



MASTER OF SCIENCE  
IN ENGINEERING

**Hes·SO**

Haute Ecole Spécialisée  
de Suisse occidentale

Fachhochschule Westschweiz

University of Applied Sciences and Arts  
Western Switzerland

Master of Science HES-SO in Engineering  
Av. de Provence 6  
CH-1007 Lausanne

# Master of Science HES-SO in Engineering

Orientation : Technologies de l'information et de la communication (TIC)

## AI-enhanced LoRa Based Indoor Localization System

Fait par

**Manon Racine**

Sous la direction de

**Dr. Nuria Pazos**

**HE-Arc**

St-Imier, HE-Arc, 21 octobre 2018



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Contexte . . . . .	1
1.2	Aspect Novateur . . . . .	2
1.3	Objectifs et tâches à réaliser . . . . .	2
1.4	Structure du rapport . . . . .	2
<b>2</b>	<b>Etat de l'art</b>	<b>3</b>
2.1	Extrait de choses importantes . . . . .	3
2.1.1	Type d'apprentissage . . . . .	3
2.1.2	Algorithme . . . . .	3
2.1.3	Outliers [FP17] . . . . .	6
2.2	Comparaison . . . . .	7
2.3	Choix . . . . .	7
	<b>Bibliographie</b>	<b>9</b>



# Table des figures

1.1	Etat de l'art des différentes méthode de positionnement intérieur [LLI]	1
2.1	Comment choisir un algorithme ML [Sci]	4
2.2	a) exemple de classification b) exemple de regression	4



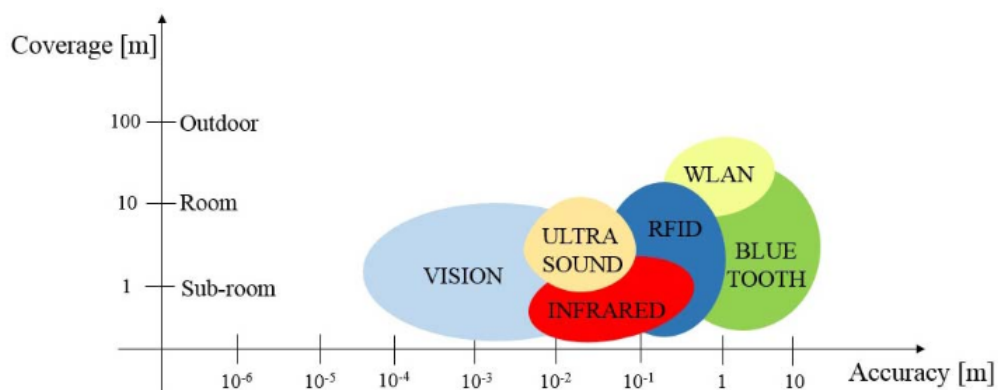
# 1 Introduction

## 1.1 Contexte

La précision du positionnement intérieur reste quelque chose de très important car il peut être utile dans plusieurs domaines comme : La gestion de stocks, la localisation de personnes âgées dans des homes, la localisation chez eux des personnes possédant un bracelet de "prisonnier", etc...

Les systèmes de localisation basés sur GPS souffrent de la détérioration de la précision et sont presque indisponibles dans les environnements intérieurs. Pour les environnements intérieurs, de nombreuses technologies de systèmes de positionnement ont été conçues sur la base de la vision, de la détection infrarouge ou ultrasonore, des champs magnétiques de la terre, des accéléromètres / gyromètres, des balises BLE ou de la communication WiFi. Bien que la création de ces nouvelles applications ait été couronnée de succès, le coût de ces récepteurs, leur consommation d'énergie et leur limitation aux environnements extérieurs excluent de nombreuses applications.

La figure 1.1 montre un graphique comparatif de la précision de positionnement concernant différentes technologie [LLI].



**Figure 1.1** – Etat de l’art des différentes méthode de positionnement intérieur [LLI]

La géolocalisation avec LoRa est une possibilité séduisante et probablement l’un des meilleurs candidats pour le positionnement intérieur. Le faible coût des infrastructures et des noeuds finaux ainsi que la disponibilité à l’échelle de la ville ou du pays pourraient permettre de nombreuses nouvelles applications. Il n’est donc pas surprenant que les chercheurs et les entités commerciales se soient mis au travail sur ce problème au cours des derniers mois. Cependant, plusieurs défis restent à relever pour qu’un tel système devienne pratique. Premièrement, la précision de localisation en extérieur qui peut actuellement être atteinte avec LoRa est comprise entre 30 et 50 mètres, ce qui n’est pas suffisant pour de nombreuses applications en milieu urbain. Deuxièmement, très peu d’expérience est disponible pour la conception de systèmes de positionnement intérieurs avec LoRa.

La portée en milieu rural est d’environ 15km alors qu’il est de 5km dans un milieu urbain cela grâce à la bonne sensibilité du récepteur (-130dBm). Une chose intéressante est la bande passante qui est plus large que d’autres technologies qui permet de distinguer différents chemins du même signal. Sagemcom ont obtenus des bons résultats au niveau de la précision qui est de environ 4 mètres.[FP17]

L’objectif général de ce projet est de développer les technologies permettant d’améliorer la précision de la géolocalisation en intérieur sur la base de la technologie LoRa.

Le positionnement intérieur reste quelque chose de très intéressant et important car il peut être utile dans plusieurs domaines comme : La gestion de stocks, la localisation de personnes âgées dans des homes, la localisation chez eux des personnes possédant un bracelet de "prisonnier", etc...

## 1.2 Aspect Novateur

Ce travail d'approfondissement doit évaluer une nouvelle approche pour améliorer le positionnement intérieur. En s'appuyant sur les capacités étendues des nouveaux circuits intégrés LoRa, ce projet développera et déploiera un système de localisation capable d'améliorer la précision de la position atteinte par les systèmes de localisation basés sur LoRa existants reposant sur des mécanismes TDOA ou de "ranging". À cette fin, une exploration et une comparaison des différentes techniques "machine learning/deep learning" pour le positionnement basé sur le "fingerprinting" seront effectuées.

L'aspect novateur du projet et d'intégrer un mécanisme d'apprentissage de la position afin d'améliorer la précision.

## 1.3 Objectifs et tâches à réaliser

1. Etudier le cahier des charges
2. Etudier l'état de l'art des techniques à utiliser dans le cadre du projet, en particulier les systèmes de localisation indoor basés sur des techniques d'apprentissage, et réunir une documentation (env. 20
3. Etablir un planning pour l'ensemble du projet.
4. Définir un plan des tests à effectuer.
5. Définir les procédures de test
6. Définir le setup pour la collecte de données de localisation
7. Prise en main de l'environnement de développement pour les phases de training et du test de la technique d'apprentissage retenue (e.g., PyTorch).
8. Implémentation de la solution ML retenue.
9. Tester le système selon le protocole préétabli.
10. Faire des propositions pour améliorer les performances de l'algorithme et, si possible, les implémenter.
11. Rédiger le rapport et documenter l'ensemble du projet.

## 1.4 Structure du rapport

Compléter cette partie en fin de rapport quand la structure est définie



## 2 Etat de l'art

Ce chapitre traite de l'état de l'art sur ce qui existe déjà en matière de "Machine Learning" concernant le positionnement intérieur. Il est important de voir ce qui se fait afin d'économiser un temps précieux pour ne pas partir dans une mauvaise direction et ainsi pouvoir être plus efficace dans la recherche.

Pour ce faire, quatre ouvrages ont été étudiés :

1. A Comparative Study on Machine Learning algorithms for Indoor Positioning [Sin+15]
2. A Machine Learning Approach to Ranging Error Mitigation for UWB Localization [Hen+12]
3. A performance guaranteed indoor positioning system using conformal prediction and the WiFi signal strength [KN17]
4. GPS-free Geolocation using LoRa in Low-Power WANs [FP17]

Les trois premiers traitent du positionnement intérieur aidé avec des algorithmes de "Machine Learning". Le quatrième a été sélectionné car il traite de la gestion de "outliers", c'est à dire comment détecter des points abstraits et qui pourraient fausser les mesures.

### 2.1 Extrait de choses importantes

Cette section permet de mettre en évidence les aspects importants qui ressortent des différentes lectures.

#### 2.1.1 Type d'apprentissage

Il existe plusieurs catégories d'estimateur et dans ces catégories, il y a plusieurs algorithmes à choix. La figure 2.1 permet de sélectionner différents algorithmes en fonction de ce qu'il est nécessaire d'obtenir.

Dans le cas de ce travail, il y a deux possibilités. D'utiliser une classification ou d'utiliser la régression. Dans la figure 2.2, une petite explication de la différence des deux types d'estimateur.

Dans le cas de la classification, on va apprendre à notre algorithme différentes positions et l'algorithme sera ensuite capable de fournir en sortie une de ses positions. Donc, en d'autres mots, il sort une région plutôt qu'une coordonnée. Si l'entraînement de l'algorithme a été fait avec un maillage fin alors le résultat sera fin. Lorsqu'un point est détecté il sera classifié selon la position la plus proche comme on le voit avec le carré vert et le carré rouge.

Dans une régression le problème est un peu différent. L'algorithme sera également entraîné avec différentes positions mais ensuite, par régression l'algorithme va tenter d'estimer la position du nouveau point et non plus de le positionner dans une classe.

Afin de clarifier ces deux options la lecture des trois documents cités ci-dessus va aider [Sin+15][Hen+12][KN17].

#### 2.1.2 Algorithme

Dans cette section, quelques algorithmes de classification utilisés dans ces études [FP17] [Hen+12] [KN17] sont brièvement décrits.

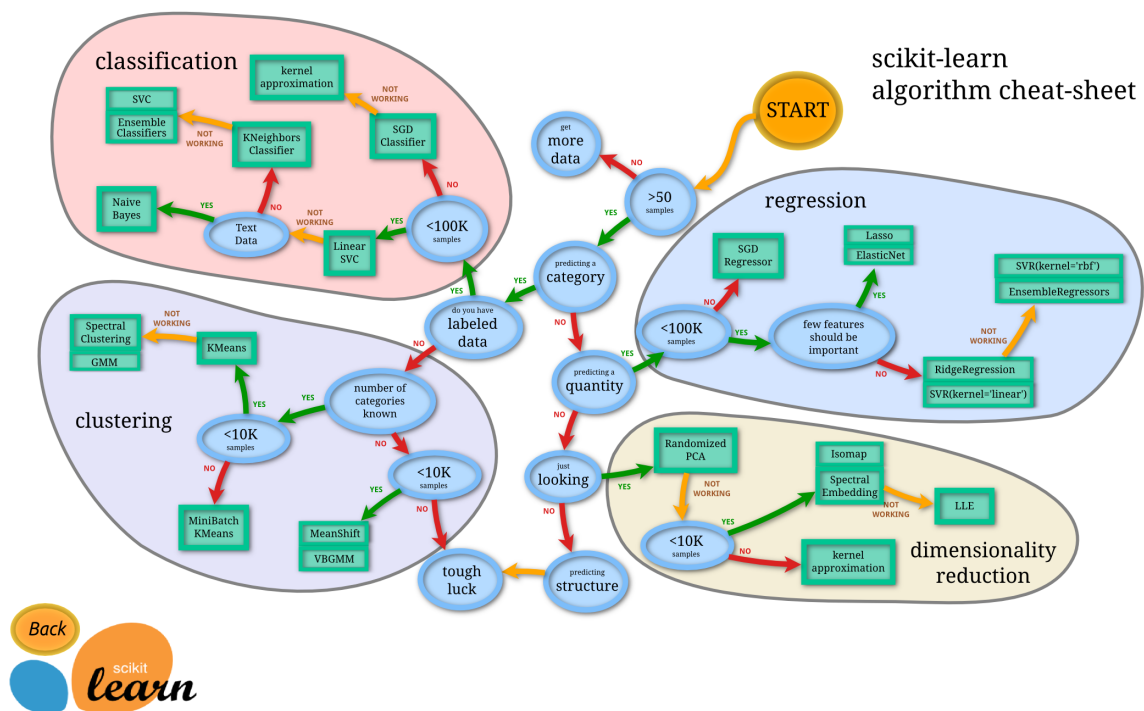


Figure 2.1 – Comment choisir un algorithme ML [Sci]

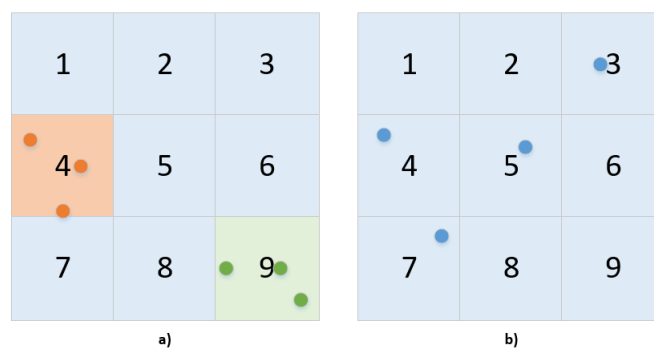


Figure 2.2 – a) exemple de classification b) exemple de regression

## Decision Tree

L'arbre de décisions est une méthode très connue en "machine learning". Il possède des noeuds de décisions (non-terminal), des branches, et des noeuds feuilles (terminal) qui représentent les caractéristiques, condition et les classes. A chaque noeud de décision on sait quelle branches suivre et lorsque l'algorithme atteint un noeud final, le label contenu dans ce même noeud est retourné comme étant la classe. L'ID3 de Quinlan et son successeur, C4.5, sont les plus populaires parmi les algorithmes d'arbre de décision [JR14].

## Naïve Bayes

Le classificateur Naïve Bayes [GHP95] basé sur le théorème de Bayes est un algorithme d'apprentissage supervisé [CS13]. Il est robuste aux données bruyantes, facile à construire, affiche une grande précision et rapidité lorsqu'il est appliqué à de grandes bases de données et exécute des modèles de classification plus complexes. Par conséquent, il est largement utilisé dans les tâches de classification. Il calcule la probabilité de chaque attribut dans les données en supposant qu'elles sont également importantes et indépendantes les unes des autres. Cette hypothèse est appelée indépendance conditionnelle de classe [WA13] [SS14].

## Bayesian Network

L'algorithme de réseau bayésien est largement utilisé pour la classification et est basé sur le théorème de Bayes où la probabilité conditionnelle sur chaque nœud est calculée et forme un réseau bayésien. Il s'appelle également réseau de croyance ou réseau occasionnel. Réseau bayésien a deux parties nommées qualitatives et quantitatives, qui sont la structure topologique du réseau bayésien et le tableau de probabilité conditionnelle (CPT), respectivement [DL].

Le réseau bayésien est un graphe acyclique dirigé où chaque nœud représente un attribut des données et un ensemble de distributions de probabilité. Ces distributions donnent les probabilités pour la valeur de chaque nœud étant donné que les parents de nœud.

## K-Nearest Neighbor

Le classificateur K-Nearest Neighbor (K-NN) [DWDMMK] est également connu sous le nom de classificateur basé sur la distance qui classe les instances en fonction de leur similarité. C'est l'un des algorithmes les plus populaires de l'apprentissage automatique. C'est un type d'apprentissage paresseux dans lequel la fonction n'est approchée que localement et tout calcul est retardé jusqu'à la classification. Le tuple inconnu dans K-NN est assigné à la classe la plus commune parmi ses K-plus proches voisins. Lorsque  $K = 1$ , le tuple inconnu se voit attribuer la classe du tuple d'apprentissage le plus proche dans l'espace des motifs [CAG].

## SMO

L'algorithme d'optimisation séquentielle minimale (SMO - Sequential minimal optimization) [J.] est représenté par John C. Platt pour la formation du classificateur de vecteurs de support à l'aide des noyaux polynomiaux ou RBF. C'est l'un des algorithmes les plus courants pour la classification des grandes marges par SVM. Il remplace globalement toutes les valeurs manquantes et transforme les attributs nominaux en attributs binaires. La SVM est une technique de classification basée sur la technologie des réseaux neuronaux utilisant la théorie de l'apprentissage statistique [PH]. Il recherche un hyperplan optimal linéaire afin de maximiser la marge de séparation entre la classe positive et la classe négative. En pratique, la plupart des données ne sont pas linéairement séparables; ainsi, pour rendre la séparation possible, la transformation est effectuée à l'aide d'une fonction du noyau. L'entrée est transformée en un espace caractéristique de dimension supérieure à l'aide d'une cartographie non linéaire [30]. Une décision sur la fonction du Kernel est nécessaire pour implémenter SVM. Le Kernel définit la classe de fonction [SMKN].

## AdaBoost

AdaBoost (Adaptive Boosting) [YRE] est un algorithme d'apprentissage d'ensemble. Généralement, il peut être utilisé avec des algorithmes de Machine learning faibles pour améliorer leurs performances. Il est simple à mettre en œuvre, rapide et moins susceptible d'avoir un overfitting. Il améliore les algorithmes de classification instables tels que J48, DecisionStump, etc. L'idée derrière cet algorithme est d'obtenir un classificateur très précis en combinant de nombreux classificateurs faibles. Il fonctionne en exécutant de manière répétée un algorithme d'apprentissage faible donné sur diverses distributions sur les données d'apprentissage, puis en combinant les classificateurs produits par l'apprenant faible en un classificateur composite unique [RRE]. Les classificateurs de l'ensemble sont ajoutés un par un, de sorte que chaque classificateur suivant est entraîné sur des données difficiles pour les membres précédents de l'ensemble. Les poids sont définis sur les instances du jeu de données, en suivant une règle selon laquelle les instances difficiles à classer prennent plus de poids. Cette règle conduit les classificateurs ultérieurs à se concentrer sur eux [SOLISP].

## Bagging

le Bagging [35] crée des sacs de données de la même taille que le jeu de données d'origine en appliquant une sélection aléatoire à différents sous-ensembles des données d'apprentissage avec de nombreux exemples qui apparaissent plusieurs fois. Ce processus est appelé réplique bootstrap des données d'entraînement. L'idée derrière cette technique est de construire différents classificateurs en utilisant ces sous-ensembles. Chaque sous-ensemble est utilisé pour entraîner un classificateur individuel. Cette approche d'ensemble utilise le nombre de classificateurs a priori [L.].

## support vector machine(SVM) pour la régression

Support vector machines (SVMs) est une méthode d'apprentissage supervisée utilisée pour la classification. Cette méthode peut être étendue pour résoudre des problèmes de régression et cette méthode s'appelle Support Vector Regression (SVR)

## Gaussian process (GP)

Gaussian processes have recently gained interest from the machine learning community, as they form an elegant framework to perform regression [CC].

### 2.1.3 Outliers [FP17]

Cette section donne un aperçu de la manière de traiter les "outliers - valeurs aberrantes", c'est-à-dire les points qui ne sont pas cohérents lors d'une mesure. Selon Barnett et Lewis [VT94], un "outliers" est défini comme étant une observation qui semble incompatible avec le reste d'un ensemble de données. Garder un "outliers" dans un set de données peut amener à de mauvais résultats, il est donc important de les détecter correctement. Il existe différentes méthodes pour déterminer ces "outliers" :

1. Grubbs' test : Détecte un "outliers" en supposant une distribution normale.
2. Tietjen-Moore test : C'est une généralisation de Grubbs' test pour détecter de multiples outliers. Il a cependant un inconvénient, il est nécessaire de connaître le nombre exact d'outliers.
3. Generalized Extreme Studentized Deviate (ESD) : C'est également une généralisation du test Grubbs' mais il n'est pas nécessaire de connaître à l'avance le nombre d'outliers. Ce test nécessite uniquement une limite supérieure pour le nombre suspect d'outliers.[Esd]

## 2.2 Comparaison

## 2.3 Choix



# Bibliographie

- [LLI] Mainetti LUCA, Patrono LUIGI et Sergi ILARIA. *A Survey on Indoor Positioning Systems*. Rapp. tech. University of Salento Lecce, ITALY (cf. p. 1).
- [FP17] Bernat Carbones FARGAS et Martin Nordal PETERSEN. « GPS-free Geolocation using LoRa in Low-Power WANs ». In : *Proceedings of 2017 Global Internet of Things Summit (GloTS)* (2017) (cf. p. 1, 3, 6).
- [Sin+15] Bozkurt SINEM et al. *A Comparative Study on Machine Learning algorithms for Indoor Positioning*. Rapp. tech. This work is supported by The Scientific et Technological Research Council of Turkey (TUBITAK) under grant number 1130024, 2015 (cf. p. 3).
- [Hen+12] Wymeersch HENK et al. *A Machine Learning Approach to Ranging Error Mitigation for UWB Localization*. Rapp. tech. Belgian American Education Foundation, the Charles Stark Draper Laboratory Robust Distributed Sensor Networks Program, the Office of Naval Research Young Investigator Award N00014-03-1-0489, the National Science Foundation under Grants ANI-0335256 and ECCS-0636519., 2012 (cf. p. 3).
- [KN17] KHUONG et NGUYEN. « A performance guaranteed indoor positioning system using conformal prediction and the WiFi signal strength ». In : *Journal of Information and Telecommunication*, 1 :1, 41-65, DOI : 10.1080/24751839.2017.1295659 (2017). URL : <https://doi.org/10.1080/24751839.2017.1295659> (cf. p. 3).
- [Sci] « Machine Learning in Python ». In : (). URL : [http://scikit-learn.org/stable/tutorial/machine\\_learning\\_map/index.html](http://scikit-learn.org/stable/tutorial/machine_learning_map/index.html) (cf. p. 4).
- [JR14] Quinlan J. R. « C4. 5 : programs for machine learning ». In : *Elsevier* (2014) (cf. p. 5).
- [GHP95] John G. H. et Langley P. « Estimating Continuous Distributions in Bayesian Classifiers ». In : *11th Conference on Uncertainty in Artificial Intelligence*, pp., 338-345 (1995) (cf. p. 5).
- [CS13] Anuradha (23) C. et Dhall S. « Software Defect Prediction Using Supervised Learning Algorithm and Unsupervised Learning Algorithm ». In : (2013) (cf. p. 5).
- [WA13] Yotsawat W. et Srivihok A. « Inbound tourists segmentation with combined algorithms using K-Means and Decision Tree ». In : *10th International Joint Conference on Computer Science and Software Engineering (IJCSE)*, pp.189-194 (2013) (cf. p. 5).
- [SS14] Ureerat S. et P. SINGSRI. « The classifier model for prediction quail gender after birth based on external factors of quail egg ». In : *IEEE 11th International Joint Conference on Computer Science and Software Engineering (IJCSE)* (2014) (cf. p. 5).
- [DL] Yang D. et Jin-lin L. « Research on personal credit evaluation model based on bayesian network and association rules ». In : *International Conference on Wireless Communications, Networking and Mobile Computing* () (cf. p. 5).
- [DWDMMK] Aha D. W., Kibler D. et Albert M. K. « Instance-based learning algorithms ». In : *Machine Learning*, vol. 6, pp., 37-66 () (cf. p. 5).
- [CAG] Shah C. et Jivani A. G. « Comparison of data mining classification algorithms for breast cancer prediction ». In : *Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)* () (cf. p. 5).
- [J.] Platt J. « Fast Training of Support Vector Machines using Sequential Minimal Optimization ». In : *Advances in Kernel Methods - Support Vector Learning* () (cf. p. 5).
- [PH] Niken P. et Ohwada H. « Applicability of machine-learning techniques in predicting customer defection ». In : *IEEE 2014 International Symposium on Technology Management and Emerging Technologies (ISTMET)* () (cf. p. 5).

- [SMKN] Obaidullah S. M., Roy K. et Das N. « Comparison of different classifiers for script identification from handwritten document ». In : *IEEE International Conference on Signal Processing, Computing and Control (ISPCC)*, pp.1-6 () (cf. p. 5).
- [YRE] Freund Y. et Schapire R. E. « Experiments with a new boosting algorithm ». In : *3th International Conference on Machine Learning, San Francisco*, pp. 148-156 () (cf. p. 6).
- [RRE] Shams R. et Mercer R. E. « Classifying Spam Emails Using Text and Readability Features ». In : *IEEE 13th International Conference on Data Mining (ICDM)*, pp. 657-666 () (cf. p. 6).
- [SOLISP] Sharif S. O., Kuncheva L. I. et Mansoor S. P. « Classifying encryption algorithms using pattern recognition techniques ». In : *IEEE International Conference on Information Theory and Information Security (ICITIS)*, , pp. 1168-1172 () (cf. p. 6).
- [L.] Breiman L. « Bagging predictors ». In : *Machine Learning*. vol. 24, no. 2, pp.123-140 () (cf. p. 6).
- [CC] Rasmussen C. et Williams C. « Gaussian Processes for Machine Learning ». In : *Springer* () (cf. p. 6).
- [VT94] Barnett V. et Lewis T. « Outliers in Statistical Data ». In : *3rd ed. Wiley Series in Probability and Mathematical Statistics* (1994). URL : <https://doi.org/10.1080/24751839.2017.1295659> (cf. p. 6).
- [Esd] « lien concernant les ESD ». In : (). URL : <https://www.itl.nist.gov/div898/handbook/eda/section3/eda35h3.htm> (cf. p. 6).