistical Planning and Inference*, 143, 2151-2159.

#### **COMPETENCES REQUISES / REQUIRED COMPETENCES**

- Master Probabilités / Statistique, ou tout diplôme de niveau équivalent
- Le candidat à cette thèse devra disposer de bonnes compétences en algorithmique et dans des langages de programmation comme R, par exemple.
- *Master in Probability and Statistics, or any equivalent degree*
- *The PhD candidate will need good algorithmic skills and the knowledge of statistical or data analysis languages such as R.*

#### **CRITÈRES D'ÉVALUATION DE LA CANDIDATURE / CRITERIA USED TO SELECT CANDIDATE**

Processus de sélection (Selection process steps):

- Constitution d'un Jury de sélection. (Establishment of the selection committee.)
- Sélection des candidats sur dossier de candidature. (evaluation of the applicants cv's)
- Audition des candidats et classement. (Interview with the selected candidates and ranking.)

Critères d'évaluation de la candidature (Criteria used in selection of the candidate):

- La motivation, la maturité scientifique et la curiosité du candidat. (The candidate's motivation, scientific maturity and curiosity.)
- Ses notes et son classement en M1 et en M2. (candidate's marks and rankings in M1 and M2.)
- Maîtrise de l'anglais. (English proficiency)

#### **CONSTITUTION DU DOSSIER DE CANDIDATURE / REQUIRED DOSSIER,**

Envoyer par email un dossier de candidature comprenant (send an e-mail with your candidature containing):

- CV (CV)
- lettre de motivation (cover letter detailing candidate's motivations)
- Copie du diplôme (copy of the diploma)
- Master ou diplôme équivalent : relevé détaillé des notes obtenues dans chaque matière et classement (candidate's MSc or equivalent : marks and ranking)
- lettres de recommandation (any letters of recommendation)
- coordonnées de personnes du milieu professionnel (minimum deux) à contacter (contact details for 2 referees)

**DATE LIMITE DE DEPOT DU DOSSIER (limiting date)**

**24/10/2020**

**CONTACTS e-mail :**

[ghislain.verdier@univ-pau.fr](mailto:ghislain.verdier@univ-pau.fr)