



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEURE ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE ABOU-BEKR BELKAID – TLEMCCEN

THÈSE LMD

Présentée à :

FACULTE DES SCIENCES – DEPARTEMENT D'INFORMATIQUE

Pour l'obtention du diplôme de :

DOCTORAT

Spécialité: *Réseaux et Systèmes Distribués*

Par :

LIAZID Hidaya

Sur le thème

**Intégration du Cloud Computing pour le stockage et l'agrégation
de données dans les réseaux de capteurs sans fil**

Soutenue publiquement le 08 Avril 2021 à Tlemcen devant le jury composé de :

Mr BENAMAR Abdelkrim	MCA	Université de Tlemcen	Président
M ^r LEHSAINI Mohamed	Professeur	Université de Tlemcen	Directeur de thèse
Mr KHALFI Mohammed Fethi	MCA	Université de Sidi Bel Abbes	Examineur
Mr MERAD BOUDIA Omar Rafik	MCA	Université d'Oran 1	Examineur
Mr HADJILA Mourad	MCA	Université de Tlemcen	Examineur

*Laboratoire Systèmes et Technologies de l'Information et de la Communication (STIC)
BP 119, 13000 Tlemcen - Algérie*

Résumé

Cette thèse traite l'intégration des réseaux de capteurs sans fil (RCSFs) au Cloud Computing. L'objectif étant de résoudre les problèmes intrinsèques que posent ces réseaux en termes de stockage de données, d'énergie limitée des batteries des capteurs et d'accès au médium (congestion). En fait, il s'agit d'améliorer les performances des RCSFs pour une plus grande durée de vie et une plus large exploitation. Comme les transmissions de données sont à la base des problèmes évoqués, nous avons commencé dans un premier temps par étudier l'impact de leur réduction à l'aide des techniques de prédiction duale et d'agrégation de données. Ainsi, nous avons mis au point un nouvel algorithme de prédiction, appelé EADPS (Extended Adaptive Dual Prediction Scheme). Ensuite, nous avons comparé l'agrégation des données à chacun des schémas ADPS (Adaptive Dual Prediction Scheme) et EADPS. Il ressort que l'agrégation est une technique bien meilleure que le schéma ADPS pour les réseaux de petite taille et quel que soit la précision des prédictions. Cependant, elle ne devient sérieusement concurrente au schéma EADPS que pour des RCSFs de grandes tailles.

Nous avons effectué différentes simulations sur des données réelles à l'aide de la prédiction seule, de l'agrégation seule et ensuite de leur combinaison. Il ressort qu'avec la combinaison de ces deux techniques nous avons maximisé le taux de réduction des transmissions. Partant de ces résultats encourageants, nous avons conçu et réalisé une plateforme de simulation pour l'intégration des RCSFs au Cloud Computing.

Mots clés : Réseaux de capteurs Sans Fil, Cloud Computing, Prédiction, Prévision, Agrégation de données.

Abstract

This thesis deals with the integration of Wireless Sensor Networks (WSNs) in Cloud Computing. The goal is to solve the intrinsic problems posed by these networks in terms of data storage, limited energy of the batteries of sensors and access to the network (congestion). In fact, it is about improving the performance of WSNs for longer network lifetime and a large number of applications. Since data transmissions are the basis of the problems mentioned, we first started by studying the impact of their reduction using dual prediction and data aggregation techniques. Thus, we have developed a new prediction algorithm, called EADPS (Extended Adaptive Dual Prediction Scheme). Next, we compared the data aggregation to each of ADPS (Adaptive Dual Prediction Scheme) and EADPS. It turns out that aggregation is a much better technique than the ADPS scheme for small networks and regardless of the accuracy of the predictions. However, it only becomes seriously competitive with the EADPS scheme for large sizes of WSNs.

We performed different simulations on real data using prediction alone, aggregation alone and then their combination. It appears that with the combination of these two techniques we have maximized the rate of reduction of transmissions. Based on these encouraging results, we designed and produced simulation software for the integration of RCSF into Cloud Computing.

Keywords : Wireless sensor networks, Cloud Computing, Prediction, Forecasting, Aggregation.

ملخص

تناول هذه الأطروحة تكامل شبكات الاستشعار اللاسلكية (WSNs) في الحوسبة السحابية. الهدف هو حل المشكلات الجوهرية التي تطرحها هذه الشبكات من حيث تخزين البيانات والطاقة المحدودة من بطاريات أجهزة الاستشعار والوصول إلى الشبكة (الازدحام). في الواقع، يتعلق الأمر بتحسين أداء شبكات الاستشعار اللاسلكية لحياة أطول وتشغيل أوسع. نظرًا لأن عمليات نقل البيانات هي أصل المشكلات المذكورة، فقد بدأنا أولاً بدراسة تأثير تقليلها باستخدام تقنيات التنبؤ المزدوج وتجميع البيانات. وبالتالي، قمنا بتطوير خوارزمية توقع جديدة تسمى (EADPS) مخطط التنبؤ المزدوج الموسع. بعد ذلك، قمنا بمقارنة تجميع البيانات بكل من مخطط التنبؤ المزدوج التكيفي (ADPS) و EADPS. اتضح أن التجميع هو أسلوب أفضل بكثير من مخطط ADPS للشبكات الصغيرة وبغض النظر عن دقة التنبؤات. ومع ذلك، فإنها تصبح منافسة بشكل جدي فقط مع مخطط EADPS للأحجام الكبيرة من شبكات.

أجرينا عمليات محاكاة مختلفة على بيانات حقيقية باستخدام التنبؤ وحده، والتجميع وحده ثم الجمع بينهما. يبدو أنه من خلال الجمع بين هاتين التقنيتين، قمنا بتعظيم معدل تقليل عمليات النقل. بناءً على هذه النتائج المشجعة، قمنا بتصميم وإنتاج أرضية محاكاة لدمج شبكات الاستشعار اللاسلكية في الحوسبة السحابية.

الكلمات المفتاحية: شبكات الاستشعار اللاسلكية، الحوسبة السحابية، التنبؤ، التوقعات، تجميع البيانات.

Remerciements

Je tiens, avant toute chose, à remercier ALLAH TOUT PUISSANT de m'avoir donné la force, le courage, la volonté et la patience afin de réaliser ce travail de thèse.

J'exprime ma profonde gratitude à mon directeur de thèse, Mr LEHSAINI Mohamed pour son esprit scientifique, sa pédagogie, sa disponibilité, sa patience, ses conseils et ses idées qui m'ont permis de mener à bien cette thèse.

Mes remerciements les plus sincères sont adressés à de me faire l'honneur de s'intéresser à ce travail et d'avoir présidé le jury

J'exprime ensuite ma plus profonde gratitude à qui m'ont fait l'honneur d'accepter d'être les rapporteurs de cette thèse.

Ce travail a été réalisé au sein du laboratoire Systèmes et Technologie de l'Information et de la Communication (STIC) de l'université de Tlemcen. Je souhaite également remercier les membres de ce laboratoire. A commencer par les responsables de ce laboratoire Mr LEHSAINI et Mr FEHAM de m'avoir accueilli au sein de ce laboratoire pour leurs disponibilités, précieux conseils et gentillesse. A Mme BELHADI Sihem pour son amitié, son aide, son encouragement et ses conseils. Enfin, à tous mes collègues du laboratoire à qui je souhaite beaucoup d'épanouissement dans leurs travaux.

Enfin, J'adresse mes sincères remerciements à toutes les personnes qui me sont très chères qui m'ont encouragé, soutenu durant toutes ces années, en commençant par ma famille : d'abord ma pensée va également vers mes très chers parents pour tous leurs sacrifices pour nous depuis toujours et pour m'avoir soutenu et encouragé dans mon choix de poursuites d'étude. Ma très chère sœur et mes très chers frères pour leurs accompagnements et encouragements, ma famille maternelle et paternelle et ma belle-famille.

A titre plus personnel, Je remercie chaleureusement mon très chers époux Imad, pour sa patience, son encouragement, ses conseils et surtout pour son soutien moral.

A vous tous, Merci !

Dédicaces

Je dédie ce travail à mes chers parents, à mes chères sœurs Djoweyda et Imane et leurs enfants. Je le dédie également à mon cher époux et à ma fille Nour El Houda. Enfin, je le dédie à mes beaux-parents.

A la mémoire de ma Grand-Mère Hadja Safia

Hidaya Liazid

Table des matières

Liste des figures	v
Liste des tableaux	viii
Introduction Générale	1
1 Généralités sur les réseaux de capteurs sans fil et le Cloud computing	5
1.1 Introduction	5
1.2 Les réseaux de capteurs sans fil	5
1.2.1 Définition d'un capteur sans fil	5
1.2.2 Structure d'un capteur sans fil	6
a) Structure physique	6
b) Structure logique	7
1.2.3 Définition d'un réseau de capteurs sans fil	7
1.2.4 Architecture de communication dans les RSCFs	8
1.2.5 Typologie d'un réseau de capteurs sans fil	8
a) Topologie plate	8
b) Topologie hiérarchique	9
1.2.6 Modèle de délivrance de données	9
1.2.7 Domaines d'application des RSCFs	10
1.2.8 Les problèmes spécifiques des RSCFs	13
1.3 Le Cloud computing	18
1.3.1 Définition du Cloud computing	18
1.3.2 Types de services du Cloud computing	19
a) Infrastructure en tant que service (IaaS)	19
b) Plateforme en tant que service (PaaS)	19
c) Logiciel en tant que service (SaaS)	19
1.3.3 Modèles de Cloud computing	20
1.3.4 Avantages du Cloud computing	20
1.3.5 Désavantages du Cloud computing	21
1.4 Conclusion	21
2 État de l'art sur l'intégration RCSF-Cloud	22
2.1 Introduction	22
2.2 Première conception	22
2.3 Deuxième conception	25
2.3.1 Stockage et traitement des données	25

a)	Les RCSFs comme entité physique perçue par l'utilisateur via le Cloud	25
b)	Les RCSFs comme entités virtuelles	32
2.3.2	Aspect sécuritaire	37
2.4	Troisième conception : Intégration orientée application	39
2.4.1	Le domaine médical	39
2.4.2	Le domaine environnemental	40
2.4.3	Le domaine agricole	41
2.4.4	Le domaine du contrôle des structures	42
2.5	Intégration entre RCSF et Cloud pour l'IoT	42
2.6	Récapitulatif	44
2.7	Résumé sur l'agrégation des données dans les RCSFs	44
2.7.1	Structure d'agrégation	45
a)	Structures basées sur la hiérarchie	45
b)	Structures basées sur l'épine dorsale	46
2.7.2	Agrégation de données sans structure	47
a)	La fonction d'agrégation	47
b)	L'agrégation des données en prévision	48
c)	La compression d'agrégation	49
2.7.3	Planification d'agrégation	49
2.7.4	Synthèse sur l'agrégation de données dans les RCSFs	50
2.8	La prédiction des données	50
2.9	Conclusion	51

3 Etude de la prédiction duale et de l'agrégation sur le modèle RCSF en anneaux 52

3.1	Introduction	52
3.2	Présentation du modèle RCSF étudié	53
3.2.1	Extension du modèle RCSF considéré	54
3.2.2	Transmissions du nœud capteur à la station de base (BS)	54
3.2.3	Transmissions de la station de base (BS) vers le nœud capteur	55
3.3	Prédictions de données dans les RCSFs	56
3.3.1	Paradigmes de communication	57
a)	Paradigme de prédiction simple	57
b)	Paradigme de prédiction duale (DPS)	57
c)	Schéma adaptatif de prédiction duale (ADPS)	59
3.3.2	Choix du modèle indépendant	60
3.3.3	Choix du modèle dans le nœud capteur	60
3.3.4	Choix du modèle dans le nœud récepteur (BS ou CH)	62
3.4	Modélisation des schémas de prédiction duale	63
3.4.1	Hypothèses	64
3.4.2	Choix du modèle de prédiction et dissémination	64
3.5	Impacts des prédictions sur le nombre de transmissions	65
3.6	Récapitulatif	68
3.7	Agrégation de données	69
3.7.1	Réseau avec deux nœuds capteurs	69

3.7.2	Généralisation : Cas d'un réseau plus grand	72
3.8	Impact des schémas d'agrégation et de prédiction de données sur le nombre de transmissions	74
3.9	Synthèse	76
3.9.1	Droite d'ajustement en corrélation linéaire	76
3.9.2	Précisions des prédicteurs	78
3.10	Conclusion	79
4	Un modèle basé sur l'agrégation et la prédiction duale de données pour les prévisions dans les RCSFs	80
4.1	Introduction	80
4.2	Amélioration du paradigme de prédiction ADPS	81
4.3	Réduction des transmissions à l'aide du schéma EADPS	83
4.4	Comparaison entre l'agrégation et l'algorithme de prédiction EADPS	86
4.5	Problématique du modèle de prédiction des données	89
4.5.1	Choix et structure du modèle	90
4.5.2	Robustesse du modèle	91
4.5.3	Sélection du modèle	92
4.5.4	Cas du prédicteur EADPS	99
4.6	Etude à l'échelle du réseau	106
4.6.1	Configuration de la simulation	106
4.6.2	Hypothèses	107
4.6.3	Impact du paradigme de communication	108
4.6.4	Résultats obtenus et discussion	110
4.6.5	Récapitulatif	116
4.7	Application du schéma EADPS au réseau étudié	117
4.7.1	Résultats de simulation	117
4.7.2	Consommation d'énergie	119
4.8	Conclusion	119
5	Apport de l'intégration RCSF-Cloud computing sur les performances des RCSFs	121
5.1	Introduction	121
5.2	Environnement de développement	121
5.2.1	Langage de programmation java	121
5.2.2	Eclipse IDE	122
5.2.3	Le langage émulé Matlab	122
5.2.4	Le simulateur CloudSim	123
a)	Architecture générale de CloudSim	123
b)	Les Classes de CloudSim	123
5.3	Architecture générale du simulateur PREDAG	125
5.4	Description de la plateforme	126
5.5	Résultats et discussion	129
5.6	Conclusion	132
	Conclusions et Perspectives	135

Bibliographie

137

Liste des figures

1.1	Les composants d'un nœud capteur [1]	6
1.2	Architecture de communication d'un RCSF	8
1.3	Fonctionnement des RCSFs	9
1.4	Services des clouds [2]	19
2.1	Architecture globale de la solution "Cloud Sensor Services" [3]	28
2.2	Architecture SOA pour Capteur-Cloud [4]	28
2.3	Architecture logicielle de Capteur-Cloud [5]	29
2.4	Architecture logicielle Cloud fournissant des services d'informations aux RCSFs [6]	31
2.5	Architecture de l'infrastructure Capteur-Cloud [7]	33
2.6	Les composants et les acteurs de l'architecture Capteur-Cloud [8]	36
2.7	Différentes structures de rail road	47
3.1	Modèle du réseau RCSF considéré	53
3.2	Une variante du schéma SPS avec génération du modèle dans le nœud capteur	58
3.3	Schéma DPS avec génération indépendante du modèle	59
3.4	Choix des modèles de prédiction dans les nœuds capteurs	61
3.5	Choix des modèles de prédiction dans la station de base	63
3.6	Impact de la taille du réseau (D) sur le nombre de transmissions dans le premier anneau	68
3.7	Exemple de corrélation entre les valeurs Y_i et Y_j d'un couple de données	70
3.8	Le rectangle bleu hachuré au centre représente la zone où les deux prédictions (y_i et y_j) sont correctes	72
3.9	Impact de la taille du réseau sur le nombre de transmissions dans le premier anneau : Application de l'agrégation et du schéma DPS séparément	75
3.10	Combinaison Agrégation-DPS	77
4.1	Nombre de transmissions en utilisant le schéma EADPS dans un RCSF avec $D = 8$	85
4.2	Impact de la taille du réseau sur le nombre de transmissions dans le premier anneau : Comparaison Agrégation - Algorithme EADPS	86
4.3	Combinaison Agrégation - Algorithme EADPS	87
4.4	Impact de la combinaison des techniques d'agrégation et de prédiction ADPS et EADPS sur le taux de transmission de données	88
4.5	Exemples de données météorologiques enregistrées durant Janvier 2017	93
4.6	Effet de la taille de la fenêtre d'estimation et de l'ordre na du modèle sur ses performances à l'aide du critère (Akk) : Cas de la température d'air	95

4.7	Exemple de résultat de prédiction des données météorologiques : Journée du 02/01/2017	96
4.8	Résultats de prédiction des données météorologiques du mois de Janvier 2017 : Superposition mesure-modèle avec indication du critère ' <i>fit</i> '	97
4.9	Résultats de prédiction des données météorologiques du mois de Janvier 2017 : Séparation mesure-modèle avec indication d'écart	98
4.10	Evolution des taux de transmission des données météorologiques prédites au niveau des capteurs en fonction de la tolérance imposée et la durée de prédiction	99
4.11	Evolution dans le temps des coefficients na et nc du modèle ARMA(2,1) sélectionné pour l'humidité relative de l'air : Algorithme EADPS	100
4.12	Évolution du critère (VC) du modèle de température ARMA(2,1) : Schéma EADPS	100
4.13	Évolution du critère (Akk) du modèle ARMA(2,1) de l'humidité relative d'air : Schéma EADPS	101
4.14	Prévisions des valeurs météorologiques sur une période de 12 heures à l'aide des modèles élaborés dans le schéma EADPS	101
4.15	Evolution des écarts absolus entre les prévisions des modèles EADPS et les valeurs mesurées sur une période de 12 heures	102
4.16	Prédiction des données météorologiques des deux premières journées du mois de Juin 2017	104
4.17	Schéma du RCSF étudié avec des données réelles	107
4.18	Taux de transmission en fonction de la taille de la fenêtre d'actualisation du modèle dans le schéma DPS	109
4.19	Evolution des taux de transmissions du réseau étudié dans le paradigme DPS	110
4.20	Evolution du nombre total de transmissions et d'extra-transmissions du réseau étudié dans le paradigme DPS	111
4.21	Evolution des taux de transmissions en utilisant le paradigme ADPS	112
4.22	Comparaison du paradigme DPS au procédé d'agrégation sans corrélation .	112
4.23	Corrélations entre les différentes grandeurs météorologiques	113
4.24	Comparaison du paradigme DPS au procédé d'agrégation avec corrélation .	114
4.25	Cas d'application de l'agrégation avec et sans corrélation	114
4.26	Comparaison du paradigme ADPS au procédé d'agrégation sans corrélation	115
4.27	Comparaison du paradigme ADPS au procédé d'agrégation avec corrélation	115
4.28	Combinaison du paradigme DPS avec l'agrégation	116
4.29	Combinaison du paradigme ADPS avec l'agrégation	116
4.30	Simulation du réseau en utilisant le paradigme EADPS	117
4.31	Taux de transmission du réseau : Application de l'agrégation avec et sans corrélation dans le schéma EADPS	118
4.32	Combinaison du paradigme EADPS avec l'agrégation avec corrélation de données	118
5.1	Architecture multicouches de CloudSim [9]	124
5.2	Diagramme de classes du simulateur CloudSim [9]	124
5.3	Architecture d'intégration entre RCSF et Cloud Computing	126
5.4	Configuration de l'environnement Cloud Computing	127

5.5	Configuration manuelle du RCSF	128
5.6	Affichage du RCSF organisé selon le modèle en anneaux	129
5.7	Consommation d'énergie dans un RCSF (sans agrégation et prédiction)	129
5.8	Gain d'énergie du réseau à l'aide du paradigme de prédiction ADPS	130
5.9	Gain d'énergie du réseau à l'aide du paradigme de prédiction DPS	130
5.10	Gain d'énergie du réseau avec agrégation de données	131
5.11	Gain d'énergie du réseau à l'aide du paradigme ADPS et l'agrégation avec corrélation de données	131
5.12	Gain d'énergie du réseau à l'aide du paradigme DPS et l'agrégation avec corrélation de données	132
5.13	Gain d'énergie du réseau à l'aide du paradigme de prédiction EADPS	132
5.14	Gain d'énergie du réseau à l'aide d'un prédicteur actualisé selon le para- dигme EADPS (Agrégation avec et sans corrélation)	133
5.15	Gain d'énergie du réseau à l'aide du paradigme EADPS (Agrégation avec et sans corrélation de données)	133

Liste des tableaux

3.1	Paramètres considérés dans le calcul du nombre de transmissions et de réceptions de données d'un capteur	55
4.1	Récapitulatif des tolérances admises	95
4.2	Taux des données non transmises en fonction du seuil de tolérance admis par l'utilisateur (Qualité du modèle EADPS)	103
4.3	Résultats de prédiction de la température de l'air en utilisant "ADPS" et "EADPS"	103
4.4	Résultats de prédiction de la pression atmosphérique en utilisant "ADPS" et "EADPS"	105
4.5	Résultats de prédiction de l'humidité en utilisant "ADPS" et "EADPS"	105
4.6	Taux de transmission de données : Comparaison entre les simulations et le modèle : Cas de la température d'air	105
4.7	Taux de transmission de données : Comparaison entre les simulations et le modèle : Cas de la pression atmosphérique	106
4.8	Taux de transmission de données : Comparaison entre les simulations et le modèle : Cas de l'humidité d'air	106

Glossaire

- ADC** Analog to Digital Converter. 6
- ADPS** Adaptive Dual Prediction schema. 3, 80
- AEMS** an Agricultural Environmental Management System. 41
- AIC** Akaike's Information Criteria. 48
- API** Application Programming Interface. 19
- AR** Auto-Regressive : Autorégressif. 89
- ARIMA** AutoRegressive Integrated Moving Average. 90
- BIC** Bayesian Information Criteria. 48
- BP** Belief Propagation. 43
- BSN** Body Sensor Network. 40
- CDMA** Code Division Multiple Access. 16
- CH** clusterhead. 9
- CIS** Cloud Information Service. 125
- CLOUDS** Cloud computing and Distributed systems. 123
- CP** Constraint Programming. 43
- CPABE** Ciphertext-Policy Attribute-Based Encryption. 38
- CPOD** Chronic Pulmonary Obstructive Disorder. 12
- CSMA** Carrier Sense Multiple Access. 16
- DARPA** Defense Advanced Research Projects Agency. 10
- DD** Directed Diffusion. 46
- DDB** Dynamic Directed Backbone. 46
- DPS** Dual Prediction Scheme. 57, 80

- DPU** Data Processing Unit. 27
- DQM** Data Quality Maximization. 47
- EADA** Energy Aware Data Aggregation. 45
- EADPS** Extended Adaptive Dual Prediction schema. 80
- EC2** Amazon Elastic Cloud Computing. 19
- EEMSRA** Energy-Efficient Mobile Sink Routing protocol Algorithm. 45
- FDMA** Frequency-Division Multiple Access. 16
- FIFO** First In First Out. 60
- GEM** Graph EMbedding. 16
- GPS** Global Positioning System. 6
- GPSR** Greedy Perimeter Stateless Routage. 82
- HEED** Hybrid Energy Efficient Distributed clustering approach. 45
- IAMU** unité de gestion des identités et des accès. 27
- IEEE** Institute of Electrical and Electronics Engineers. 18
- IoT** Internet of Things. 41, 42
- LEACH** Low Energy Adaptive Clustering Hierarchy. 45
- LMS** Least Mean Square : Moindre Carré. 89
- LOS** Line Of Sight. 13
- MAC** Media Access Control. 60
- MAE** Mean Absolute Error. 50
- MAPE** Mean Absolute Percentage Error. 50
- MCC** Mobile Cloud Computing. 29
- NaaS** Network as a Service. 34
- OMSP** Orchestration Model for Service Provisioning. 34
- QoS** qualité de service. 7, 125
- RCSF** réseau de capteurs sans fil. 1

- RF** RadioFréquence. 13
- RMSE** Random Mean Square Error. 50
- RNA** réseaux de neurones artificiels. 62
- RR** Rail road. 46
- RSA** Rivest–Shamir–Adleman. 32
- SGIM** Statistical Group Index Matching. 26
- SMPC** Stream monitoring and processing component. 26
- SOA** Service-Oriented Architecture. 27
- SOSUS** Sound Surveillance System. 10
- SPS** Single Prediction Scheme. 57
- SSC** Social Sensor Cloud. 44
- TAG** Tiny AGgregation. 45
- TASA** Time Adaptive Schedule Algorithm. 23
- TASC** Trust-Assisted Sensor Cloud. 42
- TDMA** Time Division Multiple Access. 16, 60
- TPSDT** Time and Priority-based Selective Data Transmission. 32
- VO** Virtual organisation. 25
- VSN** Virtual Sensor Network. 36

Introduction Générale

Introduction Générale

Contexte

Les capteurs sont des dispositifs de mesure qui transforment une grandeur physique en un signal électrique permettant ainsi son transport. Ils s'appuient sur divers principes physiques tels que l'induction, la piézoélectricité, la photoélectricité, la dilatation thermique, l'effet Hall, l'effet Doppler, la variation de capacité, la variation de résistance, la variation d'inductance, etc. . .

Les capteurs sont classés en deux grandes catégories : les capteurs passifs qui nécessitent une source d'alimentation électrique pour pouvoir fournir l'information et les capteurs actifs qui fonctionnent sans le besoin de cette source. À l'origine, les capteurs étaient des détecteurs électromécaniques destinés à la mesure des grandeurs physiques. Leur première utilisation remonte à 1933, dans les applications des premiers thermostats d'ambiance. Les dernières années du XX^{ème} siècle ont vu apparaître le concept de capteurs intelligents miniaturisés grâce aux progrès réalisés dans le domaine des MEMS (Micro-Electro-Mechanical-Systems). En plus de leur faculté de mesurer une grandeur physique, ces capteurs possèdent d'autres fonctionnalités comme la capacité à pouvoir communiquer entre eux et à effectuer des calculs. En effet, ils incluent désormais un processeur intégré, une mémoire et un émetteur-récepteur, le tout conçu à très petite échelle, alimenté par une source de batterie. Ces capteurs intelligents constituent un nœud dans un réseau de capteurs sans fil (RCSF).

Un réseau de capteurs sans fil est une infrastructure composée d'éléments de détection (mesure), de calcul et de communication qui donne à un administrateur la possibilité d'observer et de réagir à des événements et des phénomènes dans un environnement spécifique. L'administrateur est généralement une entité civile, gouvernementale, commerciale ou industrielle. L'environnement peut être le monde physique, un système biologique ou un cadre de technologie de l'information. Les réseaux de capteurs sans fil ont connu un fort déploiement durant ces dernières années dans une pléthore d'applications. En plus de la détection, on est souvent également intéressé par le contrôle et l'activation à l'aide des actionneurs. Les RCSF ont permis à la technologie de détection et de contrôle de faire des progrès significatifs, non seulement en science et en génie, mais tout aussi important, sur une large gamme d'applications liées à la protection et à la sécurité des infrastructures critiques, aux soins de santé, à l'environnement, à l'énergie, à la sécurité alimentaire, aux processus de production, à la qualité de vie et à l'économie. En plus de réduire les coûts et d'augmenter l'efficacité pour les industries et les entreprises, la mise en réseau de capteurs sans fil devrait apporter aux consommateurs une nouvelle génération de commodités, y compris, mais sans s'y limiter, le chauffage et l'éclairage télécommandés, la surveillance médicale, le contrôle automatisé de l'épicerie, le diagnostic de santé personnel, le bilan

automobile automatisé et la garde d'enfants. Les RCSF se caractérisent par le fait que les nœuds qui les composent présentent des limitations matérielles importantes comme la faible capacité de traitement de données et de leur stockage et l'alimentation dans de nombreux cas par des batteries non rechargeables, difficile voire impossible à remplacer. Compte tenu de ces contraintes le problème majeur dans les réseaux de capteurs sans fil demeure lié à leur durée de vie. En effet, l'épuisement d'un ou plusieurs capteurs dans le réseau aura des répercussions sur le maintien de la couverture et la connectivité de la zone à surveiller ce qui réduira la qualité de service du réseau. Des travaux de recherche se sont alors orientés vers cette problématique afin d'optimiser la consommation d'énergie au sein des réseaux de capteurs sans fil et augmenter donc leur longévité. De nombreuses méthodes ont été proposées au cours des dernières années, et de nombreuses recherches sont encore en cours sur la façon d'optimiser l'utilisation de l'énergie dans les réseaux de capteurs à batterie limitée. Cependant, aucune des solutions proposées n'est universellement applicable. Par exemple, si les applications de sécurité nécessitent un traitement rapide réactif, ce n'est pas le cas pour d'autres applications, comme en agriculture où la propriété de délai n'est pas aussi importante. Nous pensons que les problèmes d'économie d'énergie des réseaux de capteurs sans fil doivent être résolus en tenant compte des exigences de l'application de manière systématique.

De surcroît, d'autres limitations ont été perçues dans l'utilisation des réseaux de capteurs sans fil. On mentionne par exemple le nombre de capteurs qui peuvent être déployés par dizaine jusqu'à des milliers générant une multitude d'informations circulant dans le réseau, envoyées depuis chaque capteur et qui peuvent facilement atteindre un nombre exponentiel provoquant ainsi une inondation du réseau. Cette surcharge en données brutes peut facilement dépasser la capacité du réseau ce qui induit une défaillance de ses performances.

Le Cloud computing fût suggéré par plusieurs travaux de recherche comme une solution prodige au problème des réseaux de capteurs sans fil. En effet, grâce aux ressources inépuisables du cloud computing et à sa flexibilité d'intégration avec d'autres technologies, il est perçu comme une solution forte pouvant concilier les réseaux de capteurs sans fil avec leurs différentes contraintes.

Motivations de recherche

Le but de cette thèse consiste à proposer une contribution à la résolution de quelques-uns des problèmes des réseaux de capteurs sans fil bien acceptés. Il s'agira de suggérer l'extension des réseaux de capteurs sans fil vers les services du Cloud déjà disponibles pour des opérations de stockage ou d'agrégation de données dans ces types de réseaux. Les objectifs spécifiques de cette thèse se résument à :

- Mener une recherche bibliographique sur les études effectuées sur l'intégration du Cloud avec les réseaux de capteurs sans fil et le rôle que joue cette intégration dans l'amélioration des performances des réseaux de capteurs sans fil.
- Proposer sur la base de cet état de l'art une nouvelle approche de résolution des problèmes majeurs intrinsèques aux réseaux de capteurs sans fil.

- Évaluer nos propositions analytiquement et simulations et situer nos solutions par rapport aux solutions existantes.

Objectifs de recherche et Contributions

Pour atteindre nos objectifs de recherche, nous avons proposé deux contributions :

- **Contribution 1** : Une amélioration de la technique de prédiction duale des données qui est considérée dans la littérature spécialisée comme la plus performante technique dans le domaine de prédiction. Le but étant de réduire davantage le nombre de transmissions de données dans un réseau de capteur sans fil car ces transmissions sont énergivores (celles qui consomment plus d'énergie). La supériorité avérée de notre proposition permettra notamment de prolonger la durée de vie du réseau. Sa combinaison avec la technique d'agrégation de données lui offre une toute autre puissance.
- **Contribution 2** : Une nouvelle architecture d'intégration entre un réseau de capteurs sans fil et le cloud a été conçue. Dans cette deuxième partie nous avons proposé une nouvelle plateforme de simulation de l'intégration entre un réseau de capteurs sans fil et le Cloud en y ajoutant des modules spécifiques permettant de réduire le taux de transmissions au sein du réseau.

Organisation de la thèse

Le manuscrit de la thèse est structuré en cinq chapitres.

- Dans le premier chapitre intitulé "Généralités sur les réseaux de capteurs sans fil et le Cloud computing", nous donnons un aperçu général sur les réseaux de capteurs sans fil et le Cloud, leurs caractéristiques spécifiques, leurs types d'application, leurs avantages, leurs inconvénients et d'autres considérations pratiques.
- Dans le second chapitre intitulé "Etat de l'art sur l'intégration du cloud computing dans les réseaux de capteurs" : On présente un état de l'art sur les différentes approches d'intégration entre la technologie des réseaux de capteurs sans fil et la technologie du Cloud.
- Dans le troisième chapitre intitulé "Modèle RCSF en anneaux : Etude de la prédiction duale et de l'agrégation", on présente le modèle du réseau étudié, l'algorithme de prédiction duale et le processus d'agrégation de données. Différents fondements mathématiques sur lesquels ces modèles sont fondés ont été présentés et discutés.
- Dans le quatrième chapitre intitulé "Un modèle basé sur l'agrégation et la prédiction de données pour les prévisions dans les RCSFs", on s'est consacré à la présentation de notre contribution qui consiste en l'amélioration du modèle "Adaptive Dual Prediction schema (ADPS)" et à l'étude des différentes stratégies de réduction des transmissions de données. L'évaluation et la discussion de nos résultats ont été présentées.

- Dans le cinquième chapitre intitulé "Intégration entre Cloud et RCSF", une plateforme a été conçue et élaborée. Cette plateforme est de nature à intégrer un RCSF au service Cloud en vue d'étudier l'amélioration de ses performances notamment en termes d'économie d'énergie.

Finalement, nous couronnons ce manuscrit par une conclusion générale et des perspectives.

Chapitre 1

Généralités sur les réseaux de capteurs sans fil et le Cloud computing

Chapitre 1

Généralités sur les réseaux de capteurs sans fil et le Cloud computing

1.1 Introduction

Ces dernières années, les réseaux de capteurs sans fils (RCSF) n'ont pas cessé d'être au centre d'intérêt de la communauté des chercheurs scientifiques à cause de leurs capacités à récolter, à stocker, à traiter et à communiquer l'information à faible coût. Ils sont devenus un moyen incontournable dans le monde scientifique et industriel. On les retrouve dans plusieurs domaines d'application, comme le domaine médical (surveillance des patients), environnemental (surveillance des désastres), militaire (surveillance des adversaires), etc. . .

L'accroissement exponentiel du nombre d'informations fait que les réseaux de capteurs sans fil soient confrontés à diverses difficultés causées par leurs conceptions de base. Notamment les problèmes d'accès au réseau, de stockage de données et de durée de vie. Le Cloud Computing représente donc une solution prometteuse à ces problèmes. L'intégration entre ces deux technologies peut alors permettre d'alléger la charge sur les réseaux de capteurs sans fil et permettre ainsi leur meilleure exploitation. Dans ce chapitre, nous présentons l'ensemble des connaissances fondamentales liées aux réseaux de capteurs sans fil, au Cloud computing et l'intégration entre ces deux technologies.

1.2 Les réseaux de capteurs sans fil

Dans cette section, nous allons nous focaliser sur les réseaux de capteurs sans fil et sur les notions de base nécessaires à la compréhension de cette technologie.

1.2.1 Définition d'un capteur sans fil

Les progrès récents dans le domaine des micromachines, des MEMS (Micro-Electro Mechanical Systems), des communications sans fil et de l'électronique numérique ont ouvert la voie au développement de capteurs multifonctionnels à faible coût, à faible puissance et de petite taille. Un capteur sans fil a la capacité d'effectuer la détection, le

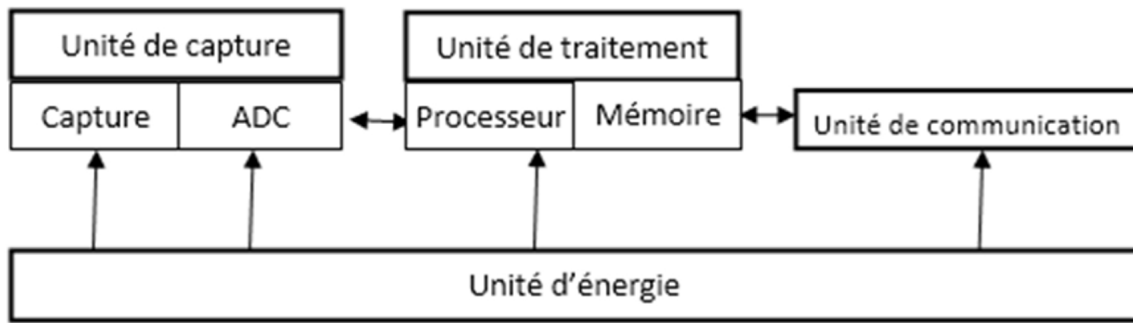


FIGURE 1.1 – Les composants d'un nœud capteur [1]

calcul sur les données et des opérations de communication sans fil. Les capteurs peuvent être déployés dans des emplacements aléatoires ou prédéfinis à l'avance.

1.2.2 Structure d'un capteur sans fil

Cette section présente l'architecture d'un nœud capteur du point de vue matériel et logiciel.

a) Structure physique

Un capteur sans fil représente l'élément fondamental d'un RCSF, il est constitué de quatre unités principales : une unité d'acquisition de données, une unité de traitement, une unité de communication et une unité de contrôle d'énergie. En plus de ces unités, un capteur sans fil peut contenir d'autres composants selon le domaine d'application considéré, par exemple il peut comporter un système de localisation (Global Positioning System (GPS)), un système générateur d'énergie (panneaux solaires) et un système mobilisateur chargé de le déplacer en cas de nécessité. La figure 1.1 illustre l'architecture d'un capteur sans fil.

1. **Unité d'acquisition** : Cette unité contient généralement des capteurs et des transducteurs (convertisseurs analogique-numériques). Les capteurs sont des dispositifs inclus dans les nœuds de capteur. Ils sont en liaison avec les phénomènes observés et permettent d'acquérir les informations et paramètres de leur environnement sous forme de signaux analogiques, qui par la suite seront convertis par Analog to Digital Converter (ADC) en des signaux numériques. Les données capturées par cette unité seront ensuite transmises vers l'unité de traitement.
2. **Unité de traitement** : L'unité de traitement est composée d'une unité de stockage, d'un microcontrôleur et d'un système d'exploitation spécifique. Il existe plusieurs systèmes d'exploitation dédiés aux capteurs sans fil comme TinyOS [10], Contiki [11], et Lorien [12]. Cette unité représente un système de commande responsable du traitement des données en provenance de l'unité d'acquisition.
3. **Unité de communication** : Les capteurs sans fil pourraient être déployés par centaines jusqu'à des milliers dans des environnements hostiles, où aucun contact

physique ne pourrait être possible pour le remplacement ou toute portée de maintenance. Par conséquent, la communication sans fil est le seul moyen d'accessibilité à distance. L'unité de communication permet d'effectuer des échanges entre les capteurs via des liens sans fil et à l'aide du module radio. Les communications sans fil offrent beaucoup d'avantages comme la mobilité et la facilité de connexion. Cette unité est donc responsable de toutes les émissions et des réceptions des données. Les capteurs sans fil sont en général géographiquement répartis sur une large zone d'interception, il est donc indispensable d'utiliser un signal suffisamment puissant pour qu'un capteur ait une portée de communication assez grande, ce qui engendre une consommation importante d'énergie. Comme la plupart des études l'ont montré [13–15] cette unité est le composant interne du capteur qui a le plus grand taux de dépense énergétique.

4. **Source d'énergie** : Cette unité représente l'un des composants les plus importants dans un capteur sans fil ; elle est responsable de l'alimentation de tous les autres composants. Chaque capteur dispose d'une source d'énergie, généralement une batterie. Cependant, compte tenu de la taille réduite du capteur, les ressources énergétiques sont souvent limitées et sont rarement remplaçables ce qui affecte négativement la durée de vie du capteur. Des solutions ont été implémentées dans certains types de capteurs afin d'améliorer leur longévité en utilisant des dispositifs de récupération d'énergie telle que des cellules solaires. Pour supprimer la contrainte énergétique, quelques recherches ont été menées sur la conversion d'énergie renouvelable ambiante (solaire, éolienne, par exemple) en énergie électrique pour alimenter certains types de capteurs.

b) Structure logique

Les nœuds capteurs exécutent des systèmes d'exploitation spéciaux, des protocoles de communication, des fonctionnalités de gestion et des applications. Les systèmes d'exploitation ont été spécifiquement développés pour les caractéristiques particulières des nœuds capteurs. Comme déjà mentionné, les plateformes matérielles sur lesquelles ces systèmes d'exploitation s'exécutent sont contraignantes. Par conséquent, ces systèmes d'exploitation ont été développés dans le but de simplifier et de donner plus d'efficacité. Certains d'entre eux conviennent également aux applications de détection d'événements, qui sont courants dans les applications des RCSF. D'un certain point de vue, les applications s'exécutant sur un nœud capteur nécessitent un bon fonctionnement du nœud et du réseau en termes de qualité de service (QoS), de gestion de données et de gestion de l'énergie. Cette fonctionnalité logicielle dans son ensemble est parfois appelée middleware.

1.2.3 Définition d'un réseau de capteurs sans fil

Un RCSF est une technologie très populaire, qui se compose d'un groupe d'entités autonomes miniaturisées et auto-organisées, appelé capteurs. Ces capteurs sont interconnectés les uns aux autres via un médium sans fil permettant ainsi d'effectuer la même fonction collectivement ou en coopération afin de surveiller et d'analyser des facteurs environnementaux ou les paramètres d'un phénomène observé.

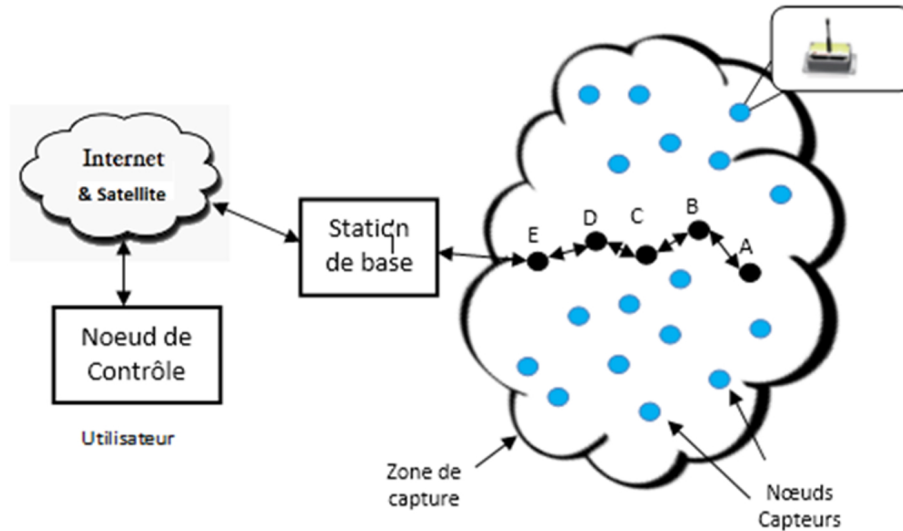


FIGURE 1.2 – Architecture de communication d'un RCSF

1.2.4 Architecture de communication dans les RCSFs

Un réseau de capteurs sans fil contient deux types de nœuds : des nœuds capteurs et des nœuds coordinateurs ou ce qu'on appelle des stations de base. Les nœuds capteurs sont éparpillés aléatoirement ou de manière prédéfinie sur une zone d'interception appelée un champ de captage. Chacun de ces nœuds collecte des paramètres correspondants au phénomène observé afin de les retransmettre aux destinataires sous forme de données. Le nœud coordinateur joue le rôle d'un puits (station de base), il est doté de ressources de calcul assez puissantes pour réceptionner les données sensorielles, d'effectuer un certain nombre de traitements si nécessaire, de stocker des données et enfin de les envoyer à l'utilisateur final via Internet ou par satellite. On note d'emblée qu'un RCSF peut contenir plusieurs nœuds coordinateurs. La figure 1.2 illustre l'architecture de communication d'un RCSF.

1.2.5 Typologie d'un réseau de capteurs sans fil

Les RCSF s'organisent généralement sous deux types de topologies, une topologie plate (où tous les nœuds sont au même niveau) et une topologie hiérarchique (où les nœuds sont organisés en différents niveaux hiérarchiques). La figure 1.3 illustre ces deux types d'architecture.

a) Topologie plate

Une topologie plate est une topologie où tous les nœuds capteurs ont le même rôle. Dans ce genre d'architecture les nœuds capteurs envoient leurs données vers la station de base, soit directement via des communications en un seul saut soit par l'intermédiaire d'un nœud de relais via des communications multi-sauts. Dans ce premier cas si la station de base ne se trouve pas dans la portée ordinaire de transmission du capteur, celui-ci dépensera plus d'énergie pour faire parvenir ses données. Une telle puissance d'émission lors de l'envoi des paquets de données se répercute négativement sur la consommation énergé-

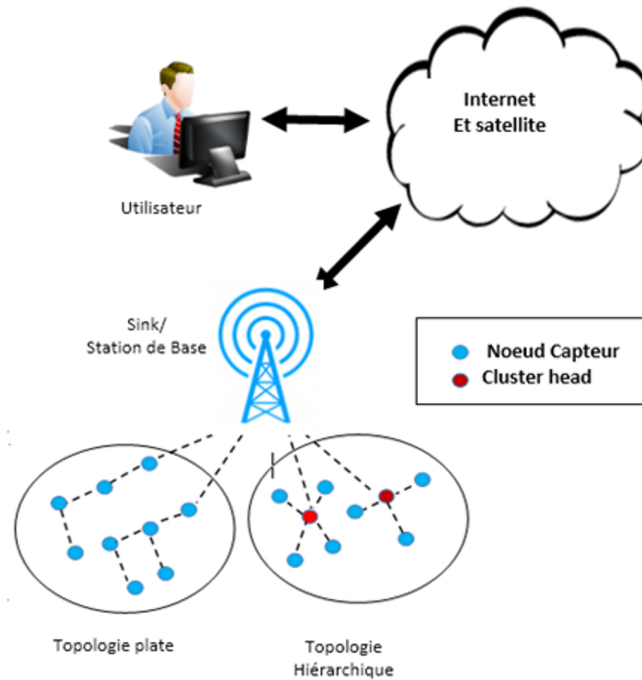


FIGURE 1.3 – Fonctionnement des RSCFs

tique du nœud et provoque une accélération dans son épuisement. D'autre part, dans des communications multi-sauts qui sont les plus fréquentes, le taux d'énergie consommé est inférieur à celui utilisé dans un mode monosaut, car un nœud capteur peut compter sur d'autres capteurs plus proches pour assurer le relais de l'information et ainsi effectuer un acheminement des données depuis le nœud source jusqu'à la station de base, la puissance d'émission est ainsi nettement réduite.

b) Topologie hiérarchique

La topologie hiérarchique repose sur le principe d'un regroupement logique des nœuds capteurs dans un RSCF en sous-groupes de capteurs appelés clusters. Le cluster est gouverné par un nœud principal appelé clusterhead (capteur principal du cluster). Ainsi, dans ce type d'architectures un nœud capteur est soit un clusterhead (CH), soit un nœud sous adjacent. Ces derniers envoient leurs données capturées à leur clusterhead correspondant lequel après réception des données effectue à son tour des traitements d'agrégation ou non puis les renvoie à la station de base allégeant ainsi le reste des nœuds de certaines tâches de traitements et diminuant aussi le trafic dans le réseau. Le clustering garantit une meilleure performance des RSCFs en termes de consommation énergétique. En cas de défaillance du clusterhead il y'a nécessité d'établir une nouvelle élection, ce qui engendre beaucoup de messages de signalisation pour effectuer le vote, ceci représente un inconvénient majeur de ce type d'architecture.

1.2.6 Modèle de délivrance de données

Les données et informations collectées par un nœud capteur doivent être délivrées à la station de base afin d'être exploitées par l'utilisateur final. Il existe plusieurs méthodes de

délivrance de données. On retrouve un modèle continu, un modèle dirigé par événements, un modèle à la demande et un modèle hybride :

1. Le modèle continu

Dans ce modèle, les données récoltées par les nœuds capteurs sont périodiquement transférées à la station de base. Tant que le nœud est vivant, l'envoi des informations s'effectue selon un intervalle de temps préprogrammé dans le nœud capteur. Ce modèle de transmission de données est généralement utilisé dans des applications qui ne nécessitent pas des données à temps critique.

2. Le modèle dirigé par événement

Dans ce type de modèle, la capture des données est actionnée par des événements, ainsi le capteur ne travaille que si un événement pertinent est arrivé. Dans ce type de modèle, les données transmises sont généralement à temps critique mais ceci peut dépendre des utilisateurs finaux.

3. Le modèle à la demande

Dans ce modèle, la transmission des données est effectuée pour donner suite à une requête envoyée par l'utilisateur au nœud capteur à chaque fois qu'il a besoin d'une mesure. Le type de requêtes est instantané et les données demandées sont en général à temps critique.

4. Le modèle hybride

Quand un nœud capteur est capable d'exécuter plusieurs modèles de délivrance de données, on peut parler de modèle hybride.

1.2.7 Domaines d'application des RCSFs

La gamme d'applications des RCSFs est très large. Ceci peut être expliqué par le fait que ces réseaux ont ouvert la possibilité d'effectuer des tâches critiques que les humains ne peuvent pas assurer. Le coût réduit des nœuds capteurs et réseaux de communication a ouvert la voie à de nombreuses applications des RCSFs. Dans ce qui suit nous allons citer brièvement quelques domaines d'applications des RCSFs.

1. Le domaine militaire

La principale motivation derrière la recherche scientifique sur les RCSFs était l'application militaire. Le système de surveillance sonore "Sound Surveillance System (SOSUS)" constitue la première application des réseaux de capteurs. Leurs utilisations remontent à la guerre froide au début des années 1950 du siècle précédent afin de détecter et suivre des sous-marins soviétiques à l'aide de capteurs acoustiques ou hydrophones [16]. Ensuite un programme de recherche sur les réseaux de capteurs distribués a été ensuite tracé par l'agence de recherche de nouveaux projets défense "Defense Advanced Research Projects Agency (DARPA)", qui est une agence du département de la défense des États-Unis, chargée de la recherche et développement des nouvelles technologies destinées à un usage militaire. Plusieurs applications militaires utilisant les RCSFs existent, on cite :

- La surveillance des champs de batailles ou de frontières,

- La détection des attaques chimiques,
- La détection de radiation

Un exemple d'application militaire est celui de BorderSense [17] qui représente une architecture de réseau de capteurs sans fil hybride pour les systèmes de patrouille frontalière. Un autre exemple d'application est le iWEDS [18] qui est un système de détection d'explosifs utilisant les RCSFs, proposé pour aider la police et les forces militaires en Inde, et enfin smart mine [19], les mines intelligentes qui sont des mines intégrées avec des capteurs sonores et des LED/buzzers.

2. La surveillance environnementale

Traditionnellement, le suivi environnemental est obtenu par un petit nombre d'unités de détection. Coûteuses, ces unités sont de haute précision et permettent la collecte des données à la fin de l'expérience directement à partir de l'équipement. La mise en place d'un RCSF fournit une solution alternative en déployant un plus grand nombre de nœuds capteurs jetables. Les nœuds sont équipés de capteurs avec moins de précision, cependant, le réseau dans son ensemble offre une meilleure résolution spatiale et permet aux utilisateurs d'accéder immédiatement aux données. Plusieurs systèmes de détection et de surveillance environnementale basés sur les RCSFs existent. Les RCSFs représentent une solution largement utilisée dans ce domaine, elle peut être utilisée pour :

- Les prévisions météorologiques,
- La surveillance de désastre naturel,
- Le suivi des animaux,
- La surveillance et la détection des inondations,
- Les études de pollution,
- La détection des risques de coulée de débris,
- L'agriculture de précision.

Le contrôle environnemental est un candidat naturel pour appliquer les RCSFs. Un exemple de ces applications est le travail rapporté par [20] où ils ont conçu un réseau réactif et événementiel dédié à la surveillance environnementale de l'humidité du sol. Un autre exemple d'application des RCSFs dans le domaine environnemental est celui du travail réalisé par [21] où un nouveau système est conçu pour effectuer une surveillance micro-environnementale en profitant des avantages des RCSFs.

3. Le suivi médical

Les RCSFs sont largement utilisés dans le domaine médical. Ils offrent de nouveaux services tels que la surveillance des signes vitaux des patients. Les capteurs sans fil portables peuvent signaler périodiquement les niveaux de plusieurs paramètres corporels, par exemple la température, la pression artérielle, l'électrocardiogramme (ECG) et le niveau d'insuline, pour un diagnostic précis. Les nœuds capteurs peuvent être utilisés pour la localisation, la surveillance et l'identification des patients à l'hôpital ou à domicile. Ce fait permet des soins à distance, par lesquels les patients, les

personnes handicapées ou âgées peuvent bénéficier de soins médicaux à domicile. D'autre part, des alarmes peuvent être activées si des événements critiques sont détectés (par exemple, les capteurs d'accélération suggèrent qu'une personne est tombée). Les soins médicaux à domicile peuvent exploiter la présence d'une domotique RCSF pour une connectivité améliorée au sein de la maison et avec des hôtes distants.

Parmi les applications existantes dans ce domaine, Sensatex [22] qui est un dispositif de surveillance de santé portable qui intègre un certain nombre de capteurs sur la carte mère portable de Georgia Tech. Un autre exemple est celui de Lifeguard, un projet de l'Université de Stanford qui est un système de surveillance physiologique composé de capteurs physiologiques (ECG/Respiration électrodes, oxymètre de pouls, tensiomètre, sonde de température), un appareil portable avec accéléromètres intégrés (Chronic Pulmonary Obstructive Disorder (CPOD)) et une station de base (Pocket PC). Le CPOD acquiert et enregistre les paramètres physiologiques mesurés par les capteurs.

Un autre domaine d'application qui peut bénéficier de capteurs portables est celui des sports et le fitness. Les capteurs de fréquence cardiaque peuvent par exemple communiquer, soit avec une montre qui peut afficher le résultat de la mesure, ou avec un appareil qui peut stocker les principales statistiques d'une séance d'entraînement. Le RCSF est, dans ce cas, composé d'un petit nombre de nœuds. En outre, il peut également être possible de transmettre ses données à une base de données distante (par exemple, connectée à Internet) où les données enregistrées peuvent être traitées et analysées et sur la base de cela, un "coach virtuel" peut fournir des commentaires et des conseils à l'utilisateur pendant que la session de formation est en cours. Enfin, en plus des paramètres liés à la santé, les capteurs biométriques peuvent rapporter des données à partir desquelles les émotions d'une personne peuvent être obtenues. La détection émotionnelle est un nouveau domaine de recherche intéressant.

4. Les applications commerciales

Des RCSF peuvent être utilisés dans le domaine commercial afin d'aider les commerçants à améliorer le processus de stockage et de livraison des marchandises. On peut par ailleurs, interroger le RCSF prévu à cet effet pour connaître à tout moment la position, l'état et la direction d'une marchandise donnée ou d'une cargaison. Ainsi, un client dans l'attente d'une marchandise donnée pourra alors avoir un avis de livraison en temps réel. En outre, les produits en fin de vie pourraient être mieux démontés et recyclés ou réutilisés avec l'utilisation de RCSF. Des entreprises manufacturières, via des RCSFs pourraient également suivre le procédé de production à partir des matières premières jusqu'à la livraison finale d'un produit donné. Ainsi, grâce à l'utilisation des RCSFs dans le domaine commercial, les entreprises pourraient offrir une meilleure qualité de service tout en réduisant leurs coûts.

5. Les applications industrielles

À mesure que l'infrastructure d'une usine vieillit, les pannes des équipements entraînent de plus en plus de temps d'arrêt imprévus. L'ARC (Advisory Group) estime que 5% de la production en Amérique du Nord est perdue à cause de temps d'arrêt. Parce que les nœuds de capteurs peuvent être profondément intégrés dans les machines et n'est pas une infrastructure, les RCSFs rendent économiquement faisable le

contrôle de l'état des machines et pour assurer un fonctionnement sûr. Les pipelines et les réservoirs vieillissants sont devenus un problème dans l'industrie du pétrole et du gaz. La surveillance de la corrosion à l'aide de processus manuels est extrêmement coûteuse en temps, consommatrice et peu fiable. Un réseau de capteurs de corrosion sans fil peut être déployé de manière économique pour identifier de manière fiable les problèmes avant qu'ils ne deviennent des pannes catastrophiques.

6. Les applications domotiques

Aujourd'hui, nos maisons sont de plus en plus numérisées et intelligentes. L'implémentation des RCSFs est utilisée afin d'avoir une base de données d'informations nécessaires à la gestion de nos maisons modernes. Par exemple dans le domaine de la surveillance de l'habitation on retrouve des capteurs de détection de mouvement, des capteurs de sons, des caméras sans fil, etc... ou encore dans d'autres applications qui permettent de gérer les systèmes domotiques comme le chauffage, la climatisation, l'éclairage etc. . . Les nœuds capteurs forment ainsi un environnement intelligent permettant de fournir toutes les informations nécessaires aux applications de confort, de sécurité et de maintenance de l'habitat. On retrouve par exemple les applications de :

- **Contrôle de la lumière** : un nouvel éclairage peut être contrôlé à partir de n'importe quel interrupteur, ce qui réduit le besoin de nouvelles connexions câblées. Les lumières peuvent également être activées en réponse à une commande d'une télécommande. En outre, ils peuvent être activés automatiquement lorsque les capteurs de la présence et la luminance détectent que les gens se trouvent dans une pièce mal éclairée.
- **Télécommande** : La technologie infrarouge a été utilisée pour la communication sans fil entre une télécommande et des appareils tels que des téléviseurs, la ventilation et la climatisation. Cependant, l'infrarouge nécessite une visibilité directe "Line Of Sight (LOS)" et une courte distance de communication. La technologie "RadioFréquence (RF)" surmonte ces limitations
- **Sécurité et sûreté** : les systèmes de sécurité avancés peuvent être basés sur plusieurs capteurs (par exemple, des détecteurs de fumées, des capteurs de bris de verre et des capteurs de mouvement) pour détecter d'éventuelles situations à risque qui déclenchent des actions appropriées en réponse. Par exemple, les détecteurs de fumée peuvent activer des alarmes incendie.

Dans une maison, le nombre de nœuds peut être de l'ordre de dizaines (et même des centaines dans certaines résidences), et la densité des nœuds peut être élevée. Les nœuds sont principalement statiques, mais certains d'entre eux (par exemple ceux attachés aux télécommandes ou même aux personnes) peuvent présenter une certaine mobilité. Le réseau domotique peut être connecté à d'autres réseaux (par exemple Internet) via une passerelle ou un routeur, selon les protocoles utilisés.

1.2.8 Les problèmes spécifiques des RCSFs

Les RCSFs ont des limites, à la fois du côté du réseau et du côté du nœud capteur. Contrairement aux réseaux traditionnels, les réseaux de capteurs doivent gérer l'énergie,

la bande passante et la puissance de traitement des ressources qui évoluent de manière dynamique. De plus, ce type de réseaux a des contraintes matérielles de conception qui influencent directement ses performances. Ces contraintes ont suggéré une grande partie des problèmes de recherche dans le domaine. Dans ce qui suit nous allons discuter des problèmes majeurs auxquels les RCSFs sont confrontés.

1. Le problème de durée de vie

La durée de vie du réseau peut être vue sous plusieurs angles, par exemple elle peut être vue comme étant l'intervalle de temps entre l'instant de la première mise en marche du réseau et l'instant d'épuisement du premier nœud. Un nœud capteur est alimenté par des ressources énergétiques très réduites, en général déployés dans des zones géographiques difficiles et inaccessibles. Il est pour ainsi dire impossible de remplacer ou de recharger la batterie d'un nœud capteur. Cette contrainte énergétique affecte directement la longévité du RCSF. Le problème de préservation énergétique est posé particulièrement si l'application doit fonctionner longtemps. Des solutions qui permettent d'équilibrer la consommation d'énergie existent, nous citons parmi elles :

- **L'optimisation du déploiement** : La distance entre le nœud capteur et la station de base influe grandement sur l'énergie consommée par la transmission de données. A ce titre, le problème de fond consiste à minimiser ces distances. Il s'agit donc du problème mathématique de la localisation que l'on peut exprimer comme suit :

Comment peut-on distribuer les capteurs de façon qu'ils soient tous à la distance minimale par rapport à la station de base? Cette question relève de l'optimisation multidimensionnelle en mathématiques. Ainsi, pour localiser une station de base par rapport à un ensemble de capteurs répartis dans l'espace plan, il faudra minimiser la fonction [23] :

$$\min\{f(x, z) = \sum_{a \in A} z_a \|x - a\|\} \quad x \in \mathbb{R}^d \quad (1.1)$$

Où $\|x - a\|$ représente la distance entre le capteur considéré "x" et la station de base "a" ; z_a est un poids que l'on peut attribuer à la base "a". Il s'agit donc du problème classique de Fermat-Weber qui consiste à minimiser la somme pondérée des distances à un ensemble de destinations fixes A.

Deux stratégies de déploiement existent, soit le déploiement déterministe ou le déploiement aléatoire. Dans la première stratégie, les nœuds capteurs sont installés manuellement dans des endroits prédéterminés, les données sont alors acheminées de manière prédéfinis, les limites de cette solution se ressentent quand le nombre de capteurs devient important et cette solution devient alors impossible. Dans l'approche aléatoire les capteurs sont éparpillés aléatoirement (exemple : largués d'un avion). Enfin, un mélange de ces deux techniques peut être envisagé.

- **L'utilisation des modèles de mobilité** : il existe des mécanismes de collecte de données à puits mobiles qui permettent d'alléger la charge de trafic des nœuds capteurs et de prolonger la durée de vie du réseau dans les RCSFs.

Cette méthode consiste à déployer un grand nombre de nœuds de micro-relais à faible coût pour le routage des données. Plusieurs travaux de recherche se sont penchés sur cette solution. Dans [24], les auteurs ont proposé l'idée de base qui consiste d'avoir un puits se déplaçant dans la zone du réseau et collectant des données à partir des capteurs. Quatre modèles de mobilités caractéristiques pour la station de base combinés avec différentes stratégies de collecte de données ont été proposés. Les résultats démontrent qu'en tirant parti de la mobilité des puits et en déplaçant le travail des capteurs vers le puits puissant, l'énergie dépensée est considérablement réduite pour relayer le trafic et ainsi prolonger considérablement la durée de vie du réseau.

- **Le contrôle de la topologie** : Les nœuds capteurs sont déployés dans des zones géographiques assez larges en dizaines ou milliers et chacun de ces nœuds possède un nombre important de voisins. Le contrôle de topologie permet de manager les relations entre chaque nœud capteur et son voisinage. Ce mécanisme permet de réduire le nombre de liens de communication en éliminant les redondances et ainsi réduire la consommation énergétique. Beaucoup de travaux ont été réalisés sur ce sujet. La technique du clustering semble être la plus importante approche permettant de rendre la communication efficace. Récemment, l'algorithme de clustering a gagné plus de popularité en raison des avantages tels que le passage à l'échelle, la facilité de regroupement, et le gain dans la consommation énergétique.
- **L'agrégation de données** : la collecte de données est le processus habituel, il est nécessaire de diminuer le nombre de paquets ou de données signalées à la station de base périodiquement. L'agrégation de données est un processus connu qui agrège le grand nombre de paquets en un seul paquet qui sera ensuite transmis à la station de base. La compression des données est également une autre technique qui corrèle les données. L'efficacité de l'agrégation des données et la compression des données est fournie par le clustering.
- **Les méthodes de routage à équilibrage énergétique** : Le routage est l'un des procédés les plus difficiles dans les RCSFs. La mise en direction se fait pour découvrir le meilleur chemin entre le nœud source et le nœud cible. Un switch est utilisé à des fins de routage afin d'envoyer le paquet ou le message du nœud source au nœud cible, une table de direction est créée pour découvrir la meilleure façon appropriée pour le faire. Le message ou le paquet est ensuite transmis par cette voie. Il est particulièrement important de concevoir un protocole de routage pour les RCSFs afin que les données de détection puissent être transmises efficacement au récepteur.
- **L'intégration avec d'autres technologies** : parmi les solutions prometteuses dans la réduction de consommation d'énergie on retrouve l'intégration des RCSFs avec d'autres technologies comme le Cloud computing. Cette solution permet de rallier les avantages du Cloud computing aux RCSFs.

2. Limitation de la capacité de stockage

Due à la limitation en espace de stockage dans un nœud capteur, les données sont transférées à une station de base centralisée. Les techniques telles que l'agrégation et

la compression réduisent la quantité des données transférées, diminuant ainsi le taux de communication et les coûts énergétiques. Ces techniques sont importantes pour le temps réel ou des applications basées sur des événements, mais elles ne sont pas toujours suffisantes. Les applications fonctionnant selon une approche d'interrogation permettent de décider d'une manière sélective quelles données sont importantes à collecter. L'optimisation du stockage des capteurs devient importante dans ce cas lorsque des données massives sont stockées dans le temps. Étant donné que l'espace de stockage est limité et la communication est coûteuse, un modèle de stockage est nécessaire pour satisfaire les contraintes de stockage et les exigences de requête. Plusieurs méthodes de stockage existent comme "Graph EMbedding (GEM)" [25], TSAR : architecture de stockage de capteur à deux niveaux [26], stockage multi-résolution [27].

3. Le problème de bande passante

La bande passante est l'une des contraintes majeures des réseaux de capteurs sans fil. La limitation de la bande passante, peut engendrer de grandes latences et par la suite un faible niveau de qualité de service. En effet, plus la charge ou la densité du réseau est grande, moins vite un nœud accédera au réseau pour transmettre ses données. L'un des plus grands problèmes des réseaux de capteurs sans fil est de savoir comment allouer et reconfigurer la bande passante disponible pour fournir une utilisation efficace de cette dernière dans le réseau. Comme la bande passante disponible dans les réseaux de capteurs sans fil est généralement insuffisante pour satisfaire le trafic hétérogène, de nombreuses mesures ont été prises pour s'adapter au trafic en constante augmentation. Les RCSFs qui partagent la même structure de base nécessitent une allocation appropriée de la bande passante afin que les nœuds du réseau puissent prendre une décision sur l'allocation de la bande passante en fonction du flux de données dans le réseau. Le mécanisme d'allocation de bande passante au niveau du nœud devrait être capable de prendre en charge le débit de données et le type de trafic qui y circule. Il devrait avoir la capacité de modifier la bande passante allouée dans le réseau en fonction des phénomènes observés. Généralement, le problème d'allocation de largeur de bande a été résolu via diverses techniques de multiplexage telles que "Time Division Multiple Access (TDMA)", "Code Division Multiple Access (CDMA)" et "Frequency-Division Multiple Access (FDMA)" dans les réseaux cellulaires et autres réseaux câblés traditionnels. Ces approches ont également été héritées des réseaux de capteurs sans fil. Mais les limitations de la consommation d'énergie et des coûts de maintenance ont restreint l'utilisation d'autres schémas de multiplexage à l'exception des schémas "Carrier Sense Multiple Access (CSMA)" et TDMA dans les applications RCSF du monde réel.

4. Le problème de sécurité des données

Les RCSFs sont largement déployés dans des domaines impliquant des informations sensibles et des services essentiels tels que l'armée. Dans ces domaines, il est nécessaire de sécuriser la communication et les données des utilisateurs. En outre, le grand déploiement des RCSFs dans des environnements difficiles augmente leurs expositions à des attaques malveillantes, telles que les attaques par déni de service et nœuds non autorisés. En plus, le support sans fil facilite l'écoute indiscrete et

l'injection de paquets contradictoire pour compromettre les fonctions des réseaux. Tous ces facteurs rendent extrêmement important le soutien de la sécurité pour les RCSFs. Dans de tels cas, les services de chiffrement de données sont nécessaires pour masquer les informations critiques. Les services de surveillance des RCSFs sont nécessaires pour la protection des attaques externes et le signalement des problèmes d'accès et d'utilisation anormale. De plus, il est important de déployer et d'appliquer une authentification forte et un contrôle d'accès pour assurer la fonctionnalité d'un RCSF sûr pour les utilisateurs. D'autre part, les nœuds capteurs ont une puissance et un traitement limités et par conséquent, les mécanismes de sécurité standards qui consomment beaucoup de ressources (traitement et communication) ne conviennent pas. Par exemple, les processus de chiffrement/déchiffrement nécessitent une puissance de calcul et une authentification très élevée puisque chaque correspondance et chaque utilisateur nécessite également une charge de communication élevée. Ces défis augmentent la nécessité de développer des solutions complètes et sécurisées qui assurent une protection plus large, tout en maintenant les performances souhaitables du réseau et une consommation d'énergie acceptable. Plusieurs solutions de sécurité dédiées aux RCSFs existent. On note parmi elles les solutions à clé partagée, le chiffrement [28], TinySec [29], E. SPINS [30].

5. Le problème de localisation des capteurs

Dans un RCSF, les nœuds capteurs déployés n'ont pas une connaissance préalable de leur emplacement physique. Le problème de la détermination de l'emplacement (position) du nœud est appelé le problème de localisation. Il existe des méthodes de localisation incluant le système de positionnement global (GPS), le nœud de petite ancre et la localisation de proximité. Équiper les nœuds capteurs d'un récepteur GPS est une solution simple au problème de localisation. Cependant, un tel système peut ne pas fonctionner lorsque les capteurs sont déployés dans un environnement rempli d'obstacles tels que des zones de feuillage denses. La méthode des nœuds de petite ancre utilise des nœuds de balise (ancre), qui connaissent leur propre position, pour aider les capteurs à déterminer leur emplacement. Cette méthode a son défaut car elle n'évolue pas bien dans les grands réseaux et des problèmes peuvent survenir en raison des conditions environnementales. La localisation basée sur la proximité utilise des nœuds voisins pour déterminer leur position, puis agir comme balises pour d'autres nœuds.

6. Le problème de passage à l'échelle

Le nombre de nœuds de capteurs déployés dans l'étude d'un phénomène peut être de l'ordre de centaines ou de milliers. Ce nombre peut atteindre une valeur extrême de millions en fonction de la demande des applications. De nouveaux schémas doivent être capables de travailler avec ce nombre de nœuds, ils doivent également utiliser la haute densité des capteurs du réseau. La densité peut varier de quelques nœuds capteurs à quelques centaines de nœuds capteurs dans une région, qui peut avoir un diamètre inférieur à 10 m, la densité μ peut être calculée selon [31] à l'aide de l'équation (1.2).

$$\mu(R) = \frac{(N * \pi * R^2)}{A} \quad (1.2)$$

Où N est le nombre de nœuds capteurs dispersés dans la région A et R est la portée de transmission radio. Fondamentalement, $\mu(R)$ donne le nombre de nœuds dans le rayon de transmission de chaque nœud de la région A .

1.3 Le Cloud computing

"Cloud Computing", pour le dire simplement, signifie "Internet Informatique". Internet est généralement visualisé comme des nuages; d'où le terme "Cloud computing" pour le calcul effectué à travers l'Internet. Grâce à cela, les utilisateurs peuvent accéder à la base de données ressources via Internet de n'importe où, aussi longtemps qu'ils y'a besoin, sans se soucier de la maintenance ou de la gestion des ressources réelles. Il facilite également l'accès des utilisateurs aux ressources partagées, des informations et des logiciels via des ordinateurs ou d'autres appareils via Internet en utilisant un navigateur Web. Les utilisateurs peuvent également développer, gérer et organiser les applications partagées. Par exemple : Gmail, Google Docs, YouTube, Hotmail etc. sont des applications du Cloud computing en direct. Le mot Cloud est utilisé pour décrire ce type de calcul à cause de la métaphore utilisée pour décrire les réseaux.

1.3.1 Définition du Cloud computing

Un Cloud sous-tend toute la technologie qui est au-dessus de l'utilisateur qui ne sait pas qu'elle existe et n'a pas besoin de le savoir. Il n'y a pas d'officielle définition de ce qu'est le Cloud computing. Pour répondre à cette question nous ferons référence à certaines définitions proposées par des organisations importantes.

- Pour l'Institute of Electrical and Electronics Engineers (IEEE), le Cloud computing est "Un paradigme dans lequel les informations sont stockées en permanence sur des serveurs internet et mis en cache temporairement sur les clients".
- Pour l'encyclopédie la plus célèbre d'Internet, Wikipedia, le Cloud computing : "consiste à utiliser des serveurs informatiques distants par l'intermédiaire d'un réseau, généralement Internet, pour stocker des données ou les exploiter. Les services de calcul, de logiciel, d'accès aux données et de stockage qui n'exigera que l'utilisateur final connaisse l'emplacement physique et la configuration du système qui fournit les services".
- L'université de Berkeley, l'une des universités les plus importantes du monde définit le terme Cloud computing comme : "Le Cloud Computing fait référence à la fois aux applications fournies en tant que services sur Internet et les logiciels, matériels et systèmes des centres de données qui fournissent ces services".

Nous pouvons donc dire que le Cloud computing est un paradigme de calcul basé sur la consommation de ressources, applications, matériel ou calcul, offerts par Internet et consommés à la demande. Ces services sont publics ou privés, gratuits ou non et ont un accord de niveau service qui les réglementent.

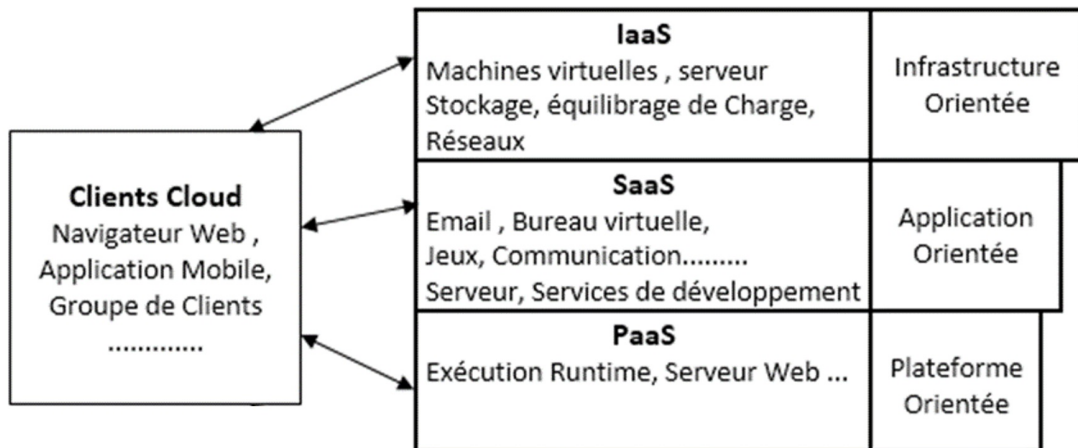


FIGURE 1.4 – Services des clouds [2]

1.3.2 Types de services du Cloud computing

Le Cloud fournit trois types de services basés sur les différentes exigences des clients comme montré dans la figure 1.4.

a) Infrastructure en tant que service (IaaS)

Ce modèle permet le partage de plusieurs ressources physiques sur le réseau. Le principal objectif de l'IaaS est de fournir un accès rapide au serveur, au stockage et au réseau par les applications et le système d'exploitation. Ainsi, il offre une infrastructure simple des services à la demande à l'aide de l'Application Programming Interface (API). L'utilisateur n'a pas besoin de gérer le matériel principal dans l'infrastructure Cloud, mais il peut contrôler le serveur, l'application et le système d'exploitation. Un exemple d'IaaS est celui d'Amazon Elastic Cloud Computing (EC2).

b) Plateforme en tant que service (PaaS)

Permet à l'utilisateur d'organiser, créer ou acquérir des applications en utilisant des langages de programmation, des bibliothèques, des services et des outils pris en charge par le fournisseur sur l'infrastructure Cloud. Il offre un grand niveau d'intégration de plateforme pour les systèmes multi-locataires. Lorsque les utilisateurs ne sont pas en mesure de gérer le réseau, les serveurs, les systèmes d'exploitation et le stockage, ils optent pour la plateforme en tant que service. Certains des exemples de PaaS sont Force.com, Google App Engine et Microsoft Azure.

c) Logiciel en tant que service (SaaS)

Dans le logiciel en tant que service, la facilité offerte à l'utilisateur consiste à utiliser différents logiciels fonctionnant sur une infrastructure Cloud. Cela conduit à une meilleure

utilisation de l'espace de stockage de notre poste de travail. Les logiciels sont accessibles à partir de divers périphériques clients via une interface client, telle qu'un navigateur Web (par exemple, messagerie électronique basée sur le Web), ou une interface de programme. Des exemples de SaaS incluent : Salesforce.com, Google Apps.

1.3.3 Modèles de Cloud computing

Le Cloud propose trois types de modèles : le Cloud public ouvert à tous ; le Cloud privé où l'accès n'est autorisé qu'à l'utilisateur de cet espace privé et le Cloud hybride qui effectue une tâche de compromis des deux types de services fournis par le Cloud public et privé. Il existe quatre modèles de déploiement définis comme suit [32] :

1. Cloud privé

Cette structure est destinée à toute organisation privée. Généralement, cette infrastructure Cloud est gérée par l'organisation elle-même ou une tierce partie et peut exister sur site ou hors site.

2. Cloud public

L'infrastructure Cloud est mise à la disposition du grand public ou un grand groupe industriel et appartient à une organisation vendant des services Cloud.

3. Cloud communautaire

L'infrastructure Cloud est partagée par plusieurs organisations et soutiennent une communauté spécifique qui a des préoccupations communes (par exemple, missions, exigences de sécurité, considérations de politique et de conformité). Tous les membres ont un accès égal à cette infrastructure. Elle peut être gérée par ces organisations ou une tierce partie, et peut exister sur place ou hors des lieux.

4. Cloud hybride

Deux ou plusieurs des modèles de Cloud ci-dessus fournissant conjointement des services efficaces au client constituent une infrastructure hybride où une partie de l'art peut être limitée pour un public particulier et certaines parties sont accessibles à tous.

1.3.4 Avantages du Cloud computing

Le Cloud Computing offre de nombreux avantages à ses utilisateurs finaux. Voici quelques bénéfices de cette technologie :

- **La rentabilité** : le Cloud computing est la meilleure méthode à faible coût pour l'utilisation de logiciels, de services de plateforme et d'infrastructure. Il est basé sur le principe du "paiement à l'utilisation", donc pas besoin de payer de l'argent supplémentaire pour l'achat de licences de logiciels ou d'infrastructure, etc...
- **Le stockage presque illimité** : Le Cloud computing offre une capacité de stockage illimitée, donc il n'existe aucune limitation pour stocker des données.
- **La sauvegarde et restauration** : Dans le Cloud computing, il est facile de sauvegarder et de récupérer.

- **Intégration automatique des logiciels** : dans le Cloud computing, le déploiement et la mise en fonctionnement d'un service nécessite peu de temps, ainsi que l'intégration des logiciels qui est généralement quelque chose qui se produit automatiquement. L'utilisateur n'a donc pas besoin de donner plus d'effort pour personnaliser l'application selon ses besoins.
- **L'accès facile aux informations et services** : dans le Cloud, une fois l'utilisateur enregistré, il peut utiliser de n'importe où ses données sans aucune limitation de localisation.
- **La conséquence positive sur l'environnement** : Le Cloud Computing, basé sur la virtualisation de serveur, la mutualisation de la puissance de calcul et la flexibilité des services s'inscrit dans une démarche éco-responsable.
- **La réversibilité** : La restitution de l'intégralité des données d'une entreprise est garantie par les fournisseurs prévoyant dans leur contrat une clause de réversibilité.

1.3.5 Désavantages du Cloud computing

Malgré ses nombreux avantages, le Cloud computing peut avoir quelques inconvénients nous citons :

- **La sécurité dans le Cloud** : La sécurité est le principal problème du Cloud computing. C'est le principal frein à l'adoption du Cloud car les utilisateurs du Cloud peuvent être confrontés à des questions relatives à la sécurité en termes d'authentification, de confidentialité et d'intégrité.
- **Problèmes techniques** : La latence et la vitesse des réseaux peuvent facilement être un goulot d'étranglement. La performance du système peut être affectée spécialement dans le service IaaS, où nous avons besoin de grandes quantités de transfert de données. Nous rencontrons également deux autres points de défaillance : la connexion de la propre organisation et la connexion du fournisseur.
- **Le temps d'indisponibilité possible** : Le Cloud computing rend la petite entreprise dépendante de la fiabilité de sa connexion Internet.

1.4 Conclusion

Dans ce chapitre nous avons passé en revue la technologie des RCSFs ainsi que celle du cloud computing. Dans chacune de ces deux technologies, nous avons présenté leurs caractéristiques, leurs architectures, leurs domaines d'application et leurs avantages et inconvénients. En outre, nous avons exposé les limitations des RCSFs et comment peut-on tirer profit de la technologie du cloud computing pour se remédier à ces limitations.

Par ailleurs, l'intégration de ces deux technologies domine le marché dans de nombreuses applications en temps réel. De nombreux analystes prévoient qu'il y aura un changement majeur vers les architectures basées sur les capteurs dans le Cloud dans les années à venir.

Chapitre 2

État de l'art sur l'intégration RCSF-Cloud

Chapitre 2

État de l'art sur l'intégration RCSF-Cloud

2.1 Introduction

Les RCSFs sont responsables d'une production massive de données. La gestion efficace de ces données n'est pas une chose aisée. Accablés par les limitations des RCSFs en termes de puissance de calcul, de capacité de stockage, de durée de vie limitée et de flexibilité de la communication ; les utilisateurs des RCSFs se contentaient de ce que cette technologie pouvait leur offrir. Avec l'avènement de l'informatique en nuage (Cloud) au début des années 2000, munie de tous ses points forts telles que la puissance de stockage et de calcul, des solutions sont venues constituer un support pouvant résoudre les problèmes propres aux RCSFs. A partir de là, l'idée d'intégrer les RCSFs au Cloud a germé et on s'est rendu compte que les limitations des RCSFs constituaient les points forts du Cloud.

Ce chapitre présente une synthèse des travaux en relation avec l'interconnexion des RCSFs au Cloud, ce qui permettra de mieux présenter les motivations de ce travail. On signale d'emblée que les activités de recherche dans le domaine d'intégration des RCSFs au Cloud ont considéré différentes conceptions que nous décrirons ci-après.

2.2 Première conception

Elle s'appuie sur le principe de considérer les RCSFs et le Cloud comme deux entités distinctes. Dans ces approches, les travaux de recherche sont s'orientés sur la manière dont le Cloud interagit avec les RCSFs pour améliorer leurs performances. Plusieurs travaux de recherche ont été effectués sur ce sujet, on cite :

W. Kurschl et al. [33] où est présentée une combinaison entre les RCSFs et le Cloud. Le modèle proposé est basé principalement sur des canaux et des filtres constituant le système hébergé sur l'infrastructure Cloud. Les canaux ne transforment pas les données, mais généralement les mettent en mémoire tampon et fournissent un mécanisme d'interconnexion des filtres. Les filtres effectuent un traitement et une transformation spécifiques sur les données d'entrée, tels que le raffinement ou la suppression des données.

Le travail de P. Zhang et al. [34] repose sur une nouvelle architecture qui permet d'améliorer les performances des RCSFs à l'aide du Cloud. Dans cette architecture un Cloud agit en tant que puits virtuel appelé nœud Cloud avec de nombreux points de

collecte qui rassemblent des données mesurées à partir des capteurs. Le RCSF est divisé en plusieurs "zones" et il existe un nombre spécifique de nœuds capteurs dans chaque zone. Ces nœuds sont responsables de la collecte des données dans leur zone. La topologie de l'organisation des capteurs dans la zone est soit une topologie plate ou une topologie hiérarchique. Les données de détection sont stockées et traitées de manière distribuée dans le Cloud. Les résultats de simulation ont montré que l'architecture proposée améliore les performances des RCSFs, en termes de réduction de l'erreur de transmission de paquets, de diminution du nombre de sauts de bout en bout et d'amélioration de l'efficacité de la consommation d'énergie.

T. Wang et al. [35] proposent de résoudre le problème de collection de données dans une intégration entre RCSF et le Cloud en considérant la réduction de la consommation d'énergie et le délai de réception des données. Un schéma de collecte de données a été élaboré en utilisant plusieurs puits mobiles responsables chacun d'un groupe de capteurs afin d'améliorer le temps de collecte de données. Un algorithme d'ordonnancement et d'adaptation temporelle "Time Adaptive Schedule Algorithm (TASA)" est élaboré. Cet algorithme contient trois parties :

1. La première partie se concentre sur la conception de la partition du réseau et de la route de livraison.
2. La deuxième partie est présentée sous-forme d'un algorithme appelé algorithme de sélection des points d'interrogation. En effet, certains capteurs sont sélectionnés pour servir de points d'interrogation dans chaque secteur.
3. La troisième partie est un algorithme, appelé algorithme de planification temporelle. Il permet d'ajuster le nombre de points d'interrogation en fonction du temps de livraison et du nombre de capteurs dans chaque secteur. Ce nombre sera ajusté pour être équilibré.

Les résultats de simulation ont démontré que l'algorithme proposé peut ajuster le délai de livraison et réduire considérablement la consommation d'énergie en améliorant la longévité du système, ce qui contribue à une meilleure intégration des réseaux et du Cloud et notamment les réseaux sociaux.

Un autre travail traitant le même problème de collecte de données depuis les RCSFs jusqu'au Cloud a été élaboré par T. Wang et al. [36]. Dans ce travail, le même principe d'utilisation de puits mobiles est utilisé. Ces puits mobiles constituent une structure d'informatique géo-distribuée (en anglais : Fog computing) et sont munis de ressources physiques supérieures à celles des puits traditionnels. Cette structure permet donc de combler en partie les insuffisances des RCSFs par rapport au Cloud. Cette solution maximise le débit et minimise la latence de transmission ; elle permet aussi de réduire la consommation d'énergie.

D. Zhang et al. [37] ont présenté une nouvelle méthode de routage multicast avec un minimum de transmissions pour les RCSFs au service Cloud dédié aux applications industrielles. L'algorithme DMTMRA a été proposé en se basant sur le chemin offset back-off et en profitant du caractère de broadcast des communications sans fil. L'algorithme établi choisi les liens de transmission aux nœuds avec le plus grand nombre de connexions multicast. Le modèle de réseau de capteurs sans fil est en multi-sauts, statique ou avec

très peu de changement sur une région plane à deux dimensions. Une comparaison du protocole proposé avec les protocoles DODMRP et ODMRP a été établie et les résultats ont montré que le protocole proposé permet d'économiser la consommation en énergie grâce à la réduction du nombre de transmissions, à la minimisation en bande passante et du délai.

N. Javaid et al. [38] présentent deux protocoles de routage pour la gestion économique en énergie et l'écartement des nœuds vides dans un réseau sous-marin de capteurs sans fil reliés à un Cloud. Il s'agit des protocoles : 'Non coopérative AVN-AHH-VBF' et 'Coopérative CoAVN-AHH-VBF'. Dans le premier protocole "AVN-AHH-VBF" le réseau passe dans un premier temps par une première phase d'initialisation où tous les nœuds doivent connaître leurs nœuds voisins et en second lieu le protocole supprime les paquets redondants afin d'éviter la répétition dans les transmissions et ainsi économiser de l'énergie dans le réseau. Ensuite, il fait la sélection efficace du nœud suivant. Dans le second protocole "CoAVN-AHH-VBF", la phase d'initialisation reste la même que celle du premier protocole, cependant, le choix du prochain saut se fera à partir du moment où un paquet de données est généré. Après une série de tests la suppression des paquets redondants aura lieu. Les schémas proposés sélectionnent efficacement les nœuds ré-expéditeurs sur les bases de moindre profondeur dans le pipeline et les régions vers la destination dans la plage du nœud source. Les résultats de simulation ont montré que les deux schémas sont efficaces en termes de coût de consommation d'énergie par paquet, un rapport de livraison élevé et un délai acceptable.

D. McLoughlin et al. [39] ont pensé à la conception d'un système basé sur le Cloud mettant diverses plateformes de simulation à la disposition de plusieurs utilisateurs souhaitant tester les protocoles de routage des RCSFs ou encore les contraintes matérielles des différentes plateformes RCSF. Cette solution consiste à créer le banc d'essai de simulation des RCSFs sur une plateforme virtuelle basée sur le Cloud afin de bénéficier de ses avantages. Les chercheurs ont utilisé un serveur Dell PowerEdge R720 afin d'héberger la plateforme virtuelle. L'accès à ce système nécessite l'installation d'un VMware. Cela accélère considérablement le processus d'apprentissage initial des différents simulateurs et offre aux utilisateurs l'accès à de nombreux autres environnements de simulation RCSF dotés de ressources informatiques et d'un stockage considérable, tout en réduisant les coûts de simulation du protocole et les besoins en matériel qui sont considérablement réduits.

F. Baza et al. [40] proposent une solution permettant de fournir des données fiables, sans erreurs de collecte de données par un RCSF et d'envoyer ensuite vers l'application Capteur-Cloud. Le but de ce travail est d'éviter la panne d'un système malgré l'occurrence de fautes. Un mécanisme de détection et de reconstitution de fautes est mis en place afin d'assurer que les données sont envoyées à la station de base en toute sécurité et fiabilité, pour ensuite être stockées correctement dans les serveurs Cloud. Un protocole FTSC est élaboré, il permet d'organiser les nœuds capteurs en groupes constitués de nœud principal doté du plus grand niveau d'énergie, de nœud de relais au nœud principal et de nœud subordonné. Ces nœuds vont recevoir les données capturées par leurs groupes respectifs et effectueront une agrégation sur ses données afin de réduire leurs flux sur le réseau. Ceci permettra de prolonger la durée de vie du réseau sur le long terme. Après avoir agrégé les données, le capteur principal va envoyer sa donnée finale à son nœud de relais, ce dernier décidera d'envoyer ou non la donnée à la station de base. Les résultats de simulations ont

montré que la solution proposée permet de réduire le taux de transmission de données et de détecter jusqu'à 90% de pannes dans les nœuds. Cependant, ces résultats montrent aussi une plus grande consommation d'énergie dans le réseau à cause des messages supplémentaires échangés entre les nœuds et les calculs additionnels effectués. La solution offre plus de précision et de fiabilité.

2.3 Deuxième conception

Dans cette conception les travaux de recherche se sont focalisés sur la gestion des données récoltées par les RCSFs. Les efforts des chercheurs sont dirigés vers le traitement, le stockage et la sécurisation des données. Dans ce qui suit nous allons citer quelques travaux en rapport avec cette conception.

2.3.1 Stockage et traitement des données

L'idée principale de ces approches s'appuie sur la façon dont les données capturées sont obtenues par les utilisateurs. Différents travaux ont été proposés avec des architectures qui gèrent la manière avec laquelle les données sont stockées et traitées après leurs collectes par les RCSFs.

a) Les RCSFs comme entité physique perçue par l'utilisateur via le Cloud

Dans ces approches les RCSF appartiennent à des organisations spécifiques. Les données récoltées sont la propriété de ces organisations et utilisées généralement par leurs propres applications. Dans ce qui suit nous citons quelques travaux réalisés dans ce contexte.

M. Hassan et al. [41] ont proposé une architecture logicielle (Framework) de connexion entre capteurs et Cloud afin de traiter, stocker et analyser des données volumineuses générées par les RCSFs. Ce travail publié en 2009 fût semble-t-il le premier à traiter l'interconnexion entre ces deux technologies (RCSF et Cloud). Le Cloud à cette époque-là n'avait pas encore adressé le problème d'intégration avec les RCSFs. A cet égard, il n'existait encore aucune infrastructure pour gérer cette intégration.

L'architecture logicielle proposée dans ce travail repose sur le principe de publication/souscription. Un modèle de publication/souscription à base de contenu encapsule les données des capteurs dans des événements et procure les services de publication et de souscription pour un échange de données asynchrone dans un système d'entité. Dans cette architecture les données capturées passent depuis une passerelle vers le courtier de publication/souscription. Ce dernier délivre les informations aux consommateurs d'application SaaS. En plus du courtier publication/souscription, trois autres composants ont été intégrés dans cette architecture :

1. Un médiateur, responsable de la négociation au sein du fournisseur Cloud et du management des opérations dans une organisation virtuelle "Virtual organisation (VO).
2. Un référentiel de politique, responsable du stockage des services web, des médiateurs et des entités virtuelles.

3. Un agent de collaboration qui représente un module de découverte de ressources dans l'environnement des fournisseurs Cloud collaborant.

Un algorithme de correspondance d'événements, rapide et évolutif, appelé correspondance d'index de groupe statique est proposé "Statistical Group Index Matching (SGIM)". Les utilisateurs enregistrent leurs informations et leurs abonnements auprès de diverses applications SaaS qui les transfèrent ensuite au registre du courtier publication/souscription. Lorsque des données de capteurs parviennent au système à partir de passerelles, le composant de surveillance et de traitement d'événements/flux "Stream monitoring and processing component (SMPC)" dans le courtier publication/souscription, détermine s'il doit être traité ou simplement stocké pour un envoi périodique ou pour une livraison immédiate. Les données capturées s'écoulent comme suit :

1. Si les données des capteurs nécessitent une livraison périodique/d'urgence, l'analyseur détermine les applications SaaS auxquelles les événements appartiennent, puis les transmet au diffuseur avec les identifiants d'application,
2. Le diffuseur, à l'aide de l'algorithme de correspondance d'événements, trouve les abonnés appropriés pour chaque application et fournit les événements à utiliser,

D'autre part, l'approche permet également une collaboration dynamique avec d'autres fournisseurs de Cloud à l'aide d'une entité de collaboration dynamique d'organisation virtuelle. Des évaluations sur les performances de l'algorithme proposé sont établies en comparaison avec des algorithmes existants dans un scénario d'application de soins de santé omniprésent dans le Cloud. Les expériences ont montré que l'algorithme proposé fonctionne mieux (en cas de distribution uniforme) que ceux existants en termes d'efficacité et d'évolutivité. Le document ne fournit pas une compréhension claire des composants et les étapes qui participent à la construction du cadre proposé. Serrano et al. [3] présentent une solution qui intègre le Cloud et les RCSFs pour le stockage, le traitement et la protection des données des RCSFs. Ces auteurs proposent dans leur architecture un middleware (intergiciel) appelé "Cloud Sensor Services", qui facilite le développement de l'arrière-plan des RCSFs, le middleware est auto-évolutif grâce aux avantages offerts par le Cloud, modulaire dans sa conception pour pouvoir ajouter facilement de nouvelles fonctionnalités et flexible pour s'adapter aux différents Cloud fournisseurs. Quatre grands services sont proposés par cette solution :

1. Le premier est un service de stockage de données des capteurs où l'on retrouve des opérations de création, de libération d'espace mémoire et l'énumération de l'espace de stockage disponible.
2. Le deuxième est un service d'information qui manipule toutes les informations concernant les capteurs utilisés par le middleware. Il contient des opérations d'inscription, de désinscription de capteurs et des opérations d'énumération de tous les capteurs enregistrés.
3. Le troisième service se focalise sur les propriétés des capteurs ; il offre des opérations de chargement et téléchargement de propriété élargie des capteurs tel que le niveau de batterie, la localisation, etc. . .

4. Le dernier service est un service de données qui contient des opérations de chargement et téléchargement de données de capteurs.

L'implémentation de cette architecture (services capteurs) s'est faite sous l'environnement *C#*, le modèle d'inversion de contrôle a été choisi afin de pouvoir supporter de multiples fournisseurs Cloud. L'idée est de créer un ensemble d'interfaces qui exposent toutes les logiques nécessaires afin de fournir des opérations basées sur le protocole REST et qui livre les services des capteurs et chaque fois que possible il crée des classes avec une logique commune qui permet de minimiser l'effort d'implémenter des adaptateurs.

Des tests d'enregistrement, de chargement et téléchargement de données ont été établis pour comparer l'implémentation de l'architecture sur deux fournisseurs Cloud : Azure (Azure Cloud Services) et Amazon (Amazon Cloud Services). Les tests de chargement sans passage à l'échelle automatique ont montré que l'usage CPU dans le système est pratiquement le même avec 5 utilisateurs ou 500, d'après les auteurs l'usage élevé de la CPU montre une meilleure utilisation des ressources Cloud. L'utilisation de l'espace mémoire aussi ne change pas significativement entre les 5 et 500 utilisateurs, les tests ont aussi montré que le temps de réponse augmente avec le nombre d'utilisateurs et que plusieurs requêtes ont échoué à cause du timeout, ce qui expliquerait l'intérêt du passage à l'échelle automatique. Le passage à l'échelle automatique réduit selon les résultats, la dégradation des performances du système proposé quand le nombre de chargements augmente. Les résultats de comparaison de l'implémentation de la solution ont montré que le fournisseur Cloud Azure était meilleur que celui d'Amazon. La figure 2.1 montre l'architecture globale de la solution implémentée.

V. Rajesh et al. [4] fournissent une solution qui permet de relier les RCSFs à l'internet en utilisant l'architecture orientée services "Service-Oriented Architecture (SOA)" via la technologie Cloud. Le rôle de l'architecture orientée services est de permettre l'interaction entre les RCSFs et le contrôleur d'intégration (par lequel les consommateurs se connectent). Le contrôleur d'intégration est responsable du stockage des données détectées et du système de récupération. Son activité principale consiste à télécharger des données des RCSFs sur Internet via la technologie Cloud. La figure 2.2 illustre le schéma de principe de cette architecture.

K. Ahmed et al. [5] offrent une architecture logicielle d'intégration des RCSFs au Cloud permettant de faciliter le passage des données d'un RCSF vers le Cloud. L'architecture inclue une unité de traitement de données (Data Processing Unit (DPU)), un courtier de publication/suscription, une unité d'abonnement à la demande (RS), une unité de gestion des identités et des accès (IAMU) et un référentiel de données (DR). Les données collectées à partir du RCSF passent par une passerelle vers l'unité de traitement qui conditionnera les données dans un format de stockage, puis les envoie au référentiel de données. Les utilisateurs se connecteraient au Cloud via l'unité de gestion des identités et des accès sécurisés, l'accès est attribué selon la politique stockée sur leurs comptes utilisateurs. Une fois l'accès accordé, les utilisateurs peuvent formuler des demandes d'accès aux données.

Les requêtes seront transmises à l'unité d'abonnement à la demande, celle-ci créerait un abonnement sur la base de cette demande et transmettrait cette souscription au Courtier publication/suscription. Les données reçues dans le Cloud seront identifiées par l'unité de traitement de données qui créera un événement de données publié et l'enverra à une file d'attente d'événements au niveau du courtier de publication/suscription. Lorsqu'un nouvel événement est publié, chaque abonnement est évalué par l'organisateur d'événements.

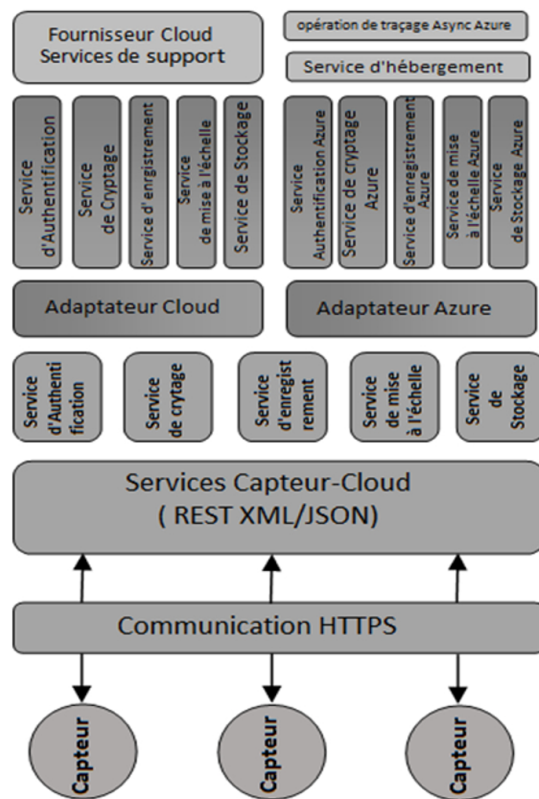


FIGURE 2.1 – Architecture globale de la solution "Cloud Sensor Services" [3]

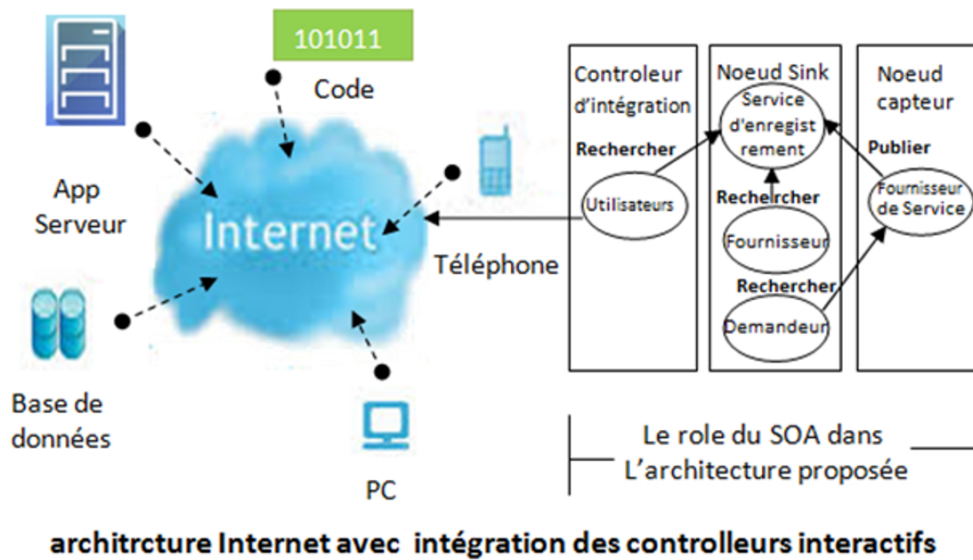


FIGURE 2.2 – Architecture SOA pour Capteur-Cloud [4]

Une fois que le processus de correspondance d'événement trouve une correspondance, les données publiées seront mises à la disposition de l'utilisateur après que le traitement soit effectué, si nécessaire. La figure 2.3 montre le schéma architectural.

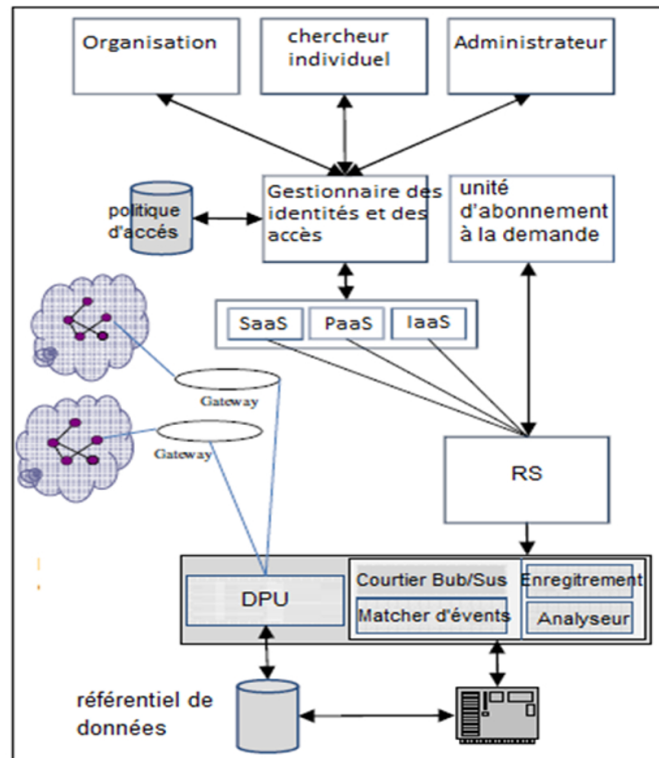


FIGURE 2.3 – Architecture logicielle de Capteur-Cloud [5]

L.P.D. Kumar et al. [42] proposent une intégration des RCSFs homogènes au Cloud afin de faciliter le stockage et le traitement des données. Le principe est le suivant : les données collectées par le RSCF seront envoyées au Cloud afin d'y être stockées ; le concept de passerelle a été utilisé dans cette approche. Les passerelles jouent le rôle de Middleware. Les capteurs ont été organisés en clusters et chaque cluster contiendrait une passerelle. Les données assemblées par les capteurs seront transmises en tant qu'entrée à un réseau de neurones formé sur la base des valeurs reçues. Le réseau de neurones représente un modèle qui peut être utilisé pour prédire la tendance future de ces données. Aussi, il filtre dans les données reçues les anomalies et jette le reste. La taille des données est ensuite réduite par compression et envoyée à la passerelle Cloud. Comme seules les anomalies dans les données sont transmises et stockées, le trafic est réduit. L'approche montre un besoin réduit en bande passante et une amélioration de la durée de vie des RCSFs.

Ce travail a été amélioré par C. Zhu et al. [43] en prenant en compte non seulement le filtrage et la compression des données, mais également la prévision, la surveillance du trafic et la sécurité des données, pour une intégration optimale pas uniquement des RCSFs homogènes mais également des RCSFs hétérogènes avec Mobile Cloud Computing (MCC).

P. You et al. [6] proposent une architecture logicielle combinant les RCSFs au Cloud afin de stocker, gérer et analyser efficacement les données des RCSFs et de les placer pour fournir facilement des services d'informations aux utilisateurs. L'architecture logicielle est connectée au RCSF grâce à un canal sans fil ou des réseaux courtiers à travers lesquels

les capteurs transmettent leurs données. La structure est constituée de trois couches : la couche physique composée des ressources informatiques nécessaires pour l'infrastructure, la couche de virtualisation qui réalise la virtualisation des ressources informatiques et la couche de service d'informations responsable de la fourniture des applications du service d'informations aux utilisateurs. La figure 2.4 montre l'architecture logicielle fournissant les services d'information pour les RCSFs.

R. Piyare et al. [44] présentent une architecture flexible et extensible qui repose sur les web services pour intégrer les RCSFs au Cloud. L'intégration des RCSFs se fait à l'aide de la plateforme "Open.Sen.se" basée sur le Cloud et dédiée pour le stockage, le traitement et l'analyse des données capturées en utilisant le protocole HTTP. Cette architecture est formée de trois couches :

1. La première couche est celle des capteurs directement déployés dans l'environnement dont les données capturées sont directement envoyées à la station de base.
2. La seconde couche est la couche coordinatrice qui est responsable du management des données reçues depuis la station de base. Les données sont temporairement rassemblées et stockées dans des buffers avant d'être envoyées à la couche suivante.
3. La troisième couche appelée couche superviseur permet de publier les données des capteurs sur Internet. Dans cette couche un web service accommodé aux stations de base avec une Web interface offre à l'utilisateur final la possibilité de manager les données et de faire des statistiques.

Les résultats de l'implémentation de cette approche ont démontré que les utilisateurs peuvent accéder aux données des capteurs n'importe où et avec n'importe quel appareil mobile relié à internet et que l'alerte email conçue pour informer l'utilisateur du niveau critique de la batterie d'un capteur a un délai d'une moyenne de 11 secondes, et enfin que les mécanismes d'endormissement des capteurs permettent une amélioration de leur durée de vie.

C. Zhu et al. [45] traitent le problème de non-sensibilisation au contexte entre l'utilisateur mobile et les RCSFs lors de l'intégration des RCSFs au MCC ce qui empêche l'utilisateur mobile d'obtenir les données souhaitées lors de cette intégration, et dégrade également les performances des RCSFs qui agissent comme source de données pour le Mobile Cloud. Une nouvelle architecture logicielle d'intégration des RCSFs au Mobile Cloud est proposée afin d'obtenir les informations sur les caractéristiques de données requises par les utilisateurs mobiles et le statut potentiel des RCSFs. L'architecture proposée effectue :

1. La recommandation de données à l'aide de l'algorithme Apriori. Cet algorithme complet de recherche utilise l'anti-monotonie des ensembles de données.
2. La prédiction de données utilisant le lissage exponentiel secondaire comme modèle de prédiction de données (SESM).
3. La surveillance du trafic de données dans le Cloud.
4. Les données (mesurées) sont recueillies avec une fréquence définie puis analysées pour vérifier s'il y a trop ou peu de données pour une période considérée.

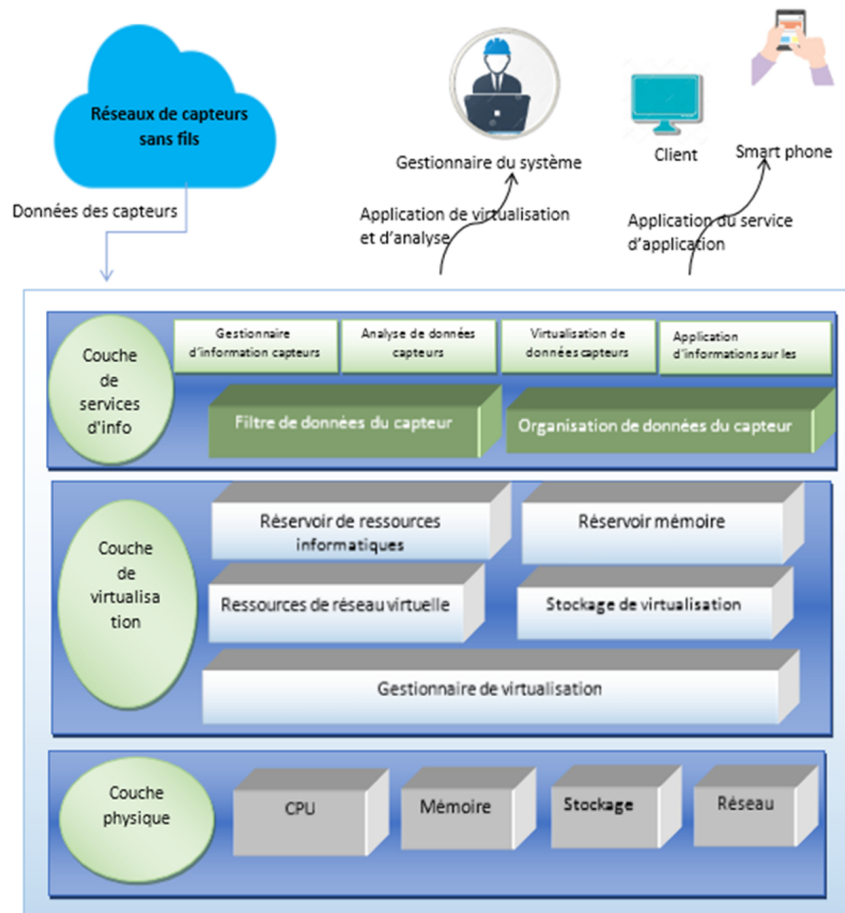


FIGURE 2.4 – Architecture logicielle Cloud fournissant des services d'informations aux RCSFs [6]

Pour l'évaluation, la plateforme "Cloud Jupiter" de l'université de technologie de Dalian située en Chine a été utilisée. Un prototype simplifié a été mis en œuvre pour vérifier l'efficacité du framework. L'évaluation et les résultats analytiques montrent que le framework proposé peut favoriser considérablement les utilisateurs mobiles pour obtenir les données qu'ils désirent quelque soit le problème de non-sensibilisation du contexte.

C. Zhu et al. [46] proposent une nouvelle architecture logicielle pour traiter les données collectées dans un système d'intégration des RCSFs au Mobile Cloud. L'architecture logicielle proposée utilise la passerelle RCSF pour effectuer les traitements suivants :

1. La surveillance du trafic de données : les capteurs collectent les données avec une fréquence définie (par exemple toutes les 30 secondes). Ainsi, la taille des enregistrements de données peut être prise en compte pour vérifier s'il y a trop ou peu de données collectées par les capteurs.
2. Le filtrage selon un ensemble de règles prédéfinies : l'unité de filtrage des données vérifie si les valeurs des données collectées par les capteurs se situent dans une plage acceptable.
3. La prévision : les données sensorielles recueillies considérées dans ce travail sont des séries chronologiques, le modèle de lissage exponentiel secondaire (SESM) est alors

appliqué.

4. La compression et décompression des données appliquées au niveau des passerelles du RCSF et du Cloud a pour but de réduire la perte de paquets due à la congestion du réseau. L'algorithme Deflate qui combine le codage Huffman avec LZ77 compression est utilisé.
5. La sécurité (le chiffrement et le déchiffrement) : le cryptosystème de "Rivest-Shamir-Adleman (RSA)" est utilisée.
6. La recommandation des données où l'algorithme d'Apriori est utilisé.

La passerelle du Cloud effectue la fonction de sécurité et de décompression des données. L'évaluation de l'architecture logicielle proposée a montré la capacité d'améliorer la durée de vie du réseau, les besoins de stockage, les performances de sécurité et de surveillance des RCSFs, la sécurité des données collectées transmises, la réduction du trafic et la largeur de bande requise pour la transmission de données mesurées ainsi que la surcharge de stockage et de traitement dans le Cloud. Cette architecture représente une amélioration du travail déjà réalisé par C. Zhu et al. [46]. Le concept de passerelle aux extrémités RCSF et Cloud a été ajouté de manière assez similaire que celle utilisée par D. Kumar et al. [42].

C. Zhu et al. [43] traitent le problème de sélection des données capturées et leur envoi au Cloud. L'envoi doit être fiable et utile pour l'utilisateur mobile. Les chercheurs proposent un nouveau schéma d'intégration entre RCSFs et Cloud mobile appelé TPSS qui est constitué de deux parties :

1. "Time and Priority-based Selective Data Transmission (TPSDT)" (transmission de données sélective basée sur le temps et les priorités) permettant à la passerelle RCSF de transmettre de manière sélective des données mesurées plus utiles au Cloud, en tenant compte des caractéristiques de temps et de priorité des données demandées par l'utilisateur mobile.
2. Programmation du sommeil basée sur les priorités pour les RCSFs afin de réduire la consommation d'énergie, afin qu'ils puissent collecter et transmettre des données de manière plus fiable.

b) Les RCSFs comme entités virtuelles

La création et le maintien de son propre RCSF fermé est très coûteux. Très vite l'idée de partager les ressources d'un RCSF sur une même infrastructure où les utilisateurs peuvent partager divers types de capteurs physiques facilement, sans se soucier de leur emplacement physique et pouvant fournir beaucoup de nouveaux services a vu le jour. On parle alors de virtualisation. Selon A. Alamri et al. [47], la virtualisation des RCSFs dans le Cloud apporte des avantages en termes de provisionnement d'une puissance de traitement évolutive sur des données massives provenant de différents RCSFs permettant divers types d'analyses, une collaboration de divers réseaux de capteurs physiques, une mise à disposition gratuite d'une capacité de stockage et de traitement de données accrue, un provisionnement dynamique et automatique de services, une optimisation des ressources et un temps de réponse rapide.

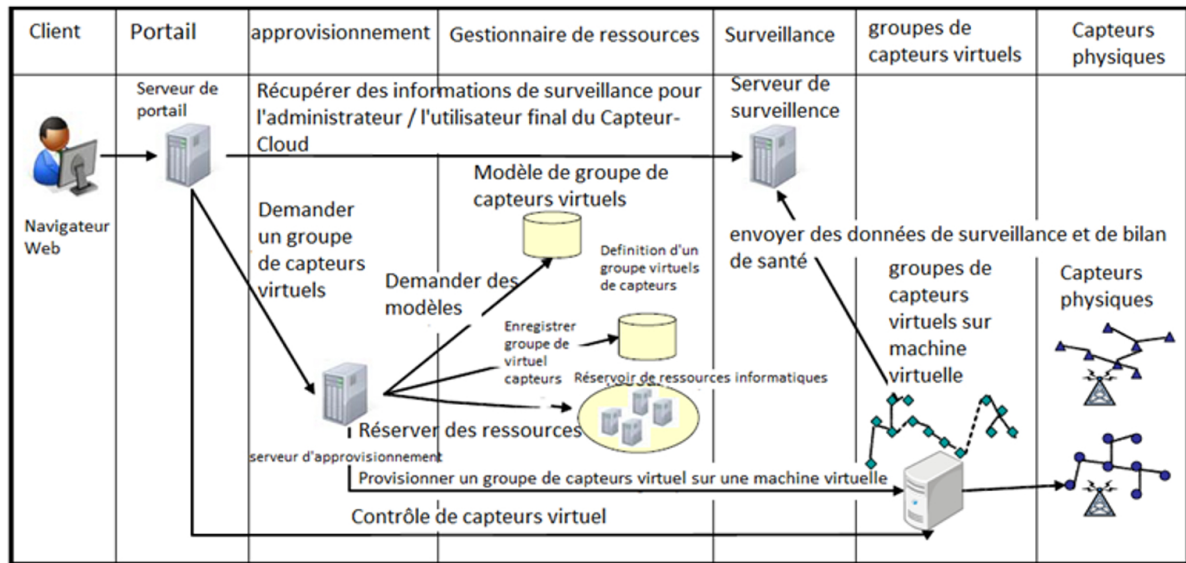


FIGURE 2.5 – Architecture de l'infrastructure Capteur-Cloud [7]

Plusieurs travaux ont été menés dans le cadre de la virtualisation des RCSFs au sein des Cloud. On peut évoquer à titre d'illustration :

Le travail présenté par M. Yuriyama et al. [7] qui fût le premier en 2010 à parler de la virtualisation des capteurs physiques dans le Cloud. L'infrastructure proposée, appelée Capteur-Cloud (en anglais : Sensor-Cloud) relie les capteurs physiques à Internet, via le Cloud. L'infrastructure Capteur-Cloud virtualise les capteurs physiques et met à la disposition des utilisateurs des modèles de groupes de capteurs virtuels. L'architecture contient des mécanismes de gestion des capteurs virtualisés, des mécanismes de gestion du flux d'approvisionnement et des mécanismes de surveillance des capteurs, ce qui améliore les délais de livraison, réduit les coûts et augmente la convivialité des utilisateurs. L'infrastructure Capteur-Cloud est axée sur la gestion du système de capteurs. La figure 2.5 résume l'architecture Capteur-Cloud proposée.

Depuis, les chercheurs ont porté un grand intérêt à la virtualisation des capteurs dans un RCSF au sein du Cloud et les études sur l'infrastructure Capteur-Cloud se sont succédées. Dans [47, 48], les auteurs ont présenté des travaux qui décrivent le paradigme Capteur-Cloud. A. Alamri et al. [47] citent différentes définitions de ce paradigme et le présentent comme un système de gestion de données collectées par des RCSFs, offrant tous les avantages et la puissance de la technologie Cloud. Ils décrivent ensuite l'architecture de l'interface Capteur-Cloud, essentiellement basée sur la virtualisation comme une architecture où l'utilisateur perçoit les capteurs comme un service suivant les demandes des différentes applications. Les capteurs physiques sont alors des entités virtuelles automatiquement aménagées selon leurs utilités dans chaque service. Ils décrivent aussi la structure du modèle comme un modèle à trois couches :

1. une couche qui traite avec les utilisateurs et leurs applications pertinentes,
2. une couche qui traite la virtualisation des capteurs physiques et des ressources dans le Cloud,

3. et la dernière couche qui traite la création de modèles de services et de définition de catalogue de services dans le menu de catalogue de formation. Les capteurs physiques sont localisés et extraits de cette couche.

R. Dwivedi et al. [49] citent les nombreux avantages d'une telle intégration en termes d'augmentation de la capacité de stockage des données, d'augmentation de la puissance de traitement, de fourniture dynamique de services, de rapidité du temps de réponse, de flexibilité, d'évolutivité, de locations multiples, d'agilité des services, d'optimisation des ressources, de collaboration, d'automatisation et de virtualisation. Ils évoquent également les difficultés qui restent encore à relever.

L. Piotrowski et al. [50] proposent de combiner les RCSFs au Cloud, en virtualisant les RCSFs réels au sein du Cloud en capteurs virtuels. Afin d'assurer une vue consistante et précise avec le système réel, la création et l'utilisation de l'intergiciel TinyDSM a été adoptée. Cette approche repose sur le concept de mémoire distribuée et partagée. Les données collectées sont répliquées sur tout le réseau ainsi que sur le Cloud afin de les rendre accessibles à tout moment.

M. Aslam et al. [51] exposent une nouvelle approche qui consiste en un système d'orchestration automatique pour la fourniture de services dans les systèmes réseaux intégrés, appelé "Orchestration Model for Service Provisioning (OMSP)", et aussi pour permettre au RCSF d'agir comme une infrastructure prête pour le Cloud qui permet des fournitures sur demande de services destinés à des individus potentiels pour l'arrière-plan de leurs systèmes. Les chercheurs introduisent le concept de RCSF-Cloud basé sur la virtualisation des RCSFs. Ils définissent cette infrastructure comme un domaine organisationnel où d'autres systèmes et entreprises se connectent et s'approvisionnent afin de délivrer des services en utilisant le paradigme "Réseaux" en tant que service "Network as a Service (NaaS)". L'infrastructure est divisée en trois parties :

1. La partie nœud réseau : constituée d'un réseau d'appareils sans fil largement distribué afin de capturer des données.
2. Une partie passerelle : constituée d'ordinateurs qui connectent le nœud réseau à l'arrière-plan du système ;
3. La dernière partie, corps de l'entreprise, est appelée l'arrière-plan : constituée d'une plateforme pour l'implémentation de composants comme le gestionnaire de plateforme et le point terminal pour les autres utilisateurs qui demandent des données d'interaction avec le domaine RCSF.

Les deux éléments clés de cette architecture sont le moteur d'orchestration utilisé pour l'approvisionnement des services à tous les niveaux du RCSF Cloud et le modèle d'orchestration qui joue le rôle de pilote pour le moteur d'orchestration. Ce travail a été étendu par S. Rea et al. [52] en utilisant le Serviceware comme approche intergicelle basée sur le fractionnement des RCSFs grâce à des services. Serviceware est élaboré pour transformer les ressources RCSF exposées en services. A.J. Hussein et al. [8] proposent la conception de système distribué pour traiter les problèmes physiques des capteurs qui résident dans divers endroits et fonctionnent dans des environnements différents. Le design proposé pour l'architecture Capteur-Cloud inclue trois couches :

1. La première couche permet la préparation de la construction du modèle de services et la définition du standard en plus de la définition des capteurs,
2. La deuxième couche est la couche la plus importante de la conception proposée, elle communique avec de nombreux groupes de capteurs physiques et tente de les placer dans un groupe classé.
3. La dernière couche correspond aux consommateurs/utilisateurs finaux et à leurs demandes applicables.

Les tests affichent des performances élevées, un bon coût total de possession et un temps de réponse efficace. La figure 2.6 illustre les composants et les acteurs de la conception proposée.

S. Misra et al. [53] proposent une solution de modélisation théorique de la virtualisation des capteurs physiques au sein d'une architecture Capteur-Cloud. Dans ce travail une expression mathématique de virtualisation est proposée afin d'étudier le comportement des applications basées sur les RCSFs sur une plateforme Capteur-Cloud. L'architecture proposée est principalement constituée de trois niveaux :

1. Le niveau organisation d'utilisateur de l'interface Capteur-Cloud : la principale motivation derrière l'idée d'utilisation des Capteur-Cloud est de permettre à l'utilisateur d'accéder directement aux données capturées sans connaître leurs emplacements physiques. La communication est établie à l'aide d'un portail Web qui permet à l'utilisateur de percevoir les capteurs comme un service et non pas comme des entités physiques. L'utilisateur formule alors sa demande de données capturées selon son application.
2. Le niveau infrastructure Capteur-Cloud qui est la vue réelle de l'architecture pour le traitement effectif requis. L'architecture Capteur-Cloud traite avec différents capteurs hétérogènes et différentes spécifications à l'aide du langage de modélisation de capteurs (SensorML). Les données détectées sont efficacement agrégées en temps réel. Dans ce niveau, on retrouve :
 - Un accès dynamique,
 - Une programmation de capteurs sur demande,
 - Une vaste capacité de stockage et de traitement spécifique,
 - Une agrégation de données en temps réel,
 - Une activation du capteur physique,
 - Une gestion d'espace mémoire.
3. Le niveau de capteurs physiques : une fois la directive de l'organisation d'utilisateurs pour le groupe de capteurs virtuels est émise elle sera interprétée par les nœuds capteurs physiques et, par conséquent, ils seront planifiés à la demande de l'application. Chaque nœud capteur physique rapporte ses données détectées au stockage du Capteur-Cloud.

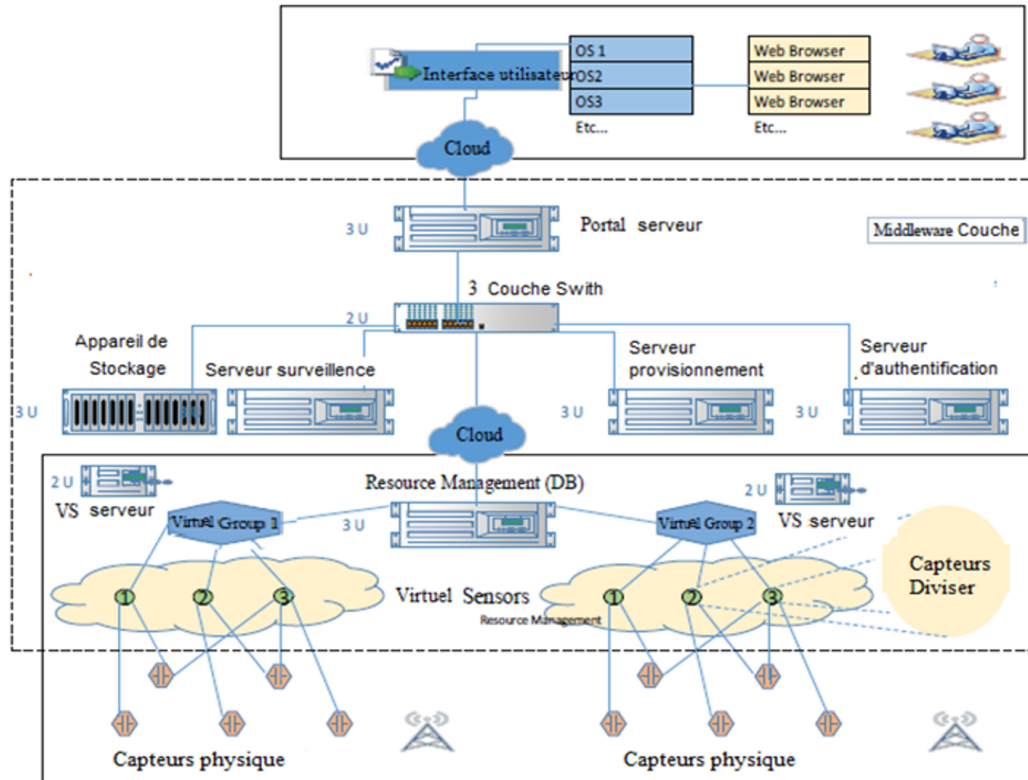


FIGURE 2.6 – Les composants et les acteurs de l'architecture Capteur-Cloud [8]

Les chercheurs ont mené une étude comparative entre l'architecture Capteur-Cloud et un réseau de capteurs traditionnel. Ils ont déduit que l'architecture Capteur-Cloud donnait de meilleurs résultats en termes de réduction de consommation d'énergie ce qui se reflète positivement sur la durée de vie des nœuds capteurs, une meilleure tolérance aux pannes et un meilleur coût d'utilisation.

M. Craciunescu et al. [54] proposent un modèle d'intégration de réseau de capteurs sans fil avec les Cloud appelé (WSN-C), afin de permettre une surveillance et un contrôle plus rapide des capteurs physiques et un flux d'informations plus souple des capteurs aux utilisateurs. En plus de la collecte des données à partir des capteurs, dans son architecture, les auteurs utilisent des actionneurs afin d'exécuter des commandes envoyées par le Cloud. Les nœuds du réseau de capteurs sont directement liés au Cloud ou à travers des points d'accès. Au sein de l'infrastructure du Cloud l'utilisateur a la possibilité de mapper des capteurs physiques à l'intérieur d'un RCSF à des capteurs virtuels à l'intérieur d'un réseau de capteurs virtuel "Virtual Sensor Network (VSN)".

Dans un très récent travail effectué par M. Zhang et al. [55], les auteurs proposent un schéma d'approvisionnement de capteurs virtuels dans une architecture Capteur-Cloud. La nouveauté dans ce travail consiste à considérer plusieurs capteurs physiques pour la création d'un seul capteur virtuel. En effet, dans les études existantes, on considère principalement la sélection de tous les capteurs physiques, ce qui entraîne une énorme consommation d'énergie. Dans le schéma proposé tous les capteurs physiques sont divisés en classes par différentes tendances changeantes, ensuite, l'algorithme de clustering k-means [56] est appliqué pour chaque classe enfin, des capteurs physiques représentatifs sont choisis pour créer un capteur virtuel correspondant. En résumé, moins de capteurs physiques

sont utilisés pour fournir des services de meilleure qualité. Les résultats expérimentaux ont montré que cette approche est efficace et adaptée aux tâches de provisionnement de capteurs virtuels. Néanmoins, la solution proposée ne traite pas le problème d’approvisionnement dans un environnement hétérogène.

Malgré les avantages apparents de la virtualisation, il reste encore des difficultés à surmonter dans cette solution. R. Hummen et al. [57] discutent de ces difficultés ; ils citent notamment les problèmes d’autorisation, le problème de sécurité, le problème de durée de vie et le problème de stockage.

2.3.2 Aspect sécuritaire

Plusieurs chercheurs se sont focalisés sur l’aspect sécuritaire des données. Comme dans chaque système, les données des RCSFs peuvent être soumises à des vulnérabilités susceptibles de compromettre leur fiabilité. La divulgation de données des RCSFs peut s’avérer critique et ainsi avoir une grande incidence dans des domaines tels que l’industrie militaire, le domaine de la santé, etc. . . . C’est pourquoi, il est primordial de garantir un niveau de sécurité et de confidentialité acceptable de l’information. Parmi les solutions actuelles proposées par la communauté scientifique, on retrouve :

Le travail de R. Hummen et al. [57] où les auteurs traitent la sécurisation des données transmises sur une infrastructure Cloud pour le traitement et le stockage à partir d’un RCSF. L’architecture proposée repose sur l’hypothèse que les données sont transférées correctement au nœud coordinateur depuis le nœud capteur et que celui-ci est capable de tracer la source de ces données. Aussi, l’infrastructure Cloud fournie soit considérée comme non hostile. Dans cette architecture les mécanismes de sécurité sont introduits au niveau du point d’accès des données entre le réseau de capteurs et l’externalisation vers le Cloud. Une nouvelle entité logique appelée point de confiance (trust point) est introduite à ce niveau-là ; elle agit comme un pont entre le domaine de sécurité du RCSF et le Cloud. Trois tâches sont réalisées par ce dernier :

1. Premièrement, elle implémente des mécanismes de sécurité pour faire suivre les données des capteurs vers le Cloud à la place du propriétaire.
2. Deuxièmement, elle protège la confidentialité et l’intégrité des données durant leur transmission vers le Cloud.
3. Troisièmement, elle applique des mesures de sécurité afin d’établir un contrôle sur le propriétaire des données même après que leurs transmissions protégées sur le Cloud soit terminée, en exécutant un management sur les services autorisés.

Des mécanismes d’isolation ont été utilisés afin de limiter le nombre de fuites de données sensibles. Après évaluation de cette architecture, les résultats ont montré une performance adéquate pour le scénario choisi et que les frais généraux d’espace mémoire et de stockage peuvent être gérés efficacement.

Dans le travail de F. Banaie et al. [58], une infrastructure pour le traitement sécurisé des données dans les services Capteur-Cloud est proposée. Le modèle de sécurité présenté utilise une cryptographie basée sur l’identité afin de simplifier la distribution et l’authentification des clés dans l’infrastructure Capteur-Cloud. L’architecture proposée simplifie la

distribution et l'authentification des clés, par conséquent, elle accélère la fourniture de services dans une infrastructure Capteur-Cloud. Cette architecture se divise principalement en trois couches principales :

1. Couche de traitement d'application : Elle encapsule les applications et une unité de traitement locale,
2. Couche de gestion : Elle prend les décisions de gestion et fournit également aux utilisateurs des services d'informations de capteurs sécurisés pour l'authentification et la gestion des accès. Cette couche permet d'empêcher les utilisateurs d'accéder aux informations confidentielles,
3. Couche de gestion de données : Elle reçoit les données de la passerelle, puis elle les traite et les ajoute au référentiel de données.

L'architecture proposée simplifie la distribution et l'authentification des clés ; donc elle accélère la fourniture de services dans une infrastructure Cloud-Capteur.

M. Ahmed et al. [59] proposent d'utiliser le modèle courtier basé sur le contenu, appliqué aux infrastructures qui intègrent les RCSFs au Cloud pour l'étude de leur sécurité. A cet égard, les chercheurs proposent d'utiliser une nouvelle approche basée sur l'algorithme en temps réel de la chaîne de Markov Monte Carlo (MCMC) pour détecter les attaques internes dédiées aux RCSFs. Les résultats de simulation ont montré que l'algorithme proposé fournit un taux acceptable de détection de nœuds suspects ou d'attaques internes.

S. Saha et al. [60] proposent un modèle de sécurité dédié aux applications centrées sur les données, c'est-à-dire les applications destinées entre les Cloud et les RCSFs. Les chercheurs ont appliqué des restrictions sur les données en fractionnant la donnée en plusieurs segments doté chacun d'un chiffrement et d'une clef attribuée à l'utilisateur autorisé pour chaque segment. Une technique hybride de cryptographie (symétrique et asymétrique) multicouche est utilisée sur les données. Le modèle de sécurité proposé assure la confidentialité, l'intégrité et le contrôle d'accès des données.

H. He et al. [61] discutent des défis que pose le stockage, la sécurisation et le partage des données dans l'intégration entre les RCSFs et le Cloud. Afin de résoudre ces problèmes un schéma de contrôle d'accès aux données, fin et léger, nommé FLAC a été proposé. Un algorithme de politique d'accès sur le message chiffré "Ciphertext-Policy Attribute-Based Encryption (CPABE)" permet d'obtenir un contrôle d'accès aux données et une confidentialité des données à granularité fine, qui prend en charge des stratégies d'accès plus diverses et gagne en efficacité par rapport aux autres algorithmes CP-ABE. De plus, les chercheurs ont conçu un algorithme ABS pour l'intégrité des données et le cas d'accès aux données à lectures multiples et écritures multiples. Dans FLAC, la plupart des tâches de calcul intensif sur des ressources sont limitées. Les capteurs et les appareils mobiles peuvent être facilement délégués aux services Cloud sans révéler le contenu des données et les clés secrètes, ce qui rend FLAC très approprié pour l'intégration entre les RCSFs et le Cloud. En outre, FLAC est prouvé sécurisé sur la base de l'hypothèse de sécurité standard et fournit une résistance à la collusion, une impossibilité de signature et anonyme.

J. Li et al. [62] proposent un schéma d'externalisation des données basé sur l'algorithme CP-ABE, afin de garantir un accès sécurisé aux données et réduire le temps de

traitement global des données. Dans leur schéma les auteurs proposent une méthode de chiffrement par bloc qui permet de chiffrer/déchiffrer un fichier volumineux et de le transmettre en blocs simultanément. Dans cette approche, le récepteur de données peut vérifier l'intégrité des données sans aucun composant clé principal. L'analyse de la sécurité montre que le schéma proposé peut répondre aux exigences du Cloud intégré aux RCSFs en matière de sécurité. Les évaluations montrent que l'approche est efficace pour le traitement des données, en particulier pour le chiffrement et la transmission des données.

T. Wang et al. [63] proposent une contribution pour une protection suffisante de la confidentialité d'un système capteur-cloud et la résolution des problèmes de stockage dans ce système. Les auteurs présentent un modèle basé sur l'informatique en périphérie (en anglais : Edge computing) pour la collecte de données, dans lequel les données brutes des RCSFs sont différenciellement traitées par des algorithmes sur des serveurs périphériques pour le calcul de la confidentialité. Une petite quantité de données de base est stockée sur les serveurs périphériques et locaux tandis que le reste est transmis au cloud pour le stockage. De cette façon, les avantages sont doubles. Tout d'abord, la confidentialité des données est préservée car les données d'origine ne peuvent pas être récupérées même en cas de fuite des données stockées dans le cloud. Le modèle basé sur l'Edge computing envoie moins de données au cloud et réduit donc le coût de communication et de stockage. Des analyses théoriques et des expériences approfondies valident cette méthode proposée.

2.4 Troisième conception : Intégration orientée application

Les RCSFs ont su attirer un grand nombre de la communauté scientifique et industrielle grâce à leurs faibles coûts d'implémentation et à leurs facilités de déploiement. Cette technologie a pu s'instaurer comme acteur incontournable dans la surveillance et l'évaluation dans de vastes domaines d'applications. Plusieurs chercheurs se sont penchés sur l'idée d'intégration entre les RCSFs et le Cloud avec comme objectif principal de répondre aux exigences d'un nombre varié d'applications. On cite :

2.4.1 Le domaine médical

X. H. Le et al. [64] proposent une architecture sécurisée pour l'intégration des RCSFs aux services cloud. Cette architecture permet la création d'un système de prise en charge et de surveillance des patients. Dans cette architecture, un capteur d'activité mobile (MA-SoL) a été développé afin de rassembler et de stocker toutes sortes de comportement des gens ainsi qu'une reconnaissance vidéo à base d'activités.

Le travail de G. Fortino et al. [65] présente l'architecture "BodyCloud". C'est une architecture système basée sur le Cloud pour la gestion et la surveillance des flux de données de capteurs corporels. Il intègre des concepts clés tels que le passage à l'échelle et la flexibilité des ressources, l'hétérogénéité des capteurs, et le déploiement dynamique ainsi que la gestion des utilisateurs et applications communautaires.

M. Guezguez et al. [66] proposent une nouvelle architecture Capteur-Cloud pour surveiller les patients à haut risque. Au-dessus des capteurs physiques, des capteurs virtuels sont construits en relation avec le dispositif médical et des demandes du personnel

médical. Ces capteurs virtuels sont créés dynamiquement et mis à jour en fonction de l'emplacement et de l'état de santé du patient et également en fonction de l'occurrence d'événements sensibles liés à l'état des capteurs physiques associés. En outre, la configuration des capteurs virtuels s'enrichit de nouvelles fonctions de surveillance, de maintenance, d'économie d'énergie et de sécurité pour améliorer leurs utilisations. Une simulation est réalisée pour évaluer l'efficacité de l'architecture proposée.

M. Muzammal et al. [67] rapportent une nouvelle approche permettant la fusion de données médicales obtenues à partir de réseaux de capteurs corporels "Body Sensor Network (BSN)" de malades souffrant de problèmes cardiovasculaires dans un environnement de Fog computing. Les auteurs ont utilisé une forêt aléatoire de noyaux dans un environnement Fog informatique pour la prédiction des maladies cardiaques. Les données d'activité quotidienne de plusieurs capteurs sont considérées. Ces données sont des entrées dans un classificateur d'ensemble après avoir effectué la fusion des données. Les ensembles sont hébergés dans un environnement Fog informatique et les calculs de prédiction sont effectués de manière décentralisée. Un nouvel ensemble de forêts aléatoires à noyau a été utilisé et des résultats de haute précision sont obtenus pour la prédiction des maladies cardiaques.

2.4.2 Le domaine environnemental

J. Cen et al. [68] proposent un système pour améliorer le schéma de surveillance des désastres en utilisant les RCSFs et le Cloud. Ce travail introduit un algorithme de réseau neuronal orthogonal basé sur une analyse des fonctions ondelettes afin d'améliorer la précision de la surveillance régionale et pour réduire le débit et la consommation en énergie des nœuds dans un réseau de capteur sans fil. Ils utilisent la plateforme Cloud pour analyser un grand nombre de données afin de prévoir un désastre. Le système utilise trois couches :

1. La première consiste en la collection et la transmission de données par le RCSF et le nœud coordinateur.
2. La deuxième couche repose sur le Cloud pour le stockage et le traitement des données. Le Cloud utilise le parallélisme afin de prédire la forme et la sévérité du désastre. L'infrastructure du Cloud est implémentée à l'aide de la plateforme Hadoop.
3. La troisième couche est dédiée à l'application de l'utilisateur final.

Les résultats montrent un bon contrôle de congestion et une meilleure précision de la surveillance.

E. Kanagaraj et al. [69] proposent une architecture pour un système de surveillance d'environnement basé sur le Cloud. Un web service est proposé aux utilisateurs de ce système à l'aide d'un serveur Cloud afin d'accéder aux données météorologiques de L'État du Perlis, en Malaisie. Ces données sont récoltées à l'aide d'un ensemble distribué de RCSFs dont les stations de base sont liées au serveur Cloud. Le système est constitué de quatre parties, une première partie appelé "mottes" constituée de réseaux de capteurs distribués, la partie serveur constituée d'un système de calcul imbriqué afin de stocker les bases de données et les inscrire, la partie Cloud constituée d'un serveur hébergeant un web serveur et une base de données centralisée hébergée au sein du fournisseur du service Cloud, la partie client constitué d'une interface web pour les utilisateurs.

D.S. Arjun et al. [70] présentent une architecture améliorée pour intégrer le Cloud avec les RCSFs afin d'analyser les données météorologiques et d'avertir les utilisateurs de SaaS d'alerte lors de catastrophes météorologiques. La nouveauté qu'apporte ce travail est la virtualisation au niveau des nœuds capteurs et de la couche "réseau".

C.B Pallavi et al. [71] proposent un modèle pour la surveillance et la détection des inondations et des probabilités d'occurrence d'inondation. Le modèle proposé permet de créer une interface conviviale d'instrument virtuel, qui peut être utilisé pour surveiller le niveau de l'eau. Ce système est divisé en trois niveaux :

1. Le premier niveau appelé nœud graine (N-mote) composé de capteurs sans fil permet de récolter les données environnementales (humidité relative, température et niveau d'eau).
2. Ces données seront envoyées au niveau supérieur appelé nœud passerelle (N-gateway), les deux niveaux bénéficient de modules Xbee assurant la communication entre eux.
3. Le troisième niveau est celui du Cloud utilisé pour le stockage et pour fournir différents services (dans ce cas de figure il s'agit de l'utilisation du Cloud Blue-mix/Ubidots). Les données reçues à travers un script python seront stockées dans des bases de données et si un certain seuil est dépassé un message d'alerte est envoyé à l'autorité concernée ou au gouvernement ou aux habitants de la région.

M. Chaitra et al. [72] livrent un système de management de désastres naturels, principalement les risques d'écoulement de débris. L'architecture de ce système repose sur des capteurs sans fil qui sont implémentés dans le site afin de récolter les données qui permettent de détecter les risques de coulée de débris. Ces capteurs sont connectés à des microcontrôleurs qui contrôlent et collectent les différentes données obtenues et qui ensuite seront envoyées à l'aide du WiFi ou de l'Ethernet à travers une passerelle vers le Cloud. Quand un certain seuil est dépassé, une alerte est envoyée aux équipes de sauvetage et au public concerné.

2.4.3 Le domaine agricole

M. Chari et al. [73] proposent un système pour améliorer la quantité et la qualité de production dans le secteur agricole. L'approche consiste à créer un système de surveillance entièrement automatisé pour les applications sous serre, ce qui réduit efficacement les efforts humains. L'internet des objets "Internet of Things (IoT)" ainsi que les RCSFs sont utilisés afin de créer des réseaux intelligents. Le système est constitué de trois parties, les RCSFs permettent de capturer les paramètres d'agriculture sur site, ces capteurs sont interconnectés au microcontrôleur (N-Mote) qui contrôle les capteurs, une passerelle est utilisée pour interconnecter deux réseaux différents connectés à Internet via une connexion Wi-Fi ou Ethernet. Les données des capteurs sont envoyées à la passerelle, puis au Cloud. Lorsque certains seuils sont atteints, les données sont envoyées aux destinataires respectifs.

M. Kassim et al. [74] présentent un système de gestion de l'environnement agricole "an Agricultural Environmental Management System (AEMS)", pratique, abordable et fiable, à l'aide des technologies RCSF et du Cloud. Cette architecture est constituée de trois niveaux :

1. Un niveau de capture de données où les données d'agriculture capturées sont envoyées au Cloud via des passerelles.
2. Le deuxième niveau est celui du service Cloud, dans ce niveau on y trouve l'acquisition, l'analyse, la conversion et le calcul des données ainsi que le stockage.
3. Le dernier niveau est le niveau de contrôle constitué de nœuds de contrôle (par exemple, des ventilateurs de refroidissement et des sprinklers).

L. Mhatre et al. [75] proposent un système dédié aux applications climatiques horticoles qui intègre la technologie du Cloud et les RCSFs ainsi que l'intelligence artificielle afin de fournir une capacité de régulation économique en énergie pour une exploitation agricole menant vers le climat intelligent.

2.4.4 Le domaine du contrôle des structures

J. M. Bindu et al. [76] proposent une solution basée sur les RCSFs et le Cloud pour la surveillance de la température et l'humidité dans des chambres froides. Le système contient deux parties, la première appelée nœud émetteur qui comprend une alimentation, un capteur de température LM-35, un capteur d'humidité HSM-20G, un microcontrôleur PIC18 et un module de Wi-Fi ESP8266. La seconde appelée nœud récepteur contient un microcontrôleur LaunchXL-CC2650 (fonctionnant avec Contiki OS), un capteur de température LM-35, un nano-ordinateur Raspberry Pi, un relais électromécanique et un compresseur. Dans ce système les valeurs de température et d'humidité capturées sont envoyées dans le Cloud par le nœud émetteur. Le nœud récepteur reçoit les données observées de différents nœuds capteurs et les stocke dans sa base de données. Les valeurs stockées sont affichées sous forme d'un graphique dans le site Web Thingspeak. La température est contrôlée par comparaison à un seuil, le compresseur sera ainsi éteint ou allumé pour maintenir la température requise.

C. Zhu et al. [77] proposent de traiter le problème d'amélioration de la qualité de service (QoS) de l'intégration RCSF au Cloud avec confiance dans les villes intelligentes. Les chercheurs proposent d'intégrer la confiance pour améliorer la qualité de service de l'intégration RCSF au Cloud. Ils présentent l'approche de confiance assistée dans le Capteur-Cloud (Trust-Assisted Sensor Cloud (TASC)) basée sur la confiance, dans laquelle des capteurs fiables (c'est à dire des capteurs dont les valeurs de confiance dépassent un certain seuil) collectent et transmettent d'abord les données sensorielles au Cloud. Ensuite, le Cloud sélectionne les centres de données de confiance (c'est à dire des centres de données dont les valeurs de confiance dépassent un certain seuil), pour stocker, traiter les données sensorielles et transmettre ensuite les données sensorielles traitées aux utilisateurs à la demande.

2.5 Intégration entre RCSF et Cloud pour l'IoT

Actuellement nos objets quotidiens sont munis d'un moyen de connexion à Internet. Nous entendons alors de plus en plus souvent parler du terme l'Internet des objets (IoT). Cette technologie permet le relais de milliards d'objets sur le web. On compte de nos jours plus de 80 milliards d'objets reliés. Les RCSFs sont intégrés à l'Internet des objets, où les

nœuds capteurs rejoignent l'Internet de manière dynamique et l'utilisent pour collaborer et accomplir leurs tâches. Plusieurs travaux se sont intéressés à l'intégration entre RCSFs et le Cloud dans des applications dédiées à l'IoT on cite :

F. Bijarbooneh et al. [78] présentent une solution basée sur le Cloud afin de minimiser la consommation d'énergie dans l'internet des objets en prenant en compte la qualité des liens réseaux et la corrélation spatio-temporelle des données. Dans cette approche un nouvel algorithme de détection adaptatif multi-phases avec un protocole de propagation des convictions appelé ASBP est utilisé. Il représente un schéma de récolte de données. Dans ce schéma s'exécutent après chaque période de temps deux phases :

1. La première phase consiste initialement à une première collecte des données où tous les nœuds capteurs sont impliqués. Ces données sont par la suite envoyées au Cloud qui calcule la matrice des coefficients de corrélations et utilise les tables de routage de la station de base pour estimer le plus court chemin des capteurs à la station de base. L'une des deux approches, la programmation par contraintes "Constraint Programming (CP)" ou algorithme glouton (en anglais : Heuristic-Based Greedy Algorithm) sera sélectionnée pour être exécutée au niveau du Cloud pour déterminer un certain nombre de capteurs retenus afin qu'ils puissent rester actifs pour la seconde phase.
2. Dans la seconde phase la station de base exécute l'algorithme de la propagation des convictions "Belief Propagation (BP)" qui prédit progressivement les données manquantes des nœuds capteurs inactifs et des paquets perdus. L'algorithme BP capture la haute corrélation spatio-temporelle dans les données à l'aide d'un modèle graphique.

Le protocole ASBP a donné de meilleurs résultats en comparaison avec une approche aléatoire avec un minimum de 30% et 70% en qualité des liens réseaux avec la station de base et ceux en termes de données pertinentes et de consommation d'énergie. Une comparaison tenant sur la précision de la prédiction des données entre l'approche des auteurs basée sur l'algorithme BP et l'approche basée sur CP en utilisant une sélection de nœuds heuristiques et aléatoires respectivement, a démontré que l'approche BP était nettement meilleure que celle de CP quand le nombre de capteurs est de 10 mais qu'ils ont donné des résultats similaires quand le nombre de nœuds capteurs augmente.

G. Dias et al. [79] présentent une architecture évolutive qui intègre les RCSFs à l'internet des objets en exploitant les services Cloud. Les auteurs visent à réduire les coûts de maintenance liés à la gestion de grands réseaux RCSFs et à s'appuyer sur le pouvoir de déplacer la plupart des calculs vers le Cloud. Dans l'architecture adoptée dans ce travail, le tableau de bord Dashboard (en anglais : Data Analytics for Sensors DAS-Dashboard) est utilisé pour déléguer la gestion et l'analyse des données des RCSFs à des mécanismes appropriés pouvant effectuer leurs opérations à distance dans le Cloud. Les résultats d'expérimentations ont montré que le taux de transmissions pouvait être réduit jusqu'à 88

Le travail de C. Chimsom et al. [80] étudie l'intégration des RCSFs au Cloud pour la conception et le développement d'un système de surveillance basé sur l'IoT. Le réseau est constitué d'un protocole ZigBee connectent les nœuds capteurs à leurs différents coordonnateurs et d'une communication avec un protocole de radio fréquence sans fil reliant les

coordinateurs à la station de base. Le système est conçu pour diviser l'environnement en des zones opérationnelles contenant chacune des nœuds capteurs avec leurs coordinateurs pour la surveillance et la détection de déclenchement de feu ou d'intrusion. La station de base est utilisée pour l'agrégation des données, la virtualisation et la connexion au Cloud, ici plateforme Azure. La logique floue est utilisée pour fusionner les différents types de capteurs et pour la prise de décision. Les résultats de simulations ont montré que le système est efficace et réduit le nombre de perte de paquets lors de la surcharge du réseau. Le système est aussi moins coûteux financièrement comparé à un système de surveillance utilisant des vidéos de surveillance.

Le travail effectué par J. Liang et al. [81] traite le problème lié aux différents types d'attaques physiques et d'attaques de communication auxquelles fait face le "Social Sensor Cloud (SSC)". Le Social Sensor Cloud (SSC) est combiné avec le réseau social, le réseau de capteurs sans fil, le cloud computing et le Fog computing, qui est actuellement un nouveau type d'Internet des objets. Pour lutter contre ce problème, un mécanisme informatique de confiance fiable (RTCM) basé sur des commentaires multi-sources et la fusion de calcul du Fog computing est proposé.

2.6 Récapitulatif

Les travaux concernant l'intégration des RCSFs au Cloud ont résulté des limitations des RCSFs. La grande puissance de calcul du Cloud et le grand espace de stockage ont constitué un support certain aux RCSFs au vu de leurs limitations. Beaucoup de travaux ont été réalisés à ce sujet comme ceux que l'on vient de résumer et qui s'articulent essentiellement autour du problème de stockage des données sensorielles, de leur sécurité et l'amélioration de la durée de vie des RCSFs. Cependant, ces travaux n'ont pas privilégié la réduction des communications entre les RCSFs et le Cloud. Quand on sait que la communication a un coût énergétique élevé, tout particulièrement en ce qui concerne l'émission de données du moment qu'un seul bit envoyé peut parfois consommer autant d'énergie que l'exécution d'un millier d'instructions par le processeur d'un capteur, alors il devient évident que la réduction du flux des communications d'un RCSF constitue un enjeu stratégique. En effet, cela contribue simultanément à la réduction de la consommation d'énergie par les capteurs, la réduction du risque de limitation d'accès au réseau sans fil des données et enfin un volume de stockage de données plus réduit. Dans cette optique un certain nombre de travaux pertinents ont été réalisés en s'appuyant sur les techniques d'agrégation et de prédiction de données. Ces deux techniques constituent des moyens fondamentaux en optimisation des performances des RCSFs. Nous résumons ci-après l'agrégation et nous donnons ensuite quelques indications sur l'application de la prédiction.

2.7 Résumé sur l'agrégation des données dans les RCSFs

L'agrégation de données est une solution fondamentale qui permet de réduire la consommation d'énergie dans les RCSFs en réduisant le nombre de données transmises du nœud capteur vers le nœud puits ou la station de base. Dans l'agrégation, les informations sont

traitées au niveau du capteur comme un condensé utile et ne transmet que le condensé au puits. Les consommations d'énergie sont réduites grâce au nombre réduit de paquets de données transmises. Sans l'utilisation de l'agrégation des données, les capteurs rapporteront toutes les données brutes au puits, bien que ces données aient tendance à être redondantes ou corrélées ce qui conduit à plusieurs inconvénients :

1. Les données redondantes n'ont aucun sens pour l'application.
2. Les chances de congestion du réseau augmentent considérablement.
3. La capacité du réseau est dégradée.
4. L'énergie de consommation augmente en conséquence.

Il existe souvent des corrélations temporelles et spatiales entre les données brutes. Pour un nœud capteur donné, il existe une corrélation temporelle entre les données collectées à différents instants, tandis que la corrélation spatiale se produit lorsque les données sont collectées à partir des nœuds capteurs voisins.

2.7.1 Structure d'agrégation

L'agrégation de données peut être divisée en trois catégories : structure d'agrégation, fonction d'agrégation et planification d'agrégation, où agréger est l'objectif de la structure d'agrégation. Cette catégorie définit la façon dont les données agrégées sont acheminées vers le récepteur en utilisant ou en créant une structure du réseau. La structure doit provoquer l'agrégation dans le réseau. Fondamentalement, la structure d'agrégation trouve plusieurs nœuds capteurs comme agrégateurs ou nœuds passerelles pour traiter les données brutes et les informations agrégées seront transférées via la structure. La structure d'agrégation ressemble plus à un routage centré sur les données, ce qui fait que les nœuds relaient les données avec plus d'informations. La structure d'agrégation rencontrée dans la littérature spécialisée propose des structures basées sur la hiérarchie ou l'épine dorsale pour transmettre des données agrégées.

a) Structures basées sur la hiérarchie

Les structures basées sur la hiérarchie font que les nœuds capteurs forment une hiérarchie, comme un arbre ou cluster. Les données sont transférées des nœuds de niveau inférieur vers les nœuds de niveau supérieur et l'agrégation sera réalisée par les nœuds de niveau supérieur. Dans une structure arborescente, les nœuds fils envoient des informations aux nœuds parents et l'agrégation est réalisée par les nœuds parents. On y retrouve des protocoles comme : "Tiny AGgregation (TAG)" et "Energy Aware Data Aggregation (EADA)". Dans une structure basée sur un cluster, les nœuds membres du cluster rapportent les données au responsable du cluster appelé clusterhead et l'agrégation sera effectuée par le clusterhead. L'un des protocoles les plus utilisés est celui de "Low Energy Adaptive Clustering Hierarchy (LEACH)" [82] ou "Energy-Efficient Mobile Sink Routing protocol Algorithm (EEMSRA)" développé à partir de LEACH. D'autres protocoles existent comme "Hybrid Energy Efficient Distributed clustering approach (HEED)" [83].

b) Structures basées sur l'épine dorsale

Les structures basées sur l'épine dorsale sont fréquemment étudiées dans les RCSF. L'idée sous-jacente est de définir le chemin dorsal ou l'ensemble dominant connecté dans le réseau. Tous les autres nœuds peuvent en transmettre des données et les nœuds de la dorsale sont responsables de l'agrégation des données. Ce type de structure facilite également le processus de diffusion de données et est utile pour les applications de puits mobiles. Il existe plusieurs protocoles basés sur la structure d'épine dorsale :

1. Directed Diffusion (DD) [84] : Ce protocole fait partie de la classe des protocoles de routage centré-données dans les RCSFs. Il vise à diffuser des données à travers des nœuds en utilisant un schéma de dénomination pour les données. Il suggère l'utilisation des paires "attribut-valeur" pour les données et interroge les capteurs à la demande en utilisant ces paires. Dans ce protocole, plusieurs chemins peuvent être construits pour que l'un d'eux soit sélectionné par renforcement. La station de base renvoie le message d'intérêt d'origine via le chemin sélectionné avec un intervalle plus petit, ce qui renforce le nœud source sur ce chemin pour envoyer des données plus fréquemment. Par ailleurs, la maintenance des chemins est possible dans le protocole "Directed diffusion". Ainsi, lorsqu'un chemin entre un nœud source et la station de base échoue, un nouveau chemin ou un chemin alternatif sera sélectionné. Cependant, le protocole "Directed Diffusion" ne peut pas être appliqué à toutes les applications des réseaux de capteurs car il est basé sur un modèle de livraison de données orienté requêtes (query-driven). Les applications qui nécessitent une livraison continue de données à la station de base ne fonctionneront pas efficacement avec un modèle de données orienté requêtes. Par conséquent, ce protocole n'est pas un bon choix en tant que protocole de routage pour des applications telles que la surveillance environnementale.
2. Rail road (RR) (La route ferroviaire) adopte une infrastructure virtuelle appelée rail qui est placée dans la zone centrale du réseau. Comme illustré sur la figure 2.7a, les nœuds à l'intérieur de ce rail sont appelés nœuds ferroviaires. Lorsqu'une source génère une nouvelle information, elle envoie des données aux nœuds rail et les nœuds ferroviaires regrouperont les informations. Dès qu'un nœud récepteur doit collecter les rapports de données générés, un message de requête est envoyé dans le rail jusqu'à ce qu'il atteigne le nœud rail qui stocke les informations pertinentes sur le nœud source. Cependant, le délai est un problème pour le protocole rail ferroviaire. De même, le routage en anneau propose une structure en anneau qui est une boucle fermée de largeur à nœud unique comme montre la figure 2.7b. L'anneau encapsule un réseau global prédéterminé, l'agrégation et la diffusion sont assurées par les nœuds sur l'anneau. L'inconvénient de ce routage est le surcoût de la construction en anneaux pour les réseaux étendus ou éparpillés.
3. Dynamic Directed Backbone (DDB) Organise les nœuds de capteurs en clusters à un seul saut. Un nœud de capteur dans chaque cluster est sélectionné en tant que leader (à savoir, tête de cluster) et les autres nœuds capteur de chaque cluster deviennent membres. Certains membres qui peuvent se connecter entre les leaders deviennent des nœuds de passerelle. DDB Construit une épine dorsale à travers les leaders et les nœuds de passerelle. Lorsqu'un récepteur mobile entre sur le réseau,

il tente de s'attacher à un membre ou à une passerelle par lesquelles il envoie une requête à son chef. Cette requête est ensuite diffusée le long de tous les nœuds de la dorsale jusqu'à atteindre le leader recevant les données reportant d'un nœud source. Si l'évier mobile bouge, il s'attache à un nouveau leader via un nouveau membre ou une nouvelle passerelle et rediffuse la requête dans toute l'épine dorsale à travers le nouveau leader. Cependant, en utilisant le clustering à un saut, DDB ne peut pas avoir suffisamment de nœuds pour remplacer un CH défaillant.

4. Data Quality Maximization (DQM) est un autre protocole basé sur un l'épine dorsale composé de passerelles. Ce protocole suppose la prévisibilité du mouvement du puits et sélectionne des passerelles adjacentes au chemin prédit du puits. Les capteurs établissent les chemins les plus courts avec les passerelles utilisant l'algorithme Floyd-Warshall. Les passerelles regroupent les données entrantes et attendent le puits. Cependant, DQM n'est applicable que dans les applications tolérantes aux retards puisque les passerelles sélectionnées diffusent les données uniquement lorsque le puits est à proximité.

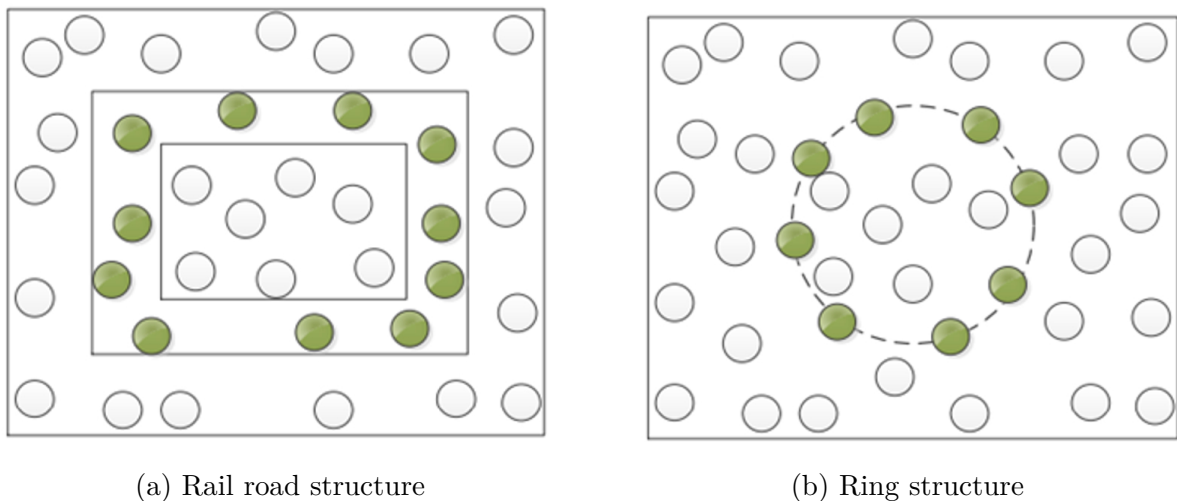


FIGURE 2.7 – Différentes structures de rail road

2.7.2 Agrégation de données sans structure

Les structures basées sur la hiérarchie ou sur l'épine dorsale facilitent l'agrégation des données selon la méthode prédéfinie, mais elles doivent toutes construire et entretenir les structures, ce qui consomme de l'énergie. Pour éviter les frais généraux de maintenance, d'autres méthodes sans structure sont proposées.

a) La fonction d'agrégation

La façon d'agréger est définie à l'aide de la fonction d'agrégation. Ce problème est la partie la plus importante pour l'agrégation de données qui se concentre sur la façon dont les données brutes des nœuds capteurs s'agrègent dans un condensé. Une fonction d'agrégation efficace et utile aide les nœuds capteurs à réduire considérablement la consommation

d'énergie. Nous avons mentionné ci-dessus qu'il existe deux corrélations (les corrélations temporelles et spatiales) entre les données brutes et la fonction d'agréger des données bénéficie principalement de ces corrélations. Fondamentalement, des opérations simples sont utilisées comme fonctions d'agrégation (basées sur la corrélation temporelle), telles que Moyenne, MAX (MIN), SUM, COUNT et MEDIAN. Les inconvénients de ces fonctions sont qu'elles ne peuvent garantir la précision selon un seuil d'erreur donné. Pour l'amélioration des exigences de précision, des fonctions d'agrégation plus complexes sont proposées. Jin Cui et al. [85] ont étudié le classement d'un groupe de fonctions d'agrégation pour déterminer comment sélectionner celle qui convient le mieux pour économiser le trafic et la capacité du réseau, prolongeant ainsi sa durée de vie. Dans cette étude, les auteurs utilisent l'exigence de précision prédéfinie par l'application, la distribution des données brutes et les performances des fonctions (modélisées par un processus de décision de Markov), et classent les fonctions dans une carte, qui montre quelle fonction fonctionne le mieux en fonction de la précision de la cible et le caractère de la distribution de données.

Les fonctions d'agrégation sont basées sur des théories mathématiques et peuvent atteindre une fidélité de récupération plus élevée que les opérations simples. Il existe deux types de fonctions d'agrégation : les fonctions de prévision et de compression. L'agrégation des données de prévision utilise des modèles de prédiction ou des données empiriques pour prédire la suivante, tandis que la compression d'agrégation est engagée pour compresser les informations.

b) L'agrégation des données en prévision

Les fonctions d'agrégation prévisionnelles bénéficient de corrélations spatio-temporelles des données brutes. La série de données d'un nœud capteur (comme la température, l'humidité, etc..) montre souvent que la valeur actuelle des données est liée à la précédente (identique ou similaire) et signifie que la variance de l'ensemble de données ne change souvent pas. Ces séries de données sont stationnaires, ce qui leur permet d'être prédites avec précision. L'agrégation en prévision est un nom générique pour les fonctions qui prédisent les données à l'aide de plusieurs modèles ou méthodes. Comme le nœud récepteur peut récupérer les données haute-fidélité des modèles, les nœuds capteurs n'ont pas besoin d'envoyer toutes les données brutes au nœud récepteur, ce qui réduit le nombre de transmissions et économiser ainsi de l'énergie.

Sur la base des modèles de prévision, les chercheurs proposent des fonctions d'agrégation de données alternatives qui peuvent être réalisées sur des nœuds capteurs. Dans [86], les auteurs utilisent le modèle autorégressif ARIMA dans les RCSFs. Ils pratiquent une méthode pilotée par le nœud récepteur, c'est-à-dire que celui-ci calcule d'abord les coefficients ϕ et ϑ de l'équation (2.1) et les envoie aux nœuds capteurs correspondants. Lorsque la précision ne peut satisfaire un certain seuil, le nœud récepteur recalcule les coefficients.

$$ARIMA \equiv y(t) = \phi_0 + \phi_1 y(t-1) + \dots + \phi_p y(t-p) - \vartheta_1 \varepsilon(t-1) - \dots - \vartheta_q \varepsilon(t-q) \quad (2.1)$$

De plus, ce travail ne s'appuie pas sur les critères "Akaike's Information Criteria (AIC)" ou "Bayesian Information Criteria (BIC)" en raison de l'énorme calcul, l'auteur définit une nouvelle métrique C , telle que :

$$C = \alpha * MAE + (1 - \alpha) * rtran \quad (2.2)$$

où MAE est l'erreur absolue moyenne définie par le seuil et $rtran$ est le rapport du nombre de données transmises sur le nombre total de données. Le paramètre α ($0 \leq \alpha \leq 1$) est utilisé pour faire le compromis entre MAE et $rtran$, c'est-à-dire entre la consommation d'énergie et la précision des données. Ce travail simplifie l'utilisation du modèle ARIMA dans les RCSFs, cependant, le noeud puits nécessite une phase de préparation pour collecter les données et calculer les paramètres, ce qui n'est pas pratique dans certains cas. De plus, il s'agit d'une méthode centralisée, qui n'est pas adaptée aux réseaux de capteurs auto-organisés.

Au contraire, J. Lu et al. [87] présentent le modèle ARMA adaptatif (A-ARMA) dans les RCSFs. A-ARMA réduit le calcul dans tous les nœuds capteurs et ne nécessite pas de phases de pré-calcul. L'idée de base du modèle A-ARMA est que chaque nœud calcule un modèle ARMA basé sur une fenêtre de taille fixe de données consécutives et il faut noter que cette fenêtre est également une fenêtre glissante. En envoyant simplement les coefficients ϕ et ϑ de l'équation (2.1) du modèle ARMA au noeud puits pour la reconstruction des données, la corrélation temporelle de ces données dans chaque fenêtre est explorée. Les coefficients du modèle sont utilisés au niveau du noeud puits pour la prévision des données à l'aide de l'équation (2.1), à moins qu'il ne reçoive de nouvelles mises à jour du modèle des nœuds capteurs. Chaque nœud vérifie localement la précision des données prévues \hat{y}_i avec les données brutes (y_i). Si la précision (erreur RMS) est acceptable selon un seuil donné, le nœud suppose que le récepteur est capable de reconstruire correctement les données et il n'y a donc pas besoin de rapporter les données. Sinon, il calcule un nouveau modèle basé sur la dernière fenêtre de données et communique les nouveaux paramètres au noeud puits afin d'ajuster la prévision. Afin de réduire la complexité du processus d'estimation du modèle tout en atteignant une grande précision, une technique de fenêtre mobile est introduite. Cela signifie que la vérification est requise chaque fois que la fenêtre avance d'un pas. La nature adaptative de la technique repose sur l'utilisation de cette fenêtre mobile. De plus, elle diminue la complexité du calcul.

c) La compression d'agrégation

La compression des informations redondantes est l'idée initiale d'agrégation et les chercheurs essaient de trouver un meilleur moyen de compresser les données brutes. Ce faisant, elle peut réduire l'énergie de consommation et ainsi prolonger la durée de vie du réseau de capteurs sans fil. La plupart des algorithmes de compression de données trouvés dans la littérature ne sont pas réalisables pour les RCSFs car la complexité de calcul est trop élevée pour être implémentée sur les nœuds capteurs. Une autre raison est que ces algorithmes nécessitent une vitesse de calcul élevée, alors que le possesseur du nœud capteur n'a pas la capacité d'exécuter ces algorithmes. Ainsi, des algorithmes à faible complexité et de légère compression sont proposés pour les RCSFs [88, 89].

2.7.3 Planification d'agrégation

Le moment de l'agrégation est concerné par la planification de l'agrégation. L'agrégation de données doit traiter les données puis les transmettre, ce qui entraîne un retard

entre les noeuds sources et le noeud puits. Une planification d'agrégation est proposée pour résoudre ce problème. Plus précisément, la planification d'agrégation définit le moment pour un noeud capteur d'agréger les données et quand le noeud doit transmettre ses données. La planification de l'agrégation a pour but de réduire le délai causé par l'agrégation des données. Il existe également de nombreux travaux sur cette question. La planification d'agrégation est considérée comme une partie de la fonction d'agrégation.

2.7.4 Synthèse sur l'agrégation de données dans les RCSFs

Nous soulignons qu'il existe plusieurs problèmes qui n'ont pas été suffisamment discutés. Premièrement, certains travaux construisent une structure d'agrégation reposant sur des données brutes (par exemple l'agrégation d'isolignes), ce qui fait que ces méthodes dépendent fortement des données brutes. En cas de variation brutale des données brutes, les performances de ces travaux sont affectées. Deuxièmement, les fonctions d'agrégation actuelles sont conçues pour des applications ciblées. Par exemple, la détection compressive convient mieux aux réseaux denses, alors que A-ARMA conserve un ordre de modèle fixe. Dans les deux cas, si la topologie du réseau évolue ou si le comportement de l'application change, de telles fonctions d'agrégation ne pourront pas saisir le nouveau contexte. Troisièmement, à notre connaissance, dans le domaine des réseaux de capteurs, il n'y a pas de travail concernant la comparaison des performances de différentes fonctions et il n'y a pas de guide sur la façon de sélectionner une fonction pour considérer une application et cibler la précision.

2.8 La prédiction des données

La prédiction des données par calcul est un domaine assez vaste de la théorie de contrôle en mathématiques. Dans le souci de ne pas surcharger ce chapitre, on se contente d'indiquer quelques travaux relatifs à la prédiction des données dans le domaine des RCSFs. Parmi les plus importants on cite :

Les travaux de F. A. Aderohunmu et al. [90] qui proposent une solution pour résoudre le problème de consommation d'énergie dans les RCSFs en se basant sur une approche fondée sur la prédiction des données à l'aide d'un modèle mathématique. Les auteurs ont choisi un modèle de prédiction très simple pour faciliter son implémentation dans les RCSFs même dans un réseau à grande échelle. Ce modèle nécessite une faible empreinte mémoire. Contrairement à d'autres travaux sur le même sujet [86, 91–95], l'idée des auteurs a consisté à implémenter le modèle de prédiction simultanément dans le noeud capteur et dans la station de base afin d'éviter les surcoûts liés à la transmission de la structure du modèle. Cela a constitué une approche innovante.

Afin d'évaluer les performances des modèles de prédiction, la sélection de critères de mesure d'erreurs ont été exploités tels que le critère "*Mean Absolute Error (MAE)*", le critère "*Random Mean Square Error (RMSE)*" et le critère "*Mean Absolute Percentage Error (MAPE)*". Ainsi, sur la base de ses critères s'effectuera le choix du bon modèle. Le coût de prédiction noté PC permet d'évaluer les performances du modèle. Il est défini par la relation :

$$PC = [a * e + (1 - a)r] * E_c \quad (2.3)$$

où :

- α : est le niveau de prédiction $0 \leq \alpha \leq 1$,
- e : est une fonction d'erreur,
- r : est le pourcentage de transmission ($r = [\text{paquetstransmis}]/[\text{nombretotaldepaquets}]$),
- E_c : est le coût de calcul dans le nœud capteur.

L'algorithme du modèle a été implémenté sous le code de calcul Matlab, les données utilisées ont été récoltées d'un bâtiment au centre de Bologne en Italie. Les résultats atteints ont été très satisfaisants.

G.M. Dias et al. [96] ont procédé à une synthèse détaillée à propos de la réduction des données basée sur les techniques de prédiction dans les RCSFs. Dans ce travail, les auteurs ont analysé et catégorisé les mécanismes de réduction des données basés sur la prédiction, conçus pour les RCSFs. Leur contribution principale est l'élaboration d'une procédure systématique de sélection d'un schéma de prédiction dédié aux RCSFs, basée sur les contraintes des RCSFs, les caractéristiques des méthodes de prédiction et des données suivies. Enfin, ils concluent leur travail par une discussion sur les défis futurs et les axes de recherche ouverts sur l'utilisation des méthodes de prédiction pour soutenir les constantes croissances des RCSFs.

2.9 Conclusion

Ce chapitre résume les différentes approches et techniques utilisées dans le cadre de l'intégration des RCSFs au Cloud. Hormis les avantages et limitations de chacune de ces méthodes, il ressort que l'aspect central relatif à l'amélioration des performances des RCSFs et donc de leurs intégrations au Cloud se résume en la minimisation de la consommation de l'énergie au niveau du nœud capteur, la faculté d'accès au réseau en minimisant la congestion des données et enfin la capacité de stockage des données. Notre étude bibliographique montre très clairement que le moyen le plus indiqué pour résoudre ces contraintes est celui de la prédiction et de l'agrégation des données car toute donnée bien prédite élimine de fait son émission et répond donc à l'objectif. En fait, ce sont les émissions qui sont énergivores. C'est pourquoi nous nous concentrons dans cette thèse sur la problématique de la prédiction et de l'agrégation des données en essayant de développer des outils capables d'améliorer les résultats atteints jusqu'ici par d'autres chercheurs dans ce domaine. Une fois ce travail achevé, l'intégration au Cloud pourra aisément exploiter une méthode appropriée parmi celles décrites dans ce chapitre.

Chapitre 3

Etude de la prédiction duale et de
l'agrégation sur le modèle RCSF en
anneaux

Chapitre 3

Etude de la prédiction duale et de l'agrégation sur le modèle RCSF en anneaux

3.1 Introduction

La croissance rapide des RCSFs et les futures applications Internet des objets (IoT) vont générer des dizaines de milliards de capteurs sans fil qui transmettent fréquemment des données au Cloud. Cependant, l'accès ou la connexion des capteurs au support sans fil du réseau est considéré comme un problème pour les prochaines générations des RCSFs. Par conséquent, le nombre de transmissions dans les RCSFs peut rapidement se transformer en un problème de congestion, perturbant la croissance exponentielle du nombre de capteurs et de données, ainsi que la quantité de données produites. Par conséquent, il est essentiel, voire indispensable de réduire le nombre de transmissions de données afin de permettre l'intégration de nouveaux nœuds capteurs, permettant ainsi la croissance des RCSFs et d'offrir en conséquence la collecte d'une plus grande variété de variables. Ainsi, le nombre de transmissions et de réceptions effectuées par les nœuds capteurs est la principale préoccupation dans ces applications, non seulement en raison du défi d'accès au support sans fil du réseau mais aussi en raison de l'énergie requise pour la communication externe et le stockage des données.

Pour étudier ce problème et répondre à cette préoccupation, nous exploitons dans ce chapitre le modèle de communication développé par K. Langendoen et A. Meier [97] pour décrire le réseau RCSF sélectionné. C'est un modèle représentatif des topologies classiques. Il est basé sur la densité des nœuds capteurs et leur zone de couverture. Le principal avantage de ce modèle est sa simplicité à identifier et à décrire le fonctionnement des congestions dans un RCSF. Ce modèle permet de décrire quantitativement l'impact attendu des techniques de prédiction et d'agrégation sur les taux de transmissions de données.

3.2 Présentation du modèle RCSF étudié

La figure 3.1 illustre un exemple de ce modèle où chaque nœud a une moyenne de cinq voisins ($C = 5$). C'est une structure organisée en un nombre d'anneaux (D), ici $D = 3$, capable de décrire un réseau avec un mode de transmission multi-sauts. Cette structure est basée sur le nombre moyen de voisins (C) d'un nœud capteur, ici $C = 5$, et le nombre de sauts entre la station de base (BS) et les nœuds capteurs les plus éloignés (i.e. situés au dernier anneau $d = D$). En supposant une densité surfacique uniforme de nœuds capteurs sur un plan, définie comme $(C + 1)$ nœuds, le premier anneau ($d = 1$) devrait avoir C nœuds.

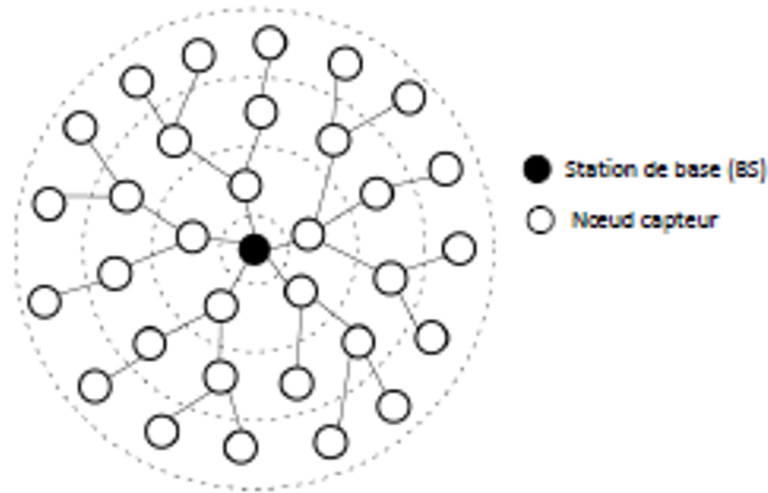


FIGURE 3.1 – Modèle du réseau RCSF considéré

Dans ce modèle, la station de base se trouve toujours au centre de l'anneau zéro et les transmissions sont effectuées soit par un nœud capteur soit par la BS. Ces transmissions peuvent atteindre les nœuds voisins de l'anneau suivant qui se situe à une unité de longueur du composant considéré c'est-à-dire la station de base ou un nœud capteur. L'ensemble des voisins du nœud capteur "i" est défini par tous les nœuds capteurs situés dans le disque unitaire centré sur "i" [98]. Le disque unitaire représente l'intervalle de transmission des nœuds capteurs et n'implique pas nécessairement que les nœuds capteurs voisins établiront une communication directe au niveau de la couche de routage. En fait, les liens de communications sont définis par un protocole de routage sous-jacent à celui décrit dans [97] où les auteurs ont supposé que ces protocoles visent à garder le plus petit nombre de sauts dans les RCSFs et que les nœuds capteurs transmettent seulement vers les nœuds capteurs du précédent anneau c'est-à-dire le prochain anneau le plus proche de la station de base. Par exemple pour atteindre la station de base (BS) à partir de l'anneau d , on peut s'attendre à d sauts de transmissions, donc la distance à partir de la station de base définit également dans quel anneau est localisé le nœud capteur. Le nombre attendu de nœuds capteurs N_d dans l'anneau d peut être calculé en fonction de la surface de la région délimitée par les deux cercles concentriques de l'anneau considéré.

$$N_d = \begin{cases} 0 & \text{si } d = 0 \\ Cd^2 - C(d-1)^2 = (2d-1)C & \text{sinon} \end{cases} \quad (3.1)$$

Le nombre de nœuds dans le RCSF est égal à CD^2 . Étant donné que le premier anneau possède C nœuds capteurs, il est prévu que C branches commencent dans la station de base avec D^2 nœuds capteurs appelés nœuds fils. Dans ce chapitre, chacune de ces branches sera désignée comme un sous-arbre du réseau. En supposant un nœud capteur situé dans l'anneau d , le nombre attendu de fils directs (I_d) peut être calculé par la relation N_{d+1}/N_d . Cette valeur ne dépend pas de C :

$$I_d = \begin{cases} 0 & \text{si } d = D \\ \frac{(2d+1)}{(2d-1)} & \text{sinon} \end{cases} \quad (3.2)$$

3.2.1 Extension du modèle RCSF considéré

Bien que les défis d'accès au support sans fil du réseau et les limites d'énergie du nœud capteur soient couramment observés dans les topologies irrégulières du monde réel, ils sont souvent négligés par d'autres modèles qui n'accordent pas une attention particulière aux nœuds capteurs pouvant avoir la charge de travail la plus élevée [99]. Cela dit, il est à souligner que dans notre travail, nous exploitons le modèle original étendu par Dias et al. [100] afin de déduire le nombre de transmissions en fonction de la position des nœuds capteurs. D'abord, l'ensemble des nœuds fils d'un nœud capteur i situé dans l'anneau d est défini comme $H_{i,d}$. Ensuite, le cardinal attendu $H_{i,d}$ est désigné comme K_d où la valeur de K_d est le nombre de fils directs multiplié par le nombre attendu de leurs fils plus un, comme mentionné par l'équation (3.3) :

$$K_d \triangleq |H_{i,d}| = \begin{cases} 0 & \text{si } d = D \\ I_d(K_{d+1} + 1) & \text{sinon} \end{cases} \quad (3.3)$$

On se rappelle que le nombre attendu de nœuds capteurs est $C * D^2$ et que le premier anneau devrait comporter C nœuds. Ainsi, le nombre attendu de fils des nœuds capteurs dans le premier anneau (K_1) est :

$$\text{Nombre attendu de fils} = (D^2 - 1) \quad \text{si } D > 1$$

3.2.2 Transmissions du nœud capteur à la station de base (BS)

En prévision de données, les nœuds capteurs mesurent et transmettent généralement leurs données à un intervalle de temps prédéfini ($t = 1/f$) pouvant varier de quelques secondes à quelques heures. Le nombre de mesures par seconde (f), et les autres paramètres indiqués dans le tableau 3.1 peuvent varier d'un cas à l'autre.

Dans l'approche la plus simple, les mesures sont transmises juste après leur création. Nous supposons ce comportement comme une base pour des comparaisons ultérieures. Sinon, ces transmissions, que nous appelons nœud à BS, peuvent ne pas se produire juste après la création des mesures si les nœuds capteurs agrègent les données reçues d'autres nœuds capteurs ou de mesures antérieures. L'impact de l'agrégation des transmissions sera également étudié dans ce modèle. Étant donné que les nœuds capteurs doivent transmettre les données de leurs fils vers la station de base, le nombre attendu de transmissions $E(S_{i,d})$ pendant une période de $(1/f)$ secondes d'un nœud capteur i situé dans l'anneau d est égal à :

TABLE 3.1 – Paramètres considérés dans le calcul du nombre de transmissions et de réceptions de données d'un capteur

Paramètre	Description
f	Nombre de mesures par période
C	Nombre attendu de voisins de chaque nœud capteur
D	Nombre attendu d'anneaux/sauts dans le RCSF
ρ	Corrélation entre les mesures dans un sous-arbre du réseau

$$E(S_{i,d}) = (K_d + 1) \quad (3.4)$$

et le nombre de réceptions est :

$$E(R_{i,d}) = K_d \quad (3.5)$$

Ainsi, le nombre total de transmissions durant une période T d'un nœud capteur i est la somme des transmissions et des réceptions :

$$\begin{aligned} E(X_{i,d}) &= E(S_{i,d}) + E(R_{i,d}) \\ &= ((K_d + 1) + K_d)fT \\ &= (2K_d + 1)fT \end{aligned} \quad (3.6)$$

En se basant sur l'équation (3.3) on peut confirmer que $K_1 > K_d$, si $d > 1$. En appliquant cette inégalité aux équations (3.4), (3.5) et (3.6), il est possible de montrer aisément que si $d > 1$, alors :

$$E(S_{i,1}) > E(S_{i,d})$$

$$E(R_{i,1}) > E(R_{i,d})$$

$$E(X_{i,1}) > E(X_{i,d})$$

Ainsi, dans un réseau de capteurs homogènes, les capteurs nœuds situés au premier anneau effectuent plus de transmissions et représentent des goulots d'étranglement qui limitent le nombre de transmissions de leurs sous-arbres en fonction de leurs capacités. Cela montre l'importance de se concentrer sur l'optimisation du nombre de transmissions dans le premier anneau car cela peut empêcher l'ajout de nouveaux nœuds capteurs au RCSF.

3.2.3 Transmissions de la station de base (BS) vers le nœud capteur

La station de base initie les transmissions vers les nœuds capteurs, par exemple, pour changer une configuration ou mettre à jour le logiciel dans les nœuds capteurs. En supposant une transmission unicast par paquet, si la station de base transmet un paquet à chaque nœud capteur dans le RCSF, un nœud capteur i situé dans l'anneau d recevra une

transmission, augmentée de K_d transmissions (une par fils), qui doivent être transmises à leurs récepteurs.

$$E(S_{i,d}^*) = K_d \quad (3.7)$$

$$E(R_{i,d}^*) = K_d + 1 \quad (3.8)$$

Dans ce cas, le nombre moyen de transmissions effectuées par la station de base vers le sous-arbre est D^2 , c'est à dire le nombre de nœuds dans chaque sous-arbre. Cependant, le nœud capteur dans le premier anneau fera ($K_1 + 1 = D^2$) receptions et ($K_1 = D^2 - 1$) transmissions.

3.3 Prédictions de données dans les RCSFs

Le terme prédiction peut faire référence au processus de déduction des valeurs manquantes dans un ensemble de données basé sur des statistiques ou une probabilité empirique, ou à l'estimation de valeurs futures sur la base de données historiques. Une méthode de prédiction P est un algorithme déterministe qui produit des prédictions basées sur deux variables d'entrée :

- un ensemble de valeurs observées X ,
- un ensemble de paramètres.

Un modèle de prédiction p est une instance d'une méthode de prédiction P , telle que $p_\theta(X) = P(X; \theta)$, c'est-à-dire que chaque modèle de prédiction est déterministe et sa sortie dépend uniquement de l'ensemble des valeurs observées. Les valeurs de θ peuvent être fournies par une fonction d'utilité qui mesure la précision des prédictions, la complexité des modèles ou la perte d'informations. Ainsi, un modèle de prédiction peut être résumé en P et θ . Choisir un modèle de prédiction signifie trouver les valeurs de θ qui résument au mieux les mesures actuelles en fonction des critères adoptés par la fonction d'utilité (par exemple, la perte d'information minimale estimée). Finalement, il est possible de créer différents modèles de prédictions qui utilisent le même algorithme (c'est-à-dire la même méthode de prédiction).

Une méthode de prédiction peut nécessiter certaines informations sur les données qui vont être prédites, par exemple, l'hypothèse que les valeurs seront normalement distribuées. Dans certains cas, cette connaissance existe déjà chez l'utilisateur avant même le déploiement du RCSF et peut être appliquée pour des calculs statistiques, telles que les régressions linéaires [101]. L'aspect positif des méthodes statistiques est qu'il est possible d'estimer au préalable le rendement du système. Par exemple, sur la base de l'hypothèse concernant la normalité des données, la probabilité de faire des prévisions précises peut être utilisée pour calculer si les gains que le système peut réaliser valent l'investissement à faire.

Pour étudier l'impact des techniques de prédiction de données sur les performances d'un RCSF, il est nécessaire de présenter d'abord les différents paradigmes de communications actuellement utilisés, c'est-à-dire le mode de communication entre les différents constituants du RCSF.

3.3.1 Paradigmes de communication

Ce sont les paradigmes de communication émergents utilisés actuellement en prédiction des données dans les RCSFs qu'on va présenter. En fait, il s'agit de présenter les différents modes de communication possibles entre les éléments constitutifs d'un RCSF tels que les nœuds capteurs et les nœuds récepteurs (CH,BS).

a) Paradigme de prédiction simple

Le paradigme de prédiction simple désigné par l'abréviation SPS (Single Prediction Scheme (SPS)) consiste à effectuer les prédictions sur un seul emplacement du réseau, qui peut être soit le nœud capteur soit le nœud récepteur (CH, BS). Par exemple, la station de base peut prédire les données mesurées par les nœuds capteurs et décider de façon autonome de l'instant où il faut prendre davantage de mesures en fonction de la fiabilité des prédictions. Ce paradigme fût utilisé par C. Liu et al. [86]. Alternativement, les nœuds capteurs peuvent prédire les variations de leur mesurande afin d'éviter des mesures inutiles et, par conséquent, leurs transmissions. Cette dernière option est particulièrement avantageuse si un nœud capteur dépense plus d'énergie pour échantillonner sa mesurande que pour calculer un ensemble d'instructions informatiques qui prédira les futures mesures. Dans [93], les auteurs ont utilisé un paradigme de communication SPS légèrement différent de celui utilisé par C. Liu et al. [86]. En fait, dans leur travaux, chaque nœud capteur construit son modèle et le communique au BS. Celui-ci utilise les paramètres de la structure du modèle sélectionné par le nœud capteur afin de prédire les données mesurées.

Le principal avantage des schémas SPS est que chaque composant du RCSF (nœud capteur ou nœud récepteur) peut décider lui-même d'adopter ou non des prédictions et il n'y a pas de surcharge (extra-transmissions) liée à la communication de leurs décisions ou à la synchronisation avec leurs voisins. Comme inconvénients, il y a une éventuelle réduction de la qualité des informations fournies par les nœuds récepteurs car ils peuvent perdre de la précision sur une partie de leurs données prédites alors qu'elles sont collectées par les nœuds capteurs. Ce paradigme présente également d'autres lacunes car il nécessite beaucoup de temps pour construire le modèle de prédiction ou pour l'actualiser lorsqu'il devient obsolète. Dans [102], les auteurs ont classé selon l'endroit où les prédictions sont faites, les travaux existants qui adoptent les schémas SPS pour réduire le nombre de transmissions. La figure 3.2 montre schématiquement une variante du paradigme de communication SPS.

b) Paradigme de prédiction duale (DPS)

Le paradigme "Dual Prediction Scheme (DPS)" est venu remédier aux inconvénients intrinsèques du schéma SPS. Les travaux basés sur le paradigme DPS consistent à effectuer simultanément les prédictions dans les nœuds capteurs et dans les nœuds récepteurs (CH ou BS). L'idée générale derrière de tels mécanismes suppose que les nœuds récepteurs fournissent les mêmes informations que les nœuds capteurs, mais ces derniers peuvent vérifier localement la précision des prédictions et éviter leurs transmissions. Pour choisir un modèle de prédiction dans un schéma DPS, les nœuds capteurs rapportent toutes les données mesurées aux nœuds récepteurs au cours d'une phase d'initialisation [92]. Le choix du modèle de prédiction peut être effectué indépendamment dans les nœuds

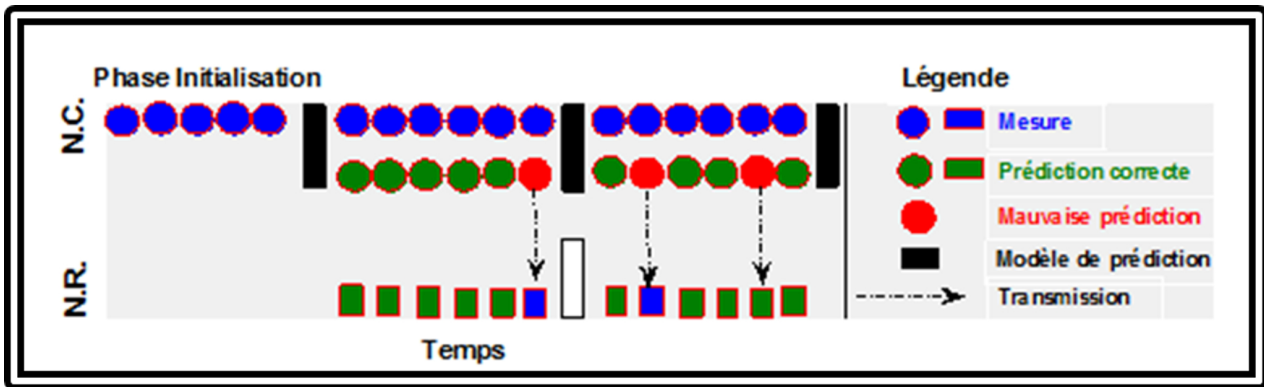


FIGURE 3.2 – Une variante du schéma SPS avec génération du modèle dans le nœud capteur

récepteurs (CH, BS) et les nœuds capteurs (Figure 3.3). Alternativement, chaque nœud capteur peut choisir son modèle de prédiction et transmettre ses paramètres au nœud récepteur. Enfin, on peut donner au nœud récepteur l'autonomie nécessaire pour choisir des modèles de prédiction pour tous les nœuds capteurs. La décision sur le choix du modèle de prédiction peut être prise au moment de l'exécution et il n'est pas nécessaire de définir une stratégie pour toute la durée de vie du RCSF [103]. Après avoir choisi (et partagé occasionnellement) les modèles de prédiction, les nœuds capteurs peuvent exploiter leur proximité avec l'origine des données. Autrement dit, ils peuvent comparer les valeurs prédites aux mesures réelles et effectuent des transmissions que si les prédictions sont inexactes. De temps en temps, si la précision des prédictions est compromise, les nœuds capteurs et les nœuds récepteurs peuvent commencer une nouvelle phase d'initialisation et choisir de nouveaux modèles de prédiction. Comme les modèles de prédiction basés sur le paradigme DPS visent à réduire le nombre de transmissions sans compromettre la qualité des informations générées par le RCSF, ils visent le compromis entre le nombre de transmissions de données et la qualité des mesures fournies par le système. Dans ce cas, l'efficacité des modèles DPS dépend non seulement de la précision des prédictions, mais aussi du nombre de transmissions nécessaires pour disséminer de nouveaux modèles de prédiction. Par conséquent, pour décider d'un nouveau modèle de prédiction, il est nécessaire d'observer quelle est la méthode de prédiction la plus appropriée, compte tenu des conditions environnementales, et si faire des prédictions (au lieu de transmettre toutes les mesures) réduira le nombre de transmissions dans le RCSF. La figure 3.3 illustre le schéma DPS avec génération indépendante du modèle dans les nœuds capteurs et dans les nœuds récepteurs.

Ainsi, on peut récapituler en soulignant qu'à partir de la littérature spécialisée, deux principales architectures ont émergé comme paradigmes de prédiction :

1. Dans les modèles basés sur le paradigme de prédiction SPS, se sont soit les nœuds récepteurs (CH, BS), soit les nœuds capteurs qui font toutes les prédictions, en s'appuyant sur la confiance évaluée par la méthode de prédiction choisie.
2. Dans les modèles basés sur le paradigme de prédiction DPS, les nœuds capteurs et les nœuds récepteurs font des prédictions simultanément et exploitent la proximité des nœuds de capteurs avec l'origine des données pour éviter les transmissions inutiles.

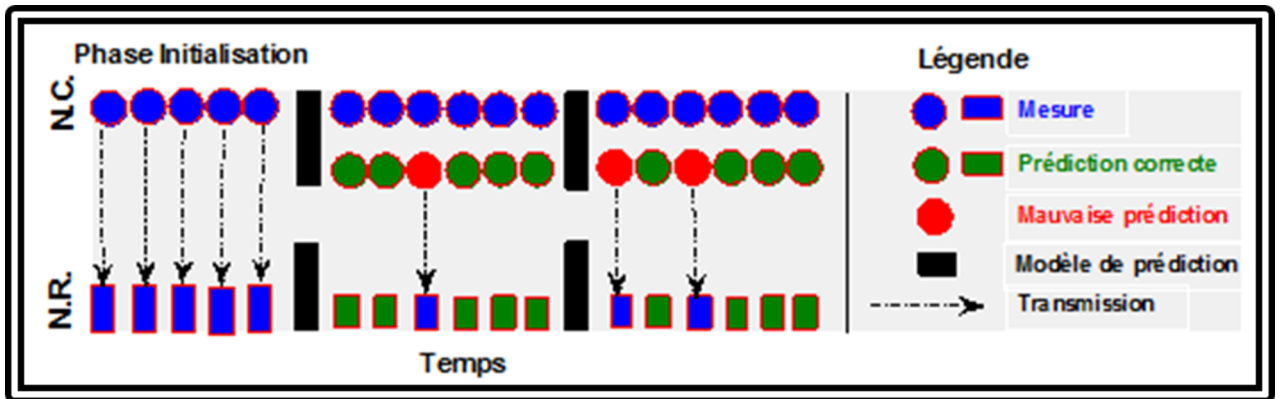


FIGURE 3.3 – Schéma DPS avec génération indépendante du modèle

La plupart des travaux récents ont tendance à utiliser des paradigmes DPS pour exploiter les meilleures caractéristiques de chaque composant RCSF, à savoir la puissance de calcul supplémentaire des nœuds récepteurs (CH et BS) et la proximité des nœuds capteurs avec les sources de données.

Par conséquent, grâce à la grande précision et à la faible complexité des algorithmes de prédiction modernes, les schémas de prédiction duale (DPS) sont des candidats potentiels pour optimiser les transmissions de données dans les RCSFs au meilleur niveau, car ils aident les nœuds capteurs à éviter les transmissions inutiles sans nuire à la qualité de leurs mesures.

c) Schéma adaptatif de prédiction duale (ADPS)

Contrairement aux travaux rapportés dans [86,91] où le nœud récepteur construit les prédicteurs et les diffuse au réseau, F.A. Aderohunmu et al. [90] ont utilisé un paradigme de communication différent, où tous les paramètres du modèle prédicteur sont calculés simultanément, selon une procédure adaptative, au niveau des nœuds capteurs et au niveau des nœuds récepteurs. Le principe repose sur l'utilisation d'un même échantillon de données pour estimer les prédicteurs. Cet échantillon des données d'estimation est appelé également fenêtre d'estimation ou table de données d'initialisation (en anglais : history table). Ce paradigme correspond en fait à celui présenté par la figure 3.3. Ces auteurs ont construit des algorithmes légers utilisant des modèles simples mis à jour également à la fois au niveau des nœuds récepteurs et des nœuds capteurs ou nœuds sources. Cela ne nécessite aucune diffusion du modèle, à l'exception de la fenêtre d'estimation nécessaire à la construction du premier modèle. Leurs algorithmes (1) et (2) s'exécutent simultanément sur les nœuds récepteurs et les nœuds capteurs (c-à-d nœuds sources) et n'utilise que les nœuds sources pour communiquer toute déviation des données prédites par rapport à l'erreur maximale e_{max} définie par l'utilisateur. En effet, les nœuds actualisent les modèles de prédiction à des instants spécifiés, par exemple à chaque intervalle de temps T . Lors de la mise sous tension, le nœud source transmet des données au nœud récepteur sur toute la longueur de la fenêtre d'estimation (i.e. durée d'initialisation) sans effectuer de prédiction. Comme le montre les algorithmes (1) et (2), on voit que depuis la ligne 8 de l'algorithme (1) jusqu'à ce que la fenêtre d'estimation soit pleine, aucune prédiction n'est effectuée au niveau du nœud récepteur BS. Chaque fois que la station de base reçoit des

données du nœud capteur, il met automatiquement à jour la fenêtre d'estimation avec les données reçues, comme montré sur la ligne 12 de l'algorithme (2). De même, au nœud source, une fois tous les paramètres initialisés, la tâche de détection démarre sur la ligne 4 (Algorithm (1)) et transmet toutes les données jusqu'à ce que la fenêtre d'estimation soit pleine. Cela garantit que le même flux de données est disponible au niveau du nœud récepteur et du nœud capteur. Une fois que la prédiction commence et jusqu'à ce que la limite d'erreur e_{max} soit violée, aucune transmission n'est effectuée, la valeur prédite est utilisée pour mettre à jour la fenêtre d'estimation et la valeur détectée (mesurée) est rejetée. La prédiction et les mises à jour au niveau du nœud récepteur se font aux lignes 16 et 17 de l'algorithme (2). Dans [90], les chercheurs ont utilisé une topologie en étoile avec un mode de communication de type "TDMA" et ont utilisé une structure de données "First In First Out (FIFO)" aux nœuds capteurs et à la station de base pour implémenter ce paradigme. Pour éviter les collisions dans le réseau, un protocole "Media Access Control (MAC)" à accès multiple par détection de porteuse non persistant (CSMA) a été utilisé dans leur architecture. Ils résument approximativement leur expérience en précisant qu'une transmission de paquet supprimée équivaut à environ 13 minutes de durée de vie supplémentaire pour chaque nœud.

Dans cette démarche, il peut être nécessaire que les nœuds sources nécessitent une fenêtre d'estimation suffisamment longue pour générer les modèles, ce qui ajoute à la charge d'exécution liée au fonctionnement du RCSF. Le nœud récepteur peut calculer les paramètres du modèle de prédiction sélectionné et les diffuse sur le réseau.

3.3.2 Choix du modèle indépendant

La phase d'initialisation garantit que le nœud récepteur (CH, BS) disposera d'informations complètes sur l'environnement avant de choisir un modèle de prédiction. Par conséquent, après cette phase, les nœuds récepteurs par exemple peuvent choisir les mêmes modèles de prédiction que les nœuds capteurs sans effectuer de nouvelles transmissions. Les algorithmes (1) et (2) illustrent les postures des nœuds capteurs et des nœuds récepteurs. Ces deux algorithmes représentent le schéma ADPS pour une réelle mise en œuvre [104].

De nouveaux modèles de prédiction peuvent être régulièrement choisis sur la base des connaissances simultanément disponibles pour les nœuds capteurs et les nœuds sources. Comme inconvénient, la variété des modèles de prédiction est limitée par les limites de mémoire et de puissance de calcul des nœuds capteurs. La méthode des moindres carrés a fourni des prédictions précises dans les simulations où les nœuds capteurs et les nœuds récepteurs ont généré leurs modèles de prédiction indépendamment [92, 105, 106]. Par exemple, dans un scénario particulier, seuls 10% des mesures seraient nécessaires pour surveiller avec précision la température ambiante [92].

3.3.3 Choix du modèle dans le nœud capteur

Alternativement, on peut également choisir des modèles de prédiction dans les nœuds capteurs, comme illustré par la figure 3.4. Comme précédemment, les nœuds capteurs commencent à transmettre toutes les mesures aux nœuds récepteurs. Cependant, une nouvelle responsabilité est attribuée aux nœuds capteurs. En effet, après avoir collecté

Algorithm 1 : Algorithm step for sensor node

```

1: Begin
2: Call Task START ;
3: Initialize parameters ;
4: Call Task Sensed_data ;
5: Call Task Transmit ;
6: Count++ ;
7: Call Task Sleep ;
8: for Every fixed sample interval do
9:   if ( $Count > Sizeof(History\_table)$ ) then
10:    Call Predictive Model ;
11:    if ( $|Predictive\_value - Sensed\_date| > Error\_Bound$ ) then
12:     Call Task Transmit ;
13:     Update History_table with Sensed_date ;
14:     Call Task Sleep ;
15:     Break ;
16:   else
17:     Update History_table with  $Predictive\_value$  ;
18:     Discard Sensed_date ;
19:     Call Task Sleep ;
20:   end if
21: end if
22: end for
23: End

```

les mesures locales et choisi un modèle de prédiction adapté à l'environnement actuel, ils doivent communiquer le modèle de prédiction aux nœuds récepteurs. L'avantage principal de ce mécanisme est que les nœuds capteurs peuvent choisir de nouveaux modèles de prédiction en utilisant toutes les données mesurées, au lieu d'utiliser uniquement les informations qu'ils partagent avec les nœuds récepteurs. D'autre part, les nœuds capteurs ont besoin de transmissions supplémentaires pour informer les nœuds récepteurs de leurs décisions.

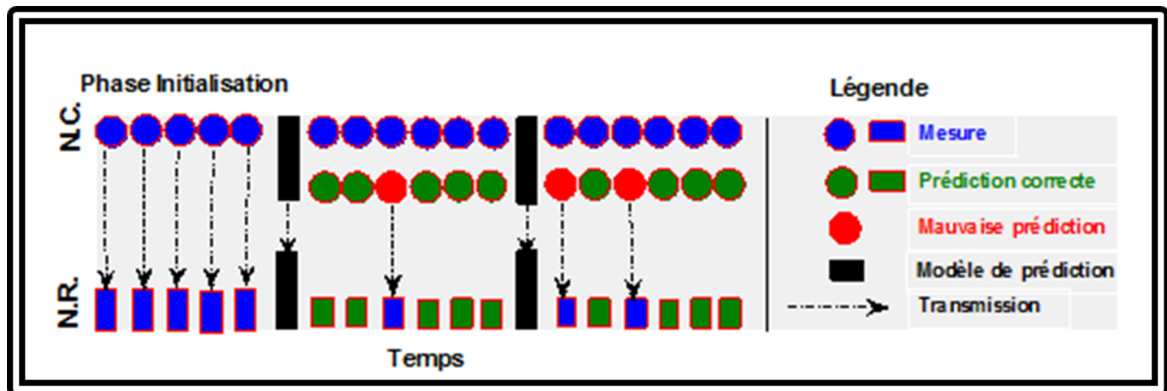


FIGURE 3.4 – Choix des modèles de prédiction dans les nœuds capteurs

Algorithm 2 : Algorithm step for base station (BS)

```
1: Begin
2: Call Task START ;
3: Initialize parameters ;
4: Wait to receive Sensed_data from leaf nodes
5: if Sensed_data is received then
6:   Count++;
7:   for Every fixed sample interval do
8:     if ( $Count < Sizeof(History\_table)$ ) then
9:       Update History_table with Sensed_date received ;
10:    else
11:      if ( $Count \geq Sizeof(History\_table)$ ) then
12:        Update History_table with Sensed_date received ;
13:        Break ;
14:      end if
15:    else
16:      Call Predictive model ;
17:      Update History_table with prediction ;
18:    end if
19:  end for
20: end if
21: End
```

Les résultats de simulation utilisant des données réelles provenant des RCSFs ont montré que cette approche pourrait réduire le nombre de transmissions de données à l'aide de modèles autorégressifs sans dépasser les capacités typiques de mémoire, ni les ressources de calcul du nœud capteur [86, 104, 107]. Alternativement, un mécanisme hybride peut améliorer la qualité des prédictions si les nœuds capteurs disposent de l'autonomie nécessaire pour adopter des méthodes de prédiction plus complexes (par exemple, réseaux de neurones artificiels (RNA)) lorsque les prédictions de plus simples modèles sont inexactes [108]. Dans le pire des cas, si la différence entre les mesures et les prédictions utilisant la méthode la plus complexe dépassent également le seuil d'acceptation, les nœuds capteurs sont responsables de la transmission des mesures réelles aux nœuds récepteurs. Plus récemment, J. Kho et al. [109] ainsi que Y. Chen et al. [110] ont donné aux nœuds capteurs la possibilité de prendre des décisions localement en utilisant la régression des processus gaussiens et la descente de gradient stochastique, qui nécessitent des calculs beaucoup plus importants que les méthodes traditionnelles. En suivant la tendance à adopter des méthodes de prédiction complexes dans les capteurs ; S. Cheng et al. [111] ont incorporé la théorie de l'information dans leur analyse et ont décrit une méthode permettant de prédire et d'évaluer avec précision leurs modèles de prévision.

3.3.4 Choix du modèle dans le nœud récepteur (BS ou CH)

Dans les schémas DPS ou ADPS, le nœud récepteur tel que le nœud CH ou le nœud BS est responsable du choix et de la transmission périodique des nouveaux paramètres du modèle de prédiction et des niveaux d'acceptation des erreurs aux nœuds capteurs,

comme illustré sur la figure 3.5. La génération des modèles de prédiction dans la station de base exploite la disponibilité asymétrique de la puissance de calcul dans les RCSFs. En effet, les stations de base ont généralement des sources d'énergie moins chères et plus de ressources (telles que la mémoire et la puissance de traitement) que les nœuds capteurs ordinaires qui mesurent et rapportent principalement des données environnementales [86, 112]. Finalement, les stations de base peuvent compter sur les services du Cloud pour analyser les données collectées et choisir des modèles de prédiction plus précis. Par exemple, les réseaux de neurones (RNA) peuvent fournir une plus grande précision que d'autres méthodes, mais ils peuvent ne pas répondre aux contraintes des nœuds capteurs, car la création d'un RNA nécessite une phase d'apprentissage intensive en calcul sur une grande quantité de données. De plus, la station de base peut assumer la responsabilité d'adapter les opérations des nœuds capteurs en fonction des économies potentielles que les prédictions peuvent générer. Dans de tels cas, la station de base peut estimer s'il est utile de faire des prédictions dans les nœuds capteurs en fonction de la relation entre la précision des prédictions, la corrélation entre les mesures et l'erreur tolérée par l'utilisateur [100]. En fonction des gains attendus, les nœuds capteurs peuvent être configurés pour :

- passer en mode veille sans effectuer de mesure,
- effectuer des mesures et transmettre toutes les mesures effectuées,
- effectuer des mesures, les transmettre au BS lorsque la prédiction diffère de plus d'une valeur d'acceptation.

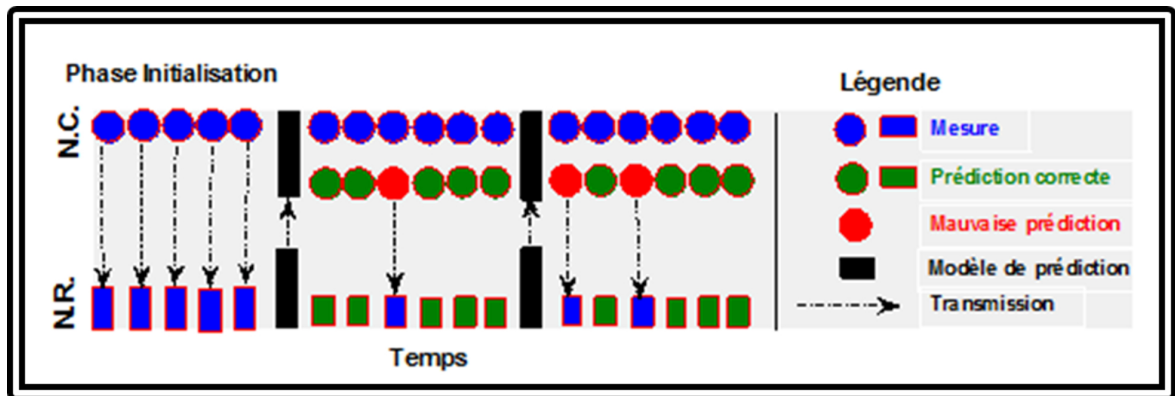


FIGURE 3.5 – Choix des modèles de prédiction dans la station de base

3.4 Modélisation des schémas de prédiction duale

Comme expliqué dans le paragraphe (b)) de la section (3.3.1) les schémas de prédiction duale DPS exploitent la proximité des nœuds capteurs aux sources des données, évitant les transmissions inutiles et gérant les limitations matérielles occasionnelles des nœuds capteurs qui pourraient réduire la durée de vie des RCSFs. Un schéma DPS possède deux tâches qui peuvent être exécutées soit par les nœuds récepteurs, soit par les nœuds

capteurs, à savoir le choix du modèle de prédiction et la diffusion du modèle de prédiction. La diffusion est le processus de transmission du modèle de prédiction soit des nœuds capteurs vers les nœuds récepteurs, ou inversement. Dans ce qui suit, nous décrivons l'impact de ces tâches sur la charge du réseau en termes de nombre de transmissions. Avant cela, nous exposons les hypothèses et les limites du schéma DPS.

3.4.1 Hypothèses

Dans le schéma DPS, il est supposé que la qualité des mesures fournies par le RCSF peut être considérée comme acceptable si les valeurs du nœud récepteur ne dépassent pas un certain seuil de tolérance. Étant donné que les nœuds capteurs peuvent comparer leurs prévisions avec des mesures réelles localement, aucune transmission ne sera nécessaire si une prédiction est précise, c'est-à-dire qu'elle ne viole pas la tolérance admise pour la valeur mesurée. Dans certains cas, les gestionnaires des RCSFs n'ont aucune information sur les statistiques des données qui seront récupérées par les nœuds capteurs. Ainsi, il peut être nécessaire de disposer d'une longue phase d'initialisation avant de commencer à faire des prédictions. Par exemple, les schémas qui utilisent des méthodes de prédiction avancées, comme les réseaux de neurones artificiels (RNA) qui nécessitent de grandes quantités de données pour trouver des modèles stables, en raison de leur grande complexité et du grand nombre de paramètres à estimer [95]. Cette phase n'est pas incluse dans ce modèle car il est supposé que le nœud récepteur tel que la station de base n'a pas de contraintes énergétiques.

Dans ce schéma DPS, on ne s'attend pas à des algorithmes distribués, c'est-à-dire que les nœuds capteurs n'ont pas à se synchroniser avec leurs voisins. Cependant, cela peut être facilement réalisé, compte tenu du nombre de voisins de chaque nœud capteur. De plus, les applications de suivi d'objets et de détection d'événements ne relèvent pas de ce schéma car elles ont généralement d'autres exigences que celles supposées dans le schéma DPS, comme une plus grande fiabilité ou des délais plus faibles [113].

Enfin, le schéma DPS est conçu pour représenter un RCSF moyen avec une connectivité entre ses nœuds capteurs au niveau de la couche d'application. Des capteurs défectueux peuvent affecter la distribution des communications et donc leur densité et forcer une nouvelle stratégie de routage. Par conséquent, les petits RCSFs avec une faible densité de nœuds capteurs nécessiteraient une attention particulière, car le nombre de fils d'un nœud capteur peut changer et donc le réseau sera surchargé avec des transmissions supplémentaires. D'un autre côté, les RCSFs grands et denses peuvent ne pas être affectés par un petit nombre de nœuds capteurs défectueux et le modèle peut toujours être valide.

3.4.2 Choix du modèle de prédiction et dissémination

Dans un schéma DPS, le même modèle de prédiction est partagé entre le nœud capteur et le nœud récepteur. Chaque nœud capteur (ou groupe de nœuds capteurs) a son modèle de prédiction, et les modèles de prédiction dans un RCSF peuvent être choisis indépendamment par les nœuds capteurs ou les nœuds récepteurs sans effectuer de nouvelles transmissions. Alternativement, les modèles de prédiction peuvent être choisis dans les nœuds récepteurs ou dans les nœuds capteurs. Dans le cas où des modèles de prédiction sont choisis dans des nœuds capteurs, les nœuds récepteurs doivent recevoir les valeurs

des paramètres et, dans certains cas, également la méthode de prédiction sélectionnée. En revanche, si les nœuds récepteurs (CH, BS) sont responsables du choix des modèles de prédiction, les nœuds capteurs doivent être informés des décisions prises.

En supposant que la diffusion d'un modèle de prédiction se fait par le biais d'une transmission unicast (monodiffusion), les nœuds capteurs dans le premier anneau recevront et retransmettront chaque information à leurs nœuds fils vers les destinations appropriées. Ainsi, si la station de base est responsable de la génération des modèles de prédiction, les nœuds capteurs du premier anneau devront communiquer les transmissions à leurs nœuds fils. Dans de tels cas, un nœud capteur dans le premier anneau recevra D^2 paquets. À partir de ces paquets, $(D^2 - 1)$ seront transmis à ses fils. Par conséquent, pour diffuser les modèles de prédiction générés dans la station de base, il y aura compte tenu de l'équation (3.9), $(2D^2 - 1)$ transmissions (somme des émissions et des réceptions) dans chaque nœud capteur du premier anneau.

$$\begin{aligned} E(X_{DIS-BS}) &= E(R_{i,1}^*) + E(S_{i,1}^*) \\ &= D^2 + (D^2 - 1) \\ &= 2D^2 - 1 \end{aligned} \quad (3.9)$$

De la même manière, dans le cas où des modèles de prédiction sont choisis dans les nœuds capteurs, chaque nœud capteur du premier anneau effectuera D^2 transmissions vers la station de base après avoir reçu des modèles de prédiction $(D^2 - 1)$ transmissions. Ainsi, le nombre total de transmissions au premier anneau sera, encore une fois, égal à $(2D^2 - 1)$.

Si les paquets du même sous-arbre sont agrégés, les nœuds capteurs du premier anneau ne recevront qu'un seul paquet qui sera divisé avant d'être retransmis à leurs nœuds fils directs dans le deuxième anneau. Dans de tels cas, un nœud capteur dans le premier anneau aura besoin de $(I_1 + 1)$ transmissions pour diffuser les modèles de prédiction comme défini dans l'équation (3.10) :

$$\begin{aligned} E(X_{DIS-BS-AGG}) &= E(X_{DIS-SN-AGG}) \\ &= I_1 + 1 \end{aligned} \quad (3.10)$$

d'où, à partir de l'équation (3.2) ($I_1 = 3$), si $d > 1$.

Enfin, si la station de base utilise des transmissions multicast (multidiffusions), les nœuds capteurs recevront et transmettront un seul paquet, c'est-à-dire :

$$E(X_{DIS-BS-BC}) = 1 \quad (3.11)$$

3.5 Impacts des prédictions sur le nombre de transmissions

Comme indiqué précédemment, l'adoption d'un schéma de prédiction de données peut bénéficier aux RCSFs pour réduire le nombre de transmissions et optimiser le contrôle d'accès au spectre du réseau, ce qui peut éventuellement réduire la consommation d'énergie et prolonger la durée de vie des RCSFs. Pour estimer le nombre de transmissions dans des réseaux homogènes, on s'appuie sur le cadre mathématique présenté par G.M. Dias et al. [100] décrivant les relations algébriques basées sur la précision des prédictions et

également la corrélation des données collectées. Une application sera détaillée au prochain chapitre.

Supposons que α_i est la précision des prédictions dans le nœud capteur i , c'est-à-dire α_i est la probabilité qu'une mesure de i corresponde à la prédiction et n'a pas à être transmise à la station de base, et $\alpha_i^c = 1 - \alpha_i$. Par conséquent, le nombre attendu de transmissions et de réceptions dans un nœud capteur i pendant un intervalle de temps de $(t = 1/f)$ secondes (c'est-à-dire entre deux mesures) est représenté respectivement par $E(S'_{i,d})$ et $E(R'_{i,d})$ tels que :

$$E(S'_{i,d}) = \alpha_i^c + \sum_{j \in H_{i,d}} \alpha_j^c \quad (3.12)$$

$$E(R'_{i,d}) = \sum_{j \in H_{i,d}} \alpha_j^c \quad (3.13)$$

Considérant une éventuelle dissémination des modèles de prédictions, le nombre attendu de transmissions et de réceptions durant une période de T secondes (temps choisi entre deux modèles de prédiction) est :

$$\begin{aligned} E(X'_{i,d}) &= E(S'_{i,d}) + E(R'_{i,d})fT + E(X_{DIS}) \\ &= (\alpha_i^c + \sum_{j \in H_{i,d}} \alpha_j^c + \sum_{j \in H_{i,d}} \alpha_j^c)fT + E(X_{DIS}) \\ &= (\alpha_i^c + 2 * \sum_{j \in H_{i,d}} \alpha_j^c)fT + E(X_{DIS}) \end{aligned} \quad (3.14)$$

On note qu'une faible précision dans les prédictions utilisées dans les nœuds capteurs éloignés de la station de base a un impact plus élevé sur le nombre total de transmissions du RCSF qu'une faible précision dans les prédictions utilisées dans les nœuds capteurs situés dans les premiers anneaux. Cependant, concernant le nombre de transmissions à un seul nœud capteur, le goulot d'étranglement du RCSF se situe toujours aux nœuds capteurs situés dans le premier anneau. Définissons la précision moyenne minimale (α_{min}) nécessaire pour réduire le nombre de transmissions en fonction de la taille du réseau et donc de son nombre d'anneaux. Cette valeur peut être utilisée pour définir le nombre maximal d'émissions $E(Max(S'_{i,d}))$ et de réceptions $E(Max(R'_{i,d}))$ dans un nœud de capteur i dans l'anneau d :

$$\begin{aligned} E(Max(S'_{i,d})) &= (1 - \alpha_{min}) + \sum_{j \in H_{i,d}} (1 - \alpha_{min}) \\ &= (1 + K_d)(1 - \alpha_{min}) \end{aligned} \quad (3.15)$$

$$\begin{aligned} E(Max(R'_{i,d})) &= \sum_{j \in H_{i,d}} (1 - \alpha_{min}) \\ &= K_d(1 - \alpha_{min}) \end{aligned} \quad (3.16)$$

En se rappelant que $K_d \triangleq |H_{i,d}|$, alors pour un nœud capteur i situé dans l'anneau d , on a :

$$E(X'_{i,d}) \leq ((1 + K_d)(1 - \alpha_{min}) + K_d(1 - \alpha_{min}))fT + E(X_{DIS}) \quad (3.17)$$

L'utilisation de prédictions réduira le nombre de transmissions si $X'_{i,d} < X_{i,d}$. Supposons un RCSF de prévision de données avec des nœuds capteurs homogènes transmettant périodiquement des mesures toutes les $(1/f)$ secondes, et un nœud capteur i situé dans l'anneau d ; selon l'équation (3.7), le nombre de transmissions en i dans une période de temps $(1/f)$ secondes est la somme $E(S_i) + E(R_i)$. Si le schéma DPS est adopté, le nombre

de transmissions sera $E(S'_i) + E(R'_i) + E(X_{DIS})$, comme défini dans l'équation (3.14). Par conséquent, un schéma DPS réduira le nombre de transmissions pendant une période T , si et seulement si :

$$(E(S'_i) + E(R'_i))fT + E(X_{DIS}) < (E(S_i) + E(R_i))fT \quad (3.18)$$

Ainsi, nous pouvons définir α_{min} comme la moyenne minimale des précisions des noeuds fils de i qui réduirait le nombre de transmissions dans un schéma DPS. Elle doit satisfaire l'équation suivante :

$$(E(S'_i) + E(R'_i))fT + E(X_{DIS}) = (E(S_i) + E(R_i))fT \quad (3.19)$$

En se basant sur les équations (3.4) et (3.5) on peut la calculer comme :

$$((K_d + 1)\alpha_{min}^c + (K_d\alpha_{min}^c))fT + E(X_{DIS}) = ((K_d + 1) + K_d)fT \quad (3.20)$$

On déduit donc l'équation :

$$\begin{aligned} ((K_d + 1)\alpha_{min}^c + (K_d\alpha_{min}^c))fT &= ((K_d + 1) + K_d)fT - E(X_{DIS}) \\ \alpha_{min}^c(D^2 + (D^2 - 1))fT &= (D^2 + (D^2 - 1))fT - E(X_{DIS}) \\ \alpha_{min}^c &= \frac{(D^2 + (D^2 - 1))fT - E(X_{DIS})}{(D^2 + (D^2 - 1))fT} \\ \alpha_{min}^c &= 1 - \frac{E(X_{DIS})}{(D^2 + (D^2 - 1))fT} \\ \alpha_{min} &= \frac{E(X_{DIS})}{(D^2 + (D^2 - 1))fT} \\ \alpha_{min} &= \frac{E(X_{DIS})}{(2D^2 - 1)fT} \end{aligned} \quad (3.21)$$

Nous arrivons à l'expression (3.22) qui donne le minimum de précision moyenne des prédictions :

$$\alpha_{min} > \frac{E(X_{DIS})}{(2D^2 - 1)fT} \quad (3.22)$$

La valeur $E(X_{DIS})$ est définie par l'équation (3.10) :

$$E(X_{DIS}) = E(S^*) + E(R^*) = 2D^2 - 1 \quad (3.23)$$

Donc :

$$\alpha_{min} = \frac{1}{fT} \quad (3.24)$$

L'équation (3.21) représente la limite inférieure de la précision minimale à atteindre si les modèles de prédictions ne sont pas générés indépendamment dans la station de base et les noeuds capteurs. Cette équation ne dépend que du nombre d'anneaux D , de la fréquence des mesures f et du temps choisi entre deux modèles de prédiction T . Les transmissions supplémentaires utilisées pour diffuser de nouveaux modèles de prédiction peuvent amener un schéma DPS à devenir inefficace. Par conséquent, l'efficacité d'un schéma DPS dépend également du nombre de transmissions nécessaires pour diffuser les modèles de prédiction parce que le nombre des transmissions sera proportionnel au nombre de sauts entre les noeuds capteurs et la station de base. De l'équation (3.21),

nous déduisons le nombre maximal de transmissions et de réceptions dans le capteur i du premier anneau comme :

$$N_b T = (1 - \alpha_{min})(2D^2 - 1)fT \quad \alpha_{min} \in [0, 1] \quad (3.25)$$

On remarque que le nombre de transmissions ne dépend pas de la densité des capteurs, mais du nombre d'anneaux D . La figure 3.6 montre l'évolution du nombre des transmissions en fonction des paramètres D et α_{min} en supposant arbitrairement $f = 2$ et $T = 24$. Par conséquent, avec des valeurs élevées de α_{min} le nombre de transmissions diminue linéairement en fonction du paramètre D .

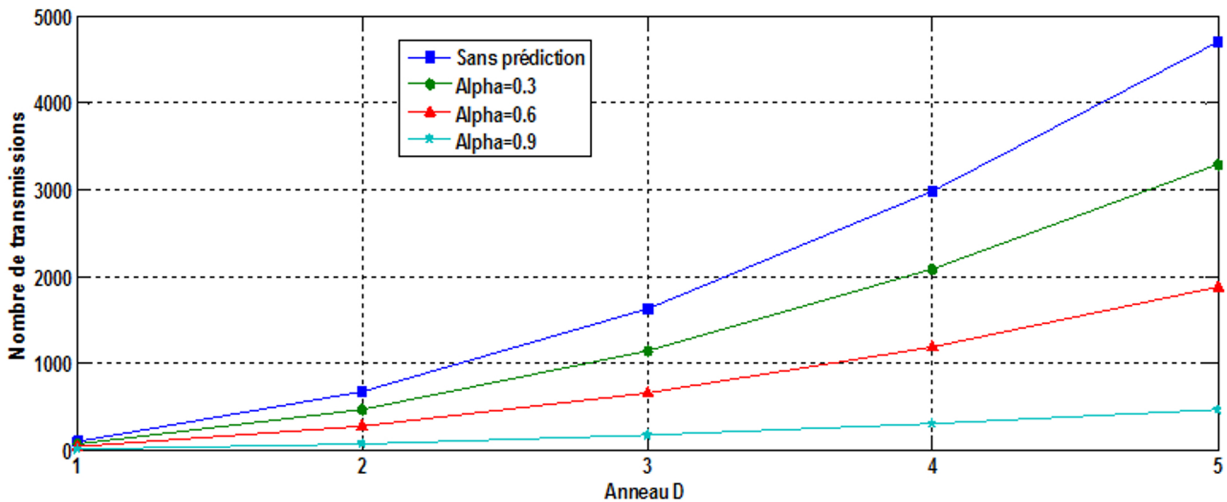


FIGURE 3.6 – Impact de la taille du réseau (D) sur le nombre de transmissions dans le premier anneau

3.6 Récapitulatif

Si les modèles de prédiction ne sont pas générés indépendamment dans les nœuds récepteurs et les nœuds capteurs, le schéma DPS requiert une précision minimale pour assurer la réduction du nombre de transmissions. Par conséquent, la précision minimale est une borne inférieure qui ne dépend que de la configuration du réseau (c'est-à-dire du nombre d'anneaux D), de la fréquence des mesures (f) et du temps entre le choix de deux modèles de prédiction (T). Si la précision des prédictions n'atteint pas cette limite, il y aura deux actions pour améliorer le fonctionnement du réseau, soit :

- Imposer de nouvelles valeurs pour f et T .
- Soit désactiver le schéma DPS.

Par ailleurs, si les modèles de prédiction sont générés indépendamment dans la station de base et les nœuds capteurs, on peut dire intuitivement que le schéma DPS ne nécessite pas de précision minimale pour assurer la réduction du nombre de transmissions.

Enfin, en se rappelant du modèle RCSF indiqué sur la figure 3.1, où chaque branche est connectée à la station de base via un nœud capteur, il a été possible de déduire le nombre de transmissions à ce nœud capteur en considérant uniquement le nombre de nœuds fils qu'il aurait. Une telle valeur a été définie comme K_1 et dépend du nombre d'anneaux dans le réseau. Par exemple, dans un RCSF avec D anneaux, ($K_1 = D^2 - 1$). Par conséquent, si on travaille avec $K_1 = 8$ nœuds capteurs, un RCSF avec trois anneaux aurait le même nombre de transmissions entre la station de base et le nœud capteur qui est le plus proche de la station de base. En d'autres termes, les neuf nœuds capteurs considérés pourraient représenter une branche du modèle avec $D = 3$ et $C = 3$. Étant donné que chaque branche du modèle devrait avoir le même nombre de nœuds capteurs et de transmissions, le pourcentage enregistré des transmissions sera également similaire au pourcentage enregistré des transmissions calculé pour un RCSF complet.

3.7 Agrégation de données

En plus des schémas DPS, il peut être possible d'adopter des schémas d'agrégation dans les nœuds capteurs, de sorte qu'un nœud capteur agrège les données reçues de ses fils et ne transmet qu'après avoir effectué sa propre mesure. Dans ce qui suit, on présente un schéma d'agrégation et on estime la réduction du nombre de transmissions si les deux techniques d'agrégation et de prédiction sont combinées. Pour être clair on introduit un scénario avec seulement deux nœuds capteurs afin de clarifier la normalisation des données et son application. Ensuite, on étend le modèle à un scénario plus complexe avec D anneaux.

3.7.1 Réseau avec deux nœuds capteurs

Considérons un réseau de capteurs composé d'une station de base (BS) et d'un nœud capteur i avec un seul nœud fils j . En raison de la topologie du réseau de capteurs, les transmissions du nœud capteur j peuvent atteindre la station de base uniquement via le nœud capteur i . Ainsi, toutes les $(1/f)$ secondes, le capteur i peut transmettre à la station de base si sa prédiction a échoué ou si celle du nœud j a échoué. On suppose que les mesures de i et j suivent des distributions normales représentées respectivement par :

$$Y_i = N(\mu_i, \sigma_i^2)$$

$$Y_j = N(\mu_j, \sigma_j^2)$$

Une telle distribution normale multivariée (NMV) peut être définie par la corrélation entre les valeurs de ces distributions, c'est-à-dire la relation entre chaque paire de mesures effectuées par i et j . Une illustration de la densité NMV contenant Y_i et Y_j est présentée sur la figure 3.7 où chaque ligne représente une densité différente de points.

Une distribution normale est caractérisée par sa fonction de densité de probabilité souvent rencontrée dans plusieurs types d'observations. Selon le théorème central limite, la distribution d'un échantillon de données de toute variable aléatoire indépendante a tendance à être normale, même si la distribution originelle à partir de laquelle la moyenne

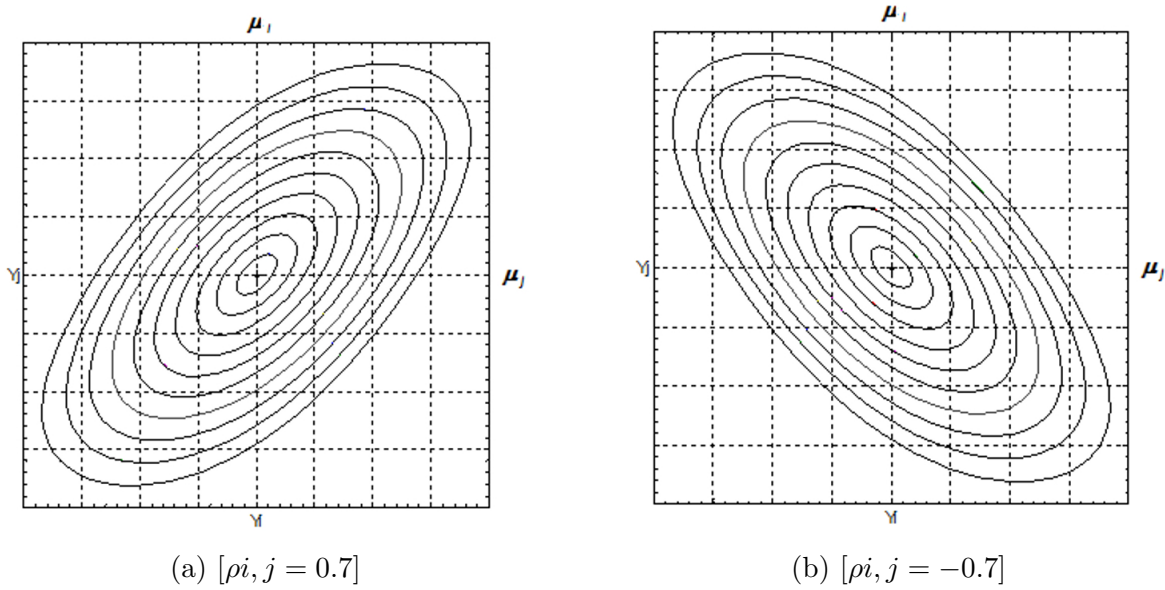


FIGURE 3.7 – Exemple de corrélation entre les valeurs Y_i et Y_j d'un couple de données

est calculée est décidément non normale. Par exemple, il est connu que les relevés environnementaux tels que la température, la pression, la luminosité et l'humidité effectués par des RCSFs en milieux externes peuvent être approchés par des distributions normales s'ils sont correctement gérés [99].

Supposons qu'un RCSF est composé d'un ensemble de nœuds capteurs S et que chaque nœud capteur $i \in S$ est responsable de la mesure d'un certain paramètre de l'environnement, de sorte que l'ensemble des observations suit une distribution normale avec une moyenne μ_i et une variance σ_i^2 . Une prédiction \bar{y}_i (par exemple, $\bar{y}_i = \mu_i$) peut être calculée par le nœud capteur i et la station de base. On définit le seuil d'acceptation ε_i , c'est-à-dire que la prédiction est considérée correcte si l'observation réelle (y_i) se trouve dans l'intervalle $[(\bar{y}_i - \varepsilon_i), (\bar{y}_i + \varepsilon_i)]$. En supposant que les données sont normalement distribuées, les chances d'observer une nouvelle valeur à l'intérieur de l'intervalle accepté peut être calculé en normalisant la valeur de ε_i , c'est-à-dire en la réécrivant en termes de variance σ_i^2 . La valeur normalisée de ε_i est représentée par z_i comme :

$$z_i = \frac{\varepsilon_i - \bar{y}_i}{\sigma_i} \quad (3.26)$$

Ainsi, la précision des prédictions (α_i) peut être calculée sur la base de la fonction de répartition multi-normale :

$$\phi_{\mu,\sigma}(x) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x - \mu}{\sigma\sqrt{2}} \right) \right] \quad (3.27)$$

Encore une fois, selon le théorème central limite, on suppose des prédictions non biaisées et les erreurs normalement distribuées. Par conséquent, le pourcentage d'observations qui tomberont en dehors de l'intervalle accepté est représenté par le test bilatéral Z (c'est-à-dire $2\phi(-|z_i|)$), et α_i est :

$$\alpha_i = 1 - 2\phi(-|z_i|) \quad (3.28)$$

En remplaçant l'équation (3.26) dans l'équation (3.27), on trouve que :

$$\alpha_i = 1 - 2\phi\left(-\left|\frac{\varepsilon_i - \bar{y}_i}{\sigma_i}\right|\right) \quad (3.29)$$

ce qui montre que la précision des prédictions dépend du seuil d'acceptation, de la moyenne et de la variance des données. En supposant que les prédictions (\bar{y}) ne sont pas biaisées (c'est-à-dire ($\bar{y} = \mu$)), on peut également les approximer en distribution normale et décider un résultat comme incorrect chaque fois qu'une mesure se situe en dehors de l'intervalle défini par le seuil de tolérance ε . Dans de tels cas, la probabilité que le nœud capteur j transmette (y compris sa mesure) après $(1/f)$ secondes est $(1 - \alpha_j)$. Par conséquent, la probabilité que i reçoive un paquet est également de $(1 - \alpha_j)$. De même, le nœud capteur i transmettra si la prédiction concernant sa mesure échoue (c'est-à-dire que la mesure est au-delà du seuil d'acceptation ε_i) ou si la prédiction dans le nœud capteur j avait échoué et une mesure a été reçue. En d'autres termes, il y aura une transmission si au moins l'une des deux prédictions échoue.

Si le capteur nœud i peut agréger des données, son nombre total de transmissions n'est pas une simple somme comme dans le cas d'une configuration sans agrégation car il dépend de la corrélation des mesures de i et j . Supposons que la corrélation entre Y_i et Y_j soit définie par le coefficient de corrélation de Pearson représenté par $\rho_{i,j}$, alors pour modéliser la probabilité d'avoir au moins une mauvaise prédiction, nous devons calculer la matrice de corrélation (Σ), qui est définie comme :

$$\begin{vmatrix} \sigma_i^2 & \rho_{i,j}\sigma_i\sigma_j \\ \rho_{i,j}\sigma_i\sigma_j & \sigma_j^2 \end{vmatrix} \quad (3.30)$$

Enfin, étant données les limites inférieures l_i et les limites supérieures u_i :

$$l_i = (\bar{y}_i - \varepsilon_i) \quad \text{et} \quad l_j = (\bar{y}_j - \varepsilon_j)$$

$$u_i = (\bar{y}_i + \varepsilon_i) \quad \text{et} \quad u_j = (\bar{y}_j + \varepsilon_j)$$

et la matrice de corrélation (Σ), il est possible de calculer la probabilité NMV suivante :

$$F(y_i, y_j) = \frac{1}{\sqrt{|\Sigma|}(2\pi)^2} \int_{l_i}^{u_i} \int_{l_j}^{u_j} e^{(\frac{1}{2}\theta^t \Sigma^{-1} \theta)d\theta} \quad (3.31)$$

La valeur de $F(y_i, y_j)$ représente la probabilité que les deux prédictions en (i et j) soient correctes et puissent être illustrées par la densité à l'intérieur du rectangle hachuré de la figure 3.8. Ainsi, la probabilité pour qu'au moins une prédiction échoue peut être calculée comme $(1 - F(y_i, y_j))$, qui, en fait, est la probabilité pour que le nœud capteur i effectue une transmission après $(1/f)$ secondes. Enfin, compte tenu des transmissions supplémentaires occasionnelles pour diffuser le modèle de prédiction, le nombre de transmissions de j pendant une période de T secondes peut être calculé comme :

$$[(1 - \alpha_j)fT + E(X_{DIS})]$$

et la somme attendue des transmissions et des réceptions au nœud capteur i pendant une période de T secondes peut être modélisée comme :

$$[(1 - F(y_i, y_j)) + (1 - \alpha_j)]fT + E(X_{DIS})$$

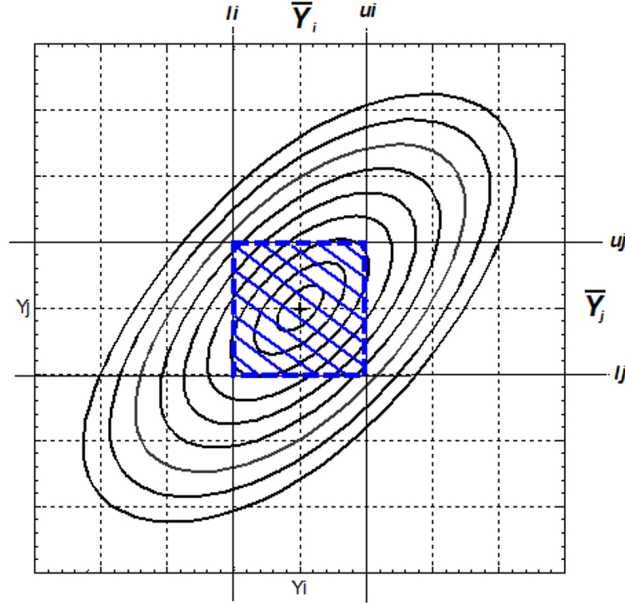


FIGURE 3.8 – Le rectangle bleu hachuré au centre représente la zone où les deux prédictions (y_i et y_j) sont correctes

3.7.2 Généralisation : Cas d'un réseau plus grand

Maintenant, on étend l'exemple précédent à de plus grands réseaux de capteurs. La matrice de corrélation (Σ) de plusieurs distributions de données peut être calculée comme :

$$\Sigma = \begin{vmatrix} \sigma_a^2 & \rho_{a,b}\sigma_a\sigma_b & \cdots & \rho_{a,z}\sigma_a\sigma_z \\ \rho_{b,a}\sigma_b\sigma_a & \sigma_b^2 & \cdots & \rho_{b,z}\sigma_b\sigma_z \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{z,a}\sigma_z\sigma_a & \rho_{z,b}\sigma_z\sigma_b & \cdots & \sigma_z^2 \end{vmatrix} \quad (3.32)$$

et, de manière similaire au modèle bidimensionnel, le nombre attendu de transmissions effectuées par le nœud capteur i dans l'anneau d (représenté par $E(S''_{i,d})$) dépend non seulement de ses prédictions mais également des prédictions utilisées dans tous ses fils. La valeur de $E(S''_{i,d})$ peut être calculé comme :

$$E(S''_{i,d}) = 1 - F(i, a, b, c, \dots, z)$$

où $(a, b, c, \dots, z) \in H_{i,d}$ et la fonction F est la fonction de probabilité NMV intégrée depuis les limites inférieures acceptées aux limites supérieures acceptées sur les $k = (1 + K_d)$ distributions.

$$F(i, a, b, c, \dots, z) = \frac{1}{\sqrt{|\Sigma|}(2\pi)^k} \int_{l_i}^{u_i} \int_{l_a}^{u_a} \int_{l_b}^{u_b} \cdots \int_{l_z}^{u_z} e^{(\frac{1}{2}\theta^t \Sigma^{-1} \theta)} d\theta \quad (3.33)$$

La fonction de répartition multi-normale $F(i, a, b, c, \dots, z)$ peut-être efficacement calculée en utilisant les méthodes de Monte Carlo pour des dimensions élevées [114].

Le nombre de réceptions en i , $E(R''_{i,d})$, est légèrement différent de l'exemple précédent, car maintenant le nœud capteur peut avoir plusieurs fils dans l'anneau suivant et leurs transmissions se produisent indépendamment. Définissons $H'_{i,d}$ comme l'ensemble des fils directs de i . Ainsi, $H'_{i,d} \triangleq I_d$. Le nombre attendu de réceptions peut être calculé comme :

$$E(R''_{i,d}) = \sum_{j \in H'_{i,d}} E(S''_{j,d+1}) \quad (3.34)$$

et le nombre total des transmissions et réceptions devient :

$$E(X''_{i,d}) = (E(S''_{i,d}) + E(R''_{i,d}))fT + E(X_{DIS}) \quad (3.35)$$

Même si la fonction $F(i, a, b, c, \dots, z)$ n'a pas d'expression approchée, il est possible de définir une borne inférieure en fonction du cas où il n'y a absolument aucune corrélation entre les valeurs mesurées par le nœud capteur i et ses fils. Lorsque la corrélation est égale à zéro (c-à-d qu'il y a indépendance en probabilité des différentes mesures), le nombre attendu de transmissions et de réceptions au nœud capteur i est maximum. Étant donné qu'il existera une transmission si au moins une prédiction échoue, la probabilité de n'avoir aucune transmission à i est :

$$P = \prod_{i=1}^{k_d+1} \alpha_i$$

En supposant que les prédictions au niveau de chacun des nœuds capteurs s'effectuent avec un même niveau de précision α alors cette probabilité P s'écrit :

$$P = \alpha^{k_d+1}$$

Donc :

$$E(\text{Max}(S''_{i,d})) = 1 - \alpha^{K_d+1} \quad (3.36)$$

En se rappelant que i devrait avoir des fils directs I_d et que chaque fils fasse partie d'un sous-arbre avec K_d/I_d capteurs nœuds. Il peut exister des réceptions indépendantes de I_d en i , et chaque réception peut ne pas se produire avec la probabilité α^{K_d+1} . Donc :

$$E(\text{Max}(R''_{i,d})) = I_d(1 - \alpha^{K_d/I_d}) \quad (3.37)$$

Par conséquent :

$$E(X''_{i,d}) \leq [(1 - \alpha^{K_d+1}) + I_d(1 - \alpha^{K_d/I_d})]fT + E(X_{DIS}) \quad (3.38)$$

Il est affirmé selon G.M. Dias et al. [100] que $E(X''_{i,d}) \leq E(X'_{i,d})$, ce qui signifie qu'un mécanisme qui agrège les données ne fera pas plus de transmissions que celui qui ne fait que des prédictions. En comparant l'équation (3.38) avec l'équation (3.17), alors pour tout $\alpha \in [0, 1]$ et $K_d \geq 0$, il peut être montré que :

$$(1 - \alpha^{K_d+1}) < (K_d + 1)(1 - \alpha)$$

Par conséquent :

$$E(\text{Max}(S''_{i,d})) < E(\text{Max}(S'_{i,d}))$$

De plus :

$$E(\text{Max}(R''_{i,d})) < E(\text{Max}(R'_{i,d})) \text{ et } I_d(1 - \alpha^{\frac{K_d}{I_d}}) < K_d(1 - \alpha)$$

Ceci peut être prouvé de façon similaire, puisque $(K_d/I_d) \geq 1$ lorsque $K_d > 0$ et $\alpha \in [0, 1]$. Dans le cas où l'on soit dans le dernier anneau, aucune réception n'est effectuée puisqu'il n'y a pas de fils ($K_d = I_d = 0$). Pour démontrer cette affirmation, supposons deux valeurs α et x tel que $\alpha \in [0, 1]$ et $x \geq 1$. On veut montrer que $(1 - \alpha^x) \leq (1 - \alpha)x$:

$$\begin{aligned} (1 - \alpha^x) &\leq (1 - \alpha)x \\ (1 - \alpha^x) &\leq (x - \alpha x) \\ \alpha^x &\geq 1 + (1 - \alpha)x \end{aligned} \tag{3.39}$$

Quand $\alpha = 0$ ou $\alpha = 1$ on peut facilement constater que l'affirmation est vraie car $x \geq 1$ par définition. Pour les autres valeurs de α , on peut utiliser l'inégalité de Bernoulli :

$$(1 + i)^j \geq 1 + i * j$$

où $i > -1$ et $i \neq 0$ est un nombre réel et $j \geq 2$ une valeur entière. En substituant les valeurs de α et x dans l'équation (3.39) respectivement par $i + 1$ et j , la preuve est faite.

Étant donné que les goulots d'étranglement d'un RCSF sont les nœuds capteurs situés dans le premier anneau et en considérant les extra-transmissions occasionnelles pour disséminer les modèles de prédictions, alors le nombre total de transmissions et de réceptions au niveau d'un nœud capteur i situé dans l'anneau $d = 1$ peut être calculé comme :

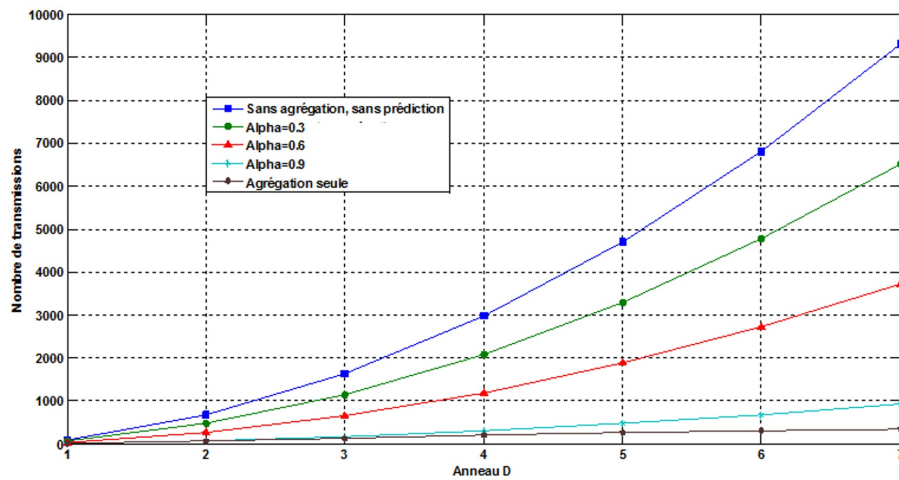
$$NbT_{ag} = [(1 - P(K_1, \alpha_{moy}, \rho)) + I_d(1 - (\alpha_{moy}^{K_d/I_d}))]fT + I_1t \tag{3.40}$$

où $P(K_1, \alpha_{moy}, \rho)$ représente la probabilité de ne faire aucune transmission dans un groupe de n nœuds capteurs mesurant des données avec une corrélation ρ si la précision moyenne des prédictions est α_{moy} . Cette probabilité est obtenue à partir du calcul de l'équation (3.33). Le paramètre I_d représente le nombre de nœuds fils du capteur situé dans le premier anneau. On se rappelle que $I_d = 3$ si $D > 1$. Le paramètre ' t ' définit la période de temps pour actualiser le modèle. Le terme I_1t représente les extra-transmissions pour disséminer le modèle de prédiction.

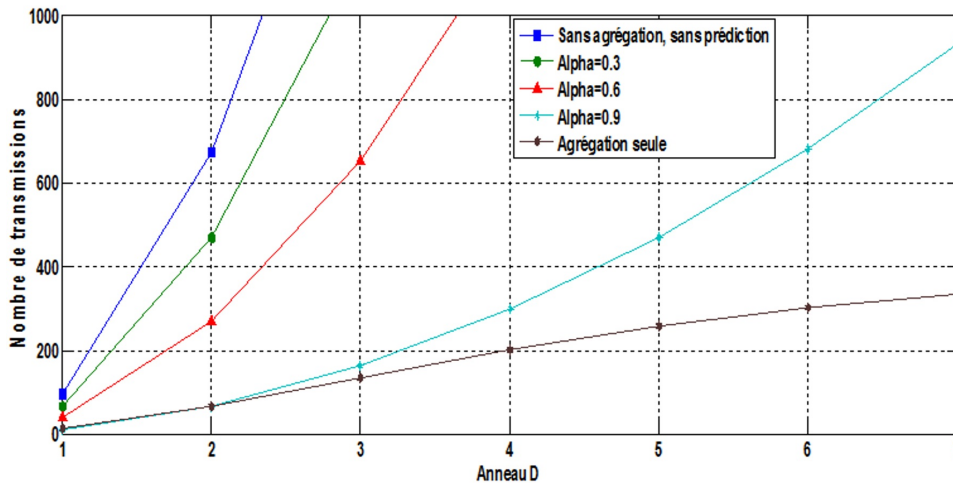
3.8 Impact des schémas d'agrégation et de prédiction de données sur le nombre de transmissions

Nous considérons le premier anneau où les nœuds capteurs constituent les goulots d'étranglement. Dans une première étape on applique séparément (c-à-d sans combinaison), le schéma d'agrégation sans corrélation de données et le schéma de prédiction duale. Les modèles de prédiction ne sont pas générés indépendamment dans la station de base et les nœuds capteurs. Après calcul des équations (3.25) et (3.38), la figure 3.9 montre l'évolution du nombre de transmissions pour un RCSF d'une taille arbitrairement choisie

égale à $D = 7$. L'évolution du nombre de transmissions selon α_{min} est tracée en supposant toujours $f = 2$ et $T = 48$. Le nombre attendu de transmissions au nœud capteur i est le maximum possible car dans un premier temps aucune corrélation n'a été considérée entre les mesures détectées. Il est utile de remarquer la supériorité de la technique d'agrégation quelque soit la taille D du réseau. Donc l'agrégation de données a un impact plus élevé que la prédiction de données sur le nombre de transmissions comme le montre la figure 3.9b. Ce résultat ne corrobore pas celui rapporté dans [100] où les auteurs concluent que l'agrégation a un impact supérieur à la prédiction uniquement lorsque la taille du réseau est assez grande ($D > 4$). Le taux de transmissions varie entre 4% et 20% selon la taille du réseau dans notre exemple.



(a)



(b)

FIGURE 3.9 – Impact de la taille du réseau sur le nombre de transmissions dans le premier anneau : Application de l'agrégation et du schéma DPS séparément

Dans une seconde étape, nous considérons la combinaison entre le schéma de prédiction duale DPS et le processus d'agrégation. Cette fois, nous analysons l'impact de la corrélation entre chaque paire de données collectées sur le nombre de transmissions quand on exécute l'agrégation. Nous considérons la corrélation linéaire par paires de Pearson. Nous

supposons que les mesures suivent une distribution normale représentée par $Y_i = N(\mu_i, \sigma_i)$ qui est caractérisée par sa fonction de densité de probabilité souvent rencontrée dans plusieurs types d'observations. On rappelle à titre d'exemple qu'il est connu que les relevés météorologiques tels que la température, l'humidité et la pression atmosphérique effectués par les RCSFs peuvent être valablement décrits par des distributions normales s'ils sont correctement gérés. Pour simplifier nos calculs, nous considérons une distribution de données autour d'une moyenne nulle avec une variance unitaire. Dans ce cas nous pouvons varier aisément le coefficient de corrélation selon la relation :

$$\rho_{xy} = \frac{COV(X, Y)}{\sigma_x \sigma_y}$$

La probabilité cumulée de la distribution normale multivariée est calculée selon l'équation (3.40). La taille du RCSF est prise égale à $D = 2$. En affectant au taux de transmissions obtenu par le seul processus d'agrégation de données un pourcentage de 100% (c'est-à-dire que le pourcentage de transmissions obtenu par la seule agrégation est considéré comme une référence), la figure 3.10 montre que la combinaison entre l'agrégation et le schéma DPS réduit davantage le taux de transmissions global. Les résultats obtenus illustrent l'évolution du nombre de transmissions en fonction du coefficient de corrélation ρ et de la précision moyenne minimale α_{min} du modèle DPS. On peut voir que les flux de transmissions de données diminuent avec l'augmentation de ces deux paramètres. On constate ici que l'impact du coefficient de corrélation n'est pas très significatif. Cependant, la précision du modèle a un impact beaucoup plus important. Le gain supplémentaire maximum obtenu ici est de 90%. Sachant que le taux de transmissions maximal atteint par l'agrégation seule est de 55.9% alors le taux de transmissions total est le produit : $10\% \times 55.9\% = 5.59\%$.

3.9 Synthèse

Deux remarques clés sont à préciser dans ce paragraphe. La première concerne la droite d'ajustement entre les variables aléatoires corrélées et la seconde concerne l'hypothèse d'identité des précisions des prédicteurs des nœuds capteurs comme nous l'avons considéré dans l'équation (3.36).

3.9.1 Droite d'ajustement en corrélation linéaire

Il est important de rappeler ici que lorsque deux variables aléatoires (ou deux distributions de données statistiques) sont linéairement corrélées, la droite d'ajustement de la variable aléatoire Y par rapport à la variable aléatoire X est approximée au sens des moindres carrés par l'expression [115] :

$$Y - \bar{Y} \cong \rho_{XY} \frac{\sigma_Y}{\sigma_X} (X - \bar{X}) \quad (3.41)$$

où ρ_{XY} est le coefficient de corrélation linéaire des variables aléatoires Y et X , \bar{Y} et \bar{X} sont les moyennes respectives de ces variables aléatoires et enfin σ_Y et σ_X leur écarts types respectifs.

Si l'on pose :

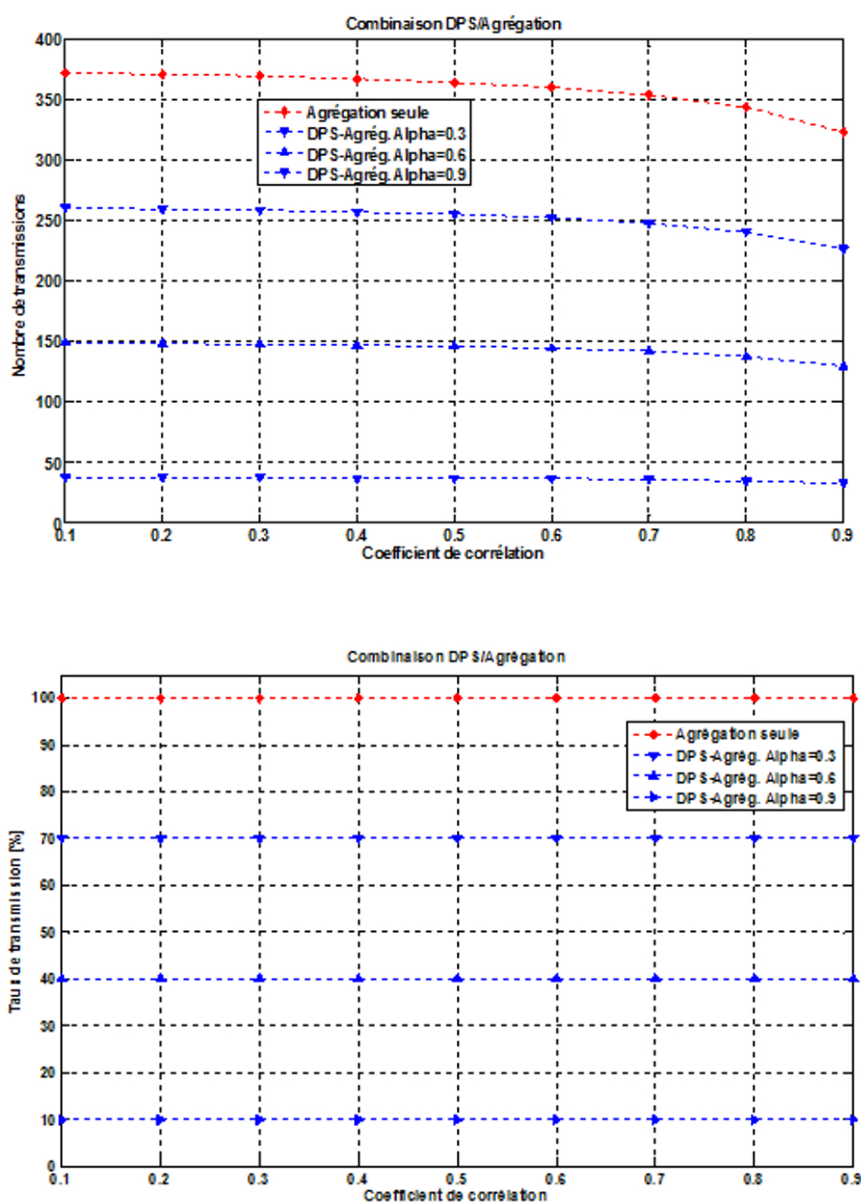


FIGURE 3.10 – Combinaison Agrégation-DPS

$$Z = (Y - \bar{Y}) - \rho_{XY} \frac{\sigma_Y}{\sigma_X} (X - \bar{X})$$

alors Z est une variable aléatoire ayant pour valeurs les erreurs de l'approximation (3.41). L'erreur moyenne quadratique Z^2 s'écrit :

$$E(Z) = E[(Y - \bar{Y}) - \rho_{XY} \frac{\sigma_Y}{\sigma_X} (X - \bar{X})]^2 = \sigma_Y^2 (1 - \rho_{XY}^2) \quad (3.42)$$

A partir de l'équation (3.42), on voit bien que l'erreur moyenne est égale à :

$$\sigma_Y \sqrt{(1 - \rho_{XY}^2)}$$

La plus grande valeur de cette erreur est égale à σ_Y si $\rho_{XY} = 0$ (c-à-d les variables Y et X sont non corrélées), et la plus petite valeur est égale à zéro si $\rho_{XY}^2 = 1$. Dans ce dernier cas les variables aléatoires Y et X prennent avec certitude (c-à-d une probabilité égale à 1), des valeurs appartenant à la droite d'ajustement :

$$Y - \bar{Y} = \rho_{XY} \frac{\sigma_Y}{\sigma_X} (X - \bar{X})$$

En conséquence, lorsqu'un processus d'agrégation avec corrélation de données est appliqué, il faudra s'assurer que la tolérance de l'utilisateur soit toujours supérieure ou égale à l'écart type de la variable à déduire par voie de corrélation. Sinon, il y a risque d'avoir des valeurs hors tolérance notamment pour des corrélations faibles, ce qui entachera d'erreur la mise à jour des modèles de prédiction. Autrement dit, la droite d'ajustement devrait nous permettre de déduire à partir des valeurs mesurées d'une distribution aléatoire celles qui lui sont corrélées avec une précision telle que les valeurs déduites soient comprises dans l'intervalle de tolérance imposée par l'utilisateur. Pour cela on doit respecter la condition :

Tolérance-utilisateur \geq Ecart type de la variable aléatoire à déduire par corrélation

3.9.2 Précisions des prédicteurs

Nous avons supposé dans la section (3.7.2) que la probabilité de n'avoir aucune transmission du nœud capteur i situé dans l'anneau d est :

$$P = \alpha^{1+k_d}$$

c'est-à-dire que tous les prédicteurs opèrent avec la même précision α . Par conséquent, si au moins un prédicteur d'un nœud capteur fils de i viole le seuil de tolérance imposé par l'utilisateur alors tous les autres nœuds capteurs fils de i violent également ce seuil et transmettent toutes leurs mesures au capteur i . Il s'agit en fait de K_d réceptions pour le nœud capteur i durant une période de temps égale à $(1/f)$. Dans ce cas le nœud capteur i agrège ces K_d réceptions et envoie le résultat vers le nœud capteur de l'anneau suivant. Cependant, si les prédicteurs opèrent avec des précisions différentes alors la probabilité de n'avoir aucune transmission du nœud capteur i situé dans l'anneau d s'écrit :

$$P = \prod_{i=1}^{k_d+1} \alpha_i$$

Dans ce cas, si un prédicteur ou un petit ensemble ' n ' de prédicteurs tel que $n < K_d$, deviennent défaillants, c'est-à-dire qu'ils violent le seuil de tolérance de l'utilisateur, le nombre de réceptions du nœud capteur i est inférieur à K_d et donc la charge du processus d'agrégation de ces réceptions serait plus légère. En conséquence, en considérant l'hypothèse $P = \alpha^{1+k_d}$, on se place dans la situation la moins favorable. Donc les résultats présentés sur la figure 3.9 concernant le nombre de transmissions dû à l'agrégation seule est maximal. Ce qui veut dire qu'en pratique l'agrégation peut être meilleure.

3.10 Conclusion

Dans ce chapitre nous avons commencé par décrire le fonctionnement des paradigmes de prédiction de données en indiquant que l'algorithme DPS exige un niveau minimum de précision α_{min} lorsque les modèles de prédiction ne sont pas sélectionnés indépendamment dans le nœud capteur et la station de base. Ceci, dans le cadre d'une topologie RCSF uniforme sous forme d'anneaux où les nœuds capteurs des premiers anneaux sont les plus exposés au goulot d'étranglement ou au bourrage de données.

Ensuite, une comparaison entre l'agrégation et le schéma DPS a clairement montré la supériorité de la technique d'agrégation quelque soit la taille du réseau considéré et ce, malgré le fait que le taux de transmissions atteint par l'agrégation était maximal car la corrélation des données n'a pas été prise en compte. Ce résultat ne corrobore pas celui rapporté par G.M. Dias et al. [100]. Après cela, nous avons supposé que les données obéissent à une distribution normale, ce qui est le cas d'une très grande majorité d'applications, afin d'analyser l'impact des corrélations de données. Ainsi, on a montré que dans ce cas la combinaison de l'agrégation et de l'algorithme DPS réduit davantage le taux de transmissions. Les simulations ont montré que la précision des modèles de prédictions n'a d'impact substantiel qu'aux faibles corrélations entre les données collectées. Au contraire, l'impact de la corrélation devient plus fort aux grandes valeurs des coefficients de corrélation. À notre connaissance, le premier travail qui a modélisé l'impact des schémas DPS sur les réseaux de capteurs et a montré leur potentiel dans toute une classe d'applications est celui publié dans [79]. Les auteurs ont comparé les valeurs obtenues en utilisant l'approche décrite plus haut avec celles qui auraient été obtenues en utilisant des données réelles. Ils ont observé que cette approche fournit une bonne estimation du nombre de transmissions qui pourrait se produire en pratique. Le prochain chapitre portera sur notre contribution quant à l'amélioration du schéma DPS et à la comparaison des résultats avec l'agrégation.

Chapitre 4

Un modèle basé sur l'agrégation et
la prédiction duale de données pour
les prévisions dans les RCSFs

Chapitre 4

Un modèle basé sur l'agrégation et la prédiction duale de données pour les prévisions dans les RCSFs

4.1 Introduction

Actuellement, seule l'analyse des travaux réalisés dans différentes applications peut aider à la sélection des schémas de prédiction appropriés pour application dans les RCSFs. Un défi pour les travaux futurs consiste à développer une méthodologie pour évaluer l'efficacité de la technique de prédiction pour la réduction des données. L'approche devrait être basée non seulement sur des considérations financières en raison des coûts de communication mais également sur la sélection du modèle de prédiction, son emplacement de mise en œuvre c'est-à-dire où les prédictions seront effectuées et sur les coûts générés impliqués dans leur calcul.

Les questions ouvertes et les défis futurs incluent le problème des prévisions à long terme. Les travaux existants se concentrent principalement sur la prévision des mesures qui vont se faire à court terme. La prévision à des intervalles de temps plus longs peut fournir une perspective sur le moment où la précision des prévisions à court terme diminuera. En tant que solution partielle, les modèles de prédiction peuvent être mis à jour ou de nouveaux schémas de prédiction peuvent être adoptés avant que la qualité ne diminue. Cela peut apporter des avantages qui ne sont pas remis en cause dans l'état actuel de la technique, en termes d'économie d'énergie, d'accès au réseau et de qualité des informations.

L'analyse comparative entre l'algorithme de prédiction DPS et l'agrégation de données, effectuée dans le chapitre précédent, nous a permis d'identifier le lest du schéma DPS. Dans ce chapitre, on va illustrer l'extension de l'algorithme "DPS" ou de sa version ADPS "ADPS" vers un nouvel algorithme désigné "Extended Adaptive Dual Prediction schema (EADPS)" [116] pour réduire davantage les taux de transmissions de données. L'idée innovante est basée sur l'élimination de la table historique utilisée dans l'algorithme ADPS pour actualiser le modèle de prédiction lorsqu'il dérive. Une application utilisant des données réelles issues de capteurs de données météorologiques est réalisée comme première expérimentation. Bien sûr, il s'agit de séries chronologiques. Le but est d'apprécier la capacité de l'algorithme innovant EADPS aidant à réduire les taux de trans-

missions de données et donc à réduire la consommation d'énergie des batteries des nœuds, à améliorer la disponibilité du stockage, à fournir plus d'informations aux utilisateurs et à améliorer l'évolutivité attendue à l'avenir. Par ailleurs, compte tenu de la pertinence des techniques d'agrégation, celle-ci devrait être combinée à ce nouvel algorithme afin d'étudier son impact. En somme, il s'agit de contrôler la communication RCSF-Cloud dans l'optique d'une intégration de nos outils pour les fournisseurs du Cloud ce qui fera l'objet de notre prochain chapitre.

4.2 Amélioration du paradigme de prédiction ADPS

On comprend à partir du chapitre (3) que le paradigme DPS est actuellement le mieux considéré comparativement au paradigme SPS car il remédie à ses inconvénients intrinsèques. Le caractère adaptatif du schéma DPS le rend encore plus efficace. Lorsqu'on utilise le paradigme ADPS, de temps à autre, la précision des prédictions est perdue. Ainsi, les nœuds capteurs et les nœuds récepteurs entament une nouvelle phase d'initialisation pour actualiser les modèles de prédiction. On élimine donc de facto les extra-transmissions dues à la dissémination des modèles mis à jour. Les paramètres de ces modèles doivent être mis à jour en raison de l'écart excessif des valeurs estimées par rapport à celles détectées par les nœuds capteurs. Par conséquent, le nœud capteur et le nœud récepteur exploitent une nouvelle et même fenêtre d'estimation de données pour mettre à jour les paramètres du modèle. Cette fenêtre d'estimation doit être suffisamment longue pour permettre une estimation correcte et précise des nouveaux paramètres du modèle.

La fréquence de ce processus adaptatif dépend essentiellement de la qualité du modèle construit. Le nœud récepteur peut assumer la responsabilité de contrôler les opérations des nœuds capteurs en fonction des économies potentielles que les prédictions peuvent introduire. Dans de tels cas, le nœud récepteur peut estimer s'il vaut la peine de faire des prédictions dans les nœuds capteurs, sur la base de la relation entre la précision des prédictions, l'erreur tolérée par l'utilisateur et éventuellement la corrélation entre les mesures.

Nous avons conçu un nouvel algorithme désigné EADPS qui repose sur le principe d'éviter la fenêtre d'estimation de données dans le schéma ADPS pour mettre à jour les paramètres du modèle lorsqu'il devient obsolète. En effet, au lieu de la fenêtre d'estimation, nous utilisons la collection des modèles obsolètes ayant déjà servis pour la prédiction des données lors des séquences antérieures de l'algorithme ADPS, et cela dès qu'il y a violation de la tolérance imposée par l'utilisateur. Le nouveau prédicteur peut être déterminé au niveau du nœud capteur et diffusé vers le nœud récepteur ou réciproquement, ou déterminé simultanément dans le nœud capteur et le nœud récepteur. On comprend tout de suite que la croissance de la population des modèles stockés améliore la qualité d'estimation des nouveaux paramètres du modèle. En résumé, nous étendons le paradigme de communication ADPS pour prolonger au maximum la durée de vie des RCSFs. Le procédé consiste à générer d'abord, l'algorithme ADPS où le nœud capteur et le nœud récepteur peuvent calculer le même modèle de prédiction car ils sont programmés pour utiliser les mêmes données, ensuite, notre algorithme se mettra en service une fois que la banque initiale des modèles obsolètes sera suffisamment riche. Cette période d'initialisation dépend de l'application et du type de modèle sélectionné. Durant cette période, le modèle de prédiction doit être mis à jour lorsqu'il dérive, puis stocké. Ainsi, lors de cette phase

d'initialisation, on collecte l'échantillon initial des modèles de prédiction éliminés qui ont déjà servis précédemment sur le RCSF. Cet échantillon constitue le premier ensemble de données nécessaires à la prédiction du nouveau modèle. La taille de cet échantillon sera déterminée en fonction d'un critère que nous choisissons comme :

$$VC = 100 \frac{1 - \text{norm}(\hat{y} - y)}{\text{norm}(\hat{y} - \text{mean}(y))} \quad (4.1)$$

$$\text{où } \text{norm}(\hat{y} - y) = \sum_{i=1}^p \sqrt{(\bar{y}_i - y_i)^2}$$

où la fonction $\text{norm}(\hat{y} - y)$ calcule la distance euclidienne, \bar{y}_i et y_i sont ici respectivement les paramètres prédits du nouveau modèle et ceux stockés au cours du temps. Le paramètre p est l'ordre du modèle. Celui-ci représente le nombre de paramètres de la structure mathématique du modèle sélectionné. Les algorithmes (1) et (2) que nous avons développés comme une extension du paradigme ADPS. Ces algorithmes mis en œuvre au niveau des nœuds capteurs et des nœuds récepteurs sont des algorithmes légers. Ils ne prennent pas en compte les données inutiles, donc la complexité en termes de stockage ou de durée de fonctionnement ne sera pas significative. De plus, nous pouvons concevoir un modèle de transmission pour les RCSFs qui prend en compte l'évolution du matériel et donc l'utilisation d'algorithmes de haute complexité pour que les prédictions n'auraient pas d'impact négatif sur les performances des algorithmes utilisés dans le cas où nous impliquons des services cloud qui ont de hautes performances et dans lesquelles l'étude de la complexité n'est généralement pas prise en compte.

Pour illustrer l'intérêt de notre travail, nous considérons les systèmes de prévision à l'aide de systèmes des RCSFs avec des nœuds capteurs homogènes organisés selon le modèle en anneau décrit dans le chapitre (3). On rappelle que les goulots d'étranglement dans ce modèle correspondent aux nœuds capteurs situés dans le premier anneau parce que ces derniers sont toujours impliqués pour relayer les données collectées par les nœuds capteurs fils vers le nœud récepteur (CH, passerelle ou station de base) ou vice-versa. De plus, le schéma de routage des données est très simple dans ce modèle d'anneaux car un nœud situé à l'anneau (d) envoie ses données collectées à l'un de ses voisins situés dans l'anneau ($d - 1$) et ce processus est répété de la même manière jusqu'à ce que les données envoyées atteignent la station de base (le nœud BS). Ce schéma de routage des données est basé sur une approche de transfert gloutonne comme dans "Greedy Perimeter Stateless Routing (GPSR)" [117]. Les capteurs mesurent et transmettent des données à un intervalle de temps prédéfini pouvant varier de quelques fractions de secondes à quelques heures. Par conséquent, les données sont transmises périodiquement avec une fréquence f à partir d'un nœud capteur i situé dans l'anneau d . De plus, les capteurs passent en mode veille pour ne pas épuiser leurs batteries rapidement. La période de ce mode est égale à la durée qui sépare deux transmissions consécutives. Ceci est illustré par les instructions 14, 22 et 27 de l'algorithme (1).

Algorithm 1 : Étapes de l'algorithme pour le nœud capteur

```

1: Begin
2: Initialiser paramètres ;
3: Détecter les données ;
4: if (le modèle de prédiction est reçu du BS) then
5:   Calculer le taux d'erreur  $e = \text{donnée prédite-donnée mesurée}$  ;
6:   if ( $e \leq e_{Max}$ ) then
7:     Continuer la détection de données ;
8:   else
9:     Transmettre ;
10:  end if
11: else
12:   Compteur++ ;
13:   Transmettre ;
14:   Se mettre en veille ;
15: end if
16: for (chaque intervalle d'échantillonnage fixe) do
17:   if (Compteur > Taille de la table historique) then
18:     Appel du modèle de prédiction ;
19:     if ((prédiction – donnée mesuré > erreur limite)) then
20:       Transmettre ;
21:       Mettre à jour la table d'historique avec la donnée détectée ;
22:       Se mettre en veille ;
23:     end if
24:   else
25:     Mettre à jour la table d'historique avec les données prédite ;
26:     Éliminer la donnée détectée ;
27:     Se mettre en veille ;
28:   end if
29: end for
30: End

```

4.3 Réduction des transmissions à l'aide du schéma EADPS

Similairement à l'équation (3.18) du chapitre (3), lorsque notre algorithme EADPS décrit par les algorithmes (1) et (2) est appliqué, le taux de transmission global du capteur i dans l'anneau d est réduit pendant une période T , si et seulement si :

$$(E(S''_i) + E(R''_i))fT + E(X''_{DIS}) \leq (E(S'_i) + E(R'_i))fT + E(X'_{DIS}) \quad (4.2)$$

Le terme de gauche de l'équation (4.2) représente le nombre de transmissions correspondant à l'algorithme EADPS tandis que le terme de droite de cette équation représente le nombre de transmissions correspondant à l'algorithme ADPS classique.

Algorithm 2 : Étapes de l'algorithme pour la BS

```

1: Begin
2: Initialiser paramètres ;
3: Attendre les données capturées des nœuds capteurs ;
4: if (donnée capturée est reçue) then
5:   Compter les modèles prédictifs dans la table de collection ;
6:   if (Taille de la table de collection  $\geq$  la taille minimal) then
7:     Construire nouveau modèle ;
8:     Transmettre le nouveau modèle ;
9:     Prédire les données ;
10:    Mettre à jour la table de collection ;
11:  else
12:    Compteur++ ;
13:  end if
14: end if
15: for (chaque intervalle d'échantillonnage fixe) do
16:   if (Compteur < Taille de la table historique) then
17:    Mettre à jour la table d'historique avec les données détectées ;
18:   else
19:    if (Compteur  $\geq$  Taille de la table historique) then
20:      Mettre à jour la table d'historique avec les données détectées reçus ;
21:      Break ;
22:    end if
23:   else
24:    Appel du modèle de prédiction ;
25:    Mettre à jour la table d'historique avec les données prédite ;
26:    Stocker le modèle prédictif dans la table de collection ;
27:   end if
28: end for
29: End

```

$$\begin{aligned} E(S'_i) &= \alpha_{min}^c D^2 fT \\ E(R'_i) &= \alpha_{min}^c (D^2 - 1) fT \end{aligned} \quad (4.3)$$

où $\alpha_{min}^c = 1 - \alpha_{min}$

Compte tenu de l'équation (4.3) on peut déduire :

$$\begin{aligned} E(S''_i) &= \alpha_{min}^c \beta_{min}^c D^2 fT \\ E(R''_i) &= \alpha_{min}^c \beta_{min}^c (D^2 - 1) fT \end{aligned} \quad (4.4)$$

La moyenne minimale entre les précisions des fils du nœud i qui réduirait le nombre de transmissions dans un schéma EADPS doit satisfaire l'équation (4.5) :

$$\underbrace{(E(S''_i) + E(R''_i)) fT + E(X''_{DIS})}_{\text{Algorithme ADPS étendu 'EADPS'}} = \underbrace{(E(S'_i) + E(R'_i)) fT + E(X'_{DIS})}_{\text{Algorithme ADPS classique 'ADPS'}} \quad (4.5)$$

En substituant les équations (4.3) et (4.4) dans l'équation (4.5) on obtient :

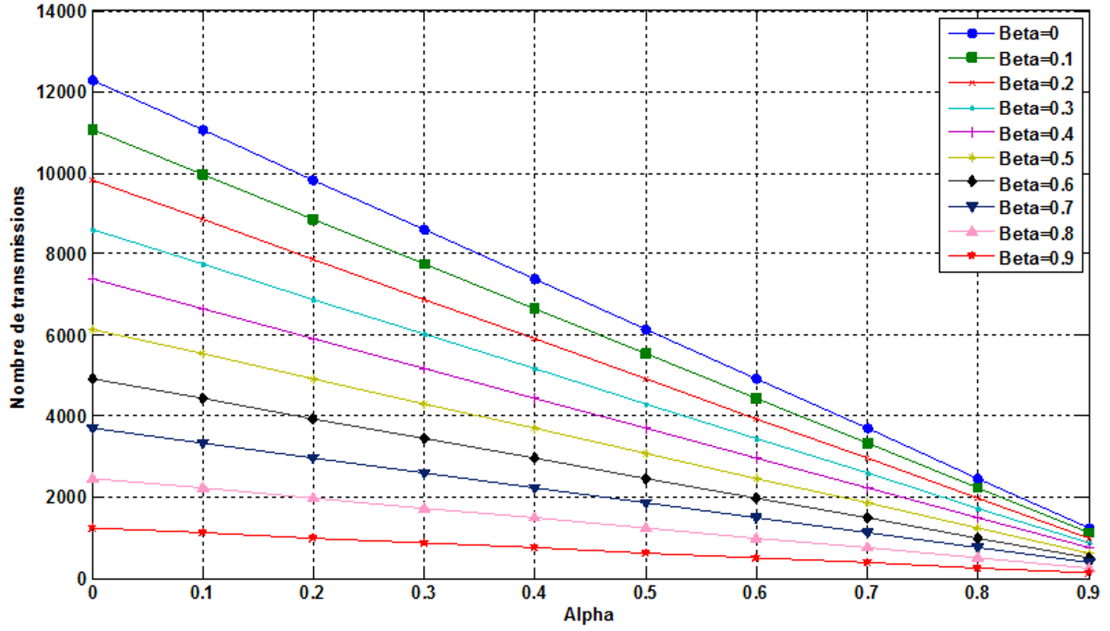


FIGURE 4.1 – Nombre de transmissions en utilisant le schéma EADPS dans un RCSF avec $D = 8$

$$\beta_{min}^c [\alpha_{min}^c (2D^2 - 1) fT] + E(X''_{DIS}) = \alpha_{min}^c (2D^2 - 1) fT + E(X'_{DIS})$$

$$\beta_{min}^c [\alpha_{min}^c (2D^2 - 1) fT] - \alpha_{min}^c (2D^2 - 1) fT = E(X'_{DIS}) - E(X''_{DIS})$$

$$\alpha_{min}^c (2D^2 - 1) fT (\beta_{min}^c - 1) = E(X'_{DIS}) - E(X''_{DIS})$$

Par conséquent, l'équation (4.6) conduit à une nouvelle précision moyenne minimale des prédictions α_{min} définie comme :

$$\beta_{min} = \frac{E(X''_{DIS}) - E(X'_{DIS})}{(1 - \alpha_{min})(2D^2 - 1) fT} \quad (4.6)$$

$$\text{où } \beta_{min} = 1 - \beta_{min}^c$$

On déduit donc le nombre de transmissions et de réceptions du nœud capteur i de l'anneau d comme :

$$NbT = (1 - \alpha_{min})(1 - \beta_{min})(2D^2 - 1) fT \quad (4.7)$$

où $\alpha_{min} \in [0, 1]$; $\beta_{min} \in [0, 1]$

La figure 4.1 montre l'évolution du nombre de transmissions pour un réseau de taille arbitrairement choisie égale à $D = 8$, en fonction de α_{min} et de β_{min} . On a également de façon arbitraire choisi $f = 2$ et $T = 24$. Sur cette figure, la courbe correspondante à $\beta = 0$ (courbe bleue) est celle qui traduit le schéma ADPS. Donc, il ressort clairement que notre algorithme réduit davantage le nombre total de transmissions et de réceptions comparativement à l'algorithme ADPS.

4.4 Comparaison entre l'agrégation et l'algorithme de prédiction EADPS

La figure 4.2 montre la comparaison des résultats trouvés quand on applique séparément (c-à-d sans combinaison) l'agrégation des données et l'algorithme EADPS. Sur cette figure, l'évolution du nombre de transmissions est tracée en fonction de α_{min} et β_{min} pour différentes tailles du réseau RCSF : $D = 2, 4, 8, 12$.

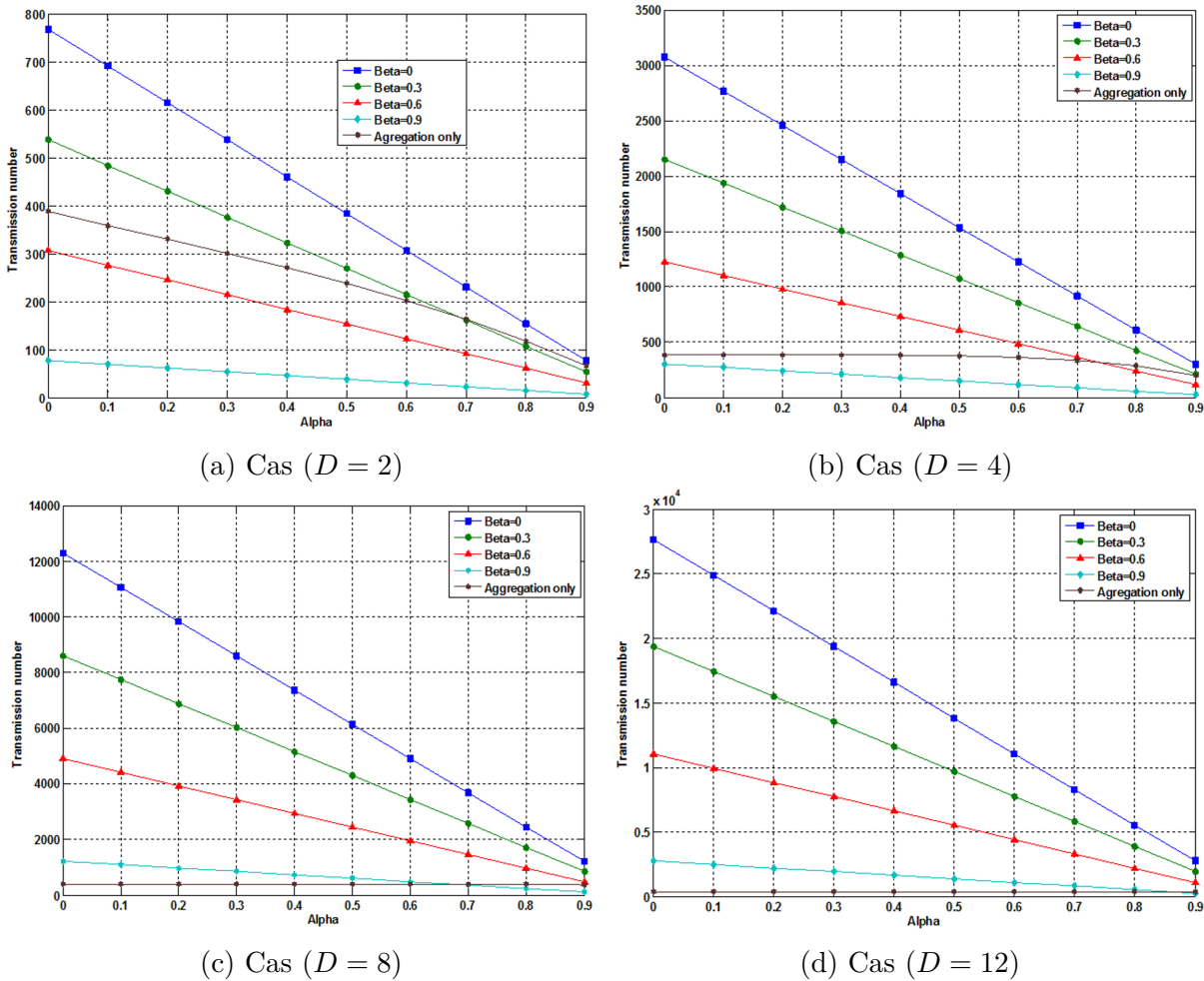


FIGURE 4.2 – Impact de la taille du réseau sur le nombre de transmissions dans le premier anneau : Comparaison Agrégation - Algorithme EADPS

L'analyse de la figure 4.2 fait ressortir que les gains en termes de réduction des transmissions obtenus par agrégation des données ne sont pas toujours meilleurs que ceux obtenus par prédiction à l'aide de l'algorithme EADPS. En effet, contrairement aux résultats trouvés dans le chapitre (3), on note ici que le processus d'agrégation de données ne devient puissant que pour des RCSFs de grandes tailles ($D > 8$). On n'en déduit que l'agrégation perd sa supériorité devant le schéma EADPS dans les RCSFs de petites tailles. Ce résultat est différent de celui présenté dans la figure 3.9 du chapitre (3) où l'agrégation était supérieure quelque soit la taille du réseau. Pour apprécier la puissance

de la combinaison de l'agrégation et du schéma EADPS, on va définir une référence qui correspond toujours au nombre de transmissions obtenu par la seule agrégation de données, auquel on attribue un taux égal à 100%. On évaluera ensuite les variations de ce taux après combinaison de l'agrégation avec le schéma EADPS. Pour approfondir cette analyse comparative, nous allons tenir compte également de l'impact des corrélations entre chaque pair de données en exploitant l'équation (3.41) du chapitre (3) et cela dans l'exécution du schéma d'agrégation. Nous considérons toujours des distributions de données avec une moyenne nulle et une variance unitaire. La fonction de répartition de la distribution multi-normale sera calculée à l'aide de l'intégrale multiple de l'équation (3.34) du chapitre (3). La figure 4.3 illustre les résultats de cette combinaison. Deux cas d'exemples sont présentés. Le premier considère une précision faible du prédicteur avec $\alpha_{min} = 0.3$ et le second considère un bon modèle avec $\alpha_{min} = 0.9$. Cette figure montre que la combinaison des techniques de prédiction et d'agrégation réduit davantage le taux de transmissions de données.

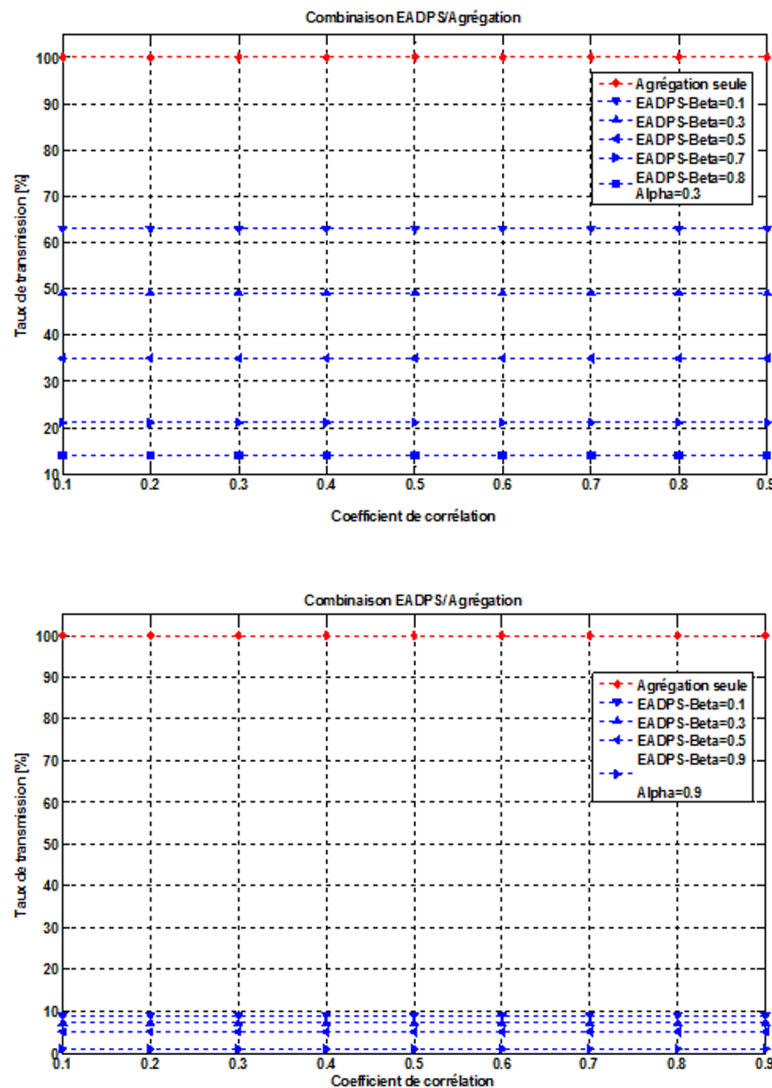


FIGURE 4.3 – Combinaison Agrégation – Algorithme EADPS

On observe l'évolution du nombre de transmissions en fonction du coefficient de corrélation ρ et de la précision β_{min} du modèle utilisé dans le schéma EADPS. On voit que la réduction du nombre de transmissions de données s'atténue surtout avec l'augmentation de β_{min} . En d'autres termes, l'impact du coefficient de corrélation sur le taux de transmission des données n'est pas significatif. Ce sont donc les précisions α_{min} et β_{min} des modèles qui impactent plus fortement le taux de transmission quelque soit ρ . Ce résultat ne corrobore pas celui trouvé antérieurement par G. M. Dias et al. [100].

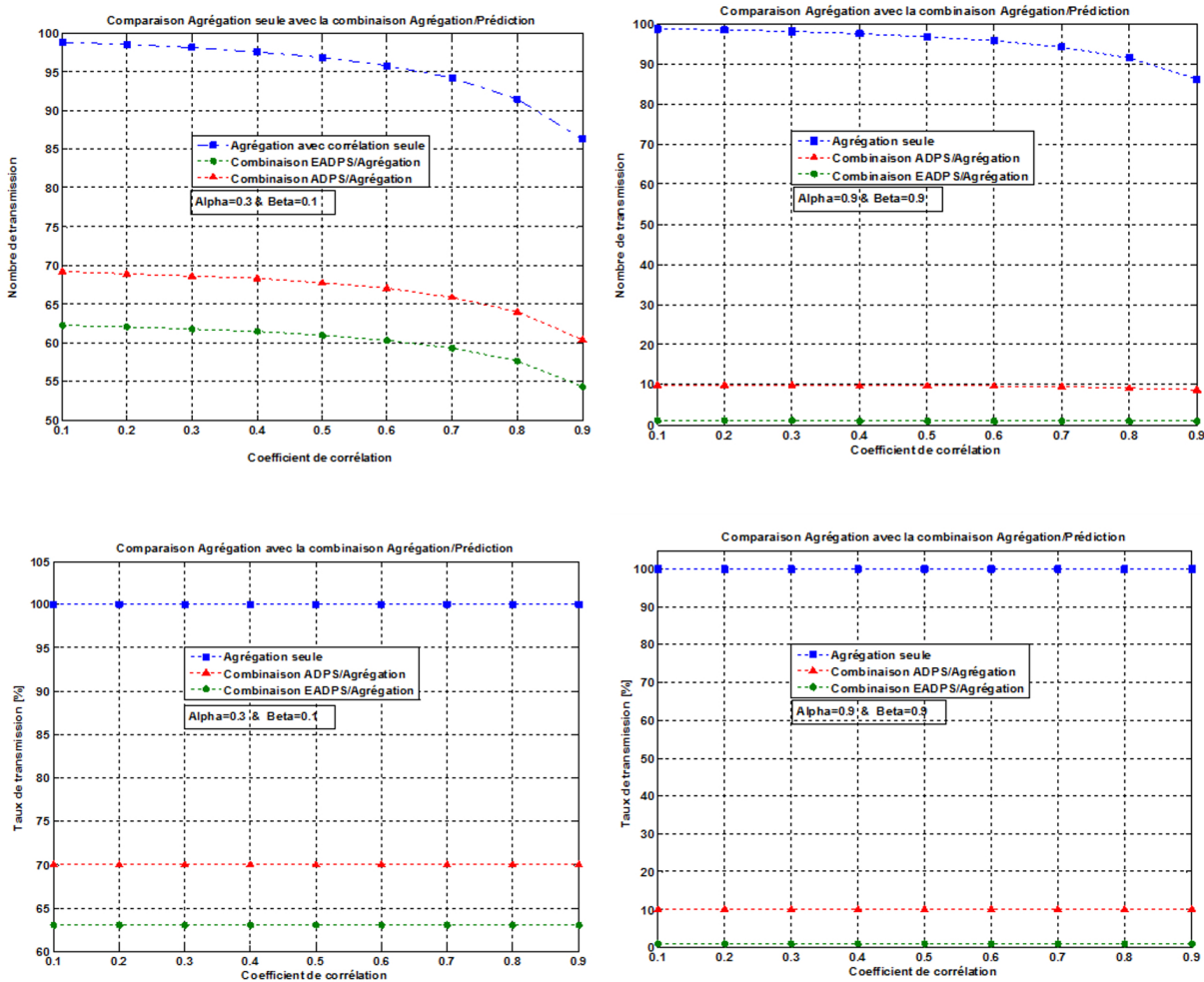


FIGURE 4.4 – Impact de la combinaison des techniques d'agrégation et de prédiction ADPS et EADPS sur le taux de transmission de données

La figure 4.4 compare la combinaison de l'agrégation de données avec les techniques de prédiction à l'aide des algorithmes ADPS et EADPS. Les résultats montrés sur cette figure correspondent aux cas suivants :

- $\alpha_{min} = 0.3$ et $\beta_{min} = 0.1$: Modèles médiocres
- $\alpha_{min} = 0.9$ et $\beta_{min} = 0.9$: Modèles de très bonne précision

On constate clairement que la combinaison agrégation-prédiction dans le cadre du paradigme EADPS est meilleure. Cela parce qu'il n'existe plus dans ce cas de fenêtre de

mise à jour du modèle. Cependant, dans les deux types de schémas ADPS et EADPS, la combinaison agrégation-prédiction de données améliore sensiblement la réduction du taux de transmission avec un gain variant entre 63% et 90% par rapport à l'application de l'agrégation seule. Ce gain impacte positivement l'économie d'énergie, le stockage des données et la longévité du RCSF.

Il y a lieu de souligner que les prédicteurs jouent un rôle fondamental. Ces prédicteurs ne peuvent en général être des modèles phénoménologiques construits sur la base de connaissances physiques car ils sont souvent lourds et compliqués et donc incompatibles avec les contraintes des nœuds capteurs. Par conséquent, ce sont les modèles prédictifs dits de l'automatique (modèles boîte-noire) qui sont exploités. La section suivante va traiter le problème spécifique de ces prédicteurs.

4.5 Problématique du modèle de prédiction des données

La sélection du modèle est une étape décisive en prédiction de données. Le choix entre des modèles concurrents doit se faire sur la base d'expérimentations ou d'avis d'experts. Il existe différents types de modèles. Par exemple, les modèles d'apprentissage automatique tels que les réseaux neurones artificiels (RNA) qui peuvent fournir une précision plus élevée que d'autres méthodes, mais ils peuvent ne pas respecter les contraintes des nœuds capteurs car la construction d'un RNA nécessite une phase intensive d'apprentissage sur une grande quantité de données. Cependant, des algorithmes légers existent et ont été testés dans différentes applications. Par exemple, dans les travaux de F. Aderohunmo et al., rapportés dans les références [90, 118] du chapitre (3), les auteurs ont expérimenté et comparé différents modèles appliqués à un système de surveillance de la température à l'intérieur d'un bâtiment et à différents niveaux d'étages. Ils ont expérimenté à la fois des simulations et des modèles adaptatifs légers mis en œuvre dans le monde réel pour obtenir une durée de vie prolongée du RCSF. Les modèles considérés sont des modèles autorégressifs et ceux à moyenne mobile exponentielle pondérée. Leur étude a montré que les modèles légers conviennent généralement à la surveillance de paramètres physiques comme la température, l'humidité, etc. . . , sans perdre beaucoup de précision par rapport aux modèles plus complexes. Des résultats similaires ont été rapportés dans les références [92, 93] en utilisant respectivement les modèles "Least Mean Square : Moindre Carré (LMS)" et "Auto-Regressive : Autorégressif (AR)" pour la prédiction de données dans les RCSFs. Le coût de la prédiction défini par l'équation (4.8) a été utilisé pour la comparaison.

$$PC = [\alpha f(e) + (1 - \alpha)r]E_c \quad (4.8)$$

où (e) est la mesure de la précision des prédictions (par exemple, l'erreur quadratique moyenne, le pourcentage d'erreur absolu moyen, l'erreur absolue moyenne, . . .). $f(e)$ est la précision selon la mesure choisie, (r) est le pourcentage de mesures transmises, α est le niveau de précision souhaité par l'utilisateur et E_c représente le coût de calcul de chaque algorithme sur le nœud capteur par rapport à l'empreinte mémoire, à l'utilisation du processeur et à la bande passante/trafic bus disponible. L'équation (4.8) peut être utile pour décider de l'adoption d'un schéma DPS ou non, mais elle est limitée à évaluer si

les prévisions permettront d'économiser de l'énergie dans un nœud capteur et ne prend pas en compte l'impact des agrégations occasionnelles, ni le travail des transmissions des nœuds fils.

Un autre exemple est celui de S. Samarah [119] qui présente un modèle linéaire de prédiction de données des séries chronologiques pour réduire le taux de transmission d'un réseau de capteurs d'une application agricole. La structure mathématique du modèle est très simple et elle est basée sur des valeurs limites (c'est-à-dire la valeur maximale et la valeur minimale) des portions temporelles du signal détecté par les nœuds capteurs. Les résultats atteints indiquent un taux de réduction de données variant entre 43% et 52%.

4.5.1 Choix et structure du modèle

Le choix et la structure du modèle de prédiction sont des questions clés. Comme nous considérons ci-après les séries chronologiques utilisées dans un environnement de RCSF, on peut alors citer trois principales approches pour élaborer un modèle de prédiction dédié à ces séries temporelles.

La première approche est basée sur la construction d'un modèle mathématique qui considère les données comme une fonction du temps $y = f(t)$. Cette ancienne catégorie de modèles est effectuée en ajustant les paramètres de la structure du modèle sélectionné à l'aide de la méthode des moindres carrés ou d'autres méthodes itératives. Une seconde catégorie de modèles cherche à déterminer chaque valeur de la série en fonction des valeurs qui la précède. C'est le cas des modèles autorégressifs "AutoRegressive Integrated Moving Average (ARIMA)". Cette catégorie de modèles a été popularisée et formalisée par Box et al. [120]. Cette catégorie de modèles cherche à déterminer chaque valeur de la série temporelle discrète comme une fonction linéaire des valeurs précédentes et d'un bruit blanc gaussien ε de variance σ^2 :

$$y(t) = f(y(t-1), y(t-2), \dots, y(t-k), \varepsilon(t), \varepsilon(t-1), \varepsilon(t-2), \dots) \quad (4.9)$$

Un modèle ARIMA est étiqueté comme modèle ARIMA (p,d,q), dans lequel :

- p est le nombre de termes autorégressifs (ordre des termes autorégressifs).
- d est le nombre de différences. En effet, l'estimation des modèles ARIMA suppose que nous travaillons sur une série stationnaire. La moyenne de la série est constante dans le temps, ainsi que la variance. La meilleure façon d'éliminer toute tendance de série est de différencier les valeurs adjacentes, c'est-à-dire de remplacer la série d'origine par la série des différences adjacentes. Une série temporelle qui a besoin d'être différenciée pour atteindre la stationnarité est considérée comme une version intégrée d'une série stationnaire (d'où le terme Integrated).
- q est le nombre de moyennes mobiles (ordre des termes du bruit $\varepsilon(t)$).

Selon les valeurs de p , q et d on retrouve les modèles autorégressifs (AR) ou les modèles à moyenne mobile (MA). Le modèle ARIMA constitue donc une mixtion des structures (AR) et (MA). Pour les structures MA, seuls les termes du bruit blanc $\varepsilon(t-i)$ sont pris en compte. Le choix du modèle AR ou MA est principalement théorique. Certains modèles ARIMA sont équivalents à d'autres méthodes, par exemple :

- ARIMA(0 ; 1 ; 0) est équivalent à une méthode simple qui suppose que la dernière observation va se répéter dans le futur,
- ARIMA(0 ; 2 ; 0) est équivalent à une méthode simple qui suppose une augmentation linéaire basée sur les deux dernières observations,
- ARIMA(0 ; 1 ; 1) est le plus simple modèle du lissage exponentiel (ES : Exponential Smoothing avec seulement un paramètre),
- ARIMA(0 ; 2 ; 2), ARIMA(0 ; 1 ; 2) et ARIMA(1 ; 1 ; 2) sont équivalents à des modèles ES plus complexes.

Par conséquent, lors du choix du modèle ARIMA qui correspond le mieux aux données, certaines des autres méthodes sont également implicitement prises en compte. Parfois, il peut être possible que l'opérateur mathématique f de l'équation (4.9) soit une combinaison non linéaire des arguments précédents, d'où la rencontre de modèles non linéaires.

La troisième approche utilise des algorithmes d'apprentissage automatique basés sur les RNA. Récemment, une étude très utile sur la réduction des données basée sur les prévisions dans les RCSFs a été rapportée dans la référence [96]. Les auteurs ont étudié et comparé les approches existantes dans le domaine de la prévision des données pour réduire les taux de transmission dans les RCSFs. Il apparaît qu'il n'y a pas a priori de modèle de prédiction préféré capable de satisfaire toutes les contraintes souvent antagonistes telles que le coût de calcul, la complexité et la précision. Sur la base d'une riche littérature spécialisée, ces auteurs ont établi une discussion sur la conception de la stratégie à adopter pour réduire les taux de transmission dans un RCSF comme défis futurs et présenté les directions de recherche ouvertes dans l'utilisation de méthodes de prédiction pour soutenir la croissance des RCSFs. Ils affirment que les recherches futures dans le domaine des RCSFs peuvent se concentrer sur les algorithmes de prédiction et la puissance de calcul pour améliorer leur précision, ce qui suggère l'intégration du Cloud computing comme solution prometteuse. De plus, ces auteurs jugent raisonnable que les outils offerts par la théorie de l'analyse statistique des données mesurées puissent aider à réaliser un mécanisme fiable qui gère les RCSFs à l'avenir. Parmi les avantages, il sera possible d'évaluer les améliorations potentielles de la congestion moyenne et de la durée de vie des RCSFs, ainsi que les limites inférieures de la qualité des informations produites.

Enfin, partant du cadre mathématique décrivant le modèle des RCSFs organisés en anneaux, présenté dans le chapitre (3), on sait qu'il faudra déterminer la condition à respecter pour réduire les transmissions de données lors de l'utilisation de l'algorithme ADPS. Cette condition concerne la précision minimale à atteindre par le modèle de prédiction de données pour que les transmissions dues à la diffusion des paramètres du modèle actualisé n'annulent pas les gains obtenus par la non-transmission des données bien prédites.

4.5.2 Robustesse du modèle

Les performances du modèle dépendent de sa robustesse lors de l'exécution. Comme décrit dans la référence [58] du chapitre (3), le modèle est garanti pour rester robuste si la variance du signal Y est positive et tend asymptotiquement vers zéro ; $\sigma^2 * Y > 0$ et $\sigma^2 * Y \rightarrow 0$ avec le temps t . Cependant, la distribution réelle d'un phénomène est généralement inconnue et nous ne pouvons fournir une estimation de $\sigma^2 * Y$ que si la loi

du grand nombre est vraie. Par conséquent, étant donné un modèle linéaire d'ordre P , alors

$$\sum_{i=1}^p |Paramètres| \leq 1 \text{ est vrai}$$

si et seulement si l'erreur de prédiction en n étapes (α) est inférieure à n fois l'erreur de prédiction en une étape. Cela signifie que l'erreur cumulative de la prédiction en n étapes n'est pas significative (autrement dit elle ne diverge pas). De toute évidence, tant que la condition illustrée par l'équation (4.10) se respecte, le modèle est garanti pour rester robuste. Cette condition est bien démontrée dans la référence [104] du chapitre (3).

$$Var[e(n)] \leq Var(UB) \tag{4.10}$$

où $Var[e(n)]$ est la variance de l'erreur de prédiction en n étapes et $Var(UB)$ est la limite supérieure de la variance du signal Y . Donc, le modèle sélectionné doit être non seulement précis mais également robuste.

4.5.3 Sélection du modèle

Dans notre application, nous utilisons les données réelles collectées par la station météorologique située à l'aéroport international Messali Hadj de Tlemcen. Les données sélectionnées sont la température locale de l'air, la pression atmosphérique et l'humidité relative. Ces paramètres sont détectés à deux mètres au-dessus du sol. L'intervalle de temps entre deux mesures consécutives est imposé à 30 minutes conformément aux normes de l'aéronautique. La figure 4.5 montre un exemple de données météorologiques enregistrées sur différentes journées du mois de Janvier 2017.

Quotidiennement, la valeur initiale commence à 0h30 min et le dernier relevé de la journée est à minuit. Par conséquent, 48 valeurs sont collectées chaque jour pour chacun des paramètres considérés (température, pression atmosphérique et humidité relative de l'air).

Partant des données recueillies durant les mois de Janvier 2017 et Juin 2017, et après une étude préliminaire concernant la stationnarité de la moyenne et la constance de la variance des signaux enregistrés, nous sélectionnons après une étude bibliographique poussée, le type de modèle autorégressif (ARMA) pour la prédiction de nos données météorologiques. Ce modèle est très efficace pour les prévisions à court terme pour deux raisons :

1. il n'est pas trop sensible au bruit des données,
2. ses prédictions suivent les tendances observées les plus récentes, ce qui représente un potentiel pour garder une grande précision même si les données présentent une variance élevée.

La structure du modèle dépend du type de série chronologique générée par le nœud capteur. Le modèle est appliqué pour représenter la relation entre les données actuelles et passées, fournies par le nœud capteur. De plus, un ensemble de critères mesurables est utilisé pour quantifier les performances dues à l'utilisation de la prédiction des données.

Dans le cas d'une série chronologique, le modèle (ARMA) est décrit par l'équation (4.11) :

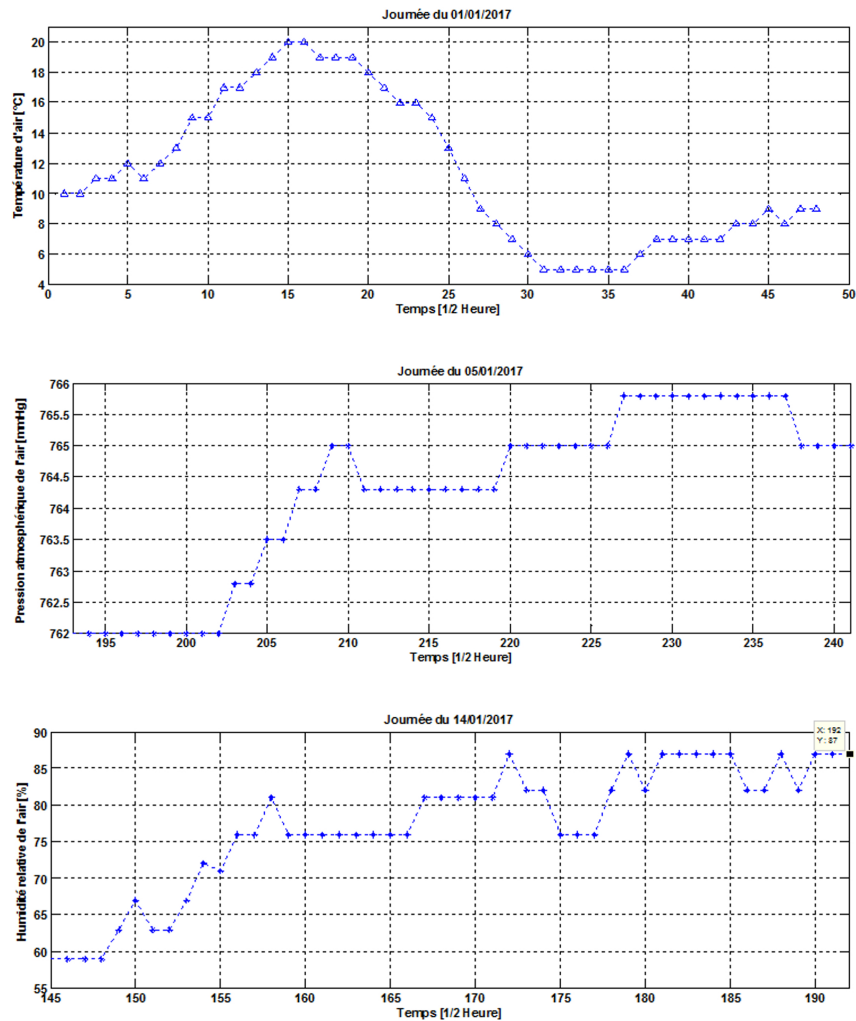


FIGURE 4.5 – Exemples de données météorologiques enregistrées durant Janvier 2017

$$A(q).y(t) = C(q).\varepsilon(t) \quad (4.11)$$

où $y(t)$ représente la valeur estimée par le modèle à l'instant ' t ' et $\varepsilon(t)$ représente la valeur de l'erreur au même instant ' t ' (écart ou perturbation considérée comme un bruit). Les équations suivantes représentent les polynômes de retard.

$$A(q) = 1 + a_1q^{-1} + \dots + a_{na}q^{-na}$$

$$C(q) = 1 + c_1q^{-1} + \dots + c_{nc}q^{-nc}$$

Les paramètres na et nc de ces polynômes représentent les ordres du modèle ARMA. Les paramètres a_i et c_i sont respectivement les coefficients d'auto-régression et du bruit. L'opérateur q^{-1} traduit le retard d'un pas (ordre 1). On note qu'un processus autorégressif ne sera stable que si les coefficients se trouvent dans un certain intervalle. Par exemple, s'il n'y a qu'un seul paramètre autorégressif, il doit être dans l'intervalle $-1 < \theta_1 < +1$ (cercle de stabilité). Dans les autres cas, les erreurs passées s'accumuleraient et les valeurs

successives des estimés $y(t)$ divergent lorsqu'on se déplace infiniment vers l'avant, ce qui signifie que la série ne serait pas stationnaire. S'il y a plus d'un paramètre autorégressif, des restrictions similaires sur les valeurs des paramètres peuvent être définies comme rapporté dans la référence [120]. L'ordre $P[na\ nc]$ du modèle est une question clé qui affecte sa performance. Le contrôle de la qualité du modèle utilise le critère de validation (VC) défini par l'équation (4.1) et/ou le critère d'Akaike (Akk) défini par l'équation (4.12). Dans les schémas adaptatifs de prédiction duale, lorsque le seuil (e_{max}) défini par l'utilisateur est violé, les coefficients du modèle sont à nouveau identifiés soit à l'aide de la fenêtre d'estimation de données dans le cas de l'algorithme ADPS, soit à l'aide de la base de données des modèles obsolètes stockés dans le cas du schéma EADPS.

$$Akk = \log(V) + \frac{2 * d}{N} \quad (4.12)$$

où V est la fonction perte, d est le nombre de paramètres estimés et N est la taille (nombre de valeurs) de l'ensemble des données d'estimation. La fonction perte est définie comme :

$$V = \det\left[\frac{1}{N} \sum_1^N \varepsilon(t, \theta_N)(\varepsilon(t, \theta_N))^T\right] \quad (4.13)$$

où θ_N représente les paramètres estimés. Pour $d \ll N$, le critère Akk s'écrira :

$$Akk = \log\left[V\left(1 + \frac{2 * d}{N}\right)\right] \quad (4.14)$$

Pour obtenir la structure du modèle nous avons étudié expérimentalement l'évolution du critère de validation (VC), défini par l'équation (4.1) en fonction de l'ordre du modèle et de la taille de la fenêtre d'estimation (c-à-d la taille de l'échantillon de données, consacré à l'estimation des paramètres du modèle). Les données météorologiques recueillies en Janvier 2017 ont servi pour cette étude. L'environnement du code de calcul Matlab a été exploité pour réaliser cet objectif. Nous avons considéré le même ensemble de données de validation pour tous les tests. La figure 4.6 illustre les résultats obtenus concernant l'ordre na . Quant à l'ordre nc , il doit être égal à na ou $(na - 1)$ selon les recommandations fournies par la documentation d'aide du code de calcul Matlab. Les meilleurs modèles correspondent aux valeurs élevées du critère (VC). La sélection du modèle est un compromis entre la précision atteinte et le temps de calcul correspondant. En effet, plus le modèle est précis, plus le coût de calcul est grand ce qui risque d'être incompatible avec les capacités des nœuds capteurs.

En résumé, le modèle ARMA(2,1) a été sélectionné. Son coût CPU n'excède pas la valeur 0.03 pour un processeur i5. Un indicateur simple et commun pour apprécier la qualité du modèle sélectionné consiste à le confronter aux observations expérimentales de la série. Les figures 4.7 et 4.8 montrent un exemple des résultats atteints. Une fois le modèle élaboré, il pourra servir comme prédicteur compte tenu des tolérances admises ou imposées par l'utilisateur sur les données considérées. Celles-ci présentent sur une base mensuelle de mesures, les écarts types suivants :

- Ecart type de la température : Std=4.61°C.
- Ecart type de la pression atmosphérique : Std= 3.51 mmHg.

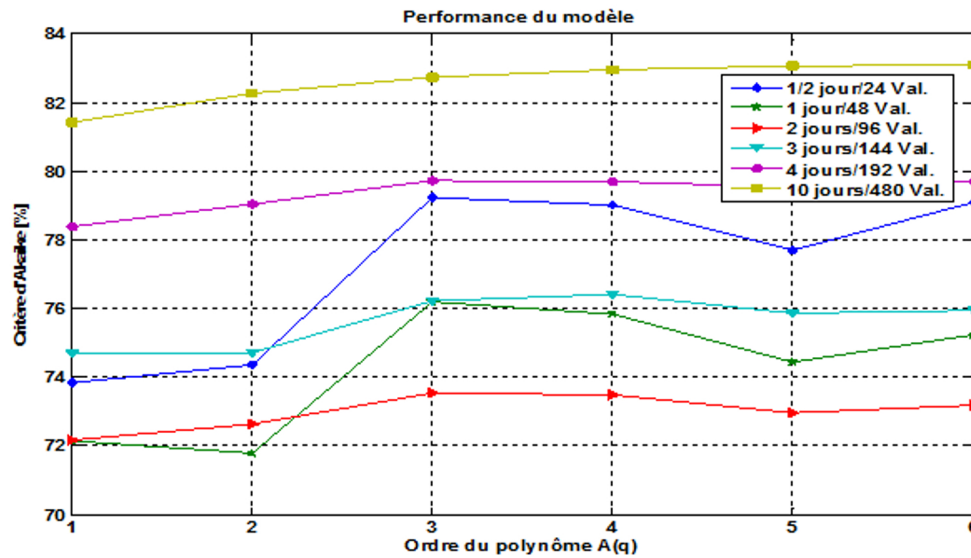


FIGURE 4.6 – Effet de la taille de la fenêtre d’estimation et de l’ordre na du modèle sur ses performances à l’aide du critère (Akk) : Cas de la température d’air

- Ecart type de l’humidité relative de l’air : Std= 18.15%.

Partant de ces résultats, nous avons choisi quatre niveaux de tolérance pour chaque grandeur météorologique détectée. Les tolérances choisies sont bien inférieures aux écarts types mentionnés ci-dessus de sorte que les précisions relatives, c’est-à-dire le rapport entre la tolérance choisie à la valeur moyenne de la grandeur collectée ne dépasse guère 0.5%. Le tableau 4.1 récapitule les valeurs des tolérances adoptées :

TABLE 4.1 – Récapitulatif des tolérances admises

Grandeur météorologique	Tol 1	Tol 2	Tol 3	Tol 4
Température	0.5 °C	1 °C	1.5 °C	2° C
Pression atmosphérique	0.5 mmHg	1 mmHg	1.5 mmHg	2 mmHg
Humidité relative	3 %	6 %	9 %	12 %

Ainsi, nous pouvons maintenant étudier la qualité de prédiction de nos modèles à l’échelle du nœud capteur. La figure 4.10 montre les résultats de prédiction exprimés en termes de taux de transmission, en fonction de la durée de prédiction et du niveau de tolérance imposée.

Quatre observations majeures ressortent de l’analyse de ces graphiques. La première est que les niveaux des transmissions diminuent lorsque le niveau de la tolérance augmente, ce qui est logique. La seconde est que le prédicteur de l’humidité relative est le plus précis, c’est-à-dire que les données numériques sont les mieux adaptées à la structure mathématique du modèle sélectionné. La troisième observation indique la robustesse des modèles puisque les taux de transmissions ne divergent pas sur toute la longueur des

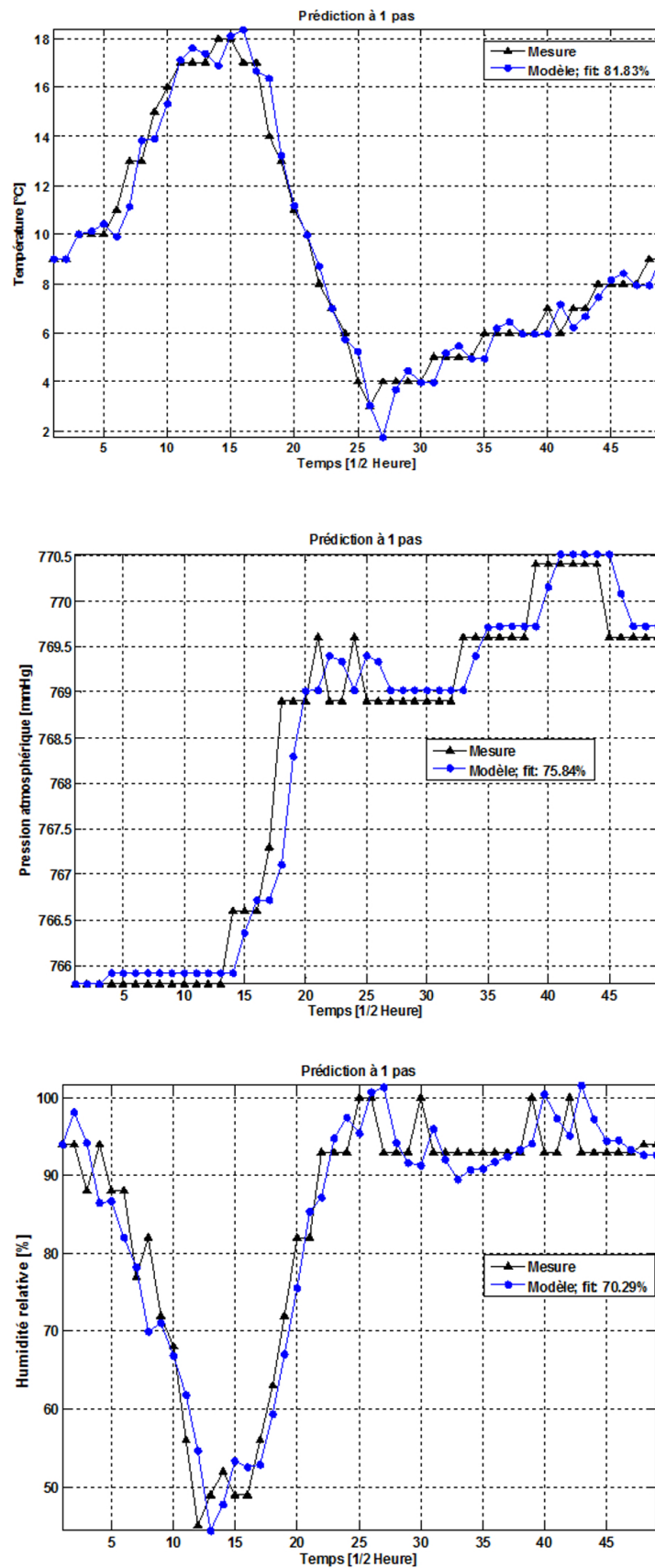


FIGURE 4.7 – Exemple de résultat de prédiction des données météorologiques : Journée du 02/01/2017

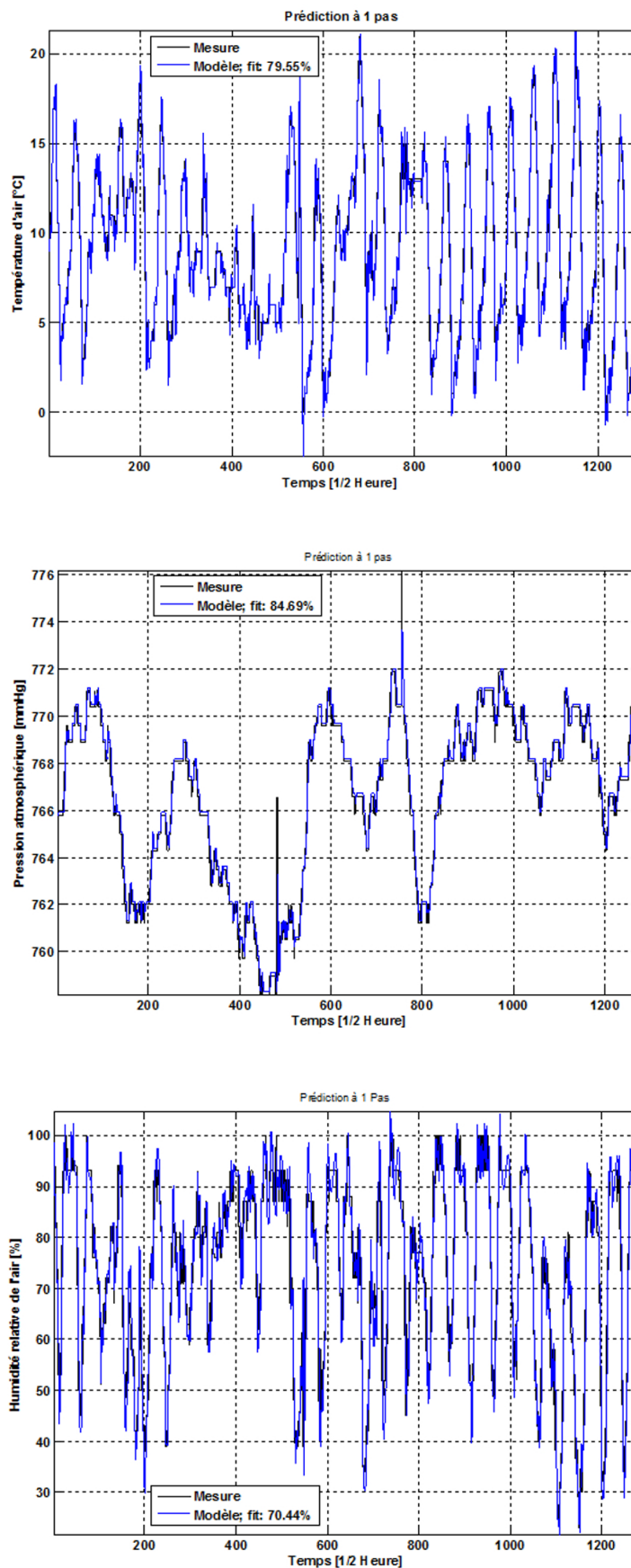


FIGURE 4.8 – Résultats de prédiction des données météorologiques du mois de Janvier 2017 : Superposition mesure-modèle avec indication du critère '*fit*'

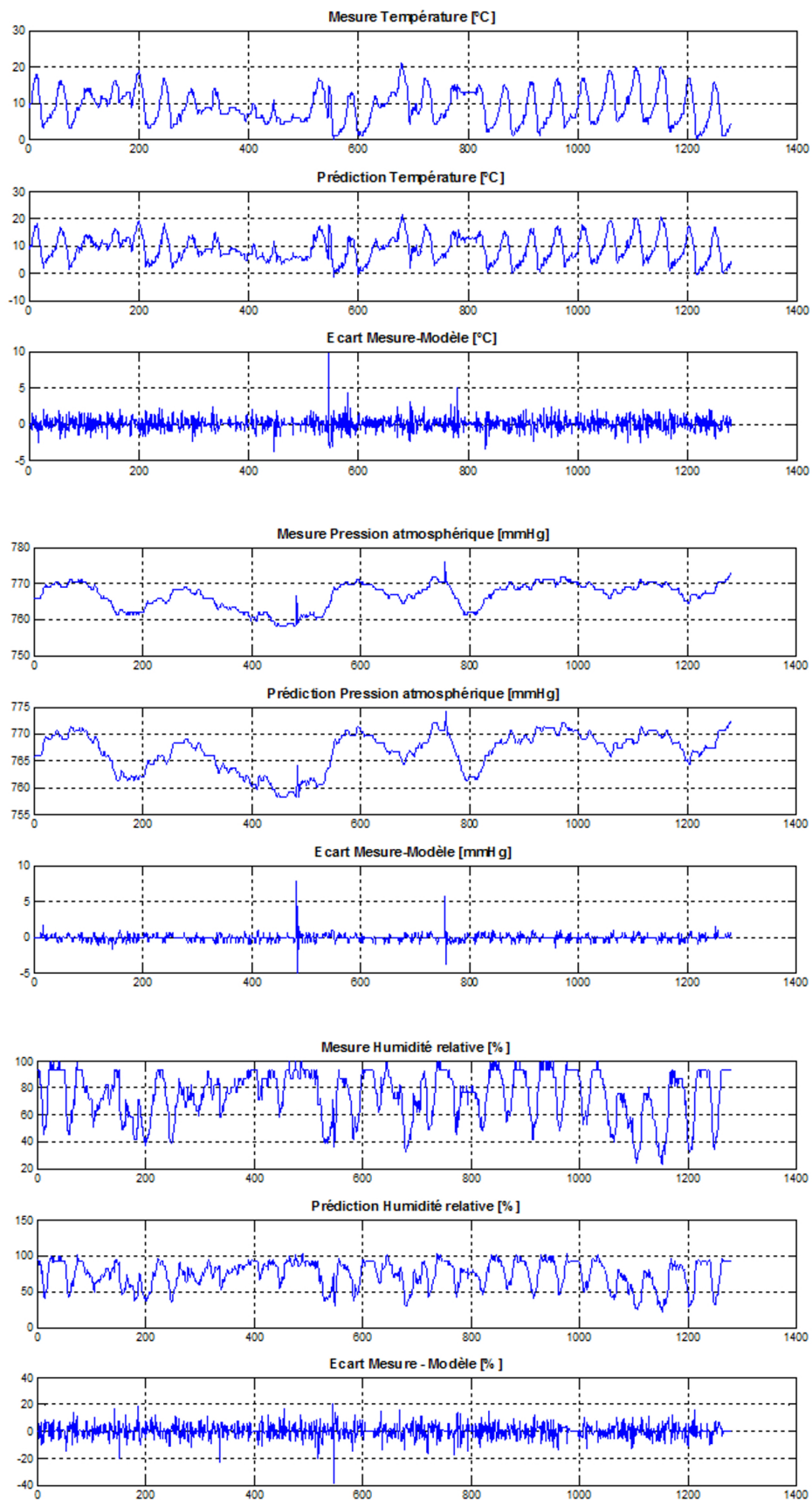


FIGURE 4.9 – Résultats de prédiction des données météorologiques du mois de Janvier 2017 : Séparation mesure-modèle avec indication d'écart

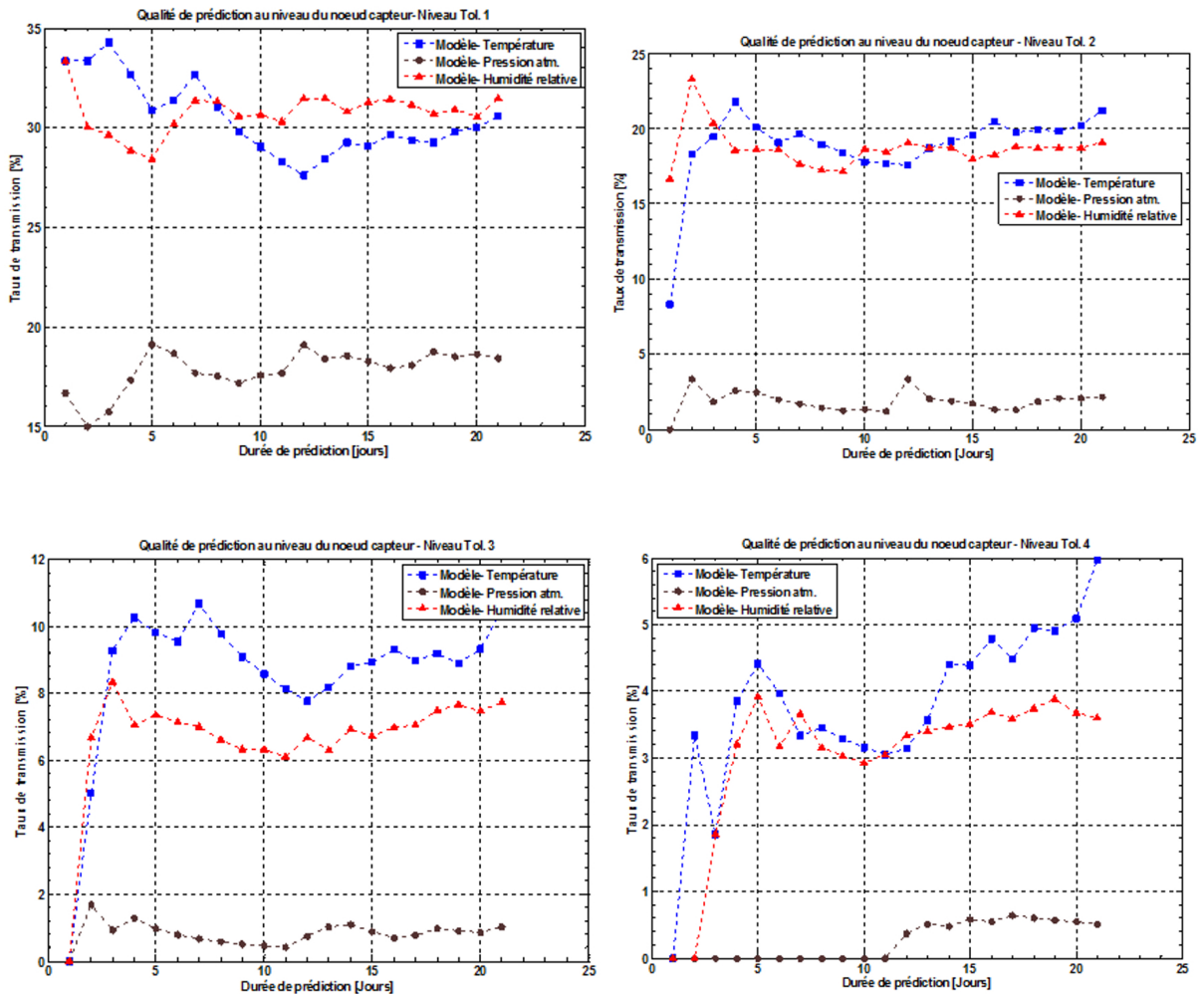


FIGURE 4.10 – Evolution des taux de transmission des données météorologiques prédites au niveau des capteurs en fonction de la tolérance imposée et la durée de prédiction

durées de transmissions. Enfin, la quatrième remarque traduit la précision des modèles puisque le taux maximal atteint ne dépasse guère 35% et que des taux de transmission nuls sont obtenus sur les petites durées de prédictions lorsque les tolérances deviennent larges. On peut donc attester de la bonne qualité de ces résultats.

4.5.4 Cas du prédicteur EADPS

On peut intuitivement admettre que le modèle issu du paradigme EADPS est susceptible de dériver avant le modèle qui est élaboré à l'aide des données physiques prédites ou mesurées. C'est pourquoi, un traitement particulier des modèles implémentés dans le schéma EADPS devient nécessaire. En effet, contrairement au procédé d'estimation des paramètres des modèles ADPS, basé sur l'approche itérative de minimisation de la prédiction d'erreur, l'estimation des paramètres du modèle EADPS s'est appuyée sur l'algorithme du maximum de vraisemblance. La figure 4.11 montre l'exemple d'évolution des paramètres du modèle ARMA retenu c'est-à-dire l'évolution dans le temps des coefficients

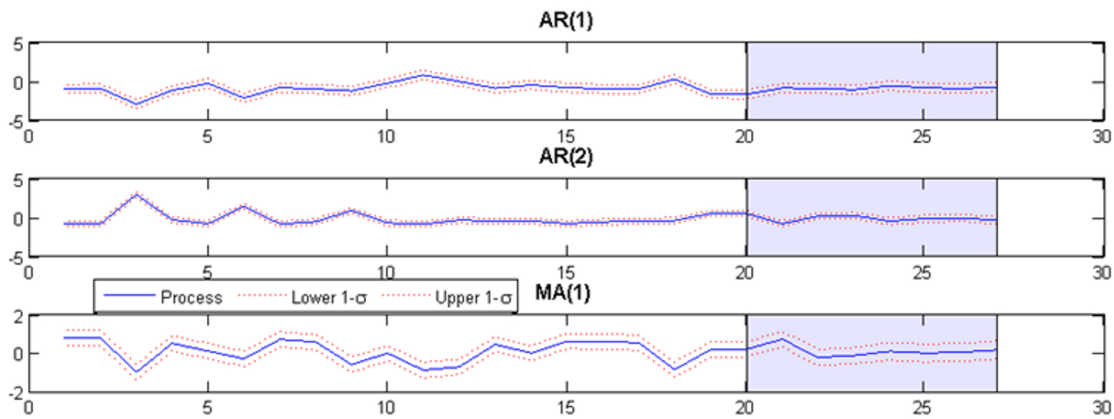


FIGURE 4.11 – Evolution dans le temps des coefficients na et nc du modèle $ARMA(2,1)$ sélectionné pour l'humidité relative de l'air : Algorithme EADPS

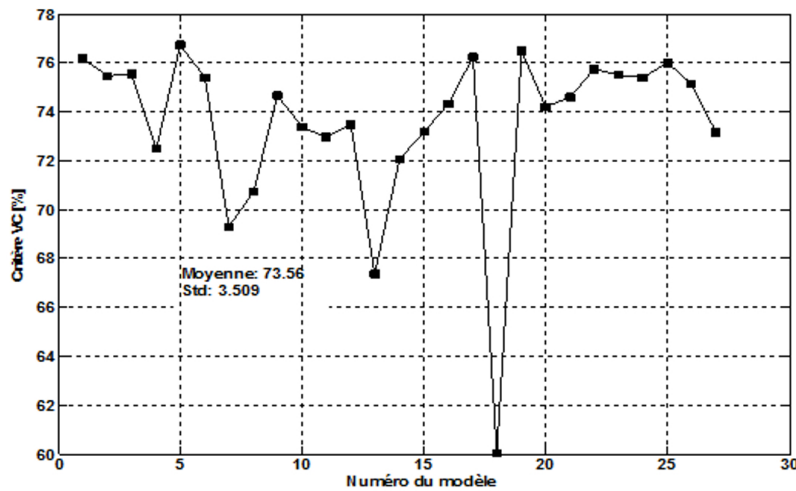


FIGURE 4.12 – Évolution du critère (VC) du modèle de température $ARMA(2,1)$: Schéma EADPS

na ($AR(1)$ et $AR(2)$) et du coefficient nc ($MA(1)$). Cette même figure montre également les prédictions futures de ces paramètres et leur intervalle de confiance.

La figure 4.12 montre l'exemple d'évolution du critère (VC) pour le modèle de prédiction de la température durant le processus EADPS où l'on voit que le modèle n° 18 était le moins performant alors que la figure 4.13 montre l'évolution du critère (Akk) pour le modèle relatif à l'humidité relative de l'air durant la même application du schéma EADPS. On remarque cette fois-ci que le meilleur modèle correspond au n°13. Ainsi, l'actualisation des modèles ne s'effectue pas toujours au même moment. En fait, lorsqu'une tolérance est violée pour une variable météorologique donnée, elle peut ne pas l'être pour une autre variable météorologique. La qualité du modèle EADPS en est responsable.

La figure 4.14 montre à la fois une prédiction des valeurs moyennes fournies par les modèles EADPS correspondants aux grandeurs météorologiques considérées ainsi que leur intervalle de confiance obtenu à partir du calcul d'erreur de la covariance. Cette prédiction est obtenue sur une durée de 12 heures qui correspond donc à 24 valeurs prédites.

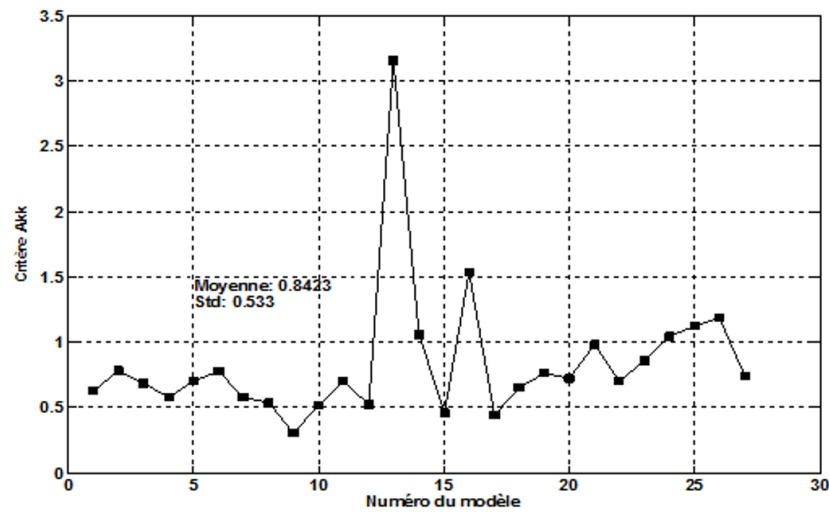


FIGURE 4.13 – Évolution du critère (Akk) du modèle ARMA(2,1) de l'humidité relative d'air : Schéma EADPS

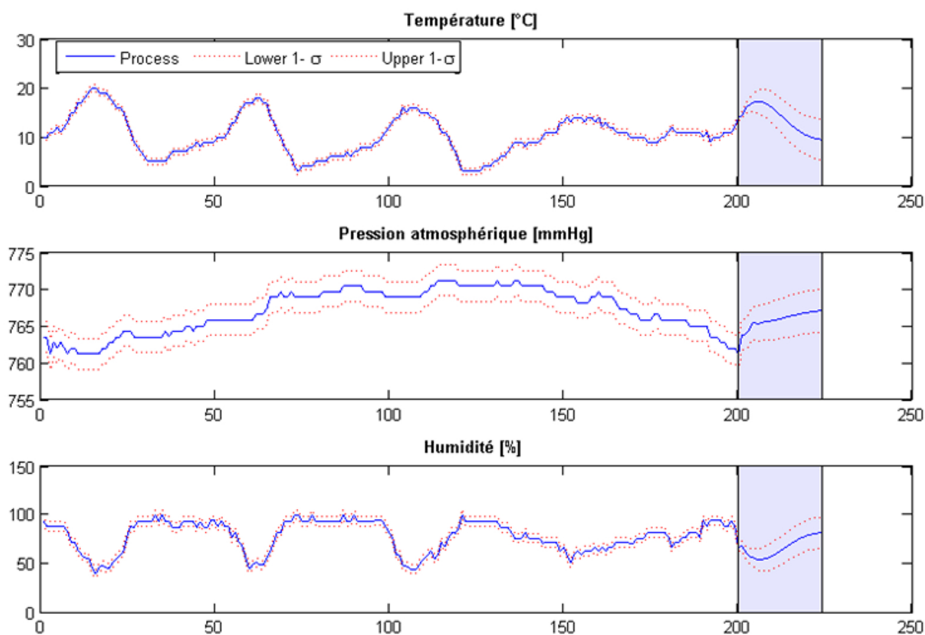


FIGURE 4.14 – Prévisions des valeurs météorologiques sur une période de 12 heures à l'aide des modèles élaborés dans le schéma EADPS

La figure 4.15 montre l'écart absolu entre les prédictions faites par ces modèles et les valeurs réelles des grandeurs météorologiques mesurées. Pour apprécier la qualité de ces modèles indépendamment du paradigme de communication, nous récapitulons sur le tableau 4.2 les taux de réduction des données (pourcentage de données non transmises) en fonction de la précision imposée par l'utilisateur. On voit que la réduction du taux de transmission des données peut dépasser 90% selon le seuil de tolérance imposé par l'utilisateur. Il est clair que le gain est toujours positif c'est-à-dire que β_{min} , équation (4.6), est strictement positif.

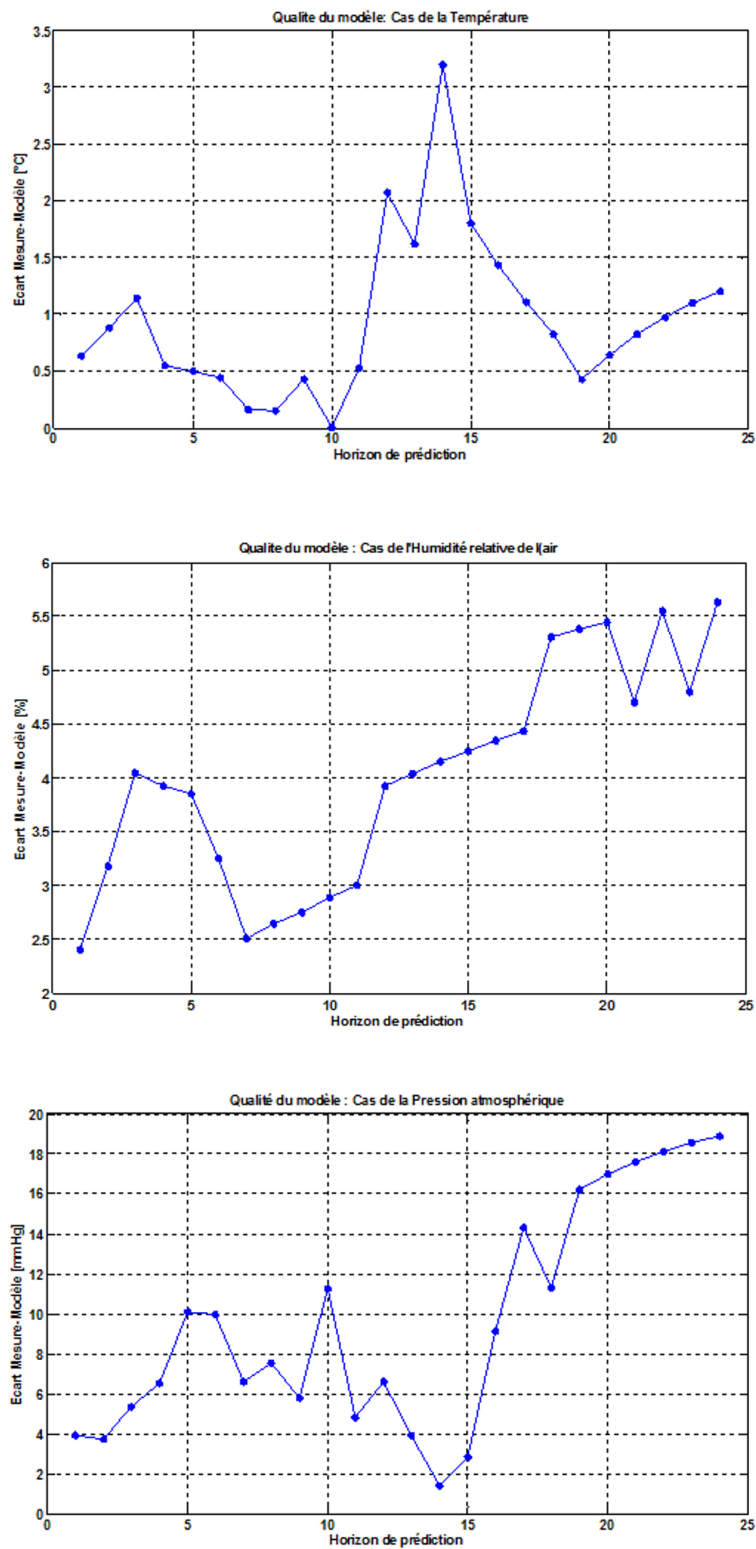


FIGURE 4.15 – Evolution des écarts absolus entre les prévisions des modèles EADPS et les valeurs mesurées sur une période de 12 heures

TABLE 4.2 – Taux des données non transmises en fonction du seuil de tolérance admis par l'utilisateur (Qualité du modèle EADPS)

Variable Météorologique	Température		Pression atmosphérique		Humidité relative	
	Tolérance \leq à	Taux de réduction de données [%]	Tolérance \leq à	Taux de réduction de données [%]	Tolérance \leq à	Taux de réduction de données [%]
Résultats obtenus sur une durée de 12 heures	0.5	29.16	6	36.36	4	45.83
	1	68.18	12	70.83	4.5	70.83
	1.5	83.33	16	75	5	79.16
	2	91.66	18	87.5	5.5	91.66

TABLE 4.3 – Résultats de prédiction de la température de l'air en utilisant "ADPS" et "EADPS"

Période Janvier 2017 Std=4.619	Taux de transmissions des données de la température : Durée (7 jours)							
	1 jour (48 mesures)				2 jours (96 mesures)			
Taille de la fenêtre d'estimation								
Tolérance imposée [°C]	0.5	1	1.5	2	0.5	1	1.5	2
Schéma ADPS [%]	32.77	16.12	13.80	12.93	35.27	26.65	23.72	22.72
Schéma EADPS [%]	21.09	12.68	6.12	1.15	21.09	12.68	6.12	1.15
Différence ADPS-EADPS	11.68	3.44	7.68	11.78	14.18	13.97	17.6	21.57

Comparons maintenant les résultats obtenus par les prédicteurs EADPS à ceux obtenus à partir du paradigme ADPS. Le tableau 4.3 montre les résultats atteints pour la prédiction de la température de l'air. Ce tableau compare le pourcentage des données transmises à l'aide de notre algorithme EADPS à celui dû à l'algorithme ADPS et cela pour différentes tolérances imposées par l'utilisateur. Aussi, nous avons fait varier la taille de la fenêtre d'estimation pour analyser son impact sur les résultats obtenus. L'analyse du tableau 4.3 montre clairement la supériorité du schéma EADPS. On note que les taux de transmissions diminuent en fonction du niveau de la tolérance admise pour les deux schémas de prédiction duale ADPS et EADPS. On note également que pour une fenêtre d'estimation de 2 jours (soit 48 valeurs collectées), les taux de transmission de données restent supérieurs, ce qui est tout à fait logique. Des résultats similaires sont obtenus pour les autres grandeurs météorologiques comme montré sur les tableaux 4.4 et 4.5. On remarque à partir de l'ensemble de ces résultats que l'algorithme de prédiction duale EADPS est capable de réduire les taux de transmissions de façon substantielle en comparaison avec l'algorithme ADPS. En effet, cette économie dépasse les 25% dans certains cas. L'algorithme EADPS a pu réaliser dans certains cas des taux de transmissions entre 0% et 2%.

La figure 4.16 montre les résultats de prédiction des données météorologiques à l'aide de l'algorithme EADPS pour les deux premières journées du mois de Juin 2017. On peut

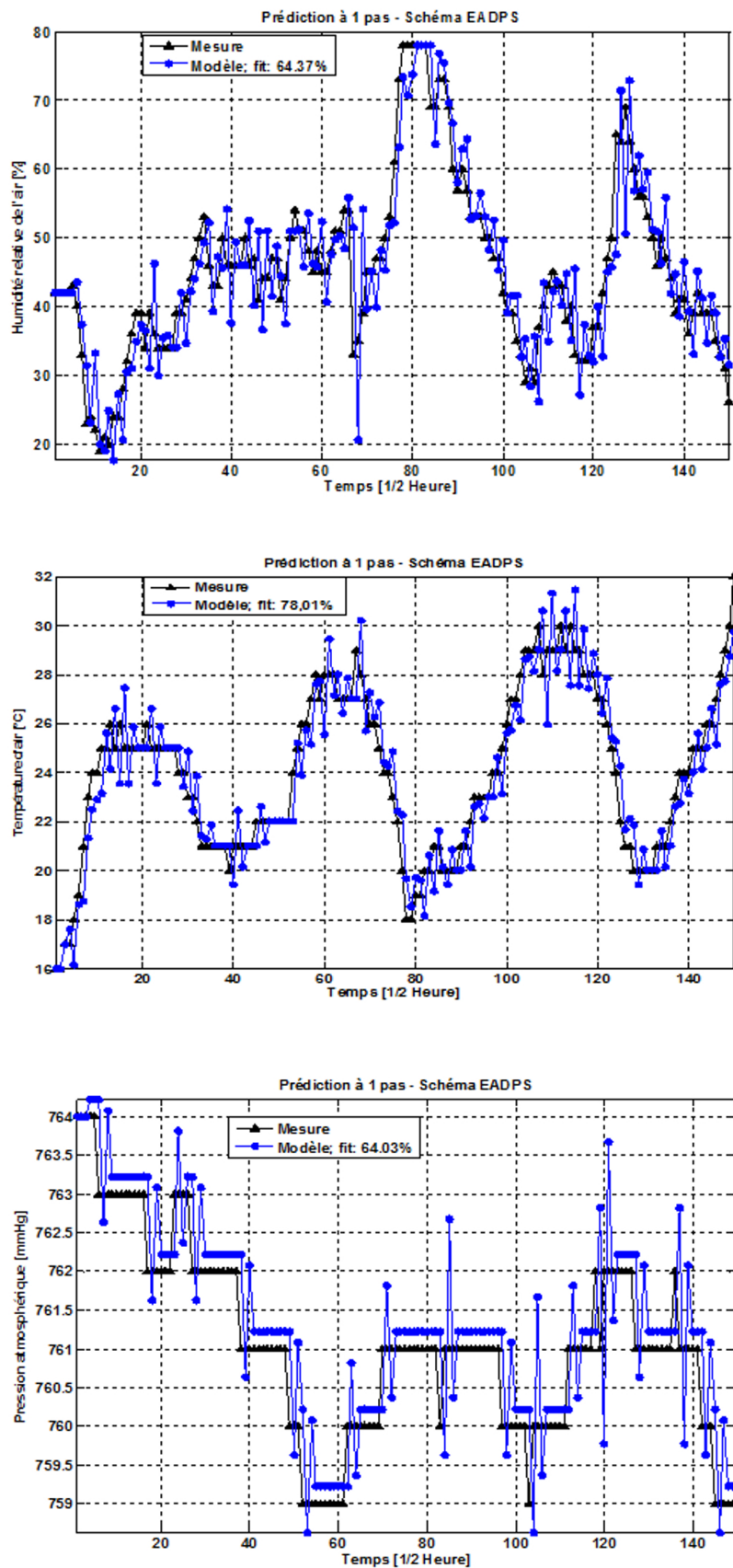


FIGURE 4.16 – Prédiction des données météorologiques des deux premières journées du mois de Juin 2017

4.5. Problématique du modèle de prédiction des données

TABLE 4.4 – Résultats de prédiction de la pression atmosphérique en utilisant "ADPS" et "EADPS"

Période Janvier 2017		Taux de transmissions des données de la pression : Durée (7 jours)							
Std=3.5139		1 jour (48 mesures)				2 jours (96 mesures)			
Taille de la fenêtre d'estimation									
Tolérance imposée [mmHg]	0.5	1	1.5	2	0.5	1	1.5	2	
Schéma ADPS [%]	15.76	12.83	12.79	12.50	28.08	23.37	22.22	22.22	
Schéma EADPS [%]	14.67	1.19	1.08	0	14.67	1.19	1.08	0	
Différence ADPS-EADPS	1.09	11.64	11.71	12.50	13.41	22.18	20.5	22.22	

TABLE 4.5 – Résultats de prédiction de l'humidité en utilisant "ADPS" et "EADPS"

Période Janvier 2017		Taux de transmissions des données de l'humidité : Durée (7 jours)							
Std=18.157		1 jour (48 mesures)				2 jours (96 mesures)			
Taille de la fenêtre d'estimation									
Tolérance imposée [%]	3	6	9	12	3	6	9	12	
Schéma ADPS [%]	33.92	17.95	15.51	14.85	48.50	35.53	30.70	29.39	
Schéma EADPS [%]	22.88	11.59	5.30	2.55	22.88	11.59	5.30	2.55	
Différence ADPS-EADPS	11.04	6.36	10.21	12.30	25.62	23.94	25.40	26.84	

TABLE 4.6 – Taux de transmission de données : Comparaison entre les simulations et le modèle : Cas de la température d'air

	$\alpha_{min} = 0$			$\alpha_{min} = 0.3$			$\alpha_{min} = 0.6$			$\alpha_{min} = 0.9$		
	Modèle	Simulation		Modèle	Simulation		Modèle	Simulation		Modèle	Simulation	
		Ex.DPS	Or.DPS		Ex.DPS	Or.DPS		Ex.DPS	Or.DPS		Ex.DPS	Or.DPS
$\beta_{min} = 0$	100%	100%	100%	70%	71%	73%	40%	40%	43%	10%	12%	14%
$\beta_{min} = 0.3$	100%	100%	100%	49%	49%	51%	28%	30%	33%	7%	7%	10%
$\beta_{min} = 0.6$	100%	100%	100%	28%	29%	31%	16%	16%	20%	4%	5%	7%
$\beta_{min} = 0.9$	100%	100%	100%	7%	7%	9%	4%	4%	7%	1%	1%	4%

remarquer le bon accord entre les mesures et la prédiction. Ces résultats indiquent encore une fois le réel potentiel de l'algorithme EADPS à réduire davantage les données transmises en raison de l'élimination de la fenêtre d'estimation précédemment utilisée pour mettre à jour les modèles obsolètes dans le schéma ADPS. Par conséquent, une meilleure économie d'énergie de la batterie du capteur est obtenue. Cependant, ce résultat reste tributaire du niveau minimum de précision que le modèle de prédiction doit assurer afin de limiter la diffusion des paramètres du modèle adaptatif. Cela reste tributaire du seuil de tolérance imposé par l'utilisateur et de l'approche utilisée pour estimer les paramètres du modèle.

TABLE 4.7 – Taux de transmission de données : Comparaison entre les simulations et le modèle : Cas de la pression atmosphérique

	$\alpha_{min} = 0$		$\alpha_{min} = 0.3$		$\alpha_{min} = 0.6$		$\alpha_{min} = 0.9$					
	Modèle	Simulation	Modèle	Simulation	Modèle	Simulation	Modèle	Simulation				
	Ex.DPS	Or.DPS	Ex.DPS	Or.DPS	Ex.DPS	Or.DPS	Ex.DPS	Or.DPS				
$\beta_{min} = 0$	100%	100%	100%	70%	71%	74%	40%	40%	42%	10%	12%	13%
$\beta_{min} = 0.3$	100%	100%	100%	49%	49%	52%	28%	30%	32%	7%	7%	9%
$\beta_{min} = 0.6$	100%	100%	100%	28%	29%	31%	16%	16%	20%	4%	5%	7%
$\beta_{min} = 0.9$	100%	100%	100%	7%	7%	9%	4%	4%	7%	1%	1%	3%

TABLE 4.8 – Taux de transmission de données : Comparaison entre les simulations et le modèle : Cas de l'humidité d'air

	$\alpha_{min} = 0$		$\alpha_{min} = 0.3$		$\alpha_{min} = 0.6$		$\alpha_{min} = 0.9$					
	Modèle	Simulation	Modèle	Simulation	Modèle	Simulation	Modèle	Simulation				
	Ex.DPS	Or.DPS	Ex.DPS	Or.DPS	Ex.DPS	Or.DPS	Ex.DPS	Or.DPS				
$\beta_{min} = 0$	100%	100%	100%	70%	70%	72%	40%	40%	42%	10%	12%	14%
$\beta_{min} = 0.3$	100%	100%	100%	49%	49%	52%	28%	29%	31%	7%	7%	8%
$\beta_{min} = 0.6$	100%	100%	100%	28%	29%	31%	16%	18%	21%	4%	5%	8%
$\beta_{min} = 0.9$	100%	100%	100%	7%	7%	9%	4%	4%	7%	1%	1%	3%

Les résultats précédents comparent les algorithmes ADPS et EADPS par rapport à un même seuil de tolérance. Dans ce cas les qualités des modèles sont différentes puisque les modèles sont élaborés à partir de sources de données différentes. En effet, pour le schéma EADPS les modèles de prédiction sont construits comme on le sait déjà à partir des modèles obsolètes. Maintenant si l'on se fixe une même qualité de modèle en ajustant les tolérances, on obtient les résultats présentés sur les tableaux 4.5 à 4.7. Comme prévu, nous avons pu observer l'avantage du schéma EADPS à même qualité de modèle que celui utilisé dans le schéma ADPS. La différence en valeur absolue entre les résultats obtenus à partir de la simulation EADPS et ceux fournis par le modèle décrit par l'équation (4.7) ne dépasse pas 5%. De plus, au vu de la différence absolue entre le schéma ADPS et le schéma EADPS, l'amélioration du taux de réduction des transmissions a atteint 4%. Par conséquent, l'algorithme EADPS permet d'atteindre un taux de réduction de transmission de 99%.

4.6 Etude à l'échelle du réseau

4.6.1 Configuration de la simulation

La simulation constitue la meilleure alternative pour comprendre en peu de temps et à faible coût le comportement d'un système RCSF. L'objectif principal est de savoir comment le nombre de transmissions varie dans les schémas de prédiction. Divers logi-

ciels de simulation tels que NS-2, OMNeT ++, Prowler, Atarraya, PiccSIM, Truetime et MATLAB®/Simulink® ont été analysés dans la littérature spécialisée [121, 122]. Il a été constaté qu'aucun des logiciels n'était le mieux adapté à l'analyse structurale. En conséquence, un cadre de simulation RCSF personnalisé a été conçu sous l'environnement Matlab.

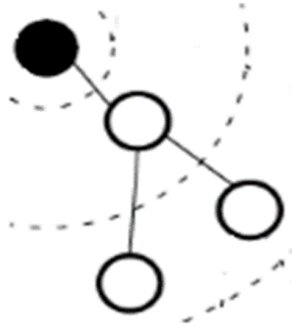


FIGURE 4.17 – Schéma du RCSF étudié avec des données réelles

Nous organisons notre RCSF selon un modèle en anneaux où chaque branche est connectée à la station de base (BS) via un nœud capteur, on rappelle qu'il est possible de déduire le nombre de transmissions à ce nœud capteur en considérant uniquement le nombre de fils qu'il aurait. Notre réseau est constitué de trois capteurs nœuds et d'un nœud récepteur (BS, CH). Différentes configurations sont possibles selon le nombre d'anneaux sélectionnés (1 anneau, 2 anneaux ou 3 anneaux). Nous décidons de sélectionner le modèle représenté par la figure 4.17, c'est-à-dire un modèle avec deux anneaux afin de permettre la simulation des techniques d'agrégation de données en plus de celles de prédiction. Bien entendu, on peut étudier chacune de ces configurations et établir une analyse comparative mais ce n'est pas l'objet de ce travail. On note que, dans le modèle RCSF considéré, lorsque de nouveaux anneaux sont ajoutés, le nombre de nœuds capteurs augmente de façon quadratique (CD^2) si aucune agrégation n'est adoptée. Cependant, le nombre de transmissions ne change pas si les nœuds capteurs les agrègent. Enfin, dans nos simulations, les nœuds capteurs ont effectué une mesure toutes les 30 minutes (1/2 heure), soit 48 mesures toutes les 24 heures, ce qui donne une fréquence de détection $f = 2$.

4.6.2 Hypothèses

Dans cette étude nous appliquerons les algorithmes de prédiction duale et les techniques d'agrégation de données. On étudiera également la combinaison de ces deux approches. On s'intéressera uniquement au nœud capteur du premier anneau car c'est le plus chargé. En effet, toutes les communications du réseau passent par ce nœud capteur qui constitue un goulot et peut empêcher le fonctionnement de tout le réseau. On ne considère pas le problème de collisions lors des transmissions des données météorologiques. On adopte l'hypothèse d'invariabilité de la structure mathématique du modèle élaboré lorsque nos algorithmes sont activés. De plus, le problème de synchronisation entre les nœuds capteurs et leurs voisins n'est pas pris en compte. Cependant, ce travail peut être facilement étendu pour prendre en compte cet aspect. Aussi, les applications de suivi

d'objets et de détection d'événements ne sont pas couvertes par ce travail en raison de leurs exigences particulières telles qu'une plus grande fiabilité ou délais plus faibles [123]. Nous supposons également que les données collectées suivent une distribution normale. Cette hypothèse est largement admise en météorologie. Enfin, en pratique, il est tout à fait possible de bénéficier des avantages potentiels d'intégration ultérieure du Cloud, en exploitant l'asymétrie du calcul du noeud BS pour qu'il soit responsable de la génération des modèles de prédiction et de leur diffusion via un mode unicast vers les nœuds capteurs.

4.6.3 Impact du paradigme de communication

Comme nous l'avons déjà mentionné au paragraphe (b) du chapitre (3), l'efficacité d'un schéma de prédiction duale est tributaire du compromis entre le nombre de données transmises et la qualité des mesures fournies. Autrement dit, le gain en nombre de transmissions de données ne doit pas être compensé par les extra-transmissions dues à la dissémination des modèles. C'est pourquoi, il y a lieu d'effectuer un évaluation préalable afin d'envisager la bonne taille de la fenêtre temporelle mobile correspondante à la durée d'actualisation du modèle. Il s'agit là d'un exercice assez délicat car il faudra rechercher le meilleur compromis entre la taille de la fenêtre de mise à jour du modèle et la précision de ce dernier. En effet, en prévision, l'efficacité de la prédiction de données ne dépend pas seulement de la qualité du modèle sélectionné car il y a lieu de s'interroger sur la façon dont ce modèle est utilisé, c'est à dire son efficacité en fonction du mode de communication adopté entre les éléments constitutifs du RCSF ainsi que de la tolérance imposée par l'utilisateur. Dans le cadre des schémas de prédiction duale, le paradigme de communication entre les nœuds capteurs et les nœuds récepteurs (station de base, passerelle et clusterhead) joue un rôle fondamental. En effet, sachant que la taille de la fenêtre d'estimation impacte directement la précision du modèle, alors la taille de la fenêtre d'actualisation ou de mise à jour du modèle impacte plutôt le taux de transmission de données. On montre à travers les résultats illustrés par la figure 4.18 l'évolution des taux de transmission de données en fonction de la taille d'actualisation du modèle. On remarque que le niveau du taux de transmission varie lentement à partir d'une fenêtre égale à 10 jours (soit 480 mesures). La taille de cette fenêtre est jugée trop longue car les taux de transmission restent élevés durant cette période, ce qui influe négativement sur la consommation de l'énergie. De plus, les extra-transmissions dues à la dissémination du modèle aggravent cette situation et risquent de rendre le paradigme DPS inefficace, sinon on sera obligé d'opérer un changement sur la durée de prédiction et/ou sur la fréquence d'acquisition des données comme expliqué au paragraphe (3.5) du chapitre (3).

Lorsqu'on considère un paradigme ADPS où le modèle est actualisé simultanément dans les nœuds récepteurs et dans les nœuds capteurs, l'impact de la taille de la fenêtre d'actualisation du modèle est similaire à celui du paradigme DPS sauf qu'ici il n'y aura pas d'extra-transmissions dues à la dissémination du modèle. Les résultats de la figure 4.18 montrent bien qu'il faut réaliser un bon compromis entre la précision du modèle et la taille de la fenêtre de sa mise à jour. Cet aspect reste une question d'optimisation assez délicate et ne peut être résolue qu'à travers une série de simulations ou des essais pratiques préalables. Cependant, pour s'affranchir de ce problème de compromis on peut adopter une politique de tailles adaptatives pour les fenêtres temporelles d'actualisation des modèles. En d'autres termes, on peut imposer une actualisation du modèle dès qu'il

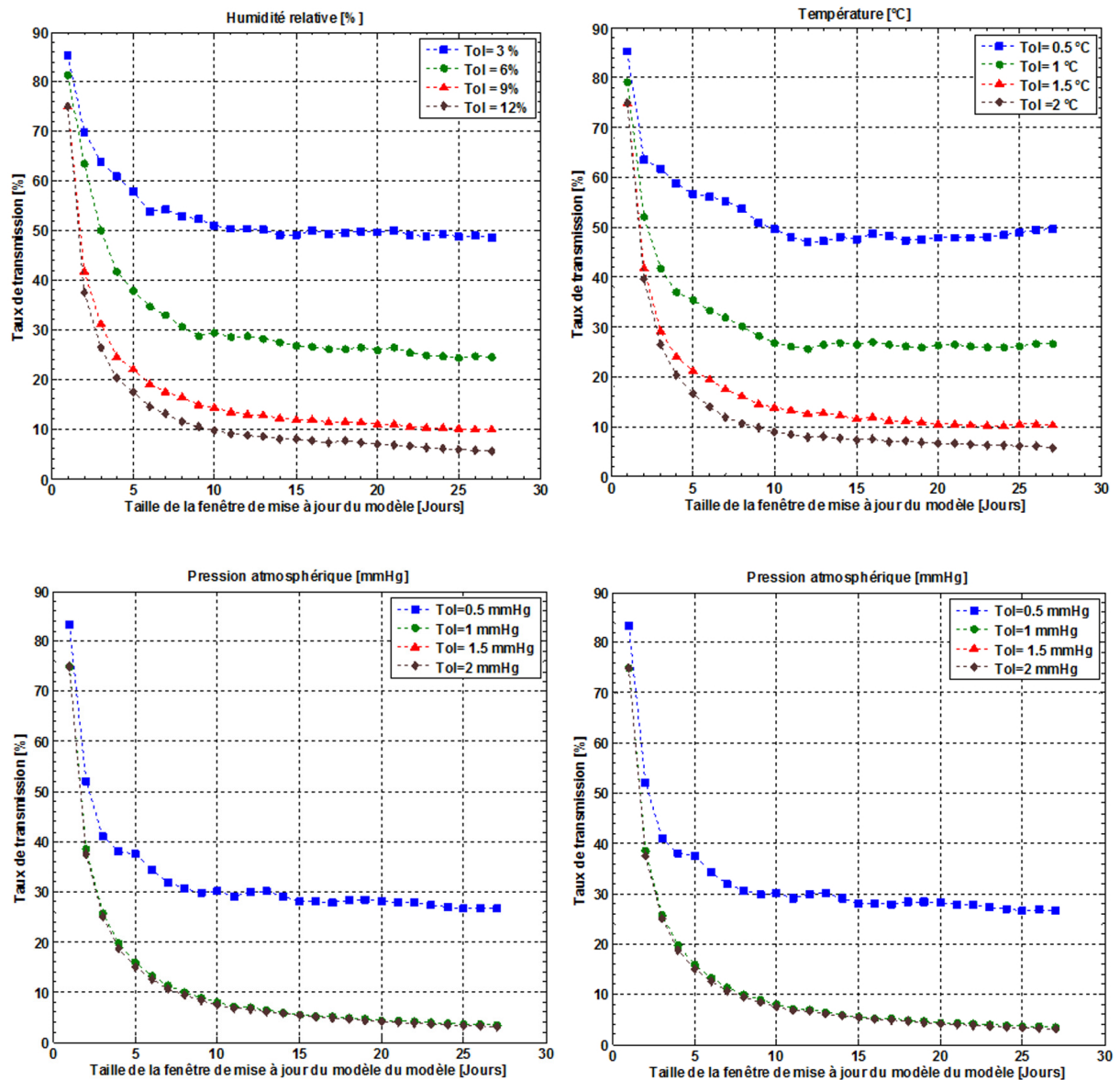


FIGURE 4.18 – Taux de transmission en fonction de la taille de la fenêtre d’actualisation du modèle dans le schéma DPS

y a violation de la tolérance imposée par l’utilisateur. C’est pourquoi nous avons décidé dans ce travail d’éviter la problématique du compromis évoqué plus haut en adoptant des paradigmes à fenêtre adaptative qui sont évidemment des fenêtres de tailles variables. Ainsi, nous sélectionnons à titre d’illustration les paradigmes suivants :

1. Paradigme DPS sans agrégation de données. Dans ce paradigme le modèle est élaboré soit dans les nœuds capteurs et transmis aux nœuds récepteurs ou inversement. Il y a donc lieu de tenir compte des extra-transmissions dues à la dissémination des modèles.
2. Paradigme DPS avec agrégation de données. C’est un mode de communication ana-

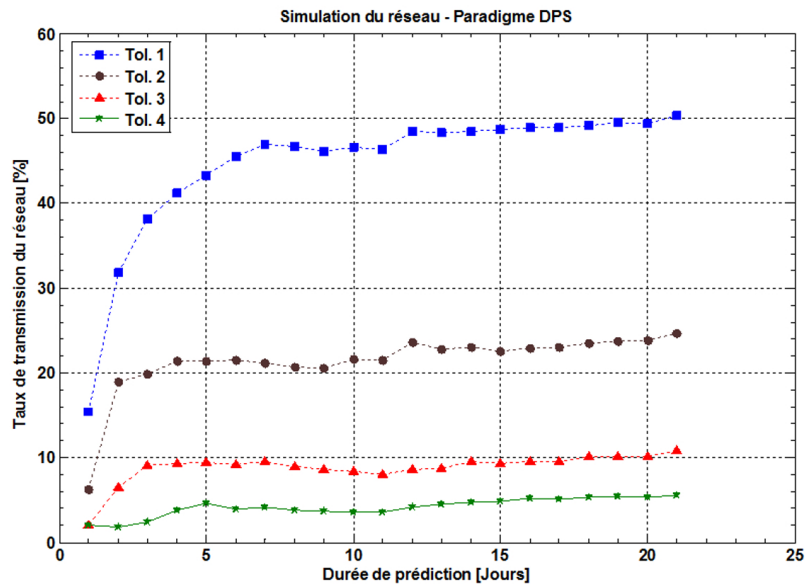


FIGURE 4.19 – Evolution des taux de transmissions du réseau étudié dans le paradigme DPS

logue au précédent mais avec intégration du procédé d'agrégation. Dans ce cas le taux de dissémination des modèles varie en fonction de la taille du RCSF comme décrit au chapitre (3).

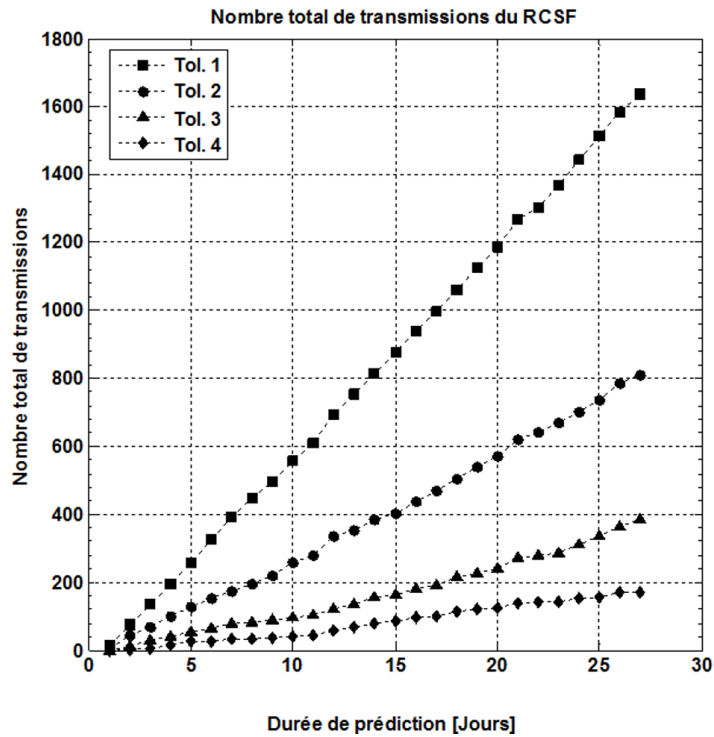
3. Paradigme ADPS. Dans ce mode de communication, nous allons appliquer une version légèrement différente de celle utilisée par F. Aderohunmu et al. [118]. En fait, en s'inspirant du paradigme EADPS, on va imposer une mise à jour au modèle de prédiction dès qu'il viole la tolérance imposée par l'utilisateur. Cette mise à jour s'effectuera à l'aide des valeurs antérieures à l'instant 't'. C'est en fait le même paradigme que le paradigme EADPS sauf qu'ici le modèle ne s'actualise pas à l'aide de la base des modèles obsolètes mais il exploite directement les valeurs des données collectées ou prédites. On s'attend donc intuitivement à ce que ce paradigme sera meilleur que celui du paradigme EADPS.
4. Paradigme ADPS avec agrégation de données. Même mode que le précédent mais avec intégration du procédé d'agrégation.

4.6.4 Résultats obtenus et discussion

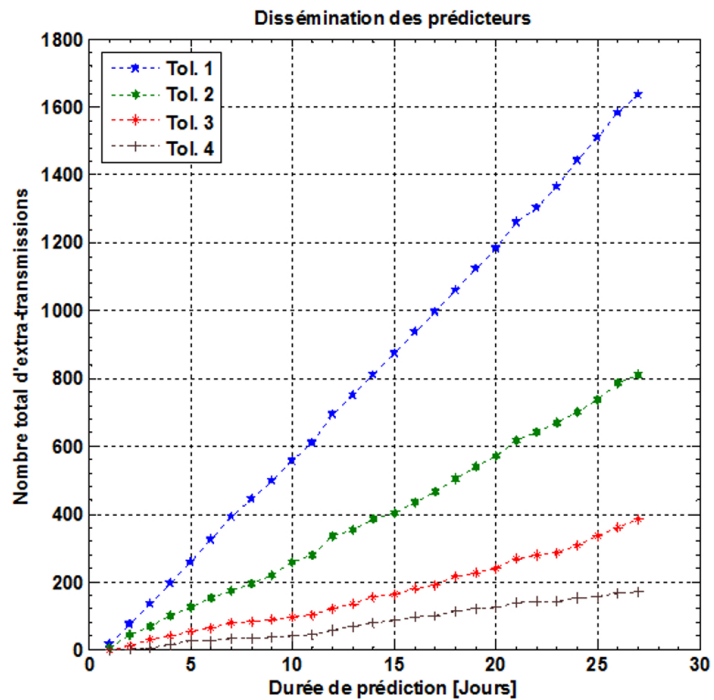
La figure 4.19 illustre l'évolution des taux de transmissions de données en fonction de la durée de prédiction et du niveau de tolérance imposée dans le cadre du paradigme DPS.

On s'aperçoit que pour les tolérances larges, les taux de transmissions restent inférieurs à 10% avec un maximum de 50% pour la tolérance la plus serrée. Aussi, au-delà d'une durée de 5 jours, les taux de transmission varient légèrement et une certaine constance semble s'installer ce qui traduit encore une fois la bonne qualité des modèles sélectionnés.

Les figures 4.20a et 4.20b, montrent l'évolution du nombre total de transmissions du réseau étudié ainsi que l'évolution du nombre total des extra-transmissions dues à la dissé-



(a) Nombre total de transmissions dans le réseau



(b) Nombre total d'extra-transmissions dans le réseau

FIGURE 4.20 – Evolution du nombre total de transmissions et d'extra-transmissions du réseau étudié dans le paradigme DPS

mination des modèles. Bien évidemment, ces deux évolutions sont identiques compte tenu de notre politique d'adoption de fenêtres adaptatives pour l'actualisation des modèles.

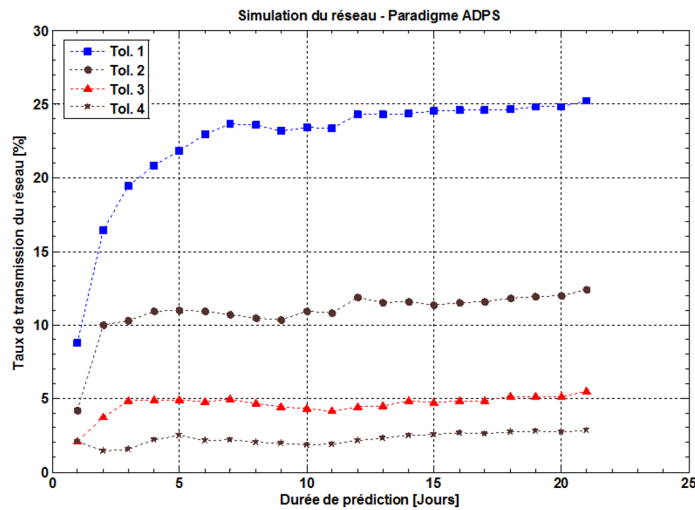


FIGURE 4.21 – Evolution des taux de transmissions en utilisant le paradigme ADPS

La figure 4.21 montre l'évolution des taux de transmission dans le cadre du paradigme ADPS. Les taux enregistrés ici sont de 50% inférieurs à ceux atteints avec le paradigme DPS car il n'y a pas d'extra-transmissions dans ce paradigme.

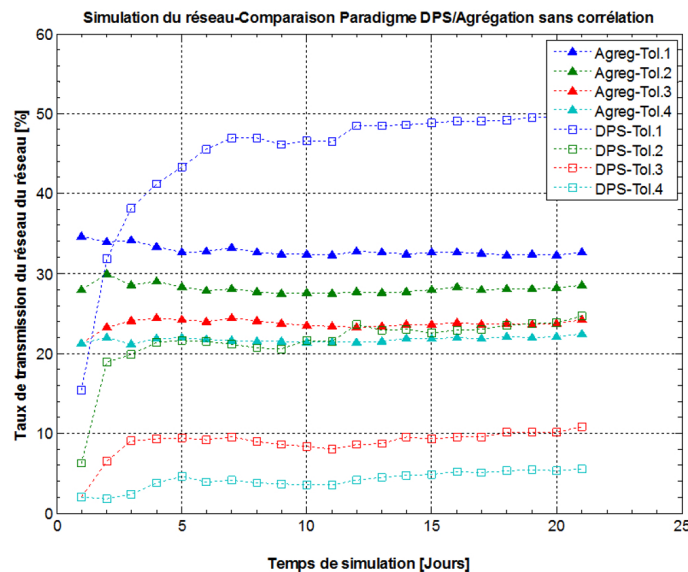
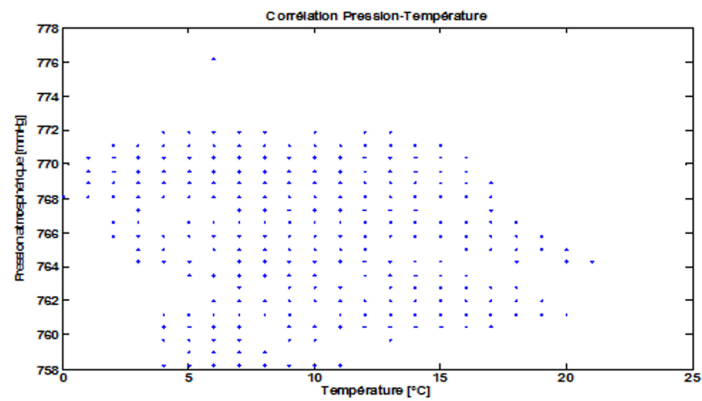
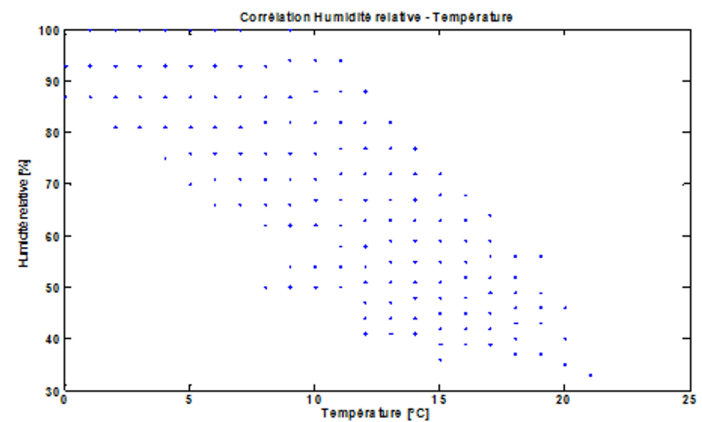


FIGURE 4.22 – Comparaison du paradigme DPS au procédé d'agrégation sans corrélation

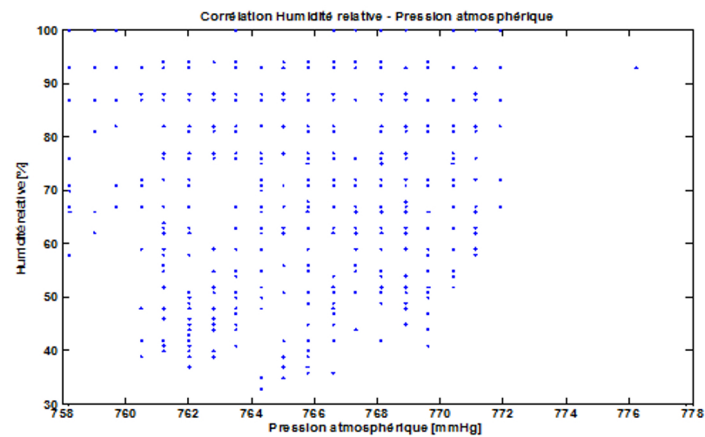
Comparons maintenant le paradigme de prédiction DPS aux procédés d'agrégation avec et sans corrélation de données. A cet effet, il faudra considérer le calcul de la matrice de covariance. Dans le cas où la corrélation des données n'est pas prise en compte, la matrice de covariance est constituée d'éléments nuls sauf sa diagonale contenant les variances des données collectées. La figure 4.22 compare l'application de l'agrégation sans corrélation de données au paradigme DPS. On voit clairement la puissance de l'agrégation des données pour la tolérance la plus serrée. Pour les autres tolérances, on note que le paradigme DPS est bien meilleur que l'agrégation compte tenu de la taille de notre réseau ($D = 2$).



(a)



(b)



(c)

FIGURE 4.23 – Corrélations entre les différentes grandeurs météorologiques

Pour comparer le paradigme DPS au procédé d'agrégation avec corrélation, il y a lieu de disposer de la matrice de covariance. Mais avant tout regardons comment sont corrélées les données à travers la figure 4.23. On voit bien sûr une collecte mensuelle que la température est bien corrélée avec l'humidité relative de l'air mais assez faiblement avec la pression atmosphérique. En revanche, l'humidité relative de l'air est assez mal corrélée

à la pression atmosphérique. Il y a lieu de noter que la matrice de corrélation varie en fonction de la durée de simulation et que les coefficients de corrélation sont supposés définis selon la méthode de Pearson.

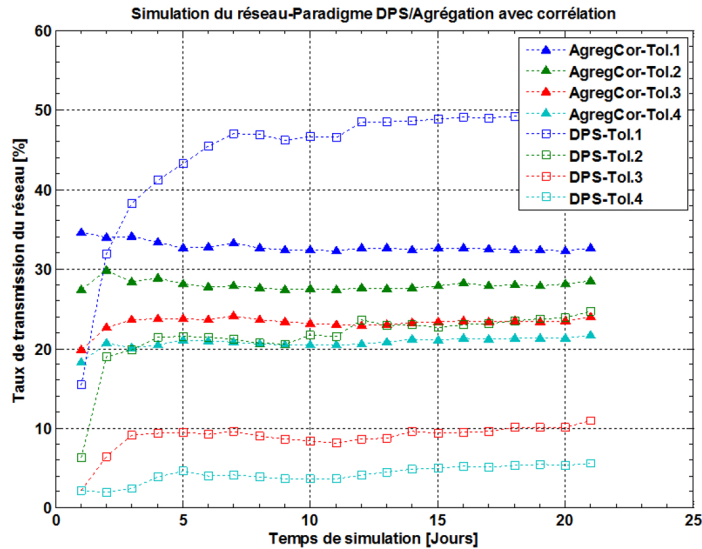


FIGURE 4.24 – Comparaison du paradigme DPS au procédé d'agrégation avec corrélation

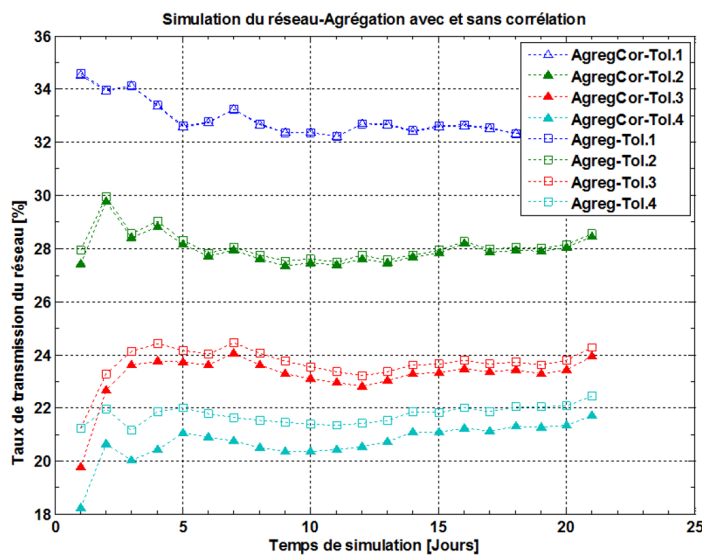


FIGURE 4.25 – Cas d'application de l'agrégation avec et sans corrélation

La figure 4.24 montre la comparaison de l'agrégation avec corrélation au paradigme DPS. On observe des résultats similaires à ceux montrés sur la figure 4.22. En effet, l'impact de la corrélation des données n'est pas très significatif au regard de la figure 4.25 qui compare l'évolution des taux de transmission de données lorsqu'on applique l'agrégation avec et sans corrélation. On voit bien que les taux de transmission générés par l'agrégation sans corrélation sont légèrement supérieurs à ceux de l'agrégation avec corrélation et ces écarts s'estompent lorsque le niveau de tolérance devient de plus en plus serré. En d'autres termes, lorsque l'agrégation avec corrélation est appliquée, elle produit

des résultats très proches de ceux générés par l'agrégation sans corrélation notamment pour les tolérances serrées.

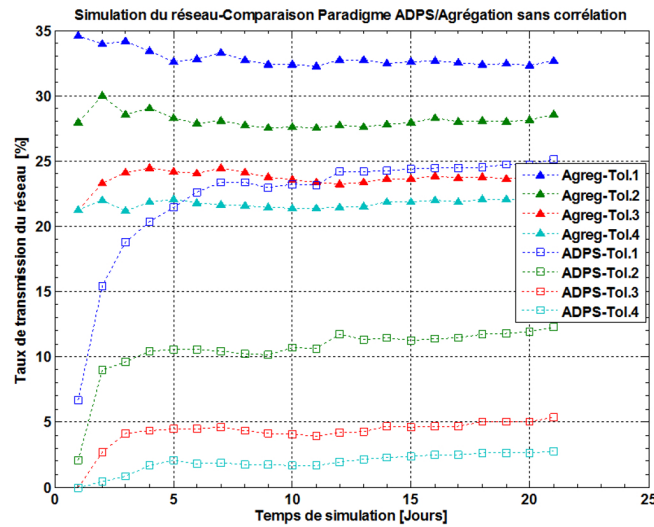


FIGURE 4.26 – Comparaison du paradigme ADPS au procédé d'agrégation sans corrélation

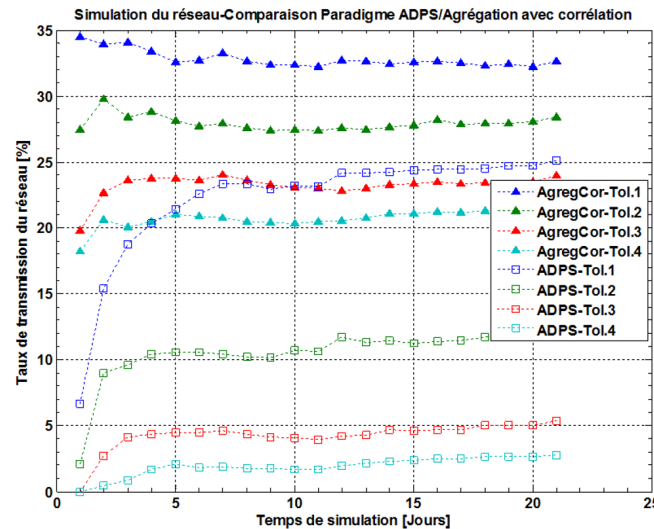


FIGURE 4.27 – Comparaison du paradigme ADPS au procédé d'agrégation avec corrélation

Comparons maintenant le paradigme ADPS au procédé d'agrégation. Les figures 4.26 et 4.27 montrent les résultats obtenus. On voit clairement la supériorité du paradigme ADPS par rapport au procédé d'agrégation. Au vu de la similitude du paradigme EADPS au paradigme ADPS avec taille de fenêtre variable, on peut dire que ce résultat confirme celui présenté par la figure 4.3 correspondant au cas $D = 2$.

A présent, nous allons combiner les algorithmes de prédiction aux procédés d'agrégation pour apprécier les gains potentiels conformément à ce qui a été présenté dans les sections (4.4) de ce chapitre et (3.8) du chapitre (3). Les figures 4.28 et 4.29 présentent successivement les résultats de la combinaison des paradigmes DPS et ADPS avec l'agrégation sans corrélation et l'agrégation avec corrélation. Sur la figure 4.29 on note de

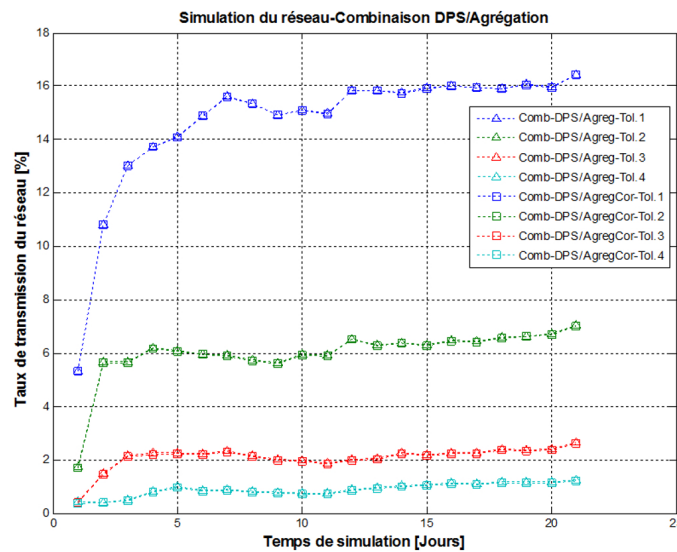


FIGURE 4.28 – Combinaison du paradigme DPS avec l'agrégation

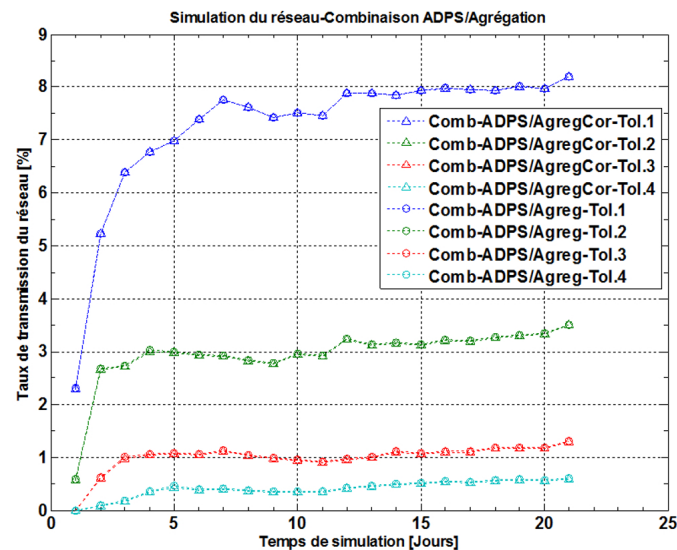


FIGURE 4.29 – Combinaison du paradigme ADPS avec l'agrégation

meilleurs résultats au profit de la combinaison du paradigme ADPS avec l'agrégation de données. Les taux maximums ne dépassent guère 8.5% et restent les plus faibles comparativement à l'agrégation seule quelque soit le niveau de tolérance. Par rapport à l'agrégation seule ou à la prédiction de données seule, la combinaison de ces deux méthodes est une stratégie bénéfique.

4.6.5 Récapitulatif

Il ressort de l'analyse des résultats présentés jusqu'ici que notre politique d'adoption d'une fenêtre variable de mise à jour des modèles de prédiction s'avère concluante et que la meilleure configuration minimisant les taux de transmission est celle qui combine le paradigme ADPS avec l'agrégation. En effet, les taux de transmission varient de 0% à

un maximum de 8.2% en fonction de la tolérance imposée par l'utilisateur. Ces résultats sont satisfaisants car ils sont en bonne concordance avec ceux obtenus analytiquement et permettent une grande réduction de la consommation d'énergie du réseau, une plus grande accessibilité au réseau de communication et un moindre volume de stockage de données.

4.7 Application du schéma EADPS au réseau étudié

Cette version est très similaire au paradigme ADPS testé plus haut, sauf qu'ici les paramètres du modèle s'actualisent à l'aide des modèles obsolètes et non pas à l'aide des données passées prédites ou mesurées (antérieures à l'instant 't').

Le modèle de prédiction qu'utilise le schéma EADPS fonctionne exactement comme celui utilisé dans le schéma ADPS. La structure du prédicteur sélectionné est restée inchangée soit un ARMA(2,1).

4.7.1 Résultats de simulation

Dans cette section, nous présentons et discutons les résultats concernant l'application au réseau du paradigme EADPS. Nous commençons par appliquer l'algorithme ADPS pour construire une première base de données des modèles obsolètes (échantillon initial). La taille initiale choisie arbitrairement étant de 12 modèles obsolètes. Cette taille constitue donc la fenêtre mobile nécessaire à l'actualisation du prédicteur EADPS. Ensuite, on déclenche l'algorithme EADPS. Dans un premier temps, seul le paradigme EADPS sera appliqué, ensuite nous le combinons avec un processus d'agrégation avec corrélation des données. Cette simulation portera sur les données météorologiques réelles du mois de Juin 2017 à l'aéroport Messali El-hadj de Tlemcen (Algérie).

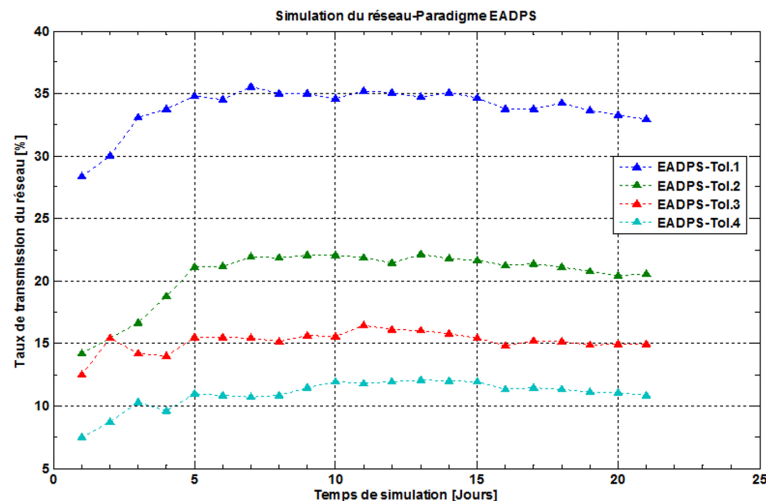


FIGURE 4.30 – Simulation du réseau en utilisant le paradigme EADPS

La figure 4.30 montre pour différents niveaux de tolérances, l'évolution du taux de transmission en fonction du temps de simulation. Comparativement au paradigme ADPS que nous avons appliqué précédemment et dont les résultats sont présentés sur la figure

4.21, on note que le paradigme EADPS est moins puissant du fait que son prédicteur est moins précis car il n'est pas élaboré directement à partir des données collectées ou prédites. En effet, le taux de transmission de données résultant du schéma EADPS varie entre 7.5% et 35% alors que celui dû au paradigme ADPS varie entre 2% et 25%.

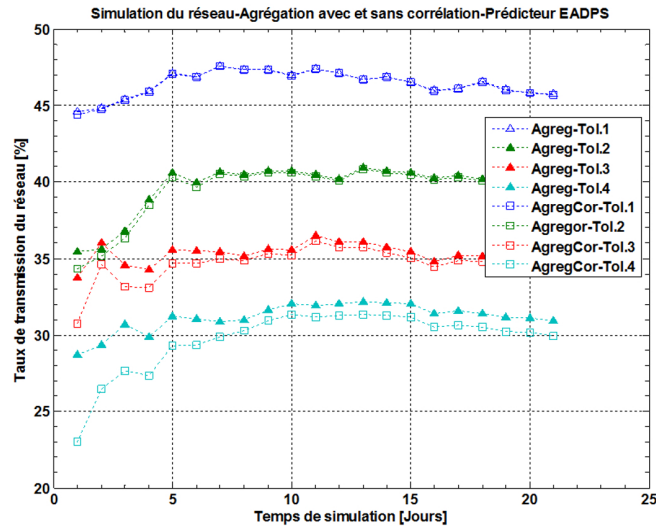


FIGURE 4.31 – Taux de transmission du réseau : Application de l'agrégation avec et sans corrélation dans le schéma EADPS

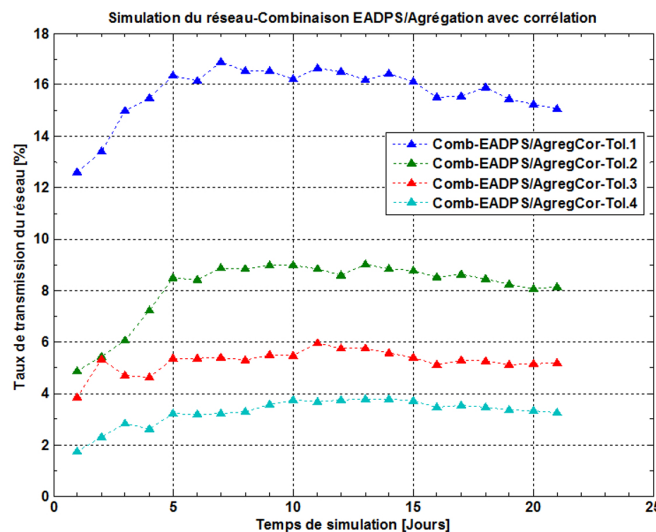


FIGURE 4.32 – Combinaison du paradigme EADPS avec l'agrégation avec corrélation de données

La figure 4.31 présente l'évolution du taux de transmission de données lorsqu'on applique une agrégation avec et/ou sans corrélation où le prédicteur résulte du paradigme EADPS. Là encore, comparativement aux cas de la figure 4.25, on note une perte de performance due essentiellement à la qualité du modèle c'est-à-dire à la valeur de α_{min} . On voit clairement dans la figure 4.31 que le taux de transmission évolue dans un intervalle

compris entre 23% et 48% alors que dans le cas de la figure 4.25 cet intervalle varie entre 18% et 35%.

La figure 4.32 traduit la combinaison du paradigme EADPS avec l'agrégation avec corrélation de données. Les mêmes conclusions se déduisent. En effet, on note une perte de performance comparativement au cas représenté par la figure 4.29 où la variation du taux de transmission évolue entre 0% et 8.2% alors que dans le cas de la figure 4.32 il évolue entre 2% et 17%.

4.7.2 Consommation d'énergie

Dans les simulations, nous avons utilisé des capteurs de type TelosB [124] dont le débit est de 250 kbps, la puissance de réception est de 38 mW, la puissance d'émission à 0 dBm est de 35 mW et la puissance de veille est de 15 μW . La consommation d'énergie du nœud i au niveau de l'anneau d pour transmettre et recevoir un paquet de données est estimée à 0.29206 $\mu J/bit$ (pour un capteur de type TelosB). Sur la base de l'équation (4.15), nous pouvons estimer la consommation d'énergie pour le nombre de transmissions (NbT) :

$$E_{Tot} = [(E_{Trans} + E_{Rec}) * k * NbT] + E_{min} \quad (4.15)$$

où E_{Tot} , E_{Trans} et E_{Rec} sont respectivement l'énergie totale, l'énergie d'émission et l'énergie de réception par bit. E_{min} est l'énergie du capteur en mode veille, c'est-à-dire l'énergie minimale nécessaire pour que le nœud capteur fonctionne sans transmettre ni recevoir quoi que ce soit. Le paramètre k représente le nombre moyen de bits par paquet. Ce paramètre est estimé à partir de notre simulation à $k = 29bits$. Enfin, à partir de l'équation (4.15), nous pouvons écrire :

$$\Delta E = (E_{Tot} - E_{min}) = 8.47 * 10^{-6} * NbT [Joules] \quad (4.16)$$

Il devient clair que la réduction de la consommation d'énergie est proportionnelle au taux de réduction des transmissions. Dans notre application, le gain d'énergie obtenu à l'aide de la combinaison ADPS-Agrégation avec corrélation a évolué dans l'intervalle 93.1% à 99.7% selon le niveau de tolérance admis.

4.8 Conclusion

Dans ce chapitre, nous avons présenté l'extension de l'algorithme ADPS en un nouvel algorithme désigné EADPS pour obtenir de meilleurs gains en termes de performances dans les RCSFs. On précise d'emblée que l'usage du processus adaptatif est en soi une contribution car il n'est pas beaucoup exploité selon notre étude bibliographique. Pour illustrer notre approche, la topologie du modèle en anneaux est utilisée afin de tirer profit du cadre mathématique récemment développé pour ce type de modèle. Les résultats obtenus sont substantiels et indiquent que l'idée proposée pour étendre le schéma ADPS a un potentiel intéressant. Dans un premier temps nous avons mené une étude analytique sur l'usage des techniques de prédiction et d'agrégation des données. Nous avons d'abord montré la supériorité de l'algorithme EADPS par rapport à l'algorithme ADPS

et nous avons ensuite montré l'impact de l'agrégation des données. Celle-ci s'avère moins performante dans les petits réseaux lorsque l'algorithme EADPS est appliqué.

Dans une seconde étape, nous avons mené une série de simulations à partir des données météorologiques réelles récoltées au niveau de l'aéroport Messali Hadj de Tlemcen. Les résultats ont montré que le nombre de transmissions des nœuds capteurs qui constituent les goulots d'étranglement (c'est-à-dire les nœuds capteurs situés dans le premier anneau et qui sont proches de la station de base) peut être réduit à près de 100% au maximum lorsque des prévisions précises sont obtenues dans le cadre de tolérances admises assez large.

Les paradigmes DPS, ADPS, EADPS et leurs combinaisons avec les techniques d'agrégation ont été testés. Il ressort que la puissance des techniques d'agrégation est fortement tributaire de la qualité ou de la précision du prédicteur. Au contraire, nous trouvons que l'impact de la corrélation des données dans notre application n'est pas très significatif. C'est pourquoi la meilleure performance atteinte dans nos essais correspond à la combinaison entre le paradigme ADPS avec taille de fenêtre variable et l'agrégation. Les taux de transmission des données varient entre 0% et 8.2%. Ces résultats génèrent un gain d'énergie de la batterie du nœud capteur dans un rapport proportionnel. Ce qui est fort intéressant dans les RCSFs qui souffrent de cet handicap (économie de l'énergie).

Chapitre 5

**Apport de l'intégration RCSF-Cloud
computing sur les performances des
RCSFs**

Chapitre 5

Apport de l'intégration RCSF-Cloud computing sur les performances des RCSFs

5.1 Introduction

L'objectif de notre application est la création d'une plateforme qui permet l'étude de plusieurs scénarios de simulation RCSF-Cloud.

Cet outil permet aux utilisateurs de décider du meilleur scénario reflétant le meilleur compromis par rapport aux objectifs fixés. Ainsi, l'utilisateur pourra étudier l'impact de différents paramètres sur les performances du réseau tels que la taille du réseau, le nombre de nœuds capteurs par anneau, le niveau de tolérance admis, la fréquence d'acquisition des données, l'ordre du prédicteur, la durée de fonctionnement et cela sous différentes stratégies de réduction des taux de transmission de données comme les techniques de prédiction, d'agrégation ou leurs combinaisons. Il s'agit dans ce chapitre de décrire l'architecture et le potentiel offert par notre logiciel. La simulation permettra l'étude d'intégration entre les RCSFs et le Cloud et permet de calculer les gains en termes de taux de transmission dans le but d'améliorer les performances des RCSFs (réduction de la consommation d'énergie des RCSFs, d'espace de stockage, de bande passante et de bourrage du réseau). Différents scénarios peuvent donc être adoptés à l'aide des algorithmes DPS, ADPS et EADPS et leur combinaison avec l'agrégation.

5.2 Environnement de développement

Avant de passer à la description de la plateforme développée, nous allons présenter le langage de programmation java, l'éditeur Eclipse, le langage émulé Matlab et le simulateur CloudSim avec son architecture et ses différentes classes de base.

5.2.1 Langage de programmation java

Java est une plate-forme informatique et un langage de programmation de haut niveau, il a été développé à l'origine par James Gosling et Patrick Naughton chez Sun Microsystems [125]. Une équipe de Sun Microsystems nommée "green team" a commencé

à travailler sur un projet de langage Java en 1991, qui a finalement été publié en 1995. Il fonctionne sur une machine virtuelle (VM), en compilant d'abord le langage Java en byte code Java VM. Java est un langage indépendant du système (facilité par les implémentations de VM sur de nombreuses plateformes), particulièrement adaptées à la programmation Web ; il a pour mantra "Ecrire une fois, exécutez n'importe où", ce qui signifie qu'il doit fonctionner de la même manière sur tout système avec une machine virtuelle Java installée. Java est un langage orienté objet en raison de ses classes et de ses objets ; il a l'avantage d'être sécurisé car il n'utilise pas de pointeurs. Il permet d'exécuter de nombreux programmes simultanément, par conséquent, le multithreading peut être réalisé. Java est robuste car il possède de nombreuses fonctionnalités telles que le Garbage collection, aucune utilisation de pointeurs explicites, la gestion des exceptions. En revanche étant un langage de haut niveau, il doit gérer les niveaux de compilation et d'abstraction d'une machine virtuelle.

5.2.2 Eclipse IDE

Eclipse est un environnement de développement logiciel multilingue, qui se compose d'un environnement de développement intégré (IDE) et d'un système de plug-in supplémentaire. Un IDE fournit de nombreux outils par exemple écriture de code, compilation, exécution, débogage, gestion de fichiers et documentation le tout sur une seule plateforme.

Eclipse est l'un des IDE les plus utilisés professionnellement pour développer des applications et solutions logicielles. Un avantage d'Eclipse par rapport aux autres IDE professionnels est qu'Eclipse est une plate-forme open source, et par conséquent, il est facile d'y ajouter de nouvelles bibliothèques et ressources. Il existe une variété de langages, dont les applications peuvent être développées dans Eclipse au moyen de systèmes de plug-ins extensibles. Certains des langages les plus programmés dans Eclipse sont JAVA, C, C ++, COBOL, Perl, Android, Python, Ruby, Groovy et Scheme.

5.2.3 Le langage émulé Matlab

Le nom MATLAB signifie Matrix Laboratory. Il a été écrit à l'origine pour fournir un accès facile au logiciel de matrice développé par le LINPACK (package de système linéaire) et projets EISPACK (Eigen system package). Matlab est un langage performant pour le calcul technique. Il intègre un environnement de calcul, de visualisation et de programmation. De plus, MATLAB est un environnement de langage de programmation moderne : il a des structures de données sophistiquées, il contient des outils d'édition et de débogage intégrés et prend en charge la programmation orientée objet. Ces facteurs font de MATLAB un excellent outil d'enseignement et de recherche.

MATLAB présente de nombreux avantages par rapport aux langages informatiques conventionnels (par exemple, C, FORTRAN) pour résoudre des problèmes techniques. MATLAB est un système interactif dont l'élément de données de base est un tableau qui ne nécessite pas de dimensionnement. Le progiciel est disponible dans le commerce depuis 1984 et est maintenant considéré comme un outil standard dans la plupart des universités et industries du monde entier. Il possède de puissantes routines intégrées qui permettent une très grande variété de calculs. Il dispose également de commandes graphiques faciles à utiliser qui rendent la visualisation des résultats immédiatement disponible. Les appli-

cations spécifiques sont rassemblées dans des packages appelés boîte à outils. Il existe des boîtes à outils pour le traitement du signal, le calcul symbolique, la théorie du contrôle, la simulation, l'optimisation et plusieurs autres domaines de la science appliquée et de l'ingénierie.

5.2.4 Le simulateur CloudSim

CloudSim a été choisi comme plateforme de simulation. C'est un Framework de modélisation et de simulation d'infrastructures et de services de cloud computing. Initialement construit au niveau du laboratoire de Cloud Computing et systèmes distribués "Cloud computing and Distributed systems (CLOUDS)" de l'Université de Melbourne, Australie.

CloudSim est devenu l'un des simulateurs de cloud open-source les plus populaires dans la recherche et le monde universitaire. CloudSim est entièrement écrit en Java.

a) Architecture générale de CloudSim

Dans cette section, nous discutons de l'architecture en couches de CloudSim. La figure 5.1 montre l'implémentation en couches de la structure logicielle CloudSim et de ses composants architecturaux. Au niveau le plus bas se trouve le moteur de simulation d'événements discrets SimJava [9] qui implémente les fonctionnalités de base requises pour les cadres de simulation de plus haut niveau tels que la mise en file d'attente et le traitement des événements, la création de composants système (services, hôte, centre de données, courtier, machines virtuelles), la communication entre les composants et la gestion de l'horloge de simulation.

La couche CloudSim fournit une nouvelle prise en charge pour la modélisation et la simulation de centre de données dans l'environnement des Cloud computing, y compris les interfaces de gestion pour les machines virtuelles, la mémoire, le stockage et la bande passante. La couche CloudSim gère l'instanciation et l'exécution des entités principales (VM, hôtes, centres de données, application) pendant la période de simulation. Cette couche est capable d'instancier simultanément et de gérer de manière transparente une infrastructure Cloud à grande échelle composée de milliers de composants système. Les problèmes fondamentaux tels que l'approvisionnement des hôtes vers les VM en fonction des demandes des utilisateurs, la gestion de l'exécution des applications et la surveillance dynamique sont traités par cette couche.

La couche la plus élevée de la pile de simulation est le code utilisateur (user code) qui expose les fonctionnalités liées à la configuration des hôtes (nombre de machines, leurs spécifications, etc.), des applications (nombre de tâches et leurs exigences), des machines virtuelles, du nombre d'utilisateurs et de leur application types et des politiques de planification des courtiers. Un développeur d'applications cloud peut générer un mélange de configurations d'application et de scénarios de disponibilité du cloud à ce stade. Dans cette couche des tests robustes basés sur les configurations Cloud personnalisées déjà prises en charge dans CloudSim sont effectués.

b) Les Classes de CloudSim

La figure 5.2 illustre le diagramme de conception de classes pour le simulateur CloudSim. Dans cette section, nous présentons des détails plus fins liés à quelques classes fon-

de "QoS" des usagers et déploie les tâches de service dans les Clouds. Le courtier agissant au nom des usagers identifie les fournisseurs de services Cloud appropriés via le "Cloud Information Service (CIS)" et négocie avec eux une allocation de ressources qui répond aux besoins de QoS des usagers. Les chercheurs et les développeurs de systèmes doivent étendre cette classe pour mener des expériences avec leurs politiques de placement d'applications développées sur mesure.

3. SANStorage : Cette classe modélise un réseau de stockage qui est généralement disponible pour les data centers basés sur le Cloud pour stocker de grandes quantités de données. SANStorage implémente une interface simple qui peut être utilisée pour simuler le stockage et la récupération de toute quantité de données, à tout instant en fonction de la disponibilité de la bande passante du réseau. L'accès aux fichiers d'un SAN au moment de l'exécution entraîne des délais supplémentaires pour l'exécution de l'unité de tâches, en raison du temps écoulé pour transférer les fichiers de données requis via le réseau interne du centre de données.
4. VirtualMachine : Cette classe modélise une instance d'une VM, dont la gestion au cours de son cycle de vie est de la responsabilité du composant Host (hôte). Un hôte peut simultanément instancier plusieurs machines virtuelles et allouer des processeurs en fonction des politiques de partage de processeurs prédéfinies (espace partagé, temps partagé). Chaque composant de VM a accès à un composant qui stocke les caractéristiques liées à une VM, telles que la mémoire, le processeur, le stockage et la politique de planification interne de la VM, qui est étendue à partir du composant abstrait appelé VMScheduling.
5. Cloudlet : Cette classe modélise les services applicatifs basés sur le Cloud (livraison de contenu, réseautage social, workflow d'entreprise), qui sont couramment déployés dans les centres de données. CloudSim représente la complexité d'une application en termes de besoins de calcul. Chaque composant d'application a une longueur d'instruction pré-assignée (héritée du composant Gridlet de GridSim) et la quantité de transfert de données (avant et après l'extraction) qui doit être entreprise pour héberger l'application avec succès.

Dans ce qui suit nous allons décrire l'architecture générale de notre simulateur que nous avons nommés PREDAG.1.0, nous apporteront ensuite une description de notre logiciel ainsi son mode de fonctionnement et des étapes à suivre pour dérouler un scénario de simulation et enfin nous décrirons le diagramme de conception de classe pour notre simulateur avec un brève tour d'horizon sur quelques classes fondamentales de notre simulateur.

5.3 Architecture générale du simulateur PREDAG

Dans cette partie l'architecture globale de notre logiciel est illustrée, comme le montre la figure 5.3. L'architecture proposée est composée principalement de trois unités fondamentales :

1. Unité d'acquisition de données : dans cette unité on représente le RCSF où les capteurs sont directement implémentés dans l'environnement. Ces capteurs récoltent les

données et les envoient directement à la passerelle. Aucun traitement n'est effectué à ce niveau, les données brutes sont envoyées directement à la station de base.

2. Unité de gestion des données : cette unité est responsable de la gestion et du traitement des données. Le but de ce traitement est d'améliorer la durée de vie et les performances du réseau. Sachant que la consommation d'énergie des capteurs est essentiellement due aux transmissions des données, il est donc nécessaire de diminuer les taux de transmission pour réduire la consommation d'énergie. Deux techniques peuvent être utilisées. La technique de prédiction où les données sont périodiquement prédites et donc non transmises et la technique d'agrégation où les données sont regroupées avant d'être transmises au nœud récepteur (la station de base). Les données traitées seront ensuite stockées dans les serveurs Cloud.
3. Environnement Cloud : L'environnement Cloud offre à l'utilisateur des services IaaS ainsi qu'un SaaS. Les opérations de traitement et de stockage seront effectuées sur les données reçues au sein du Cloud.

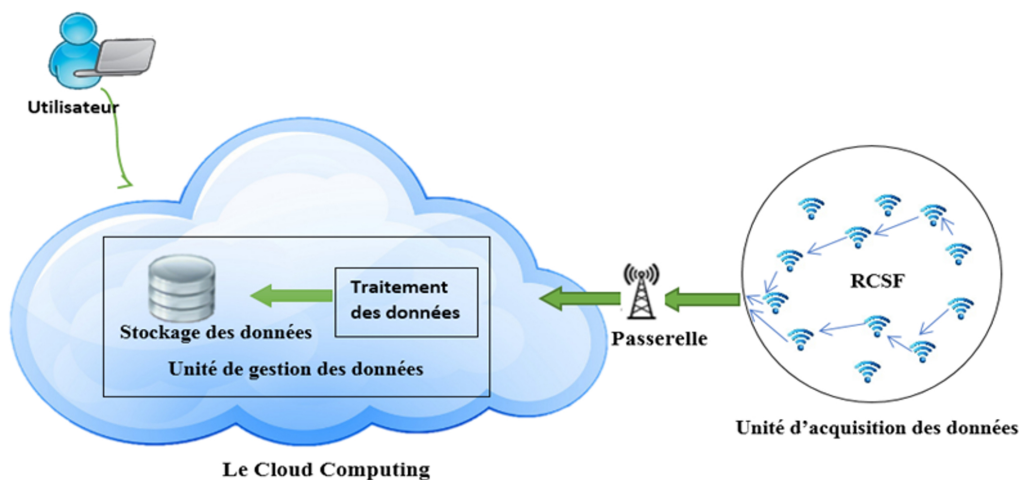


FIGURE 5.3 – Architecture d'intégration entre RCSF et Cloud Computing

5.4 Description de la plateforme

Avant de lancer la simulation l'utilisateur doit parcourir les étapes suivantes :

1. **La configuration de l'environnement Cloud** : Dans cette partie et comme le montre la figure 5.4 l'utilisateur doit saisir des informations dont le simulateur a besoin afin de créer l'environnement Cloud :
 - **Data center** : afin de créer un centre de données l'utilisateur doit remplir les informations suivantes : le nom du data center, son architecture, son système d'exploitation (SE), la politique d'ordonnancement qu'il utilise, le nom du gestionnaire des machines virtuelles (VMM), et les différents coûts (mémoire, traitement, bande passante et traitement, intervalle d'ordonnancement).

- **Machines physiques** : afin de créer une ou plusieurs machines physiques la configuration suivantes est nécessaire : le nombre des machines, le nombre des processeurs, la capacité de la RAM, le nombre des MIPS, la bande passante et la capacité de stockage. Le simulateur permet à l'utilisateur de créer des machines physiques hétérogènes.
- **Machines Virtuelles** : afin de créer des machines virtuelles il faut remplir les champs suivants : le nombre des VMs, le nombre des processeurs, la capacité de la RAM pour chaque VMs, le nombre des MIPS, la bande passante et la capacité de stockage. Comme les machines physique le simulateur offre aussi la possibilité à l'utilisateur de créer des machines virtuelles hétérogènes.
- **Les Cloudlets** : pour les Cloudlets l'utilisateur doit saisir la taille des Cloudlets, le nombre des Cloudlets et le nombre de processeurs.

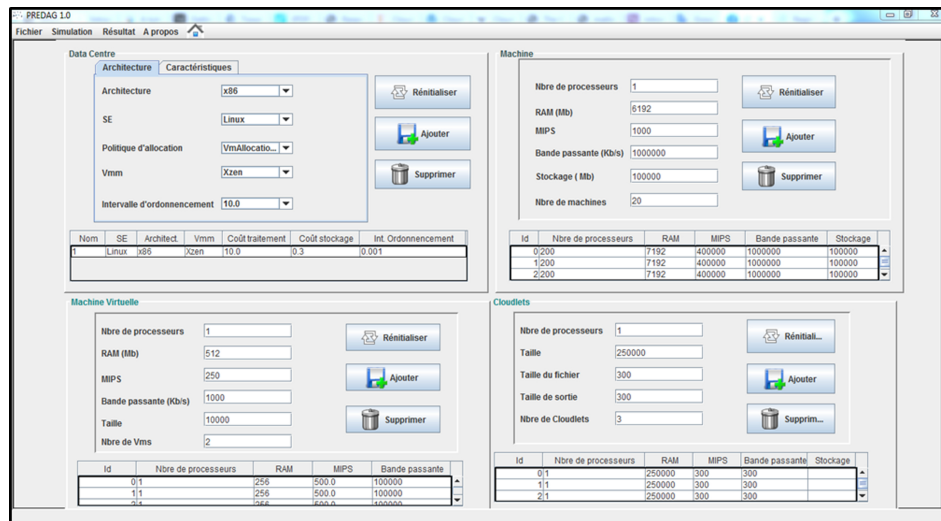


FIGURE 5.4 – Configuration de l'environnement Cloud Computing

2. La création du RCSF : la figure 5.5 illustre l'interface qui permet à l'utilisateur de gérer la partie réseau de la simulation. Cette interface donne la possibilité à l'utilisateur de créer un réseau de capteurs sans fil adaptables à plusieurs types d'application. On y retrouve la création des nœuds capteurs et la génération du réseau.

- **Caractéristiques des nœuds capteurs** : L'utilisateur devra saisir les différentes caractéristiques techniques du nœud capteur à savoir la capacité mémoire de chaque nœud capteur, le niveau de la batterie, l'énergie de transmission, l'énergie de réception, l'énergie minimale de veille et le débit. Les capteurs créés sont stockés dans un tableau de caractéristiques et peuvent être hétérogènes ou homogènes.
- **Caractéristique du réseau de capteur** : le simulateur permet à l'utilisateur de choisir entre une configuration automatique et une configuration manuelle pour la création du réseau modélisé en anneau. Dans la première configuration le réseau de capteurs sans fil est généré automatiquement selon le modèle en

anneau. Il suffit à l'utilisateur de remplir des informations élémentaires sur le réseau, là aussi le simulateur donne le choix à l'utilisateur entre fournir le nombre de capteurs désiré dans le réseau ou bien la taille du réseau (le nombre d'anneaux contenus dans le réseau). La densité du réseau doit être remplie. Elle représente le nombre de capteurs présent dans le premier anneau et enfin la période d'acquisition des données. En revanche dans la deuxième configuration l'utilisateur devra remplir manuellement un tableau plaçant le nombre d'anneaux et le nombre de nœuds capteurs dans chaque anneau, d'autres paramètres comme le nombre de fils directs (Id_1), le nombre de fils total (Kd_1) et la période d'acquisition devront aussi être remplis. Une fois le réseau est créé, un tableau s'affiche avec le nombre de lignes représentant le nombre de capteurs dans le réseau et chacun de ses capteurs affichera l'identifiant de son père et le nombre de ses fils directs. Une représentation visuelle d'organisation du réseau s'affichera après que l'utilisateur ait appuyé sur le bouton affiché réseau. Un exemple d'affichage est représenté dans la figure 5.5. Ce réseau sera prêt pour les prochaines étapes de la simulation.

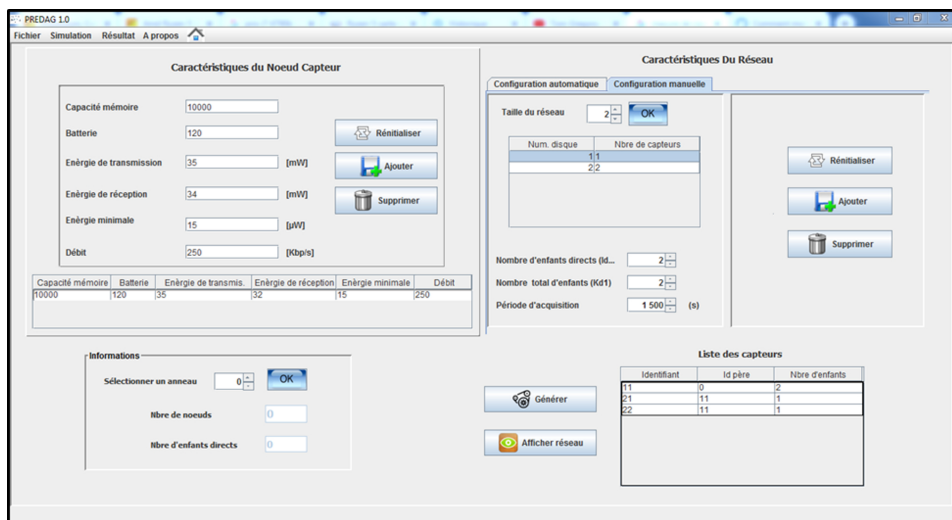


FIGURE 5.5 – Configuration manuelle du RCSF

3. **Configuration des paramètres de simulation** : Après la création du RCSF selon le modèle en anneaux et la configuration de l'environnement Cloud avec ses différentes ressources, l'utilisateur doit passer par le paramétrage du paradigme de prédiction à exécuter dans l'environnement Cloud. L'utilisateur aura le choix entre le modèle EADPS et le modèle ADPS. Afin de simuler le modèle EADPS et comme le montre la figure 5.6 l'utilisateur devra fournir les données suivantes : le modèle de prédiction à utiliser, l'ordre de ce modèle, l'ordre du bruit du modèle de prédiction, la durée de simulation du déroulement du modèle, le coefficient de la taille d'estimation initiale, le nombre de jours (N), le nombre et les valeurs des tolérances admises sur les données dans les différents capteurs.
4. Enfin, la simulation sera exécutée selon les paramètres introduits et les schémas de prédiction : EADPS et ADPS.

Integration entre Cloud et RCSF

Programme de simulation
Veuillez sélectionner un programme de simulation

EADPS ADPS

Données de simulation

Modèle de prédiction: ARMA Ordre AR: 2

Durée de simulation: 4 (Jour) Ordre MA: 1

Nbre de Tolérances: 4 Coef. taille initiale d'estimation: 0.75

Nombre de jours: 5

Tolérances admises			
Tolérances - Capteurs	1	2	3
1	0.5	0.5	3
2	1	1	6
3	1.5	1.5	9
4	2	2	12

Fichier de données: C:\Users\Dell\Documents\MATLAB\VAL\new.xlsx

Valider

FIGURE 5.6 – Paramétrage de la simulation

5.5 Résultats et discussion

La figure 5.7 montre clairement que la consommation d'énergie croît linéairement en fonction de la durée de simulation, ce qui indique la nécessité de recourir à une solution permettant de stabiliser cette consommation dans le temps. Sur la figure 5.8, on perçoit le gain d'énergie du réseau lorsqu'on applique la technique de prédiction de données à l'aide du paradigme ADPS. On voit clairement que le gain en énergie augmente proportionnellement avec l'accroissement du niveau de la tolérance sélectionnée. Ce gain peut atteindre 100% au maximum. L'évolution de ce gain est stable à partir du 5^{ème} jour de simulation ce qui traduit la robustesse du prédicteur. On remarque que durant la période de simulation allant de 1 à 5 jours les gains en énergie sont plus élevés car le taux de transmissions de données est le plus faible.

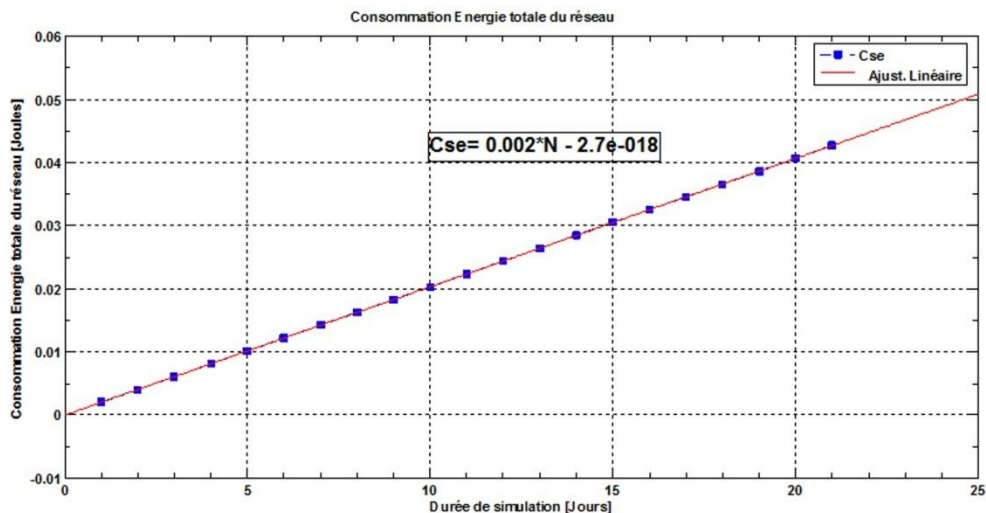


FIGURE 5.7 – Consommation d'énergie dans un RCSF (sans agrégation et prédiction)

La figure 5.9 montre le gain d'énergie du réseau durant l'application de la technique de prédiction des données à l'aide du paradigme DPS. Comme dans le cas du paradigme ADPS on remarque que le gain en énergie augmente lorsque le niveau de tolérance aug-

mente et peut atteindre un maximum de 98%. Cependant, le niveau du gain reste plus faible que celui atteint pour le paradigme ADPS. La figure 5.10 illustre l'évolution du gain d'énergie lorsqu'on applique l'agrégation de données. On voit d'abord que le gain est meilleur lorsque l'agrégation est appliquée avec corrélation de données et cela parti-

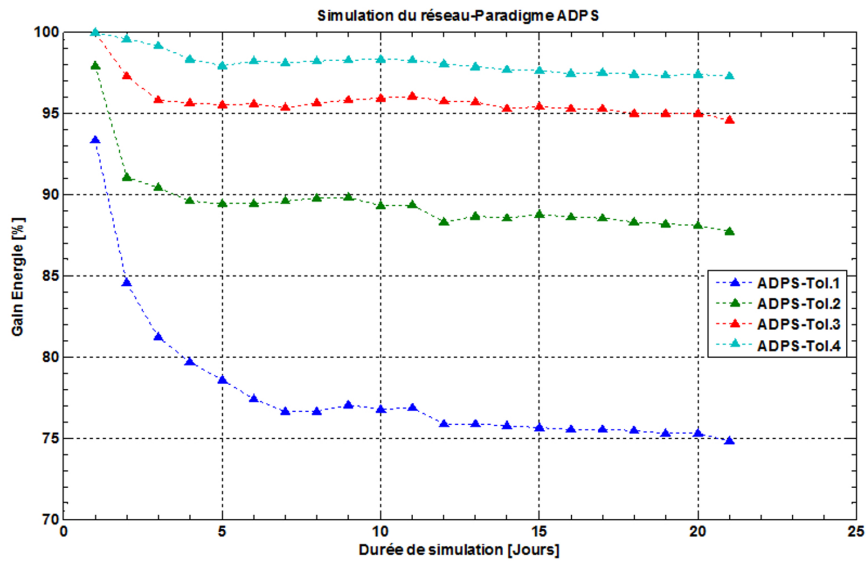


FIGURE 5.8 – Gain d'énergie du réseau à l'aide du paradigme de prédiction ADPS

culièrement pour des tolérances larges. Là encore on constate que le gain en énergie est relativement stable au-delà du 5^{ème} jour de simulation. Cependant, on constate que ces résultats sont moins bons que ceux obtenus à l'aide des algorithmes de prédictions.

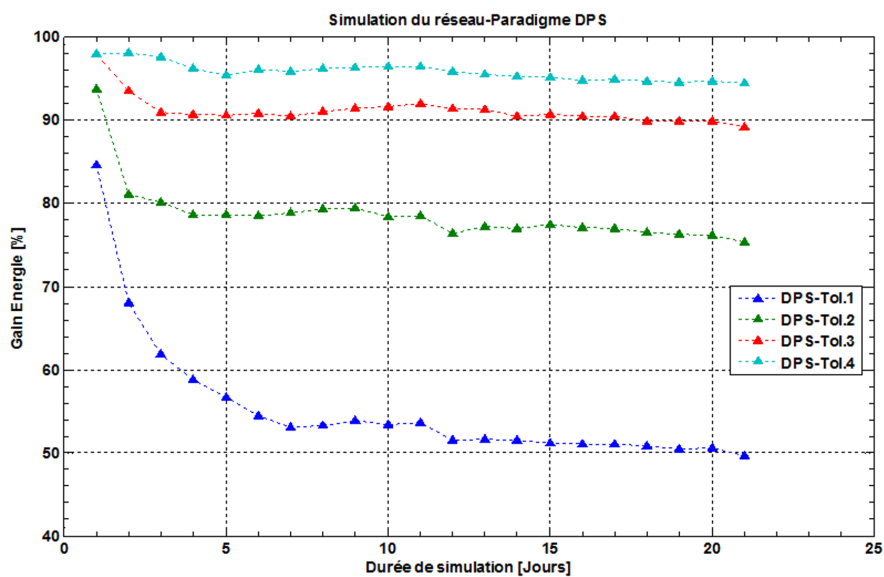


FIGURE 5.9 – Gain d'énergie du réseau à l'aide du paradigme de prédiction DPS

La figure 5.11 montre l'évolution du gain en énergie lorsqu'on combine la prédiction de données à l'aide de l'algorithme ADPS et l'agrégation avec corrélation de données. Dans ce cas les gains en énergie s'améliorent nettement puisque le plus faible niveau reste supérieur à 91%. La figure 5.12 montre l'évolution du gain en énergie lorsqu'on combine le paradigme DPS avec l'agrégation avec corrélation de données. Même si les résultats s'améliorent par rapport à chacune des techniques appliquées séparément, il y a lieu de constater qu'ils restent moins bons que ceux présentés sur la figure 5.11.

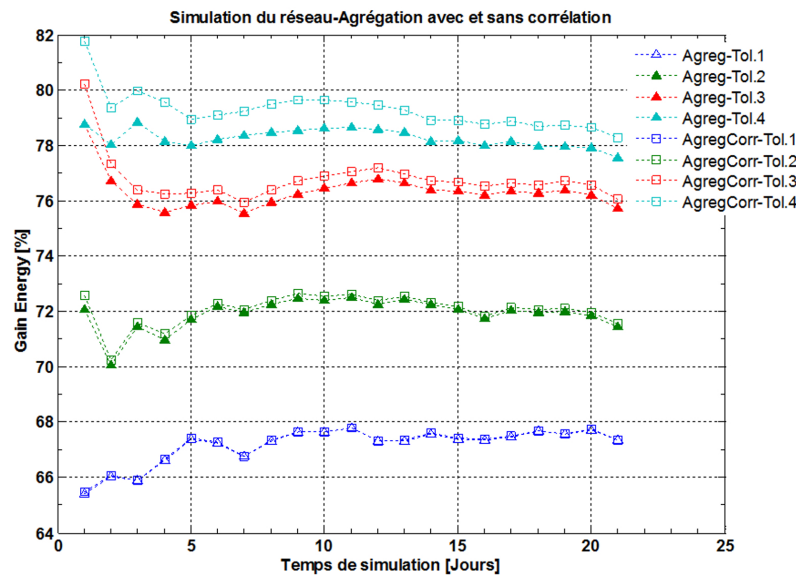


FIGURE 5.10 – Gain d'énergie du réseau avec agrégation de données

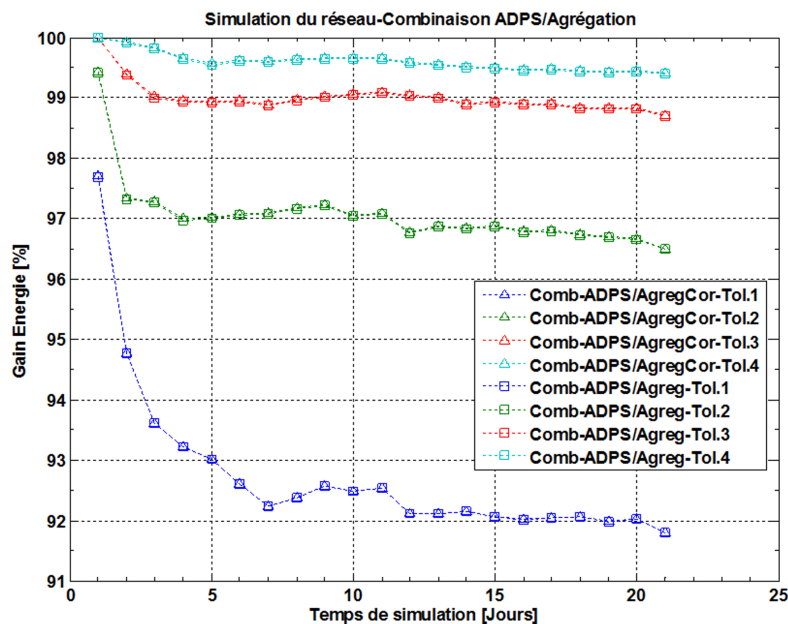


FIGURE 5.11 – Gain d'énergie du réseau à l'aide du paradigme ADPS et l'agrégation avec corrélation de données

La figure 5.13 illustre les résultats obtenus lorsqu'on applique le paradigme EADPS. On rappelle que dans ce paradigme, l'actualisation du prédicteur s'effectue à l'aide des modèles de prédiction obsolètes. Bien que ces résultats soient encourageants, ils restent quelque peu moins bons que ceux atteints à l'aide du paradigme ADPS et cela est évidemment dû à la qualité de l'actualisation du prédicteur. La figure 5.14 traduit les résultats lorsqu'on applique l'agrégation avec et sans corrélation de données lorsque la prédiction de données s'effectue à l'aide d'un prédicteur actualisé selon le paradigme EADPS. Les résultats sont évidemment moins bons que ceux présentés sur la figure 5.11 à cause de

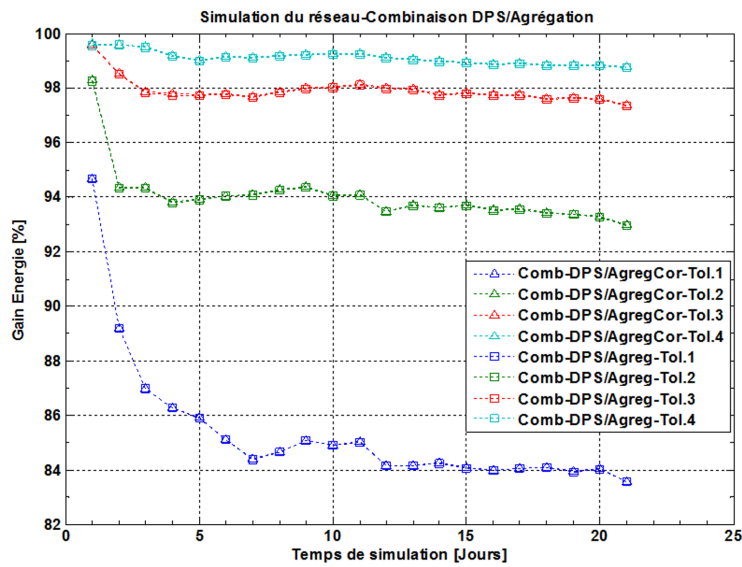


FIGURE 5.12 – Gain d'énergie du réseau à l'aide du paradigme DPS et l'agrégation avec corrélation de données

la qualité du prédicteur. Enfin la figure 5.15 montre l'évolution des résultats lorsqu'on combine le paradigme EADPS avec l'agrégation avec et sans corrélation de données.

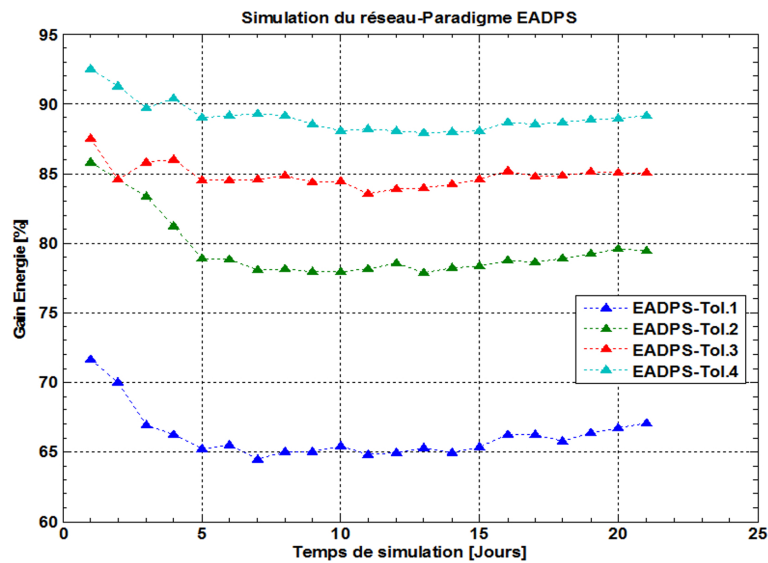


FIGURE 5.13 – Gain d'énergie du réseau à l'aide du paradigme de prédiction EADPS

5.6 Conclusion

Dans ce chapitre nous avons présenté l'architecture et les différents outils utilisés dans l'élaboration de la plateforme de simulation. Une série de simulations a été effectuée dans différents scénarios et avec différentes techniques de prédiction. Les résultats obtenus à

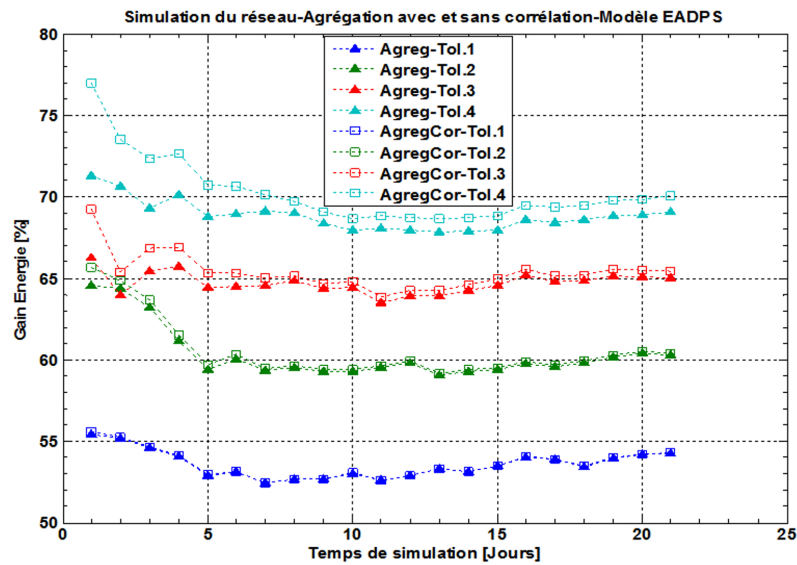


FIGURE 5.14 – Gain d'énergie du réseau à l'aide d'un prédicteur actualisé selon le paradigme EADPS (Agrégation avec et sans corrélation)

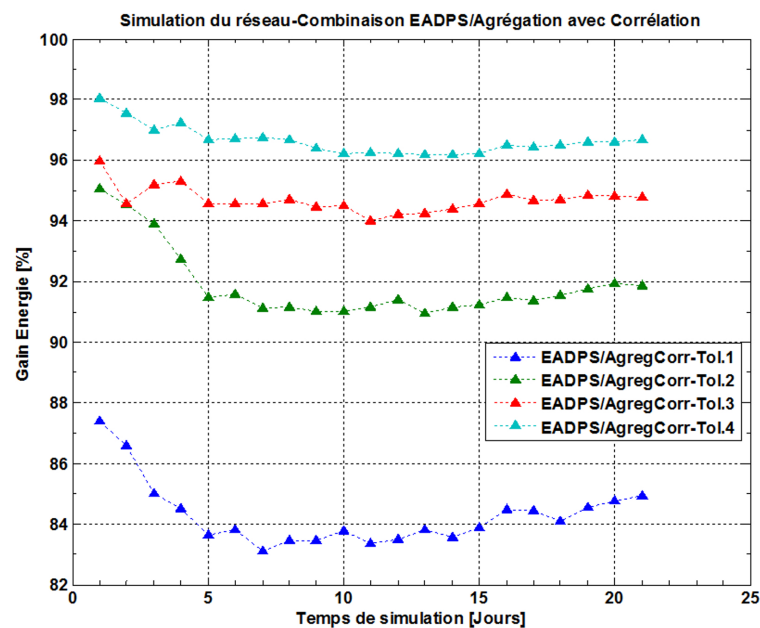


FIGURE 5.15 – Gain d'énergie du réseau à l'aide du paradigme EADPS (Agrégation avec et sans corrélation de données)

travers ces simulations ont justifié les résultats obtenus par analyse mathématique présentés dans le chapitre précédent.

Il ressort que la plateforme développée constitue un outil intéressant pour les utilisateurs désireux de simuler l'intégration entre les RSCFs et le cloud computing afin d'étudier le gain en énergie en faisant appel aux différentes techniques de prédiction et/ou agrégation.

Conclusions et Perspectives

Conclusions et Perspectives

Le travail effectué dans le cadre de cette thèse concerne l'amélioration des performances des réseaux de capteurs sans fil (RCSF). Il s'agit de réduire la consommation d'énergie des batteries des nœuds capteurs, d'améliorer la disponibilité du stockage, d'améliorer l'accès au réseau et de réduire les problèmes de congestion des données, de fournir plus d'informations aux utilisateurs et enfin d'améliorer l'évolutivité attendue à l'avenir. La réduction des taux de transmissions et l'intégration du réseau RCSF au Cloud computing ont constitué deux leviers pour atteindre l'objectif défini par notre cahier des charges.

Après un premier chapitre dédié aux généralités sur les RCSF et le Cloud computing, une étude bibliographique sur le sujet a fait l'objet du second chapitre de cette thèse. Trois différentes conceptions sur l'intégration RCSF-Cloud computing existent aujourd'hui :

- La première s'appuie sur le principe de considérer les RCSF et le Cloud comme deux entités distinctes. Dans ces approches, les travaux de recherche sont orientés sur la manière dont le Cloud interagit avec les RCSF pour améliorer leurs performances. Différentes architectures ont donc été proposées par plusieurs chercheurs.
- Dans la seconde conception, les travaux de recherche se sont focalisés sur la gestion des données récoltées par les RCSF. Les efforts des chercheurs sont dirigés vers le traitement, le stockage et la sécurisation des données.
- La troisième conception repose sur la virtualisation. Il s'agit du principe de partage des ressources d'un RCSF sur une même infrastructure où les utilisateurs peuvent accéder à divers types de capteurs facilement, sans se soucier de leur emplacement physique et pouvant fournir beaucoup de nouveaux services.

Cette étude bibliographique nous a indiqué qu'au fond, quelque soit la conception considérée, la réduction des taux de transmission de données du RCSF reste la voie privilégiée, capable de répondre efficacement à plusieurs problèmes simultanément d'où l'intérêt que nous avons porté aux algorithmes de prédiction de données et au processus d'agrégation. Ceci a constitué le contenu du chapitre (III). Ce sont les algorithmes de prédiction duale (DPS) qui s'avèrent aujourd'hui les plus efficaces et notamment ceux à caractère adaptatif. Dans un premier temps et dans le cadre de la réduction maximale des données collectées par les RCSF, nous avons mené une étude comparative entre les schémas DPS et l'agrégation. Pour réaliser cela, nous avons sélectionné un modèle en anneau pour le RCSF. Ce choix justifié par les avantages de ce modèle nous a permis de mener à bien l'étude comparative désirée. Il ressort que l'agrégation des données est supérieure à l'algorithme de prédiction DPS en matière de réduction des taux de transmission quelque soit la taille du réseau et quelque soit le niveau de précision des modèles. Ce résultat

vient infirmer celui rapporté par le premier chercheur ayant travaillé sur cette question en l'occurrence G.M. Dias et al. [100]. Ayant compris le lest des schémas DPS dû au procédé d'actualisation du modèle de prédiction, nous avons par une contribution originale fait évoluer les schémas DPS ou ADPS (Adaptive Dual Prediction Scheme) selon le cas d'application, vers un nouvel algorithme de prédiction désigné EDPS ou EADPS (Extended Dual Prediction Scheme). Ce fût l'objet du chapitre (IV) où une nouvelle comparaison prédiction-agrégation fût effectuée. Cette nouvelle comparaison a fait ressortir que l'algorithme EADPS est supérieur à l'agrégation dans les réseaux de petite taille notamment lorsque la précision du modèle est bonne. Ensuite, nous avons étudié l'impact de la corrélation des données en combinant tour à tour l'agrégation aux schémas de prédiction ADPS et EADSP. Cette combinaison donnait l'avantage au couple agrégation-ADPS. A partir de là, il devient nécessaire de bien sélectionner la stratégie de réduction du taux de transmission des données en fonction de la taille du RCSF et du type d'algorithme sélectionné. Le chapitre IV a été également consacré a une série de simulations des schémas de prédiction duale et de leur combinaison avec l'agrégation à l'aide cette fois-ci de données réelles concernant les relevés météorologiques collectées au niveau de l'aéroport international Messali Hadj de Tlemcen. Ces simulations ont permis de quantifier les gains pour chacun des cas étudiés. En résumé, il ressort qu'en prédiction, l'algorithme EADPS était supérieur à l'algorithme ADPS et dans leur combinaison avec l'agrégation, la légère supériorité agrégation-ADPS était confirmée. A la lumière de l'ensemble des résultats obtenus, nous avons conçu et réalisé un logiciel de simulation décrivant l'intégration des RCSF au Cloud.

Comme perspectives on peut citer :

1. La conception d'un modèle de transmission RCSF qui prend en charge l'évolution du matériel et donc l'utilisation d'algorithmes de haute complexité.
2. La prise en charge du problème de synchronisation entre les nœuds capteurs et leurs voisins.
3. Les développements des applications de suivi d'objets et de détection d'événements.
4. La prise en compte des problèmes de collisions lors des transmissions des données.

Références Bibliographiques

Bibliographie

- [1] G. Anastasi, M. Conti, M. D. Francesco, and A. Passarella, “Energy conservation in wireless sensor networks : A survey,” *Ad Hoc Networks*, vol. 7, no. 3, pp. 537–568, 2009.
- [2] Q. Zhang, L. Cheng, and R. Boutaba, “Cloud computing : state-of-the-art and research challenges,” *Journal of Internet Services and Applications*, vol. 1, pp. 7–18, 2010.
- [3] E. M. S. O. Serrano, “Cloud framework for wireless sensor networks,” Master’s thesis, Técnico Lisboa, Portugal, July 2014.
- [4] V. Rajesh, J. M. Gnanasekar, R. S. Ponmagal, and P. Anbalagan, “Integration of Wireless Sensor Network with Cloud,” in *Proceedings of IEEE International Conference on Recent Trends in Information, Telecommunication and Computing*, pp. 321–323, March 2010.
- [5] K. Ahmed and M. Gregory, “Integrating Wireless Sensor Networks with Cloud Computing,” in *Proceedings of the 7th IEEE International Conference on Mobile Ad-hoc and Sensor Networks*, pp. 364–366, December 2011.
- [6] P. You, Y. Peng, and H. Gao, “Providing Information Services for Wireless Sensor Networks through Cloud Computing,” in *Proceedings of IEEE Asia-Pacific Conference on Services Computing*, pp. 362–364, December 2012.
- [7] M. Yuriyama and T. Kushida, “Sensor-Cloud Infrastructure - Physical Sensor Management with Virtualized Sensors on Cloud Computing,” in *Proceedings of the 13th IEEE International Conference on Network-Based Information Systems*, pp. 1–8, September 2010.
- [8] A. J. Hussein, A. Riadh, M. Alsultan, and A. Tareq, “Applications and Design for a Cloud of Virtual Sensors,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 2, pp. 229–240, 08 2016.
- [9] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. F. De Rose, and R. Buyya, “Cloud-Sim : A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms,” *Software : Practice & Experience*, vol. 41, p. 23–50, January 2011.
- [10] P. Levis, S. Madden, J. Polastre, R. Szewczyk, K. Whitehouse, A. Woo, D. Gay, J. Hill, M. Welsh, E. Brewer, and D. Culler, *TinyOS : An Operating System for*

- Sensor Networks*, pp. 115–148. Berlin, Heidelberg : Springer Berlin Heidelberg, 2005.
- [11] A. Dunkels, B. Gronvall, and T. Voigt, “Contiki - a lightweight and flexible operating system for tiny networked sensors,” in *Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks*, pp. 455–462, 2004.
- [12] B. Porter and G. Coulson, “Lorien : A Pure Dynamic Component-Based Operating System for Wireless Sensor Networks,” in *Proceedings of the 4th ACM International Workshop on Middleware Tools, Services and Run-Time Support for Sensor Networks (MidSens '09)*, MidSens '09, (New York, NY, USA), p. 7–12, Association for Computing Machinery, 2009.
- [13] Y. Xu, J. Heidemann, and D. Estrin, “Geography-Informed Energy Conservation for Ad Hoc Routing,” in *Proceedings of the 7th Annual ACM International Conference on Mobile Computing and Networking*, MobiCom '01, (New York, NY, USA), p. 70–84, Association for Computing Machinery, 2001.
- [14] Y. Yu, R. Govindan, and D. Estrin, “Geographical and Energy Aware Routing : a recursive data dissemination protocol for wireless sensor networks,” tech. rep., UCLA Computer Department. Technical Report. UCLA-CAD TR-01-0023. 2001, 2001.
- [15] L. Wang and Y. Xiao, “Energy saving mechanisms in sensor networks,” in *Proceedings of the 2nd IEEE International Conference on Broadband Networks*, pp. 724–732 Vol. 1, 2005.
- [16] H. Deng and H. Xu, “Acoustic threatening sound recognition system,” in *SPIE Defence Security Symposium*, vol. 9-13, April 2007.
- [17] Z. Sun, P. Wang, M. C. Vuran, M. A. Al-Rodhaan, A. M. Al-Dhelaan, and I. F. Akyildiz, “BorderSense : Border patrol through advanced wireless sensor networks,” *Ad Hoc Networks*, vol. 9, no. 3, pp. 468–477, 2011.
- [18] B. Hariharan and A. Sasidharan, “iWEDS-An Intelligent Explosive Detection and Terrorist Tracking System Using Wireless Sensor Network,” *International Journal of Computer Science Issues*, vol. 8, pp. 550–554, 07 2011.
- [19] B. Seth and K. Sivalingam, “Wireless sensor node-based design,” in *Proceedings of the 4th IEEE International Conference on Communication Systems and Networks (COMSNETS)*, pp. 1–6, January 2012.
- [20] R. Cardell-Oliver, K. Smettem, M. Kranz, and K. Mayer, “Field testing a wireless sensor network for reactive environmental monitoring [soil moisture measurement],” in *Proceedings of IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing Conference*, pp. 7–12, 2004.
- [21] X. Cao, J. Chen, Y. Zhang, and Y. Sun, “Development of an integrated wireless sensor network micro-environmental monitoring system,” *ISA Transactions*, vol. 47, no. 3, pp. 247–255, 2008.

-
- [22] G. Cho, *Smart Clothing : Technology and Applications*. CRC Press, <https://explore-recourses.stanford.edu>, 2009.
- [23] F. Plastria and M. Elosmani, “On the convergence of the Weiszfeld algorithm for continuous single facility location–allocation problems,” *Top (2008)* 16, p. 388–406, 2008.
- [24] I. Chatzigiannakis, A. Kinalis, and S. Nikolettseas, “Efficient data propagation strategies in wireless sensor networks using a single mobile sink,” *Computer Communications*, vol. 31, no. 5, pp. 896 – 914, 2008. Mobility Management and Wireless Access.
- [25] J. Newsome and D. Song, “GEM : Graph EMbedding for Routing and Data-Centric Storage in Sensor Networks without Geographic Information,” in *Proceedings of the 1st International Conference on Embedded Networked Sensor Systems (SenSys ’03)*, (New York, NY, USA), p. 76–88, Association for Computing Machinery, 2003.
- [26] D. Peter, G. Deepak, and S. Prashant, “TSAR : a two tier sensor storage architecture using interval skip graphs,” in *Proceedings of the 3rd ACM International Conference on Embedded Networked Sensor Systems (SenSys’2005)*, pp. 39–50, November 2005.
- [27] D. Ganesan, B. Greenstein, D. Perelyubskiy, D. Estrin, and J. Heidemann, “An Evaluation of Multi-Resolution Storage for Sensor Networks,” in *Proceedings of the 1st IEEE International Conference on Embedded Networked Sensor Systems, SenSys ’03*, (New York, NY, USA), p. 89–102, Association for Computing Machinery, 2003.
- [28] S. Kalpana, G. M. K., K. Deepak, S. Raja Peeyush Kumar, and P. Vikas Kumar, “A Comparative Study of Various Security Approaches Used in Wireless Sensor Networks,” *International Journal of Advanced Science and Technology*, vol. 17, pp. 31–44, 2010.
- [29] C. Karlof, N. Sastry, and D. Wagner, “TinySec : A Link Layer Security Architecture for Wireless Sensor Networks,” in *Proceedings of the 2nd IEEE International Conference on Embedded Networked Sensor Systems (SenSys’2004)*, SenSys ’04, (New York, NY, USA), p. 162–175, Association for Computing Machinery, 2004.
- [30] P. Adrian, S. Robert, T. J. D., W. Victor, and C. David E., “SPINS : Security Protocols for Sensor Networks,” *Wireless Networks*, vol. 8, p. 521–534, 2002.
- [31] B. Nirupama, E. Deborah, G. Lewis, and H. John, “Scalable Coordination for Wireless Sensor Networks : Self-Configuring Localization Systems,” in *Proceedings of the 6th International Symposium on Communication Theory and Applications (ISCTA ’01)*, pp. 1–6, 2001.
- [32] P. Mell and T. Grance, “The NIST Definition of Cloud Computing,” tech. rep., National Institute of Standards and Technology, Department of Commerce, USA, Special Publication 800-145, 2011.

- [33] W. Kurschl and W. Beer, “Combining Cloud Computing and Wireless Sensor Networks,” in *Proceedings of the 11th International Conference on Information Integration and Web-Based Applications & Services, iiWAS '09*, (New York, NY, USA), p. 512–518, Association for Computing Machinery, 2009.
- [34] P. Zhang, H. Sun, and Z. Yan, “A Novel Architecture Based on Cloud Computing for Wireless Sensor Network,” in *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering, ICCSEE 2013*, pp. 0472–0475, Atlantis Press, Paris, France, 2013.
- [35] T. Wang, Y. Li, G. Wang, J. Cao, M. Z. A. Bhuiyan, and W. Jia, “Sustainable and Efficient Data Collection from WSNs to Cloud,” *IEEE Transactions on Sustainable Computing*, vol. 4, no. 2, pp. 252–262, 2019.
- [36] T. Wang, J. Zeng, Y. Lai, Y. Cai, H. Tian, Y. Chen, and B. Wang, “Data collection from WSNs to the cloud based on mobile Fog elements,” *Future Generation Computer Systems*, vol. 105, pp. 864 – 872, 2020.
- [37] D.-G. Zhang, K. Zheng, T. Zhang, and X. Wang, “A Novel Multicast Routing Method with Minimum Transmission for WSN of Cloud Computing Service,” *Soft Computing*, vol. 19, no. 7, p. 1817–1827, 2015.
- [38] N. Javaid, T. Hafeez, Z. Wadud, N. Alrajeh, M. S. Alabed, and N. Guizani, “Establishing a Cooperation-Based and Void Node Avoiding Energy-Efficient Underwater WSN for a Cloud,” *IEEE Access*, vol. 5, pp. 11582–11593, 2017.
- [39] D. McLoughlin, E. O’Connell, W. Elgenaidi, J. Coleman, and T. Newe, “Review and evaluation of WSN simulation tools in a cloud based environment,” in *Proceedings of the 10th IEEE International Conference on Sensing Technology (ICST)*, pp. 1–6, 2016.
- [40] F. Baza, S. Hammoudi, Z. Aliouat, and M. Aliouat, “Ensuring reliable sensed data in WSN for a reliable sensor cloud,” in *Proceedings of IEEE International Conference on Applied Smart Systems (ICASS)*, pp. 1–5, 2018.
- [41] M. M. Hassan, B. Song, and E.-N. Huh, “A Framework of Sensor-Cloud Integration Opportunities and Challenges,” in *Proceedings of the 3rd ACM International Conference on Ubiquitous Information Management and Communication, ICUIMC '09*, (New York, NY, USA), p. 618–626, Association for Computing Machinery, 2009.
- [42] L. P. Dinesh Kumar, S. Shakena Grace, A. Krishnan, V. M. Manikandan, R. Chinraj, and M. R. Sumalatha, “Data filtering in wireless sensor networks using neural networks for storage in cloud,” in *Proceedings of IEEE International Conference on Recent Trends in Information Technology*, pp. 202–205, April 2012.
- [43] C. Zhu, X. Li, H. Ji, and V. C. M. Leung, “Towards Integration of Wireless Sensor Networks and Cloud Computing,” in *Proceedings of the 7th IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, pp. 491–494, 2015.

-
- [44] R. Piyare, Sun Park, Se Yeong Maeng, Sang Hyeok Park, Seung Chan Oh, Sang Gil Choi, Ho Su Choi, and Seong Ro Lee, "Integrating Wireless Sensor Network into Cloud services for real-time data collection," in *Proceedings of IEEE International Conference on ICT Convergence (ICTC'2013)*, pp. 752–756, October 2013.
- [45] C. Zhu, V. C. M. Leung, H. Wang, W. Chen, and X. Liu, "Providing Desirable Data to Users When Integrating Wireless Sensor Networks with Mobile Cloud," in *Proceedings of the 5th IEEE International Conference on Cloud Computing Technology and Science*, vol. 1, pp. 607–614, December 2013.
- [46] C. Zhu, H. Wang, X. Liu, L. Shu, L. T. Yang, and V. C. M. Leung, "A Novel Sensory Data Processing Framework to Integrate Sensor Networks With Mobile Cloud," *IEEE Systems Journal*, vol. 10, pp. 1125–1136, January 2016.
- [47] A. Alamri, W. S. Ansari, M. M. Hassan, M. S. Hossain, A. Alelaiwi, and M. A. Hossain, "A Survey on Sensor-Cloud : Architecture, Applications, and Approaches," *IEEE Systems Journal*, vol. 2013, pp. Article ID 917923, 18 pages, February 2013.
- [48] R. K. Dwivedi and R. Kumar, "Sensor Cloud : Integrating Wireless Sensor Networks with Cloud Computing," in *Proceedings of 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, pp. 1–6, November 2018.
- [49] R. K. Dwivedi, S. Singh, and R. Kumar, "Integration of Wireless Sensor Networks with Cloud : A Review," in *Proceedings of the 9th IEEE International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pp. 114–119, January 2019.
- [50] P. Langendoerfer, K. Piotrowski, M. Diaz, and B. Rubio, "Distributed Shared Memory as an Approach for Integrating WSNs and Cloud Computing," in *Proceedings of the 5th IEEE International Conference on New Technologies, Mobility and Security (NTMS)*, pp. 1–6, 2012.
- [51] M. S. Aslam, S. Rea, and D. Pesch, "Service Provisioning for the WSN Cloud," in *Proceedings of the 5th IEEE International Conference on Cloud Computing*, pp. 962–969, June 2012.
- [52] S. Rea, M. S. Aslam, and D. Pesch, "Serviceware - A service based management approach for WSN cloud infrastructures," in *Proceedings of IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pp. 133–138, March 2013.
- [53] S. Misra, S. Chatterjee, and M. S. Obaidat, "On Theoretical Modeling of Sensor Cloud : A Paradigm Shift From Wireless Sensor Network," *IEEE Systems Journal*, vol. 11, no. 2, pp. 1084–1093, 2017.
- [54] M. Craciunescu, S. Mocanu, and G. Manea, "Towards Practical Integration of WSN in Cloud Dedicated to Smart Environments," in *Proceeding of the 21st IEEE International Conference on Control Systems and Computer Science (CSCS'2017)*, pp. 447–452, May 2017.

- [55] M. Z. Zhang, L. M. Wang, and S. M. Xiong, "Using Machine Learning Methods to Provision Virtual Sensors in Sensor-Cloud," *Sensors (Basel)*, p. 1836 : 17 pages, April 2020.
- [56] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, vol. 1, p. 281–297, 1967.
- [57] R. Hummen, M. Henze, D. Catrein, and K. Wehrle, "A Cloud design for user-controlled storage and processing of sensor data," in *Proceedings of the 4th IEEE International Conference on Cloud Computing Technology and Science*, pp. 232–240, December 2012.
- [58] F. Banaie and S. A. H. Seno, "A cloud-based architecture for secure and reliable service provisioning in wireless sensor network," in *Proceedings of 4th International Conference on Computer and Knowledge Engineering (ICCKE'2014)*, pp. 96–101, October 2014.
- [59] M. R. Ahmed, H. Cui, and X. Huang, "Smart integration of cloud computing and MCMC based secured WSN to monitor environment," in *Proceedings of the 4th IEEE International Conference on Wireless Communications, Vehicular Technology, Information Theory and Aerospace Electronic Systems (VITAE)*, pp. 1–5, May 2014.
- [60] S. Saha, R. Das, S. Datta, and S. Neogy, "A cloud security framework for a data centric WSN application," in *Proceedings of the 17th International Conference on Distributed Computing and Networking (ICDCN '16)*, pp. 1–6, January 2016.
- [61] H. He, J. Zhang, J. Gu, Y. Hu, and F. Xu, "A fine-grained and lightweight data access control scheme for WSN-integrated cloud computing," *Cluster Computing*, vol. 20, p. 1457–1472, July 2017.
- [62] J. Li, Z. Guan, X. Du, Z. Zhang, and Z. Zhou, "A Low-Latency Secure Data Outsourcing Scheme for Cloud-WSN," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, March 2017.
- [63] T. Wang, Y. Mei, W. Jia, X. Zheng, G. Wang, and M. Xie, "Edge-based differential privacy computing for sensor–cloud systems," *Journal of Parallel and Distributed Computing*, vol. 136, pp. 75 – 85, 2020.
- [64] X. H. Le, S. Lee, P. T. H. Truc, L. The Vinh, A. M. Khattak, M. Han, D. V. Hung, M. M. Hassan, M. Kim, K. Koo, Y. Lee, and E. Huh, "Secured WSN-integrated cloud computing for u-Life Care," in *Proceedings of the 7th IEEE Consumer Communications and Networking Conference*, pp. 1–2, June 2010.
- [65] G. Fortino, M. Pathan, and G. Di Fatta, "BodyCloud : Integration of Cloud Computing and body sensor networks," in *Proceedings of the 4th IEEE International Conference on Cloud Computing Technology and Science*, pp. 851–856, December 2012.

-
- [66] M. J. Gueguez, S. Rekhis, and N. Boudriga, "A sensor cloud architecture for healthcare applications," in *Proceedings of the 31st Annual ACM Symposium on Applied Computing (SAC '16)*, p. 612–617, April 2016.
- [67] M. Muzammal, R. Talat, A. H. Sodhro, and S. Pirbhulal, "A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks," *Information Fusion*, vol. 53, pp. 155–164, 2020.
- [68] J. Cen, T. Yu, Z. Li, S. Jin, and S. Liu, "Developing a disaster surveillance system based on wireless sensor network and cloud platform," in *Proceedings of IET International Conference on Communication Technology and Application (ICCTA 2011)*, pp. 757–761, October 2011.
- [69] E. Kanagaraj, L. M. Kamarudin, A. Zakaria, R. Gunasagaran, and A. Y. M. Shakkaff, "Cloud-based remote environmental monitoring system with distributed WSN weather stations," in *Proceedings of IEEE International Conference on SENSORS*, pp. 1–4, November 2015.
- [70] D. S. Arjun, A. Bala, Dwarakanath V, Sampada K S, Prahlada Rao B B, and H. Pasupuleti, "Integrating cloud-WSN to analyze weather data and notify SaaS user alerts during weather disasters," in *Proceedings of IEEE International Advance Computing Conference (IACC)*, pp. 899–904, 2015.
- [71] C. Pallavi and V. Chandrakala, "Development of Flood Monitoring System using WSN and IoT based on Cloud," *International Research Journal of Engineering and Technology*, vol. 4, no. 5, pp. 2937–2943, 2017.
- [72] M. Chaitra and B. Sivakumar, "Cloud Based Disaster Detection & Management System using WSN," *International Research Journal of Engineering and Technology*, vol. 4, no. 5, pp. 1162–1167, 2017.
- [73] S. M. Chari and B. Sivakumar, "Development of Smart Network Using WSN and IoT for Precision Agriculture Monitoring System on Cloud," *International Research Journal of Engineering and Technology*, vol. 4, no. 5, pp. 1502–1505, 2017.
- [74] M. R. M. Kassim and A. N. Harun, "Wireless sensor networks and cloud computing integrated architecture for agricultural environment applications," in *Proceedings of the 11th IEEE International Conference on Sensing Technology (ICST)*, pp. 1–5, 2017.
- [75] L. Mhatre and N. Rai, "Integration between wireless sensor and cloud," in *Proceedings of IEEE International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pp. 779–782, 2017.
- [76] J. M. Bindu, Nikitha, M. Namitha, and H. S. Pradeep, "WSN based Online Parameter Monitoring in Cold Storage Warehouses in Cloud using IOT concepts," *International Research Journal of Engineering and Technology*, vol. 5, no. 7, pp. 767–773, 2018.

- [77] C. Zhu, L. Shu, V. C. M. Leung, S. Guo, Y. Zhang, and L. T. Yang, "Secure Multimedia Big Data in Trust-Assisted Sensor-Cloud for Smart City," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 24–30, 2017.
- [78] F. H. Bijarbooneh, W. Du, E. C. . Ngai, X. Fu, and J. Liu, "Cloud-Assisted Data Fusion and Sensor Selection for Internet of Things," *IEEE Internet of Things Journal*, vol. 3, no. 3, pp. 257–268, 2016.
- [79] G. M. Dias, C. B. Margi, F. C. P. de Oliveira, and B. Bellalta, "Cloud-Empowered, Self-Managing Wireless Sensor Networks : Interconnecting Management Operations at the Application Layer," *IEEE Consumer Electronics Magazine*, vol. 8, no. 1, pp. 55–60, 2019.
- [80] I. Chukwuemeka Chimsom and M. K. Habib, "Design of a Two-Tier WSN-based IoT Surveillance System with Cloud Integration," in *Proceedings of the 20th IEEE International Conference on Research and Education in Mechatronics (REM)*, pp. 1–7, 2019.
- [81] J. Liang, M. Zhang, and V. C. M. Leung, "A Reliable Trust Computing Mechanism Based on Multisource Feedback and Fog Computing in Social Sensor Cloud," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5481–5490, 2020.
- [82] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, vol. 2, pp. 1–10, January 2000.
- [83] O. Younis and S. Fahmy, "HEED : a Hybrid, Energy-Efficient, Distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on mobile computing*, vol. 3, no. 4, pp. 366–379, 2004.
- [84] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed diffusion : a scalable and robust communication paradigm for sensor networks," in *Proceedings of the 6th IEEE Annual International Conference on Mobile Computing and Networking (MobiCom'00)*, 2000.
- [85] J. Cui, K. Boussetta, and F. Valois, "Classification of data aggregation functions in wireless sensor networks," *Computer Networks*, vol. 178, p. 107342, 2020.
- [86] C. Liu, K. Wu, and M. Tsao, "Energy efficient information collection with the ARIMA model in wireless sensor networks," in *Proceedings of IEEE International Conference on Global Telecommunications (GLOBECOM '05)*, vol. 5, pp. 5 pp.–2474, 2005.
- [87] J. Lu, F. Valois, M. Dohler, and M. Wu, "Optimized Data Aggregation in WSNs Using Adaptive ARMA," in *Proceedings of the 4th IEEE International Conference on Sensor Technologies and Applications*, pp. 115–120, 2010.
- [88] N. Kimura and S. Latifi, "A survey on data compression in wireless sensor networks," in *Proceedings of IEEE International Conference on Information Technology : Coding and Computing (ITCC'05) - Volume II*, vol. 2, pp. 8–13 Vol. 2, 2005.

-
- [89] T. Srisooksai, K. Keamarungsi, P. Lamsrichan, and K. Araki, "Practical data compression in wireless sensor networks : A survey," *Journal of Network and Computer Applications*, vol. 35, no. 1, pp. 37–59, 2012. Collaborative Computing and Applications.
- [90] F. A. Aderohunmu, G. Paci, D. Brunelli, J. D. Deng, and L. Benini, "Prolonging the lifetime of wireless sensor networks using light-weight forecasting algorithms," in *Proceedings of the 8th IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, pp. 461–466, 2013.
- [91] A. Jain, E. Y. Chang, and Y.-F. Wang, "Adaptive Stream Resource Management Using Kalman Filters," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, SIGMOD '04, (New York, NY, USA), p. 11–22, Association for Computing Machinery, 2004.
- [92] S. Silvia and R. Kay, "An adaptive strategy for quality-based data reduction in wireless sensor networks," in *Proceedings of the 3rd IEEE International Conference on networked sensing systems (INSS 2006)*, pp. 29–36, 2006.
- [93] D. Tulone and S. Madden, "PAQ : Time Series Forecasting for Approximate Query Answering in Sensor Networks," in *European Workshop on Wireless Sensor Networks* (K. Römer, H. Karl, and F. Mattern, eds.), (Berlin, Heidelberg), pp. 21–37, Springer Berlin Heidelberg, 2006.
- [94] C. Yang, R. Cardell-Oliver, and C. McDonald, "Combining temporal and spatial data suppression for accuracy and efficiency," in *Proceedings of the 7th IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, pp. 347–352, 2011.
- [95] H. Jiang, S. Jin, and C. Wang, "Prediction or Not ? An Energy-Efficient Framework for Clustering-Based Data Collection in Wireless Sensor Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 6, pp. 1064–1071, 2011.
- [96] G. M. Dias, B. Bellalta, and S. Oechsner, "A Survey About Prediction-Based Data Reduction in Wireless Sensor Networks," *ACM Computing Surveys*, vol. 49, pp. Paper n°58, 35 pages, November 2016.
- [97] K. Langendoen and A. Meier, "Analyzing MAC Protocols for Low Data-Rate Applications," *ACM Transactions on Sensor Networks*, vol. 7, p. Article n°19 (40 pages), Septembre 2010.
- [98] N. C. Brent, J. C. Charles, and S. J. David, "Unit disk graphs," *Discrete Mathematics*, vol. 86, no. 1-3, pp. 165–177, 1990.
- [99] A. Deshpande, C. Guestrin, S. R. Madden, J. M. Hellerstein, and W. Hong, "Model-Driven Data Acquisition in Sensor Networks," in *Proceedings of the 30th ACM International Conference on Very Large Data Bases, Volume 30, VLDB '04*, p. 588–599, VLDB Endowment, 2004.

- [100] G. M. Dias, B. Bellalta, and S. Oechsner, “The impact of dual prediction schemes on the reduction of the number of transmissions in sensor networks,” *Computer Communications*, vol. 112, pp. 58 – 72, 2017.
- [101] H. Timm Neil, *Applied multivariate analysis*. Springer, <http://link.springer.com/content/pdf/10.1007/b98963.pdf>, 2002.
- [102] G. M. Dias, T. Adame, B. Bellalta, and S. Oechsner, “A self-managed architecture for sensor networks based on real time data analysis,” in *Proceedings of IEEE International Conference on Future Technologies (FTC)*, pp. 1297–1299, 2016.
- [103] I. Lazaridis and S. Mehrotra, “Capturing sensor-generated time series with quality guarantees,” in *Proceedings the 19th IEEE International Conference on Data Engineering (Cat. No.03CH37405)*, pp. 429–440, 2003.
- [104] D. J. McCorrie, E. Gaura, K. Burnham, N. Poole, and R. Hazelden, *Predictive Data Reduction in Wireless Sensor Networks Using Selective Filtering for Engine Monitoring*, pp. 129–148. New York, NY : Springer New York, 2015.
- [105] S. Biljana, S. Dimitar, and D. Danco, “Data Prediction in WSN using Variable Step Size LMS Algorithm,” in *Proceedings of 5th IEEE International Conference on Sensor Technologies and Applications (SENSORCOMM 2011)*, pp. 191–196, 2011.
- [106] M. Wu, L. Tan, and N. Xiong, “Data prediction, compression, and recovery in clustered wireless sensor networks for environmental monitoring applications,” *Information Sciences*, vol. 329, pp. 800–818, 2016. Special issue on Discovery Science.
- [107] Y.-A. Le Borgne, S. Santini, and G. Bontempi, “Adaptive model selection for time series prediction in wireless sensor networks,” *Signal Processing*, vol. 87, no. 12, pp. 3010–3020, 2007. Special Section : Information Processing and Data Management in Wireless Sensor Networks.
- [108] R. Askari Moghadam and M. Keshmirpour, “Hybrid ARIMA and Neural Network Model for Measurement Estimation in Energy-Efficient Wireless Sensor Networks,” in *International Conference on Informatics Engineering and Information Science (A. Abd Manaf, S. Sahibuddin, R. Ahmad, S. Mohd Daud, and E. El-Qawasmeh, eds.)*, (Berlin, Heidelberg), pp. 35–48, Springer Berlin Heidelberg, 2011.
- [109] J. Kho, A. Rogers, and N. R. Jennings, “Decentralized Control of Adaptive Sampling in Wireless Sensor Networks,” *ACM Transactions on Sensor Networks*, vol. 5, p. Article n° 19 (35 pages), june 2009.
- [110] Y. Chen, Z. Wang, Z. Zhao, D. Li, and L. Cui, “Online model-driven data acquisition for wireless sensor networks,” in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1572–1577, 2015.
- [111] S. Cheng, Z. Cai, J. Li, and X. Fang, “Drawing dominant dataset from big sensory data in wireless sensor networks,” in *Proceedings of IEEE International Conference on Computer Communications (INFOCOM)*, pp. 531–539, 2015.

-
- [112] S. Goel and T. Imielinski, "Prediction-Based Monitoring in Sensor Networks : Taking Lessons from MPEG," *ACM SIGCOMM Computer Communication Review*, vol. 31, p. 82–98, October 2001.
- [113] J. S. Armstrong, *Principles of Forecasting : A Handbook for Researchers and Practitioners*. Springer Science & Business Media, <https://www.gwern.net/docs/predictions/2001-principlesforecasting.pdf>, 2001.
- [114] G. Alan and B. Frank, "Comparison of Methods for the Computation of Multivariate t Probabilities," *Journal of Computational and Graphical Statistics*, vol. 11, no. 4, pp. 950–971, 2002.
- [115] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*, pp. 977–979. Berlin, Heidelberg : Springer Berlin Heidelberg, 2011.
- [116] H. Liazid, M. Lehsaini, and A. Liazid, "An Improved Adaptive Dual Prediction Scheme for Reducing Data Transmission in Wireless Sensor Networks," *Wireless Networks*, vol. 25, no. 6, p. 3545–3555, 2019.
- [117] B. Karp and H. T. Kung, "GPSR : Greedy Perimeter Stateless Routing for Wireless Networks," in *Proceedings of the 6th IEEE Annual International Conference on Mobile Computing and Networking*, MobiCom '00, (New York, NY, USA), pp. 243–254, 2000.
- [118] F. A. Aderohunmu, G. Paci, D. Brunelli, J. D. Deng, L. Benini, and M. Purvis, "An Application-Specific Forecasting Algorithm for Extending WSN Lifetime," in *Proceedings of IEEE International Conference on Distributed Computing in Sensor Systems*, pp. 374–381, 2013.
- [119] S. Samarah, "A Data Predication Model for Integrating Wireless Sensor Networks and Cloud Computing," *Procedia Computer Science*, vol. 52, pp. 1141 – 1146, 2015. The 6th International Conference on Ambient Systems, Networks and Technologies (ANT-2015), the 5th International Conference on Sustainable Energy Information Technology (SEIT-2015).
- [120] B. George E. P., J. Gwilym M., R. Gregory C., and L. Greta M., *Time Series Analysis : Forecasting and Control*. Wiley–Blackwell; 5th Edition, 2015.
- [121] S. Mohanty, *Nanoelectronic Mixed-Signal System Design*. McGraw-Hill Professional, ISBN-13 : 978-0071825719, 2015.
- [122] M. L. Rajaram, "Comparative Analysis and Implementation of High Data Rate Wireless Sensor Network," Master's thesis, University of North Texas, USA, Denton, Texas, 2015.
- [123] F. Wang and J. Liu, "Networked Wireless Sensor Data Collection : Issues, Challenges, and Approaches," *IEEE Communications Surveys Tutorials*, vol. 13, no. 4, pp. 673–687, 2011.

BIBLIOGRAPHIE

- [124] J. Polastre, R. Szewczyk, and D. Culler, “Telos : enabling ultra-low power wireless research,” in *Proceedings of the 4th IEEE International Symposium on Information Processing in Sensor Networks (IPSN 2005)*, pp. 364–369, 2005.
- [125] L. DEKHICI, *Programmation Orientée Objet (POO)*. Cours en ligne, University of Sciences and Technology of Oran, Algeria, 2019.

Résumé

Cette thèse traite l'intégration des réseaux de capteurs sans fil (RCSF) au Cloud Computing. L'objectif étant de résoudre les problèmes intrinsèques que posent ces réseaux en termes de stockage de données, d'énergie limitée des batteries de capteurs et d'accès au réseau (congestion). En fait, il s'agit d'améliorer les performances des RCSFs pour une plus grande durée de vie et une plus large exploitation. Comme les transmissions de données sont à la base des problèmes évoqués, nous avons commencé dans un premier temps par étudier l'impact de leur réduction à l'aide des techniques de prédiction duale et d'agrégation de données. Ainsi, nous avons mis au point un nouvel algorithme de prédiction, appelé EADPS (Extended Dual Prediction Scheme). Ensuite, nous avons comparé l'agrégation des données à chacun des schémas ADPS (Adaptif Dual Prediction Scheme) et EADPS. Il ressort que l'agrégation est une technique bien meilleure que le schéma ADPS pour les réseaux de petite taille et quel que soit la précision des prédictions. Cependant, elle ne devient sérieusement concurrente au schéma EADPS que pour de RCSFs de grandes tailles.

Nous avons effectué différentes simulations sur des données réelles à l'aide de la prédiction seule, de l'agrégation seule et ensuite de leur combinaison. Il ressort qu'avec la combinaison de ces deux techniques nous avons maximisé le taux de réduction des transmissions. Partant de ces résultats encourageants, nous avons conçu et réalisé une plateforme de simulation pour l'intégration des RCSFs au Cloud Computing.

Mots clés : Réseaux de capteurs Sans Fil, Cloud Computing, Prédiction, Prévision, Agrégation de données.

Abstract

This thesis deals with the integration of Wireless Sensor Networks (WSNs) in Cloud Computing. The goal is to solve the intrinsic problems posed by these networks in terms of data storage, limited energy of the batteries of sensors and access to the network (congestion). In fact, it is about improving the performance of WSNs for longer network lifetime and a large number of applications. Since data transmissions are the basis of the problems mentioned, we first started by studying the impact of their reduction using dual prediction and data aggregation techniques. Thus, we have developed a new prediction algorithm, called EADPS (Extended Dual Prediction Scheme). Next, we compared the data aggregation to each of ADPS (Adaptive Dual Prediction Scheme) and EADPS. It turns out that aggregation is a much better technique than the ADPS scheme for small networks and regardless of the accuracy of the predictions. However, it only becomes seriously competitive with the EADPS scheme for large sizes of WSNs.

We performed different simulations on real data using prediction alone, aggregation alone and then their combination. It appears that with the combination of these two techniques we have maximized the rate of reduction of transmissions. Based on these encouraging results, we designed and produced simulation software for the integration of RCSF into Cloud Computing.

Keywords : Wireless sensor networks, Cloud Computing, Prediction, Forecasting, Aggregation.

ملخص

تناول هذه الأطروحة تكامل شبكات الاستشعار اللاسلكية (WSNs) في الحوسبة السحابية. الهدف هو حل المشكلات الجوهرية التي تطرحها هذه الشبكات من حيث تخزين البيانات والطاقة المحدودة من بطاريات أجهزة الاستشعار والوصول إلى الشبكة (الازدحام). في الواقع، يتعلق الأمر بتحسين أداء شبكات الاستشعار اللاسلكية لحياة أطول وتشغيل أوسع. نظرًا لأن عمليات نقل البيانات هي أصل المشكلات المذكورة، فقد بدأنا أولاً بدراسة تأثير تقليلها باستخدام تقنيات التنبؤ المزدوج وتجميع البيانات. وبالتالي، قمنا بتطوير خوارزمية توقع جديدة تسمى (EADPS) مخطط التنبؤ المزدوج الموسع. بعد ذلك، قمنا بمقارنة تجميع البيانات بكل من مخطط التنبؤ المزدوج التكيفي (ADPS) و EADPS. اتضح أن التجميع هو أسلوب أفضل بكثير من مخطط ADPS للشبكات الصغيرة وبغض النظر عن دقة التنبؤات. ومع ذلك، فإنها تصبح منافسة بشكل جدي فقط مع مخطط EADPS للأحجام الكبيرة من شبكات.

أجرينا عمليات محاكاة مختلفة على بيانات حقيقية باستخدام التنبؤ وحده، والتجميع وحده ثم الجمع بينهما. يبدو أنه من خلال الجمع بين هاتين التقنيتين، قمنا بتعظيم معدل تقليل عمليات النقل. بناءً على هذه النتائج المشجعة، قمنا بتصميم وإنتاج أرضية محاكاة لدمج شبكات الاستشعار اللاسلكية في الحوسبة السحابية.

الكلمات المفتاحية: شبكات الاستشعار اللاسلكية، الحوسبة السحابية، التنبؤ، التوقعات، تجميع البيانات.