MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE ABDELHAMID IBN BADIS - MOSTAGANEM

Faculté des Sciences Exactes et d'Informatique

Département de Mathématiques et informatique

Filière : Informatique

MEMOIRE DE FIN D'ETUDES

Pour l'Obtention du Diplôme de Master en Informatique

Option : **Ingénierie des Systèmes d'Information**

Présenté par :

Daou Mama Bino

Bouali Youcef Zakaria

THEME :

Estimation des données manquantes dans les réseaux de capteurs sans fil

Soutenu le : 19/06/2021

Devant le jury composé de :

BENAMEUR AEK	Grade	Université de Mostaganem	Président
ABID M.	Grade	Université de Mostaganem	Examinateur
HABIB ZAHMANI	Grade	Université de Mostaganem	Encadreur
SANAA K. GHALEM	Grade	Université de Mostaganem	Co-Encadrante

Année Universitaire 2020-2021

DÉDICACES

Je dédie ce travail :

À mes chers parents, qui m'ont doté d'une éducation digne, qui ont conjugué efforts et sacrifices sans cesse pour mon instruction et mon bien-être. Aucune dédicace ne serait à la hauteur pour exprimer ce que vous méritez.

À mes frères : Yacouba, Amadou, Abib, Mamadou, Mohamed.

À mes Sœurs : Awa et Fatoumata.

À mes amis.

Les mots ne suffissent guerre pour exprimer l'attachement, l'amour, le respect que j'ai pour vous. Puisse Dieu vous combler de santé, de bonheur et d'une longue vie.

Daou Mama Bino

DÉDICACES

Je dédie ce travail :

À mes chers parents, qui m'ont doté d'une éducation digne, qui ont conjugué efforts et sacrifices sans cesse pour mon instruction et mon bien-être. Aucune dédicace ne serait à la hauteur pour exprimer ce que vous méritez.

À mes frères : Mohamed, Youcef, Abde-Elkader, Ahmed, Miloud.

À mes Sœurs, À ma famille.

À mes amis.

Les mots ne suffissent guerre pour exprimer l'attachement, l'amour, le respect que j'ai pour vous. Puisse Dieu vous combler de santé, de bonheur et d'une longue vie.

Zakaria Bouali Youcef

REMERCIEMENTS

Nous remercions Dieu pour la santé, la volonté, le courage, la détermination qui nous ont permis de réaliser ce travail modeste.

Nous ne saurions trouver les termes qu'ils faut pour exprimer nos profondes reconnaissances envers ***Madame Ghalem Sanaa*** *et* ***Monsieur Habib Zahmani*** *pour la qualité de l'encadrement dont ils nous ont fait bénéficier, pour avoir bien guidé et bien structuré ce travail en conjuguant disponibilité, orientations, conseils et critiques constructives. Nous vous remercions.*

Un grand merci à nos familles qui nous apportaient tous les jours de l'équilibre à travers des soutiens et des encouragements.

Nous remercions profondément Mr Benameur AEK d'avoir accepter de présider notre soutenance.

Nous adressons nos reconnaissances à Mme Abid Meriem d'avoir accepter d'examiner et de valoriser notre travail.

Nous remercions également tous ceux dont les discussions, les remarques et les suggestions ont conduit à l'amélioration de ce travail.

LISTE DES FIGURES

LISTE DES TABLEAUX

LISTE DES ABRÉVIATIONS

ADC	Analog Digital Converter
AR	Artificial Retina
ART	Adaptive Resonnance Theorie
BRL	Block Random Lost
CARM	Closed Itemsets based Association Rule Mining
CH	Cluster Head
DMLA	Dégénérescence Maculaire Liée à L'âge
EAR	Energy and Activity Aware Routing
EFRL	Element Frequent Loss in a Row
ERL	Element Random Lost
FARM	Freshness Association Rule Mining
HUM	Humidité
IEEE	Institute Of Electrical and Electronics Engineering
KNN	K Nearest Neighbor
LEACH	Low Energy Adaptive Clustering Hierarchy
LOCF	Last Observation Carried Forward
MAC	Medium Access Control
MAE	Mean Absolute Percentage Error
MANET	Mobile Ad hoc Network
MAR	Missing At Random
MCAR	Missing Complety At Random

MEMS	Micro Electro Mechanical Systems
ML	Machine Learning
MLP	Perceptrons Multi Couche
MNAR	Missing Not At Random
MVLM	Multi-View Learning Method
NAWMS	Non Intrusive Autonomous Water Monitoring System
NFC	Near Field Communication
PCP	The percentage of correct predictions
QDS	Qualité de Service
RCSF	Réseaux de capteurs sans-fil
RMSE	Root Mean Square Error
RP	Rétinite Pigmentosa
SAR	Sequentiel Assignement Routing
SE	Système d'exploitation
SELR	Successive Elements Loss in a Row
SELR	Combination Loss
SMAC	Sensor MAC
SMP	Sensor Management Protocol
SVM	Support Vector Machine
TADAP	Task Assignement and Data Advertissement Protocol
TKCM	Top-k Case Matching
TMP	Température
UDP	User Datagram Protocol
WARM	Window Association Rule Mining
WSN	Wireless Sensor Network
XMAC	Expected Mac

Résumé

Avec le développement de l'électronique, de la communication sans fil, le souci d'observer et éventuellement de contrôler certains phénomènes physiques (température, la pression, la luminosité) est devenu possible et plus que facile grâce à la naissance d'une nouvelle technologie du nom de WSN. Cette technologie n'a cessée de croitre depuis son apparition grâce à ses différents domaines d'applications. Elle consiste à un ensemble de micro-capteurs, capable de communiquer entre eux, déployés dans une zone d'intérêt. Ces micro-capteurs perçoivent leur environnement et récoltent des données, et les transmette à un point de récolte appelé Sink. Le Sink à son tour transmet ces données à un centre pour des fins d'exploitation. Cette technologie est utiliser dans différents domaines comme le domaine militaire, médical, d'agriculture, de surveillance, etc.

Avec leurs perspectives d'utilisation facile et attrayante, les RCSFs ne sont pas parfaits. Ils sont soumis à plusieurs contraintes qui peuvent entraver leur bon fonctionnements comme les ressources matérielles limitées , une bande passante faible, des capacités de capture et de communication réduites, les conditions d'environnement et de déploiement sévères. Ces limites sont la cause de différent problèmes dans cette technologie comme l'absence de données dans l'ensemble des données collectées. Dans certaines application prendre des décisions directement avec des telles données peut conduire a des erreurs. De ce fait, les données récoltées par les RCSFs doivent subir différents types d'analyses avant leurs exploitations.

Pour remédier à ce problème, le travail suivant porte sur *l'estimation des données manquantes* présentent dans les données collectées par les RCSFs. Pour se faire, après analyse des différentes méthodes d'estimations existantes, nous allons énumérer les problèmes auxquels ces méthodes sont confrontées avec l'objectif de construire une méthode d'estimation plus efficace répondant aux critères des RCSFs. Nous évaluerons aussi les performances de la méthode en faisant une étude comparative entre elle et certaines méthodes d'estimations existante sur un jeu de données des RCSFs, en générant des données manquantes suivant les différents modèle de pertes des RCSFs.

Mots-clés : RCSF, données manquantes, méthodes d'estimation, performance, modèle de perte, estimation.

Abstract

The development of electronics, wireless communication, the concern to observe and eventually control certain physical phenomena (temperature, pressure, brightness) has become possible and more than easy thanks to the birth of a new technology called WSN.

This wireless technology has continued to grow since its inception thanks to these different areas of application. It consists of a set of micro-sensors called sensor nodes, capable of communicating with each other, deployed in an area of interest in a predetermined or random manner to achieve an objective. These sensor nodes are able to perceive their environment deployment and collect data that will be sent autonomously to a collection point called Sink, then transmitted by the latter to the control center. Ones the collected data received at the database it can be used to analyze the environment of deployment and make decisions. It has been used in different fields such as : military, medical, agricultural, surveillance, etc.

WSN show flows despite its appealing advantages. They are subject to several constraints that can hinder their proper functioning such as limited material resources (energy, storage capacity, processing), low bandwidth, reduced capture and communication capacities, environmental conditions, deployment, etc. Those limitations can increase the missing data rate in WSN (data not received), leading to a misrepresentation of the environment of deployment. Thus the data collected by the WSN must undergo different types of analyzes before their exploitation.

To answer this problem, we are interested at *the estimation of missing data in the WSN*. To do so, after analyzing the different existing estimation methods, we then enumerate the various problems with which these methods are confronted with the objective of constructing a more efficient estimation method which meets the criteria of the WSN. We will also evaluate the performance of our method by making a comparative study between certain estimation methods.

Keywords : Wireless sensor networks, missing data, estimation methods, performance, missing patterns.

TABLE DES MATIÈRES

INTRODUCTION GÉNÉRALE

Suite aux progrès récents de la technologie des Systèmes Micro-Electro-Mécaniques (MEMS), de la communication sans-fil et l'électronique numérique, la conception et le développement de systèmes multi-fonctionnels à faible coût des nœuds de capteurs de petite taille communiquant via des liaisons sans-fil a de courtes distances sont devenues réalisables [1]. Ces minuscules nœuds de capteurs sont capables d'interagir avec leur environnement et détecter des phénomènes, de traiter des paramètres physiques de façon autonome. Ces facteurs ont donné naissance à une nouvelle classe de réseau nommée les réseaux de capteurs sans-fil ou Wireless Sensors Network en Anglais.

Un RCSF est un réseau MANET basé sur l'effort collaboratif d'un grand nombre de nœuds capteurs, appelés «motes». Ces dispositifs sont moins couteux, et possèdent une faible consommation d'énergie. Ils coopèrent ensemble pour extraire différents types de données (scalaire et multimédia) de l'environnement étudié et l'envoient à un point central appelé station de base ou Sink. Le Sink peut à son tour utiliser ses informations localement ou les transmettre à travers une connexion Internet ou Satellite à un utilisateur final pour des fins d'exploitations. Ce type de réseau a été utiliser dans différents domaines comme le domaine militaire, le domaine médical, le domaine de surveillance, etc.

Les nœuds de capteurs de ce type de réseau sont équipés de ressources matériels comme un ou plusieurs capteurs, d'un micro-processeur, d'une source d'énergie, d'une antenne. Leurs dispositions dans une zone de surveillance peuvent être de façon aléatoire ou déterministe suivant son domaine d'application. Mais malgré les grandes avancées de la technologie de miniaturisation, ces capteurs sont dotés de ressources limitées. De ce fait les données collectées par ce type de réseau contiennent fréquemment de valeurs de mesures manquantes dues à de nombreux facteurs comme des pannes matérielles, capacité de communication des nœuds de capteurs limitée, conditions environnementales sévères, etc. Alors il serait dommage voir impardonnable de prendre des décisions sur ces ensembles de données incomplètes. Cependant la mise en place d'approche de restauration de ces valeurs manquantes prend tout son importance.

Ces approches doivent restaurer ces valeurs manquantes en minimisant le plus possible

la différence entre les valeurs manquées et les valeurs estimées. Notre objectif consiste à mettre en place des approches répondant à ces critères. Pour atteindre cet objectif, nous avons divisé notre travail en plusieurs parties. Une première consacrée à une étude générale sur les RCSFs, une seconde faisant l'objet d'étude d'estimation des données manquantes dans les RCSFs, une troisième consacrée à la proposition d'approches d'estimation et à une analyse du jeu de données choisi pour l'évaluation de ces approches, et une dernière faisant l'objet de discussions des résultats obtenus.

CHAPITRE 1

ÉTAT DE L'ART DES RÉSEAUX DE CAPTEURS SANS-FIL

1.1 Introduction

Les réseaux de capteurs sans-fil représentent une nouvelle génération de réseau, l'une des technologies la plus importante du XXIe siècle, un domaine de recherche en constante évolution qui a attiré l'attention de nombreux chercheurs grâce aux succès grandioses qu'a connu leur application dans les milieux académiques, industriels, sociaux, constitutionnels, etc. Ils ont vu le jour grâce aux progrès rapides de la technologie des Systèmes Micro-Electro-Mécanique [2]. Typiquement ils consistent à un grand nombre de nœuds capteurs déployés dans un environnement, qui collaborent, et mesurent de grande quantité de données, les stocke et les transmette à d'autres nœuds ou à une station de base appelée Sink. Ils sont capable de s'auto-configurer, de se gérer de façon autonome et dispose pour cela d'une réserve énergétique. Dans ce présent chapitre, nous allons présenter brièvement cette technologie, quelque domaines d'applications d'elle, ainsi que certains défis auxquels elle est confrontée.

1.2 Les réseaux de capteurs sans-fil

Un RCSF est composé d'un grand nombre de petites entités appelées nœuds capteurs ou « mote » en Anglais, déployés dans ou près d'une zone d'intérêt (champ de déploiement), organisés en champs (sensors fields). Ces nœuds communiquent entre elles via une connexion

sans-fil, dans le but d'accomplir une tâche commune. Chaque nœud déployé dans la zone d'intérêt surveille l'environnement afin de mesurer d'éventuel évènement sur cet environnement et capturer des mesures. Ces dernières peuvent être stockées, traitées ou transmises à un nœud passerelle ou puits appelé «Sink» en utilisant un mécanisme de routage qui est le plus souvent un routage multi-sauts [1]. Le Sink à son tour pourra les utiliser ou les transmettre à l'utilisateur final par le biais d'Internet ou par satellite pour des fins d'exploitation. Les nœuds de capteur peuvent jouer à la fois le rôle de capteur de données ou le rôle de routeur (relais) et leur mode de déploiement dépend du domaine d'application. Il peut être complètement aléatoire (largués par un engin) pour certains domaines et être précis (placés manuellement) pour d'autres domaines. La figure 1.1 représente le fonctionnement d'un RCSF.

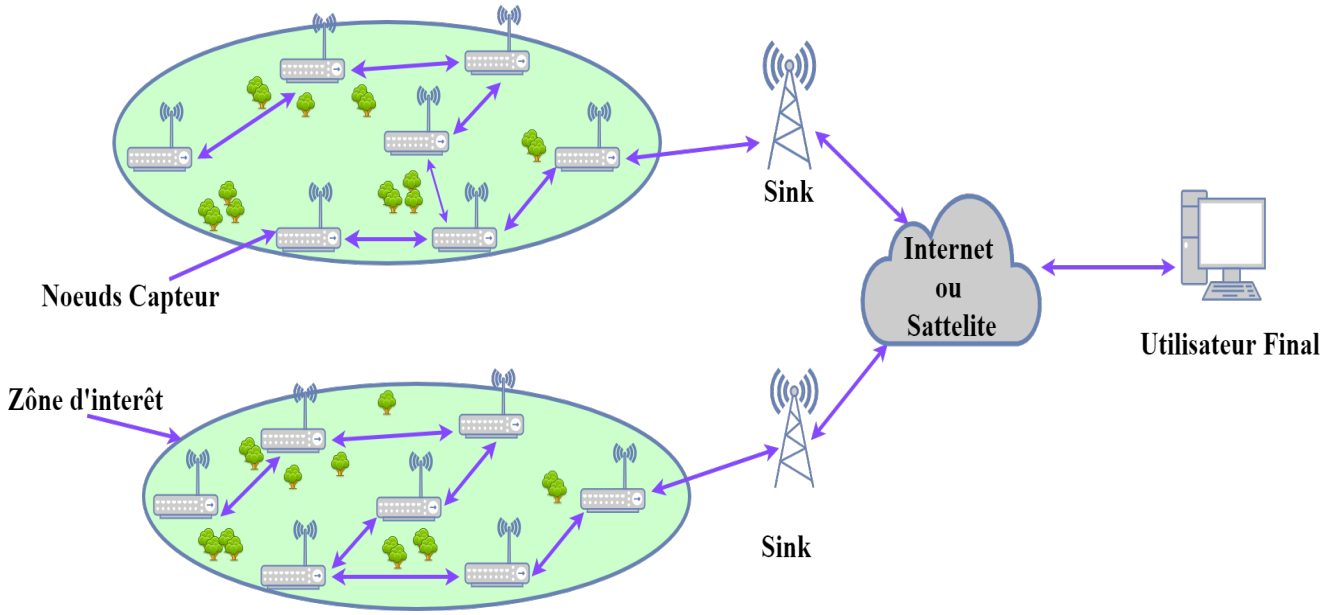


FIGURE 1.1 – Fonctionnement d'un RCSF

1.2.1 Description des composants d'un RCSF

Un RCSF est composé principalement d'un ensemble de capteurs, d'un ou plusieurs nœuds Sink, d'un ou plusieurs gestionnaires de tâches (utilisateurs), et un composant facultatif appelé agrégateur [40].

- **Les nœuds capteurs (sensors) :** Ils sont chargés de capter les données au sein de l'environnement de déploiement et les acheminer vers le Sink. Ils sont généralement immobiles, mais d'autres nœuds sont dotés d'un système de mobilité garantissant ainsi une grande couverture de la zone de déploiement.
- **La passerelle (Sink) :** Est un nœud particulier qui reçoit l'ensemble des données mesu-

rées par les autres nœuds du réseau et les achemine vers l'ordinateur central par Internet ou par Satellite. Il assure aussi la liaison entre le réseau et l'ordinateur central ainsi sa défaillance provoque la perte du réseau tout entier.

- **Utilisateur (Ordinateur central) :** Il est le dernier niveau d'un RCSF. Il reçoit les données transmises par le Sink, et peut aussi communiquer (faire une requête) directement sur les nœuds capteurs en utilisant le Sink comme passerelle afin de collecter des données dont il a besoin.
- **L'agrégateur (aggregator) :** C'est un composant facultatif dans un RCSF. Il fait un amas des messages qu'il reçoit des nœuds de capteurs puis les transmette en un seul message au Sink. Son objectif est de minimiser le trafic dans le réseau.

1.3 Architecture des réseaux de capteurs sans-fil

On distingue deux types d'architectures pour les RCSFs : l'architecture à plat et l'architecture hiérarchique [41].

1.3.1 Architecture à plat

Un RCSF fonctionnant sur une architecture à plat est un réseau homogène dont les nœuds sont identiques en termes de complexité du matériel et possèdent le même rôle dans l'exécution des tâches, mis à part le Sink qui joue le rôle de passerelle entre l'utilisateur final et les autres nœuds du réseau. Dans cette architecture, les nœuds de capteurs peuvent communiquer directement avec le centre de traitement en utilisant une forte puissance d'émission, ou par l'intermédiaire d'un mode de communication multi-sauts en utilisant une puissance d'émission beaucoup plus faible [41]. Dans le cas de la communication directe, la consommation énergétique du nœud pour l'envoi des données au centre de traitement est plus importante. Par conséquent ces nœuds peuvent rapidement épuiser leur énergie. Dans le cas de la communication multi-sauts qui est le plus fréquent, un nœud capteur qui veut transmettre ses données à la station de base passe par d'autres nœuds du réseau (routeur). Ce type d'architecture présente différent avantages comme une conservation d'énergie des nœuds du réseau par conséquent la durée de vie du réseau, la tolérance aux pannes. Mais comporte aussi des inconvénients comme les problèmes de routage, le temps de latence élevé et les problèmes de sécurité [41]. La figure 1.2 illustre le fonctionnement d'un RCSF suivant l'architecture à plat.

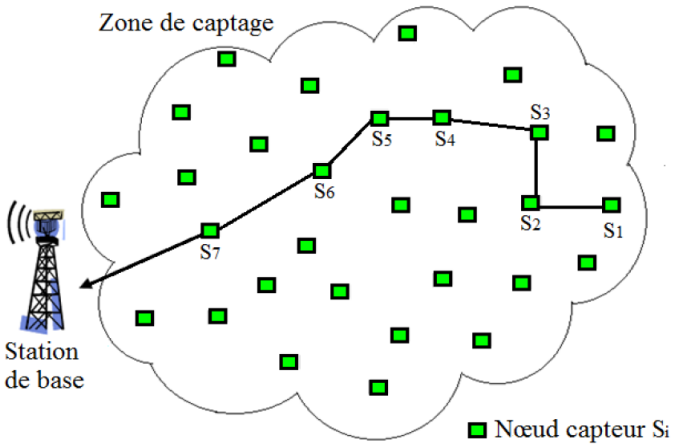


FIGURE 1.2 – Fonctionnement d'un RCSF suivant l'architecture à plat [41]

1.3.2 Architecture hiérarchique

Dans l'architecture hiérarchique, le réseau est partitionné en plusieurs niveaux de responsabilité. L'une des méthodes la plus utilisée est le clustering dans lequel les nœuds capteurs du réseau sont partitionnés en groupes ou chaque nœud est soit chef du groupe (CH) ou membre du groupe [41]. Dans cette architecture, les nœuds membres ne peuvent pas communiquer directement avec le Sink mais doivent passer par leur CH qui à son tour communique avec le Sink. L'inter-action entre les nœuds membre d'un cluster est aussi gérée par le CH. Les CHs ont aussi la capacité d'agréger les données des clusters membres et les acheminer en un seul message vers le Sink. Cette opération permet de réduire le trafic dans le réseau et de faciliter la tâche du CH. Les CHs peuvent aussi communiquer ensemble pour acheminer les données au Sink. Ce type d'architecture a pour avantages une faible consommation d'énergie, la scalabilité, la réduction de collision, mais possède aussi des inconvénients comme la difficulté de sélection et maintenance du CH [1]. La figure 1.3 illustre le fonctionnement d'un RCSF suivant l'architecture hiérarchique.

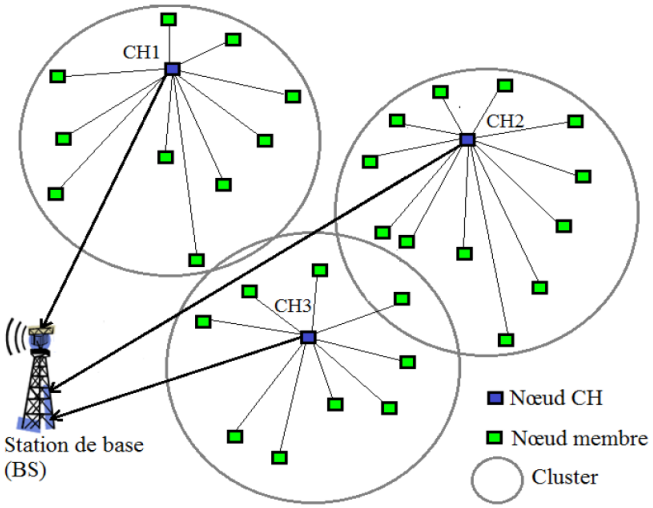


FIGURE 1.3 – Fonctionnement d'un RCSF suivant l'architecture hiérarchique [41]

1.4 Composants des nœuds capteurs

Les nœuds capteurs sont les éléments de base formant un RCSF. Ils possèdent des composants matériels et logiciels.

1.4.1 Composants matériels

Un nœud capteur est composé principalement de 4 unités de base : l'unité d'acquisition, l'unité de traitement, l'unité de communication et l'unité d'énergie [1, 40]. Ces unités sont illustrées dans la figure suivante.

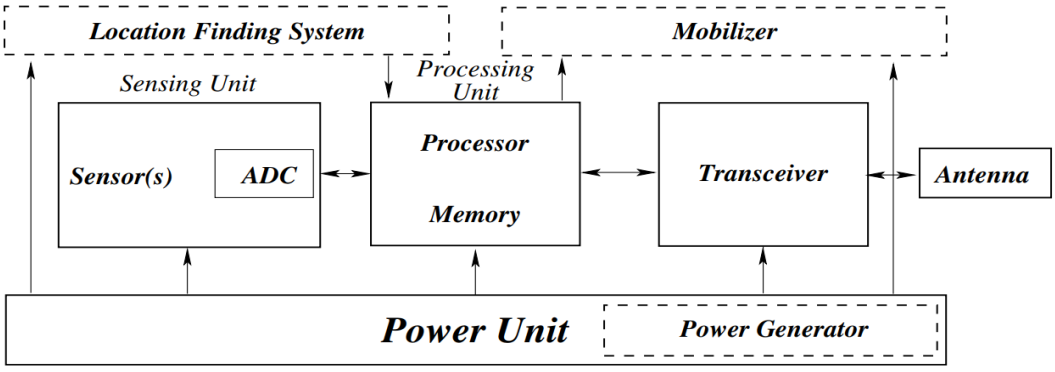


FIGURE 1.4 – Composants matériels d'un nœud capteur [2]

- **L'unité d'acquisition (Sensing unit) :** Elle est le composant principal d'un nœud capteur. Elle est composée de deux sous-unités, le capteur et le convertisseur analogique/numérique ADC. Le capteur fournit des mesures de l'environnement transformées en signaux ana-

logiques à l'ADC, ce dernier les transforme en des signaux numériques compréhensibles par l'unité de traitement.

- **L'unité de traitement (Processing unit) :** Elle est le contrôleur principal du nœud de capteur, et à travers elle tous les composants du nœud de capteur sont gérés [1]. Elle est composée de deux interfaces, une avec l'unité d'acquisition et une autre avec l'unité de communication. Elle fonctionne sur un système d'exploitation conçu uniquement pour les micros capteurs (Tiny OS). Elle possède une petite mémoire pouvant stocker les données collectées par le nœud de capteur et peut éventuellement les analyser.
- **L'unité de communication (Transceiver) :** Est un module radio (émetteur et récepteur) sans-fil, équipé d'une antenne assurant la communication entre les nœuds du RCSF.
- **L'unité d'énergie (Power unit) :** Elle fournit l'énergie nécessaire au bon fonctionnement du dispositif. A cause de la petite taille des nœuds, cette unité n'est souvent pas rechargeable et constitue la principale contrainte des RCSFs. Son utilisation nécessite un fonctionnement économe du nœud de capteur.
- **Unités supplémentaires :** D'autres capteurs peuvent contenir des unités supplémentaires comme l'unité de localisation permettant de connaitre la position du nœud capteur, une unité de mobilité permettant au nœud capteur de se déplacer et de garantir une meilleure couverture, et une unité de régénération d'énergie permettant de recharger la batterie du nœud de capteur quand celle-ci s'épuise.

1.4.2 Composants logiciels

Un nœud capteur est composé de deux types de logiciels : l'intergiciel (Middleware) et le Système d'exploitation (OS) [42].

- **Intergiciel :** C'est un ensemble de programme qui a pour objectif de supporter le développement, la maintenance, l'exécution des applications utilisant les RCSFs [42]. Il existe plusieurs intergiciels comme TinyDB [3], Cougar [18] et SensorWare [4].
- **Système d'exploitation :** Ils existent plusieurs SE pour les RCSFs, mais les plus répandus sont TinyOs [5] et Contiki [12].
  - **TinyOs :** Développé par les chercheurs de l'université Américaine de Berkeley est un SE intégré, modulaire, et open source conçu pour les systèmes embarqués, et contraignants en termes de ressources comme les nœuds d'un RCSF. Il est écrit en grande partie en nesC un dérivé du langage de programmation C.

— **Contiki :** Crée par une équipe de recherche du centre Suédois, c'est un système d'exploitation multi-tâches, open source, flexible, portable. Il supporte les protocoles IPv6 prenant en charge le mode de transmission TCP ou UDP. Il est destiné pour les petits capteurs disposant des ressources limitées.

Les systèmes d'exploitation des RCSFs ne se limitent pas à ceux susmentionnés, il existe d'autres comme : Mantis [6], LiteOs [7], SenSmart [8], SenSpire [9], SOS [16].

1.5 Les types de RCSF

En fonction de l'environnement où ils seront déployés, les RCSFs son confrontés à différents défis et contraintes. Avec le développement rapide de cette technologie au fil des ans, une solution de conception d'une grande diversité de types de RCSF faisant face à différents type de défis ont fait leurs apparition. Il existe : les RCSFs terrestres, les RCSFs souterrains, les RCSFs sous-marins, les RCSFs multimédias et les RCSFs mobiles [10].

1.5.1 Les RCSFs terrestres (Terrestrial WSNs)

Ce type de réseau se compose de centaines ou de milliers de nœuds de capteurs. Ces nœuds de capteur sont peu coûteux et sont déployés dans une zone d'une manière aléatoire ou suivant une stratégie (pré-planifié) [10]. Dans ce type de réseau, une communication fiable dans l'environnement est très importante et les nœuds de capteurs doivent être en mesure de capturer et communiquer efficacement les données collectées à la station de base.

La puissance de la batterie des nœuds de capteur étant limitée, ils peuvent être équipés d'une source d'énergie secondaire telle que les cellules solaires [10]. Ce type de réseau doit conserver l'énergie en utilisant un mode de communication optimal (routage) multi-sauts, une courte portée de transmission, une agrégation de données collectées, etc.

1.5.2 Les RCSFs souterrains (Underground WSNs)

Les RCSFs peuvent également être déployés dans le sous-sol. Ce type de RCSF est composé d'un ensemble de nœuds de capteurs enfouis sous terre, dans une mine, dans une grotte afin de surveiller leurs conditions [19], et est déployé de façon pré-planifiée. Pour acheminer les données collectées à la station de base, d'autres nœuds de capteurs supplémentaires sont placés

au-dessus du sol. Contrairement aux RCSFs terrestres, les nœuds d'un RCSF souterrain sont beaucoup plus chers car ils nécessitent des équipements assurant une communication fiable à travers le sol, l'eau, les roches et d'autres contenus minéraux. L'environnement souterrain rend la communication sans-fil un défit du fait des pertes de signal et du niveau élevé d'atténuation.

1.5.3 Les RCSFs sous-marins (Underwater WSNs)

Les RCSFs peuvent également être déployés dans l'eau. Ce type de RCSF consiste en un certain nombre de nœuds capteurs et de véhicules déployés dans un milieu aquatique dans le but d'effectuer des tâches collaboratives de surveillance et de collecte de données [20], utilisant des ondes acoustiques pour la communication. Pour atteindre ces objectifs, les nœuds de capteurs et les véhicules doivent être capables de s'auto-organiser, de s'auto configurer et s'adapter aux conditions sévères de l'environnement océanique. Les véhicules déployés sont autonomes et servent à la collecte ou à l'exploration des données à partir des nœuds de capteurs. À l'inverse d'un RCSF terrestre, les nœuds de ce type de réseau sont couteux, et sont aussi déployés en nombres limités. Ce type de réseau est contraint en termes d'énergie, de bande passante réduite, de long délai de transmission, d'affaiblissement de signal.

1.5.4 Les RCSFs multimédias(Multi-media WSNs)

Les RCSFs multimédias ont fait leur apparition au cours des dernières années [21]. Ils ont étés proposer pour permettre aux capteurs la surveillance et suivi des évènements sous forme de multimédia (image, vidéo, son). Les nœuds de ce type de réseau sont déployés de façon pré-planifié, et sont équipés de caméra et/ou de microphone. Ils peuvent acquérir, stocker, traiter et communiquer si besoin du contenu multimédia riche à partir de l'environnement de déploye-ment. Cependant atteindre ces objectifs n'est pas une tâche facile, car ils s'accompagnent de défis comme la nécessité d'une grande bande passante, une consommation d'énergie élevée, la garantie d'une qualité de service, la mise en place de techniques de traitement et de compression des données.

1.5.5 Les RCSFs mobiles (Mobile WSNs)

Ce type de RCSF consiste en un ensemble de nœuds de capteurs mobiles qui peuvent se déplacer et interagir avec l'environnement physique [10] dans lequel ils sont déployés. Le

déploiement peut commencer avec un certain nombre de nœuds et peut s'étendre par la suite en augmentant le nombre de nœuds. Ils peuvent transmettre les données capturées à d'autres nœuds s'ils se trouvent à porté l'un de l'autre. En plus de leur capacité à pouvoir se déplacer, ils peuvent se repositionner, calculer, détecter et communiquer comme les nœuds d'un RCSF statique. A l'opposé d'utilisation du routage fixe, les RCSFs mobiles utilisent un routage dynamique. Ils sont contraint en termes de déploiement, de localisation, d'énergie, de gestion de mobilité, de contrôle des nœuds mobiles, de maintient d'une couverture de détection, etc.

1.6 La pile protocolaire des RCSFs

Pour la gestion de la communication dans les RCSFs, l'approche qui gère celle des réseaux traditionnels est celle qui a été adoptée, celle de découper la communication en plusieurs niveaux ou couches qui sont la couche physique, la couche liaison de données, la couche réseau, la couche transport, la couche application [1]. En plus d'eux, différentes autres couches ont étés ajoutées et qui sont propres aux RCSFs : le plan de gestion de d'alimentation, le plan de gestion de la mobilité et le plan de gestion des tâches [1], comme indiqué sur la figure 1.5.

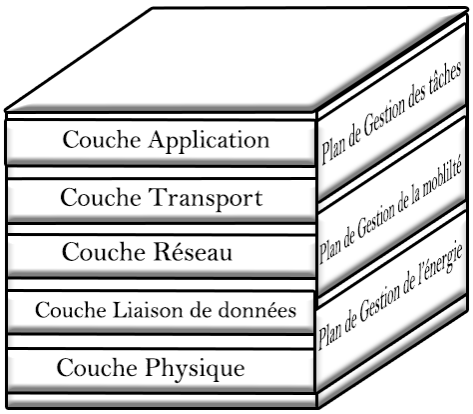


FIGURE 1.5 – Pile protocolaire des RCSFs

- **Couche application (Application Layer) :** Elle est la couche la plus proche de l'utilisateur, contenant l'application principale et les fonctionnalités de gestion assurant l'interaction du nœud avec les applications. Les fonctionnalités de traitement de requête, de gestion réseau résident aussi dans cette couche [1]. Elle utilise les protocoles SMP, TADAP.
- **Couche transport (Transport Layer) :** Elle est responsable du bon transport des données, de leur qualité, de leur fiabilité ainsi que la gestion d'éventuelles erreurs de transmission. Elle utilise le protocole UDP-LIKE qui fonctionne comme le protocole UDP.

- **Réseau (Network Layer) :** Elle permet d'établir des chemins entre les nœuds capteurs et le Sink pour l'acheminement des données issues de la couche adjacente (transport) et de sélectionner le meilleur en fonction de certaines contraintes comme l'énergie, cout du transport, etc. Parmi ses protocoles, nous citons : LEACH [24], et SAR[17].
- **Liaison de données (Data Link Layer) :** Elle est responsable de l'accès au support physique, du contrôle d'erreur, de la mise en place d'une communication point à point entre les nœuds. Il peut aussi détermine les liens de communication entre les nœuds [1], et utilise les protocoles SMAC [13], XMAC [14] ainsi que EAR [15].
- **Physique (Physical Layer) :** Elle est responsable de la sélection et de la régénération de fréquence, de la détection de signal, de la modulation, du cryptage ainsi que de l'acheminement et réception des données sur le support physique [1].
- **Le plan de gestion de l'alimentation :** Il gère la façon dont les nœuds de capteurs utilisent leur énergie (batterie).
- **Le plan de gestion de la mobilité :** Il détecte et enregistre le mouvement des nœuds capteurs, de sorte qu'une route de retour vers l'utilisateur est toujours maintenue, et les nœuds capteurs peuvent garder une trace de leurs voisins [1].
- **Le plan de gestion des tâches :** Il équilibre et planifie les tâches de détection attribuées à une région spécifique [1].

1.7 Domaines d'applications des RCSFs

Avec leurs émergence, leur caractéristiques, leurs disponibilités en différents gamme ont déclenchée des recherches sur de nombreux aspects d'eux, leurs applications ont longuement été discutées. Cette discussion a inspirée de nombreuses applications, certaines d'entre elles sont futuristes tandis qu'un grand nombre d'entre elles sont déjà en cours d'utilisation. Dans cette section quelques un de leur applications seront présenter.

1.7.1 Application militaire

Par leur caractéristiques d'auto-organisation, de déploiement rapide et simple, de tolérance aux fautes ainsi que de petite taille des nœuds de capteurs font de ce type de réseau une technologie assez importante dans le domaine militaire [40]. Ils peuvent être utiliser dans le commandement, la surveillance des forces amies ou ennemies, la reconnaissance des champs de

bataille, évaluation des dommages de combat, ainsi que dans la surveillance des équipements et des munitions [42, 40]. Ils peuvent être aussi utiliser pour la détection de produit chimiques, des attaques nucléaire ou biologiques.

Une de leur application existante est le système de détection de tireur d'élite Boomerang développé pour une détection précise de l'emplacement de tireur d'élite. Il a été utilisé par l'armée [1]. Le système utilise des capteurs acoustiques passifs distribués pour détecter les incidents entrants. L'audio détecté par les microphones est traité pour estimer la position exacte du tireur. Il peut être porté par un soldat ou par une voiture. La figure suivante illustre ce système.

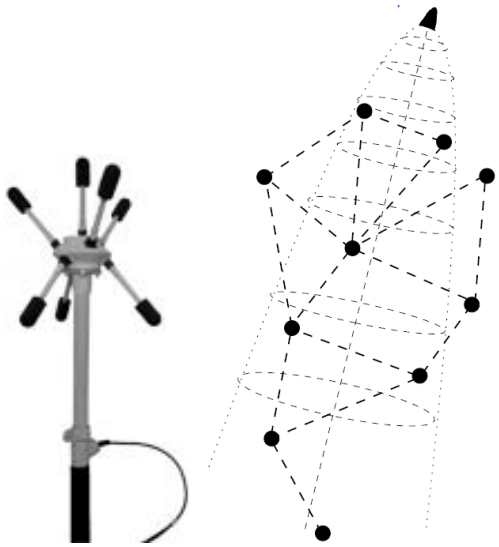


FIGURE 1.6 – Système de détection de Spniper distribué [1]

1.7.2 Application médicale

Ils consistent à des nœuds de capteurs implantés ou avalés par des patients permettant de relever des données comme la température corporelle, la glycémie, la pression artérielle, etc. Ainsi ces informations sont envoyées à une station de base se situant au domicile du patient ou à un médecin. On peut également les utiliser pour surveiller certains patients en réadaptation, les personnes âgées, les fonctions vitales des patients souffrant d'une certaine maladie.

Le projet Artificial Retina (AR) vise à construire une rétine artificielle pour les personnes malvoyantes. Plus précisément, le projet se concentre sur le traitement de deux maladies rétiniennes : la dégénérescence maculaire liée à l'âge (DMLA) et la rétinite pigmentosa (RP) [1], qui entraînent une perte de vision sévère au centre de la rétine. L'objectif du projet AR est de remédier à cet problème de vision. La figure suivante illustre le système de rétine artificielle.

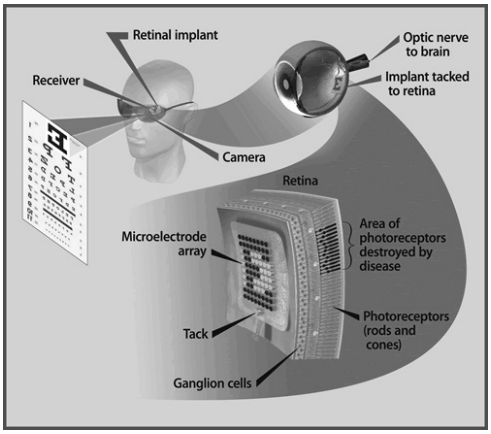


FIGURE 1.7 – Système de rétine artificiel [1]

1.7.3 Application domotique

Les RCSFs peuvent être déployés dans les appareils Electro-ménager afin d'augmenter leurs performances et les rendre intelligents. Ces capteurs peuvent être enfouis dans les réfrigérateurs, les fours à micro-ondes, les aspirateurs, les systèmes de sécurité, etc. Ils capturent des informations et interagissent entre eux et avec le réseau du domicile afin d'envoyer les informations collectées aux utilisateurs (habitants de la maison) permettant de gérer la maison localement ou à distance.

NAWMS, est un système autonome de surveillance d'eau pour les maisons. Son objectif principal est de localiser le gaspillage dans l'utilisation d'eau et d'informer les locataires pour une utilisation plus efficace [1]. La figure suivante illustre ce système.

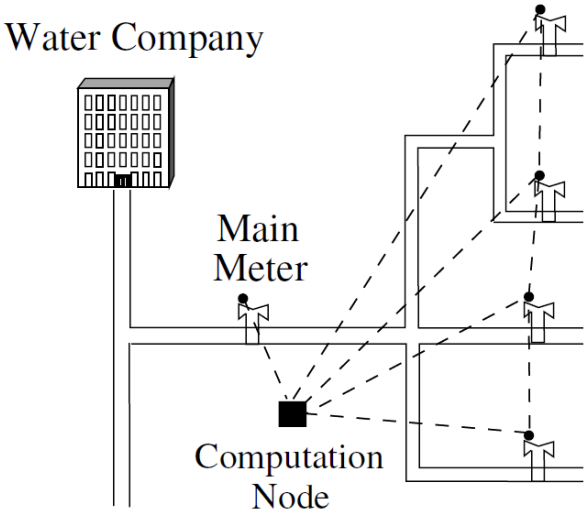


FIGURE 1.8 – Système de surveillance d'eau NAWS [1]

L'application des RCSFs ne se limitent pas à ceux décrient dessus, ils s'insèrent aussi

dans la surveillance environnementale [1], la précision d'agriculture , la construction de villes intelligentes [40], l'industrie [1].

1.8 Les défis des RCSFs

Bien que les RCSFs possèdent de nombreuses caractéristiques intrinsèques qui attirent beaucoup l'attention, comme toute autre technologie informatique ils possèdent aussi des limites particulières dont nous aborderons certaines dans cette section.

1.8.1 Les ressources limitées

La contrainte la plus importante dans les RCSFs est celle de l'énergie (puissance de la batterie limitée) du à leur petite taille. Cette miniaturisation influence d'autres capacités des nœuds capteurs comme le stockage, la puissance de calcul, la communication. La durée de vie d'un capteur étant directement liée à celle de sa batterie ainsi celle du réseau est liée à l'ensemble de ces nœuds, car l'indisponibilité de certains nœuds peut provoquer le dysfonctionnement de tout le réseau. Alors les composants du nœud doivent utiliser ses ressources de façon efficace.

1.8.2 L'auto-organisation

Cette caractéristique est l'un des défis des RCSFs de fonctionner dans des régions éloignées, dans les environnements difficiles, sans support d'infrastructure ni possibilité de maintenance ou de réparation [11]. De ce fait les nœuds de capteurs doivent s'auto-gérer, s'auto-configurer, opérer, collaborer avec d'autres nœuds capteurs, s'adapter aux changements d'environnement sans intervention humaine.

1.8.3 Le coût des nœuds capteurs

Un RCSF se compose d'un ensemble de nœuds capteurs, pouvant atteindre une grande échelle pour un projet. Le coût d'un nœud de capteur est critique pour la métrique financière globale du réseau tout entier, compte tenu du nombre élevé des nœuds composant un réseau. Certains nœuds composant un réseau de capteurs ne coûtent souvent pas plus de 1$ tandis-que d'autres coûtent 10$.

1.8.4 La sécurité

La sécurité étant l'un des points importants dans les réseaux sans-fil et dans plusieurs domaines en particulier le domaine militaire [40]. Le fonctionnement à distance et sans surveillances, la communication sans-fil des nœuds de capteurs augmentent leurs expositions aux intrusions malveillantes et aux attaques. Il existe de nombreuses méthodes permettant de sécuriser les réseaux de capteurs, mais qui nécessitent de grandes ressources et qui ne peuvent être satisfaites due à la limitation en ressources. En conséquence, il est primordial d'introduire de nouvelles solutions pour assurer les contraintes de sécurité.

1.8.5 Autres défis pour les RCSFs

Les défis des RCSFs ne se limitent pas à ceux susmentionnés, ils sont contraints aussi en termes de fiabilité [40], de performance en temps réel [40], de conception, de déploiement, de gestion décentralisée, de tolérance aux fautes, de scalabilité, etc.

1.9 Conclusion

En parcourant ce chapitre nous notons une large étude sur les RCSFs qui nous a permis de découvrir les caractéristiques remarquables que possède ce type de réseau, mais aussi de découvrir certains défis auxquels ils sont confrontés malgré les avancées technologiques. Le chapitre suivant fera l'objet d'étude du sujet sur lequel nos travaux se portent qu'est l'estimation des données manquantes dans les RCSFs.

CHAPITRE 2

ESTIMATION DES DONNÉES MANQUANTES DANS UN RCSF

2.1 Introduction

Rare sont les jeux de données parfaitement renseignés. Leurs collecte se fait fréquemment avec quelques imprévus entrainant différents phénomènes dans le jeu de donnée comme la présence des données manquantes. Ce problème existe depuis les premières tentatives d'exploitation de données comme connaissance [40], et les RCSFs sont aussi confrontés à ce problème.

Les nombreuses contraintes des RCSFs comme la mode de transmission non fiable, la défaillance de signal, l'épuisement d'énergie des nœuds de capteur, il arrive assez fréquemment que des observations soient incomplètes. Or la plupart de ces données collectées feront l'objet d'analyse et éventuellement de prendre des décisions avec. En présence de ces données manquantes, l'analyse peut être biaiser considérablement, les performances du système utilisant ces données peuvent être réduites, car la plupart des méthodes de traitements et d'analyses exigent des ensembles de données complètes. D'ou la nécessité de mettre en place des approches permettant d'estimer de façon sûr et efficace ces données manquantes dans les RCSFs.

On peut penser résoudre pour un premier temps ce problème de données manquantes en éliminant certaines contraintes sur les nœuds capteurs comme assurer la fiabilité de la transmission, intégrer des modules de régénération d'énergie, etc. Mais ces solutions font soulever d'autres problèmes comme les capteurs de grandes taille, un surcout du nœud capteur, un vite épuisement de la batterie des nœuds capteurs, une durée de communication élevée, etc.

Une autre solution serait de mettre en place des méthodes permettant d'estimer des données manquantes dans les RCSFs. Ces approches doivent, estimer les données perdues avec un taux d'erreur minimal. Dans ce sens, plusieurs méthodes d'estimation de données manquantes ont été proposées. Ce chapitre fera l'objet d'étude de certaines d'entre elles, et des données manquantes.

2.2 Définition des données manquantes(missing data)

Avant d'exposer des concepts relatifs aux données manquantes, il est nécessaire de les comprendre. Une donnée manquante est définie comme un phénomène qui se produit lorsqu'une valeur ou information n'est obtenue dans une observation. De même elle peut aussi être définit comme étant l'absence d'une valeur ou d'un ensemble de valeurs dans une observation qui seraient significatives si elles sont observées et existent dans beaucoup de jeu données du monde réel. C'est un problème qui existe toujours dans les ensembles de données pour diverses raisons. Elles peuvent être une information incomplète, des fichiers manquants, des erreurs de saisie etc. Sa présence rend un ensemble de données incomplet, par conséquent non fiable et l'exploitation direct de cet ensemble de données en sa présence peut entrainer des complications. Ainsi les données issues du monde réel doivent passer par des processus de préparation pour pouvoir être utilisées facilement sans complication. La figure suivante illustre l'aspect visuel de données manquantes.

	t1	t2	t3	t4	t5	t6	Tm
S1	1	2	3	4	5		7
S2	6	3	4	5	6	7	1
S3	5	6	8	9	4	5	2
S4	4	7		2	4	9	2
S5	3	8	4	77		10	3
S6	2	9	5		7	11	5
S7	1	16	15	14	13	12	6
Sn	6	5	4	3	2	1	7

FIGURE 2.1 – Aspect visuel de données manquantes

2.3 Les causes de perte des données dans les RCSFs

Plusieurs raisons expliquent la présence de données manquantes dans les jeux de données collectés par les RCSFs comme : la capacité de communication des nœuds capteurs limitée,

la présence d'interférences environnementale, les conditions environnementales sévères (pluie, tonnerre, foudre), la limitation d'énergie des nœuds de capteurs [37]. De même selon [28, 27], l'endommagement du nœud de capteur, les collisions de paquets envoyés, la mauvaise performance des algorithmes de routage et la communication sans fil en tant que t'elle sont aussi des causes qui entrainent des pertes de données dans les RCSFs.

2.4 Typologie des données manquantes

Connaitre le type des données manquantes sont des conditions primordiales pour aborder quelle approche utilisera t-on pour leur estimation. Dans [43, 29], une typologie a été proposer par Little & Rubin(1987) pour les données manquantes en les répartissant en trois catégories.

2.4.1 Missing Completely at Random (MCAR)

MCAR est l'hypothèse la plus fréquente dans les jeux de données de la vie réel. Une donnée manquante est de ce type, c'est à dire manquante de façon complètement aléatoire, si son absence dépend uniquement des paramètres extérieurs [43]. Elle se produit généralement en raison d'une défaillance d'équipement de mesure, ou si l'échantillon est insatisfaisant.

2.4.2 Missing at Random (MAR)

Le type de données manquantes MAR n'est pas fréquent. Il survient lorsque les données ne manquent pas suivant MCAR [43]. En d'autres termes les manques de données sur une variable sont liées à certaines autres variables observées.

2.4.3 Missing Not at Random (MNAR)

MNAR est la forme de données manquante la plus difficile à rencontrée selon [30]. appelé également manquement non ignorable dans [30, 29], elle se produit si les données ne manquent pas suivant le type MCAR ou MAR. C'est à dire la probabilité d'absence dépend de la variable.

2.5 Modèle de pertes de données des RCSFs

Les travaux traditionnels supposent généralement que la perte de données suit une distribution aléatoire [31], cette revendication n'est pas correcte dans tout les cas de données manquantes. Dans les RCSFs la manque de données suit certains modèles.

Selon [31, 29], il existe cinq modèles de perte de données dans les RCSFs (ERL, BRL, EFLR, SELR, CL) . Afin de bien illustrer ces modèles considérons une matrice M qui enregistre les données brutes collectées par un RCSF, les données de la matrice sont composées d'une valeur $M(i, j)$ si elles ne sont pas manquantes et "NAN" sinon.

2.5.1 Element Random Lost (ERL)

Ce modèle a pour cause profonde les bruits et les collisions dans les RCSFs [31, 29]. Les données de la matrice M sont perdues de façon indépendante et au hasard. Les valeurs manquantes sont réparties de manière aléatoire sur les nœuds de capteur. La figure 2.2a l'illustre.

2.5.2 Block Random Lost (BRL)

Dans ce modèle de perte, les données des nœuds adjacents ou proches dans un intervalle de temps adjacents se perdent ensemble. La principale cause de ce type de modèle est la congestion [31, 29]. La figure 2.2b l'illustre.

2.5.3 Element Frequent Loss in a Row (EFLR)

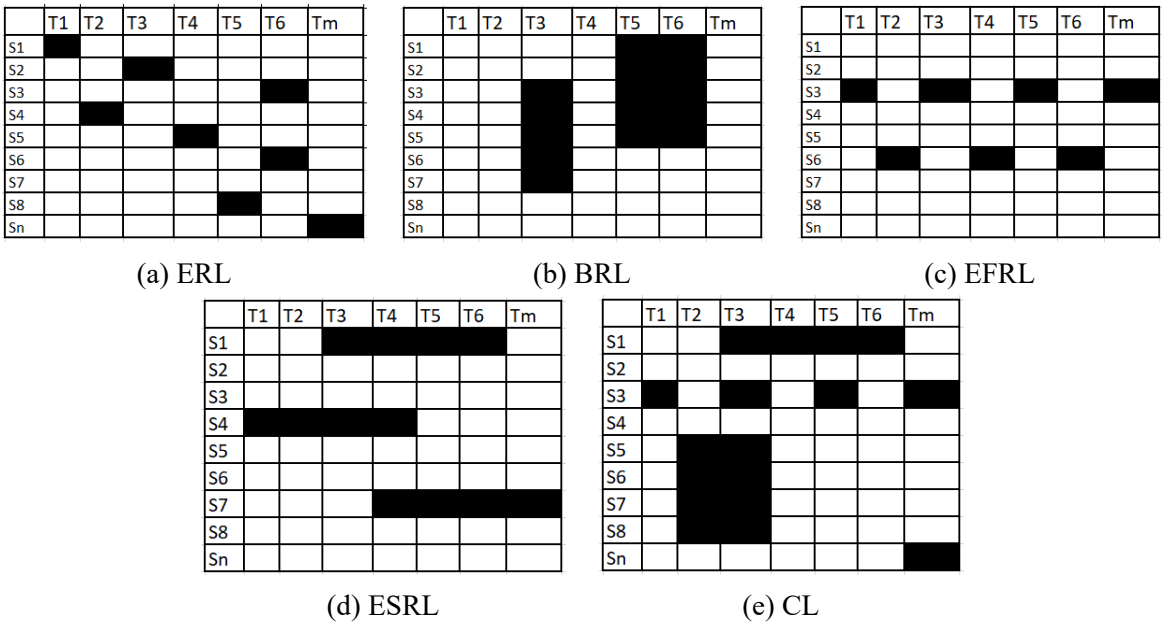
Dans ce modèle, les données d'un capteur se perdent fréquemment dans le temps. Les liaisons non fiables, l'intermittence de la transmission, la mauvaise qualité des liens sont à l'origine de ce modèle [31, 29]. La figure 2.2c l'illustre.

2.5.4 Successive Elements Loss in a Row (ESLR)

Ce modèle de perte survient quand certains nœuds commencent à perdre des données de façon consécutive dans le temps. Ceci est dû en un endommagement, une manque d'énergie du nœud capteurs [31, 29]. La figure 2.2d illustre ESLR.

2.5.5 Combinational Loss (CL)

Ce modèle de perte est la combinaison des autres modèles de perte susmentionnés.



(a) ERL (b) BRL (c) EFRL

(d) ESRL (e) CL

FIGURE 2.2 – Les modèles de perte de données des RCSFs

2.6 Répartition des données manquantes

Soit $Y = (y_{ij}) \in R^{n*p}$ représentant la matrice rectangulaire des données pour p variables $y_1, ..., y_p$ et n observations. Considérons $M = (m_{ij})$ la matrice d'indication de valeurs manquantes. Cette matrice sera utilisée pour définir la répartition des valeurs manquantes et selon [43], il existe trois types.

2.6.1 Répartition univariée

Les valeurs manquantes sont univariées, s'il existe une valeur d'observation y_{ki} manquante pour une seule variable Y_k. Ceci implique qu'il n'y aura plus d'observation de cette variable. La figure 2.3a l'illustre.

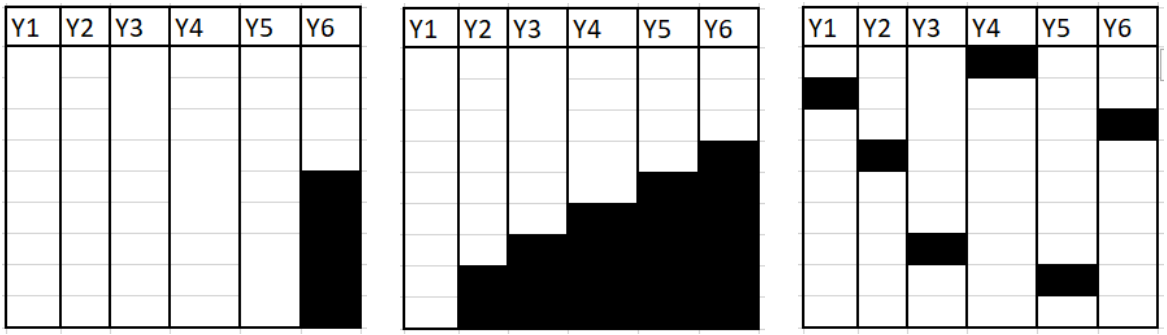
2.6.2 Répartition monotone

Les valeurs manquantes sont monotones s'il existe une variable Y_j manquante pour un nœud de capteur i, alors ceci implique que toute les variables suivantes $\{Y_k\}_{k>j}$ sont man-

quantes pour ce nœud de capteur. L'indicateur de valeurs manquantes M est alors un entier $M \in (1, 2, ..., p)$ pour chaque nœud, indiquant le plus grand j pour lequel Y_j est observé. La figure 2.3b l'illustre.

2.6.3 Répartition non monotone ou arbitraire

Les valeurs manquantes sont arbitraires si la matrice d'indication de valeurs manquantes est définie par $M = (m_{ij})$ avec $m_{ij} = 1$ si y_{ij} est manquant et 0 sinon. La figure 2.3c l'illustre.



(a) Répartition univariée (b) Répartition monotone (c) Répartition arbitraire

FIGURE 2.3 – Schéma de répartition des données manquantes

2.7 Probabilité d'absence

La probabilité d'absence selon le type de données manquantes (MCAR, MAR, MNAR) peut être exprimer en fonction de la matrice M d'indication de données manquantes [43]. Les données sont réparties en deux selon M. On définit donc $Y_{obs} = Y_{1_{\{M=0\}}}$ les données observées et $Y_{mis} = Y_{1_{\{M=1\}}}$ les données manquantes telle que $Y = \{Y_{obs}; Y_{mis}\}$. Le mécanisme de données manquantes est caractérisé par la distribution conditionnelle de M sachant Y donné par la probabilité $p(M/Y)$. Suivant les types des données manquantes, il y aura trois probabilités d'absence [43].

- Dans le cas des données de type MCAR, l'absence de données est indépendante des valeurs de Y donc : $p(M/Y) = p(M)$ pour tout Y [40, 43].
- Suivant MAR, l'absence de données dépend uniquement de Y_{obs} et non de Y_{mis} avec : $p(M/Y) = p(M/Yobs)$ pour tout Y_{mis} [40, 43]
- Suivant MNAR, la distribution de M dépend aussi de Y_{mis} [40, 43].

2.8 Estimation des données manquantes dans les RCSFs

Un RCSF est un ensemble de nœuds capteurs qui communiquent entre eux pour collecter des données brutes dans un environnement et les envoyées à un utilisateur final pour des fins d'utilisation. Ces utilisations peuvent être d'ordre d'exploration et d'analyse. Donc les données collectées par les RCSFs doivent bien représentées l'environnement dans laquelle elles ont étés collecter. Mais les contraintes telles que l'utilisation du protocole de communication non fiable, couplé à d'autres contraintes matérielles, la défaillance du signal, l'endommagements du nœuds capteurs, affectent cette collecte entrainant ainsi des pertes de données des nœuds de capteurs. Avec la présence de ces valeurs manquantes, l'utilisation du jeu de données devient difficile et peut même conduire à des résultats erronés. Donc la perte de données devient l'un des défis principaux de ce type de réseau. Pour remédier à ce problème, différentes solutions ont été proposées comme l'utilisation de protocole de transmission fiable [32], perceptrons multi couche (MLP), Adaptive resonnance theorie (ART), etc. Mais ces solutions soulèvent d'autres inconvénients comme une grande consommation d'énergie par les capteurs dégradant ainsi leur durée de vie, augmentent le coût des nœuds capteurs, et aussi entraine un retard dans l'ensemble du réseau. Pour éviter ces problèmes nous avons pris le chemin d'utiliser des méthodes permettant d'estimer les données manquantes. La section suivante fera l'objet d'étude de méthode d'estimation et de certaines mises en œuvre dans le cadre des RCSFs.

2.8.1 Qu'est ce qu'une méthode d'estimation

Une méthode d'estimation est l'ensemble des techniques utilisées pour compléter un jeu de données avant toute utilisation en remplaçant ses valeurs manquantes. Face au problème de perte de données, plusieurs méthodes d'estimation ont vu le jour et peuvent être classer en 2 grandes catégories : les méthodes d'imputation modernes et traditionnelle [29].

Les méthodes d'estimation traditionnelles

En présence de données manquantes, ces méthodes les supprime ou les estime avec une seule valeur et contient les méthodes comme :

— **Suppression par liste :** Dans [29, 30], elle supprime les observations contenant des données manquantes.

— **Moyenne de classe :** Dans [29, 30], elle estime les données manquantes en les replaçant

par la moyenne de l'observation sur laquelle la donnée est manquante.

— **Imputation aléatoire :** Dans [30], elle estime les données manquantes en les remplaçant par une valeur choisi de façon aléatoire dans le flux de données.

— **MEDIAN :** Elle remplace la donnée manquante par la médian de l'observation.

Ils existent d'autres méthodes d'imputation traditionnelle comme : imputation séquentielle hot-deck, l'imputation déductive, imputation hiérarchique Hot-Deck [30]. Ces méthodes sont facile à utilisées, mais ne peuvent être appliquer que sur les données manquantes de types MCAR, aussi si la taille du jeu données n'est pas assez importante, et peuvent entrainer des biais.

Les méthodes d'estimation modernes

Elles utilisent des calcules mathématiques ou statistiques pour imputer les valeurs manquantes d'un jeu de données et contient les méthodes comme :

— **Imputation multiple :** Dans [26], elle utilise la corrélation spatiale et temporelle mais aussi les variations globale et locale pour l'estimation. Elle fait m imputation et choisi la meilleur valeur pour l'estimation.

— **K plus proche voisin :** Dans [27], elle utilise les valeurs des objets qui sont similaire ou proche pour l'imputation en calculant la distance ou la similarité entre eux.

— **Méthode ML :** Elle utilise les algorithmes d'apprentissage automatique pour gérer les données manquantes.

Cette catégorie de méthodes ne se limite pas à ceux susmentionnées mais ils existent d'autres comme : l'algorithme génétique, régression, règles d'association [30]. Elles possèdent l'avantage de pouvoir s'appliquer sur tous les types de données manquantes mais quelques une d'entre elles sont difficiles d'utiliser. La figure suivante illustre une classification de certaines d'entre elles.

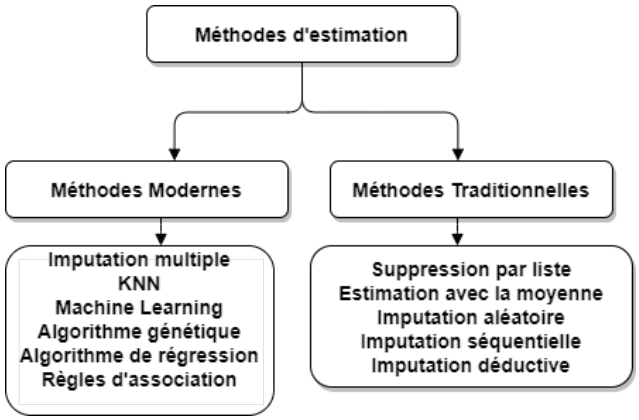


FIGURE 2.4 – Diagramme de classement des méthodes d'estimation

2.9 Les mesures de performance

Avec l'émergence des techniques d'imputations de données manquantes, un grand nombre d'études comparatives visant à évaluer leur performances ont vu le jour. Elle est l'étape qui suit directement l'imputation et quantifient l'effet de la méthode d'imputation. Dans [33], il existe différente méthodes pour évaluer le résultat d'une méthode d'imputation. Mais nous nous sommes intéressé aux méthodes d'évaluations directes.

Évaluation directe : Elle est la méthode d'évaluation la plus utilisée, elle évalue directement la différence entre les valeurs d'origine dans l'ensemble des données collectées et les valeurs estimées dans l'ensemble des données incomplètes simulées [33]. Nous utiliserons 2 méthodes (MAE, RMSE) pour ce type d'évaluation. Dans les équations de ces méthodes nous considérons que x_i est la valeur d'une mesure supposée manquante et x'_i est cette valeur estimée, et n est le nombre total de valeurs manquantes ou estimer.

— **MAE :** Elle est obtenue à partir des mesures liées à l'erreur absolue moyenne en pourcentage et cela peut être calculé par la formule suivante :

$$MAE = \frac{100}{n} \sum_{i=1}^{n} |\frac{x_i - x'_i}{x_i}| \tag{2.1}$$

— **RMSE :** Elle est obtenue à partir de l'erreur quadratique moyenne et peut être calculée par la formule suivante :

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - x'_i)^2} \tag{2.2}$$

Il existe d'autres méthodes d'évaluation comme selon la précision de la classification [33], le temps d'imputation, stratégie de simulation pour l'imputation [33], la complexité de l'algorithme d'imputation etc.

2.10 Étude comparative de quelques méthodes d'estimation

Le problème de données manquantes dans les RCSFs est devenu omniprésent, de ce fait plusieurs méthodes d'estimations de ces données ont fait leur apparition. Ces méthodes permettent toutes d'estimer les données manquantes, mais elles ne procèdent pas toutes de la même

façon, d'autres se basent sur la corrélation temporelle, d'autres aussi sur la corrélation spatiale et d'autres sur les deux corrélations, et d'autres sur des calcules mathématiques ou statistiques. Avec cette diversité de façon de procéder à l'imputation, elles peuvent être performantes l'une que l'autre.

Dans [37], une approche LOCF utilise la corrélation spatiale pour faire l'estimation est décrit ou les valeurs manquantes sont estimées en les remplaçant par la dernière observation non manquante. L'algorithme n'utilisant que les dernières observations pour l'estimation réduit la précision et la fiabilité des données, et l'aspect spatiale est négligée. Dans [27], une méthode d'estimation KNN est mise en place ou les données du nœud le plus proche sont utilisées pour faire l'estimation des données manquantes d'un autre nœud. Cette méthode n'utilisant que la corrélation spatiale, elle néglige l'aspect temporelle des données et fournit des résultats d'estimation de faible précision, et nécessite des cordonnées spatiales. De même dans [28], un algorithme d'estimation WARM est décrit ou le capteur voisin du capteur défaillant est utilisé pour l'estimation des données de celui dont la données est manquantes. Il utilise le principe de fenêtre glissante, alors seuls les derniers w tour de rapport de données sont utilisés pour l'estimation. Cet algorithme est limité en choix de la taille de la fenêtre (petite w comporte un risque de perte et grand w entraine une surcharge d'espace), il ignore aussi l'aspect temporel des données. Un autre algorithme similaire FARM a été proposé dans cite [39], il utilise des règles d'associations pour trouver les capteurs. Ainsi l'estimation d'une lecture manquante d'un capteur est faite à partir des données des capteur trouver. Ce algorithme souffre aussi car il néglige l'aspect temporel des données des RCSFs. De même dans [35] l'algorithme TSCA et dans [36] l'algorithme STC utilisent la corrélation spatiaux-temporelle pour l'estimation. Ces méthodes ont montré leur supériorité par rapport aux autres mais sont complexe dans les applications de la vie réel.

Les méthodes d'estimation ne se limitent pas à ceux susmentionnées mais plusieurs autres existent comme CARM, TKCM [34], Matrix Factorization, Support Vector Machine(SVM), Tensor Decomposition, imputation multiple [26], méthode utilisant les techniques de ML.

2.11 Conclusion

En parcourant ce chapitre, nous avons fournit un large aperçu sur les données manquantes en général et dans le cadre des RCSFs. Pour ces données manquantes, nous avons discuté de leurs estimations, les causes de leurs perte dans le cadre des RCSFs, mais aussi de leurs différents types, de leurs modèles de perte, ainsi que de leurs répartitions, etc. Le chapitre a été achevé

en faisant une étude comparative entre certaines différentes méthodes d'estimations existantes. Suite aux limites de ces méthodes, deux nouvelles méthodes seront proposée dans le chapitre suivant.

CHAPITRE 3

PROPOSITION D'UNE APPROCHE D'ESTIMATION DE DONNÉES MANQUANTES DANS LES RCSFS

3.1 Introduction

Avec les caractéristiques inhérentes des RCSFs, la présence de données manquantes dans les données collectées par ce type de réseau est inévitable. Pour l'estimation de ces données manquantes différents travaux ont vu le jour. Mais ces derniers présentent encore des problèmes, leur application est souvent possible que sur des jeux de données comprenant un faible taux de données manquantes, leurs complexités de calcul importantes, la négligence de certaines caractéristiques des données des RCSFs, etc. Le présent chapitre fera l'objet d'analyse du jeu de données choisi pour l'étape d'évaluation de nos méthodes d'estimation.

Le jeu de données qui a été choisi est celui d'Intel Berkeley [44], il est composé de plusieurs enregistrements collectés par 54 capteurs qui sont déployés dans l'Intel Berkeley Research Lab entre la période du 28 février et le 5 avril 2004. Ces capteurs sont de types Mica2Dot, occupés d'une carte météorologique, d'un SE TinyOs, et d'un middleware TinyDB. Ils ont collecté des informations relatives à l'humidité, à la température, à la lumière, et à la tension une fois toutes les 31 secondes. Ce dataset possède un taux de données manquantes de 23% [31, 36], les capteurs de ce RCSF sont organisés comme indiqué dans la figure ci-après.

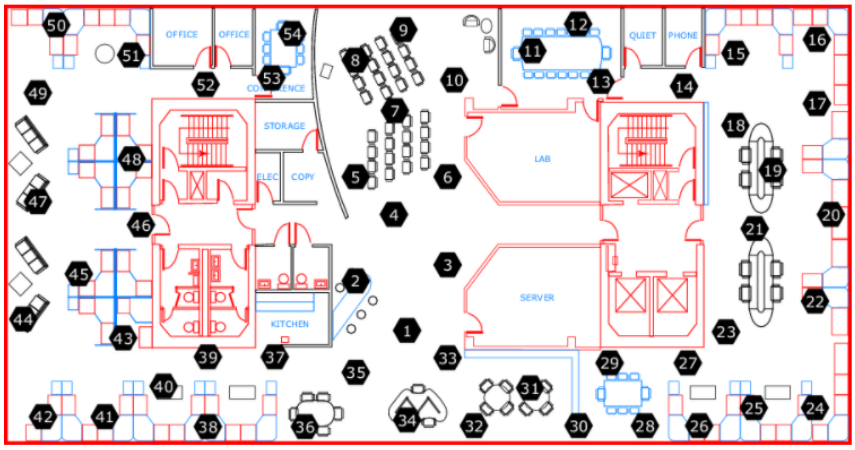


FIGURE 3.1 – Disposition des capteurs déployés dans Intel Berkeley Research Lab [44]

3.2 Analyse complète du jeu de données

L'analyse des données est un processus crucial dans les domaines utilisant les ensembles de données comme source de connaissance. Elle consiste en la découverte d'informations et d'interprétation du sens des données afin d'avoir un aperçu sur l'ensemble du jeu de données. Pour se faire, nous passerons par différentes séries d'analyses.

3.2.1 Analyse descriptive

Le jeu de données choisi contient des informations collectées sur plusieurs attributs.

— **La date (date) :** Elle est la date à la quelle la mesure a été faite, et est de la forme yyyy-mm-dd.

— **L'heure (time) :** Elle est l'heure à la quelle la mesure a été faite, et est de la forme hh :mm :ss.xxxx.

— **epoch (epoch) :** Il est l'attribut qui stocke le numéro de séquence envoyé par chaque capteur de façon croissante et monotone après chaque capture. Il est de type entier.

— **L'identifiant des nœuds (moteid) :** Comme son nom l'indique, il est l'identifiant des capteurs déployés pour la collecte et prend uniquement des valeurs entières comprises entre 1 et 54.

— **La température (temperature) :** Elle stocke la température de l'environnement dans lequel sont déployés les capteurs. Elle est de type réel, et est exprimée (°C).

— **L'humidité (humidity) :** Elle est l'humidité de l'environnement dans lequel sont dé-

ployés les capteurs, et est liée à la température. Elle est de type réel et sa valeur est comprise entre 0 et 100%.

— **La lumière (light) :** Elle est l'intensité lumineuse de l'environnement dans lequel son déployés les capteurs. Elle est de type réel et est exprimée en Lux.

— **La tension (voltage) :** De type réel, elle est exprimée en volts et est fortement liée à la valeur de l'attribut température, sa valeur varie entre 2 à 3.

Le tableau suivant résume la description et le format des attributs décrient ci-dessus.

date : yyyy-mm-dd	time : hh :mm :ss.xxx	epoch : int	moteid : int
temperature :real	humidity :real	light : real	voltage : real

TABLE 3.1 – Résumé de la description et du format des attributs du dataset

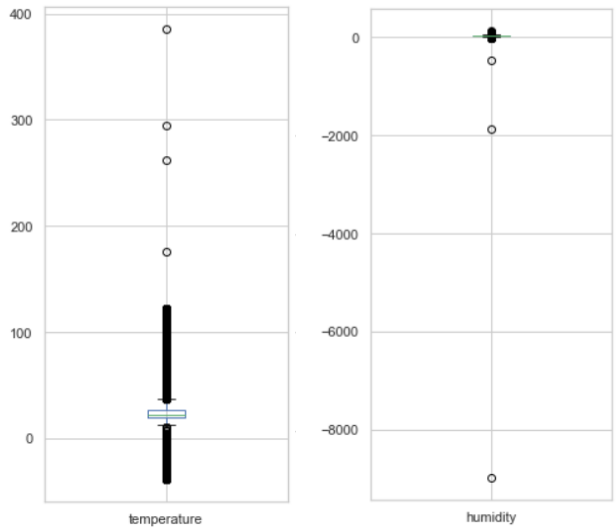
Le tableau suivant donne un aperçu d'une partie du jeu de données.

	date	time	epoch	moteId	temperature	humidity	light	voltage
1	2004-03-31	03 :38 :15.757551	2	1.0	122.1530	-3.9190	11.04	2.0339
2	2004-02-28	00 :59 :16.02785	3	1.0	19.9884	37.0933	45.08	2.6996
3	2004-02-28	01 :03 :16.33393	11	1.0	19.3024	38.4629	45.08	2.6874
4	2004-02-28	01 :06 :16.013453	17	1.0	19.1652	38.8039	45.08	2.6874
5	2004-02-28	01 :06 :46.778088	18	1.0	19.1750	38.8379	45.08	2.6996
6	2004-02-28	01 :08 :45.992524	22	1.0	19.1456	38.9401	45.08	2.6874
7	2004-02-28	01 :09 :22.323858	23	1.0	19.1652	38.8720	45.08	2.6874
8	2004-02-28	01 :09 :46.109598	24	1.0	19.1652	38.8039	45.08	2.6874
9	2004-02-28	01 :10 :16.6789	25	1.0	19.1456	38.8379	45.08	2.6996
10	2004-02-28	01 :10 :46.250524	26	1.0	19.1456	38.8720	45.08	2.6874

TABLE 3.2 – Aperçu d'une partie du jeu de données

3.2.2 Analyse statistique

Après une analyse descriptive du jeu de données, nous allons procéder à une analyse statistique afin de fournir graphiquement à partir d'indicateurs des résumés sur les séries de valeurs présentes dans le dataset. Ces indicateurs peuvent être la moyenne, la médiane, les quartiles, etc. Nous nous sommes intéressés dans cette analyse statistique uniquement aux attributs (température et humidité) dont leur valeurs sont issues de l'environnement. Les figures suivantes résument la série de valeur présente dans ces 2 attributs.



(a) température (b) humidité

FIGURE 3.2 – Résumé des variables (observation de l'environnement) du dataset

Nous pouvons constater sur la figure 3.2a la variation de la température entre la valeur de -38,4°C et 125°C dans l'environnement où sont déployés les capteurs. La température moyenne de l'environnement est de 39,20°C.

De même sur la figure 3.2b, l'humidité de la zone de captage varie entre la valeur de -150% et 100%, alors que la moyenne est de 33,90%. Le tableau suivant résume la description de ces indicateurs.

	Nombre d'enregistrement	Maximum	Minimum	Moyenne	Médiane
température	2312781	385,56°C	-38,40°C	39,20°C	22,43°C
humidité	2312780	137,51%	-8983.13%	33.90%	39,28%

TABLE 3.3 – Résumé de l'analyse statistique du jeu de données

3.2.3 Analyse de la corrélation

Après une analyse descriptive et statique, une analyse de la corrélation permettra de faire une étude entre les attributs d'observation, et ainsi permettre de voir à quel point nos données varient ensemble. La figure suivante représente le résultat de cette analyse.

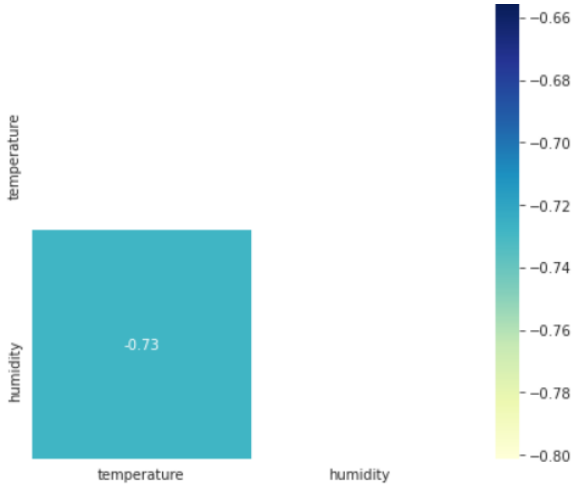


FIGURE 3.3 – Matrice de corrélation

Sur la figure 3.3, on constate qu'il existe une forte relation entre la température et l'humidité de l'environnement ou son déployés les capteurs. De ce fait, la température de cet environnement dépend fortement de l'humidité et vis versa.

3.3 Squelette de la méthode d'estimation

Une méthode d'estimation est une procédure qui permet d'estimer les valeurs manquantes d'un jeu de données tout en minimisant le plus possible le taux d'erreur d'estimation.

3.3.1 Représentation des données des RCSFs

Un RCSF est composé d'un ensemble de capteurs $S = \{s_1, s_2, .., s_n\}$ dispersés dans un environnement et collectant des données sur cet environnement pendant des tranches de temps $T = \{t_1, t_2, .., t_m\}$. Ces données collectées peuvent être représentées par une matrice M de taille $n * m$ où S est l'ensemble des capteurs de mesures, et T est l'ensemble des temps de mesure. $M(s_i, t_j)$ désigne la donnée prélevée par le capteur s_i au temps t_j ou $i \in 1, .., S$ et $j \in 1, .., T$ [31, 35]. Ces données collectées sont non manquantes si elles sont présentes dans M sinon elles sont considérées comme manquantes. La matrice suivante illustre la représentation des données collectées par un RCSF.

$$M = \begin{pmatrix} (s_1, t_1) & (s_1, s_2) & \cdots & (s_1, t_m) \\ (s_2, t_1) & (s_2, s_2) & \cdots & (s_2, t_m) \\ \vdots & \vdots & \ddots & \vdots \\ (s_n, t_1) & (s_n, t_2) & \cdots & (s_n, t_m) \end{pmatrix}$$

3.3.2 Formulation du problème d'estimation de données manquantes

Soit M la matrice contenant les données envoyées par les capteurs d'un RCSF, certaines d'entre elles peuvent ne pas être reçues et par conséquent sont considérées absentes. Soit S_{miss} le capteur dont la donnée est manquante, V_{miss} est manquante au temps T_{miss}. Le but ici est de trouver EV la valeur estimée de cette donnée manquante, et qui s'approche le plus possible de la valeur originale (manquante) [31]. De même dans [36], la méthode d'estimation doit minimiser le plus possible la différence entre la valeur manquée et celle estimée afin d'être précise dans l'estimation comme indiqué dans l'équation suivante .

$$min \, \|V_{miss} - EV| \tag{3.1}$$

3.4 Les méthodes des moyennes

Elles consistent à utiliser la valeur moyenne d'une observation pour remplacer les données manquantes. Nous allons pour cela utilisé la moyenne harmonique et géométrique dans d'autre contextes.

3.4.1 Méthode d'estimation HMEAN

Elle est notre première approche. Elle consiste à estimer les valeurs manquantes d'un capteur par la moyenne harmonique des n valeurs suivantes non manquantes enregistrées après la donnée manquante ou les n valeurs précédentes non manquantes enregistrées avant la donnée manquante avec $n \in \{2, 10\}$. Elle est définie comme l'inverse de la moyenne arithmétique, et est calculée en divisant le nombre d'observations total par la somme des inverses de chaque mesure dans la série et est obtenue à partir de la formule suivante :

$$\bar{x} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \tag{3.2}$$

x_i est une valeur de mesure non manquante du capteur S_{miss} parmi les valeurs à utiliser pour l'estimation, et n est le nombre total de valeurs utilisées pour l'estimation. Cette méthode d'estimation

1. Définit les paramètres dont elle aura besoin pour l'estimation. À savoir une liste $ListNotMiss$ de n valeurs non manquantes suivantes enregistrées par le capteur à partir du temps $T_{miss}+1$ ou précédentes enregistrées par le capteur à partir de $T_{miss}-1$.
2. Calcule la moyenne harmonique des valeurs présentent dans la liste précédemment définie ($ListNotMiss$)
3. Remplace la donnée manquante du capteur S_{miss} au temps T_{miss} par la valeur estimée.

Avant d'écrire l'algorithme de cette méthode d'estimation, nous allons définir certains paramètres dont les méthodes auront besoin lors du processus d'estimation. Ces paramètres sont la liste des n valeurs non manquantes $ListNotMiss$

Algorithme 1 : Algorithme de définition de paramètres pour l'estimation

Algorithme 1 : Algorithme de *définition de paramètres*

Entrées : S_{miss}, T_{miss}, matriceDeDonnees, direction, nValeur

Sorties : La liste des nValeurs $ListNotMiss$

si *direction ==”suivant”* **alors**
- debut ← $T_{miss} + 1$
- taille ← **taille**(matriceDeDonnees(S_{miss}))
- **tant que** *nValeurs >0 && debut<taille* **faire**
  - **si** *matriceDeDonnees(S_{miss},debut) n'est pas ”NAN”* **alors**
    - Ajouter la valeur matriceDeDonnees(S_{miss},debut) à liste $ListNotMiss$
    - nValeurs ← nValeurs-1
  - **fin**
  - debut ← debut+1
- **fin**

fin

sinon
- debut ← $T_{miss} - 1$
- **tant que** *nValeurs >0 && debut>=0* **faire**
  - **si** *matriceDeDonnees(S_{miss},debut) n'est pas ”NAN”* **alors**
    - Ajouter la valeur matriceDeDonnees(S_{miss},debut) à liste $ListNotMiss$
    - nValeurs ← nValeurs-1
  - **fin**
  - debut ← debut-1
- **fin**

fin

L'algorithme suivant illustre le fonctionnement de la méthode d'estimation HMEAN.

Algorithme 2 : Algorithme de la méthode d'estimation HMEAN

3.4.2 Méthode d'estimation GMEAN

Elle est notre seconde approche et consiste à estimer une donnée manquante d'un capteur par la moyenne géométrique de ses n données précédentes ou suivantes enregistrées, avec $n \in \{2, 10\}$. Elle est définie comme la racine Nième du produit des valeurs de mesures, et est calculée

Algorithme 2 : Algorithme *d'estimation HMEAN()*

Entrées : S_{miss}, T_{miss}, matriceDeDonnees, ListNotMiss
Sorties : matriceDeDonneesImputer
n ← **taille**(ListNotMiss)
i ← 0
sommeInverse ← 0
moyenneHarmonique ← 0
pour *i allant de 0 à n* **faire**
  si *ListNotMiss(i)* $\neq$ *0* **alors**
    sommeInverse ← sommeInverse+$\frac{1}{ListNotMiss(i)}$
  fin
fin
moyenneHarmonique ← $\frac{n}{sommeInverse}$
matriceDeDonnees(S_{miss}, T_{miss})← moyenneHarmonique

à partir de la formule suivante :

$$\bar{x} = \sqrt[n]{\prod_{i=1}^{n} x_i} \tag{3.3}$$

x_i est une valeur de mesure non manquante dans les n valeurs à utiliser pour l'estimation, et n est le nombre de valeurs de mesure du capteur S_{miss} utilisées pour l'estimation de la donnée manquante.

1. La méthode d'estimation définit les paramètres dont elle aura besoin dans le processus d'estimation. À savoir une liste $ListNotMiss$ de n valeurs non manquantes suivantes enregistrées par le capteur S_{miss} à partir de T_{miss}+1 ou précédentes sauvegardées à partir de T_{miss}-1.
2. Calcule la moyenne géométrique des valeurs présentes dans la liste précédemment définie ($ListNotMiss$)
3. Remplace la donnée manquante du capteur S_{miss} au temps T_{miss} par la valeur précédemment calculée.

La méthode utilisera l'algorithme 1 pour la définition des paramètres dans le processus d'estimation et l'algorithme suivant illustre le fonctionnement de GMEAN.

Algorithme 3 : Algorithme *d'estimation GMEAN*

Entrées : S_{miss}, T_{miss}, matriceDeDonnees, ListNotMiss
Sorties : matriceDeDonneesImputer
n ← **taille**(ListNotMiss)
i ← 0
produitTerme ← 1
moyenneGeometrique ← 0
pour *i allant de 0 à n* **faire**
  produitTerme ← produitTerme*ListNotMiss(i)
fin
moyenneGeometrique ← $\sqrt[n]{produitTerme}$
matriceDeDonnees(S_{miss}, T_{miss})← moyenneGeometrique

Algorithme 3 : Algorithme de la méthode d'estimation GMEAN

3.5 Conclusion

Après avoir analysé le dataset, nous avons pu confirmer la présence significative de valeurs manquantes, mais aussi comprendre les données qui constituent ce dernier. Nous avons par la suite proposé deux approches qui permettent d'estimer ces valeurs manquantes. Le chapitre suivant fera l'objet d'implémentation et d'évaluation de ces approches proposées.

CHAPITRE 4

IMPLÉMENTATION, RÉSULTATS ET DISCUSSION

4.1 Introduction

Dans ce chapitre nous présentons les outils utilisés dans le cadre de la réalisation de ce projet. Suivra ensuite une présentation de certaines fonctionnalités de l'application. Enfin nous présentons les résultats des expérimentations.

4.2 Les outils de développement

Le choix d'un langage de programmation très utilisé permet de bénéficier d'un meilleur support au moment du codage, avec l'avantage de produire un outil robuste. Pour cela nous avons utilisé *Python* sous l'environnement de développement *Pycharm* et *Anaconda*.

— **Python** : Ce langage présente plusieurs avantages comme sa facilité d'utilisation, la disponibilité de nombreuses bibliothèques dans différents domaines (ML,WEB, etc), ce qui a motivé notre choix de l'utiliser.

4.3 Conception

Pour la conception de l'application, nous avons utilisé deux types de diagrammes. Un use-case permettant de définir les fonctionnalités qu'un utilisateur peut faire dans l'application

et le diagramme de séquence montrant le déroulement de ces fonctionnalités.

4.3.1 Diagramme de cas d'utilisation

Il décrit le système du point de vue des acteurs sous forme d'actions et réaction, le comportement d'un système du point de vue des acteurs.

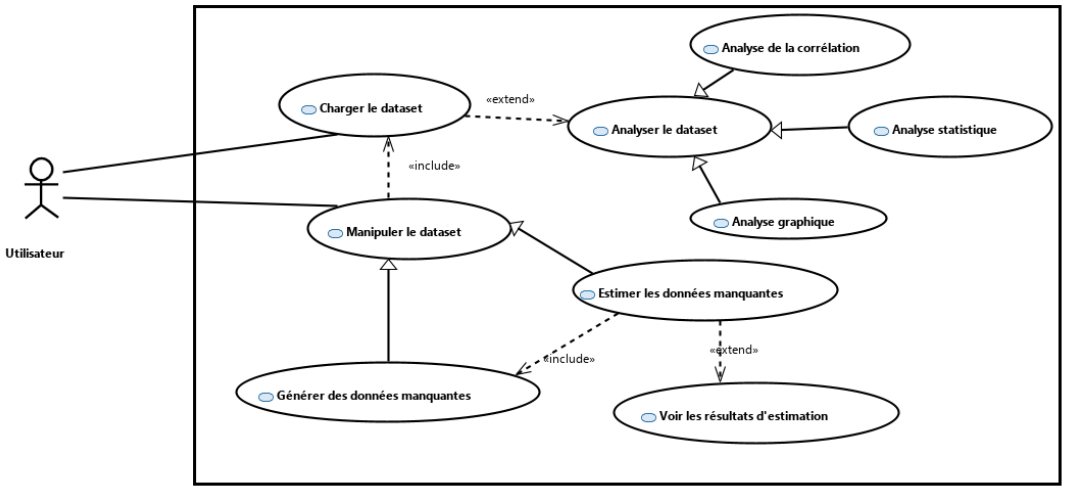


FIGURE 4.1 – Diagramme de cas d'utilisation de l'application

4.3.2 Diagramme de séquence

Il décrit le déroulement d'une fonctionnalité du système de façon séquentielle.

— **Lecture du jeu de données :** Elle est la fonctionnalité de l'application permettant de charger un jeu de données pour les autres étapes et est illustrée dans la figure 4.2.

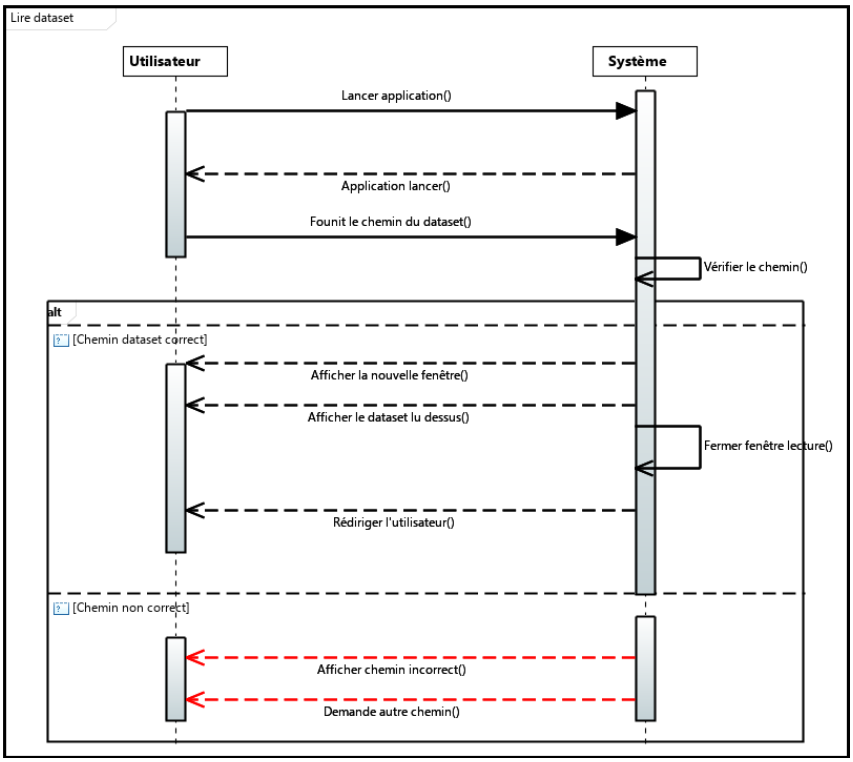


FIGURE 4.2 – Diagramme de séquence de lecture du jeu de données

— **Génération de données manquantes :** Elle est la procédure permettant de générer des données manquantes et est illustrée dans la figure 4.3.

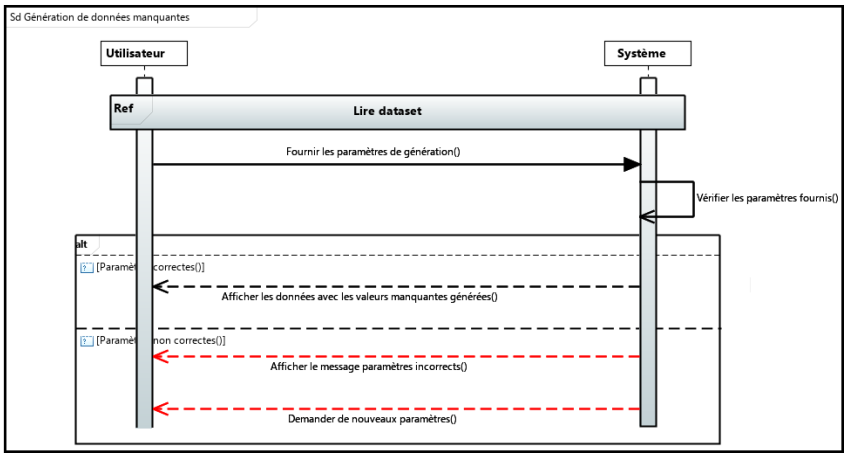


FIGURE 4.3 – Diagramme de séquence de génération de données manquantes

— **Estimer les données manquantes :** Elle est la procédure permettant d'estimer les données manquantes générées et est illustrée dans la figure 4.4.

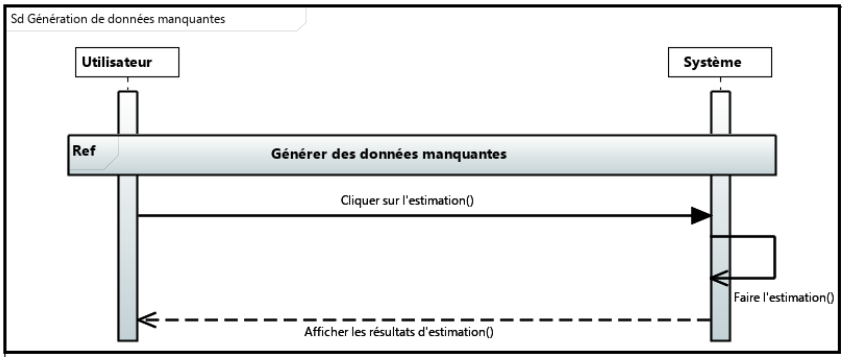
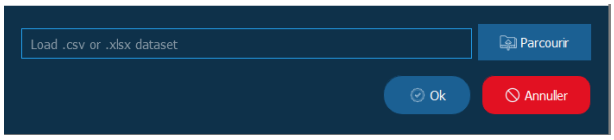


FIGURE 4.4 – Diagramme de séquence d'estimation des données manquantes

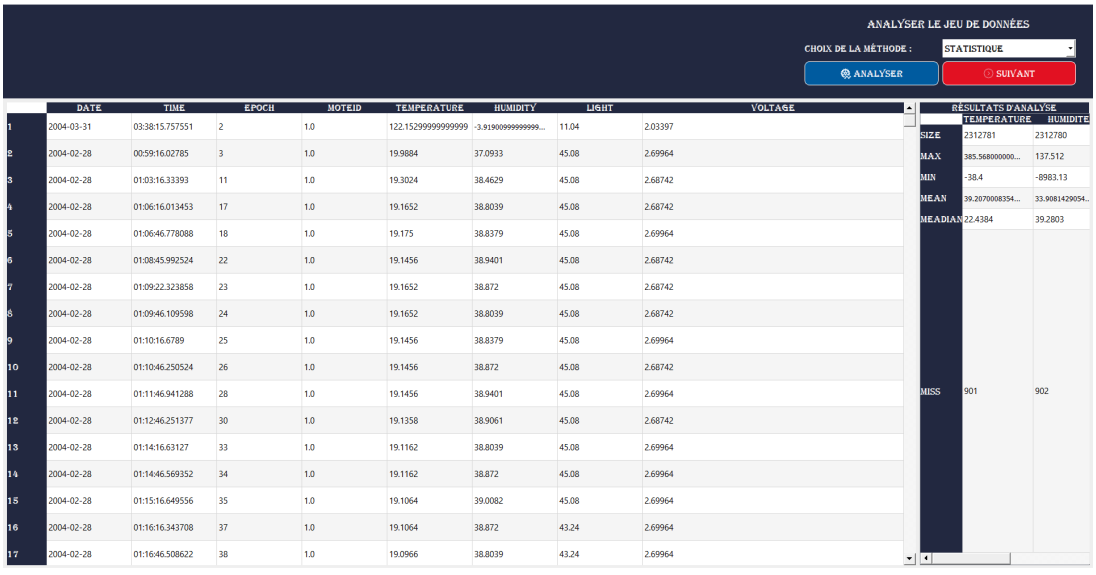
4.4 Présentation de l'application

4.4.1 Lecture du jeu de données

Les utilisateurs doivent passer par cette interface pour le chargement du jeu de données pour les autres étapes. Elle est illustrée dans la figure suivante.



(a) Lecture du jeu de données



	DATE	TIME	EPOCH	MOTEID	TEMPERATURE	HUMIDITY	LIGHT	VOLTAGE
1	2004-03-31	03:38:15.757551	2	1.0	122.15299999999999	-3.91909999999999...	11.04	2.03397
2	2004-02-28	00:59:16.02785	3	1.0	19.9884	37.0933	45.08	2.69964
3	2004-02-28	01:03:16.33393	11	1.0	19.3024	38.4629	45.08	2.68742
4	2004-02-28	01:06:16.013453	17	1.0	19.1652	38.8039	45.08	2.68742
5	2004-02-28	01:06:46.778088	18	1.0	19.175	38.8379	45.08	2.69964
6	2004-02-28	01:08:45.992524	22	1.0	19.1456	38.9401	45.08	2.68742
7	2004-02-28	01:09:22.323858	23	1.0	19.1652	38.872	45.08	2.68742
8	2004-02-28	01:09:46.109598	24	1.0	19.1652	38.8039	45.08	2.68742
9	2004-02-28	01:10:16.6789	25	1.0	19.1456	38.8379	45.08	2.69964
10	2004-02-28	01:10:46.250524	26	1.0	19.1456	38.872	45.08	2.68742
11	2004-02-28	01:11:46.941288	28	1.0	19.1456	38.9401	45.08	2.69964
12	2004-02-28	01:12:46.251377	30	1.0	19.1358	38.9061	45.08	2.68742
13	2004-02-28	01:14:16.63127	33	1.0	19.1162	38.8039	45.08	2.69964
14	2004-02-28	01:14:46.569352	34	1.0	19.1162	38.872	45.08	2.69964
15	2004-02-28	01:15:16.649556	35	1.0	19.1064	39.0082	45.08	2.69964
16	2004-02-28	01:16:16.343708	37	1.0	19.1064	38.872	43.24	2.69964
17	2004-02-28	01:16:46.508622	38	1.0	19.0966	38.8039	43.24	2.69964

(b) Affichage avec analyse statistique

FIGURE 4.5 – Lecture et Affichage du jeu de données

4.4.2 Génération de données manquantes

Elle est la fenêtre sur laquelle les données manquantes sont générées en les marquant en rouge, elle affiche aussi les capteurs concernés par les données manquantes. Elle est illustrée dans la figure suivante.

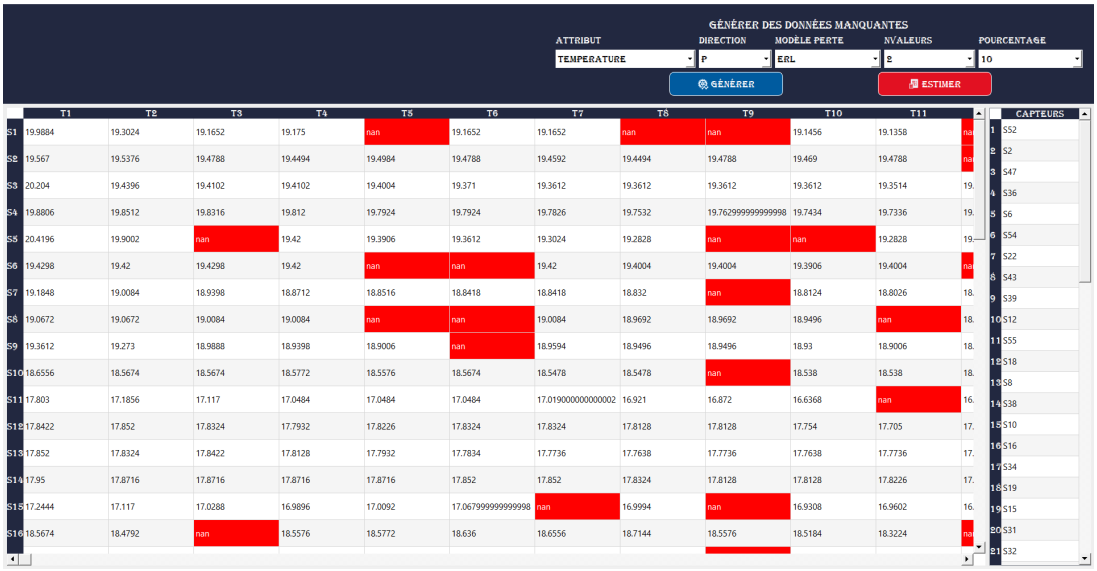


FIGURE 4.6 – Fonctionnalité de génération de données manquantes

4.4.3 Estimation de données manquantes générées

Elle est la fenêtre qui affiche les résultats d'estimation avec les paramètres utilisés pour générer les données manquantes et est illustrée dans la figure suivante.

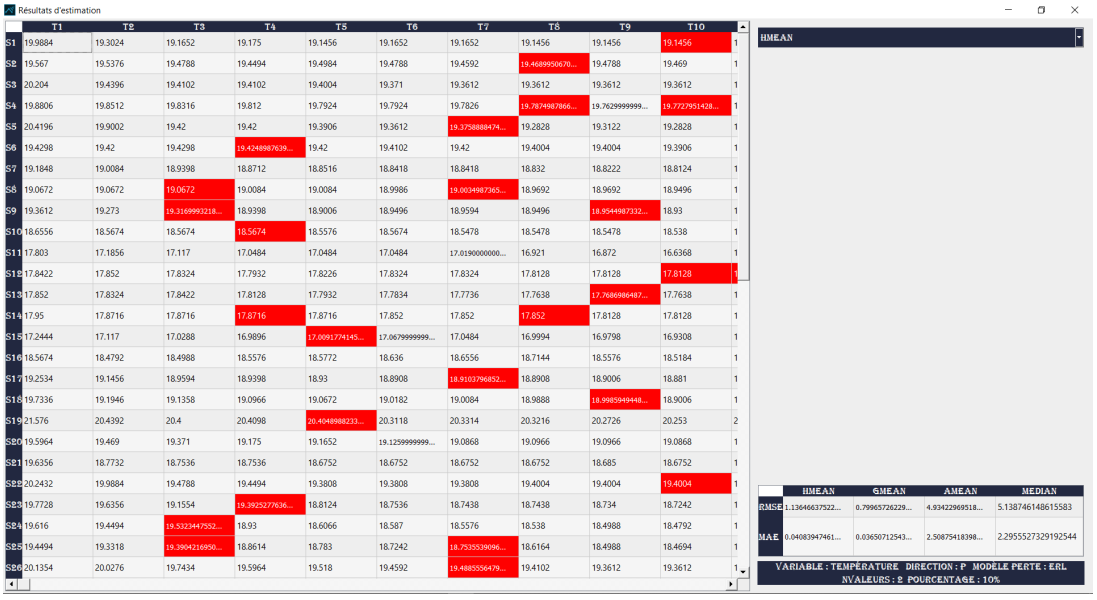


FIGURE 4.7 – Résultats d'estimation

4.5 Résultats et discussions

L'ensemble des tests ont étés effectuer sous l'environnement Python de version 3.9.4, sur une machine possédant une mémoire RAM de 8GO et un CPU I5. Les méthodes d'estimations ont étés évaluées en générant des données manquantes artificielles suivant les modèles de pertes de données des RCSFs. Ces taux varient entre 10 à 50%, 10 à 40% et 5 à 30% respectivement pour ERL et ESRL, BRL, EFRL. Ensuite nos deux approches proposées estiment les données manquantes avec les n précédentes valeurs non manquantes, enregistrées par le capteur S_{miss} avant T_{miss} ou avec les n valeurs suivantes non manquantes enregistrées par le capteur S_{miss} après T_{miss}. Le n variant de 2 à 10 afin d'obtenir un meilleur résultat d'estimation.

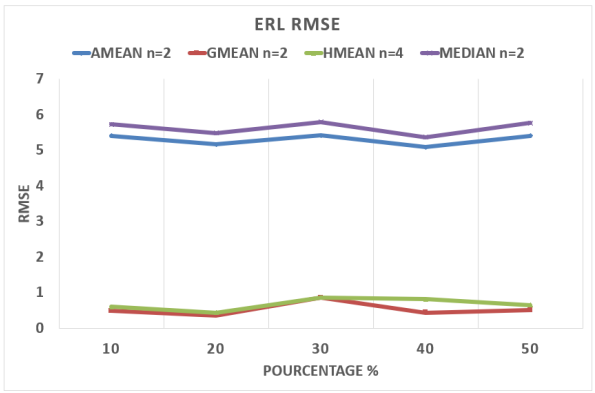
Les performances de nos approches ont été évaluée en utilisant les métriques RMSE et MAE, et comparer ces résultats aux résultats d'autres méthodes existantes AMEAN, MEDIAN. Nous avons fourni 8 graphiques par modèle de perte avec un total de 32 graphiques sur lesquelles le pourcentage de données manquantes est représenté sur l'axe X, et sur l'axe Y est représentée les valeurs de précision des méthodes d'estimations en une certaine valeur de n.

ERL

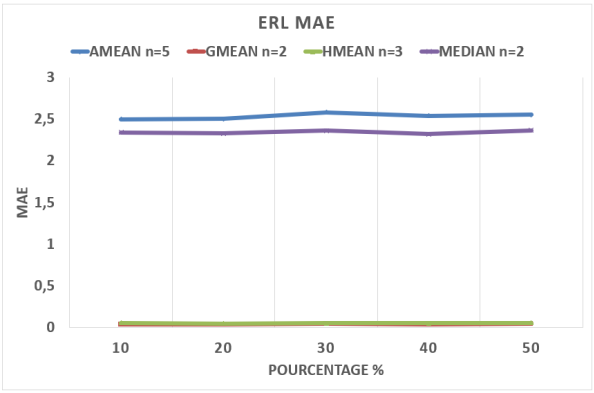
Les résultats obtenu par les méthodes d'estimation suivant ce modèle sont renseignés dans les tableaux de cette section en marquant les meilleurs résultats des méthodes par un*, ces résultats sont aussi représentés graphiquement sur les figures de cette section.

	Estimation avec les valeurs précédentes sur la température							
	HMEAN		GMEAN		AMEAN		MEDIAN	
Taux de pertes	RMSE n=4	MAE n=3	RMSE n=2	MAE n=2	RMSE n=2	MAE n=5	RMSE n=2	MAE n=2
10	0,6013	0,0525	0,4802	**0,0314***	5,4130	**2,5007***	5,7282	2,3440
20	**0,4361***	**0,0436***	**0,3564***	0,0319	5,1717	2,5064	5,4916	2,3296
30	0,8604	0,0502	0,8510	0,0426	5,4186	2,5832	5,7945	2,3634
40	0,8142	0,0475	0,4362	0,0378	**5,0943***	2,5430	**5,3717***	**2,3210***
50	0,6504	0,0515	0,5016	0,0405	5,4125	2,5559	5,7719	2,3675

TABLE 4.1 – Résultats d'estimation avec les valeurs précédentes sur la température ERL



(a) RMSE

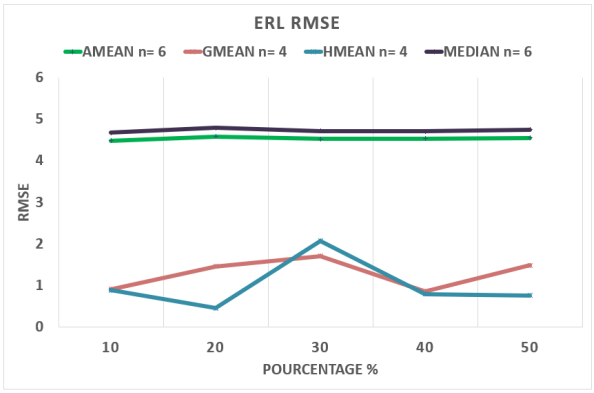


(b) MAE

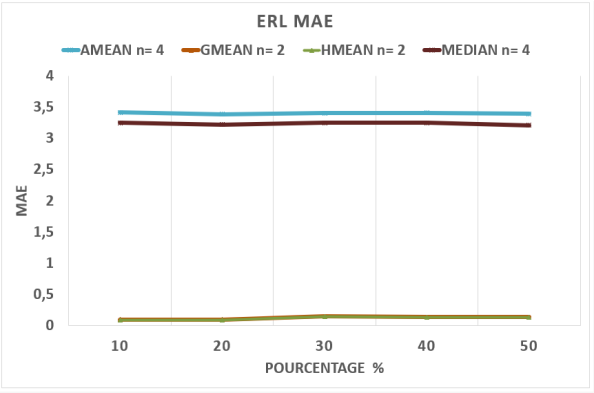
FIGURE 4.8 – Estimation avec les valeurs précédentes sur la température ERL

	Estimation avec les valeurs précédentes sur l'humidité							
	HMEAN		GMEAN		AMEAN		MEDIAN	
Taux de pertes	RMSE n=4	MAE n=2	RMSE n=6	MAE n=2	RMSE n=6	MAE n=4	RMSE n=2	MAE n=4
10	0,8803	**0,0828***	1,5090	**0,0919***	**4,4885***	3,4197	**4,6762***	3,2522
20	**0,4461***	0,0844	1,2340	0,0949	4,5915	**3,3917***	4,7970	3,2266
30	2,0727	0,1362	1,4372	0,1443	4,5272	3,4065	4,7189	3,2532
40	0,7910	0,1273	1,1760	0,1347	4,5384	3,4128	4,7115	3,2524
50	0,7583	0,1268	**0,9925***	0,1327	4,5544	3,3967	4,7555	**3,2162***

TABLE 4.2 – Résultats d'estimation avec les valeurs précédentes sur l'humidité ERL



(a) RMSE

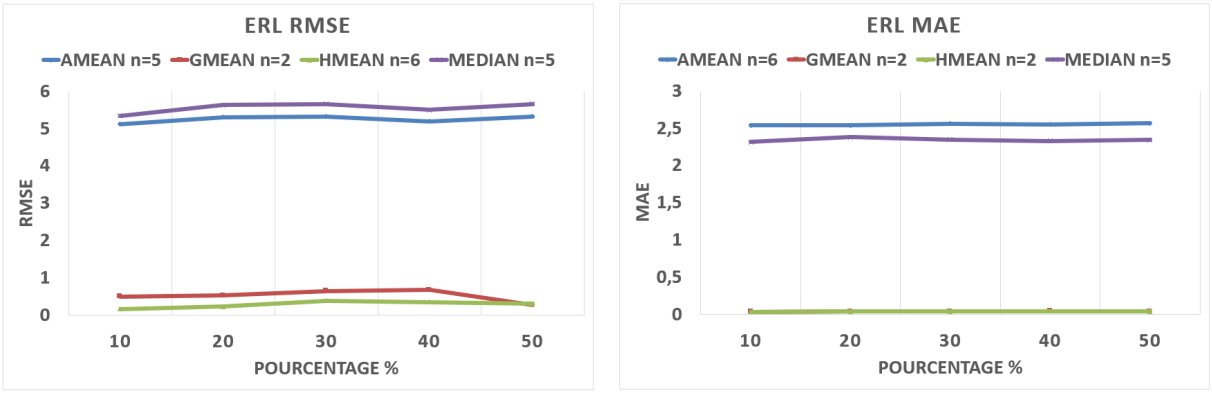


(b) MAE

FIGURE 4.9 – Estimation avec les valeurs précédentes sur l'humidité

	Estimation avec les valeurs suivantes sur la température							
	HMEAN		GMEAN		AMEAN		MEDIAN	
Taux de pertes	RMSE n=6	MAE n=2	RMSE n=2	MAE n=2	RMSE n=5	MAE n=6	RMSE n=5	MAE n=5
10	**0,1612***	**0,0358***	0,4945	**0,0321***	**5,1183***	2,5325	**5,3410***	**2,3106***
20	0,2211	0,0376	0,52	0,0347	5,2994	**2,5322***	5,6282	2,3777
30	0,3803	0,0365	0,6406	0,0369	5,3212	2,5577	5,6510	2,3446
40	0,3393	0,0402	0,6680	0,0399	5,1857	2,5477	5,4930	2,3271
50	0,2984	0,03653	**0,2692***	0,0365	5,3237	2,5603	5,6526	2,3387

TABLE 4.3 – Résultats d'estimation avec les valeurs suivantes sur la température ERL

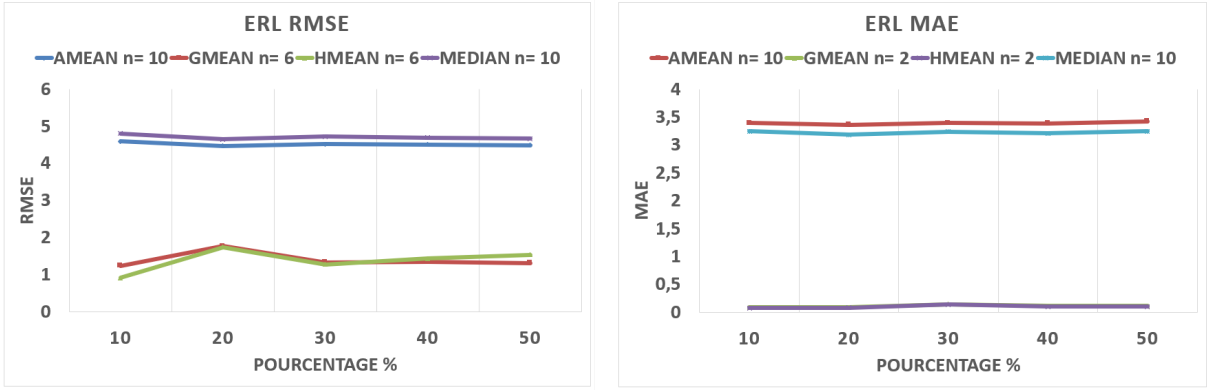


(a) RMSE (b) MAE

FIGURE 4.10 – Estimation avec les valeurs suivantes sur la température

	Estimation avec les valeurs suivantes sur l'humidité							
	HMEAN		GMEAN		AMEAN		MEDIAN	
Taux de pertes	RMSE n=6	MAE n=2	RMSE n=6	MAE n=2	RMSE n=10	MAE n=10	RMSE n=10	MAE n=10
10	**0,9120***	**0,0751***	**1,2384***	**0,0832***	4,5847	3,3932	4,8026	3,2454
20	1,7311	0,0807	1,7662	0,0878	**4,4610***	**3,3613***	4,6417	**3,1808***
30	1,2747	0,1358	1,3287	0,1421	4,5091	3,3934	4,7176	3,2352
40	1,4251	0,0967	1,3491	0,1123	4,5020	3,3823	4,69	3,2080
50	1,5319	0,1003	1,3035	0,1117	4,4844	3,4208	**4,6584***	3,24

TABLE 4.4 – Résultats d'estimation avec les valeurs suivantes sur l'humidité ERL



(a) RMSE (b) MAE

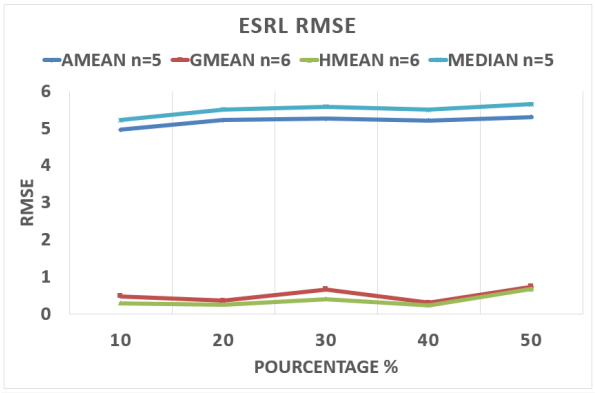
FIGURE 4.11 – Estimation avec les valeurs suivantes sur l'humidité

ESRL

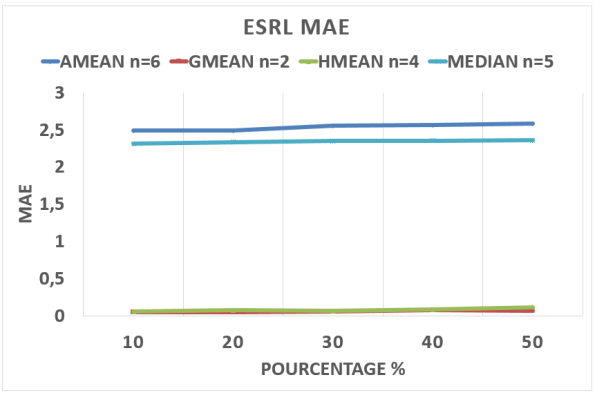
Reportés sur les figures de cette section les résultats d'estimation obtenu par les différentes méthodes dont les valeurs sont renseignées dans les tableaux de cette section en marquant les meilleurs résultats de chaque approche par un *.

	Estimation avec les valeurs précédentes sur température							
	HMEAN		GMEAN		AMEAN		MEDIAN	
Taux de pertes	RMSE n=6	MAE n=4	RMSE n=6	MAE n=2	RMSE n=5	MAE n=6	RMSE n=5	MAE n=5
10	0,2905	**0,0546***	0,4722	**0,0473***	**4,9576***	2,4865	**5,2213***	**2,31***
20	0,2459	0,0737	0,3543	0,0480	5,2185	**2,4864***	5,5042	2,3367
30	0,4023	0,0633	0,6612	0,0571	5,2556	2,5514	5,5753	2,3471
40	**0,2324***	0,0806	**0,2947**	0,0754	5,2077	2,5619	5,4990	2,3519
50	0,6646	0,1140	0,7349	0,0637	5,3037	2,5790	5,6414	2,3592

TABLE 4.5 – Résultats d'estimation avec les valeurs précédentes sur température ESRL



(a) RMSE

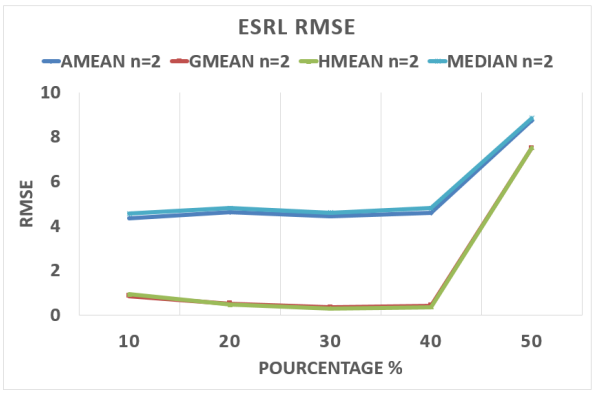


(b) MAE

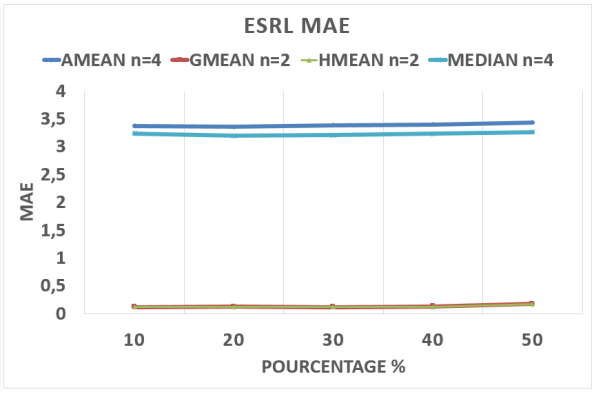
FIGURE 4.12 – Estimation avec les valeurs précédentes sur la température

	Estimation avec les valeurs précédentes sur humidité							
	HMEAN		GMEAN		AMEAN		MEDIAN	
Taux de pertes	RMSE n=2	MAE n=2	RMSE n=2	MAE n=2	RMSE n=2	MAE n=4	RMSE n=2	MAE n=4
10	0,9241	0,1168	0,8475	**0,1183***	**4,3472***	3,3715	**4,5418***	3,2388
20	0,4698	0,1165	0,5106	0,1215	4,6079	**3,3552***	4,7926	**3,1938***
30	**0,2923***	**0,1156***	**0,3312***	0,1187	4,4408	3,3864	4,5919	3,2068
40	0,3573	0,1247	0,4093	0,1301	4,5829	3,3999	4,7976	3,2275
50	7,4993	0,1717	7,4929	0,1746	8,7333	3,4324	8,8278	3,2623

TABLE 4.6 – Résultats d'estimation avec les valeurs précédentes sur humidité ESRL



(a) RMSE

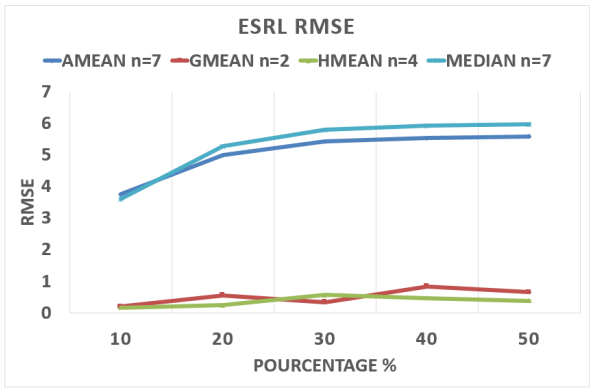


(b) MAE

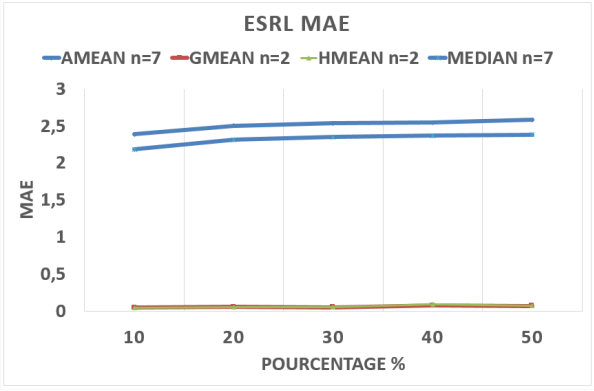
FIGURE 4.13 – Estimation avec les valeurs précédentes sur l'humidité

	Estimation avec les valeurs suivantes sur la température							
	HMEAN		GMEAN		AMEAN		MEDIAN	
Taux de pertes	RMSE n=4	MAE n=2	RMSE n=2	MAE n=2	RMSE n=7	MAE n=7	RMSE n=7	MAE n=7
10	**0,1603***	**0,0477***	**0,1983***	**0,0477***	**3,75***	**2,3840***	**3,5945***	**2,1834***
20	0,2344	0,0530	0,5483	0,0554	4,9817	2,5001	5,2589	2,3110
30	0,5668	0,0524	0,3350	0,0524	5,4192	2,5332	5,7874	2,3476
40	0,4692	0,0892	0,8378	0,0718	5,5204	2,5456	5,9139	2,3656
50	0,3820	0,0684	0,6485	0,0684	5,5814	2,5813	5,9568	2,3822

TABLE 4.7 – Résultats d'estimation avec les valeurs suivantes sur la température ESRL



(a) RMSE

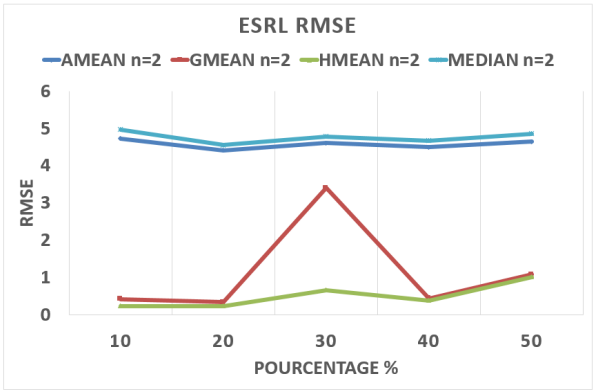


(b) MAE

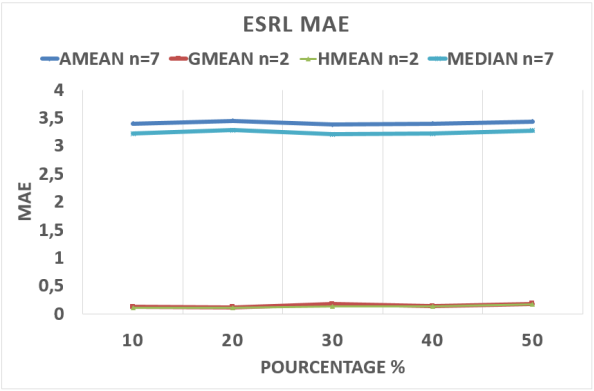
FIGURE 4.14 – Estimation avec les valeurs suivantes sur la température

	Estimation avec les valeurs suivantes sur humidité							
	HMEAN		GMEAN		AMEAN		MEDIAN	
Taux de pertes	RMSE n=2	MAE n=2	RMSE n=2	MAE n=2	RMSE n=2	MAE n=7	RMSE n=2	MAE n=7
10	**0,2314***	**0,1078***	0,4169	0,1228	4,7230	3,3986	4,9702	3,2163
20	0,2320	0,1104	**0,3330***	**0,1175***	**4,4077***	3,4453	**4,5530***	3,2816
30	0,6588	0,1310	3,4002	0,1749	4,6011	**3,3787***	4,7790	**3,2068***
40	0,3815	0,1265	0,4387	0,1352	4,4934	3,4002	4,6720	3,2173
50	1,0086	0,1645	1,0740	0,1774	4,6442	3,4323	4,8570	3,2734

TABLE 4.8 – Résultats d'estimation avec les valeurs suivantes sur humidité ESRL



(a) RMSE



(b) MAE

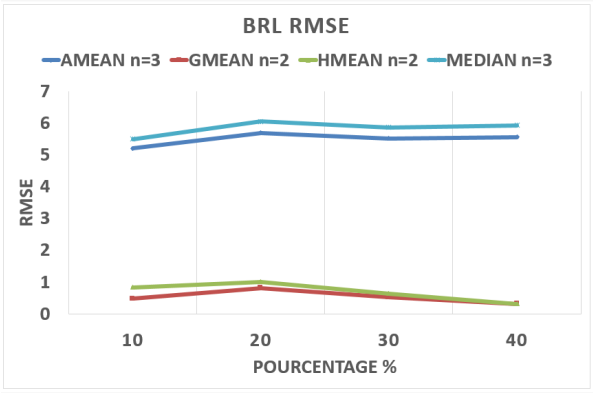
FIGURE 4.15 – Estimation avec les valeurs suivantes sur l'humidité

BRL

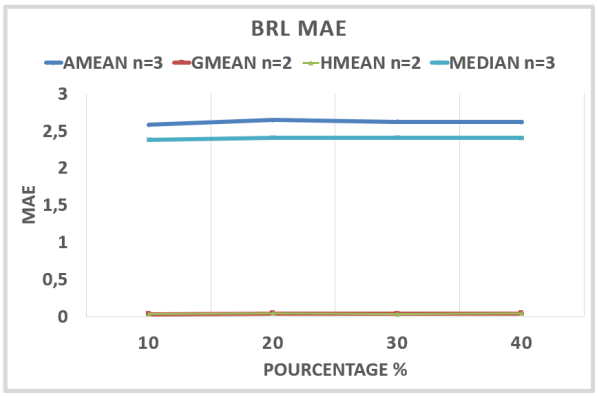
Les tableaux de cette section contiennent les résultats d'estimation des différentes approches en marquant le meilleur résultat pour chaque méthode par un *, et les figures de cette section donnent un aperçu graphique de ces résultats.

	Estimation avec les valeurs précédentes sur température							
	HMEAN		GMEAN		AMEAN		MEDIAN	
Taux de pertes	RMSE n=2	MAE n=2	RMSE n=2	MAE n=2	RMSE n=3	MAE n=3	RMSE n=3	MAE n=3
10	0,8285	**0,0341***	0,4736	**0,0310***	**5,1975***	**2,5815***	**5,4894***	**2,3796***
20	1,0071	0,0428	0,8146	0,0409	5,6884	2,6437	6,0441	2,4091
30	0,6417	0,0356	0,5233	0,0346	5,5099	2,6210	5,8620	2,4091
40	**0,3090***	0,0391	**0,3088***	0,0391	5,5584	2,6162	5,9232	2,4058

TABLE 4.9 – Résultats d'estimation avec les valeurs précédentes sur température BRL



(a) RMSE

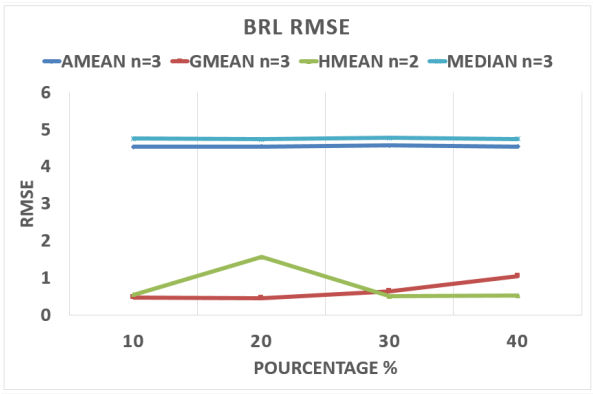


(b) MAE

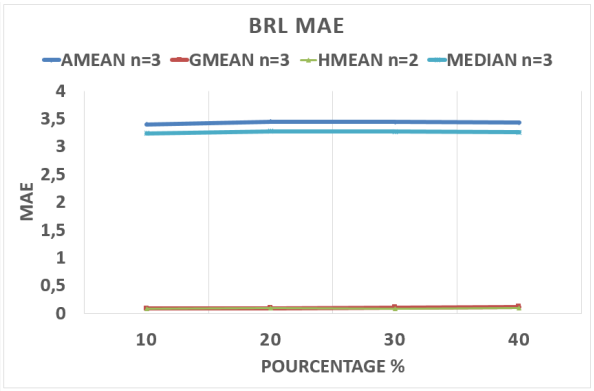
FIGURE 4.16 – Estimation avec les valeurs précédentes sur la température

	Estimation avec les valeurs précédentes sur l'humidité							
	HMEAN		GMEAN		AMEAN		MEDIAN	
Taux de pertes	RMSE n=2	MAE n=2	RMSE n=3	MAE n=3	RMSE n=3	MAE n=3	RMSE n=3	MAE n=3
10	0,5462	**0,0789***	0,4767	**0,0868***	**4,5332***	**3,3992***	4,7518	**3,2342***
20	1,5599	0,0955	**0,4550***	0,0899	4,5364	3,4451	**4,7371***	3,2719
30	**0,5033***	0,0869	0,6365	0,1016	4,5669	3,4440	4,7694	3,2704
40	0,5160	0,0963	1,0512	0,1145	4,5389	3,4261	4,7446	3,2569

TABLE 4.10 – Résultats d'estimation avec les valeurs précédentes sur l'humidité BRL



(a) RMSE

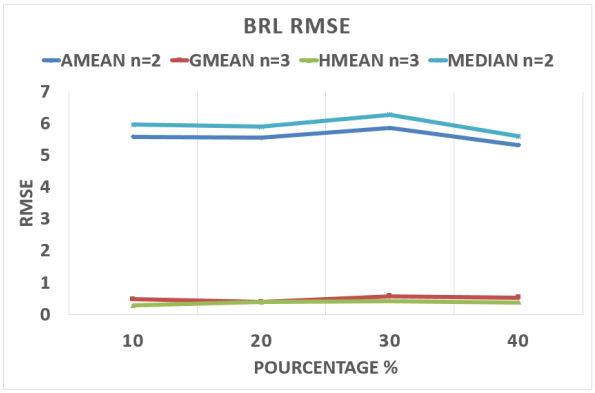


(b) MAE

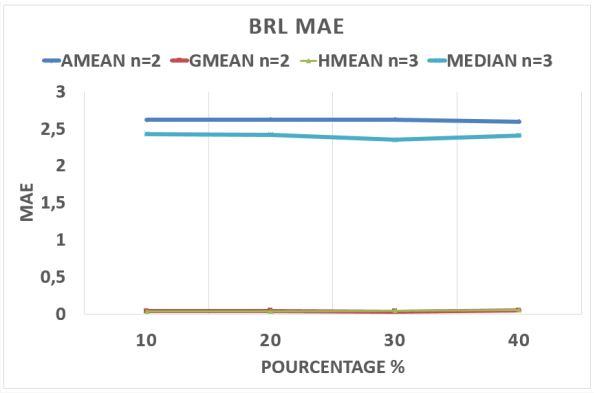
FIGURE 4.17 – Estimation avec les valeurs précédentes sur l'humidité

	Estimation avec les valeurs suivantes sur température							
	HMEAN		GMEAN		AMEAN		MEDIAN	
Taux de pertes	RMSE n=3	MAE n=3	RMSE n=3	MAE n=2	RMSE n=2	MAE n=2	RMSE n=2	MAE n=3
10	**0,2823***	**0,035***	0,4764	0,0350	5,5752	2,6180	5,9608	2,4294
20	0,39	0,0376	**0,3901***	0,0403	5,5503	2,6216	5,8876	2,4161
30	0,4191	0,0432	0,5676	**0,0340***	5,85	2,6214	6,2754	**2,3479***
40	0,3743	0,0496	0,5306	0,0469	**5,3126***	**2,5969***	**5,5922***	2,4073

TABLE 4.11 – Résultats d'estimation avec les valeurs suivantes sur température BRL



(a) RMSE

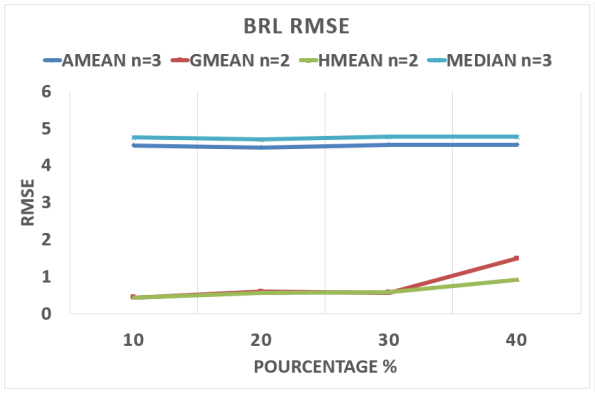


(b) MAE

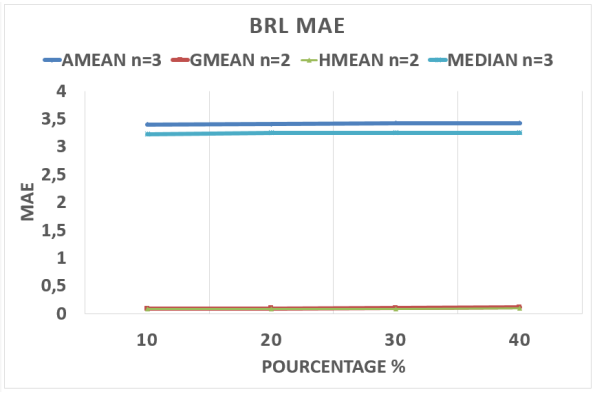
FIGURE 4.18 – Estimation avec les valeurs suivantes sur la température

	Estimation avec les valeurs suivantes sur l'humidité							
	HMEAN		GMEAN		AMEAN		MEDIAN	
Taux de pertes	RMSE n=2	MAE n=2	RMSE n=2	MAE n=2	RMSE n=3	MAE n=3	RMSE n=3	MAE n=3
10	**0,4320***	**0,0766***	**0,4391***	**0,0830***	4,5405	**3,3984***	4,7562	**3,2267***
20	0,5620	0,0856	0,5961	0,0948	**4,4827***	3,4053	**4,6940***	3,2475
30	0,5830	0,0886	0,5634	0,0966	4,5583	3,4202	4,7753	3,2496
40	0,9203	0,0991	1,4840	0,1095	4,5599	3,4201	4,7743	3,2462

TABLE 4.12 – Résultats d'estimation avec les valeurs suivantes sur l'humidité BRL



(a) RMSE



(b) MAE

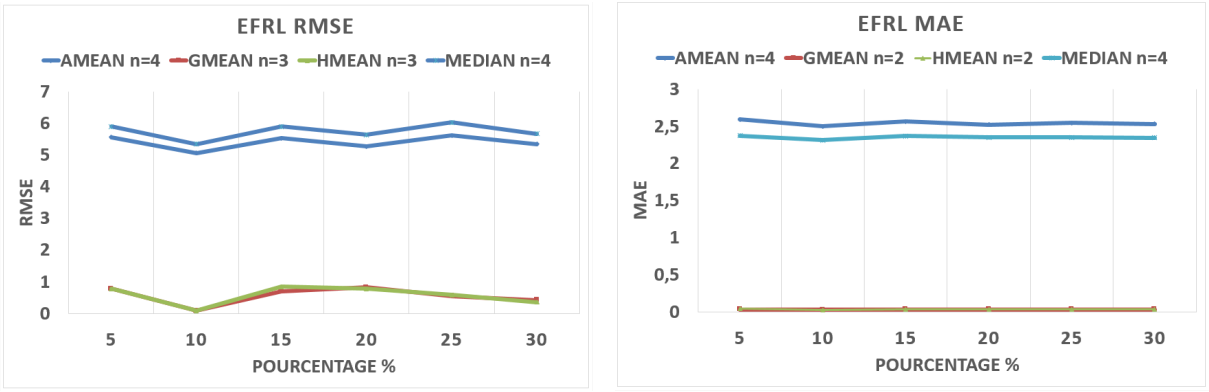
FIGURE 4.19 – Estimation avec les valeurs suivantes sur l'humidité

EFRL

Les résultats obtenu par les méthodes d'estimation dans nos évaluation sont renseignés dans les tableaux de cette section en marquant les meilleurs résultats des méthodes par un*, ces résultats sont aussi représentés graphiquement sur les figures de cette section.

	Estimation avec les valeurs précédentes sur température							
	HMEAN		GMEAN		AMEAN		MEDIAN	
Taux de pertes	RMSE n=3	MAE n=2	RMSE n=3	MAE n=2	RMSE n=4	MAE n=4	RMSE n=4	MAE n=4
5	0,7798	0,0396	0,7798	0,0321	5,5552	2,5917	5,8926	2,3720
10	**0,0831***	**0,0263***	**0,0831***	**0,0263***	**5,0529***	**2,4956***	**5,3442***	**2,3128***
15	0,8383	0,0317	0,6929	0,0316	5,5188	2,5632	5,8876	2,3676
20	0,7746	0,0338	0,8190	0,0337	5,2599	2,5188	5,6286	2,3491
25	0,5791	0,0328	0,5512	0,0309	5,6053	2,5423	6,0270	2,3504
30	0,3572	0,0311	0,4205	0,0307	5,3355	2,5299	5,6614	2,3449

TABLE 4.13 – Résultats d'estimation avec les valeurs précédentes sur température EFRL

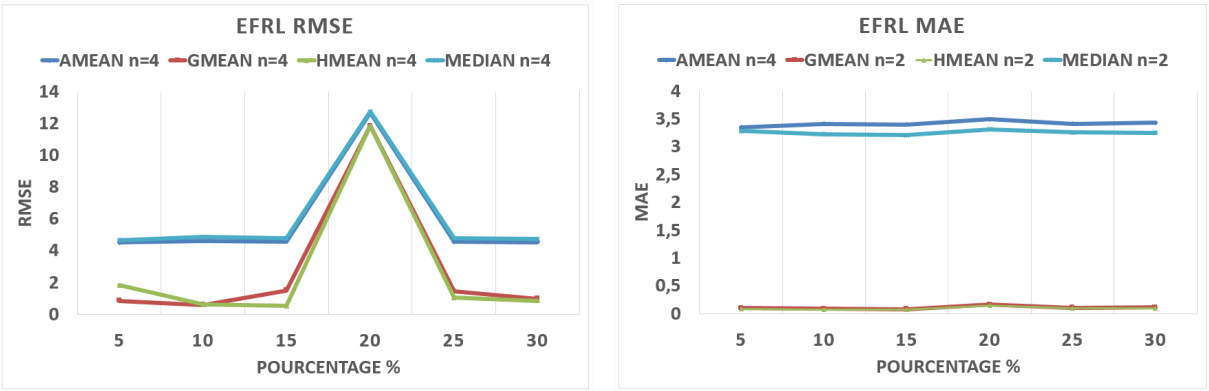


(a) RMSE (b) MAE

FIGURE 4.20 – Estimation avec les valeurs précédentes sur la température

	Estimation avec les valeurs précédentes sur l'humidité							
	HMEAN		GMEAN		AMEAN		MEDIAN	
Taux de pertes	RMSE n=4	MAE n=2	RMSE n=4	MAE n=2	RMSE n=4	MAE n=4	RMSE n=4	MAE n=2
5	1,8273	0,0814	0,8467	0,0976	**4,4965***	**3,3491***	**4,6603***	3,2817
10	0,5987	0,0764	**0,5745***	0,0877	4,6119	3,4125	4,8407	3,2271
15	**0,5193***	**0,0733***	1,4849	**0,0796***	4,5586	3,3989	4,7641	**3,2055***
20	11,8397	0,1485	11,8264	0,1586	12,6602	3,4897	12,7416	3,3123
25	1,0341	0,0949	1,4498	0,1059	4,5483	3,4125	4,7531	3,2566
30	0,8485	0,0983	0,9676	0,1080	4,5338	3,4284	4,7262	3,2504

TABLE 4.14 – Résultats d'estimation avec les valeurs précédentes sur l'humidité EFRL

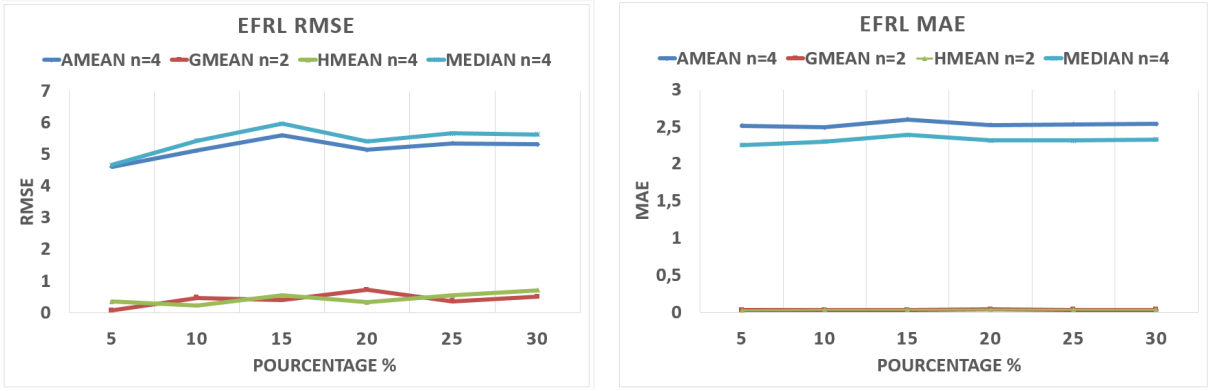


(a) RMSE (b) MAE

FIGURE 4.21 – Estimation avec les valeurs précédentes sur l'humidité

	Estimation avec les valeurs suivantes sur température							
	HMEAN		GMEAN		AMEAN		MEDIAN	
Taux de pertes	RMSE n=4	MAE n=2	RMSE n=2	MAE n=2	RMSE n=4	MAE n=4	RMSE n=4	MAE n=4
5	0,3412	**0,0250***	**0,0603***	**0,0250***	**4,6025***	2,5112	**4,6555***	**2,2506***
10	**0,2280***	0,0323	0,4668	0,0294	5,1154	**2,4858***	5,4275	2,2966
15	0,54	0,0311	0,3966	0,0290	5,5940	2,5946	5,9559	2,3848
20	0,3244	0,0378	0,7113	0,0361	5,1298	2,5226	5,4022	2,3127
25	0,5446	0,0315	0,3602	0,0299	5,3367	2,5286	5,6509	2,3112
30	0,7045	0,0332	0,4981	0,0311	5,3054	2,5329	5,6228	2,3263

TABLE 4.15 – Résultats d'estimation avec les valeurs suivantes sur température EFRL

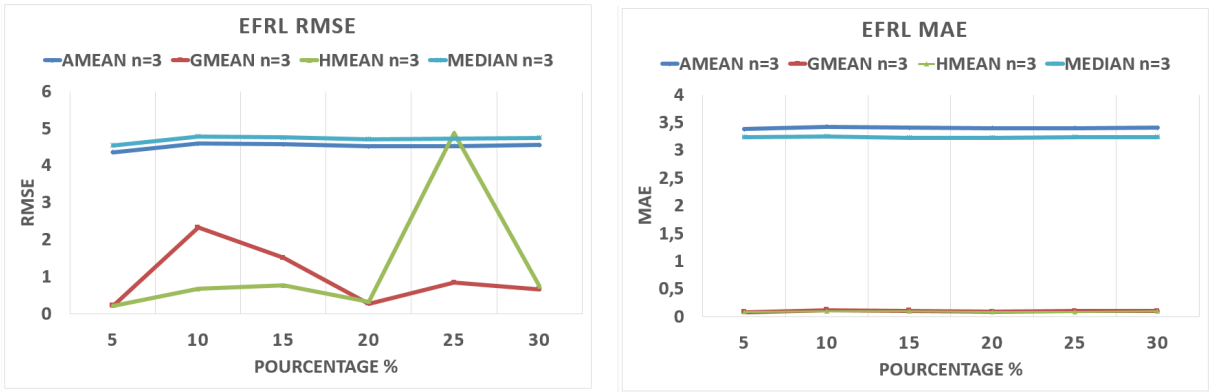


(a) RMSE (b) MAE

FIGURE 4.22 – Estimation avec les valeurs suivantes sur la température

	Estimation avec les valeurs suivantes sur l'humidité							
	HMEAN		GMEAN		AMEAN		MEDIAN	
Taux de pertes	RMSE n=3	MAE n=3	RMSE n=2	MAE n=3	RMSE n=3	MAE n=3	RMSE n=3	MAE n=3
5	**0,2090***	**0,0788***	**0,2087***	**0,0789***	**4,3542***	**3,3770***	**4,5367***	3,2338
10	0,6685	0,0937	2,3227	0,1173	4,5822	3,4244	4,7764	3,2522
15	0,7592	0,0962	1,4989	0,1043	4,5642	3,4052	4,7518	**3,2240***
20	0,3282	0,0863	0,2715	0,0853	4,5152	3,3961	4,7044	3,2255
25	4,8721	0,0888	0,8373	0,0964	4,5242	3,4007	4,7140	3,2339
30	0,7373	0,0912	0,6595	0,1008	4,5522	3,4066	4,7440	3,2339

TABLE 4.16 – Résultats d'estimation avec les valeurs suivantes sur l'humidité EFRL



(a) RMSE (b) MAE

FIGURE 4.23 – Estimation avec les valeurs suivantes sur l'humidité

4.6 Interprétation des résultats

Les résultats obtenus dans les expérimentations par nos algorithmes montrent une meilleure précision comparée aux autres algorithmes d'estimations. Dans les sections suivante, une comparaison entre nos approches d'estimation montrant laquelle est plus meilleure mais aussi à quelle valeur de n sera faite.

4.6.1 Schéma de perte ERL

Suivant ce modèle, dans nos résultats obtenu nous comptons que HMEAN a été 6 fois meilleure que GMEAN, 3 fois avec les valeurs précédentes avec $n = 2, 3, 4$, et 3 fois avec les valeurs suivantes en $n = \{2, 6\}$. GMEAN n'a été que 2 fois meilleure, une en estimation avec les valeurs précédentes et une en estimation avec les valeurs suivantes le tout en $n = 2$. De par ces résultats HMEAN est efficace en estimation que GMEAN sur ce modèle de perte. Il serait donc intéressant de l'utiliser dans les estimations(précédentes ou suivantes) avec des valeurs de $n = \{2, 4, 6\}$.

4.6.2 Schéma de perte ESRL

Suite aux résultats obtenus dans les expérimentations l'approche HMEAN devance de loin GMEAN en termes d'efficacité après avoir été meilleur qu'elle 3 fois en estimation avec les valeurs précédentes avec $n = \{6, 2\}$, et 3 fois avec les valeurs suivantes avec $n = \{4, 2\}$. GMEAN a enregistré un score supérieur à 2 fois avec les valeurs précédentes et une autre en estimation avec les valeurs suivantes le tout en $n = 2$. Ainsi HMEAN est la meilleure approche en estimation suivant ce modèle avec $n = \{2, 6\}$ pour l'estimation avec les valeurs précédentes, et $n = \{4, 2\}$ pour l'estimation avec les valeurs suivantes.

4.6.3 Schéma de perte BRL

Suivant ce modèle, les approches HMEAN et GMEAN sont meilleures avec un score similaire (4 fois). GMEAN a fourni 3 en estimation avec les valeurs précédentes et 1 en estimation avec les valeurs suivantes avec $n = \{2, 3\}$. HMEAN en a fourni 3 en estimation avec les valeurs suivantes et 1 avec les valeurs précédentes en $n = \{2, 3\}$. De par ces résultats, HMEAN serait efficace pour l'estimation avec les valeurs suivantes avec $n = \{2, 3\}$ sur ce modèle. Pour

l'estimation avec les valeurs précédentes il serait bien d'utiliser GMEAN avec $n = \{2, 3\}$.

4.6.4 Schéma de perte EFRL

Comme dans le modèle BRL, les deux approches ont fourni les mêmes valeurs de performance soit 3 en estimation avec les valeurs précédentes avec $n = \{2, 3, 4\}$ et 1 en estimation avec les valeurs suivantes avec $n = 3$ pour HMEAN. GMEAN en a fourni 3 en estimation avec les valeurs suivantes pour $n = \{2, 3\}$ et 1 en estimation avec les valeurs précédentes pour $n = 2$. Il serait bien d'utiliser HMEAN pour l'estimation avec les valeurs précédentes, car elle est efficace dans ce sens avec les valeurs de $n = \{2, 3, 4\}$. GMEAN aussi serait bien à utiliser pour l'estimation avec les valeurs suivantes avec $n = \{2, 3\}$.

En somme pour les projets d'estimation il serait bien d'utiliser l'approche HMEAN avec les petites valeurs de n, car plus n est grand plus la plage d'erreur de la méthode devient grande dû aux variations des n valeurs utilisées pour l'estimation.

4.7 Conclusion

Dans ce chapitre, nos approches proposées ont étés évaluer sur le dataset d'Intel Berkely par rapport à d'autres méthodes existantes suivant les modèles de perte de données que font face les RCFs. Les résultats des tests ont démontré une meilleure performance de nos approches par rapport à ceux dont elles ont été comparer. Elles offrent plusieurs résultats possible par rapport aux autres méthodes ainsi d'imputer la valeur manquante avec une valeur meilleure égale ou plus proche de la valeur manquante, ce qui fait la force de ces méthodes. En fonction des données non manquantes elles offrent un meilleur résultat.

CONCLUSION GÉNÉRALE ET PERSPECTIVES

Les RCSFs sont devenus un axe de recherche incontournable, une technologie en constante évolution avec leurs diversités de pouvoir être appliqués dans différents domaines. Ces perceptives d'applications peuvent être d'ordre médical, de surveillance. Ils permettent de collecter des données pour des fins d'exploitation. Ces données collectées sont d'une importance capitale, car c'est par leurs analyses qu'une décision sera prise et éventuellement par la suite agir sur l'environnement où ils ont été collectés. Bien que cette technologie soit en constante évolution, elles comportent encore plusieurs problèmes à résoudre comme la présence de données manquantes dans les informations collectées. Ces pertes sont dûs a un ou des dysfonctionnement(s) dans les processus d'acquisition de ces données. Ainsi les analyses et les prises de décisions effectuées sur ces données peuvent ne pas être claires, voire même conduire à des erreurs. Pour remédier à cela, la mise en place de méthodes permettant de reconstruire ces données manquantes était primordiale. Ce qui était l'objet de notre travail.

Dans une première partie, nous nous sommes concentrés sur les RCSFs afin de bien comprendre ce qu'est cette technologie, une seconde partie portant sur le cœur du sujet dans laquelle nous avons discuté de différentes notions sur les données manquantes en générale et dans le cadre des RCSFs, mais aussi des méthodes d'estimation existantes. Certaines d'eux n'étaient pas adaptées aux RCSFs ou étaient trop complexes. Pour résoudre ce problème nous avons mis en place deux méthodes qui ont étés évaluer sur le jeu de données d'Intel Berkeley par rapport à d'autres approches existantes en générant des données manquantes artificielles suivant les modèles de perte des RCSFs. Les résultats obtenus ont étés prometteurs et satisfaisants.

Comme perspectives nous projetons d'améliorer nos travaux, en mettant en place une méthode d'estimation de données manquantes dans les RCSFs se basant sur la corrélation qui existe entre les attributs du jeu de données, de faire un hybride de cette méthode avec nos approches déjà proposées.

BIBLIOGRAPHIE

Livres

[1] Ian F. Akyildiz, Mehmet Can Vuran, *Wireless Sensor Networks, Georgia Institute of Technology, University of Nebraska-Lincoln, USA, 2010*, 10-310

[2] Akyildiz, Ian F, Metmet Can Vuran, *Wireless Sensor Networks. Vol 4. John Wiley & Sons 2010*

[3] Madden, Samuel R, al, *"TinyDB : An acquisitional query processing system form sensor networks." ACM transactions on database system(TODS) 30.1, 2005* , 122-173

[4] Boulis, Athanassios, et al, *"SensorWare : Programming sensor networks beyond code update and querying." Pervasive and mobile computing 3.4, 2007*, 386-412

[5] Levis, Philip, et al, *"TinyOs : An operating system for sensor networks." Ambient intelligence. Springer, Berlin, Heidelberg, 2007*, 115-148

[6] Bhatti, Shah, et al, *"MANTIS OS : An embedded multithreaded operating system for wireless micro sensor platforms." Mobile Networks and Applications 10.4, 2005.*

[7] Cao, Qing, and Tarek Abdelzaher, *"LiteOs : a lightweight operating system for C++ software development in sensor networks." Proceedings of the 4th international conference on Embedded networked sensor systems., 2006*, 137-150

[8] Chu, Rui, et al, *"SenSmart : adaptive stack management for multitasking sensor networks." IEEE transactions on computers 62.1, 2011.*, 137-150

[9] Dong, Wei, et al, *"SenSpire Os : A predicable, flexible, and efficient operating system for wireless sensor networks." IEEE transactions on computers 60.12, 2011.*, 1788-1801

[10] Jennifer Yick, Biswanath Mukherjee, Dipak Ghosal, *Wireless sensor network survey, Department of Computer Science, University of California, Davis, CA 95616, United States, 2008*, 2293-2296

[11] Waltenegus Dargie, Christian Poellabauer, *FUNDAMENTALS OF WIRELESS SENSOR NETWORKS THEORY AND PRACTICE, Technical University of Dresden, Germany, University of Notre Dame, USA, 2010*

Articles de conférence

[12] Dunkels, Adam, Bjorn Gronvall, and Thiemo Voigt., *"Contiki a lightweight and flexible operating system for tiny networked sensors." 29th annual IEEE internationnal conference on local computer network. IEEE, 2004*

[13] Song, Wen-Miao, Yan-Ming Liu, and Shu-E. Zhang, *"Research on SMAC protocol for WSN", 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing. IEEE, 2008.*

[14] Buettner, Micheal, et al., *"X-MAC : a short preamble MAC protocol for duty-cycled wireless sensor networks.", Proceeding of the 4th international conference on Embedded networked sensor systems. 2006.*

[15] Juneja, Dimple, Sandhya Bansal, and Arora N, Gurpreet Kaur., *"Design and implementation of EAR algorithm for detectting routing attacks in WSN.", International Journal of Engineering Science and Technology 2.6, 2010.*, 1677-1683

[16] Han, Chih-Chieh, et al., *"A dynamic operating system for sensor nodes." Proceedings of the 3rd international conference on Mobile systems, applications, and services.*

[17] Yang, Ting, and Zhiqun Li., *"A 7-bit 26-MS/s SAR ADC in 0.18 μm CMOS process for WSN application." 2012 4th International High Speed Intelligent Communication.*

Articles

[18] Yao, Yong, and Johannes Gehrke, *"The cougar approach to in-network query processing in sensor networks." ACM Sigmod record 31.3, 2002*, 9-18

[19] Erich P. Stuntebeck (corresponding), Dario Pompili, Tommaso Melodia, *Wireless Underground Sensor Networks using Commodity Terrestrial Motes, Broadband and Wireless Networking Laboratory School of Electrical and Computer Engineering Georgia Institute of Technology Atlanta, GA 30332, 2008*, 111-114

[20] Ian F. Akyildiz, Dario Pompili, Tommaso Melodia, *Challenges for Efficient Communication in Underwater Acoustic Sensor Networks, Broadband and Wireless Networking Laboratory School of Electrical and Computer Engineering Georgia Institute of Technology, Atlanta, GA 30332*

[21] Mariam AlNuaimi, Farag Sallabi and Khaled Shuaib, *A Survey of Wireless Multimedia Sensor Networks Challenges and Solutions, Faculty of Information Technology United Arab Emirates University United Arab Emirates, International Conference on Innovations in Information Technology ,2011*

[22] Taner Cevik, Alex Gunagwera, Nazife Cevik, *A SURVEY OF M1,Department of Computer Engineering, Fatih University, Arel University Istanbul, Turkey, International Journal of Computer Networks and Communications (IJCNC) Vol.7, No.5, September 2015*

[23] YACINE CHALLAL, *Réseaux de Capteurs Sans Fils,2015*, 26-29

[24] Losci, V.,G. Morabito, and Salvatore Marano., *"A two-levels hieerarchy for low-energy adaptative clustering hierarchy(TL-LEACH)." IEEE vehicular technology conference. Vol. 62. No 3. IEEE ; 1999, 2005.*

[25] http://wikistat.fr/pdf/st-m-app-intro.pdf, *Imputation de données manquantes.*

[26] L. Li, J. Zhang, Y. Wang, and B. Ran, *Missing Value Imputation for Traffic-Related Time Series Data Based on a Multi-View Learning Method, 2018.*

[27] Liqiang Pan, Jianzhong Li, *K-Nearest Neighbor Based Missing Data Estimation Algorithm in Wireless Sensor Networks, School of Computer Science and Technology. 2009*

[28] Mihail Halatchev, Le Gruenwald , *Estimating Missing Values in Related Sensor Data Streams ,The University of Oklahoma School of Computer Science Norman.*

[29] Geeta Chhabra, Vasudha Vashisht, Jayanthi Ranjan, *A Review on Missing Data Value Estimation Using Imputation Algorithm, India. 2019*

[30] Shylaja B, Dr. R. Saravana Kumar , *TRADITIONAL VERSUS MODERN MISSING DATA HANDLING TECHNIQUES : AN OVERVIEW,2Department of Computer Science and Engineering, Dayananda sagar Academy of Technology and Management Bengaluru, India*

[31] Linghe Kong, Mingyuan Xia, Xiao-Yang Liu, Guangshuo Chen, Yu Gu, Min-You Wu, Xue Liu, *Data Loss and Reconstruction in Wireless Sensor Networks, IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS*

[32] Kirti Kharb, Bhisham Sharma, Dr. Trilok C. Aseri, *Reliable and Congestion Control Protocols for Wireless Sensor Networks, 2015* ,1-11

[33] Wei-Chao Lin1, Chih-Fong Tsai, *Missing value imputation : a review and analysis of the literature, Department of Information Management, Taoyuan, Taiwan. 2019*

[34] Kevin Wellenzohn, Hannes Mitterer, Johann Gamper, M. H. Böhlen, Mourad Khayati, *Missing Value Imputation in Time Series using Top-k Case Matching, Taoyuan, Free University of Bozen-Bolzano and University of Zurich, 2014*, pages 2-6

[35] Jingfei He and Yatong Zhou, *Real-Time Data Recovery in Wireless Sensor Networks Using Spatiotemporal Correlation Based on Sparse Representation, The School of Electronics and Information Engineering and Key Laboratory of ElectronicMaterials and Devices of Tianjin, Hebei University of Technology, Tianjin 300401, China*, 2019

[36] Mariem Ahmed, Walied Saeed, and Ashraf El-Sisi, *Simple Missing Data Estimation Algorithm in WSN Based on Spatial and Temporal Correlation, Computer Science dept, Faculty of computers and Information, Menofia University*

[37] Hong Zhou, Kun-Ming Yu, Ming-Gong Lee and Chin-Chuan Han, *The Application of Last Observation Carried Forward Method for Missing Data Estimation in the Context of Industrial Wireless Sensor Networks*

[38] Zhipeng Gao, Weijing Cheng, Xuesong Qiu, and Luoming Meng, *A Missing Sensor Data Estimation Algorithm Based on Temporal and Spatial Correlation*

[39] Le Gruenwald, Hamed Chok, Mazen Aboukhamis, *Using Data Mining to Estimate Missing Sensor Data, School of Computer Science The University of Oklahoma Norman, OK 73019, U.S.A.*

Thèses

[40] Sanaa Kawther Ghalem, Thèse de doctorat, *Gestion des incertitudes dans les réseau de capteurs sans-fil,Université d'Oran, 2019*, 2-109

[41] Diery Ngom, Thèse de doctorat, *Optimisation de la durée de vie dans les réseaux de capteurs sans fil sous contraintes de couverture et de connectivité réseau, Réseaux et télécommunications[cs.NI].Université de Haute Alsace -Mulhouse ;Université Cheikh Anta Diop de Dakar, 2016. Français. NNT : 2016MULH9134. tel-01531464*, 23-30

[42] Kechar Bouabdellah, Thèse de doctorat, *Problématique de la consommation d'énergie dans les réseaux de capteurs sans-fil, Université d'Oran, 2010*, 28-37

Documents web

[43] HTTP://WIKISTAT.FR/PDF/ST-M-APP-INTRO.PDF, *Imputation de données manquantes.*

[44] HTTP://DB.CSAIL.MIT.EDU/LABDATA/LABDATA.HTML, *Berkeley Intel Lab Data.*