



HAL
open science

Validation de données débitmétriques issues de réseaux d'assainissement

Farida Hamioud

► **To cite this version:**

Farida Hamioud. Validation de données débitmétriques issues de réseaux d'assainissement. Autre. Institut National Polytechnique de Lorraine, 2007. Français. NNT : 2007INPL007N . tel-01752944

HAL Id: tel-01752944

<https://hal.univ-lorraine.fr/tel-01752944>

Submitted on 29 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>



Validation de Données Débitmétriques Issues de Réseaux d'Assainissement

THÈSE

Soutenue le 25 Janvier 2007

pour l'obtention du

Doctorat de l'Institut National Polytechnique de Lorraine

(spécialité automatique et traitement du signal)

par

Farida HAMIPOUD

Composition du jury

<i>Président :</i>	M. ROBERT	Professeur à l'ESSTIN
<i>Rapporteurs :</i>	B. OULD-BOUAMAMA	Professeur à l'Université des Sciences et Technologies de Lille
	P. CHATELLIER	Directeur de Recherche LCPC
<i>Examineurs :</i>	C. JOANNIS	Ingénieur LCPC
	J. RAGOT	Professeur à l'INPL

*A la mémoire de ma chère mère,
à mon cher père,
à mon cher mari,
je dédie ce travail.*

Remerciements

Les travaux présentés dans ce mémoire ont été réalisés au Laboratoire Central des Ponts et Chaussées et au Centre de recherche en Automatique de Nancy. Ces travaux ont été dirigés par Monsieur J. RAGOT.

Je voudrais tout d'abord exprimer toute ma reconnaissance à Monsieur M. ROBERT pour avoir accepté la présidence du jury de thèse.

Je tiens à remercier Monsieur B. OULD-BOUAMAMA et Monsieur P. CHATELLIER pour l'attention qu'ils ont portée à ce travail en me faisant l'honneur d'être rapporteurs.

Je voudrais témoigner toute ma gratitude à Monsieur Claude JOANNIS pour sa disponibilité et sa bienveillance constantes.

Je remercie vivement Monsieur José RAGOT pour avoir assuré la tâche de directeur de thèse et dont les nombreux conseils scientifiques ont permis de guider cette étude.

Je remercie également Monsieur Michel AUMOND du LCPC qui n'a cessé de soutenir et valoriser ce travail, par les moyens expérimentaux.

Je tiens à remercier tous les membres de l'équipe eau et environnement qui m'ont soutenu par leurs conseils et remarques.

Table des matières

Introduction	1
1 Les réseaux de mesure en assainissement	7
1.1 Introduction.....	8
1.2 Objectifs d'obtention de mesures en réseau d'assainissement.....	8
1.2.1 Objectifs liés aux questions techniques.....	8
1.2.2 Objectifs liés aux questions réglementaires et juridiques.....	12
1.2.3 Objectifs liés aux questions financières.....	15
1.3 Conception du réseau de mesure.....	16
1.3.1 Grandeurs mesurées.....	16
1.3.2 Position des points de mesure sur le système d'assainissement.....	16
1.3.3 Stratégie d'acquisition.....	20
1.3.4 Fréquence de transmission.....	21
1.3.5 Stratégie d'échantillonnage.....	21
1.3.6 Transmission et gestion des données.....	22
1.4 Méthodes et technologie pour l'évaluation du débit.....	24
1.4.1 Mesure de débit.....	24
1.4.2 Mesure de la hauteur.....	25
1.4.3 Mesure de la vitesse.....	26
1.5 Anomalies affectant les mesures.....	27
1.5.1 Anomalies de fonctionnement.....	27
1.5.2 Anomalies sur le processus.....	29
1.6 Données.....	32
1.6.1 Réseau de mesure de Nantes.....	32
1.6.2 Acquisition et gestion des données sur le système de Nantes.....	33
1.6.3 Données disponibles.....	33
1.6.4 Données utilisées.....	34
1.7 Conclusion.....	34
2 Méthodes de validation des données de mesure	37
2.1 Introduction	38
2.2 Techniques de validation des données	39
2.3 Méthode de validation des données de mesure à base de modèles analytiques	46
2.4 Généralités sur le diagnostic des systèmes.....	47
2.5 Conclusion.....	50

3 Modèles	51
3.1 Introduction.....	52
3.2 Présentation des données utilisées.....	53
3.3 Prétraitement des données.....	56
3.4 Analyse des données	57
3.5 Etablissement des modèles.....	59
3.6 Structure des modèles.....	62
3.6.1 Choix de la structure du modèle.....	64
3.6.2 Sélection des variables explicatives	65
3.7 Identification paramétrique par la méthode des moindres carrés ordinaire.....	70
3.8 Validation des modèles	70
3.8.1 Evaluation de l'écart type de l'erreur et du coefficient de détermination.....	71
3.9 Conclusion.....	76
4 Détection et localisation des défauts de mesures	79
4.1 Introduction.....	80
4.2 Génération de résidus.....	82
4.3 Caractéristiques des résidus.....	85
4.3.1 Caractéristiques théoriques des résidus.....	85
4.3.2 Caractéristiques expérimentales des résidus.....	85
4.3.3 Calcul des tolérances associées aux résidus.....	86
4.4 Matrice des signatures des défauts	86
4.5 Procédure de détection des défauts	89
4.5.1 Choix des paramètres pour la procédure de détection.....	91
4.5.2 Application de la procédure de détection.....	99
4.6 Localisation des défauts.....	101
4.6.1 Localisation des défauts basée sur le calcul de distance.....	101
4.6.2 Localisation de défauts par la méthode DMP.....	104
4.6.3 Localisation de défauts basée sur les concepts de la logique floue.....	108
4.7 Comparaison des trois procédures de détection et de localisation considérées	121
4.8 Conclusion.....	122
Conclusion	123
Références	127
Annexe A Points de mesures	135
Annexe B Prétraitement des données	139

Annexe C	Résultats de l'application de la méthode OLS	145
Annexe D	Validation croisée des modèles	147
Annexe E	Signature des défauts théorique	171

Liste des figures

Figure 1.1. Schéma d'un système d'assainissement.....	18
Figure 1.2. Schéma d'un système de mesure	18
Figure 1.3. Synoptique d'acquisition et de gestion des données.....	23
Figure 1.4. Vitesse en fonction du temps fournie par un capteur en état d'ensablement.....	28
Figure 1.5. Vitesse en fonction du temps fournie par un capteur en état d'encrassement temporaire (bouchage).....	28
Figure 1.6. Schéma du système d'assainissement de Nantes	32
Figure 2.1. Scéma descriptif des étapes nécessaire pour la validation.....	47
Figure 3.1. Emplacement des mesures pour le réseau d'assainissement, h_i hauteur d'eau, v_j vitesse d'écoulement d'eau et Q_k débit d'eau, avec i, j, k' numéro du capteur $\in \{1, 2, \dots, 40\}$	54
Figure 3.2. Allure des signaux hauteur, vitesse, débit et pluie dans le réseau sur une période d'une semaine.....	55
Figure 3.3.a. Intercorrélation entre les variables v_5 et h_4 , $\Delta_{v_5 h_4} = 0 \text{ min}$	60
Figure 3.3.b. Intercorrélation entre les variables h_{18} et h_2 , $\Delta_{h_{18} h_2} = 30 \text{ min}$	60
Figure 3.4.a. Schéma de principe du modèle hydrologique	63
Figure 3.4.b. Schéma de principe du modèle hydraulique	63
Figure 3.5. Critère d'arrêt AIC en fonction des variables sélectionnées	68
Figure 3.6.a. Signal h_1 en fonction du temps.....	73
Figure 3.6.b. Signal h_4 en fonction du temps	73
Figure 3.7. Identification du modèle sur période 1 et validation sur période 2.....	74
Figure 3.8. Identification du modèle sur période 2 et validation sur période 1.....	74
Figure 3.9. La mesure h_4 et son estimée \hat{h}_4 en superposition en fonction du temps	75
Figure 4.1 Principe général de diagnostic à base de modèles analytiques	80
Figure 4.2. Ecart-type σ_r en fonction du nombre des échantillons du résidu r_4	91
Figure 4.3. Allure du taux de variation des défauts dépendants de la mesure	94
Figure 4.4. Tracé de J	95
Figure 4.5. Tracé de U_n	96
Figure 4.6. Tracé de $U_n - m_n$ et λ	96
Figure 4.8. Fonction sigmoïde générale utilisée dans DMP. $\rho = 2, 10$	105
Figure 4.9. Facteurs de vraisemblance des hypothèses de défaillance.....	107
Figure 4.10. Fonction sigmoïde générale utilisée dans la localisation à base de concepts de la logique floue, $\rho = 4, 12$	108
Figure 4.11. Degrés de véracité des différentes hypothèses, opérateur d'agrégation « minimum ».	112
Figure 4.12. Degrés de véracité des différentes hypothèses, opérateur d'agrégation selon Mandel	113

Figure 4.13. Degrés de véracité des différentes hypothèses, nouvel opérateur d'agrégation.....	113
Figure 4.14. Degrés de véracité de chaque hypothèse sans prise en considération de la persistance de défaut.....	118
Figure 4.15. Degrés de véracité de chaque hypothèse avec prise en considération de la persistance de défaut.....	119
Figure 4.16. Degrés de véracité des différentes hypothèses concernant la présence des défauts	120
.....	
Figure B.1. Interpolation constante-avant	141
Figure.B.2. Signaux bruts et signaux rééchantillonnés, pas de temps d'une minute.	142
.....	
Figure B.3. Signaux filtrés avec un filtre "trim" et puis un filtre "passe-bas".	143
Figure D.1. Signaux $h_1, h_2, h_4, v_5, h_7, h_{10}, h_{11}, h_{12}$ en fonction du temps.	148
Figure D.2. Signaux $h_{14}, h_{15}, h_{18}, h_{20}, h_{22}, h_{37}, h_{40}$ en fonction du temps.	149
Figure D.3. Signaux débits.	150
Figure D.4. Identification du modèle v_5 sur période 1 et validation sur période 2.	151
Figure D.5. La mesure v_5 et son estimée \hat{v}_5 en superposition, données en fonction du temps.	151
Figure D.6. Identification du modèle h_7 sur période 1 et validation sur période 2.	152
Figure D.7. La mesure h_7 et son estimée \hat{h}_7 en superposition, données en fonction du temps.	152
Figure D.8. Identification du modèle h_{10} sur période 1 et validation sur période 2.	153
Figure D.9. La mesure h_{10} et son estimée \hat{h}_{10} en superposition, données en fonction du temps.	153
Figure D.10. Identification du modèle h_{12} sur période 1 et validation sur période 2.	154
.....	
Figure D.11. La mesure h_{12} et son estimée \hat{h}_{12} en superposition, données en fonction du temps.	154
Figure D.12. Identification du modèle h_{15} sur période 1 et validation sur période 2.	155
.....	
Figure D.13. La mesure h_{15} et son estimée \hat{h}_{15} en superposition, données en fonction du temps.	155
Figure D.14. Identification du modèle h_{18} sur période 1 et validation sur période 2.	156
.....	
Figure D.15. La mesure h_{18} et son estimée \hat{h}_{18} en superposition, données en fonction du temps.	156
Figure D.16. Identification du modèle h_{20} sur période 1 et validation sur période 2.	157
.....	
Figure D.17. La mesure h_{20} et son estimée \hat{h}_{20} en superposition, données en fonction du temps.	157
Figure D.18. Identification du modèle h_{22} sur période 1 et validation sur période 2.	158
.....	

Figure D.19. La mesure h_{22} et son estimée \hat{h}_{22} en superposition, données en fonction du temps.	158
Figure D.20. Identification du modèle h_{37} sur période 1 et validation sur période 2.	159
Figure D.21. La mesure h_{37} et son estimée \hat{h}_{37} en superposition, données en fonction du temps.	159
Figure D.22. Identification du modèle h_{40} sur période 1 et validation sur période 2.	160
Figure D.23. La mesure h_{40} et son estimée \hat{h}_{40} en superposition, données en fonction du temps.	160
Figure D.24. Identification du modèle Q_{24} sur période 1 et validation sur période 2.	167
Figure D.26. Identification du modèle Q_{30} sur période 1 et validation sur période 2.	168
Figure D.27. Allure de Q_{30} et \hat{Q}_{30} en superposition en fonction du temps	168
Figure D.28. Identification du modèle Q_{36} sur période 1 et validation sur période 2.	169
Figure D.29. Allure de Q_{36} et \hat{Q}_{36} en superposition en fonction du temps	169

Liste des Tableaux

Tableau 1. Typologie des sources d'anomalies par nature de capteur	31
Tableau 3.1. Capteurs et signaux correspondants sur le réseau d'assainissement de Nantes.....	53
Tableau 3.2. Corrélation entre les différentes variables sur la période de pluie	57
Tableau 3.3. Corrélation entre les différentes variables sur la période sans pluie	58
Tableau 3.4. Corrélation entre les variables "débits et pluies".....	58
Tableau 3.5. Décalage temporel en minutes pour les variables hauteurs et vitesses..	61
Tableau 3.6. Décalage temporel en minutes pour les variables débits.....	61
Tableau 3.7. Critère d'arrêt <i>AIC</i> et variables sélectionnées	69
Tableau 3.8. Validation du modèle en considérant la période 1 comme jeu-identification.....	72
Tableau 3.9. Validation du modèle en considérant la période 2 comme jeu-identification.....	73
Tableau 3.10. Expressions des modèles établis.....	76
Tableau 4.1. P_d % des défauts constants introduits sur le signal $h_4(k)$	98
Tableau 4.2. P_d % des défauts dépendants des mesures et introduits sur $h_4(k)$ (taux constant)	98
Tableau 4.3. P_d % des défauts dépendants des mesures et introduits sur $h_4(k)$ (taux variable).....	98
Tableau 4.4. P_d % des défauts dépendants des mesures et introduits sur $s_1(k)$ (taux constant)	99
Tableau 4.5. P_d % des défauts dépendants des mesures et introduits sur $s_1(k)$ (taux constant)	100
Tableau 4.6. P_d % des défauts introduits sur $s_4(k)$ pour τ_d de 10%.....	100
Tableau 4.7. Instants et durées des ruptures détectées dans les résidus	102
Tableau 4.8. Structure de la base des règles des inférences	111
Tableau 4.9. P_d , P_{fa} et τ_{fa} de défaut réel, opérateur « minimum ».....	114
Tableau 4.10. P_d , P_{fa} et τ_{fa} de défaut réel, opérateur proposé par Mandel.....	114
Tableau 4.11. P_d , P_{fa} et τ_{fa} de défaut réel, nouvel opérateur	114
Tableau 4.12. P_d des défauts introduits sur $h_4(k)$ (taux constant), opérateur « minimum »	115
Tableau 4.13. P_d des défauts introduits sur $h_4(k)$ (taux constant), opérateur proposé par Mandel.....	115
Tableau 4.14. P_d des défauts introduits sur $h_4(k)$ (taux constant), nouvel opérateur	115
Tableau 4.15. P_d , P_{fa} et τ_{fa} pour les trois procédures de détection et localisation d'un défaut réel.....	121

Tableau A. Différents points de mesure et équipements correspondants.....	135
Tableau C.1. Résultats de l'application de la technique OLS pour le débit Q_{24}	145
Tableau C.2. Résultats de l'application de la technique OLS pour le débit Q_{30}	145
Tableau C.3. Résultats de l'application de la technique OLS pour le débit Q_{36}	145
Tableau D.1.a. Evaluation des modèles avec : période 1 considérée comme jeu-identification.....	161
Tableau D.1.b. Evaluation des modèles avec : période 2 considérée comme jeu-test	162
Tableau D.1.c. Evaluation des modèles avec : période 3 considérée comme jeu-test	162
Tableau D.2.a. Evaluation des modèles avec : période 2 considérée comme jeu-identification.....	163
Tableau D.2.b. Evaluation des modèles avec : période 1 considérée comme jeu-test	163
Tableau D.2.c. Evaluation des modèles avec : période 3 considérée comme jeu-test	164
Tableau D.3.a. Evaluation des modèles avec : période 3 considérée comme jeu-identification.....	164
Tableau D.3.b. Evaluation des modèles avec : période 1 considérée comme jeu-test	165
Tableau D.3.c. Evaluation des modèles avec : période 2 considérée comme jeu-test	165
Tableau D.4. Evaluation des modèles débits avec : période 1 considérée comme ..	170
jeu-identification	170
Tableau D.5. Evaluation des modèles débits avec : période 2 considérée comme jeu-test	170
Tableau E.1. Les sensibilités $S_{i,j}$ pour l'ensemble de modèles établis	171
Tableau E.2. Matrice des signatures des défauts théoriques $\Lambda_{\tilde{y}}$, pour l'ensemble de modèles établis.....	172

Introduction

La dépollution des eaux usées avant leur rejet dans le milieu récepteur est devenue un impératif réglementaire [Directive, 91 ; Directive, 00] à l'échelle de l'Europe, justifié par la protection de l'environnement et de la santé publique. Cette dépollution est réalisée dans des usines de traitement (ou stations d'épuration), mettant en œuvre divers procédés, mécaniques, biologiques et physico-chimiques. En France, la construction des stations d'épuration s'est fortement développée depuis les années 60. Le parc installé atteint aujourd'hui plus de 16000 unités, dont une grande majorité de petites installations : seulement 7,5 % des usines ont une capacité permettant de desservir plus de 10000 habitants [IFEN, 06].

Cependant le processus de dépollution n'est pas confiné dans ces usines. Il s'étend en réalité à tous les réseaux de collecte permettant d'acheminer les effluents vers une usine de traitement. Ces réseaux représentent un linéaire cumulé de 250000 km, dont 80000 km desservent des agglomérations de plus de 10000 habitants. Ils ont un fonctionnement complexe qui se répercute sur la quantité et la qualité des effluents soumis au traitement. En particulier, ils peuvent être le siège d'entrées et de sorties plus ou moins bien maîtrisées et se comporter comme des réacteurs physico-chimiques spontanés, présentant des phases d'accumulation, de transformation, et de relargage des polluants. On s'attache donc maintenant à évaluer et améliorer la performance globale de l'ensemble réseau et station d'épuration, désormais identifié par le terme "système d'assainissement" [Décret, 94]. Ainsi sont nées les études diagnostic, tout d'abord pour répondre aux exigences techniques des agences de l'eau, puis comme élément des procédures réglementaires. Enfin, l'arrêté interministériel [Arrêté, 94] introduit le concept d'autosurveillance des stations d'épuration et des réseaux d'assainissement.

Toutes ces démarches de surveillance et de diagnostic s'appuient sur des résultats de mesures, qui peuvent, selon le cas, être recueillis au cours de campagnes ponctuelles, ou rapatriés en permanence depuis un réseau de capteurs fixes. Un tel réseau permet d'assurer la surveillance en temps réel du fonctionnement du système d'assainissement, ainsi que

l'évaluation hors ligne de ses performances, en établissant des bilans à des échelles de temps allant de la journée à l'année.

La surveillance des usines d'épuration est une application classique des techniques de supervision industrielle, portant sur des équipements conçus pour fonctionner ensemble, avec la présence régulière de personnel. Elle est mise en œuvre depuis des décennies sur certains sites. En revanche la supervision des réseaux est plus récente et plus difficile. L'extension géographique du système (quelques dizaines à quelques centaines de km), impose une faible densité de capteurs, décrivant le fonctionnement du système à une échelle grossière. La construction des ouvrages, quant à elle, s'est échelonnée sur de longues périodes (dépassant parfois le siècle), pour répondre à des objectifs divers, et la connaissance que l'on a du système et de son fonctionnement (y compris ses dysfonctionnements) est très incomplète. Enfin, les opérations quotidiennes d'exploitation (curage, réparations...) modifient ce fonctionnement d'une manière encore assez peu traçable.

Notre travail porte sur la fiabilité des informations fournies par les capteurs fixes équipant les réseaux d'assainissement, en cherchant à appliquer des techniques qui sont utilisées en supervision industrielle, et qui ont déjà été testées sur d'autres réseaux urbains, tels que les réseaux de distribution d'eau potable.

Les conditions des mesures dans les réseaux d'assainissement sont telles que la fiabilité des résultats de mesures ne peut être garantie en permanence. Les résultats faux peuvent considérablement perturber un bilan. C'est pourquoi il est indispensable d'implanter un niveau de surveillance dont le rôle est de connaître à chaque instant l'état du système et fournir des données validées. Dans ce contexte, la validation des mesures a pour but de limiter l'effet perturbateur que celles-ci sont susceptibles d'induire lorsqu'elles sont entachées d'erreurs. Elle est donc une méthode d'identification des conditions anormales, liées à des dysfonctionnements dans les réseaux de mesure, en particulier ceux liés aux capteurs. Elle permet d'évaluer la fiabilité de chaque point de mesure et de caractériser les performances du réseau de mesure, et de l'améliorer.

La validation des données dans le cas du réseau d'assainissement porte essentiellement sur les mesures de hauteur, de vitesse et de débit d'écoulement d'eau dans le réseau. Ces mesures sont utilisées pour déterminer des contextes hydrologiques (temps sec/temps de pluie), pour calculer des grandeurs intégrées (volumes d'effluents rejetés), et pour calculer des indicateurs (coefficient de ruissellement, taux de collecte, etc.). Elles représentent donc une

masse considérable d'informations qui, en raison de l'importance des enjeux environnementaux et financiers que recouvre son usage, doit être représentative, pertinente et fiable.

Les méthodes de validation permettent de déceler les résultats suspects, et de les soumettre à l'expertise d'un opérateur humain. Ces méthodes utilisent la redondance globale liant les divers points de mesure : la structure arborescente du réseau de collecte, et l'influence générale des précipitations sur les débits font qu'une partie de l'information est commune à plusieurs points de mesure. Ces méthodes visent à détecter des incohérences entre les différents capteurs, et à localiser le (ou les) capteurs à l'origine de ces incohérences, en s'appuyant sur des jeux de données réelles issues du système de mesure de Nantes.

La motivation de l'étude découle du constat que l'autosurveillance des réseaux est basée sur des méthodes semi-automatisées (l'acquisition des données est automatisée, mais pas la validation) et qui sont consommatrices de temps.

L'originalité de ce travail de thèse repose sur la validation automatisée des données de mesure issues du réseau d'assainissement. Cette validation utilise des outils permettant la détection et la localisation des défaillances qui se présentent sur l'ensemble des capteurs installés sur le réseau. Cette détection et localisation se basent sur un traitement de l'ensemble des signaux délivrés par ces capteurs. Ce traitement est effectué à un pas de temps d'une minute. D'autre part, nous avons amélioré le taux de fausses alarmes au niveau de la prise de décision dans la phase de localisation basée sur le concept de la logique floue.

L'ensemble des résultats obtenus dans cette étude, est présenté dans ce mémoire. Celui-ci est organisé comme suit :

Chapitre 1 : Les réseaux de mesure en assainissement

Dans ce chapitre, nous présenterons l'ensemble des objectifs pour lesquels les mesures sur les réseaux d'assainissement sont effectuées et nous décrirons l'ensemble de ces mesures (hauteurs, vitesses et les débits d'écoulement de l'eau dans ces réseaux). Nous indiquerons par la suite les méthodes et les techniques de mesure utilisées. Puis, nous exposerons les différentes anomalies qui peuvent affectées les mesures issues des réseaux d'assainissement. Comme le réseau de mesure de Nantes représente le domaine d'application de l'étude menée dans ce mémoire, nous terminerons ce chapitre par une description de ce réseau.

Chapitre 2 : Généralités sur les méthodes de validation des données de mesure

Dans ce chapitre, nous dresserons un état de l'art sur les techniques de validation des résultats de mesure, surtout celles appliquées au domaine de l'environnement. Comme nous nous sommes intéressés à la méthode de validation basée sur le diagnostic des défauts à base de modèles analytiques, nous décrirons celle-ci plus particulièrement. A la fin de ce chapitre, nous présenterons des généralités sur le diagnostic des systèmes et quelques notions utiles pour notre travail.

Chapitre 3 : Modèles

La méthode de détection et de localisation des défauts qui se présentent sur les capteurs installés sur le réseau, nécessite des modèles analytiques. Dans ce chapitre, nous allons présenter l'ensemble de ces modèles établis pour la représentation de l'ensemble des signaux considérés. Ces modèles permettent de générer des relations de redondance de l'information. Celle-ci est dite aussi redondance analytique et permet de remédier à l'absence de redondance matérielle. Dans ce chapitre, nous désignerons par modèle la description mathématique des relations qui existent entre différentes variables sur le réseau de mesure considéré. Nous exploiterons les corrélations qui existent entre les diverses données disponibles. Ce sont des corrélations débits-débits, hauteurs-vitesses et hauteurs-hauteurs.

Le choix de la structure de modèles s'effectuera en se basant d'une part sur les corrélations existantes entre les différents signaux considérés et d'autre part sur une procédure statistique pour la sélection des variables depuis l'ensemble présélectionné. Les structures que nous allons utiliser, seront données sous forme de relations linéaires entre chacune des variables à expliquer et les variables explicatives. Une procédure d'identification des paramètres de ces modèles sera appliquée par la suite. Elle est basée sur les techniques de régression linéaire. Les modèles que nous allons établir représenteront un ensemble de relations analytiques utilisables dans une phase de détection et de localisation des défaillances.

Chapitre 4 : Méthodes de détection et de localisation des défaillances appliquées aux données issues du réseau d'assainissement de Nantes

Dans ce chapitre, nous allons présenter un ensemble de méthodes de détection et de localisation des défaillances qui peuvent se présenter au niveau des capteurs installés sur le réseau d'assainissement. Ces méthodes se basent principalement sur les modèles établis dans

le chapitre 3. Ils seront utilisés pour générer des indicateurs des défauts de mesure dits résidus et générer la matrice des signatures des défauts théoriques. Ces résidus seront utilisés dans trois procédures de détection et de localisation. Une première procédure de détection et de localisation que nous avons appelé ‘‘Page-distance’’, utilise le test statistique de Page pour la détection, et le calcul de distance *Euclidienne* pour la localisation. Une deuxième procédure de détection et de localisation basée sur la méthode dite DMP (pour Diagnostic Model Processor), proposée dans [Petti et *al.*, 90; Petti et Djurjati, 91], et une troisième basée sur les concepts de la logique floue [Mandel 98, Giuliani 98, Hamioud et *al.*, 05]. À la fin de ce chapitre, nous allons présenter quelques comparaisons entre ces méthodes.

Chapitre 1 Les réseaux de mesure en assainissement

Sommaire

1.1	Introduction	8
1.2	Objectifs d'obtention de mesures en réseau d'assainissement.....	8
1.2.1	Objectifs liés aux questions techniques.....	8
1.2.2	Objectifs liés aux questions réglementaires et juridiques	12
1.2.3	Objectifs liés aux questions financières	15
1.3	Conception du réseau de mesure	16
1.3.1	Grandeurs mesurées	16
1.3.2	Position des points de mesure sur le système d'assainissement.....	16
1.3.3	Stratégie d'acquisition.....	20
1.3.4	Fréquence de transmission	21
1.3.5	Stratégie d'échantillonnage	21
1.3.6	Transmission et gestion des données	22
1.4	Méthodes et technologie pour l'évaluation du débit	24
1.4.1	Mesure de débit	24
1.4.2	Mesure de la hauteur	25
1.4.3	Mesure de la vitesse	26
1.5	Anomalies affectant les mesures	27
1.5.1	Anomalies de fonctionnement.....	27
1.5.2	Anomalies sur le processus	29
1.6	Données.....	32
1.6.1	Réseau de mesure de Nantes	32
1.6.2	Acquisition et gestion des données sur le système de Nantes.....	33
1.6.3	Données disponibles.....	33
1.6.4	Données utilisées.....	34
1.7	Conclusion.....	34

1.1 Introduction

Parmi les objectifs principaux des réseaux d'assainissement, la protection de la population contre les inondations et la réduction des polluants rejetés dans les milieux (eaux) récepteurs, essentiellement aux temps de pluie. Les systèmes de mesure sur les réseaux d'assainissement sont généralement basés sur un réseau de points permanents (ensemble des appareils) de mesure de pluie et de débit. Malheureusement, le bon fonctionnement de système de mesure est limité par des facteurs (anomalies) qui rendent la surveillance de réseau d'assainissement sensible.

Notre travail porte sur une application de méthodes de validation des données de mesures. Ces méthodes doivent être adaptées à la nature de ces données. Il est donc utile de présenter le contexte dans lequel ces mesures sont obtenues et les méthodes mises en œuvre pour les recueillir.

Dans ce chapitre, nous présenterons tout d'abord, les objectifs pour lesquels les mesures dans les réseaux d'assainissement sont collectées. Nous décrirons par la suite le réseau de mesure ainsi que des méthodes et techniques de mesure utilisées pour obtenir les différentes mesures hauteurs, vitesses et débits d'écoulement d'eau dans les réseaux d'assainissement. Nous présenterons à la fin de ce chapitre, le réseau de mesure de Nantes et les différentes anomalies qui peuvent affecter ces mesures.

1.2 Objectifs d'obtention de mesures en réseau d'assainissement

Les objectifs liés à l'utilisation des capteurs en assainissement sont regroupés selon des besoins en trois grandes catégories [Bertrand-Krajewski et *al.*, 00] :

1.2.1 Objectifs liés aux questions techniques

Ces objectifs sont relatifs à la programmation de l'assainissement, à des besoins d'étude ou de recherche, et aux besoins du gestionnaire.

a. Objectifs liés à la programmation d'assainissement

Il s'agit de définir des orientations et options techniques d'un système d'assainissement, de leurs mises en œuvre et de leurs actualisations. Ces objectifs correspondent aux attributions du maître d'ouvrage. Les résultats de mesure sont généralement utiles, voire indispensables à l'élaboration des schémas directeurs d'assainissement qui consistent à définir pour le court et le long terme, les modalités de collecte et de traitement des eaux usées sur une commune.

- Evaluation des performances du système

Cette évaluation est indispensable et doit être confrontée aux objectifs assignés au système d'assainissement, en particulier ceux relatifs à la protection des milieux aquatiques superficiels et souterrains. Elle permet la prise en considération des indicateurs de performance [Le Gauffre et *al.*, 06], par exemple le taux de débit des eaux parasites (donné en m^3/an : pour $4000 \text{ m}^3/\text{an}$, nous pouvons dire que le système est performant et pour un taux de $10000 \text{ m}^3/\text{an}$ ou plus, le système est mauvais), le taux de saturation des ouvrages, la fréquence de débordement ou d'apparition d'autres défaillances et le temps de séjour des eaux dans les ouvrages.

Parmi ces indicateurs, nous avons le taux physique de renouvellement du réseau qui représente le rapport entre la longueur du réseau renouvelée (ou réparée) et la longueur totale du réseau de l'année, et le taux d'obstruction sur réseau qui représente le nombre de désobstructions réalisées sur le réseau par la longueur totale du réseau.

- Suivi de l'évolution de la pluie

En pratique, l'évolution de la pluie est souvent évaluée à partir des valeurs des débits transités en certains points du réseau d'assainissement et peut être associée à des objectifs concernant plus directement les performances. Notons également que le développement de l'urbanisation et de l'activité économique mérite un suivi spécifique concernant son effet sur le fonctionnement des ouvrages d'assainissement. Ce suivi permet d'ajuster les hypothèses des schémas directeurs, et de déclencher en temps convenable les études et les aménagements nécessaires.

b. Objectifs relatifs à des besoins d'étude et de maîtrise d'œuvre

Dans les réseaux d'assainissement, il existe de fortes interactions entre les diverses échelles spatiales : le dimensionnement d'un ouvrage particulier peut nécessiter une modélisation d'un ensemble beaucoup plus vaste afin d'étudier et de prendre en compte les influences réciproques des diverses parties du système d'assainissement.

- Compréhension du fonctionnement, conception et dimensionnement

Les objectifs liés à la compréhension du fonctionnement du système d'assainissement font appel à des résultats de mesure nombreux et variés (pluie, hauteur et vitesse d'écoulement d'eau dans le réseau), de plus en plus fréquemment liés à des modèles de simulation numérique dont ils permettent la validation. Ces modèles permettent non seulement d'analyser des modifications du système d'assainissement et d'en affiner le

dimensionnement, mais aussi d'atteindre un certain niveau de compréhension du fonctionnement du système, permettant de dépasser l'aspect local de la conception des ouvrages et d'imaginer des restructurations cohérentes à des échelles plus globales.

D'autres résultats de mesures plus sommaires, peuvent également être utilisés pour les tâches de conception et surtout de dimensionnement simple, sans pouvoir contribuer beaucoup à l'objectif de compréhension. Ces résultats de mesure reposent en général sur une exploitation et une analyse statistique des données expérimentales acquises et sont d'autant plus fiables que les séries de données sont plus longues.

- Vérification de l'efficacité des aménagements

Ce type d'objectif concerne le plus souvent des aménagements ou ouvrages particuliers. L'évaluation et la validation de techniques nouvelles ou en cours de développement (traitement des eaux pluviales au fil de l'eau, réhabilitation de réseaux, etc.) peut être envisagée par la réalisation de tranches successives de travaux et le suivi des résultats obtenus. Ceux ci permettent alors d'orienter et d'améliorer les solutions techniques à retenir dans les projets ou extensions futurs.

c. Objectifs relatifs à l'exploitation des systèmes d'assainissement

L'exploitation des systèmes d'assainissement assure le bon fonctionnement et l'entretien de l'ensemble du réseau (collecteurs, branchements, déversoirs d'orage, vannes, etc.), en accomplissant les missions suivantes :

- Relevé des caractéristiques de l'état des ouvrages et de leur vieillissement ;
- Manœuvres diverses du réseau de façon à maintenir la permanence hydraulique des écoulements ;
- Surveillance dans le cadre de la police du réseau (exemple : déversements délictueux, non conformes) ;
- Contact avec les Services Techniques des Mairies et les habitants usagers du service.

- Détection d'anomalies

Pour cet objectif, les techniques de mesure utilisées sont généralement simples et correspondent à des besoins élémentaires : détection de niveaux hauts, bas ou de surverses, durées de pompage, etc. Néanmoins, des technologies plus élaborées utilisées dans des applications plus complexes sont mises en œuvre de plus en plus fréquemment : mesurages de

hauteur d'eau par capteur de pression ou à ultrasons, débitmètres électromagnétiques sur conduites de refoulement, etc.

- Gestion d'ouvrages en temps réel

Cet objectif nécessite des moyens de mesure plus élaborés, par exemples pour gérer le remplissage et la vidange de capacités de collecte (bassins de retenue, collecteurs de grand diamètre) ou pour orienter et répartir les flux collectés vers différentes destinations (déversoirs d'orage, station d'épuration, station de relèvement ou de refoulement, ouvrages de traitement spécifiques, etc.). Il faut des moyens de mesure adaptés pour connaître les entrées (pluie, eaux usées collectées), les sorties et l'état des ouvrages (on se limite souvent à leur taux de remplissage, mais des données plus spécifiques peuvent être nécessaires, tels que les débits instantanés, la turbidité des effluents, leur salinité, etc.).

- Surveillance des écoulements en temps réel

Cet objectif fait appel aux mêmes moyens de mesure qui viennent d'être évoqués au paragraphe précédent. Les points de mesure sont choisis pour être représentatifs du fonctionnement du réseau par temps sec et/ou par temps de pluie. Ils renseignent sur l'état de charge du réseau et sur la « normalité » de son fonctionnement. Des exemples d'application les plus fréquentes sont : la surveillance de non déversement par temps sec ou le suivi de points de débordement par temps de pluie.

- Accessibilité des ouvrages et sécurité du personnel

Cet objectif est généralement fixé en complément des trois précédents, notamment dans le cadre de la gestion en temps réel. En effet, ce type de gestion peut générer des risques spécifiques supplémentaires : ouvertures automatiques de vannes, modifications de seuils, du sens des écoulements, mises en charge, etc., pouvant avoir des conséquences importantes sur l'accessibilité des ouvrages et la sécurité du personnel.

- Bilan annuel d'exploitation

Ce bilan peut comporter une synthèse des dysfonctionnements (pannes, surverses, etc.) et de leurs conséquences (plaintes, impacts sur les milieux aquatiques et leurs usages, etc.), l'évolution du taux de collecte, ainsi que les indicateurs plus économiques faisant appel à l'évaluation ou à la quantification de l'énergie consommée, des réactifs utilisés, des matières collectées (produits de curage, matières de vidange, etc.). Il sera utilisé pour améliorer et orienter l'exploitation : planification des curages, des renouvellements, automatisation, inspections vidéo, suivi des conventions de rejets industriels, etc.. Les informations collectées

pour les objectifs précédents pourront bien sûr être mises à profit dans le cadre du bilan annuel, après une exploitation synthétique et statistique en temps différé. Elles devront le plus souvent être complétées par la détermination des flux en certains points clés du réseau, notamment ses principales entrées ou sorties (station d'épuration, déversoir d'orage, etc.).

- Gestion de la maintenance et de l'entretien des ouvrages et des équipements

La mise en œuvre des stratégies d'exploitation définies dans le cadre de l'objectif précédent peut nécessiter l'évaluation d'indicateurs permettant de déclencher des actions telles que le curage, le renouvellement de matériels électromécaniques, des campagnes de mesure de débits instantanés. Ces indicateurs sont par exemple des taux de remplissage d'ouvrages pour une pluie donnée, des taux d'ensablement des collecteurs, des débits ou des durées de pompage, des débits observés à l'aval du réseau.

- Gestion et fiabilisation du capteur

Certains résultats de mesure seront recherchés ou utilisés pour en valider d'autres ou s'y substituer en cas de défaillance. Par exemple, une détection de seuil par contact pourra compléter un suivi permanent de la hauteur d'eau. Sous cet objectif peut être regroupée l'obtention de toutes les informations, données et résultats de mesure permettant de connaître l'état du capteur et de pouvoir le gérer. La redondance des informations qui passe par la redondance des capteurs ou appareils de mesure, si possible à partir des technologies différentes pour éviter les mêmes causes de défaillances ou d'erreur, constitue un élément très important de fiabilisation du capteur.

1.2.2 Objectifs liés aux questions réglementaires et juridiques

a. Respect des obligations relatives à l'autosurveillance

Ces obligations sont définies pour l'essentiel, par l'arrêté du 22 décembre 1994 [Arrêté, 94] et par les recommandations ministérielles du 12 mai 1995 [Circulaire, 95].

Ces textes ne sont pas très contraignants sur ce qu'il faut mesurer au niveau des réseaux d'assainissement (le cas des stations d'épuration est différent, et les exigences réglementaires correspondantes y sont beaucoup plus détaillées), mais s'attachent plutôt à :

- Définir divers objectifs pouvant être assignés à la collecte de résultats de mesure ;
- Inventorier les divers acteurs concernés par la mise en place des dispositifs de mesure ou par les résultats que ces dispositifs fournissent.

L'arrêté du 22 décembre 1994 [Arrêté, 94] évoque trois types d'objectifs pouvant être assignés à un dispositif de suivi permanent du fonctionnement d'un système d'assainissement :

- La surveillance des rejets et éventuellement du milieu aquatique à l'échelle locale. Cette surveillance se réfère implicitement à une obligation de résultats, explicitée généralement sous la forme d'autorisations de rejets ;
- l'évaluation du bon fonctionnement et de la fiabilité des ouvrages. Il s'agit davantage ici d'un contexte de vérification des moyens mis en œuvre, c'est-à-dire de l'adéquation entre les ouvrages et leur exploitation d'une part, et les résultats souhaités d'autre part;
- l'évaluation de l'impact des dépassements des niveaux (débits, concentrations, etc.) fixés dans des autorisations de rejets (en cas de « circonstance exceptionnelles », pluvieuse notamment). Cet objectif s'apparente au premier, mais il va au delà puisqu'il prend en compte l'éventualité d'un fonctionnement dans des conditions « non nominales » : cas de force majeure (pluies exceptionnelles, ouvrage en chômage technique, réparation, etc.), ou des conditions de fonctionnement anormales ou non réglementaires. Les informations obtenues permettent ainsi de tirer des enseignements des expériences et peut être, de faire évoluer les objectifs, les spécifications ou les caractéristiques du système d'assainissement.

Seul le premier objectif, la surveillance des rejets, est assorti d'obligations précises en termes de moyens de mesure, principalement en ce qui concerne les stations d'épuration :

- Fréquence et nature des analyses des rejets des stations d'épuration en fonction de leur capacité et du milieu aquatique ;
- suivi en continu du débit et estimation de la charge polluante rejetée par temps de pluie au niveau des déversoirs d'orage et des dérivations éventuelles situés sur un tronçon destiné à collecter une charge brute de pollution organique par temps sec supérieur à 600 kg par jour (10000 EH) ;
- surveillance permettant d'estimer les périodes de déversement et des débits rejetés au niveau des déversoirs d'orage et des dérivations éventuelles situées sur un tronçon destiné à collecter une charge brute de pollution organique par temps sec comprise entre 120 et 600 kg par jour (200 à 10000 EH).

Il n'en est pas de même de l'évaluation du fonctionnement, qui peut inclure des opérations très diverses : l'arrêté mentionne, outre les inspections télévisées, des enregistrements de débits horaires sur les principaux émissaires, ainsi que l'obligation d'un suivi régulier des effluents des établissements les plus polluants, ainsi qu'une évaluation du taux de collecte. Enfin il précise qu'en cas de dépassement des seuils autorisés, le maître d'ouvrage devra accompagner la transmission des résultats d'un commentaire sur les causes de ces dépassements et sur les actions correctrices adaptées. Ceci suppose de disposer d'informations relativement précises sur ce qui se passe à l'amont des points de rejet.

A ces objectifs pour lesquels l'aspect « police des eaux » est prépondérant et qui sont assignés au gestionnaire des ouvrages, les recommandations du 12 mai 1995 [Circulaire, 95] apportent des compléments en vue de favoriser la diffusion des informations entre les différents acteurs de l'assainissement :

- détection des anomalies, alerte de l'exploitant ;
- élaboration des statistiques nationales et européennes dans le cadre de la directive européenne du 21 mai 1991 [Directive, 91];
- réponse aux besoins (non précisés) des agences de l'eau ;
- information du maître d'ouvrage ;
- échange d'information entre services ;
- information du public.

Ce texte ne définit pas explicitement de nouveaux objectifs (sauf l'alerte de l'exploitant), mais détaille les différents destinataires des informations et se réfère à leurs besoins.

b. Respect des critères de rejets et des conventions de déversements.

Il s'agit des conditions générales que doivent satisfaire des effluents résiduels industriels pour être considérés comme « effluents domestiques » et donc être admis dans le réseau public d'évacuation des eaux résiduelles urbaines, ou des conditions particulières sont fixées par des conventions de déversements. Cet objectif rejoint une prescription de l'annexe II de l'arrêté du 22 décembre 1994 relative à la surveillance des ouvrages de collecte des

établissements rejetant des substances dangereuses pour le système de traitement. Il ne s'agit pas nécessairement de mettre en place un capteur spécifique permanent faisant double emploi avec l'autosurveillance des établissements industriels.

c. Catastrophes naturelles

Un réseau de capteurs pluviométriques permet une appréciation objective des volumes précipités et de l'état de charge d'un réseau d'assainissement. Il constitue ainsi un capteur global qui permet d'évaluer le caractère exceptionnel des événements pluvieux selon une échelle d'appréciation souvent rattachée aux courbes IDF (Courbes Intensité de pluie-durée-fréquence : courbes donnant la distribution fréquentielle des valeurs maximales d'intensité de pluie sur une durée quelconque) de l'instruction technique interministérielle de 1977 [Ministère de l'Équipement, 77]. Les résultats de mesure obtenus fournissent à une collectivité ou au gestionnaire de réseau des éléments quantifiés pour juger de contentieux ou pour établir des dossiers de demande de déclaration d'état de catastrophe naturelle. Pour que ces éléments ne soient pas discutables, il est impératif que la qualité des résultats de mesure soit garantie et prouvée par la mise en place d'une démarche métrologique rigoureuse.

d. Justification des rejets des stations d'épuration et des déversoirs

Dans le cas des stations d'épuration et des déversoirs d'orage, la collectivité et son service gestionnaire doivent pouvoir prouver qu'ils respectent les autorisations de rejet qui leur ont été fixées, par exemple en cas de conflit ou de procédure juridique. En ce qui concerne les stations d'épuration et les déversoirs d'orage, l'arrêté du 22 décembre 1994 est assez précis sur les moyens à mettre en œuvre. Mais ceux-ci peuvent prendre davantage d'extension, en particulier en ce qui concerne l'amont de la station d'épuration, dans l'éventualité d'un dépassement des seuils de rejet, dont il conviendra de préciser les causes et les effets. Des résultats de mesure et des informations permettant de comprendre le fonctionnement de l'ensemble du système d'assainissement sont alors nécessaires et indispensables.

1.2.3 Objectifs liés aux questions financières

Il s'agit de déterminer une répartition des coûts de fonctionnement et d'investissement [Alexandre et Grand d'Eston, 98], le plus souvent dans un contexte intercomunal-industriel(s). Les résultats de mesure permettent de fixer cette répartition en fonction des débits et des flux polluants et en fonction du type d'ouvrage. Nous entrons ici dans le

domaine de la politique tarifaire, pour laquelle la fiabilité et l'incertitude des résultats de mesure sont peut être moins importantes que leur mode d'exploitation.

1.3 Conception du réseau de mesure

Le réseau de mesure représente l'ensemble des capteurs installés sur le réseau d'assainissement.

La définition des objectifs conduit à identifier et décrire les éléments techniques du système de mesure. Les questions à traiter peuvent être données comme suit :

- Grandeurs mesurées ;
- position des points de mesure sur le réseau d'assainissement ;
- stratégie d'acquisition ;
- gestion des données.

1.3.1 Grandeurs mesurées

Dans la section 1.2 relative à la définition des objectifs, un certain nombre de grandeurs à étudier a été déjà évoqué : pluies, hauteurs d'eau dans des ouvrages, vitesses, débits, qualité des milieux aquatiques, etc. Certaines de ces grandeurs ne sont pas accessibles par un mesurage direct : les débits, en particulier représentent des résultats de mesures obtenus à partir des valeurs de hauteurs et de vitesses, qui peuvent également avoir un intérêt propre vis à vis d'autres objectifs. Nous verrons plus loin dans ce chapitre comment le débit est mesuré.

1.3.2 Position des points de mesure sur le système d'assainissement

a. Réseaux d'assainissement et réseau de mesure

Les réseaux d'assainissement ont pour fonction l'évacuation des eaux usées et des eaux pluviales loin des habitations, et le plus souvent leur acheminement vers une station d'épuration. Un réseau d'assainissement est constitué par de nombreux ouvrages tels que :

- Des collecteurs et canalisations, fonctionnant en général selon le principe des écoulements gravitaires à surface libre ;
- des branchements, et autres ouvrages de collecte et d'acheminement des eaux de chaque bâtiment et chaque parcelle vers les collecteurs ;

- des bassins de retenue qui sont destinés à stocker temporairement les eaux urbaines excédentaires lors des pluies avant de les restituer au réseau ou au milieu récepteur dans des conditions acceptables ;
- des déversoirs d'orages qui permettent le rejet direct d'une partie des effluents au milieu naturel lorsque les débits à l'amont de ces déversoirs dépassent les capacités d'évacuation ou ce traitement à l'aval ;
- des siphons utilisés principalement pour franchir les obstacles (rivière, voie ferrée, etc.)
- des vannes qui représentent des systèmes permettant de réguler le débit dans un ouvrage hydraulique ;
- des ouvrages de traitements : stations d'épurations (à l'aval du réseau), dessableurs et bassins d'orage (sur le réseau).

L'ensemble réseau d'assainissement et station d'épuration, constitue un système d'assainissement.

Le réseau d'assainissement a souvent une structure arborescente : en temps sec il n'y a qu'un chemin pour aller d'une source amont jusqu'à la station d'épuration ou le milieu récepteur. En temps de pluie des sorties intermédiaires peuvent se faire au niveau des déversoirs d'orages. Le réseau peut être unitaire et collecter eaux usées et eaux pluviales dans les mêmes canalisations ou séparatif, il y a alors deux réseaux séparés, l'un pour les eaux usées, l'autre pour les eaux pluviales.

Un réseau de mesure est constitué d'un ensemble de capteurs. Sa structure reflète celle du réseau d'assainissement. Un point de mesure rassemble les mesures (hauteur, vitesse ou débit d'écoulement d'eau) issues d'un ou de plusieurs capteurs. On peut distinguer divers types de points de mesure :

- Les points de transfert (PT) permettent de mesurer les débits ou les masses de polluants transférés de l'amont à l'aval du réseau jusqu'à une station d'épuration ou un exutoire;
- les entrées et sorties d'ouvrage de traitement et en particulier les stations d'épuration (SE) permettent de réaliser des bilans d'eau et de pollution ;
- les points de déversement (correspondants aux déversoirs d'orages mobiles ou fixes notés respectivement DO et DF, qui permettent l'évaluation des débits rejetées au milieu récepteur sans traitement) ;

- des pluviomètres qui permettent de mesurer une des entrées du système, à savoir les précipitations.

Nous avons représenté dans les figures 1.1 et 1.2 respectivement, deux schéma de base d'un système d'assainissement et de son système de mesure.

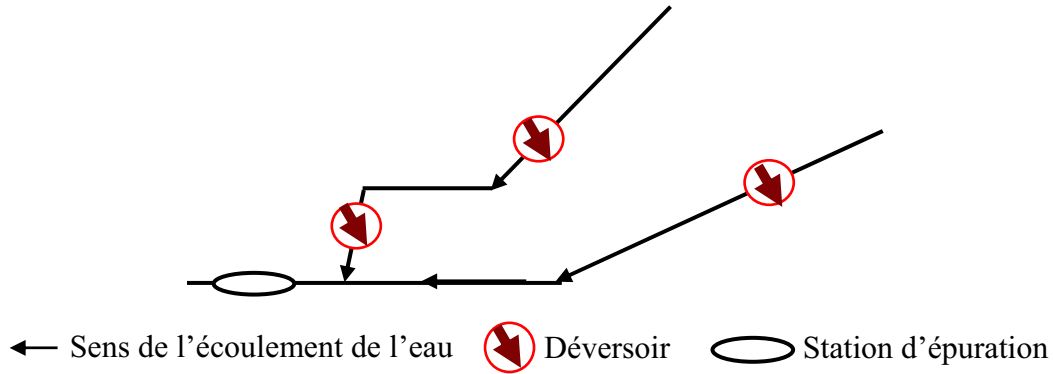


Figure 1.1. Schéma d'un système d'assainissement

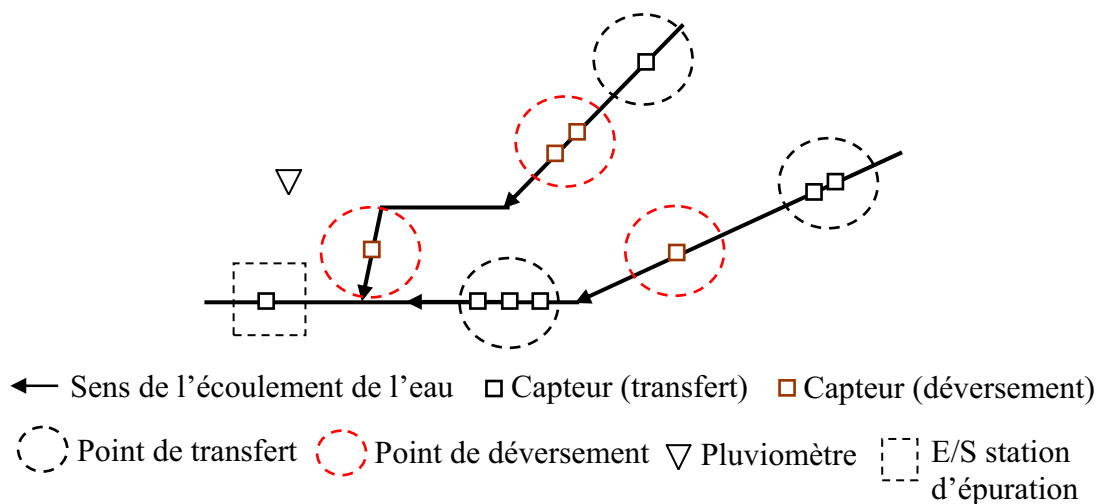


Figure 1.2. Schéma d'un système de mesure

b. Densité spatiale des points de mesure

Un réseau de mesure en assainissement comporte quelques dizaines de points de mesure répartis en fonction des objectifs.

Des objectifs de type réglementaire (autorisations de rejets et autosurveillance) mettront l'accent sur les sorties du système (déversoirs d'orage et stations d'épuration). Des objectifs comptage sur les entrées, des objectifs d'exploitation sur les flux ou plus simplement les stocks ou les rendements dans les ouvrages conduisent également à une localisation quasiment imposée des points de mesure. Il en est de même pour la surveillance des points de déversement ou de débordement. Des objectifs plus globaux tels que l'évaluation des

performances ou l'aide à la conception et au dimensionnement doivent prendre en compte des points de mesure plus nombreux afin de disposer d'informations permettant d'accéder à une réelle compréhension du fonctionnement du système d'assainissement dans toute son étendue spatiale.

La répartition des points de mesure au vu des contraintes financières inévitables, peut relever d'un compromis à fixer entre deux types de stratégie et en fonction des objectifs visés : soit de nombreux points de mesure peu coûteux et donnant une information limitée mais spatialement distribuée, soit un nombre limité de points de mesure plus élaborés fournissant une information plus complète. La première stratégie est assez peu utilisée sauf pour la détection des défauts ou les alarmes, domaine où elle se banalise avec le développement des systèmes de télésurveillance.

La densité spatiale des pluviographes est un problème spécifique compte tenu de la variabilité spatio-temporelle importante des événements pluvieux. En général, l'augmentation de la densité spatiale de pluviographes a pour intérêt de fournir une information plus détaillée (meilleure évaluation de la lame d'eau précipitée sur la surface du bassin versant) plutôt que d'en augmenter la richesse ou d'en diminuer l'incertitude au niveau local.

c. Complémentarité des points de mesure

Pour des objectifs à caractère juridique ou financier, chaque point de mesure est en général peu dépendant des autres (comptage de débit en vue d'une répartition financière par exemple). Il n'en va pas de même dès que l'on cherche à comprendre le fonctionnement du système d'assainissement dans sa globalité, à en évaluer les performances ou à concevoir des aménagements, ou plus simplement pour son suivi. Dans ces cas, seule une approche intégrant des points de mesure complémentaires permet d'atteindre les objectifs visés.

La complémentarité la plus fréquente entre les points de mesure apparaît pour les calculs de bilan, que se soit à l'échelle de l'ensemble du système, d'un bassin versant ou d'un ouvrage particulier. De tels bilans ne peuvent être établis qu'à partir de plusieurs points de mesure en associant souvent des grandeurs différents (pluie, débits).

La complémentarité voire la redondance, apparaît également nécessaire pour les opérations de validation des données et pour l'exploitation du capteur. Dans la définition des objectifs attachés à chaque point de mesure, on précisera donc ceux qui apparaissent complémentaire avec d'autres points que l'on identifiera. Cette complémentarité des points de

mesure joue également un rôle dans l'organisation de la maintenance et pour assurer une fiabilité mutuelle des données.

1.3.3 Stratégie d'acquisition

La stratégie d'acquisition détermine les conditions de la prise d'information et de son archivage. Elle comporte les éléments suivants :

- Période de mesure (ou mesure permanente) ;
- fréquence de transmission (temps réel ou temps différé) ;
- fréquence d'auscultation ;
- intégration ;
- fréquence d'enregistrement.

Les trois derniers points peuvent être regroupés sous terme de stratégie d'échantillonnage.

a. Période de mesure

Les suivis dans le temps des mesures réalisés sur une ou plusieurs périodes de temps bien définies représentent l'approche par campagne de mesures. Cette approche permet d'avoir une information sur le fonctionnement du système d'assainissement spécifique à un contexte donné (temps de pluie, temps sec). Pour quelques objectifs cités précédemment tels que les objectifs liés à la maîtrise d'œuvres, notamment dans les études liées à des projets d'aménagement, d'extension ou de rénovation, cette approche est recommandée.

Pour la majorité des objectifs cités précédemment, les suivis permanents des mesures représentent une exigence pour les atteindre.

Nous nous intéressons donc à ce deuxième type d'approche. Il nous permet d'une part, d'obtenir une meilleure représentativité et d'autre part, il nous permet d'exploiter la cohérence globale basée sur la simultanéité des mesures. Ceci peut être obtenu rarement par la première approche.

1.3.4 Fréquence de transmission

Il s'agit ici de définir le mode d'acquisitions et de transmissions des données. Pour les objectifs liés à l'exploitation, le mode d'acquisitions et de transmissions en temps réel est nécessaire. Pour les autres objectifs, l'utilisation des résultats de mesure est effectuée pour l'essentiel en temps différé. En effet, ce ne sont pas les valeurs instantanées des grandeurs qui sont pertinentes mais le plus souvent des synthèses, des bilans ou des signaux représentant l'évolution de ces grandeurs. Cependant, même dans ce cas, une disponibilité très rapide des données est appréciable, ne serait-ce que pour juger le bon fonctionnement d'un capteur et détecter rapidement les anomalies, les dysfonctionnements et les pannes.

1.3.5 Stratégie d'échantillonnage

La stratégie d'échantillonnage des données est réalisée par des algorithmes programmés dans les enregistreurs et les appareils de mesure. Elle comprend trois éléments essentiels :

- La fréquence de scrutation (ou d'auscultation) ;
- le traitement des informations avant enregistrement ;
- la fréquence d'enregistrement (ou de transmission).

La scrutation correspond à l'interrogation des transmetteurs des capteurs et à la mémorisation temporaire d'une ou plusieurs valeurs successives avant l'enregistrement définitif qui peut être réalisé avec une fréquence beaucoup plus faible. En effet la fréquence de scrutation détermine la capacité du système à détecter les variations des grandeurs mesurées. Il y a tout intérêt à utiliser cette capacité, sauf si on doit la limiter pour des raisons de disponibilité d'énergie (alimentation par batteries). Pour les systèmes de mesure permanents, nous disposons en général d'une alimentation par le secteur et la question ne se pose pas.

La fréquence d'enregistrement ou de transmission est quant à elle limitée pour éviter de surcharger inutilement les capacités de mémorisation, de transfert et de traitement des informations. Les valeurs usuelles en assainissement s'échelonnent en général entre une minute et une heure. Les fréquences les plus élevées correspondent à des objectifs de surveillance ou d'analyse des phénomènes, les fréquences les plus basses correspondent plutôt à des objectifs de bilan ou de dimensionnement. On peut chercher à optimiser l'enregistrement en diminuant sa fréquence lors des périodes les moins intéressantes, pour l'augmenter lorsqu'on détecte l'apparition de conditions moins banales. Bien entendu, la notion de

« phénomène intéressant » dépend étroitement des objectifs visés. Pour des objectifs d'évaluation, de conception ou de dimensionnement, en particulier lorsqu'on vise une certaine compréhension ou la modélisation des phénomènes, on s'intéresse à l'évolution temporelle des grandeurs mesurées. L'apparition de variations significatives des valeurs d'une grandeur est alors un critère d'intérêt et peut être détecté par le processus de scrutation. Pour d'autres objectifs, on pourra préférer augmenter la fréquence d'enregistrement sur un critère de dépassement d'une valeur seuil, en particulier pour les objectifs liés aux aspects juridiques ou réglementaires.

La fréquence d'enregistrement des données étant généralement plus faible que la fréquence de scrutation, la valeur stockée est représentative de plusieurs valeurs élémentaires. Il importe donc de définir le traitement permettant de passer des valeurs élémentaires scrutées aux valeurs stockées, et la manière dont la fréquence d'enregistrement peut dépendre du traitement ou de son résultat. Ce traitement vise à extraire une ou plusieurs valeurs caractéristiques ou représentatives de l'ensemble des valeurs scrutées. Il peut s'agir de paramètres statistiques classiques : moyenne, cumul, minimum, maximum, écart type, etc., ou du résultat d'un filtrage numérique. Une transformation préalable des grandeurs issues des capteurs peut être nécessaire pour que ces calculs aient un sens physique. En particulier, et à titre d'exemple, il faut d'abord transformer les hauteurs en débits avant d'en calculer la moyenne.

1.3.6 Transmission et gestion des données

Tout ce qui précède concernait la partie du système de mesure implantée sur le terrain.

Toutes les installations locales sont intégrées dans un système de gestion centralisé, permettant de rassembler toute les informations en un même endroit.

a. Transmission des données

L'enregistrement des informations est effectué localement sur des stations de gestion locale (dites SGL) après passage par des transmetteurs comme le montre la figure 1.3. Ces stations assurent l'échantillonnage, l'enregistrement temporaire et éventuellement des fonctions de calcul sur les valeurs enregistrées, avant télétransmission. Quelques enregistrements se font sur des enregistreurs autonomes relevés périodiquement à l'aide d'un micro-ordinateur portable. Les satellites sont liés au poste central le plus souvent par un réseau téléphonique commuté (dit RTC) pour transmettre leurs données régulièrement au

superviseur, mais disposent d'une autonomie de mémoire leur permettant de stocker plusieurs jours de données. Dans le cas d'application en temps réel, le schéma général est identique, mais la capacité de stockage local est moindre et la transmission beaucoup plus fréquente, et les lignes sont souvent spécialisées.

Dans les deux cas les satellites peuvent générer des alarmes à partir de chaque entrée capteur, sur des critères relatifs à l'étendue de mesure et à leur dynamique (rémanence, gradient). Il s'agit alors d'une prévalidation en temps réel permettant d'intervenir rapidement pour supprimer les erreurs sur les données les plus flagrantes.

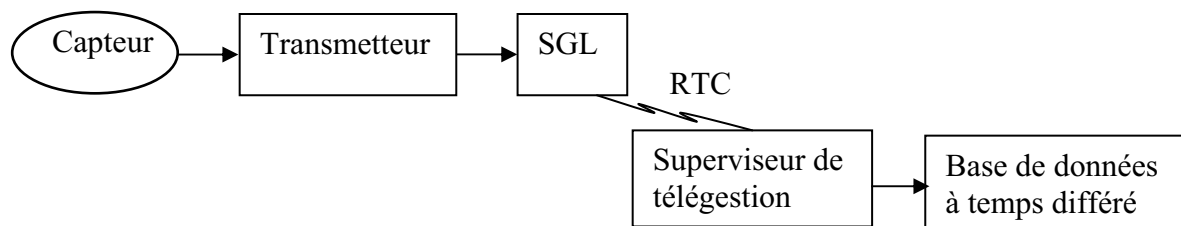


Figure 1.3. Synoptique d'acquisition et de gestion des données

Les données sont ensuite intégrées dans une base de données, qui permet un accès facile aux informations nécessaires à l'exploitation des données.

b. Gestion des données

Une base de données permet d'enregistrer de manière structurée divers types d'informations :

- Des informations plutôt statiques, décrivant la structure du réseau d'assainissement et du réseau de mesure ;
- des données dynamiques : résultats de mesure bien sûr, mais aussi historique des divers événements et interventions dont la connaissance est utile pour l'interprétation de ces résultats ;
- des valeurs de paramètres, intervenant dans les formules et algorithmes de calcul ;
- des résultats des calculs qu'ils soient effectués sur les résultats de mesures, ou issus d'analyses plus complexes.

Les outils permettant d'exploiter le contenu de la base de données couvrent l'ensemble du processus, qui va de la transformation des données brutes (hauteurs d'eau, vitesses, basculements d'augets.) en données exploitables (intensités pluvieuses, débits.) jusqu'à l'analyse et la synthèse des résultats.

La plupart du temps, les données sont échantillonnées d'une manière asynchrone et à fréquence variable. Les outils associés à la base de données doivent donc permettre un rééchantillonnage des données, permettant de reconstituer des valeurs pour n'importe quel instant. La formule d'interpolation doit tenir compte du mode d'échantillonnage d'origine, et en particulier de mode d'intégration avant enregistrement. Ces outils peuvent aussi permettre d'identifier les données manquantes, à condition que la stratégie d'échantillonnage comporte une fréquence fixe permettant d'identifier ces manques.

Les points de mesure, les types de mesure préconisés sur chaque point ainsi que le type d'équipement utilisé sont récapitulés dans le tableau A.1 de l'Annexe A.

1.4 Méthodes et technologie pour l'évaluation du débit

Le débit est une grandeur très souvent évaluée car sa connaissance est essentielle vis-à-vis du fonctionnement du système d'assainissement à plusieurs titres :

1. La mesure continue des débits rejetés est quasi obligatoire pour les déversoirs d'orage ;
2. les débits constituent une base de dimensionnement pour la plupart des ouvrages, soit sous forme « instantanée » (ouvrages de transit, décantation intensive), soit sous forme intégrée (volumes, temps de séjour dans des ouvrages de traitements) ;
3. le débit dans un réseau d'assainissement est une grandeur physique qui est mesurée en passant par une ou plusieurs grandeurs mesurables. La validité de cette mesure est la plupart du temps, soumise à certaines hypothèses relatives aux conditions d'écoulement dans la section de mesure. Il convient de vérifier autant que faire se peut la validité de ces hypothèses.

1.4.1 Mesure de débit

Le débit est mesuré par un débitmètre. Ce dernier fournit une estimation de la mesure de débit à partir d'un calcul interne qui s'effectue au niveau du transmetteur en exploitant les deux mesures hauteur et vitesse dans le même point. Le débit peut être évalué par les deux méthodes suivantes :

1. Une méthode basée sur les mesures de la hauteur d'eau seulement.

La relation expérimentale de Manning-Strickler [Chow V. T, 73] est utilisée pour définir le débit en fonction de la hauteur d'eau seulement. Cette relation est donnée comme suit :

$$Q(k) = h(k)K_s (r_s(k))^{2/3} \sqrt{tg(\varphi)} \quad (1.1)$$

où Q le débit d'eau dans le réseau, h la hauteur d'eau, K_s coefficient de Manning Strickler (il vaut 70 pour le béton), r_s rayon hydraulique, $tg(\varphi)$ la pente d'écoulement d'eau par rapport à l'horizontal, φ est l'angle entre la canalisation et l'horizontal (comme φ est petite, nous avons $tg(\varphi) \cong \sin(\varphi) \cong \varphi$ et où k représente le temps. Le rayon hydraulique est donné par :

$$r_s(k) = \frac{S(k)}{r_p(k)},$$

avec $S(k)$ la surface mouillée et $r_p(k)$ le périmètre mouillé (la partie du collecteur qui se trouve au contact de l'eau). Ces deux grandeurs sont exprimées en fonction de la hauteur d'eau selon la forme du collecteur comme indiquée dans [Chow V. T, 73].

2. Une deuxième méthode utilise les mesures de hauteur et de vitesse d'eau. Tout d'abord la valeur de la section mouillée est évaluée puis le débit avec la relation (1.2) donnée ci après [Fontana et Mizier, 95].

$$Q(t) = v(k)S(k) \tag{1.2}$$

où Q le débit d'eau dans le réseau, v la vitesse d'écoulement de l'eau, S la surface mouillée dans une canalisation et où t représente le temps.

Nous allons décrire brièvement par la suite les méthodes de mesure de vitesse et de hauteur d'eau dans un réseau d'assainissement.

1.4.2 Mesure de la hauteur

La mesure de la hauteur d'eau se fait par les deux méthodes suivantes :

a. Mesure de la pression

Le principe consiste à évaluer la hauteur d'eau à partir de la pression exercée par la colonne d'eau située au dessus du capteur de mesure de cette hauteur. La pression peut être soit mesurée directement (capteur piézométrique), soit par l'intermédiaire de systèmes pneumatiques (bulle à bulle).

b. Mesure acoustique

Des capteurs à ultrasons permettent d'avoir la mesure de la hauteur d'eau. La mesure de hauteur repose sur la mesure du temps nécessaire à l'onde ultrasonore pour faire un parcours aller-retour entre le capteur et la surface de l'eau. Ces capteurs sont utilisés selon deux modalités : immergés ou aériens.

Dans le cas des capteurs ultrasons aériens, l'onde ultrasonore se propage dans l'air, et le capteur est placé au-dessus du niveau d'eau. Ces capteurs ne demandent que peu d'entretien et sont d'un fonctionnement fiable.

Dans le cas des capteurs immergés, l'onde ultrasonore se propage dans l'eau et le capteur placé au fond de l'ouvrage.

Les incertitudes liées à ces capteurs sont de l'ordre de 1% de l'étendue de la mesure et d'au moins 1 cm pour les hauteurs d'eau inférieures à 1 m [Bertrand-Krajewski et *al.*, 00].

1.4.3 Mesure de la vitesse

Des capteurs de différentes technologies permettent la mesure de la vitesse.

a. Techniques acoustiques : Ultrasons à « temps de transit »

La mesure de vitesse repose sur la mesure du temps nécessaire à l'onde ultrasonore pour faire un parcours aller-retour sur la largeur de la conduite. Ce temps est directement proportionnel à la vitesse d'écoulement. La précision que nous pouvons obtenir peut être de l'ordre de 1% [Fontana et Mizier, 95 ; Bertrand-Krajewski et *al.*, 00].

b. Techniques acoustiques : effet Doppler

Les capteurs de vitesse à effet Doppler mesurent la vitesse d'écoulement de l'eau usée en l'assimilant à la vitesse des particules présentes dans les eaux usées. Cette vitesse est mesurée en comparant la variation de fréquence de l'onde réfléchi sur la particule par rapport à la fréquence de l'onde émise [Fontana et Mizier, 95 ; Mizier, 96]. Pour ces capteurs l'imprécision est de l'ordre de 10 à 15% [Fontana et Mizier, 95].

c. Techniques électromagnétiques

Le champ magnétique créé autour de l'écoulement induit une force électromotrice qui est directement proportionnelle à la vitesse d'écoulement. Cela permet donc d'avoir la vitesse d'écoulement d'eau.

Les capteurs de vitesse posent problèmes car, pour la plupart, une maintenance fréquente est nécessaire et donc contraignante et leur implantation dans les réseaux d'assainissement peut être difficile. Donc la possibilité d'avoir un débit depuis des hauteurs d'eau seulement, est plus intéressante que celle permettant d'obtenir le débit depuis des vitesses et des hauteurs d'eau si les conditions sont favorables, par exemple pas d'effet aval.

Les indications fournies par un réseau de mesure peuvent être perturbées de diverses manières. Les différents types d'anomalies qui peuvent affecter un réseau de mesure, sont présentés dans le paragraphe qui suit.

1.5 Anomalies affectant les mesures

Un résultat de mesure n'est pas toujours une image fidèle de la réalité, et ce pour différentes raisons :

- Le capteur de mesure de hauteur d'eau ou de vitesse d'écoulement d'eau dans le réseau peut perdre le contact avec le phénomène mesuré (encrassement, colmatage, blocage, opération d'entretien) ;
 - ce capteur peut être influencé par un autre phénomène (température, pression, vitesse, parasites électromagnétiques) ;
 - la règle de correspondance entre le signal de sortie et la grandeur mesurée peut ne pas être correcte (défaillance générale du capteur, dérive, décalage, erreur de paramétrage) ;
 - il peut y avoir un mauvais repérage temporel (dérive d'horloge, mauvaise mise à l'heure, erreur de paramétrage, problème d'enregistrement);
 - possibilité d'avoir un résultat de mesure correct mais qui ne représente pas les phénomènes auxquels on s'intéresse, car ceux-ci sont masqués par des événements perturbateurs : curage d'un collecteur, arrosage, déviation des effluents à l'amont, rejet accidentel, etc. Même si les mesures sont justes, elles nécessitent une interprétation distincte de celles qui correspondent à des circonstances plus habituelles.

Les différents types d'anomalies affectant les mesures peuvent être regroupés dans deux grandes catégories :

1.5.1 Anomalies de fonctionnement

Ces anomalies sont généralement liées aux défauts de capteurs de mesure installés dans le réseau. Il y'a celles qui se maintiennent une fois apparues et celles qui se manifestent furtivement. Les quelques défauts les plus fréquents pour cette première catégorie sont :

- Défaut pérenne qui nécessite une intervention tels que : dérive qui correspond à un décalage du signal en fonction du temps, micro encrassement qui se traduit par une atténuation du signal en fonction du temps, bouchage (macro encrassement) d'un capteur, capteur hors service, ensablement, désétalonnage, panne électrique, problème sur l'électronique d'acquisition, mesure hors gamme, déphasage vitesse/hauteur.

Un exemple de ce type d'anomalies est considéré dans les figures 1.4 donnée ci-dessous.

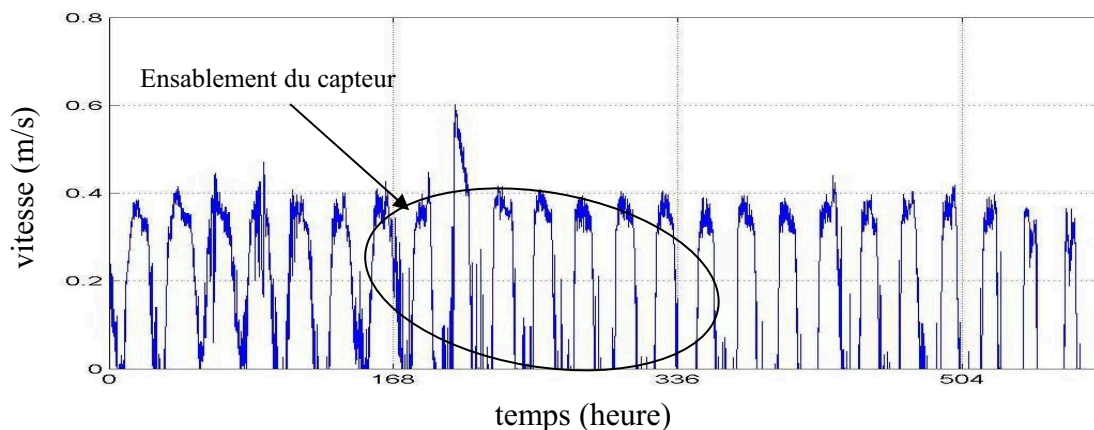


Figure 1.4. Vitesse en fonction du temps fournie par un capteur en état d'ensablement

La figure 1.4 représente un défaut de capteur de mesure de vitesse d'eau dans le réseau. Les mesures considérées sont définies sur la période du 01/09/02 au 25/09/02. Nous remarquons selon la figure, que la vitesse nocturne (valeurs de vitesse les plus faibles) est bruitée au début de cette période, c'est-à-dire, pendant 168 heures. Elle tombe à zéro au fil du temps et la durée où elle est nulle est de plus en plus significative. Cela est remarqué pendant la période allant de 168 à 336 heures. Le phénomène s'explique par l'ensablement progressif du capteur de vitesse. En conséquence, un fonctionnement incorrect du capteur se produit et s'aggrave au cours du temps. La durée du défaut dépend de la durée d'ensablement du capteur. Cette situation ne demande pas nécessairement une intervention de nettoyage.

- Défaut temporaire qui ne nécessite pas d'intervention tel que : encrassement temporaire du capteur (bouchage, macro encrassement).

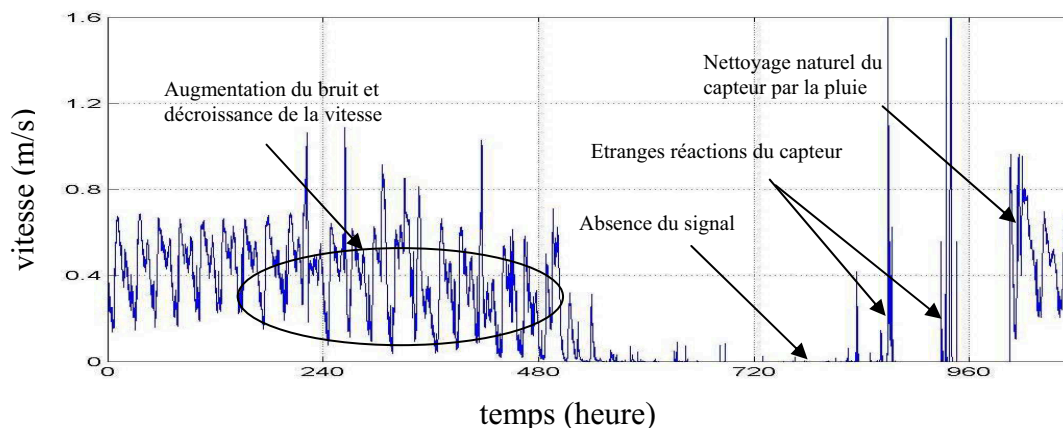


Figure 1.5. Vitesse en fonction du temps fournie par un capteur en état d'encrassement temporaire (bouchage)

A la figure 1.5, nous présentons les mesures effectuées sur la période du 17/11/01 au 27/12/01. Nous remarquons sur la figure ce qui suit :

- une augmentation de bruit et une baisse de la vitesse entre 240 et 480 heures,
- une perte quasi-complète du signal de vitesse à partir de 504 jusqu'à 960 heures,
- une apparition de quelques valeurs de vitesse à la fin du signal sans la moindre intervention.

Les deux premières remarques s'expliquent par l'encrassement progressif du capteur de vitesse. La dernière s'explique par la reprise progressive de fonctionnement du capteur à cause de son nettoyage naturel par les eaux de pluies. Cette situation représente un défaut temporaire qui ne nécessite pas une intervention.

1.5.2 Anomalies sur le processus

Des anomalies peuvent se présenter sur le processus lui-même. Les mesures effectuées sont justes dans le cas où les capteurs ne représentent pas de défauts mais elles ne sont pas représentatives du fonctionnement normal du système. Nous distinguons deux types d'anomalies.

a. Anomalies de conception

Ce sont des anomalies liées aux problèmes posés par des conditions d'écoulement (temporairement) incompatibles avec les hypothèses ayant conduit au choix du capteur, de son emplacement et du traitement des informations qui en sont issues. Cette situation apparaît par exemple si le niveau d'eau est trop faible.

b. Anomalies de représentativité du contexte de mesure

Ce type d'anomalies met en cause la représentativité de la mesure par rapport aux conditions de fonctionnement habituel du réseau.

Pour cette catégorie d'anomalies, sept situations sont possibles :

- entretien du capteur ;
- entretien du réseau ;
- modification de la configuration du réseau, et donc du cheminement des eaux usées. La configuration du réseau peut changer si nous avons par exemple des autorisations

d'effectuer des déversements temporels (augmenter la quantité des eaux déversées dans les milieux naturels) ;

- événements hydrologiques identifiables par des informations extérieures, par exemple : pluie, marée, rejets accidentels dont une source potentielle est identifiée ;
- événements hydrologiques non identifiés, mais dont on a tout lieu de penser que les capteurs en fournissent une image fidèle ; par exemple : rejets accidentels (les déversoirs peuvent s'ouvrir sous certaines conditions inhabituelles non souhaitables (pluie forte, oubli de mise en marche normal l'hors d'intervention d'agent, etc.) ;
- influence aval (qui peut être par exemple un écoulement non libre en aval. Cela fausse les résultats de calcul de débit dans le cas de l'application de la relation de Manning Strickler (qui exige un écoulement libre pour être appliquée) ;
- intrusion du milieu extérieur.

Seul le premier cas de figure relève d'une élimination des résultats dans les analyses, car la période d'entretien d'un capteur est enregistrée. Pour les autres cas, le diagnostic des anomalies pourra être difficile mais sera facilité si nous disposons d'informations sur l'exploitation du réseau. En pratique les anomalies sont détectées sans être pour autant identifiées.

Les deux types d'anomalies liées aux capteurs et au processus, peuvent se présenter alternativement ou simultanément. Nous pouvons donc distinguer les situations suivantes :

- présence des anomalies liées au processus,
- présence des anomalies sur un capteur de mesure,
- présence des anomalies liées au processus et des anomalies sur un capteur de mesure.

Selon le type d'anomalies observées, nous pouvons suspecter certains capteurs pouvant expliquer la présence de ces anomalies. Le tableau 1 récapitule les causes les plus fréquentes par type de capteur.

Origine	Type d'anomalie	Paramètre	pluie	hauteur d'eau			Vitesse		débit				
		Technologie	Auget	piezo	us aériens	us immergés	Doppler	capteurs us	Pompes	électro-magnétique	hauteur + vitesse	f(h)	
Défauts temporaires	Dérive	Dérive du matériel de mesure (capteurs et transmetteurs)		X	X	X	X	X	X	X	X	X	
		Dérive du contexte de la mesure (évolution de section) : ensablement									X	X	
	Encrassement	micro encrassement	X	X		X	X	X			X	X	
		macro encrassement		X			X		X		X	X	
		Ensablement		X			X				X	X	
	événements perturbateurs	Intervention sur capteur	X	X	X	X	X	X	X	X	X	X	
		curages		X	X	X	X	X			X	X	
		arrêts de postes de pompage		X	X	X	X	X	X	X	X	X	
		intrusion de mer ou de rivière		X	X	X	X	X	X	X	X	X	
		Arrosage	X										
		manœuvres de vannes		X	X	X	X	X	X	X	X	X	
		matériel HS	pannes électriques : surtensions (foudre, ...), problème sur alimentation, composants défectueux	X	X	X	X	X	X	X	X	X	X
		pannes « mécaniques » : comportements « agressifs » (vandalisme, rongeurs,...), chocs	X	X	X	X	X	X			X	X	
		corrosion		X	X	X	X	X	X	X	X	X	
	horodatage	problème d'horodatage et de synchronisation	X	X	X	X	X	X	X	X	X	X	
	Défauts pérennes : conception ou de mise en œuvre	Conditions d'écoulement temporairement défavorables	bulles				X		X			X	X
			turbulences					X	X			X	X
			champ de vitesse « atypique »					X				X	
			charge en (matières en suspension) (hors problème d'encrassement)					X				X	
			faibles vitesses					X				X	
fortes vitesses				X									
faibles hauteurs d'eau (y compris émergence de capteur)				X				X	X			X	X
fortes hauteur d'eau					X								
transitoires (stratégie d'acquisition)				X	X	X	X	X	X			X	X
		influence aval										X	
interférences		physiques (hors conditions d'écoulement : pression atmosphérique)		X	X	X	X	X			X	X	
		électriques (parasites, isolation)		X	X	X	X	X		X	X	X	
Paramétrage		conversion de 4-20 en grandeurs physiques		X	X	X	X	X		X	X	X	
		mise en forme d'impulsions	X							X			
		calculs de débits							X	X	X	X	
	stratégie d'acquisition		X	X	X	X	X	X	X	X	X		

Tableau 1. Typologie des sources d'anomalies par nature de capteur

Dans ce tableau, nous présentons dans la première colonne, l'origine d'une anomalie pouvant affecter une mesure et qui peut être une anomalie temporaire ou pérenne. Dans la deuxième colonne, nous indiquons par type d'anomalie rencontrée, la cause de sa présence (dérive, encrassement, etc.) et nous indiquons dans la troisième colonne, les anomalies liées à chaque cause. Nous présentons d'autre part, la grandeur mesurée (pluie, hauteur, vitesse et débit d'eau) ainsi que la technologie utilisée pour obtenir les mesures ou la méthode d'évaluation considérée. La présence d'une anomalie sur une mesure ou une grandeur évaluée est indiquée par une croix dans le tableau.

1.6 Données

1.6.1 Réseau de mesure de Nantes

Le syndicat d'assainissement de Nantes s'est doté en 1997 d'un réseau de points de mesure (voir figure 1.6) permettant la télésurveillance dans le système de collecte des eaux usées. Ce réseau fonctionne depuis 2001, et comporte 25 points de mesure de débit et 8 pluviomètres pour un total de 53 capteurs.

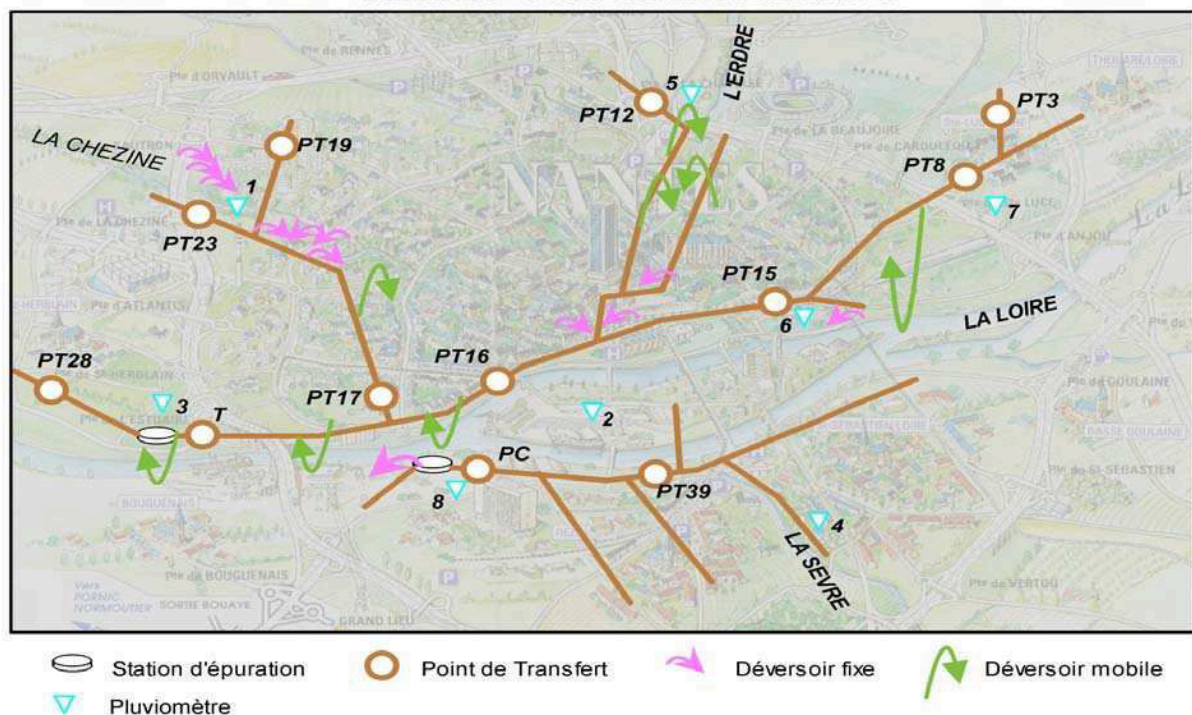


Figure 1.6. Schéma du système d'assainissement de Nantes

Nous avons indiqué par PT un point de transfert. T et PC représentent les deux points de mesures des deux stations d'épurations indiquées sur la figure.

1.6.2 Acquisition et gestion des données sur le système de Nantes

L'acquisition et la gestion des données dans le cas du système de Nantes sont les mêmes que celles décrites dans la section 1.3.6, mais nous utilisons pour la base de donnée indiquée dans la figure 1.3, le logiciel de mesure informatisé pour l'autosurveillance des réseaux (MINAUTOR) et pour le superviseur de télégestion, un ordinateur.

Le pas de temps d'acquisition est fixe pour les points de transfert (5 minutes). Il est variable pour les points de rejet (une heure en l'absence de déversement, cinq minutes au delà d'un seuil de variation de hauteur). Les basculements des augets des pluviomètres sont comptés et l'intensité de la pluie est calculée toutes les 5 minutes.

Dans notre travail, nous utilisons MINAUTOR pour l'exportation des données collectées depuis le système de mesure de Nantes. Ces données sont exportées à des pas de temps variables, elles sont horodatées et enregistrées sous format texte. Elles contiennent des données manquantes dont un code est attribué pour pouvoir les détecter dans une phase de prétraitement ultérieure.

1.6.3 Données disponibles

La station de gestion locale (SGL) permet l'acquisition, l'enregistrement et le transfert des données depuis un point de mesure vers la base de données (qui centralise des SGL(s)). Ces données sont ensuite exportées vers MINAUTOR (voir figure 1.3 qui décrit la stratégie d'acquisition des données et qui a été détaillée dans la section 1.3.3).

Les données disponibles sont :

- Des données issues des capteurs installés sur les points de transfert (notés PT pour abréviation) suivants :

PT 16 : Quai fosse

PT 15 : Duchesse Anne

PT 3 : Chemin du Château

PT 17 : Bouhier

PT 23 : Boulevard du Massacre (St Herblain)

Ces points sont généralement basés sur des technologies de mesures simultanées de la hauteur et de la vitesse d'écoulement.

- Des données issues des capteurs installés sur les déversoirs. Ces déversoirs peuvent être équipés de capteurs de hauteurs pour les déversoirs mobiles ou de simple détection de surverse en mode tout ou rien pour les déversoirs fixes. Ces déversoirs mobiles ou fixes sont notés DO et DF respectivement. Comme ils sont installés à proximité d'une rivière, ils peuvent être cités selon les rivières comme suit :

Loire

DO 26 : Cordon Bleu (déversoir mobile)

DO 27 : Quai Fosse (déversoir mobile)

DO 30 : Broutelle (déversoir mobile)

Erdre

DO 31 : St Mihiel (déversoir mobile)

DO 25 : St Mihiel (déversoir mobile)

DF 112 : St Mihiel (déversoir mobile)

DF 104 : Duchesse Anne

L'ensemble des grandeurs mesurées ou évaluées correspondant à chaque point de mesure est récapitulé dans le tableau de l'Annexe A.

1.6.4 Données utilisées

Les données utilisées sont des grandeurs mesurées au niveau de chaque point de mesure. Dans le cas des points de transfert d'eau, il s'agit des hauteurs, des vitesses et des débits d'écoulement d'eau dans le réseau. Dans le cas des points déversoirs d'orages, il s'agit des hauteurs d'eau seulement. Ces données sont utilisées avec une période d'échantillonnage d'une minute.

1.7 Conclusion

Les mesures effectuées sur les réseaux d'assainissement sont difficiles à obtenir en raison des perturbations aléatoires souvent présentes.

Dans le domaine de l'assainissement, nous avons présenté les différentes grandeurs mesurées et les emplacements des points de mesure, ce qui nous permettra de voir comment les mesures sont liées entre elles d'un point de mesure à un autre selon leurs variations.

Nous avons présenté le réseau de mesure de Nantes qui est le domaine d'application de notre travail de recherche. Nous avons pris en considération les méthodes et les techniques d'application ainsi que les méthodes et les techniques permettant les mesures des hauteurs, des vitesses et des débits d'eau. Nous avons aussi classifié les différents types d'anomalies pouvant affecter ces mesures.

Ce chapitre nous a permis de présenter :

- Le réseau de mesure;
- les grandeurs mesurées et celle évaluée;
- les types de capteur utilisé pour obtenir ces mesures et les différents types d'anomalies pouvant affecter les mesures.

Chapitre 2 Généralités sur les méthodes de validation des données de mesure

Sommaire

2.1	Introduction	38
2.2	Techniques de validation des données	39
2.3	Méthode de validation des données de mesure à base de modèles analytiques	46
2.4	Généralités sur le diagnostic des systèmes.....	47
2.5	Conclusion.....	50

2.1 Introduction

Dans ce chapitre, nous présentons tout d'abord, un descriptif des techniques de validation des données de mesures. Puis nous décrivons celle que nous avons choisie, pour la validation des données issues du réseau d'assainissement de Nantes. Comme le diagnostic de défauts de mesures représente une étape nécessaire dans la procédure de validation sollicitée, nous avons consacré une partie à la fin de ce chapitre pour exposer des généralités sur le diagnostic et des notions utiles pour notre étude.

La validation ou la critique des données en assainissement est effectuée essentiellement par comparaison du résultat d'une ou plusieurs mesures avec d'autres informations liées aux phénomènes mesurés. Les connaissances et informations utilisées sont de nature très diverse.

- a. Connaissance physique générale sur la grandeur mesurée (valeurs possibles, vraisemblables, vitesse de variation) et sur le capteur de mesure (étendue des valeurs mesurables, stratégie d'échantillonnage etc.);
- b. connaissance des mesures ponctuelles et simultanées de la même grandeur par un autre capteur que le capteur habituel;
- c. résultats issus des modèles construits en se basant sur le principe de la redondance de l'information.

Dans le cadre de cette thèse, nous nous sommes intéressés plus particulièrement aux informations de type (c), en distinguant deux niveaux de validations [Mourad et *al.*, 05] :

- Un premier niveau de validation consiste à s'assurer de la cohérence du signal issu de chaque capteur. Il a pour but de détecter les résultats manifestement aberrants. Il est basé sur une analyse « locale » des mesures fournies par un capteur. Il permet donc de déceler des dysfonctionnements de capteur relativement « francs ». Les anomalies qui se présentent sur les mesures qui peuvent être mises en évidence par ce premier niveau correspondent, par exemple, à des dépassements d'échelle ou des défauts de capteurs. Les algorithmes classiques de recherches automatiques d'anomalie permettent de détecter les mesures trop aberrantes : dépassement d'un seuil par exemple [Perrault et al. 91], mais sont souvent mis en échec par des anomalies moins évidentes, lorsque les mesures sont entachées d'erreurs de faibles amplitudes.
- Un deuxième niveau de validation repose sur une analyse de la cohérence globale des données à plus large échelle (relations débits-débits, vitesses-hauteurs, hauteurs-hauteurs). L'utilisation des modèles traduisant cette cohérence permet d'analyser l'ensemble des signaux fournis par les capteurs en même temps. Ce deuxième niveau a pour but de distinguer les

résultats de mesure qui ont de bonnes chances de représenter fidèlement la réalité de ceux qui sont suspects.

2.2 Techniques de validation des données

La validation des données hydrométriques est un sujet relativement peu étudiée (par rapport aux publications scientifiques et techniques peu nombreuses). Plusieurs expériences concernant les données pluviométriques et débitmétriques ont été analysées. Les techniques utilisées pour la validation des données sont souvent manuelles avec, dans peu de cas, une assistante automatique très limitée.

Des méthodes générales existent pour la validation de données telles que les tests statistiques, la théorie de décision, le filtrage des signaux [Ragot 90; Barnett et Lewis 90 ; Brunet 90; Bennis 95] et les techniques de redondance [Brunet 90] qui offrent de nombreuses méthodes pour la détection des défauts et de valeurs aberrantes dans les séries de données. Ces techniques rencontrent un problème d'adéquation entre les modèles théoriques proposés et les contextes rencontrés en pratique. Le choix d'une technique dépend fortement de la nature des données et des défauts à détecter.

Les travaux de [Bennis et *al.*, 96; Berrada et *al.*, 96] présentent d'une part l'adaptation à l'hydrologie de méthodes de filtrage et de validation utilisée traditionnellement dans d'autres disciplines. D'autre part, elles proposent de nouvelles méthodes univariées et multivariées de validation des données.

La méthode proposée par [Berrada et *al.*, 96], traite le sujet de la validation des données hydrologiques historiques. Elle permet d'utiliser plusieurs séries historiques réelles relatives aux mesures des niveaux d'eau dans un réservoir, des pluies et des débits d'entrée et de sortie. L'objectif principal de la méthode est de valider les données de pluie. Cette méthode utilise plusieurs techniques de filtrage (filtres : médian, fréquentiel, Fraser¹, ébarbeur winsoriseur, de Kalman) pour valider les données hydrométriques. Ces techniques sont appliquées sur des séries de données journalières simulées avec introduction de bruits et sur des données réelles (hauteurs d'eau dans les réservoirs, débits et pluies). Les résultats sont ensuite comparés à l'aide de critères qualitatifs et quantitatifs. Ces techniques de filtrage utilisées dépendent du type de données traitées, de l'amplitude et du nombre d'erreurs qui les affectent. Ainsi, pour

¹ Le filtre de Fraser est basé sur un modèle de type ARMA (Autoregressif à moyenne mobile [Box et Jenkins, 76]) couplé au filtre de Kalman [Radix, 84].

valider les niveaux à variation lente des grands réservoirs, le filtre de Fraser est approprié pour détecter les valeurs aberrantes isolées (quand le coefficient d'autocorrélation est faible, le modèle autorégressif est non approprié et le filtre de Kalman diverge). Dans le cas des séries de données de hauteurs d'eau et de pluie très bruitées, les filtres fréquentiel, médian, ébarbeur winsoriseur, linéaire et quadratique sont appropriés.

- Filtre linéaire

Un filtre à régression linéaire sur une fenêtre mobile, permet d'approximer par une droite, la tendance de $2l + 1$ mesures successives représentant la largeur de la fenêtre ; l est un entier positif choisi a priori.

- Filtre quadratique

Un filtre à régression quadratique permet d'approximer la tendance sur les $2l + 1$ points représentant la largeur de la fenêtre mobile par un polynôme du second degré; l est un entier positif choisi a priori.

- Filtre ébarbeur winsoriseur

Ce filtre conserve les qualités de la médiane, comme estimateur d'un paramètre central d'une population, plus robuste que la moyenne. Ce filtre comme son nom l'indique effectue deux opérations. La première est l'ébarbage « trimming » qui consiste à exclure un certain nombre d'observations les plus éloignées de la médiane, pour caractériser un échantillon. La deuxième opération est le « winsorizing » qui permet la reconstitution des valeurs exclues par les valeurs les plus proches conservées.

Le but est de détecter les valeurs aberrantes et de les remplacer par une valeur « raisonnable ».

Ebarbage :

Soit le signal mesuré y , et soit l'échantillon $Y = [y_{-l}, y_{-l+1}, \dots, y_{+l-1}, y_l]$, où $2l + 1$ représente la taille d'une fenêtre mobile et l est un entier positif choisi a priori. Le filtrage est effectué en deux étapes. Lors de la première, l'échantillon Y est classé selon un ordre croissant. Dans la deuxième étape, les valeurs aberrantes doivent être éliminées. Cette élimination se fait suivant un rejet a priori où le nombre de points à rejeter est fixé à l'avance. Pour écarter systématiquement de l'échantillon un certain pourcentage a (du nombre total

$2l + 1$) de premières valeurs et un autre pourcentage b (du nombre total $2l + 1$) de dernières valeurs, les nombres exacts des valeurs exclus seront :

$$r = \text{ent}((2l + 1)a), \quad (2.1)$$

pour les plus faibles valeurs de l'échantillon et

$$s = \text{ent}((2l + 1)b), \quad (2.2)$$

pour les plus fortes valeurs de l'échantillon,

avec $\text{ent}(z)$ la partie entière du nombre réel z .

Winsorizing :

Pour une fenêtre mobile les proportions a et b sont prises égales (filtre symétrique) avec une reconstitution des points sans moyennage. Ainsi pour une fenêtre donnée, les r points les plus faibles sont simplement remplacés par la valeur immédiatement supérieure et les r points les plus forts par la valeur immédiatement inférieure. Le but est de repérer ‘automatiquement’ les valeurs aberrantes et les remplacer par une valeur ‘raisonnable’.

-Filtre passe-bas

Ce filtre est bien approprié en hydrologie pour éliminer les fluctuations rapides. Il est peu coûteux en temps de calcul. Notons que la nécessité d'effectuer un tel filtrage dépend des résultats obtenus après le filtrage « trimming », suivant l'existence ou pas des fluctuations rapides.

Soit le signal mesuré $y(k)$, où k représente le temps. Le signal filtré $y_f(k)$ en utilisant un filtre passe-bas d'ordre un est donné par la relation suivante :

$$y_f(k) = \alpha y_f(k - 1) + (1 - \alpha)y(k), \quad (2.3)$$

avec $0 \leq \alpha < 1$.

L'application de tous les filtres sur une série de données perturbées artificiellement, donne de bons résultats. Certains résultats sont plus performants que d'autres selon les caractéristiques du signal et du bruit, mais cela demande la connaissance de toutes les caractéristiques de données synthétiques ce qui limite cette méthode.

La validation en temps réel des données des hauteurs d'eau et des débits dans les systèmes hydriques est proposée dans [Berrada et *al.*, 03]. Cette méthode est basée sur la redondance matérielle et analytique et utilise deux modèles :

- 1- Le premier, spatial, permet de relier à l'aide d'une équation de régression multiple, les données de mesures fournies par différents capteurs.
- 2- Le second, temporel, permet de déterminer à l'aide d'un modèle autorégressif, les tendances de variations données par les différents capteurs.

Ces deux modèles sont incorporés dans un système de diagnostic de pannes basé sur un principe de vote logique. Parmi les valeurs consistantes avec la tendance de variation indiquée par le modèle de régression linéaire, la plus consistante avec la tendance de variation indiquée par le modèle autorégressif est retenue. Le filtre de Kalman est utilisé pour filtrer les mesures et identifier en temps réel les paramètres des modèles utilisés. La méthodologie proposée s'est révélée concluante aussi bien sur les données réelles mesurées que sur des données synthétiques.

La méthode de détection de défauts dans les réseaux d'assainissement par combinaison du filtre de Kalman et le test de rapport de probabilité séquentiel binaire proposée par [Piatyszec et *al.*, 00] est applicable en temps réel sur le réseau d'assainissement et en temps de pluie. Elle consiste à obtenir une estimation du débit considéré, en utilisant l'estimateur d'état « filtre de Kalman », de calculer le débit par un modèle déterministe et de calculer une « innovation » qui représente la différence entre la mesure et l'estimation. L'« innovation » est une évaluation de la capacité du filtre de suivre le comportement observé. Quand le système est dans un état de comportement anormal, l'« innovation » dévie significativement de zéro et elle est presque nulle dans le cas contraire. Cette innovation est considérée comme une variable aléatoire gaussienne définie par son écart type et sa moyenne qui est nulle dans le cas de fonctionnement normal du système et différente de zéro dans le cas de son fonctionnement anormal. Deux hypothèses H_0 et H_1 sont considérées respectivement pour les deux cas. La détection des anomalies considère le changement de la moyenne du processus aléatoire. Le test de rapport de probabilité séquentiel de Wald est appliqué. Une fois un défaut est détecté sur un point donné du réseau, une méthode de diagnostic permettant de déterminer la cause de la présence d'anomalie est appliquée. Cette méthode considère deux nœuds I et J où le débit est évalué et entre lesquels il se propage, une anomalie se présente au niveau du

nœud I qui se trouve en amont du nœud J . La méthode permet d'abord de calculer la somme des logarithmes de probabilités appelée « vraisemblance calculée ». Cette somme quantifie la capacité du filtre à prédire ou suivre le débit calculé (à partir du modèle mentionné ci-dessus) au nœud J , pour une période donnée, considérant la propagation du débit calculé entre les nœuds I et J . Ensuite, la somme des logarithmes de probabilités appelée « vraisemblance mesurée » est calculée. Cette somme quantifie la capacité du filtre à prédire ou suivre le débit évalué (à partir des mesures effectuées) au nœud J , pour une période donnée, considérant la propagation du débit évalué entre les nœuds I et J . Finalement, la somme la plus significative permet d'identifier les causes de l'anomalie qui se présente au niveau du nœud I .

Cette méthode donne de bons résultats durant l'analyse de plusieurs pluies sur le réseau de mesure de Saint Denis (France). L'application de cette méthode est limitée par la complexité d'application en temps réel. En plus, elle demande la connaissance des conditions de fonctionnement du système et la détection rapide du comportement anormal du réseau ou /et les mesures erronées.

Le problème de la validation des données de débits a été traité par [Nguyen et Bisson 98], afin de fournir les valeurs les plus fiables pour la gestion optimale et sécuritaire des systèmes de ressources hydriques. Des procédures de validation des débits basées sur les méthodes de régression et sur la technique de filtrage passe bas (de type : filtrage exponentiel) des séries de données sur des périodes bien définies sont élaborées et comparées. Les résultats de l'application de ces méthodes à un bassin versant bien défini ont permis de conclure que la méthode basée sur le filtrage exponentiel est la plus appropriée. De plus, cette méthode de filtrage exponentiel proposée ne contient qu'un seul paramètre dont l'estimation se fait à l'aide d'une méthode pratique et efficace.

Le travail proposé par [Bennis et Kang 00] développe une méthode automatisée pour la validation des séries de données des débits d'eau dans les réservoirs. Cette méthode permet de valider et de corriger le niveau d'eau dans le réservoir selon le nombre de capteurs installé dans le réservoir. Quand le réservoir est équipé avec seulement un capteur, le filtrage univarié est utilisé voir [Berrada et *al.*, 96]. Par exemple, le filtrage fréquentiel élimine systématiquement toutes les variations correspondantes aux fréquences supérieures à la fréquence seuil prédéterminée. Dans le cas de plusieurs capteurs dans le même réservoir, le filtrage multivarié est utilisé pour valider les séries de données du niveau d'eau pour chaque capteur. Le principe de base de cette méthode implique de donner plus de poids pour les

capteurs qui fournissent les valeurs les plus consistantes avec les valeurs précédentes et antécédentes validées dans une série de mesure donnée. Ainsi, les variations isolées, positives ou négatives observées dans les séries de données considérées seront éliminées dans le cas où les valeurs correspondantes pour d'autres capteurs produisent des variations plus consistantes. Pour analyser le degré de persistance spécifique à une valeur donnée, cette valeur est comparée à une valeur prédite obtenue à partir d'un modèle autorégressif établi à partir des données précédemment validées. En pratique cette méthode de filtrage compare, d'une part, la différence entre la valeur estimée par le modèle autorégressif et la valeur historique mesurée, d'une autre part, la différence entre la valeur estimée par le même modèle autorégressif et la valeur estimée par le modèle régressif pour d'autres capteurs. Le modèle régressif est un modèle de régression linéaire avec des coefficients variables décrits dans [Bennis et al., 97]. Parmi les valeurs mesurées et celles estimées par le modèle régressif, celle qui est proche de la valeur estimée par le modèle autorégressif est retenue. Si l'on dispose de M valeurs successives pour valider la valeur de la hauteur d'eau dans le réservoir à un instant donné k , on utilise les valeurs de $k - 1$ à 1 et de $k + 1$ à $k + M$ fournies par le même capteur donnant la valeur à valider et par les capteurs voisins. Cela est effectué en appliquant deux fois la méthode de validation (filtrage multivarié), c'est-à-dire un aller-retour dans le temps par rapport à la valeur considérée. Pour un aller, le filtrage est appliqué du jour 1 au jour k . Pour un retour, le filtrage est appliqué du jour M au jour k . La valeur historique considérée au temps k est rejetée seulement si elle est mise à l'écart dans les deux applications de validation aller-retour. Cette méthode est du premier niveau de validation et est applicable localement.

Dans le cadre du projet OTHU (Observatoire de Terrain en Hydrologie Urbaine), une procédure automatique d'aide à la validation de données a été développée et proposée par [Mourad et Bertrand-Krajewski 01]. Elle concerne la critique et la validation de données environnementales en hydrologie urbaine. Cette procédure consiste à appliquer, en temps différé, un ensemble de sept tests (portant sur les valeurs, les gradients, les corrélations entre grandeurs, etc. qui sont mentionnés par la suite). Ces tests sont appliqués sur les séries de données afin de détecter les valeurs qui semblent fausses ou douteuses. Cette étape de pré-validation attribue à chaque donnée et pour chaque test trois notes partielles A, B, C qui représente respectivement « donnée fiable, donnée douteuse et donnée non valide ». Ces tests considèrent les critères suivants :

1. état du fonctionnement du capteur qui permet de savoir si le capteur est en fonctionnement ou hors fonctionnement. Ce repérage d'état du capteur est lié à une valeur bien définie du courant électrique (généralement égale à zéro) délivré lors de mise hors fonctionnement du capteur,
2. étendue physique de la mesure du capteur par exemple pour un capteur Doppler, nous avons (0-2m/s),
3. étendu réaliste locale qui représente un intervalle dans lequel se situent habituellement les valeurs de la grandeur mesurée (avec un niveau de confiance de 95%),
4. délai de maintenance : une valeur enregistrée dans des périodes de maintenance ou après maintenance d'un capteur donné est juger bonne, douteuse ou mauvaise,
5. vitesse de variation de la grandeur mesurée : chaque capteur possède un intervalle de vitesse de variations plus ou moins défini et stable. La fixation des bornes de cet intervalle est liée à la détection des défauts. On applique un filtre moyeneur aux variations du signal considéré puis on calcule la différence entre ces variations et le résultat du filtrage. Cette différence représente un résidu (signal à moyenne nulle) qui est un indicateur de variations anormales. Le choix d'un seuil permet de repérer les instants où les valeurs du résidu s'écartent significativement de sa moyenne. Les valeurs mesurées correspondantes à ces valeurs reçoivent une note B,
6. redondance matérielle où l'on compare les résultats de mesures obtenus par deux capteurs. Le résidu obtenu est mis sous test de Page. Les valeurs détectées comme anormales reçoivent une note B,
7. redondance analytique où l'on compare deux grandeurs corrélées (pluie-débit, hauteur et vitesse de l'écoulement), en utilisant un modèle empirique pour cette application. Le test de Page est appliqué par la suite et la note B est attribuée aux valeurs suspectes.

Une note globale est ensuite attribuée à chaque donnée, qui correspond à la note partielle la plus mauvaise. La validation finale est effectuée par un opérateur expert assisté par un outil de visualisation graphique automatique mettant en évidence les données notées B ou C détectées par la pré-validation automatique. Ils permettent cependant de repérer la plupart des défauts et des valeurs douteuses ou aberrantes avec un faible taux de fausse alarme. Cette méthode de validation appartient à un premier niveau de validation et ne considère pas la cohérence

globale entre les différentes mesures. Les tests et les critères de cette méthode de validation des données sont relativement simples.

Le travail proposé dans [Boukhris, 98] traite le problème de la validation des données de pluie fournies par des pluviomètres installés à l'entrée d'un bassin versant et des hauteurs d'eau dans le bassin. Les hauteurs d'eau permettent d'évaluer des débits et les données pluie permettent d'obtenir des débits estimés en utilisant un multi-modèles.

Les tolérances sur les hauteurs d'eau sont fixées, alors que celles correspondantes aux débits peuvent être calculées.

La première méthode utilisée consiste à comparer les débits deux à deux pour obtenir des différences qui représentent des résidus. La détection des défauts de mesures est effectuée en comparant ces résidus aux tolérances calculées pour le débit. La localisation des défauts est effectuée par consultation de tous les résidus en supposant que l'on ne peut avoir de défauts multiples.

Une deuxième méthode consiste à utiliser l'approche espace de parité [Ragot et *al.*, 90] pour les données de débits mentionnées ci-dessus. Cette méthode permet de localiser le débit affecter par des défauts.

2.3 Méthode de validation des données de mesure à base de modèles analytiques

Une façon efficace pour valider les mesures est de multiplier les chaînes de mesure. Cette redondance des mesures est dite matérielle. Dans le cas de deux capteurs qui mesurent une même variable, une différence entre les deux mesures supérieure à un certain seuil, est un indicateur de présence de défaut. Lorsque plusieurs capteurs sont utilisés le capteur en défaut peut, de plus, être désigné. Cette redondance est utilisée dans les domaines à haut niveau de risque. Cette méthode est fiable [Berrada et *al* 96 ; Bennis et Kang 00]. Malheureusement, compte tenu des contraintes budgétaires, la métrologie des réseaux d'assainissement urbains reste encore très limitée. Ainsi, la redondance matérielle sur laquelle reposent les méthodes précitées est quasiment inexistante et dans le meilleur des cas, une seule station de mesure existe à un endroit donné du réseau.

Notre travail consiste en la validation des données de mesure selon une technique de diagnostic hors ligne. Ce diagnostic permet la détection et la localisation des défauts qui peuvent se présenter aux niveaux des capteurs de mesure de hauteurs, de vitesse et de débits

d'écoulement d'eau dans le réseau d'assainissement. Cette technique est basée sur l'utilisation de modèles qui permettent de produire des relations de redondance de l'information.

La méthode de validation, comme indiqué dans le schéma de la figure 2.1, comporte selon sa définition donnée ci-dessus, deux étapes nécessaires et successives suivantes :

- Etablissement des modèles ;
- Diagnostic des défaillances.

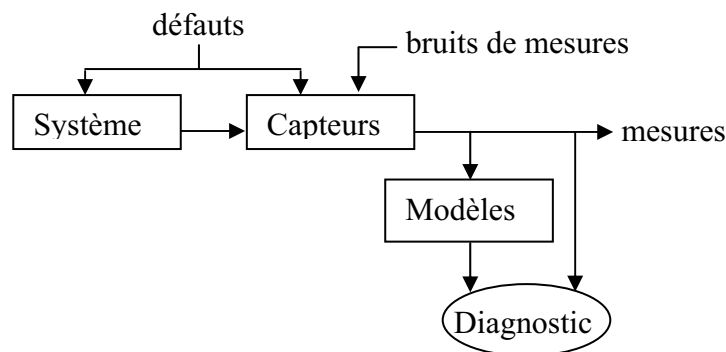


Figure 2.1. Scéma descriptif des étapes nécessaires pour la validation

Il est essentiel pour la suite de ce mémoire d'utiliser des définitions précises pour le diagnostic, les défaillances, les défauts, et les anomalies.

2.4 Généralités sur le diagnostic des systèmes

Dans les disciplines de l'automatique et du traitement du signal, plusieurs auteurs [AFNOR, 88 ; Zwingelsein, 95 ; Basseville et Cordier, 96 ; Isermann et Ballé, 97 ; Van Schrick, 97] ont permis d'extraire les principaux concepts et de définir les principes du diagnostic. A partir de ces références, nous avons considéré quelques notions liées au diagnostic.

Notion de diagnostic :

Nous avons considéré la définition retenue par les instances nationales et internationales de normalisation (AFNOR, CEI) : « Le diagnostic est l'identification de la cause probable de la (ou les) défaillance(s) à l'aide d'un raisonnement logique fondé sur l'ensemble d'information provenant d'une inspection, d'un contrôle ou d'un test ».

Cette définition fait apparaître les deux éléments essentiels en diagnostic :

- observation des symptômes de la défaillance,

- identification de la cause de la défaillance à l'aide d'un raisonnement logique fondé sur les observations.

Le diagnostic est une tâche essentielle de la surveillance du fonctionnement des systèmes. Il consiste à détecter un fonctionnement anormal au sein du système et à déterminer sa cause en localisant le ou les composants du système présentant une anomalie de fonctionnement et éventuellement en caractérisant l'anomalie (instant d'apparition, durée, etc.). Le diagnostic commence par une comparaison entre le comportement et le fonctionnement théorique attendu fourni par un modèle.

Un modèle d'un système est une description de sa structure physique et une représentation comportementale et/ou fonctionnelle, abstraite de chacun de ces composants [De Kleer et Williams, 87].

Notion de défaillance :

De nombreuses définitions ont été proposées dans la littérature, nous avons considéré celle proposée par l'AFNOR [AFNOR, 88] et repris dans [Zwengelstein, 95] :

Une défaillance est « l'altération ou la cessation de l'aptitude d'un ensemble à accomplir sa ou ses fonctions requise(s) avec les performances définies dans les spécifications fonctionnelles du matériel ». Un ensemble défaillant présente donc une altération fonctionnelle.

Notion du défaut :

Un défaut se rapporte à une anomalie de comportement au sein d'un système physique. Il est souvent défini comme étant un écart entre la caractéristique observée et une caractéristique de référence [AFNOR, 94 ; Isermann et Ballé, 97].

La notion de défaut et celle de défaillance sont voisines, la première est liée au comportement et la deuxième est liée à la fonction. Selon le dictionnaire IEEE [Jay, 88], un défaut n'implique pas forcément une défaillance car il peut ne pas altérer le fonctionnement d'un système mais peut éventuellement présager d'une défaillance.

Notion d'anomalie :

Une anomalie désigne tout fonctionnement qui s'éloigne de ce qui est considéré comme normal.

Le diagnostic nécessite un ensemble de tests permettant de révéler les incohérences entre les observations et le modèle, signes de défauts ou symptômes de défaillances.

L'objectif est ensuite de localiser le ou les composants ayant un comportement ou un fonctionnement anormal à partir des incohérences révélées par les tests de cohérence.

Pour cela, un ensemble d'hypothèses sur l'origine possible des anomalies est généré. Chacune des hypothèses est formulée en terme de changements dans la structure, le comportement ou le fonctionnement du système. La localisation consiste à rechercher les hypothèses permettant d'expliquer l'ensemble des incohérences observées.

Pour la génération des hypothèses sur l'origine des défaillances et la réalisation du diagnostic, Chatain dans son travail [Chatain, 93] distingue les deux approches suivantes :

- *Approche orientée bon fonctionnement* :

Dans cette approche seule une description précise du fonctionnement normal du système est nécessaire.

Le diagnostic est effectué en vérifiant tout d'abord la cohérence des observations avec le modèle de bon fonctionnement et en extrayant les symptômes.

- *Approche orientée mauvais fonctionnement* :

Dans cette approche, le raisonnement s'appuie sur une liste des symptômes, une liste des défaillances, et sur des relations entre défaillances et symptômes. Le diagnostic s'effectue par un raisonnement inductif, par l'inversion des relations défaillances-symptômes. Les symptômes observés sont utilisés pour invoquer directement une ou plusieurs hypothèses de défaillance spécifiques à partir des relations établies.

En pratique, le diagnostic peut faire appel aux deux approches. L'approche orientée bon fonctionnement peut être utilisée dans un premier temps, pour la détection globale d'une anomalie au sein du système, la localisation étant assurée ensuite par une approche orientée mauvais fonctionnement.

Les principes du diagnostic étant précisés, nous présentons par la suite les approches à base de modèles analytiques pour la détection et la localisation de défauts.

Le diagnostic à base de modèles s'appuie uniquement sur les données d'un modèle de fonctionnement normal du système et procède par comparaison des comportements du modèle et des observations, traduisant l'état réel du système afin de détecter des incohérences et de déduire les causes de dysfonctionnement.

Les méthodes de diagnostic à base de modèles analytiques sont souvent désignées dans la littérature par « analytical redundancy-based FDI » (pour Fault Detection and Isolation).

Ces méthodes sont développées principalement dans les disciplines de l'automatique et du traitement du signal. Les études menées dans [Gertler, 88 ; Basseville, 88 ; Frank, 90 ; Patton, 97 ; Chen et Patton, 99 ; Russel et *al.*, Frank et *al.*, 00] présentent des références de synthèse des progrès dans ces méthodes. Par exemple dans le travail de Frank et *al.*, une discipline détaillée des différentes approches à base de modèle analytique est présentée. Un autre travail qui est plus représentatif de toutes les techniques disponibles pour le diagnostic des défaillances a été présenté dans [Kramer et Mah, 93]. Il concerne la validation des données, rectification et les problèmes du diagnostic des défaillances.

2.5 Conclusion

Dans ce chapitre, nous avons défini dans un premier temps, la validation des données et nous avons présenté deux niveaux distincts de validation, dont le deuxième est celui que nous développons dans cette étude. Ensuite, nous avons dressé un état de l'art sur les techniques de validation des données de mesures, en particulier les données débitométriques. Toutes ces méthodes de validation citées sont généralement des méthodes appliquées sur des banques de données de mesures réelles. A la fin, nous avons présenté les étapes nécessaires comprises dans la méthode de validation, que nous avons adoptée et nous avons exposé des généralités sur le diagnostic des systèmes ainsi que les notions utiles pour la suite de ce travail.

Chapitre 3 Modèles

Sommaire

3.1	Introduction	52
3.2	Présentation des données utilisées.....	53
3.3	Prétraitement des données	56
3.4	Analyse des données	57
3.5	Etablissement des modèles.....	59
3.6	Structure des modèles.....	62
3.6.1	Choix de la structure du modèle.....	64
3.6.2	Sélection des variables explicatives	65
3.7	Identification paramétrique par la méthode des moindres carrés ordinaire	70
3.8	Validation des modèles	70
3.8.1	Evaluation de l'écart type de l'erreur et du coefficient de détermination.....	71
3.9	Conclusion.....	76

3.1 Introduction

Un modèle est un ensemble de relations mathématiques qui présente, au moins dans un certain domaine de fonctionnement, un comportement identique à celui du procédé réel. Il peut être de connaissance ou de représentation [Walter et Pronzato, 94 ; Ben Abdennour et *al.*, 01].

Le modèle de connaissance est basé sur les lois de la physique. Bien qu'il offre une description précise facilitant la compréhension du comportement des processus réels, ce type de modèle est souvent complexe et peu utilisé en automatique. En fait, pour la majorité des processus réels, ce modèle est souvent difficilement applicable et inexploitable pour leur surveillance. Par ailleurs, ces processus sont le plus souvent non linéaires, mais peuvent être fréquemment approximés avec une précision suffisante par des modèles linéaires.

Le modèle de représentation est obtenu expérimentalement en exploitant des mesures par des méthodes d'identification ou d'apprentissage. Il est peu apprécié par les physiciens car ses paramètres n'ont aucune signification physique.

Le processus étudié est dynamique, les résultats de mesure sont discrets, et des corrélations significatives existent entre plusieurs variables. Nous nous sommes donc intéressés aux modèles dynamiques, discrets et linéaires. Ces modèles paramétriques nécessitent une phase d'identification. Parmi les méthodes d'identification des paramètres des modèles [Walter, 94,97 ; Landau et Besançon-Voda, 01 ; Ben Abdennour et *al.*, 01], nous avons considéré celle des moindres carrés adaptée aux modèles linéaires.

Le réseau de surveillance implanté à Nantes permet de fournir un ensemble de mesures qui sert, après un prétraitement [Ragot et *al.*, 90 ; Bennis et *al.*, 96 ; Berrada et *al.*, 96] complété par une validation préalable par expertise [Aumond et Joannis, 01 ; Aumond et *al.*, 03], à construire des modèles de références. Ces modèles permettent de produire des relations de redondance de l'information qui réduisent ou évitent le besoin de la redondance matérielle [Bennis et Kang, 00 a] et permettent d'expliquer des variables représentant les mesures de type hauteurs et des vitesses d'eau dans le réseau d'assainissement, en fonction d'autres variables de même type. Ces modèles sont utilisés dans une procédure de détection et de localisation des défauts de mesures dans le réseau.

Dans ce qui suit, nous présentons tout d'abord, l'ensemble de données, leur prétraitement et leur analyse. Ensuite, nous exposons la méthode d'obtention des modèles, la technique de sélection des variables explicatives et la procédure d'identification des paramètres de ces modèles. Puis, nous présentons une méthode de validation de ces modèles en calculant leur pouvoir explicatif.

3.2 Présentation des données utilisées

Les données que nous avons utilisées sont les résultats de mesure issus du réseau d'assainissement de Nantes. Ces résultats représentent les mesures de hauteurs d'eau, de vitesses et de débits dans le réseau. Un prétraitement présenté dans la section 3.3 doit être appliqué à ces données avant leur utilisation.

Nous avons établi un tableau récapitulatif (tableau 3.1) de l'ensemble des stations de mesure utilisées dans le réseau, de l'ensemble de points de mesure, et de la variable représentative du résultat de mesure obtenu par chaque capteur.

Pour plus de souplesse dans les notations, nous n'avons pas indiqué dans le tableau donné ci-dessus et dans tous les tableaux donnés par la suite, la variable temps.

Station de mesure	Point de mesure	Variable	
Quai fosse	Do27	h_1	
		h_2	
	Pt16	h_4	
		v_5	
		Q_6	
		h_7	
Saint Mihiel	Do25	h_{10}	
		h_{11}	
	Df112	h_{12}	
		Do31	h_{14}
			h_{15}
Broutelle-Guai Robert	Do30	h_{18}	
Cordon bleu	Do26	h_{20}	
		Q_{21}	
Bouhier	Pt17	h_{22}	
		Q_{24}	
Bd du Massacre	Pt23	Q_{30}	
Chemin du château	Pt3	Q_{36}	
Duchesse Anne	Pt15	h_{37}	
		h_{40}	
	Df104	h_{40}	
Pluviomètre Duchesse Anne	P02	Pl_2	

Tableau 3.1. Capteurs et signaux correspondants sur le réseau d'assainissement de Nantes

Nous rappelons que, les données fournies par des capteurs installés dans des points différents pas éloignés les uns aux autres, sont acquises dans un seul point appelé point de mesure. Ces capteurs ne représentent pas de redondance matérielle mais ils sont référencés à un même point de mesure.

Deux capteurs qui ne sont pas éloignés géographiquement ne mesurent pas les mêmes quantités car les deux capteurs sont installés selon trois cas, comme suit :

1. dans deux canalisations différentes,
2. dans la même canalisation, un premier capteur installé en entrée, un deuxième en sortie où le débit d'eau est plus bas à cause d'un déversement, ou une rupture de pente,

3. dans la même canalisation qui est séparée par un siphon. Ce dernier crée une différence de niveau d'eau à ces extrémités.

Nous avons représenté un schéma (figure 3.1) du réseau de mesure où nous avons indiqué l'ensemble des capteurs considéré dans le tableau 3.1. Les termes : transfert et déversement indiqués sur la figure 3.1, sont définis dans le premier chapitre.

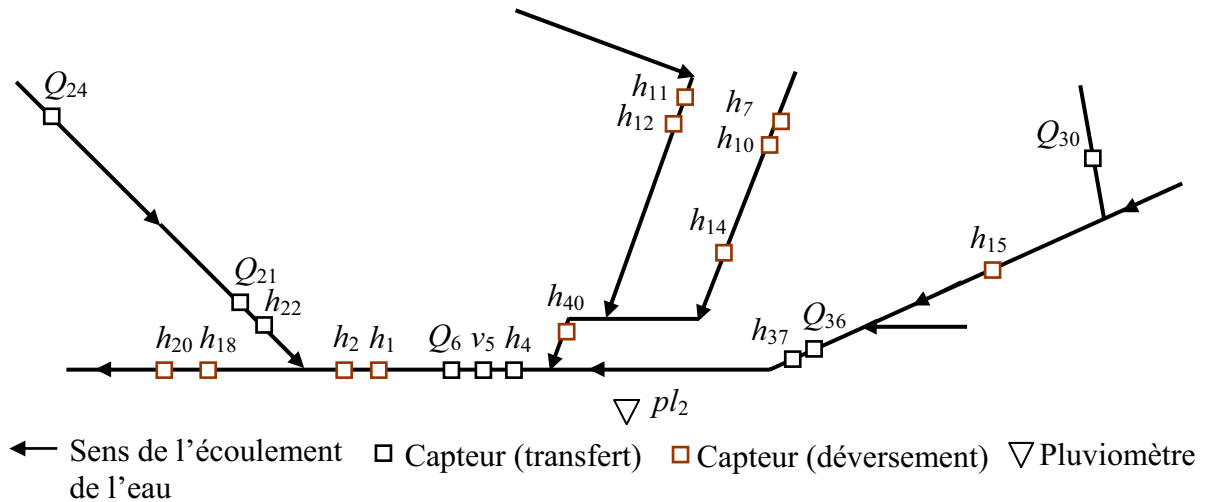


Figure 3.1. Emplacement des mesures pour le réseau d'assainissement, h_i hauteur d'eau, v_j vitesse d'écoulement d'eau et Q_k , débit d'eau, avec i, j, k numéro du capteur $\in \{1, 2, \dots, 40\}$

Nous avons représenté dans la figure 3.2, l'allure des signaux, $h_4(k)$, $v_5(k)$, $Q_6(k)$ et la pluie en fonction du temps k , sur une période d'une semaine (mois de novembre de l'année 2001). L'emplacement des capteurs fournissant ces mesures est indiqué sur la figure 3.1.

Le débit eaux usées, indiqué par la variable Q_6 dans le troisième graphe de la figure 3.2, est produit en grande partie par l'activité humaine, d'où des cycles journaliers et hebdomadaires que nous remarquons sur les valeurs du débit. Pendant les jours de la semaine, les variations du débit sont presque identiques (les coefficients de corrélation entre les valeurs du débit d'un jour à un autre sont supérieurs à 0.80). Nous remarquons également que le débit atteint une valeur maximale de $1.24 \text{ m}^3 / \text{s}$, sur la période considérée. Cette valeur correspond à une valeur maximale de pluie, présentée dans le quatrième graphe de la figure 3.2. Cela est justifié par le fait que les pluies génèrent aussi des débits d'eaux qui sont évacués par le même réseau de collecte. Ainsi à toute intensité élevée de pluie correspond un pic de débit.

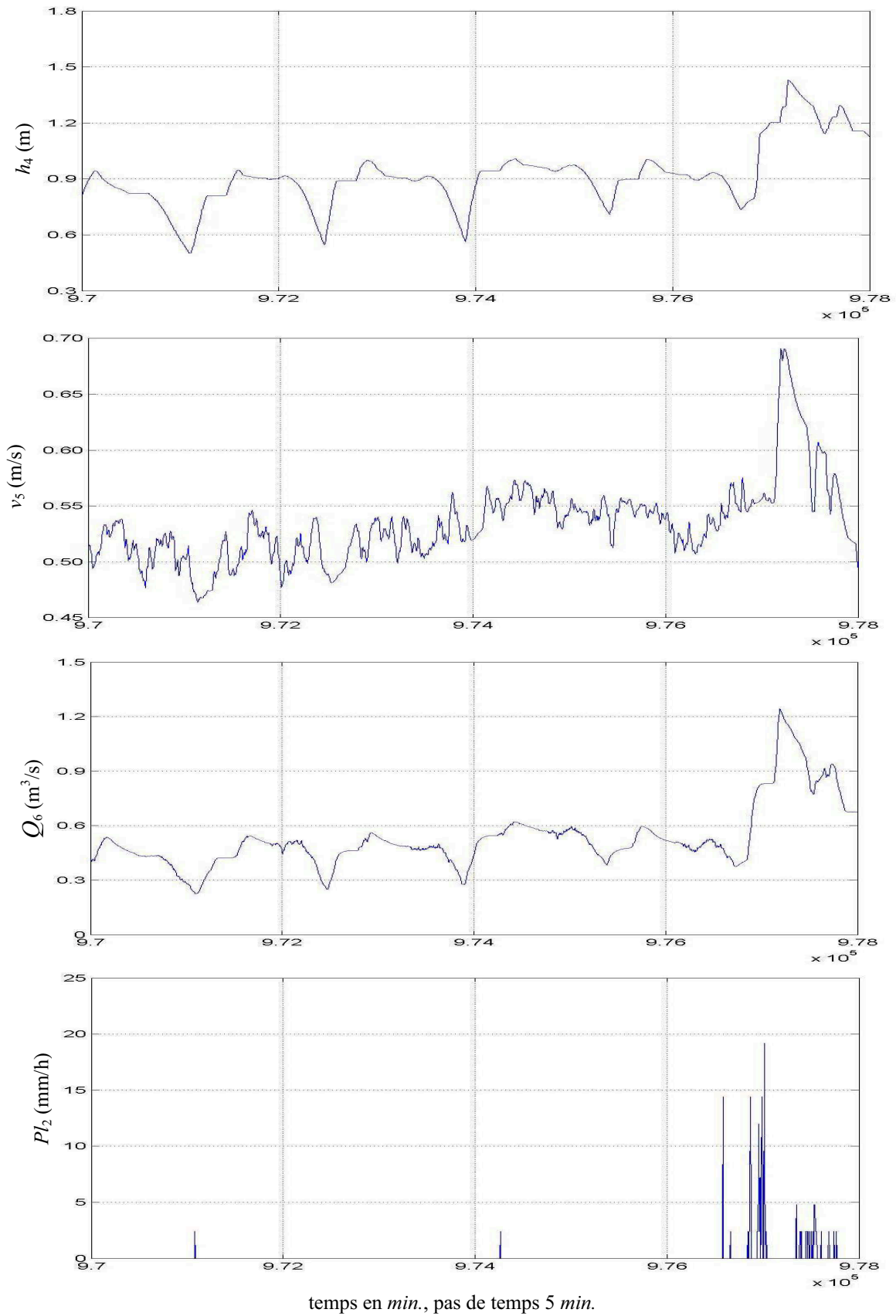


Figure 3.2. Allure des signaux hauteur, vitesse, débit et pluie dans le réseau sur une période d'une semaine

Comme pour le débit, nous remarquons des fluctuations journalières sur les valeurs de la hauteur d'eau h_4 et peu remarquées sur les valeurs de la vitesse v_5 , qui sont représentées respectivement dans le premier et deuxième graphes de la figure 3.2.

Les particularités constatées pour les données du point de mesure Quai fosse, peuvent être généralisées. Leur connaissance est utile pour la validation manuelle des données qui constitue une étape nécessaire dans une phase de prétraitement de données, présentée dans la section suivante.

3.3 Prétraitement des données

Pour établir un modèle, nous avons besoin d'un ensemble de mesures validées ou « crédibles », autrement dit, qui ne contiennent pas de données aberrantes.

Dans quelques travaux de recherches [Ragot et *al.*, 90 ; Berrada et *al.*, 96 ; Mourad et *al.*, 05], les techniques de filtrage sont utilisées pour valider les données hydrométriques.

Nous nous sommes basés sur cette idée de filtrage de données, pour préparer un ensemble de mesures validées. Comme ces filtres ne nous permettent pas d'éliminer les perturbations, une expertise [Aumond et Joannis, 01 ; Aumond et *al.*, 03] est nécessaire pour les enlever. En fait, l'expert connaît le fonctionnement du système ainsi que les variations habituelles des signaux traduisant ce fonctionnement. Il commence par visualiser chaque signal à part, puis comparer les signaux entre eux, s'il remarque des fluctuations inhabituelles ou des valeurs aberrantes et ne peut en aucun cas les expliquer ou les justifier, il décide la présence des anomalies. Par exemple : Dans le cas de la validation d'une hauteur d'eau au niveau du déversoir situé à proximité de l'Erdre (cette hauteur représente en réalité le niveau d'eau dans « l'Erdre ») l'expert peut remarquer des fluctuations journalières et quelques fluctuations lentes. Selon sa connaissance a priori du fonctionnement du système, il justifie la présence des fluctuations lentes par la présence des intensités élevées de pluie. Par contre, il ne peut justifier la présence des fluctuations journalières qui ne doivent pas exister en cas de fonctionnement normal du système. Ainsi, il les considère comme étant des anomalies.

Nous avons appliqué le filtre ébarbeur winsoriseur dont les relations d'ébarbage sont données par (2.1) et (2.2), et le filtre passe-bas exprimé par la relation (2.3), sur toutes les données (celles utilisées pour établir des modèles et celles à valider en utilisant ces modèles établis).

Nous rappelons que, l'objectif principal du filtrage, est l'élimination du bruit de mesure. Il a l'inconvénient de déformer le signal original. Certaines méthodes de filtrage sont adaptées à l'élimination des bruits de faibles amplitudes. D'autres par contre, permettent la détection, l'élimination, puis l'estimation des valeurs aberrantes.

3.4 Analyse des données

Pour établir des relations entre les différentes variables, nous avons calculé la corrélation entre toutes les variables. Nous avons respectivement regroupé les corrélations entre les variables hauteurs et vitesses d'eau pendant une période de temps sec et une période de temps de pluie dans le tableau 3.2 et le tableau 3.3. La période de temps de pluie est définie sur une durée de trois semaines allant du 01 au 21 octobre, ce qui correspond à 30000 échantillons. La période sans pluie est définie du 14 au 21 novembre, ce qui correspond à 13000 échantillons.

En examinant ces deux tableaux, nous remarquons de fortes corrélations entre certaines variables, indiquées en gras. Ces corrélations ne varient que légèrement en passant de la période sèche à la période de pluie. Nous pouvons donc dire que la pluie n'a pas de grande influence sur ces corrélations.

	h_1	h_2	h_4	v_5	h_7	h_{10}	h_{11}	h_{12}	h_{15}	h_{17}	h_{18}	h_{20}	h_{22}	h_{37}	h_{40}
h_1	1.00														
h_2	-0.22	1.00													
h_4	0.99	-0.19	1.00												
v_5	0.71	0.13	0.69	1.00											
h_7	0.32	-0.09	0.31	0.20	1.00										
h_{10}	0.27	-0.10	0.26	0.15	0.98	1.00									
h_{11}	0.50	0.06	0.49	0.36	0.91	0.88	1.00								
h_{12}	0.22	-0.14	0.22	0.10	0.84	0.87	0.82	1.00							
h_{15}	0.76	-0.14	0.78	0.45	0.62	0.57	0.73	0.45	1.00						
h_{17}	0.57	-0.08	0.55	0.50	0.76	0.71	0.80	0.58	0.76	1.00					
h_{18}	-0.24	0.96	-0.23	-0.09	-0.05	-0.01	0.05	-0.01	-0.04	0.03	1.00				
h_{20}	-0.18	0.91	-0.04	-0.03	0.02	0.02	0.08	0.01	-0.01	0.06	0.77	1.00			
h_{22}	0.89	-0.09	0.93	0.53	0.60	0.54	0.75	0.42	0.90	0.83	-0.02	-0.06	1.00		
h_{37}	0.84	-0.17	0.85	0.57	0.53	0.47	0.68	0.36	0.93	0.78	-0.07	-0.03	0.94	1.00	
h_{40}	0.81	-0.17	0.80	0.60	0.57	0.50	0.70	0.39	0.91	0.82	-0.06	-0.03	0.95	0.99	1.00

Tableau 3.2. Corrélation entre les différentes variables sur la période de pluie

	h_1	h_2	h_4	v_5	h_7	h_{10}	h_{11}	h_{12}	h_{15}	h_{18}	h_{20}	h_{22}	h_{37}	h_{40}
h_1	1.00													
h_2	-0.40	1.00												
h_4	0.98	-0.45	1.00											
v_5	0.55	0.10	0.53	1.00										
h_7	-0.05	0.10	-0.05	-0.10	1.00									
h_{10}	0.09	0.10	0.08	0.24	0.96	1.00								
h_{11}	0.22	0.10	0.28	-0.26	-0.01	0.75	1.00							
h_{12}	0.05	-0.02	0.04	0.10	-0.01	0.12	-0.27	1.00						
h_{15}	0.81	-0.39	0.83	0.02	-0.08	-0.01	0.54	-0.13	1.00					
h_{18}	-0.46	0.95	-0.50	0.09	0.06	0.10	-0.01	0.06	-0.46	1.00				
h_{20}	-0.26	0.81	-0.29	-0.05	-0.01	0.04	0.03	-0.02	-0.26	0.73	1.00			
h_{22}	0.94	-0.17	0.90	-0.07	-0.07	0.01	0.76	-0.21	0.92	-0.28	-0.10	1.00		
h_{37}	0.89	-0.38	0.91	0.07	-0.07	0.05	0.39	-0.06	0.84	-0.47	-0.22	0.82	1.00	
h_{40}	0.78	-0.26	0.75	0.11	-0.06	0.08	0.35	-0.10	0.88	-0.32	-0.19	0.80	0.94	1.00

Tableau 3.3. Corrélation entre les différentes variables sur la période sans pluie

Nous avons par la suite, établi le tableau 3.4 récapitulatif des corrélations entre les variables débits d'eau dans le réseau et la pluie (pl_2). Nous remarquons de fortes corrélations entre certaines variables débits, indiquées en gras et quelques corrélations entre les variables pluies. Les corrélations pluie-débits sont très faibles.

	Q_6	Q_{21}	Q_{24}	Q_{30}	Q_{36}	Pl_2
Q_6	1.00					
Q_{21}	0.81	1.00				
Q_{24}	0.72	0.75	1.00			
Q_{30}	0.50	0.53	0.84	1.00		
Q_{36}	0.94	0.81	0.77	0.57	1.00	
Pl_2	0.16	0.08	0.07	-0.04	0.10	1.00

Tableau 3.4. Corrélation entre les variables "débits et pluies"

Nous avons calculé d'une part, les corrélations entre les variables, leurs carrés et leurs racines aux carrés et d'une autre part les corrélations entre les variables définies aux instants k et et celles définies aux instants $k-1$ et $k-2$. Dans les deux cas, nous avons remarqué

que ces corrélations, même parfois de valeurs significatives, sont moins bonnes que celles données dans les tableaux 3.2, 3.3, et 3.4.

Nous nous sommes basés sur les fortes corrélations (coefficient de corrélation supérieur à 0.65) qui existent entre les variables, pour sélectionner celles pour lesquelles des relations linéaires statiques peuvent exister. En ce qui concerne les corrélations moyennes (coefficient de corrélation variant entre 0.50 à 0.60), des relations non linéaires existent entre certaines variables et peuvent être approximées par des relations linéaires.

3.5 Etablissement des modèles

Les informations apportées par les mesures recueillies en divers points dans le réseau de mesure, présentent des redondances. Ces redondances sont dues au fait que la structure du réseau d'assainissement est arborescente et se traduit par la contribution des débits en amont aux débits en aval, et de leur réaction aux mêmes facteurs (activité humaine, évènements pluvieux, etc.).

Nous avons exploité les corrélations qui existent entre les diverses données disponibles. Ce sont des corrélations débits-débits, hauteurs-vitesses et hauteurs-hauteurs.

Comme il a été déjà mentionné ci-dessus, le débit qui se trouve en amont d'un point de mesure, contribue au débit de ce point de mesure. En effet, des corrélations existent entre les débits des différents points de mesure.

Une relation de redondance analytique est une équation dans laquelle toutes les variables sont connues. Elle permet d'exprimer une variable en fonction d'une ou d'un ensemble de variables qui sont fortement corrélées avec celle-ci.

Comme le réseau d'assainissement occupe une surface de 50km^2 , des décalages existent entre les débits qui se propagent d'un point à un autre du réseau. Ces décalages sont pris en considération par la suite.

Un simple calcul des fonctions d'intercorrélation entre les différentes variables, nous a permis de quantifier ces décalages. Un décalage temporel entre deux variables peut être estimé par la valeur maximale de la fonction d'intercorrélation des deux variables.

Soit la fonction d'intercorrélation $R_{y_i y_j}(k_i, k_j)$ définie entre les variables $y_i(k)$ et $y_j(k)$, où k représente le temps, avec $i, j = 1:N$, et N est le nombre de variable

considérées. Lorsque les variables sont stationnaires, cette intercorrélacion ne dépend que de la différence $k_i - k_j$, que nous avons noté τ .

Nous avons représenté la fonction d'intercorrélacion entre les différentes variables en fonction de τ . Dans les deux figures 3.3.a et 3.3.b, nous avons considéré quelques exemples pour illustrer le décalage temporel noté Δ_{y_i, y_j} qui existe entre les deux variables $y_i(k)$ et $y_j(k)$, et correspond au coefficient de corrélation maximal.

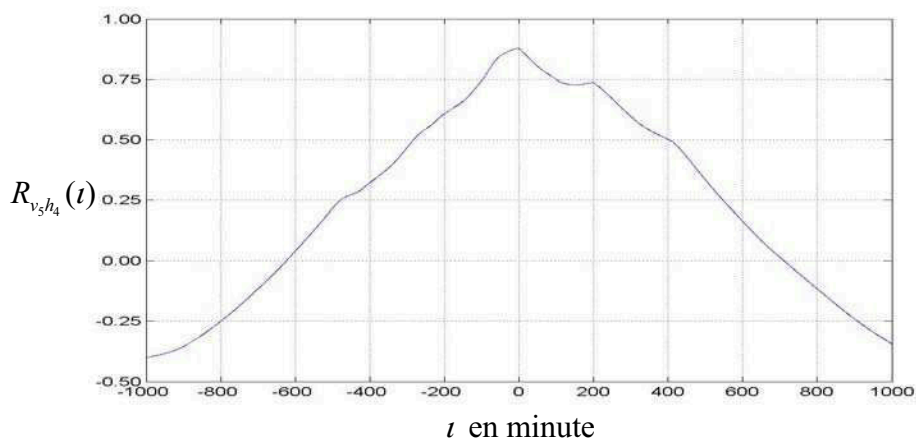


Figure 3.3.a. Intercorrélacion entre les variables v_5 et h_4 , $\Delta_{v_5, h_4} = 0 \text{ min}$.

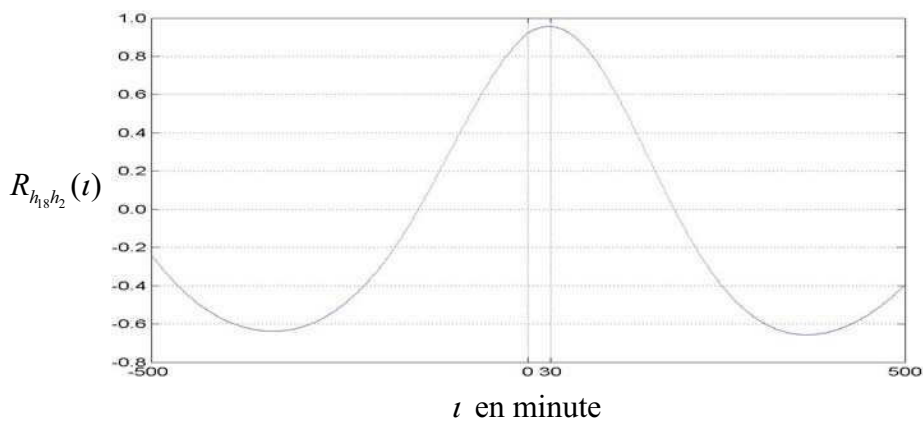


Figure 3.3.b. Intercorrélacion entre les variables h_{18} et h_2 , $\Delta_{h_{18}, h_2} = 30 \text{ min}$.

Dans les tableaux 3.5 et 3.6, nous avons indiqué les décalages temporels qui existent entre les différentes variables.

	h_1	h_2	h_4	v_5	h_7	h_{10}	h_{11}	h_{14}	h_{15}	h_{18}	h_{20}	h_{22}	h_{25}	v_{26}	h_{37}	h_{40}
h_1	0															
h_2	451	0														
h_4	0	0	0													
v_5	0	0	0	0												
h_7	0	0	0	0	0											
h_{10}	0	0	0	0	0	0										
h_{11}	0	0	0	0	0	0	0									
h_{14}	0	0	0	0	0	0	0	0								
h_{15}	0	0	0	0	0	0	0	0	0							
h_{18}	0	-30	0	0	0	0	0	0	0	0						
h_{20}	0	30	0	0	0	0	0	0	0	0	0					
h_{22}	-140	0	-100	0	0	0	0	0	0	0	0	0				
h_{25}	0	0	0	0	0	0	0	0	0	0	0	0	0			
v_{26}	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
h_{37}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
h_{40}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tableau 3.5. Décalage temporel en minutes pour les variables hauteurs et vitesses

	Q_6	Q_{21}	Q_{24}	Q_{30}	Q_{33}	Q_{36}
Q_6	0					
Q_{21}	-131	0				
Q_{24}	-137	-8	0			
Q_{30}	-107	8	15	0		
Q_{33}	-84	46	56	38	0	
Q_{36}	-116	12	16	-8	8	0

Tableau 3.6. Décalage temporel en minutes pour les variables débits

Avant de définir la structure des modèles, nous avons considéré les deux conditions suivantes :

Condition 1. Pas de déversement sur la période d'identification du modèle.

Quelques déversements peuvent ne pas avoir d'influence sur les modèles établis. L'influence des déversements sur la validité du modèle dépend des variables considérées et pas seulement de la quantité déversée.

Nous pouvons vérifier si un déversement a eu lieu ou pas, en comparant sur le même point de mesure, la hauteur d'eau au niveau d'un déversoir et la hauteur d'eau au niveau de la vanne qui commande le déversement.

Condition 2. Respect du domaine de définition des mesures.

La hauteur d'eau dans le réseau est mesurée à partir d'un point de référence (le zéro, qui est représenté physiquement par le radier d'un collecteur). Mais les plages mesurées peuvent être différentes d'une variable à une autre. Par exemple, dans le cas des deux variables $h_2(k)$ et

$h_{18}(k)$ fortement corrélées et où k représente le temps, $h_2(k)$ est mesurée à partir du zéro (le radier) et $h_{18}(k)$ est mesurée à partir d'une hauteur (la hauteur de la lame du déversoir) égale à $1.5m$. Ainsi, la plage de comparaison des deux hauteurs d'eau se situe au dessus de la valeur $1.5m$. Cette plage représente le domaine de définition du modèle qui peut être établi entre les deux variables.

3.6 Structure des modèles

Pour établir un modèle, il faut :

- Définir sa structure, qui représente le choix entre modèles linéaires, non linéaires, statiques ou dynamiques.
- Choisir des variables et des combinaisons de variables à utiliser dans le modèle et définir des ordres de structures.

Les modèles peuvent être classés en deux types :

- **Modèles de connaissance**

Ces modèles sont basés sur des considérations physiques. Les modèles utilisés en assainissement comprennent :

1. Des modèles dits "hydrologiques", qui permettent de transformer une entrée pluviométrique en sortie débitométrique. Cela se fait sur une surface de quelques dizaines à quelques centaines d'hectares pour laquelle tous les apports (eaux usées et eaux pluviales) sont acheminés vers un point unique, appelé exutoire. Un schéma donnant le principe de ces modèles est présenté à la figure 3.4.a. Ces modèles permettent de prévoir les débits calculés sur chaque point en fonction de mesures de pluie (supposées validées par ailleurs) et d'un historique des valeurs de débit sur le point considéré (celui-ci permet de construire un modèle de profil journalier des débits d'eaux usées). Ils ne correspondent pas à l'objectif de notre travail qui est de vérifier la cohérence globale des différentes mesures réalisées sur un réseau, et non de vérifier la cohérence de chaque mesure avec une mesure de référence constituée par la pluviométrie, car celle-ci est également délicate à valider et, par ailleurs, les modèles sont relativement imprécis.

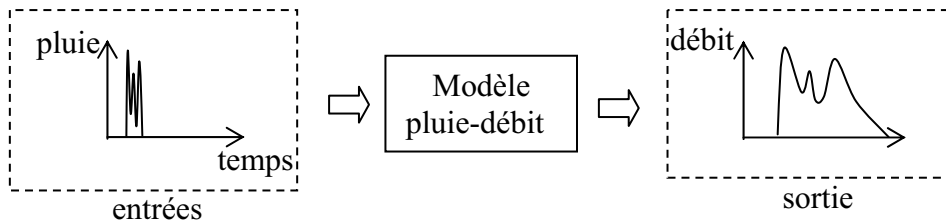


Figure 3.4.a. Schéma de principe du modèle hydrologique

2. Des modèles dits "hydrauliques", décrivant les phénomènes de stockage, de déstockage, et d'amortissement dans les ouvrages principaux d'un réseau d'assainissement. Ces modèles sont basés sur la résolution des équations différentielles régissant la dynamique des fluides, éventuellement exprimées sous une forme simplifiée, dont les équations de Barré de Saint Venant [Barré de Saint Venant, 71] constituent la formulation la plus courante. L'échelle géographique de ces modèles est celle de l'ouvrage (bassin, collecteur). Les données d'entrée sont les conditions aux limites de l'ouvrage, exprimées en termes de hauteur d'eau ou de débit. Les données de sortie sont la hauteur d'eau et la vitesse à l'intérieur de l'ouvrage. Un schéma donnant le principe de ces modèles est présenté à la figure 3.4.b. Ces modèles ne sont guère applicables pour notre travail, car nous avons besoin des conditions aux limites. Ces dernières et en particulier les débits d'entrée sont inconnus. La question est généralement résolue en considérant comme débits d'entrée, les sorties de modèles hydrologiques, sur des bassins versants élémentaires. On se retrouve alors ramené à la situation exposée pour les modèles hydrologiques en ce qui concerne leur adéquation aux objectifs de cette thèse.

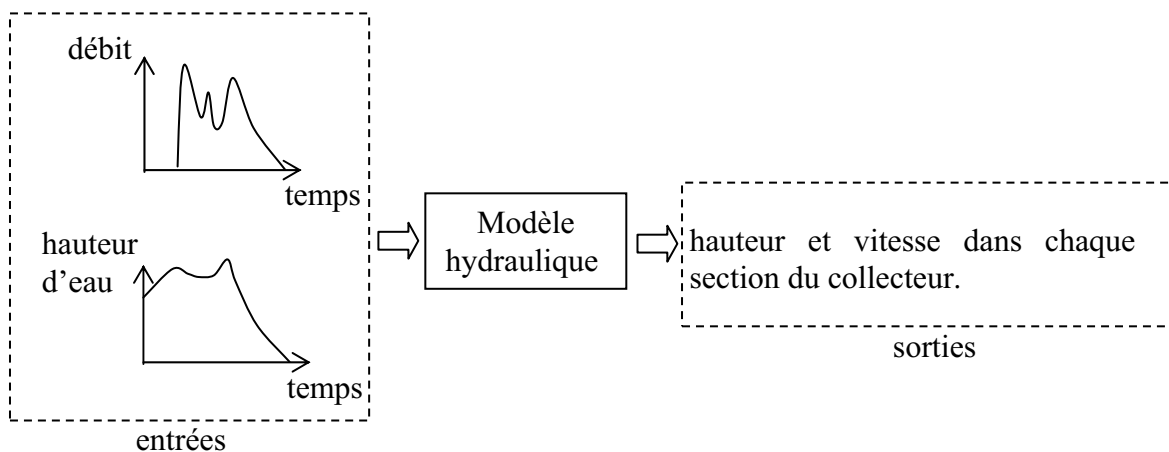


Figure 3.4.b. Schéma de principe du modèle hydraulique

- **Modèles de représentation**

Ces modèles ne sont pas fondés sur les caractéristiques physiques du système étudié mais sur les mesures enregistrées au passé et les corrélations qui existent entre les différentes grandeurs. Ces modèles sont donc des modèles statistiques. Le choix de la structure des modèles nécessite en général des hypothèses simplificatrices afin de limiter leur complexité. Ce choix doit prendre en compte :

- le but fixé à la modélisation,
- les possibilités d'application de ces modèles (nous n'avons pas besoin d'élaborer des modèles très complexes, avec beaucoup de paramètres si leur application ne nous fournit pas des résultats satisfaisants).

3.6.1 Choix de la structure du modèle

Nous avons remarqué dans une étape antérieure que des corrélations significatives existent entre quelques variables (voir tableau 3.2, 3.3, 3.4). Ainsi, des modèles linéaires statiques ont été pris en considération. La relation permettant d'exprimer les mesures en fonction de l'état du système est donnée par :

$$Y = X\theta + \varepsilon, \quad (3.1)$$

où X représente la matrice d'état du système, $Y(n,1)$ représente un vecteur de n mesures,

$\theta(p,1)$ est le vecteur des paramètres inconnus du modèle,

$X(n,p)$ est une matrice de p variables explicatives et de n mesures,

$\varepsilon(n,1)$ représente le vecteur bruit, qui est considéré comme étant un vecteur gaussien de valeur moyenne nulle et de variance σ^2 .

Pour l'ensemble des variables considérées dans le réseau d'assainissement, cette matrice est donnée comme suit : $X = [h_1 \ h_2 \ h_4 \ v_5 \ Q_6 \ h_7 \ h_{10} \ h_{12} \ h_{14} \ h_{15} \ h_{18} \ h_{20} \ Q_{21} \ h_{22} \ Q_{24} \ Q_{30} \ Q_{36} \ h_{37} \ h_{40}]'$.

La sortie et les entrées du modèle représentent respectivement, la variable à expliquer et les variables explicatives.

Présélection des variables

La nature d'une variable peut être une hauteur, une vitesse ou un débit. Nous nous intéressons aux corrélations débits-débits, hauteurs-hauteurs et vitesses-hauteurs dans le même point de mesure, comme il a été indiqué précédemment. La nature des variables

explicatives, est fixée par la nature de la variable à expliquer. Ainsi, nous pouvons établir des relations entre :

- les hauteurs et les vitesses locales (même station de mesure),
- l'ensemble des hauteurs,
- l'ensemble des débits.

Ces relations permettent de faire une présélection des variables explicatives à introduire dans un modèle, pour expliquer la variable considérée. Si cette dernière est un débit, l'ensemble des variables explicatives représente toutes les variables débits se trouvant dans toutes les stations de mesure. Si elle est une vitesse, seules les variables locales c'est-à-dire celles qui se trouvent dans la même station de mesure peuvent être considérées comme variables explicatives. Dans un dernier cas, si elle représente une hauteur, les variables locales au niveau d'une station de mesure, sont considérées comme variables explicatives. Si nous ne disposons pas de variables au niveau de la station de mesure, nous considérons les variables sur tout le réseau de mesure, par exemple dans le cas de la variable à expliquer h_{15} , nous avons besoin des variables hauteurs sur tout le réseau.

3.6.2 Sélection des variables explicatives

Généralement, la sortie d'un modèle peut être expliquée par un ensemble de p variables explicatives. Parfois, il est possible d'expliquer cette sortie par un sous-ensemble comportant seulement quelques variables de tout l'ensemble. Pour cela, il est important de définir une méthode permettant la sélection de ce sous-ensemble de variables.

Un modèle est un outil de description permettant de préciser des relations entre ces variables explicatives. Généralement, un conflit existe entre le souhait d'obtenir une description précise de la sortie du modèle et le souci d'avoir la description la plus simple : la précision entraîne l'inclusion d'un grand nombre de variables qui est incompatible avec la simplicité.

Nous avons exposé brièvement les techniques les plus usuellement utilisées pour la sélection d'un sous-ensemble de variables explicatives afin d'obtenir le modèle le plus approprié. Ces techniques [Tomassone, 92; Kleinbaum, 98] sont :

- La technique de toutes les régressions possibles : consiste à évaluer chaque sortie du modèle avec p variables explicatives. Pour p variables explicatives, le nombre de modèles à identifier est de $2^p - 1$. Une fois ces modèles obtenus, le meilleur modèle est choisi selon le

critère (erreur quadratique minimale et coefficient de corrélation maximal entre la variable mesurée et celle expliquée par la régression).

- Les techniques pas à pas : consistent à introduire les variables explicatives ou les supprimer dans la régression l'une après l'autre. Il existe trois techniques de sélection pas à pas :

- Technique de sélection ascendante : consiste à introduire les variables une par une dans la régression. Le critère d'introduction des variables explicatives peut être le coefficient de corrélation (le plus élevé) ou, l'erreur quadratique moyenne (la plus faible), entre la variable mesurée et celle expliquée par la régression. Une statistique t_d de *Student* peut être testée, elle est utilisée pour décider de retenir la première variable introduite, le test est répété chaque fois que nous introduisons une nouvelle variable explicative. Quand ce test devient non significatif, la sélection des variables est arrêtée.

- Technique de sélection descendante : consiste, en partant de la régression complète où toutes les variables explicatives sont considérées, à éliminer la variable correspondant à la statistique t_d de *Student* la plus petite, ou celle qui augmente le moins la somme des carrés des écarts entre la variable à expliquer et son estimation donnée par la régression. Dans le cas de l'utilisation de la statistique t_d de *Student*, la suppression d'une variable explicative ne peut se faire que si cette statistique est inférieure à une valeur fixée.

- Technique de sélection progressive avec possibilité d'élimination : c'est une version modifiée de la technique de sélection ascendante. Elle permet de réexaminer, à chaque étape l'introduction des variables incorporées dans le modèle dans les étapes précédentes. Des variables introduites tout au début de la sélection peuvent, à cause de leur liaison avec d'autres variables introduites ultérieurement, devenir non significatives après un certain nombre de pas. La technique de sélection progressive utilise ce fait pour éliminer de telles variables.

Dans notre travail, nous avons considéré une procédure basée sur la technique de sélection ascendante pour la sélection des variables explicatives à introduire dans les modèles à partir des ensembles des variables définis dans la phase de présélection. Cette technique à régression multiple, est la méthode des moindres carrés orthogonale. Elle consiste à classer les variables explicatives selon la quantité d'information (définie par la suite) apportée par chacune d'entre elles dans le modèle considéré. Chaque fois que nous introduisons une variable explicative, un écart, entre la variable à expliquer et celle mesurée, est calculé. La quantité d'information apportée par chaque variable introduite représente la différence entre le

premier écart calculé pour la première variable explicative introduite et l'écart correspondant à chaque introduction de variable explicative. Nous avons fixé pour cette technique, un critère d'arrêt (critère d'Akaike, noté, AIC) pour la sélection des variables. Ce critère nous permet de sélectionner un nombre limité de variables explicatives classées par cette méthode.

La régression des moindres carrés orthogonale

Cette procédure de sélection [Chen, 89, 91] est considérée comme l'une des méthodes de réduction orthogonale [Miller, 02]. Elle permet de sélectionner les variables explicatives significatives selon un rapport de réduction de l'erreur de l'estimation de la sortie du modèle calculé pour chacun des régresseurs. Ce rapport de réduction d'erreur est calculé à partir de l'ensemble des variables excepté la variable à expliquer. Il représente le taux de réduction de la variance (qui doit être maximisé) pour la variable à expliquer. Nous avons considéré le critère d'arrêt AIC , pour lequel nous arrêtons la sélection quand il atteint son minimum comme il a été traité dans [Hamioud *et al.*, 03].

Cette méthode est intéressante car elle permet, d'une part, de connaître la contribution de chaque variable explicative dans le modèle considéré et d'autre part, de sélectionner quelques variables explicatives parmi l'ensemble des variables explicatives candidates.

Soit la relation (3.2), où ε représente le vecteur de bruit, il est considéré comme étant un vecteur de variables aléatoires indépendantes de moyenne nulle et de variance constante, $\varepsilon \sim N(0, \sigma^2 I_n)$, où $\sigma^2 I_n$ est la matrice de variance-covariance.

Considérons la décomposition orthogonale de la matrice X donnée comme suit :

$$X = WB, \quad (3.2)$$

où

$$B = \begin{bmatrix} 1 & b_{1,2} & \cdots & b_{1,p} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & b_{p-1,p} \\ 0 & \cdots & 0 & 1 \end{bmatrix},$$

est une matrice triangulaire supérieure, de dimensions (n, p) et $W = [w_1 \dots w_p]$, est une matrice de dimensions (p, p) avec des colonnes orthogonales, qui satisfont $w_i^t w_j = 0$, pour $i \neq j$, avec t l'indicateur de transposée.

La relation (3.2) peut être réécrite comme suit :

$$Y = Wg + \varepsilon, \quad (3.3)$$

où g est le vecteur de poids tel que :

$$B\theta = g . \quad (3.4)$$

Connaissant B et g , θ peut être calculé avec la relation (3.4) par un algorithme de substitution descendante et une estimation \hat{Y} de Y , peut être calculée.

La somme des carrés de la variable à expliquer Y est donnée comme suit :

$$Y'Y = g'W'Wg + e'e .$$

Le taux de réduction d'erreur dû à w_i , est donné par : $\tau_i = (g_i'w_i'w_i g_i) / Y'Y$, où i indique le numéro de la variable explicative considérée. La variable explicative pour laquelle τ_i est maximum, est sélectionnée.

Nous arrêtons la sélection quand le critère de sélection AIC défini comme suit :

$$AIC = -2\ln(\hat{\sigma}^2) + p' , \quad (3.5)$$

atteint son minimum, avec $\hat{\sigma}^2$: la variance de l'erreur d'estimation de Y , définie par

$\hat{\sigma}^2 = \frac{\sigma^{2'}\sigma^2}{n - p'}$, et où $\sigma^{2'}$ est la transposée de σ^2 , et p' est le nombre des variables explicatives sélectionnées.

Considérons l'exemple de la variable Q_6 , qui représente le débit d'eau dans le réseau (voir figure 3.1). La procédure de présélection des variables explicatives, indiquée précédemment, nous a permis d'obtenir l'ensemble des variables suivant : $\{Q_{21}, Q_{24}, Q_{30}, Q_{36}, C\}$, avec C une constante. Nous représentons à la figure 3.5, le critère d'arrêt AIC .

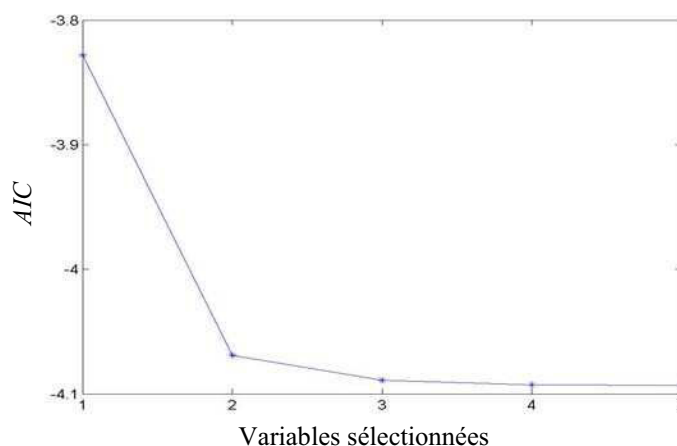


Figure 3.5. Critère d'arrêt AIC en fonction des variables sélectionnées

Quand ce critère atteint son minimum ($AIC = -4.09$), la sélection des variables est arrêtée et les variables sélectionnées sont indiquées dans le tableau 3.7.

Dans le tableau 3.7, nous avons regroupé les résultats de l'application de la technique de sélection pour la variable à expliquer. Nous avons considéré six variables comme entrées pour le modèle explicatif de la variable Q_6 . Pour chaque étape, une sélection d'une variable de forte contribution pour l'explication de Q_6 , est introduite et le critère AIC est calculé. Nous avons souligné la valeur minimale du critère et l'étape d'arrêt correspondante.

étape	débit sélectionné	AIC	taux d'explication
1	Q_{36}	-3.83	0.94
2	C	-4.07	0.95
3	Q_{24}	-4.09	0.95
4	Q_{21}	-4.09	0.95
5	Q_{30}	-4.09	0.95

Tableau 3.7. Critère d'arrêt AIC et variables sélectionnées

Cette procédure de sélection nous a permis de choisir les variables nécessaires pour expliquer la variable considérée.

Comme il a été indiqué dans le paragraphe de présélection des variables explicatives, les variables vitesses ne peuvent être expliquées que par des variables locales et les variables hauteurs et débits peuvent être expliquées par l'ensemble des variables. Par conséquent, le nombre de variables nécessaires pour expliquer une variable vitesse et dans certains cas la variable hauteur est faible ; il est de l'ordre de deux variables, et celui nécessaire pour expliquer les débits et dans certains cas des hauteurs est plus élevé ; il est parfois de l'ordre de cinq variables. Ainsi, dans le cas des débits et certains cas de hauteur à expliquer, la méthode OLS est nécessaire pour faire la sélection des variables explicatives utiles et pour identifier les paramètres du modèle correspondant à la variable à expliquer. Par contre elle n'est pas nécessaire pour le cas de vitesse et certains cas de hauteur et nous utilisons la méthode des moindres carrés, ordinaire pour identifier les paramètres du modèle expliquant cette variable.

Les cas où la hauteur ne demande pas l'application de la OLS est celui de la variable h_{40} car celle-ci peut être expliquée par une seule variable locale qui est h_{37} et nous n'avons besoin que de la méthode des moindres carrés ordinaire pour identifier le modèle. Pour la

vitesse v_5 seules les deux variables h_4 et h_1 peuvent l'expliquer, la variable h_2 étant physiquement indépendante des trois variables h_4 , h_1 et v_5 , elle ne peut être considérée comme variable explicative pour v_5 . Dans le cas d'un débit Q_6 , les variables explicatives sont les variables débits indiquées dans le tableau 3.7. Dans le cas de la variable h_{15} , toutes les variables hauteurs sont présélectionnées pour être introduites dans la procédure de sélection OLS.

Les résultats de l'application de la OLS sur les autres variables sont indiqués dans l'annexe C.

3.7 Identification paramétrique par la méthode des moindres carrés ordinaire

Après la sélection de la structure du modèle, l'identification constitue une phase importante dans la définition des modèles.

Considérons le vecteur des paramètres estimés, $\hat{\theta}$, du modèle donné par la relation (3.4). En remplaçant ce vecteur dans cette relation, la sortie du modèle \hat{Y} qui représente le vecteur des estimations de la sortie du système, peut être écrite sous forme :

$$\hat{Y} = X\hat{\theta}, \quad (3.6)$$

En minimisant l'écart quadratique entre les valeurs de Y et de son estimé \hat{Y} , appelé critère quadratique, une estimation $\hat{\theta}$ des paramètres du modèle considéré est exprimée comme suit :

$$\hat{\theta} = (X'X)^{-1} X'Y, \quad (3.7)$$

3.8 Validation des modèles

La validation des modèles est une étape d'analyse critique des résultats obtenus pour mesurer leur qualité. De nombreuses techniques ont été développées, pour évaluer et valider les modèles en fonction de différents objectifs.

Nous proposons dans ce paragraphe une procédure statistique permettant d'évaluer la validité des modèles établis précédemment. Il s'agit d'utiliser au mieux les données dont nous disposons pour juger de la pertinence du modèle. Cette procédure est basée en partie sur les travaux établis par [Johnson et Wichern, 01]. Elle comporte trois étapes :

Dans une première étape, l'ensemble des données disponibles (trois mois de mesures avec une période d'échantillonnage d'une minute) est partitionnée en trois jeux de données. Nous avons considéré un jeu de données sur lequel le modèle est identifié, appelé jeu-identification et deux jeux de données pour lesquels le modèle est validé, appelé jeux-test. Dans un premier cas, nous avons considéré le premier jeu de données comme jeu-identification, le deuxième et le troisième, jeux de données comme jeux-test. Dans un deuxième cas, nous avons considéré le deuxième jeu de données comme jeu-identification, le premier et le troisième, jeux de données comme jeux-test. Dans un dernier cas, nous avons considéré le troisième jeu de données comme jeu-identification, le premier et le deuxième, jeux de données comme jeux-test. Nous avons calculé pour chaque modèle et pour chaque jeu de données pour les trois cas, les estimations de Y . Les deux étapes qui suivent sont valables pour chaque cas cité ci-dessus.

Une deuxième étape consiste à calculer des critères d'évaluation des modèles. Nous avons considéré, le coefficient de détermination r^2 qui représente le pourcentage de variation de Y , expliquée par le modèle linéaire et l'écart type s_e de l'erreur entre la variable considérée et son estimation \hat{Y} donnée par le modèle.

Une troisième étape consiste à comparer ces deux grandeurs avec des valeurs que nous fixons a priori pour en conclure si le modèle correspondant est valide ou pas : pour une valeur $s_e < 10\%$ de \bar{Y} , avec \bar{Y} la moyenne de Y et pour une valeur de $r^2 \geq 80\%$, le modèle est considéré valide.

3.8.1 Evaluation de l'écart type de l'erreur et du coefficient de détermination

Considérons la somme des carrés des écart des valeurs de Y par rapport à la moyenne, notée SY . Cette somme est partitionnée comme suit :

$$(Y - \bar{Y})(Y - \bar{Y}) = (Y - \hat{Y})(Y - \hat{Y}) + (\hat{Y} - \bar{Y})(\hat{Y} - \bar{Y}), \quad (3.8a)$$

ou bien sous la forme suivante :

$$SY = SE + SR. \quad (3.8b)$$

où SE représente la portion de Y qui n'est pas expliquée par le modèle appelée la variation due à l'erreur d'estimation et SR représente la portion expliquée par le modèle.

La quantité r^2 est donnée par la relation suivante :

$$r^2 = 1 - \frac{SE}{SY}, \quad (3.9)$$

Un estimateur de s_e peut être obtenu comme suit :

$$s_e^2 = \frac{SE}{n - p}, \quad (3.10)$$

avec $n - p$ degré de liberté attaché à SE . Il correspond au nombre de données indépendantes disponibles pour calculer la variance après avoir estimé les p paramètres du modèle.

Nous avons tracé sur les mêmes graphes de la figure 3.7 et la figure 3.8, l'estimation \hat{Y} en fonction de Y pour chaque modèle.

Exemple

Considérons la relation (3.4) qui permet d'expliquer la variable h_4 . Dans ce cas $y(k) = h_4(k)$, et $x(k) = [1 \ h_1(k)]$, avec $n = 87835$ mesures (période de deux mois). Nous n'avons considéré dans cet exemple que deux jeux de données : période 1 de 10500 valeurs correspondant au jeu-identification et période 2 de 77335 valeurs correspondant au jeu-test. Le nombre de valeurs considéré pour chaque période est choisi arbitrairement.

À l'instant donné k , le modèle considéré par la relation (3.4), s'écrit comme suit :

$$h_4(k) = \theta_1 h_1(k) + \theta_0. \quad (3.11)$$

Le vecteur paramètre θ est identifié en utilisant la relation (3.10). En remplaçant le résultat de l'identification dans la relation (3.1), celle-ci devient :

$$\hat{h}_4(k) = 0.98h_1(k) - 0.015. \quad (3.12)$$

Nous considérons pour décider si le modèle est valide ou pas, un premier tableau 3.8 où nous avons considéré la période 1 comme jeu-identification et la période 2 comme jeu-test.

période considérée	r^2 en %	s_e	10% de \bar{h}_4
période 1 est le jeu-identification	98	0.024	0.097
période 2 est le jeu-test	98	0.025	0.090

Tableau 3.8. Validation du modèle en considérant la période 1 comme jeu-identification

Dans un deuxième tableau 3.9, nous avons considéré la période 2 comme jeu-identification et la période 1 comme jeu-test.

période considérée	r^2 en %	s_e	10% de \bar{h}_4
période 2 est le jeu-identification	97	0.024	0.090
période 1 est le jeu-test	98	0.026	0.097

Tableau 3.9. Validation du modèle en considérant la période 2 comme jeu-identification

Comme $r^2 > 80\%$ et $s_e < 10\%$ de \bar{h}_4 , nous pouvons dire que le modèle établi est valide pour l'estimation.

Nous avons tracé à la figure 3.6.a, $h_1(k)$ et à la figure 3.6.b, $h_4(k)$, en fonction du temps.

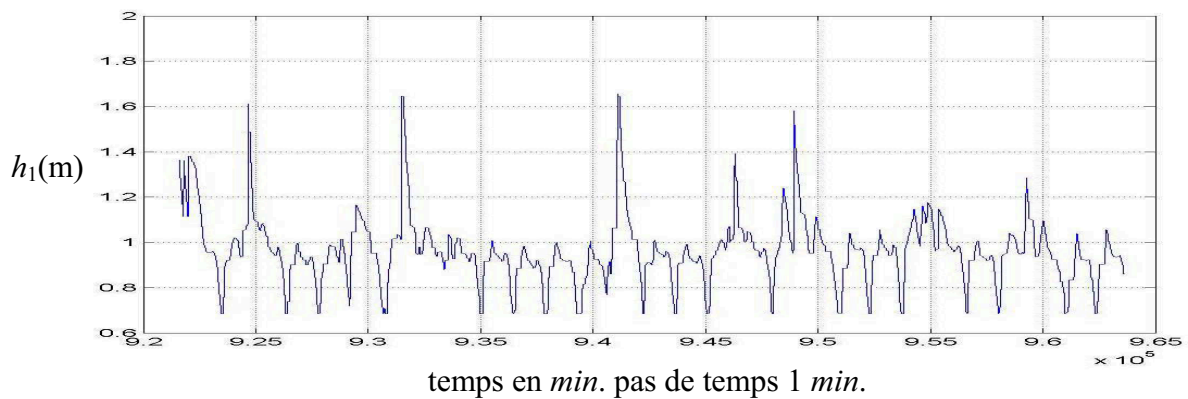


Figure 3.6.a. Signal h_1 en fonction du temps

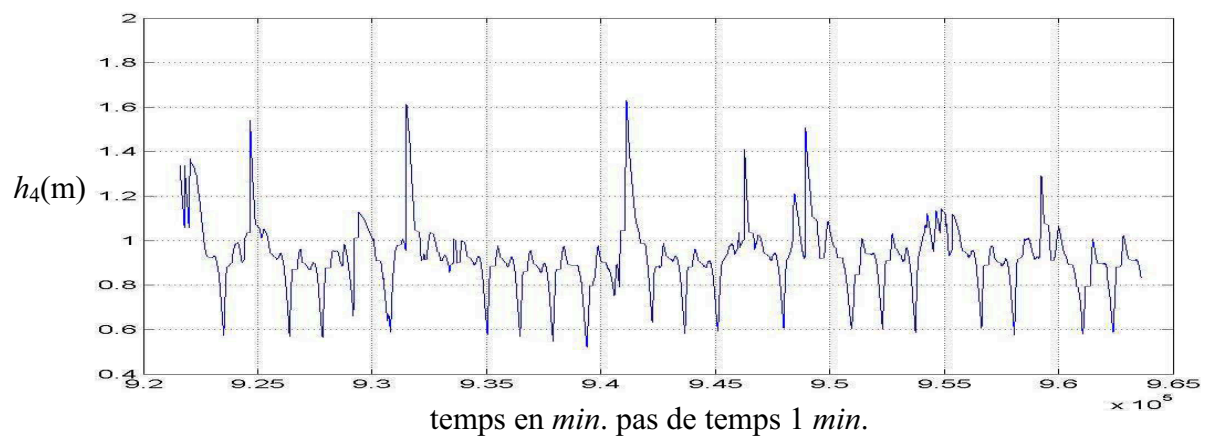
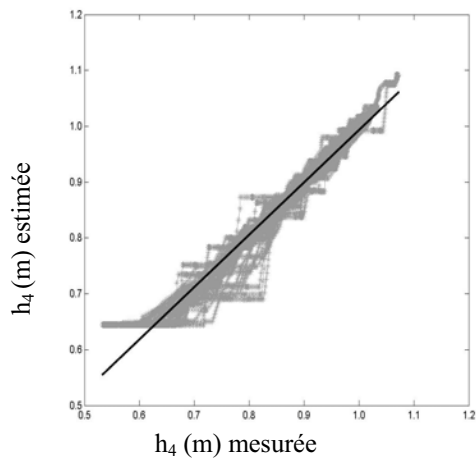
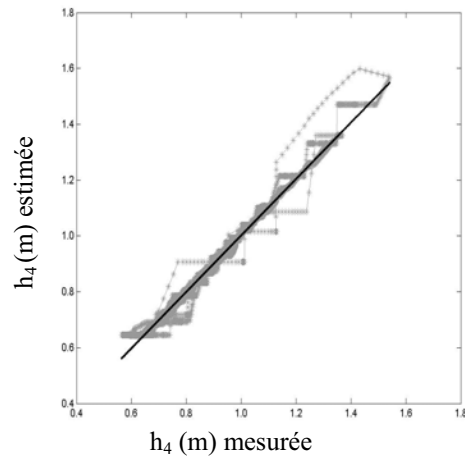


Figure 3.6.b. Signal h_4 en fonction du temps

Sur la même figure 3.7.a, et la figure 3.7.b, nous avons représenté $h_4(k)$ et son estimation pour le cas où la première période est considérée comme jeu-identification et la seconde comme jeu-test. A la figure 3.8, nous avons représenté les mêmes graphes que dans la figure 3.7 mais, en considérant la deuxième période comme jeu-identification et la première comme jeu-test.

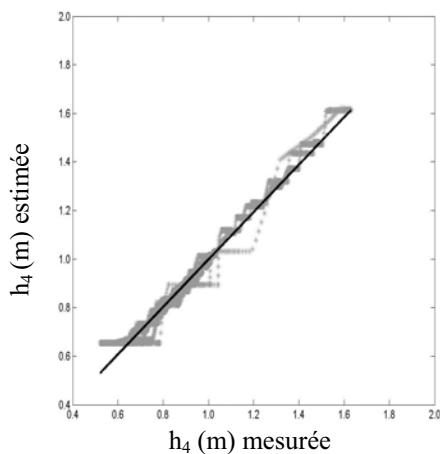


a. période 1 \equiv jeu-identification

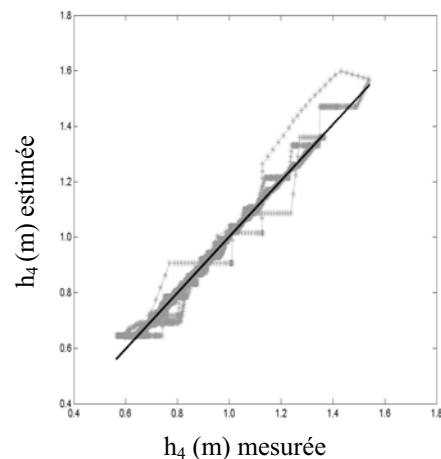


b. période 1 \equiv jeu-test

Figure 3.7. Identification du modèle sur période 1 et validation sur période 2



a. période 2 \equiv jeu-identification



b. période 1 \equiv jeu-test

Figure 3.8. Identification du modèle sur période 2 et validation sur période 1

Nous remarquons, à partir des deux figures 3.7 et 3.8, que les résultats obtenus restent comparables avec les deux jeux de données utilisés. Par conséquent, la variable $h_4(k)$ peut être expliquée par un modèle linéaire en fonction de la variable $h_1(k)$.

Nous avons représenté à la figure 3.9, la variable h_4 en fonction du temps en superposition avec son estimation donnée par le modèle sur une période de deux semaines.

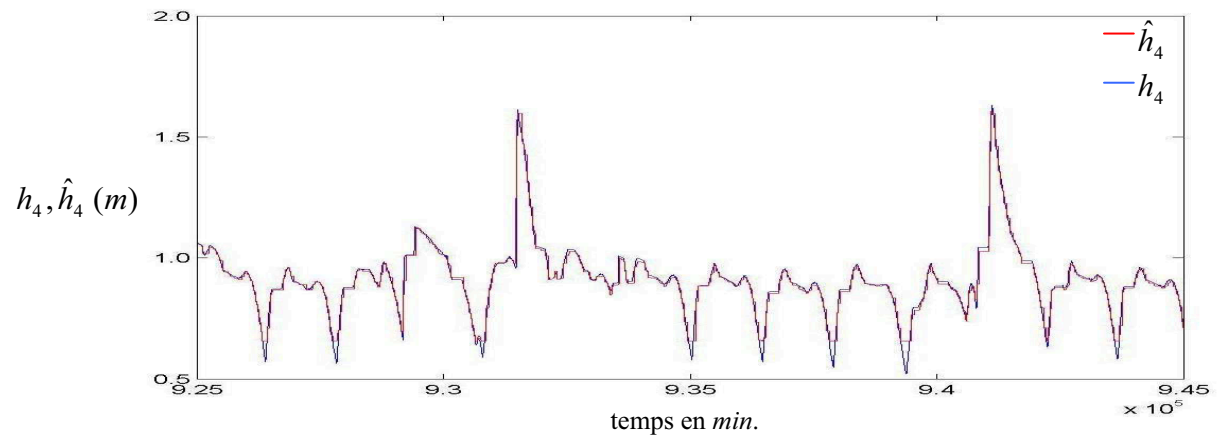


Figure 3.9. La mesure h_4 et son estimée \hat{h}_4 en superposition en fonction du temps

La même procédure de validation est appliquée pour l'ensemble des modèles établis, et les résultats obtenus sont donnés dans l'annexe D.

Les expressions de l'ensemble des modèles obtenus sont récapitulées dans le tableau de synthèse 3.9 donné ci-dessous. Elles sont présentées avec plus de détails dans l'annexe D.

Variable à expliquer	Modèle
h_4	$\hat{h}_4(k) = 0.99h_1(k) - 0.015$
v_5	$\hat{v}_5(k) = 0.44h_1(k) - 0.27h_4(k) + 0.29$
Q_6	$Q_6(k) = 1.60Q_{36}(k - 116) + 0.28Q_{24}(k - 137) + 0.028$
h_7	$\hat{h}_7(k) = 1.06h_{10}(k) + 0.57$
h_{10}	$\hat{h}_{10}(k) = 0.98h_7(k) - 0.05h_{11}(k) - 0.55$
h_{12}	$\hat{h}_{12}(k) = 0.87h_{14}(k) - 0.10$
h_{15}	$\hat{h}_{15}(k) = 0.88h_{37}(k) - 0.04$
h_{18}	$\hat{h}_{18}(k) = 1.03h_2(k + 28) + 1.22$
h_{20}	$\hat{h}_{20}(k + 30) = 0.79h_2(k) - 0.09$
h_{22}	$\hat{h}_{22}(k) = 0.20h_1(k + 140) + 0.75h_4(k + 100) - 0.21$
Q_{24}	$Q_{24}(k) = 1.5624Q_{30}(k + 15) + 0.0433Q_{21}(k + 8) - 0.0436$
Q_{36}	$Q_{36}(k) = 0.1884Q_6(k + 116) + 0.2713Q_{21}(k - 12) + 2.3948Q_{30}(k - 10) + 0.0157$
h_{37}	$\hat{h}_{37}(k) = 0.32h_1(k) + 0.68h_{15}(k) - 0.020$
h_{40}	$\hat{h}_{40}(k) = 0.85h_{37}(k) + 0.20$

Tableau 3.10. Expressions des modèles établis

3.9 Conclusion

Le travail exposé dans ce chapitre consistait à présenter les données utilisées, les analyser et leur appliquer des prétraitements. Ces données ont été utilisées pour établir des modèles capables d'expliquer un ensemble de variables hauteurs, vitesses et débits en fonction de même type de variables, selon des conditions mentionnées dans ce chapitre.

Nous avons constaté depuis les tables de corrélations établies entre les différentes variables, et présentées dans ce chapitre, des corrélations significatives (supérieur à 0.8) entre quelques variables. Ainsi, des modèles de régression linéaire ont été pris en considération. Les structures des modèles ont été effectuées en se basant sur une analyse de ces corrélations. Par la suite, l'identification des paramètres des modèles a été réalisée par l'application d'une

technique de régression. Nous avons appliqué après avoir obtenu ces modèles, une procédure de validation pour mesurer la qualité de ces modèles ainsi que leur validité.

Ces modèles établis représentent une base de données de référence pour les méthodes de détection et de localisation, présentées dans le chapitre suivant.

Chapitre 4 Détection et localisation des défauts de mesures

Sommaire

4.1	Introduction	80
4.2	Génération de résidus	82
4.3	Caractéristiques des résidus	85
4.3.1	Caractéristiques théoriques des résidus.....	85
4.3.2	Caractéristiques expérimentales des résidus	85
4.3.3	Calcul des tolérances associées aux résidus	86
4.4	Matrice des signatures des défauts	86
4.5	Procédure de détection des défauts	89
4.5.1	Choix des paramètres pour la procédure de détection.....	91
4.5.2	Application de la procédure de détection	99
4.6	Localisation des défauts	101
4.6.1	Localisation des défauts basée sur le calcul de distance	101
4.6.2	Localisation de défauts par la méthode DMP	104
4.6.3	Localisation de défauts basée sur les concepts de la logique floue.....	108
4.7	Comparaison des trois procédures de détection et de localisation considérées	121
4.8	Conclusion.....	122

4.1 Introduction

Les mesures issues du réseau d'assainissement doivent être soigneusement examinées et validées avant toute utilisation. Comme les capteurs installés sur le réseau d'assainissement sont soumis à des perturbations, il est nécessaire d'appliquer une procédure de validation des mesures. Cette procédure est effectuée selon deux étapes. Une étape de détection qui permet de détecter des dysfonctionnements dans le réseau ainsi que des défauts de capteurs et donc d'identifier les données suspectes (anomalies) qui caractérisent ces dysfonctionnements et ces défauts. Une étape de diagnostic qui permet de déterminer l'origine d'une anomalie lorsqu'elle a été détectée et de préciser l'instant de son apparition, sa durée ainsi que son amplitude.

Les méthodes de détection et de diagnostic se répartissent en deux grandes catégories selon que nous disposons ou non d'un modèle du processus étudié. Dans le premier cas, on utilise les redondances d'informations et la connaissance fournie par un modèle de référence pour caractériser le mode de fonctionnement ou l'état du système, puis décider s'il est normal ou anormal. Dans le deuxième cas, c'est l'analyse des données fournies par le système qui permet de décider de son état. Dans notre cas d'étude, la détection et le diagnostic des défauts sont à base de modèles. Ces modèles ont été établis au cours du chapitre 3 et représentent les modèles de références. La figure 4.1, montre la structure de la procédure de diagnostic à base de modèles.

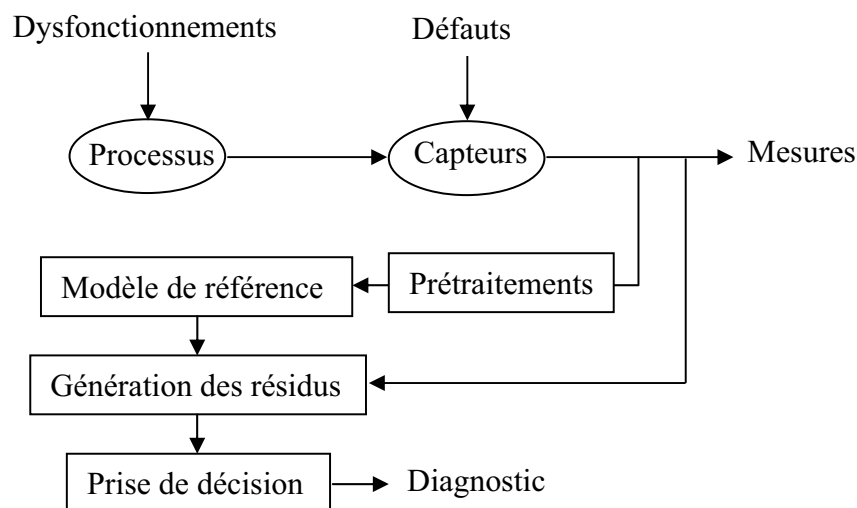


Figure 4.1 Principe général de diagnostic à base de modèles analytiques

Les premiers travaux concernant la détection datent de 1977 [Potter et Suman, 77] à propos de l'utilisation de la redondance pour la détection de défauts de capteurs ou encore les travaux de Himmeblau et de Pau, respectivement publiés en 1978 et 1981, posant les principes fondamentaux de la détection et du diagnostic de défauts. Ces développements ont non seulement concerné la formalisation théorique des problèmes rencontrés et leur résolution mais également la mise en œuvre sur des processus réels.

Une grande variété de méthodes de diagnostic de fonctionnement de capteurs a été proposée dans les domaines où la sécurité des systèmes est fondamentale. Deux grandes familles existent :

- Les méthodes basées sur le traitement du signal qui permettent de détecter des défauts de capteurs considérés indépendamment les uns des autres (traitement mono-signal). Elles consistent à comparer les valeurs du signal issu du capteur à des références (seuils définis par la connaissance a priori sur la grandeur mesurée) [Brunet et *al.*, 90; Zwingelstein, 95]. Des tests statistiques sur les signaux eux-mêmes ou sur les paramètres des modèles les représentant permettent également de détecter des défauts de capteurs [Basseville, 86 ; Ragot et *al.*, 90 ; Basseville et Nikifirov, 93]. Il s'agit des tests de saut de moyenne, de variance sur un horizon glissant.
- Les méthodes basées sur la redondance de l'information [Dorr, 95 ; Giraud, 98 ; Boukhris, 98 ; Simani et Patton, 02]. Le principe de la méthode considérée dans ce travail contenant la redondance d'information consiste en un test de cohérence entre les données de mesures. Cette méthode a été choisie parce qu'elle permet d'exploiter les relations analytiques qui existent entre les différentes données de mesure.

La détection de défaut à base de modèles exploite les signaux d'entrée et de sortie du système pour générer des résidus [Maquin et Ragot, 91 ; Combastel et *al.*, 02]. Ces résidus sont analysés pour prendre une décision sur la présence ou l'absence de défauts. Cette analyse peut se faire selon la méthode de localisation des défauts choisie au niveau du diagnostic. Ce dernier vient compléter l'étape de détection. Il permet de localiser et d'identifier le défaut.

La prise de décision pour la détection des défauts peut s'effectuer par un simple test de dépassement de seuil sur les valeurs instantanées ou sur des moyennes mobiles des résidus [Page, 54; Hinkley, 71] ou faire appel à la théorie de la décision statistique [Basseville et Nikifirov, 93].

En ce qui concerne la localisation des défauts, la seule analyse des résidus générés n'est pas suffisante. Une analyse des relations ayant permis de générer les résidus est également nécessaire. Nous nous intéressons donc aux expressions qui ont permis de générer ces résidus. Les valeurs des résidus à un instant donné, constituent une signature de défaut expérimentale reflétant le comportement observé. Les résidus sont définis d'une manière théorique, en fonction des mesures et des défauts. A partir des expressions donnant ces résidus, nous pouvons déterminer les symptômes correspondants à chaque situation de défaut. En d'autres termes, chaque signature de défaut permet de définir le sous-ensemble de résidus affecté par ce défaut en théorie, pour chaque situation de défaut envisagée. Une signature de défaut théorique est donc associée à chacune de ces situations. La situation de défaut la plus vraisemblable est désignée suite à la comparaison de la signature expérimentale avec chaque signature théorique de défaut.

Ce chapitre est consacré à la description de la méthode de détection et de localisation des défauts de mesure sur le réseau d'assainissement que nous avons adoptée. Nous décrivons tout d'abord comment les résidus sont générés et comment la matrice des signatures des défauts est établie. Par la suite, nous présentons l'approche de Page Hinkley pour la détection. Elle est appliquée sur les valeurs des résidus générés, sur les résidus filtrés et sur les normes des résidus. Puis, nous décrivons les méthodes de localisation utilisées : celle à base de calcul de distance, celle dite Diagnostic Model Processor et celle basée sur les concepts de la logique floue. À la fin de ce chapitre, nous présentons quelques comparaisons entre ces méthodes.

4.2 Génération de résidus

La phase de génération des résidus consiste à produire des signaux révélateurs d'une incohérence entre le fonctionnement courant du réseau et un fonctionnement de référence. Ces résidus sont des signaux indicateurs de défauts. L'état de fonctionnement courant est représenté par la valeur des mesures. Le fonctionnement de référence correspond au bon fonctionnement du réseau. Dans ce cas, les résidus générés restent statistiquement nuls en fonctionnement normal et tout résidu qui s'écarte notablement de la valeur zéro est révélateur de la présence d'un défaut.

Les méthodes usuelles de génération des résidus ont été proposées dans la littérature, et sont basées sur les observateurs [Patton et Chen, 97] et les équations de parité [Gertler, 97]. La génération des résidus pour les systèmes linéaires dynamiques peut être effectuée par deux méthodes. Ces deux méthodes sont basées sur l'exploitation des modèles établis lors du bon

fonctionnement du système. La première méthode est basée sur l'estimation d'état. Les résidus sont générés en comparant les variables à leurs estimations correspondantes aux sorties des modèles. La deuxième consiste en la génération directe des résidus. Elle est basée sur l'exploitation des relations de redondance analytique. Ces relations ne nécessitent que les variables mesurées et permettent donc de calculer à chaque instant, un vecteur de résidus. Quand toutes les variables sont mesurées, les équations des modèles sont toutes des relations de redondances. Dans le cas où toutes les variables ne sont pas mesurées, les équations de modèles doivent être analysées pour éliminer les variables non mesurées [Maquin et Ragot, 00].

Les modèles établis dans le chapitre 3 servent à générer l'ensemble des résidus.

La forme générale des modèles est donnée par la relation suivante :

$$Mx(k) = 0, \quad (4.1)$$

avec $M \in \mathfrak{R}^{l \cdot n}$, $x \in \mathfrak{R}^n$, où M est une matrice de contrainte du modèle du système et $x(k)$ est le vecteur d'état du système à l'instant k .

L'équation de mesure correspondante à ce modèle s'exprime selon :

$$y(k) = Cx(k) + \varepsilon(k) \quad (4.2)$$

où $y(k)$ est le vecteur de mesure des variables à l'instant k , C est la matrice de sélection des mesures et $\varepsilon(k)$ est le vecteur des bruits de mesure supposés centrés.

Dans le cas des systèmes statiques (4.1), un vecteur de résidus peut être généré par combinaison linéaire des mesures des variables. Il s'écrit comme suit :

$$r(k) = Gy(k), \quad (4.3)$$

où G est une matrice de projection des mesures dans l'espace des résidus.

La matrice G est construite à partir du modèle du système en état de bon fonctionnement défini par (4.1) et (4.2). L'expression (4.3) permettant le calcul des résidus ne fait intervenir que des variables connues. Comme ces relations sont appelées relations de redondances analytiques, la détermination de G permet la génération des relations de redondance analytique.

La matrice G est définie telle que, en fonctionnement normal, l'on ait :

$$GCx(k) = 0. \quad (4.4)$$

Soit pour avoir une solution indépendante de l'état : $GC = 0$.

Les résidus ne dépendent donc plus que des bruits de mesure :

$$r(k) = GC(k) + G\varepsilon(k) = G\varepsilon(k). \quad (4.5)$$

Dans le cas de présence de défauts, la relation (4.2) peut être réécrite comme suit :

$$y(k) = Cx(k) + \varepsilon(k) + \delta(k), \quad (4.6)$$

avec $y \in \mathfrak{R}^n$, $\varepsilon \in \mathfrak{R}^n$, $\delta \in \mathfrak{R}^n$, où $y(k)$ est le vecteur de mesures, $\varepsilon(k)$ est le vecteur des bruits de mesures, $\delta(k)$ est le vecteur des défauts qui peuvent affecter les mesures. Dans le cas de défauts uniques, une seule composante δ_i du vecteur δ est non nulle. L'indice i correspond à la mesure y_i affectée par le défaut.

Par conséquent, en présence de défaut le vecteur des résidus est défini par :

$$r(k) = G\varepsilon(k) + G\delta(k). \quad (4.7)$$

En l'absence des erreurs de mesures $\varepsilon(k)$ et des défauts $\delta(k)$, le vecteur des résidus $r(k)$ est nul.

Dans le cas où toutes les variables sont mesurées (cas de notre système étudié), les équations du modèle constituent des relations de redondance analytique. Le vecteur résidu est donc généré comme suit :

$$r(k) = My(k). \quad (4.8)$$

avec $r(k) = [r_1(k) \ r_2(k) \ \dots \ r_t(k)]^t$, t est l'indicateur de transposée.

La relation (4.3) permet le calcul du vecteur des résidus à partir du vecteur des mesures et la relation (4.7) traduit l'influence des erreurs de mesures et des défauts ; elle permet donc d'évaluer le résidu. En présence des défauts ($\delta(k) \neq 0$), $r(k)$ est non nul. Dans ce cas, nous

devons résoudre le problème inverse, c'est-à-dire identifier à partir des résidus, les variables affectées par les défauts (localisation).

Le résidu est un signal indicateur des défauts. Il doit être statistiquement nul en cas de fonctionnement normal du système et dévie de zéro en présence d'un événement anormal.

4.3 Caractéristiques des résidus

Il est important de présenter les caractéristiques statistiques de chaque résidu. Elles sont résumées par sa distribution statistique, sa moyenne et son écart-type. Ces caractéristiques sont à définir dans le cas théorique et expérimental.

4.3.1 Caractéristiques théoriques des résidus

Selon la relation (4.7), les résidus sont exprimés en fonction des erreurs de mesures et des défauts. Dans le cas de fonctionnement normal du système (cas de non présence de défauts), ces résidus correspondent à une combinaison linéaire des bruits de mesures de valeurs moyennes nulles. Si ces bruits de mesures sont caractérisés par une loi normale, la distribution théorique des résidus est une loi normale.

La connaissance de la variance des bruits de mesures permet leur caractérisation précise. Si la matrice variance-covariance des bruits de mesures est connue, la valeur théorique de la matrice de variance-covariance des résidus notée V_r , peut être calculée. A partir de la relation (4.7) et en cas de non présence de défauts, les résidus sont exprimés comme suit : $r(k) = M\varepsilon(k)$ et la relation donnant sa matrice V_r est donnée par : $V_r = GVG^t$. Sous ces conditions de fonctionnement normal du système, les résidus sont des variables aléatoires normales centrées. Les éléments de la diagonale de V_r correspondent aux variances des différents résidus.

4.3.2 Caractéristiques expérimentales des résidus

Dans le cas où la distribution des bruits de mesures est mal connue voire inconnue, les résidus peuvent être caractérisés de manière expérimentale. Cette approche est incontournable en présence d'incertitudes de modèle non négligeables et impossibles à quantifier de manière théorique. Dans ce cas, les valeurs des résidus sont calculées à partir des mesures enregistrées sur une période de fonctionnement normal du système suffisamment longue. Les caractéristiques statistiques (moyenne et écart-type) des résidus peuvent être estimées par la

suite à partir des échantillons de valeurs de résidus ainsi obtenus (la moyenne nulle en théorie ne l'est pas forcément en pratique).

4.3.3 Calcul des tolérances associées aux résidus

Une tolérance d'un résidu représente les limites du domaine de son évolution en l'absence de défauts. Elle est fixée en fonction des caractéristiques statistiques du résidu en fonctionnement normal du système. Cette tolérance notée τ_i peut, par exemple, être déterminée à partir de la valeur de la variance v_i du résidu r_i :

$$\tau_i = \alpha \sqrt{v_i}, \quad (4.9)$$

où α est un coefficient dont la valeur est fixée sur la base de considérations statistiques.

Connaissant les propriétés statistiques des résidus, la valeur de α peut être calculée de telle sorte qu'un pourcentage donné des échantillons $r_i(k)$ soit compris dans l'intervalle $[-\tau_i, \tau_i]$. Dans le cas où le résidu suit une distribution normale centrée, la valeur de α est le plus couramment fixée à 1.96. Pour cette valeur de α , selon la table de Laplace-Gauss, 95% des échantillons de $r_i(k)$ sont compris dans l'intervalle $[-\tau_i, \tau_i]$.

Les caractéristiques des résidus théoriques ou expérimentales sont nécessaires pour la mise en œuvre de méthodes de détection et de localisation des défauts.

Les valeurs des résidus sont non suffisantes pour la localisation des défauts, qui nécessite la connaissance de l'influence de chaque variable sur les valeurs des résidus c'est-à-dire la connaissance de la relation entre les effets (appelés aussi symptômes par analogie au domaine médical) et leurs causes.

4.4 Matrice des signatures des défauts

À partir de l'ensemble des résidus générés, la matrice des signatures des défauts théoriques, Λ , peut être établie. Elle représente une codification binaire des occurrences des variables dans les équations des résidus, ou peut être des coefficients des différentes variables qui interviennent dans les équations générant les résidus r_j . Pour conserver plus d'informations, le codage binaire peut être remplacé par un codage ternaire. Chaque colonne de cette matrice représente un vecteur de valeurs binaires appelé signature de défaut théorique. Ces signatures traduisent en fait l'influence des défauts sur les résidus.

Dans le cas d'une matrice de valeurs ternaires, nous avons :

$$\Lambda_{ij} = \text{sign}(S_{ij}), \quad (4.10)$$

où S_{ij} représentent les éléments de la matrice M , avec $i=1:l, j=1:q$, $\Lambda_{.j}$ est la signature de défaut affectant la variable y_j et $\text{sign}(\cdot)$ est la fonction indiquant le signe de chaque élément S_{ij} . Elle est définie par :

$$\text{sign}(S_{ij}) = \begin{cases} -1 & \text{si } S_{ij} < 0 \\ 0 & \text{si } S_{ij} = 0 \\ 1 & \text{si } S_{ij} > 0 \end{cases} . \quad (4.11)$$

La matrice des signatures des défauts expérimentales, peut être générée en associant à chaque résidu testé une valeur binaire 0 ou 1, ou bien ternaire +1, 0, ou -1, selon qu'il est affecté ou pas par un défaut. Cela se traduit par la détection des défauts au niveau d'un résidu en effectuant un test logique. Dans le cas de codification binaire nous avons :

$$\begin{aligned} \text{si } |r_j(k)| \leq \kappa & \text{ alors aucun défaut n'affecte le résidu } r_j \\ \text{si } |r_j(k)| > \kappa & \text{ alors le résidu } r_j \text{ est affecté par un défaut } \end{aligned} \quad (4.12)$$

où κ représente un seuil de détection défini a priori et où $|r_j(k)|$ représente la valeur absolue du résidu à l'instant k .

Dans le cas de codification ternaire, le test est exprimé comme suit :

$$\begin{aligned} \text{si } r_j(k) > \kappa & \text{ alors le résidu } r_j \text{ est affecté par un défaut positif} \\ \text{si } \kappa < r_j(k) < \kappa & \text{ alors aucun défaut n'affecte le résidu } r_j \quad , \\ \text{si } r_j(k) < -\kappa & \text{ alors le résidu } r_j \text{ est affecté par un défaut négatif} \end{aligned} \quad (4.13)$$

Les éléments de la matrice des signatures des défauts expérimentales sont définis par la relation (4.14) pour le cas d'une codification binaire et par la relation (4.15) pour une codification ternaire comme suit :

$$sb_j(k) = \begin{cases} 0 & \text{si } r_j(k) \text{ n'est pas affecté par un défaut} \\ 1 & \text{si } r_j(k) \text{ est affecté par un défaut} \end{cases} , \quad (4.14)$$

$$st_j(k) = \begin{cases} 0 & \text{si } r_j(k) \text{ n'est pas affecté par un défaut} \\ 1 & \text{si } r_j(k) \text{ est affecté par un défaut positif} \\ -1 & \text{si } r_j(k) \text{ n'est pas affecté par un défaut négatif} \end{cases} . \quad (4.15)$$

Pour que les défauts puissent être détectés, aucune colonne de la matrice des signatures de défauts ne doit être nulle.

Pour que tous les défauts puissent être localisés, toutes les signatures théoriques de défauts doivent être distinctes. Les matrices de signatures de défauts théoriques satisfaisant ces deux contraintes engendrent des résidus dont les défauts sont dits « faiblement isolables » [Gertler et Singer, 85] ou « isolable de manière déterministe » [Gertler et Anderson, 92]. Dans le cas où le seuil κ fixé est élevé, quelques valeurs à « 1 » dans la signature binaire des défauts expérimentales se transforment en « 0 » conduisant à une signature dégradée. Pour garantir une localisation correcte des défauts, la matrice des signatures de défauts théoriques ne doit pas comporter de signature théorique égale à une signature dégradée. Dans ce cas, les défauts seront dits « fortement isolables » ou « statistiquement isolables ».

Exemple 4.1

Considérons le vecteur $x(k) = [h_1(k), h_4(k), v_5(k), h_{15}(k), h_{37}(k)]$.

Dans une première étape, nous considérons le sous-ensemble de modèles qui relie ces variables à partir de l'ensemble de modèles établis dans le chapitre 3.

Les équations des modèles définis par 4.1 sont directement utilisées pour générer les résidus. Comme toutes les variables sont mesurées, les équations du modèle constituent des relations de redondance analytique. Le vecteur des résidus est généré à partir de l'expression 4.4, avec $r(k) = [r_4(k) \ r_5(k) \ r_{37}(k)]$.

Les résidus générés sont définis :

$$\begin{aligned} r_4(k) &= h_4(k) - 0.99h_1(k) + 0.014, \\ r_5(k) &= v_5(k) - 0.44h_1(k) + 0.17h_4(k) - 0.28, \\ r_{37}(k) &= h_{37}(k) - 0.32h_1(k) - 0.68h_{15}(k) - 0.007. \end{aligned} \tag{4.16}$$

Pour cet ensemble de résidus et selon le tableau E.2 de l'annexe E, la matrice de contrainte M peut être déduite :

$$M = \begin{bmatrix} -0.99 & 1 & 0 & 0 & 0 \\ -0.44 & 0.17 & 1 & 0 & 0 \\ -0.32 & 0 & 0 & -0.68 & 1 \end{bmatrix}. \quad (4.17)$$

La matrice des signatures des défauts théoriques ternaire est ensuite établie. Elle est donnée par la relation (4.12) comme suit :

$$\begin{bmatrix} & f_1 & f_4 & f_5 & f_{15} & f_{37} \\ r_4 & -1 & +1 & 0 & 0 & 0 \\ r_5 & -1 & +1 & +1 & 0 & 0 \\ r_{37} & -1 & 0 & 0 & -1 & +1 \end{bmatrix}, \quad (4.18)$$

où f_j indique la présence d'un défaut sur la variable y_j .

Nous remarquons d'après la matrice donnée par (4.18) que les défauts f_1, f_4 et f_5 sont isolables et que les défauts f_{15} et f_{37} ne le sont pas.

Cette matrice nous a permis d'identifier les défauts non isolables. Nous pouvons par exemple considérer un quatrième résidu qui permet d'obtenir des signatures des défauts f_{15} et f_{37} où la condition d'isolabilité est remplie. Le résidu que nous pouvons considérer est celui correspondant au modèle (C.40), donné comme suit :

$$r_{40}(k) = h_{40}(k) - 0.85h_{37}(k) - 0.20. \quad (4.19)$$

Le problème est que cette équation n'est valide que pour des valeurs mesurées bien définies de la variable h_{40} .

4.5 Procédure de détection des défauts

La procédure de détection permet de prendre une décision sur la présence d'un ou de plusieurs défauts de mesures dans le système. Elle est basée sur le test de Page [Basseville, 86] qui appliqué sur un résidu, permet de détecter un changement dans la moyenne de ce résidu.

Selon la relation (4.7), un résidu r peut être défini en fonction du bruit et des défauts de mesures ε et δ respectivement. Le bruit de mesure ε est caractérisé par une moyenne nulle et un écart-type σ_ε . En absence de défauts ($\delta = 0$), la moyenne de r est nulle et en présence de défauts ($\delta \neq 0$), la moyenne du résidu est non nulle.

En fait, nous considérons les deux hypothèses H_0 et H_1 comme suit :

H_0 : Le système est en état normal de fonctionnement, le résidu suit une distribution normale de variance connue σ_r^2 et de moyenne nulle.

H_1 : Le système est en état normal de son fonctionnement, le résidu présente une distribution normale de variance σ_r^2 , de moyenne nulle jusqu'à l'instant inconnu k_r où l'état du système change et devient anormal. Dans ce cas, le résidu présente une distribution normale de variance connue σ_r^2 , et une moyenne supérieure en valeur absolue à une valeur dite saut et notée μ , de l'instant inconnu k_r à l'instant connu k .

L'application de ce test sur les valeurs du résidu permet donc de détecter soit une augmentation, soit une diminution de moyenne. Elle suppose la connaissance d'une valeur absolue minimale du saut, notée μ_m . Ce test consiste à calculer les deux entités U_n et T_n comme suit :

$$U_n = \sum_{k=1}^n \left(r(k) - \frac{\mu_m}{2} \right), \quad n \geq 1 \text{ et } U_0 = 0 \quad (4.20)$$

$$m_n = \min_{0 \leq i \leq n} (U_i), \quad n \geq 1,$$

$$T_n = \sum_{k=1}^n \left(r(k) + \frac{\mu_m}{2} \right), \quad n \geq 1 \text{ et } T_0 = 0 \quad (4.21)$$

$$M_n = \max_{0 \leq i \leq n} (T_i), \quad n \geq 1,$$

avec μ_m la valeur absolue de μ et n le nombre des échantillons du résidu considéré.

Une augmentation ou une diminution dans la moyenne que nous appelons aussi rupture dans les valeurs du résidu considéré, est détectée quand : $U_n - m_n \geq \lambda$ ou $T_n - M_n \geq \lambda$ respectivement, où λ est un seuil de détection à fixer.

Quand une rupture est détectée les entités U_n , m_n , T_n et M_n sont réinitialisés afin de détecter d'autres ruptures éventuelles.

La détection d'une rupture dans la moyenne d'un résidu se traduit par une détection de l'instant d'apparition et de la durée de cette rupture. Elle permet de prendre une décision sur

la présence de défaut sur l'ensemble des variables entrées-sortie du modèle qui a permis la génération de ce résidu.

4.5.1 Choix des paramètres pour la procédure de détection

Dans un premier temps, nous avons à définir la valeur de σ_0 pour chaque résidu. Pour ce, nous avons considéré un exemple.

Exemple 4.2

Considérant le résidu r_4 sans défauts, défini par un nombre n d'échantillons. Nous avons fait varier un entier naturel e de 1 à n et pour chaque e échantillons de r_4 , nous avons calculé l'écart-type, σ_r . Le tracé de σ_r en fonction du nombre des échantillons considérés e est donné à la figure 4.2

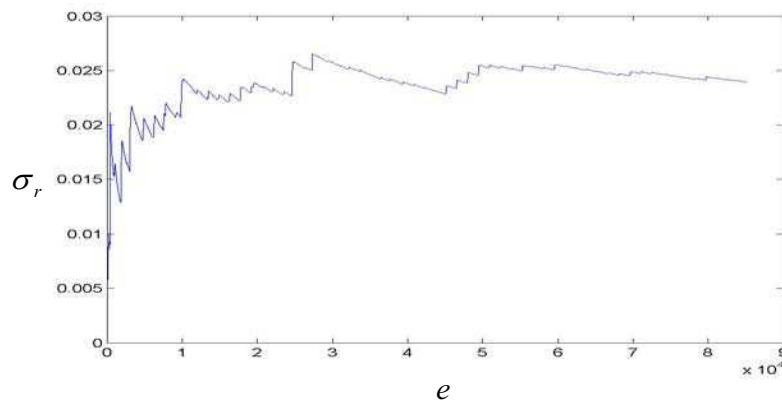


Figure 4.2. Ecart-type σ_r en fonction du nombre des échantillons du résidu r_4

Nous remarquons d'après la figure 4.1, des fluctuations très faibles dans les valeurs de σ_r qui tend vers une limite pour un nombre d'échantillons e bien défini. Ainsi, la valeur de σ_0 considérée est celle pour laquelle σ_r est constante à 2.10^{-7} près.

En ce qui concerne le seuil λ , sa valeur est calculée par l'expression $2d/p$, où d est égale à 2 pour des distributions normales et p est la valeur du saut μ_m , exprimé en nombre d'écart-type du résidu $\mu_m = p\sigma_0$.

Après avoir fixé les paramètres de la procédure de détection et les valeurs initiales, nous avons essayé de mesurer sa capacité de détection. Pour ce faire, nous avons calculé la

probabilité de détection des défauts simulés. Ces défauts sont caractérisés par des instants et des durées bien définis.

Notons pour le test de détection les quatre cas suivants :

D_d : le cas de détection d'un défaut alors qu'il est réellement présent,

F_d : le cas de détection d'un défaut alors qu'il n'existe pas en réalité,

Z_d : le cas de non détection d'un défaut alors qu'il n'existe pas en réalité,

M_d : le cas de non détection d'un défaut alors qu'il existe en réalité,

Supposons que N représente le nombre de tests possibles. Alors, il s'exprime :

$$N = D_d + F_d + Z_d + M_d$$

Une bonne détection représente le cas de détection d'un défaut sachant que celui-ci est réellement présent. A cette bonne détection est associée une probabilité de détection, que nous

avons noté P_d , et qui s'exprime : $P_d = \frac{D_d}{D_d + M_d}$.

Cette probabilité est définie donc comme un rapport de nombre de défauts introduits et détectés sur le nombre de défauts introduits. Elle est calculée comme suit :

$$P_d = \frac{n_{i_d}}{N_i}, \quad (4.22)$$

où n_{i_d} est le nombre de défauts introduits et détectés et N_i est le nombre total de défauts introduits.

Le nombre n_{i_d} est déterminé par une simple comparaison de chaque instant de défaut détecté aux instants des défauts introduits. Chaque fois qu'un défaut détecté coïncide avec un défaut introduit, un compteur initié à zéro est incrémenté.

Une fausse alarme représente le cas d'une décision de présence de défaut alors que ce dernier n'existe pas réellement. A cette fausse alarme est associée un taux de fausses alarmes que nous avons noté τ_{fa} , et qui s'exprime : $\tau_{fa} = \frac{F_d}{D_d + F_d}$.

Ce taux est défini comme le rapport du nombre de fausses alarmes sur le nombre de détection effectuées. Il est calculée comme suit :

$$\tau_{fa} = \frac{n_d - n_{id}}{n_d}, \quad (4.23)$$

où n_d représente le nombre de défauts détectés (introduits ou pas).

A une fausse alarme, peut aussi être associée une probabilité de fausses alarmes que nous avons noté P_{fa} , et qui s'exprime : $P_{fa} = \frac{F_d}{Z_d + F_d}$.

Cette probabilité est définie comme un rapport de nombre de défauts introduits et détectés sur le nombre de défauts introduits. Elle est calculée comme suit :

$$P_{fa} = \frac{n_d - n_{id}}{N - N_i} \quad (4.24)$$

Différents types de défauts simulés ont été considérés :

a. défauts constants

Ce type de défaut est caractérisé par une amplitude δ constante et une durée d_d . Il est défini comme suit :

$$\delta = \tau_d \cdot y', \quad (4.25)$$

avec $y' = \max(y(k))$ dans notre application, où $\max(y(k))$ est la valeur maximale de $y(k)$ et τ_d est le taux de variation de défaut à fixer.

b. défauts dépendant de la mesure.

Ce type de défauts dépend des valeurs entrées-sortie du modèle, il est caractérisé par une amplitude dépendant de l'amplitude d'une des variables entrées-sortie du modèle. Nous avons donc défini deux cas.

$$\delta(k) = \tau_d \cdot y(k). \quad (4.26)$$

b.1. taux constant

Dans ce cas, le taux de variations des défauts est fixé à une valeur constante.

b.2. taux variable

Dans ce cas, le taux de variation des défauts est variable en fonction de la durée du défaut.

où τ_d est défini par une fonction triangulaire comme montrée dans la figure 4.3.

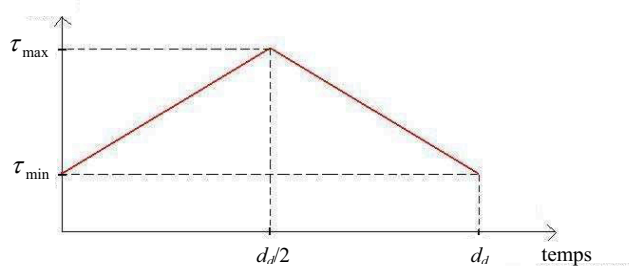


Figure 4.3 Allure du taux de variation des défauts dépendants de la mesure

Nous avons considéré l'exemple 4.3, pour illustrer comment le seuil de détection λ et comment la procédure de détection est appliquée.

Rappelons que P_d et τ_{fa} croient avec la diminution du seuil λ . Ainsi, pour obtenir une bonne détection des défauts avec de faibles fausses alarmes, il faut trouver un compromis, qui permet de déterminer le seuil λ tout en maximisant P_d avec une valeur de τ_{fa} acceptable. Pour ce faire, nous avons défini la fonction coût J comme suit :

$$J = 100 - P_d + \tau_{fa}.$$

Comme nous ne disposons pas de relations analytiques exprimant P_d et τ_{fa} en fonction de λ , nous avons calculé numériquement J et nous l'avons tracé en fonction de λ .

Exemple 4.3

Considérant le résidu $r_4(k)$ défini pour $k = 1:42000$ échantillons et donné par la première équation de la relation (4.17).

Nous avons introduit des défauts sur la variable h_4 (r_4 dépend de cette variable). La durée entre deux défauts a été fixée à 60 minutes. Ces défauts sont définis par la relation (4.25) avec un τ_d constant, égale à 10% et une durée d_d à 30 minutes. La valeur de l'écart type σ_0 est considérée comme il a été indiqué dans le paragraphe 4.4.1 et le saut $\mu = p\sigma_0$, où p est un entier naturel considéré égale à 2 pour cet exemple. Nous avons tracé J en fonction de λ dans la figure 4.4, donnée ci-dessous. La valeur de λ considérée correspond à la valeur minimale de J . Dans ce cas, $\lambda = 0.3$, la P_d est de 91.6% et le τ_{fa} est de 5.9%.

Les valeurs de P_d et τ_{fa} sont obtenues en utilisant les relations (4.22) et (4.23) respectivement.

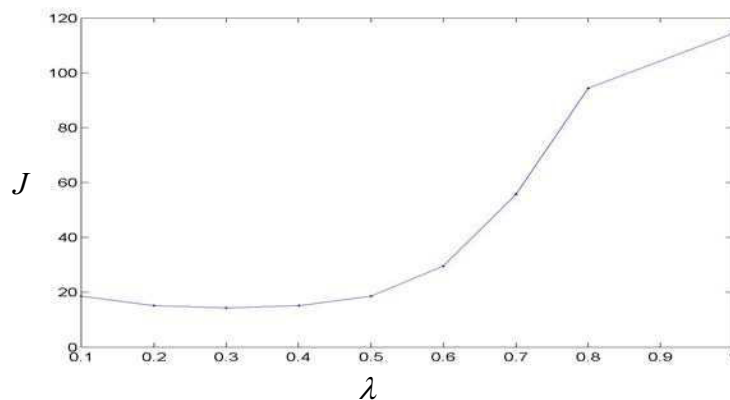
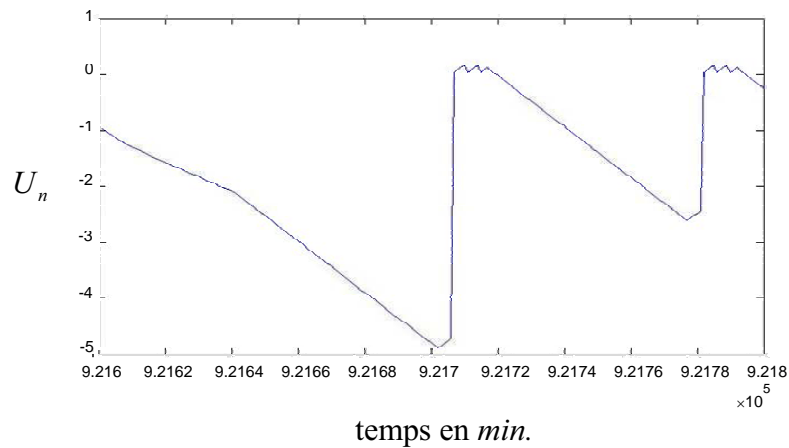
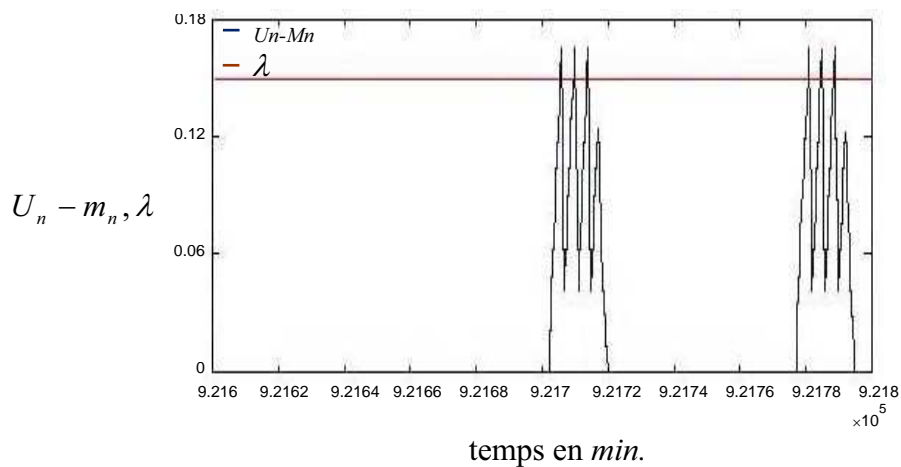


Figure 4.4. Tracé de J

Nous avons appliqué le test de Page en faisant varier p de 2 à 6 et pour chaque p , nous avons calculé λ qui minimise J . Nous avons considéré le cas où la P_d est la plus élevée. Dans ce cas $p = 4$ et $\lambda = 0.15$ avec une valeur de P_d de 90% et celle de τ_{fa} de 3.4%. Nous avons représenté l'évolution de U_n et de $U_n - m_n$ dans les graphes des figures 4.5 et 4.6, respectivement.

Figure 4.5 Tracé de U_n Figure 4.6 Tracé de $U_n - m_n$ et λ

Nous remarquons à partir de la figure 4.5. que U_n décroît au début, jusqu'à l'instant 921714 minutes, où elle croît brutalement. Une fois remis à zéro, U_n décroît de nouveau jusqu'à l'instant 921784 minutes, où elle croît brutalement une deuxième fois. Cette croissance brutale permet l'apparition d'un minimum dans les valeurs de U_n .

Pour cela, nous avons consulté comment croît le facteur U_n par rapport au minimum m_n et si cette différence dépasse le seuil de détection λ , nous sommes dans une situation de présence d'un défaut et les valeurs U_n et m_n sont réinitialisées (la même explication peut se faire pour une diminution de moyenne dans le cas de présence de défauts négatifs).

Nous avons représenté dans le même graphique de la figure 4.6, l'évolution de la quantité $U_n - m_n$ et le seuil de détection λ . Nous remarquons depuis cette figure deux dépassements de seuil aux deux instants 921714 et 921784 minutes.

Les deux croissances brutales représentent des dépassements de saut δ de moyenne dans les valeurs de r_4 .

Nous pouvons conclure que cette méthode permet de détecter les défauts (selon les paramètres de détection choisis) mais avec un retard de détection. Comme nous travaillons hors ligne, ce retard ne représente pas un inconvénient pour cette méthode.

La règle de décision dans ce test de détection de ruptures dans le résidu considéré est une règle d'arrêt. Cet arrêt correspond à un instant que nous avons noté k_a et qui représente également l'instant d'alarme sur la présence de rupture. Cet instant vérifie la relation $k_r < k_a$, où k_r , est l'instant de rupture détectée. Le retard à la détection s'exprime par la relation suivante :

$$\tau_{rd} = k_a - k_r \quad (4.27)$$

L'évaluation théorique de l'efficacité de la méthode de détection est liée à cet instant de rupture. Le test de détection étant influencé par k_r et par la fonction de test $U_n - m_n$ (ou $T_n - M_n$), le choix du seuil est déterminant dans l'évaluation. D'une part, pour un seuil de détection λ fixé, si la valeur de $U_n - m_n$ est faible (ou $T_n - M_n$ est élevée) à l'instant $k_r - 1$ alors le retard à la détection sera grand. D'autre part, la diminution du seuil de détection λ apportera une diminution du retard mais une augmentation de fausses alarmes. Ainsi, l'efficacité d'une méthode de détection est liée aux deux quantités : taux de fausses alarmes et retard à la détection.

Exemple 4.4

Pour le résidu $r_4(k)$ où $k = 1 : 42000$ minutes, nous avons introduit 1500 défauts dans le signal $h_4(k)$.

Pour chaque type de défauts cité ci-dessus, nous avons établi un tableau où nous avons récapitulé les caractéristiques (instants d'apparition et durées) des défauts introduits et la probabilité de détection calculée avec la relation (4.22) pour chaque durée et pour chaque taux fixés.

$\tau_d \%$ \backslash d_d	5	10	15	20	25	100
2	0.5	0.9	1.6	3.4	2.6	13.3
4	0.5	1.3	10.4	52.2	73.5	99.5
6	0.8	2.4	89.4	92.7	94.1	99.6
8	1.3	77.1	95.9	96.2	96.2	99.2
10	1.3	93.6	96.5	96.7	96.6	98.3

Tableau 4.1. $P_d \%$ des défauts constants introduits sur le signal $h_4(k)$

$\tau_d \%$ \backslash d_d	5	10	15	20	25	100
5	0.5	0.8	2.5	7.0	9.4	37.8
10	0.9	3.2	81.3	87.8	89.6	95.8
15	1.3	78.0	92.6	93.5	94.3	98.8
20	1.2	89.9	95.2	95.4	95.6	99.0
25	1.3	94.3	96.4	96.8	96.7	98.5
30	1.3	95.8	96.3	96.7	96.5	98.1
35	1.3	96.0	96.3	96.7	96.6	97.7
40	1.3	96.0	96.2	96.7	96.4	97.1

Tableau 4.2. $P_d \%$ des défauts dépendants des mesures et introduits sur $h_4(k)$ (taux constant)

d_d	5	10	15	20	22	25	30
$P_d \%$	0.9	0.9	12.7	72.0	80.9	86.9	91.8

Tableau 4.3. $P_d \%$ des défauts dépendants des mesures et introduits sur $h_4(k)$ (taux variable)

Nous avons marqué en gras les valeurs de P_d qui sont supérieures ou égales à 72%.

Nous remarquons que la probabilité de détection est significative à partir des taux de variations des défauts d'au moins 10%.

En ce qui concerne le premier cas où des défauts constants ont été introduit sur $h_4(k)$, nous remarquons d'après les résultats donnés dans le tableau 4.1 que P_d est significative à partir de τ_d à 4% et d'une durée de 25 minutes.

Pour le cas des défauts dépendants des mesures avec un τ_d constant pour tous les défauts introduits et défini sur l'intervalle [5%, 40%], nous remarquons que P_d est significative pour τ_d d'au moins 10% et de durée de 15 minutes.

Pour le cas de défauts dépendants des mesures, avec un τ_d variable d'un défaut à un autre et défini dans l'intervalle $[5\%, 15\%]$, nous remarquons d'après le tableau 4.3, que P_d est significative pour une durée de défaut d'au moins 22 minutes.

Nous avons considéré des défauts dépendants des mesures avec un τ_d constant, introduits sur la variable $h_1(k)$, les résultats obtenus sont récapitulés dans le tableau 4.4.

τ_d % \ d_d	5	10	15	20	25	100
5	0.6	1.1	4.0	5.1	7.5	25.7
10	1.1	3.8	75.7	88.1	92.1	98.2
20	1.9	94.2	96.2	96.5	96.6	98.4
30	1.4	96.0	96.2	96.7	96.4	97.1
40	1.3	96.0	96.2	96.7	96.4	96.9

Tableau 4.4. P_d % des défauts dépendants des mesures et introduits sur $s_1(k)$ (taux constant)

Nous remarquons d'après le tableau 4.4 que les résultats obtenus sont légèrement différents de ceux donnés dans le tableau 4.2.

Nous pouvons conclure, d'après les résultats obtenus dans cette section, que la détection ne dépend pas de la forme du défaut (pointe, carrée,...etc.), mais de son amplitude et de sa durée. Elle n'est significative que pour une durée de défaut supérieure à 15 minutes et un taux de variation de ce défaut d'au moins 10% du signal considéré.

4.5.2 Application de la procédure de détection

Nous avons considéré deux cas d'étude de la procédure de détection basée sur le test de Page pour mesurer sa qualité. Ces deux cas consistent en l'application de cette procédure sur :

1. la norme du résidu, (nous avons considéré la valeur absolue du résidu).
2. le résidu filtré par le filtre passe bas donné par la relation (2.3).

Cette application a permis d'obtenir les résultats récapitulés dans les deux tableaux 4.5 et 4.6 respectifs donnés ci-dessous.

Nous avons considéré pour cela des défauts dépendants des mesures avec un taux τ_d constant pour les deux cas d'étude cité ci-dessus. Pour le deuxième cas, nous avons fixé les

durées des défauts à 15 et à 30 minutes et nous avons fait varier le coefficient α du filtre passe-bas.

τ_d % \ d_d	5	10	15	20	25	100
5	0.5	0.8	2.5	7.0	9.3	37.8
10	0.9	3.2	81.3	87.8	89.4	95.5
20	1.3	90.3	95.6	95.7	95.4	98.6
30	2.0	95.5	96.1	96.6	96.2	97.9
40	1.4	95.5	95.9	96.6	96.1	96.9

Tableau 4.5. P_d % des défauts dépendants des mesures et introduits sur $s_1(k)$ (taux constant)

En comparant les résultats donnés dans le tableau 4.5 avec ceux récapitulés dans le tableau 4.2 où la procédure de détection est appliquée directement sur des résidus, nous remarquons que les résultats sont quasiment identiques. Nous pouvons donc conclure que l'application de la procédure de détection sur le résidu ou sa norme ne change pas les performances de la détection.

d_d \ α	0	0.2	0.5	0.8	0.9	0.92	0.95	0.98
15	92.4	91.7	91.5	90.9	85.3	77.9	39.5	4.9
30	97.2	97.2	97.6	97.7	97.2	96.7	95.1	87.2

Tableau 4.6. P_d % des défauts introduits sur $s_4(k)$ pour τ_d de 10%

La valeur de $\alpha = 0$ dans le tableau 4.6, correspond au cas où le résidu n'a pas été filtré.

Dans le cas de l'application de la procédure de détection sur le résidu $r_4(k)$ non filtré (voir tableau 4.2), nous avons remarqué que P_d est significative pour un taux et une durée de défaut de 10% et de 15 minutes respectivement. Pour cela, nous avons considéré dans le tableau 4.6 une durée de défaut supérieure ou égale à 15. Nous remarquons donc d'après la première ligne du tableau correspondant à une durée de défaut de 15 minutes que pour un résidu filtré, plus le facteur α est grand, plus la détection est mauvaise et pour une durée de défaut de 30 minutes, nous remarquons que les P_d sont meilleures mais cela n'améliore pas pour autant la détection en comparant ces P_d à celles données dans le tableau 4.2.

Nous pouvons conclure que l'application de la procédure de détection sur un résidu filtré ou sur la norme du résidu n'améliore pas la détection par rapport à son application sur un résidu non filtré.

4.6 Localisation des défauts

La localisation consiste à chercher dans la matrice des signatures théoriques des défauts, la signature correspondante à une signature de défaut expérimentale observée.

À chaque instant k , les éléments de la signature expérimentale peuvent être déterminés, en utilisant la relation (4.14) pour des signatures binaires et la relation (4.15) pour des signatures ternaires.

4.6.1 Localisation des défauts basée sur le calcul de distance

La localisation des défauts est effectuée à l'aide de calcul de distances entre la signature de défaut expérimentale et les signatures théoriques. Le défaut le plus vraisemblable est celui correspondant à la signature théorique dont la distance à la signature expérimentale est la plus faible. Nous avons considéré la distance euclidienne définie par :

$$d_j(k) = \left(\sum_{i=1}^N (\Lambda_{ij} - st_i(k))^2 \right)^{1/2}. \quad (4.28)$$

Cette phase de localisation est d'autant plus performante que les signatures théoriques sont différentes.

Exemple 4.5

Considérons le même exemple 4.1, le vecteur $x(k) = [h_1(k), h_4(k), v_5, h_{15}(k), h_{37}(k)]$ et $y(k) = [h_4(k), v_5(k), h_{37}(k)]$.

Le sous-ensemble de modèles qui relie ces variables à partir de l'ensemble de modèles établis dans le troisième chapitre a permis d'obtenir l'ensemble des résidus générés, donné dans (4.16) et la matrice des signatures des défauts théoriques, indiquée dans (4.18).

Un défaut de mesure réel est présent sur la variable v_5 entre les instants 978000 et 980200 minutes. L'application de la procédure de détection sur l'ensemble de ces résidus nous a permis d'obtenir les résultats récapitulés dans le tableau 4.7. Ils représentent les durées et les

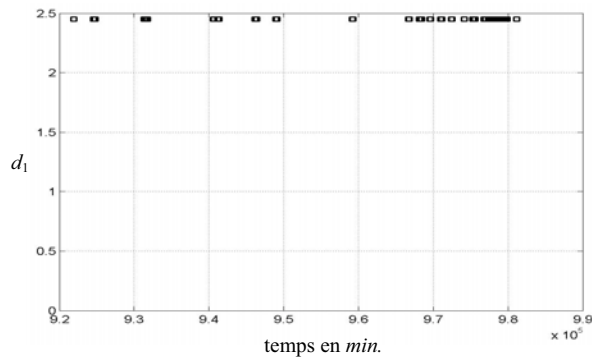
instants d'apparition des ruptures dans les résidus. Les résultats sont notés en gras, souligné et en caractère normal pour les résidus r_4 , r_{15} et r_{37} respectivement. L'ensemble des résultats est classé selon un ordre croissant des instants d'apparition des ruptures.

	Instant d'apparition de rupture en minutes	Durée de rupture en minutes
r_4	921910	33
	924520	37
r_{37}	<u>924740</u>	<u>34</u>
	r_5	
	931350	21
	<u>931590</u>	<u>70</u>
	<u>931680</u>	<u>38</u>
	<u>931740</u>	<u>20</u>
	<u>940610</u>	<u>39</u>
	<u>941180</u>	145
	946160	19
	946210	159
	948970	42
	949040	20
	949080	40
	959180	90
	966730	66
	968110	86
	968430	20
	969560	112
	971020	109
	972450	20
	974120	70
	975320	86
	975570	57
	976770	77
	976940	45
	977030	17
	977410	49
	978030	75
	978140	112
	978310	<u>86</u>
	978520	<u>20</u>
	978590	<u>295</u>
	979040	<u>492</u>
	979750	<u>40</u>
	979810	<u>74</u>
	981120	55

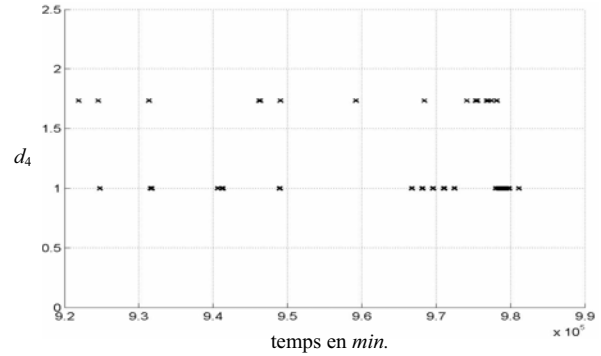
Tableau 4.7. Instants et durées des ruptures détectées dans les résidus

Nous remarquons depuis les résultats donnés dans ce tableau que les ruptures détectées dans les valeurs des résidus sont caractérisées par des durées supérieures ou égales à 17 minutes. Ces ruptures sont des conséquences de présence de défauts dans les mesures ou des fausses alarmes.

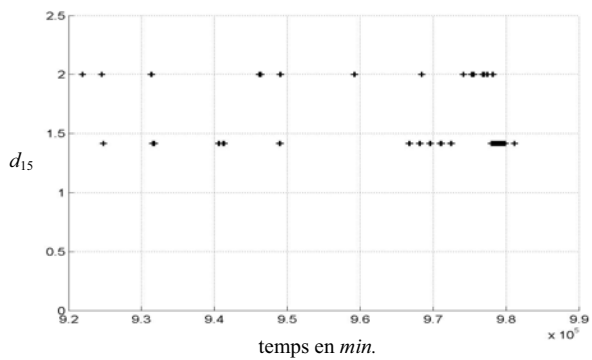
Nous avons appliqué pour compléter cette étape de détection, la procédure de localisation à base de calcul de distance mentionnée ci-dessus. Celle-ci nous a permis de localiser les défauts présents sur l'une ou l'autre des variables considérées. Nous avons calculé une distance pour chaque signature de défaut théorique qui peut être présent sur l'une des variables appartenant à x . Ce calcul est effectué en utilisant la relation (4.28). Nous n'avons considéré les distances qu'aux instants de ruptures détectées. Ces distances sont représentées dans les graphes de la figure 4.7.



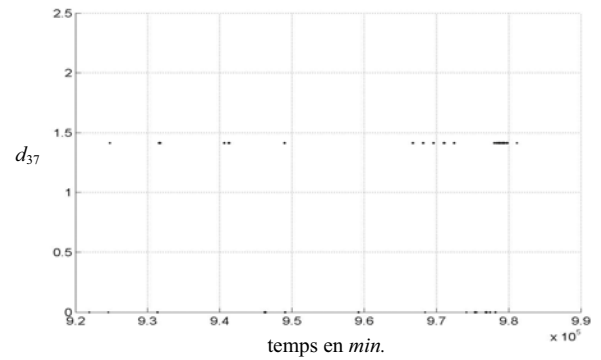
a. Distance Euclidienne d_1 .



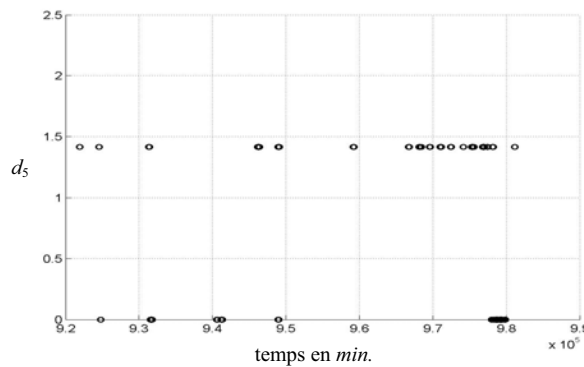
b. Distance Euclidienne d_4 .



c. Distance Euclidienne d_{15} .



d. Distance Euclidienne d_{37} .



e. Distance Euclidienne d_5 .

Figure 4.7. Distance Euclidienne calculée pour chaque signature de défaut théorique

D'après ces graphes, nous remarquons que la distance $d_1(k)$ est maximale pour tout instant k appartenant à toutes les périodes de ruptures mentionnées dans le tableau 4.7. Ce qui permet de conclure qu'aucun défaut n'est présent sur la variable $h_1(k)$. Nous remarquons aussi depuis le graphe 4.7.b que la distance $d_4(k)$ est minimale et vaut un, pour des instants k appartenant à quelques périodes de ruptures, données dans le tableau 4.7. Comme cette distance est éloignée de zéro, la décision sur la présence de défaut étant prise sur le critère de distance minimale, ce qui implique que cette détection ne représente en réalité que des fausses alarmes. En consultant le graphe 4.7.e qui donne la distance d_5 , nous remarquons que celle-ci est minimale et quasiment nulle pendant la période allant de 978030 à 979810 minutes. Cela nous permet de confirmer la présence d'un défaut sur la variable $v_5(k)$. Nous avons bien localisé le défaut réel et donc nous avons une bonne détection mais avec un taux de fausses alarmes, calculé avec la relation (4.27), de 5.5%.

Cette méthode de détection et de localisation ne nous a pas permis de voir la direction du défaut présent sur la variable $v_5(k)$. Autrement dit, le signe du défaut n'est pas défini. En plus le choix des paramètres dans la procédure de détection doit se faire soigneusement pour aboutir à des bons résultats où la probabilité de détection est significative et le taux de fausses alarmes est faible. Les fausses alarmes ont été réduites dans la phase de localisation selon les résultats obtenus ci-dessus.

Deux autres méthodes de détection et de localisation de défaut basées sur des techniques non booléennes de localisation des défauts, permettent de fournir les signes des défauts détectés. Elles sont présentées dans les sections qui suivent. Quelques exemples ont été choisis, pour illustrer ces deux méthodes.

4.6.2 Localisation de défauts par la méthode DMP

La méthode appelée DMP (Diagnostic Model Processor) proposée dans [Petti et al. 90; Petti et Djurjati, 91 ; Petti, 92], s'inspire en partie des travaux donnés dans [Kramer, 87]. Cette méthode a été reprise dans un travail récent [Rengaswamy, 01]. Son principe repose sur une évaluation non booléenne des résidus. Cette évaluation se traduit par le calcul d'un facteur de satisfaction des équations des modèles. La localisation des défaillances est effectuée en combinant les différents facteurs de satisfaction à l'aide d'une somme pondérée par les rapports sensibilité/tolérance de chaque équation de modèle. Cela permet d'éliminer,

ou de réduire l'influence des équations dans lesquelles le capteur n'intervient pas ou intervient avec une sensibilité faible.

Vérification de la cohérence de mesures :

La vérification de la cohérence des observations avec le modèle est effectuée par le calcul de facteurs de satisfaction. Ce dernier noté sf_i , est calculé en fonction de la valeur du résidu r_i prise à chaque instant k par l'expression suivante :

$$sf_i(k) = \text{sign}(r_i(k)) \frac{(|r_i(k)|/\tau_i)^\rho}{1 + (|r_i(k)|/\tau_i)^\rho}, \quad i = 1 \dots l \quad (4.29)$$

avec ρ réel pair et exprime la raideur de la fonction sigmoïde générale donnée par la relation (4.29) comme illustré dans la figure 4.8, τ_i est la tolérance associée au résidu r_i et $\text{sign}(\cdot)$, est la fonction définie par la relation (4.11).

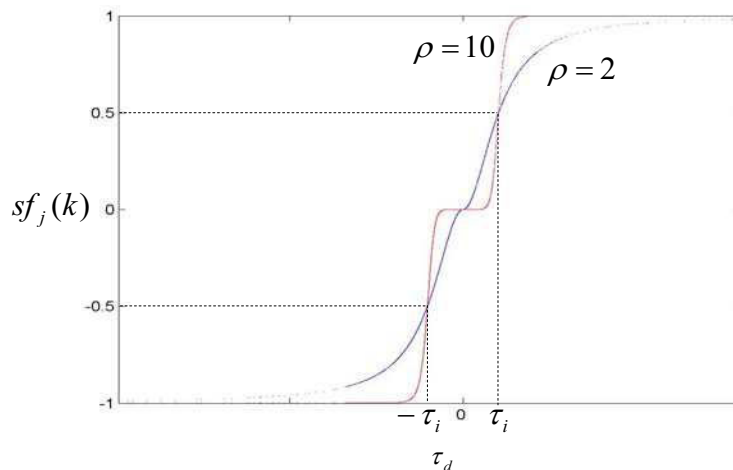


Figure 4.8. Fonction sigmoïde générale utilisée dans DMP. $\rho = 2, 10$

L'expression de ce facteur de satisfaction est en fait inspirée de la fonction de croyance proposée dans [Kramer, 87]. Il s'interprète de la façon suivante : si sf_i tend vers 0, alors la $i^{\text{ème}}$ équation de modèle est parfaitement cohérente avec les observations sinon, s'il tend vers -1 ou +1, l'équation est fortement incohérente, négativement ou positivement.

Signatures théoriques des défauts :

Les signatures théoriques des différentes défaillances δ_j sont établies selon la relation (4.10). Les coefficients S_{ij} des différentes variables qui interviennent dans les équations générant les résidus r_i , présentées dans le tableau 3.10, sont établis. Ils sont rassemblés dans

le tableau E.1 de l'annexe E. La mise en œuvre de la méthode DMP suppose que les effets des défaillances se traduisent par des anomalies sur les mesures.

Localisation des défauts :

Selon [Petti et *al.*, 90; Petti et Djurjati, 91], la localisation des défauts se base sur un critère de discrimination simple. Les facteurs de satisfaction sf_i définis en (4.29) et les sensibilités S_{ij} sont combinées pour déterminer les hypothèses des défauts les plus vraisemblables.

La sensibilité normalisée par rapport aux tolérances est définie par :

$$\tilde{S}_{ij}(k) = \frac{S_{ij}}{\tau_i(k)}, \quad i = 1 \dots l, \quad j = 1 \dots q \quad (4.30)$$

Le facteur de vraisemblance du défaut δ_j est donné par :

$$lf_j(k) = \sum_{i=1}^l sig_{ij}(k) sf_i(k), \quad j = 1 \dots q, \quad (4.31)$$

où les poids $sig_{i,j}(k)$ sont donnés par :

$$sig_{ij}(k) = \frac{\tilde{S}_{ij}(k)}{\sum_{i=1}^l |\tilde{S}_{ij}(k)|}, \quad i = 1 \dots l, \quad j = 1 \dots q. \quad (4.32)$$

La prise de décision concernant la vraisemblance d'une des hypothèses de défaillances δ_j ne dépend donc que des facteurs de satisfaction des équations de modèles qui ne sont affectées que par ce défaut. Lorsque lf_j tend vers 1 ou -1, nous pouvons conclure que l'hypothèse de défaillance est δ_j et la direction du défaut correspondant est positive ou négative respectivement.

Pratiquement et d'après [Surgenor et Jofriet, 92], pour un facteur de vraisemblance lf_j d'une hypothèse de présence du défaut δ_j supérieur en valeur absolue à 0.5, le défaut est vraisemblable et s'il dépasse 0.8, alors la présence du défaut est certaine.

Exemple 4.7

Considérons l'ensemble des résidus donné dans l'exemple 4.1, par la relation (4.16), où un défaut réel δ_5 négatif est présent sur la variable v_5 , dans l'intervalle de temps $[9.780 \cdot 10^5, 9.802 \cdot 10^5]$ en minutes. La matrice M des coefficients S_{ij} est donnée par (4.17) et celle des signatures des défauts théoriques dans (4.18). Les facteurs de vraisemblance sont calculés selon la relation (4.21). Les résultats obtenus sont représentés dans la figure 4.9.

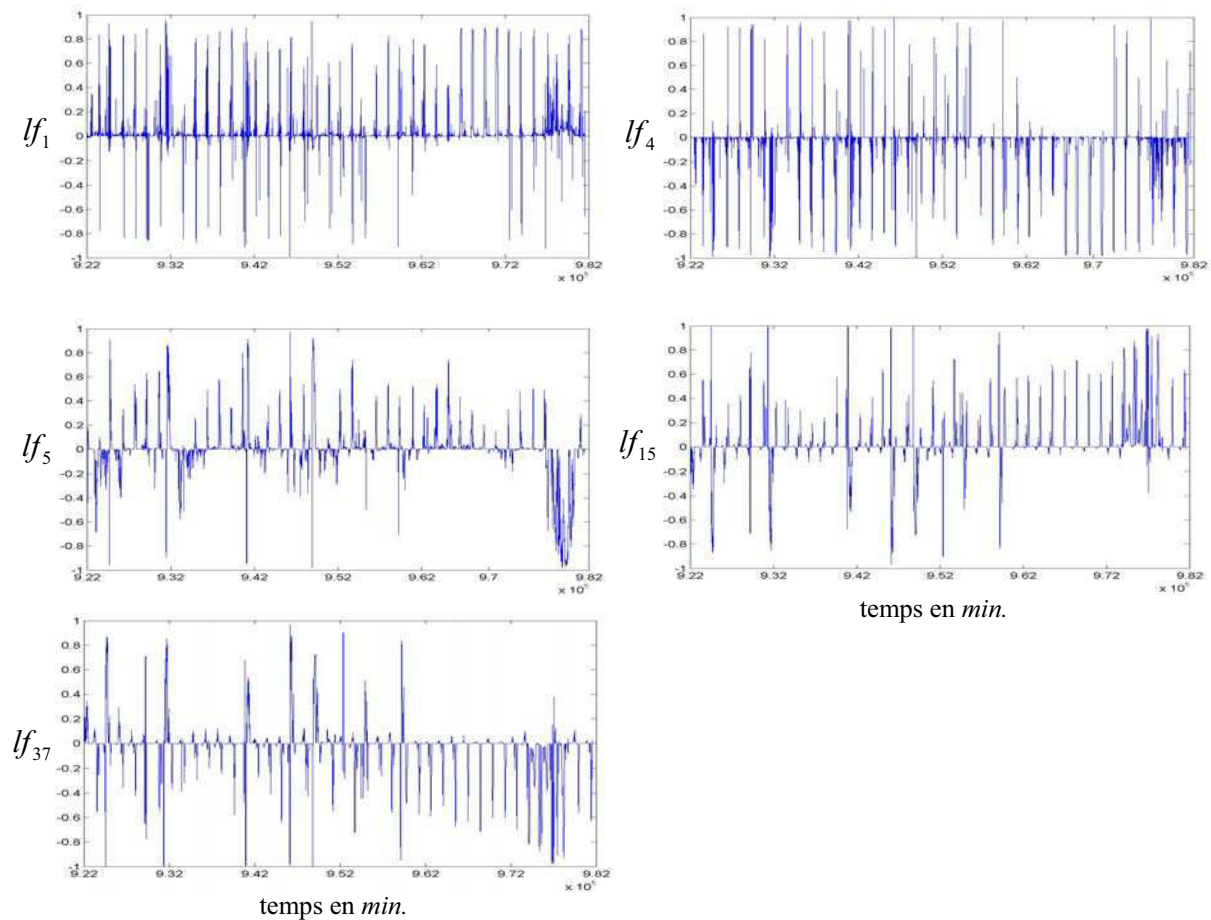


Figure 4.9. Facteurs de vraisemblance des hypothèses de défaillance

Nous remarquons que quelques valeurs des facteurs de vraisemblance dépassent en valeur absolue le seuil 0.8, ce qui signifie la présence des défauts à ces instants de dépassement. La durée de dépassement de ce seuil du facteur de vraisemblance Lf_5 du défaut δ_5 , est la plus significative, elle est de l'ordre de 5000 minutes. Pour les autres facteurs de vraisemblance de défauts, les durées de dépassement de ce seuil, sont inférieures à 100 minutes. Par conséquent, la présence d'un défaut sur la variable v_5 est certaine. Les autres détections ne représentent que des fausses alarmes.

Avec cette méthode, nous remarquons que la valeur du degré de vraisemblance d'un défaut, ne dépend que de la valeur des résidus qui sont sensibles à ce défaut. L'efficacité de cette méthode peut être améliorée en prenant en compte les résidus qui ne sont pas sensibles à ce défaut.

4.6.3 Localisation de défauts basée sur les concepts de la logique floue

Cette méthode d'évaluation des résidus s'appuie sur des notions élémentaires relatives à la théorie des ensembles flous [Chen et *al.*, 00]. Ainsi, chaque résidu peut être caractérisé par trois ensembles flous appelés modalités r_i^+ , r_i^- et r_i^0 ‘positif’, ‘négatif’ et ‘nul’ respectivement. Ces fonctions prennent des valeurs dans $[0, 1]$.

A chaque instant k , chaque résidu r_i appartient à une des trois modalités décrites ci-dessus. Des fonctions d'appartenance du résidu r_i , à ces modalités sont de forme sigmoïde comme indiqué dans la figure 4.10. Elles sont définies par les relations (4.33), (4.34) et (4.35).

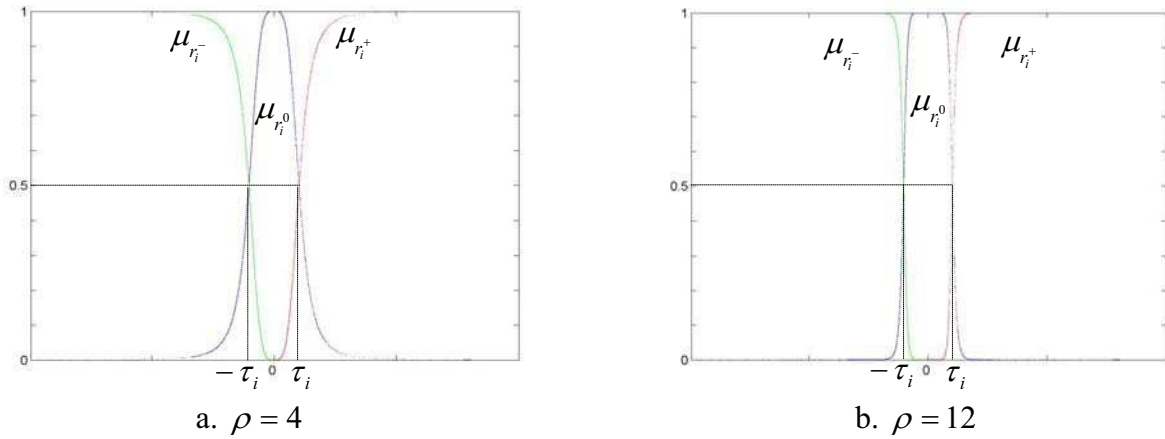


Figure 4.10. Fonction sigmoïde générale utilisée dans la localisation à base de concepts de la logique floue, $\rho = 4, 12$

$$\mu_{r_i^0}(k) = 1 - \frac{(r_i(k)/\tau_i)^\rho}{1 + (r_i(k)/\tau_i)^\rho}, \quad \text{avec } \rho \text{ est un réel pair} \quad (4.33)$$

$$\mu_{r_i^+}(k) = \begin{cases} 0 & \text{si } r_i(k) \leq 0 \\ 1 - \mu_{r_i^0}(k) & \text{si } r_i(k) > 0 \end{cases}, \quad (4.34)$$

$$\mu_{r_i^-}(k) = \begin{cases} 1 - \mu_{r_i^0}(k) & \text{si } r_i(k) \leq 0 \\ 0 & \text{si } r_i(k) > 0 \end{cases}, \quad (4.35)$$

où ρ est un facteur de forme et τ_i est la tolérance associée au résidu r_i . Cette tolérance représente le seuil pour lequel les valeurs du résidu sont acceptées (les valeurs pour lesquelles le résidu n'est pas affecté par des ruptures). Elle est fixée en fonction de la variance du résidu.

La localisation des défauts est basée sur l'analyse conjointe des résidus et des signatures ternaires théoriques. Elle permet de déduire de la valeur des différents résidus, la variable affectée par un ou plusieurs défauts. Pour cela, on construit une base de règles (règles d'inférence floue) liant symptômes (déviations de résidus) et causes (présence d'un défaut). Chaque règle comporte :

- une partie prémisse relative aux différents résidus de la forme :
si $r_1(k)$ est $r_1^{g_{1j}}$ et $r_2(k)$ est $r_2^{g_{2j}}$ et... $r_i^{g_{ij}}$ et... $r_l(k)$ est $r_l^{g_{lj}}$,

où $j = 1 : q$, $r_i^{g_{ij}}$ représente l'une des trois modalités r_i^+ , r_i^- ou r_i^0 associée au résidu r_i selon la valeur de g_{ij} définie par :

$$g_{ij} = \begin{cases} +, & \text{pour } \Lambda_{ij} > 0 \\ -, & \text{pour } \Lambda_{ij} < 0, \\ 0, & \text{pour } \Lambda_{ij} = 0 \end{cases} \quad (4.36)$$

avec Λ_{ij} est définie par la relation (4.10).

- une partie conséquence de la forme : *alors* δ_j .

La quantification des différentes propositions s'effectue au moyen des valeurs des fonctions d'appartenance des résidus appelées aussi degrés d'appartenance. Le coefficient $\mu_{r_i^{g_{ij}}}(k)$ désigne la quantification de la proposition « $r_i(k)$ est $r_i^{g_{ij}}$ ».

A chaque instant et pour chaque règle de base, le degré de véracité d'une conséquence doit être évalué en fonction de la validité des prémisses. Le « et » dans une proposition est un opérateur de conjonction ou d'agrégation qui permet de combiner des degrés d'appartenance des différentes propositions. Différents opérateurs peuvent être utilisés, les plus courants étant les opérateurs « minimum » indiqué par « min » et « moyenne ». L'utilisation de ces opérateurs conduit aux expressions suivantes du degré de véracité $\mu_\delta(k)$ d'une règle :

$$\mu_{\delta_j}(k) = \min\left(\mu_{r_i^{g_{ij}}}(r_i(k))\right), \quad (4.37)$$

$$\mu_{\delta_j}(k) = \frac{1}{l} \sum_{i=1}^l \mu_{r_i^{g_{ij}}}(r_i(k)), \quad (4.38)$$

où $j = 1 : q$, δ_j est le défaut affectant la variable y_j (δ_j^-, δ_j^+ sont des défauts négatif et positif qui peuvent affectés chaque variable y_j) et g_{ij} est définie par la relation (4.36) ; elle représente l'influence du défaut δ_j sur r_i .

Cependant, ces deux opérateurs sont mal adaptés au contexte de localisation de défauts. En effet, si un résidu est de faible sensibilité au défaut (le cas par exemple de présence de défaut de faible amplitude, l'utilisation de l'opérateur « minimum » conduit à un degré de véracité relative à ce défaut très faible. L'opérateur « moyenne » est moins sensible à ce phénomène. Cependant, si le nombre de résidus insensibles à un défaut est trop important, les valeurs du degré de véracité de la règle relative à ce défaut seront relativement élevées en l'absence du défaut considéré. Afin d'éviter ces inconvénients et selon [Mandel et Ragot, 98], les modalités r_i^0 d'une part et r_i^+ et r_i^- sont traitées séparément en combinant minimum et moyenne.

L'application de cet opérateur d'agrégation sur les fonctions d'appartenance permet d'obtenir le degré de véracité donné par :

$$\mu_{\delta_j}(k) = \min \left(\frac{\sum_{i \in \{i / g_{ij} \in \{0\}\}} \mu_{r_i^{g_{ij}}}}{|\{i / g_{ij} \in \{0\}\}|}, \frac{\sum_{i \in \{i / g_{ij} \in \{-,+\}\}} \mu_{r_i^{g_{ij}}}}{|\{i / g_{ij} \in \{-,+\}\}|} \right), \quad j = 1 : q \quad (4.39)$$

où g_{ij} est défini par la relation (4.33) $|\{e\}|$ représente le cardinal de l'ensemble $\{e\}$, $\mu_{r_i^{g_{ij}}}$ est la fonction d'appartenance du résidu r_i à la modalité $r_i^{g_{ij}}$ définie par les relations (4.33), (4.34) ou (4.35), pour $g_{ij} = 0, +, \text{ ou } -$.

Dans le cas où le nombre des résidus sensibles à un défaut est supérieur à celui des résidus insensibles à ce défaut, l'opérateur minimum est le plus approprié par rapport à celui proposé dans [Mandel et Ragot, 98] et donné par (4.39). Pour conserver les qualités des deux

opérateurs, nous avons considéré un nouvel opérateur d'agrégation [Hamioud et al., 05a] qui permet d'obtenir un degré de véracité comme suit :

$$\mu_{\delta_j}(k) = \min \left(\frac{\sum_{i \in \{i / g_{ij} \in \{0\}\}} \mu_{r_i^{g_{ij}}}}{|\{i / g_{ij} \in \{0\}\}|}, \min(\mu_{r_i^{g_{ij}}} / g_{ij} \in \{-, +\}) \right), \quad j = 1 : q \quad (4.40)$$

Nous avons considéré l'exemple 4.8 et l'exemple 4.9 pour comparer les trois opérateurs d'agrégation définis par (4.37), (4.39) et (3.40). Une partie de cette comparaison a été effectuée dans [Hamioud et al., 05b] où nous n'avons considéré que les deux expressions (4.39) et (4.40).

Exemple 4.8

Considérons l'ensemble des résidus donné dans l'exemple 4.1, par la relation (4.17). Dans une première étape, la matrice des signatures des défauts théoriques donnée par (4.18) est utilisée pour générer la base des règles d'inférence. Cette base est décrite dans le tableau 4.8 où δ_j^- et δ_j^+ désignent respectivement des défauts négatif et positive affectant la variable y_j .

règles	modalités									défauts
	r_4			r_5			r_{37}			
	r_4^-	r_4^0	r_4^+	r_5^-	r_5^0	r_5^+	r_{37}^-	r_{37}^0	r_{37}^+	
1			•			•			•	δ_1^-
2	•			•			•			δ_1^+
3	•			•				•		δ_4^-
4			•			•		•		δ_4^+
5		•		•				•		δ_5^-
6		•				•		•		δ_5^+
7		•			•				•	δ_{15}^-
8		•			•		•			δ_{15}^+
9		•			•		•			δ_{37}^-
10		•			•				•	δ_{37}^+
11		•			•			•		Sans défauts

Tableau 4.8. Structure de la base des règles des inférences

Dans une deuxième étape, nous représentons la « fuzification » c'est-à-dire la représentation floue des valeurs des résidus, et calculons les valeurs des trois fonctions d'appartenance $\mu_{r_i^-}$, $\mu_{r_i^+}$ et μ_{r_i} définies par les relations (4.33), (4.34) et (4.35) pour chaque résidu r_i ainsi que les degrés de véracités définis par les relations (4.37), (4.39) et (4.40).

Nous avons représenté ces degrés de véracité respectivement dans les graphes des figures 4.11, 4.12 et 4.13, pour chaque hypothèse de présence de défaut négatif ou positif sur chacune des variables h_1, h_4, v_5, h_{15} , et h_{37} .

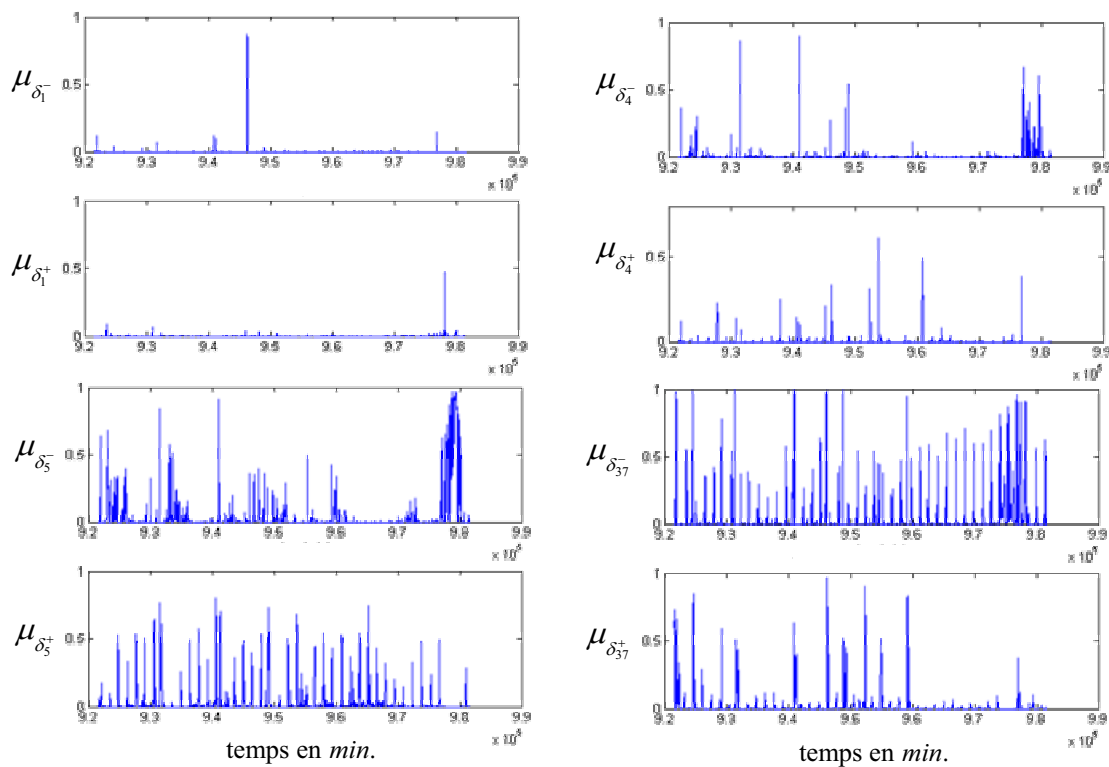


Figure 4.11. Degrés de véracité des différentes hypothèses, opérateur d'agrégation « minimum ».

Nous remarquons d'après ces graphes, que le degré de véracité $\mu_{\delta_5^-}$ de présence de défaut négatif affectant la variable v_5 prend des valeurs proches de 1 sur la période allant de 978000 à 980200 minutes. Ce défaut est réel et est confirmé par un superviseur.

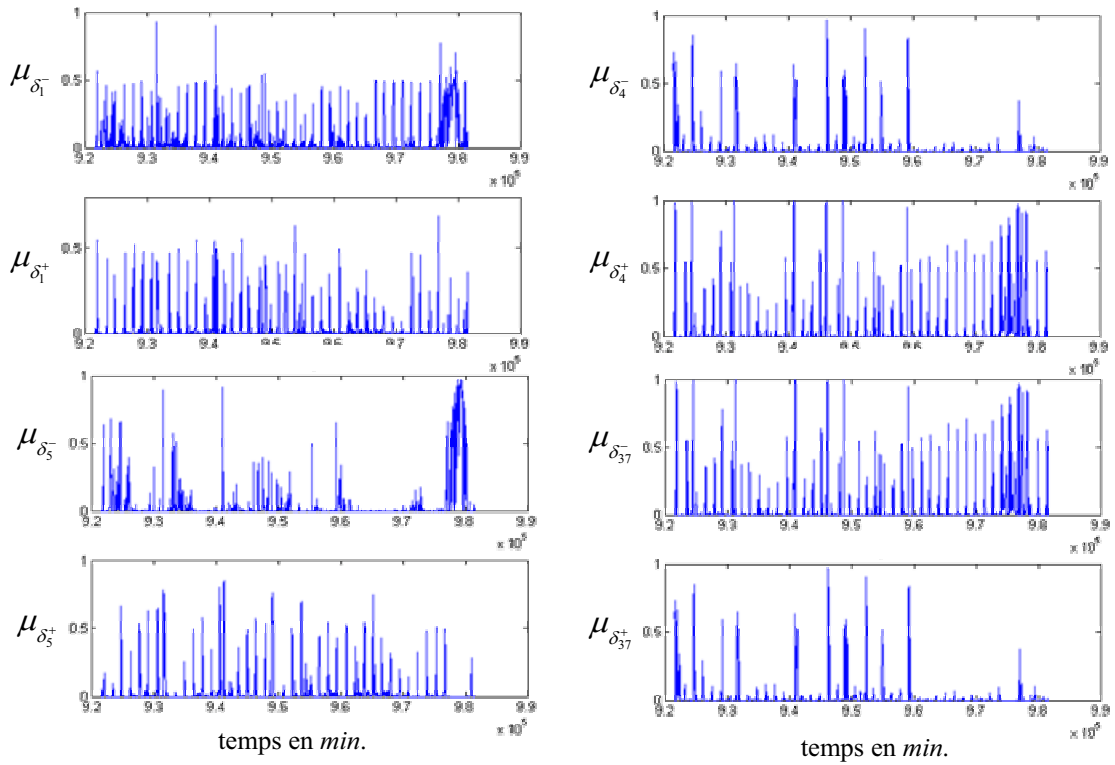


Figure 4.12. Degrés de véracité des différentes hypothèses, opérateur d'agrégation selon Mandel

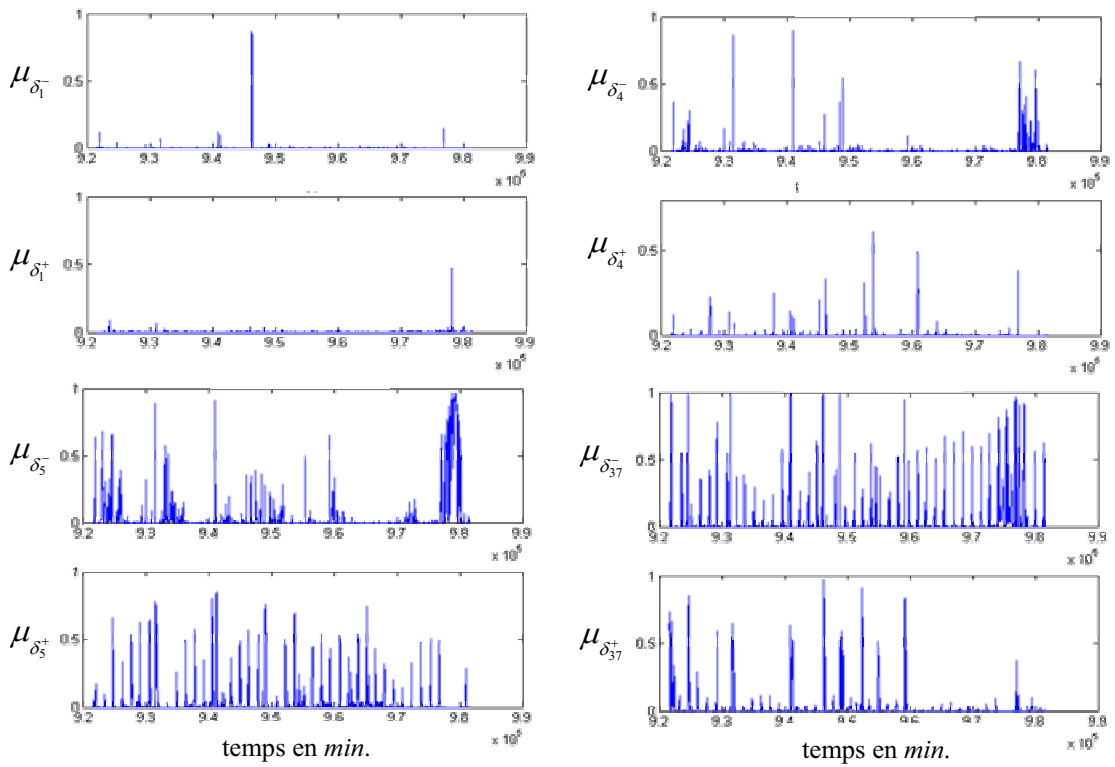


Figure 4.13. Degrés de véracité des différentes hypothèses, nouvel opérateur d'agrégation

Nous avons calculé pour cet exemple les probabilités de détection, les probabilités de fausses alarmes et les taux de fausses alarmes calculées par les relations (4.22) (4.23) et (4.24) respectivement et ce pour les trois cas d'agrégation décrits précédemment. Les résultats obtenus sont récapitulés dans les trois tableaux (4.9), (4.10) et (4.11) respectifs.

S	P_d	P_{fa}	τ_{fa}
0.8	0.30	0.02	0.66
0.6	0.52	0.06	0.74
0.5	0.67	0.11	0.80

Tableau 4.9. P_d , P_{fa} et τ_{fa} de défaut réel, opérateur « minimum »

S	P_d	P_{fa}	τ_{fa}
0.8	0.32	0.03	0.69
0.6	0.55	0.08	0.78
0.5	0.69	0.14	0.84

Tableau 4.10. P_d , P_{fa} et τ_{fa} de défaut réel, opérateur proposé par Mandel

S	P_d	P_{fa}	τ_{fa}
0.8	0.32	0.03	0.68
0.6	0.55	0.07	0.77
0.5	0.70	0.12	0.82

Tableau 4.11. P_d , P_{fa} et τ_{fa} de défaut réel, nouvel opérateur

Nous remarquons d'après les trois tableaux donnés ci-dessus ce qui suit :

- les P_d sont légèrement élevées pour le cas de l'opérateur Mandel et du nouvel opérateur par rapport au cas de l'opérateur « minimum »,
- l'opérateur Mandel permet d'obtenir une probabilité de détection élevée mais aussi un taux de fausses alarmes, élevé par rapport au cas de l'opérateur « minimum »,
- l'opérateur « minimum » permet de diminuer le taux de fausses alarmes mais avec une plus faible probabilité de détection,
- le nouvel opérateur permet d'obtenir une probabilité de détection élevée avec un taux de fausses alarmes faible par rapport à celui obtenu dans le cas de l'opérateur Mandel.

Nous avons considéré un autre exemple où les défauts sont simulés cette fois-ci en considérant seulement le calcul de la probabilité de détection.

Exemple 4.9

Reprenons le même exemple 4.5 en introduisant un ensemble de défauts dans les valeurs de la variable h_4 . Ces défauts dépendent des mesures de la variable h_4 et sont définis selon la relation (4.26). Ils ont une durée de 15 minutes et un taux, τ_d fixé à 5%, 10%, 20% ou 30%.

L'application de la procédure de détection et de localisation nous a permis d'établir les trois tableaux (4.12), (4.13) et (4.14), où nous avons récapitulé les probabilités de détection calculées par la relation (4.22) pour les trois cas d'agrégation.

τ_d % \ S	0.8	0.7	0.6
5	1.0	2.0	10.5
10	67.5	74.0	78.0
20	76.0	80.5	84.0
30	78.0	82.0	85.5

Tableau 4.12. P_d des défauts introduits sur $h_4(k)$ (taux constant), opérateur « minimum »

τ_d % \ S	0.8	0.7	0.6
5	1.5	4.0	13.0
10	71.0	80.0	84.0
20	79.5	86.5	91.0
30	82.0	88.0	92.5

Tableau 4.13. P_d des défauts introduits sur $h_4(k)$ (taux constant), opérateur proposé par Mandel

τ_d % \ S	0.8	0.7	0.6
5	1.5	4.0	13.0
10	71.0	80.0	84.0
20	79.5	86.5	91.0
30	82.0	88.0	92.5

Tableau 4. 14. P_d des défauts introduits sur $h_4(k)$ (taux constant), nouvel opérateur

Nous remarquons d'après les trois tableaux donnés ci-dessus, que les P_d sont élevées pour le cas de l'opérateur Mandel et du nouvel opérateur.

Prise en compte de la persistance du défaut

La persistance temporelle des symptômes observés est un indicateur important dans la localisation des défauts.

Comme il a été déjà indiqué précédemment, un défaut réel persiste et sa durée est significative. Il est donc intéressant de considérer un indicateur de persistance temporelle de défauts. En plus si un défaut d'amplitude positive faible se présente, la valeur du degré de véracité associé à ce défaut est faible. Dans le cas où ce défaut est persistant, le degré de véracité associé peut être significatif en prenant en compte l'indicateur de persistance de ce défaut. Ainsi, la considération de cet indicateur, permet d'obtenir des degrés de véracité de valeurs élevées pour des défauts de faibles amplitudes à condition que ces défauts soient persistants.

Pour chaque résidu $r_i(k)$ sont associés les indicateurs de persistance de défaut $p_i^-(k)$, $p_i^0(k)$ et $p_i^+(k)$ pour une persistance de défaut négatif, pas de persistance et une persistance de défaut positif respectivement. Ils correspondent à la persistance des degrés d'appartenance $\mu_{r_i^-}(k)$, $\mu_{r_i^0}(k)$ et $\mu_{r_i^+}(k)$ respectifs. Ces indicateurs sont donnés par la relation (4.38) et sont définis par un rapport du nombre de valeurs des degrés d'appartenance (qui dépassent un certain seuil S_p , à fixer) sur la largeur d'une fenêtre bien définie (notée L).

$$\begin{aligned} p_i^-(k) &= \frac{N(\mu_{r_i^-}(j))}{L}, \\ p_i^0(k) &= \frac{N(\mu_{r_i^0}(j))}{L}, \\ p_i^+(k) &= \frac{N(\mu_{r_i^+}(j))}{L}, \end{aligned} \quad (4.41)$$

où $j = k : L + k$, et

$$\begin{aligned} N(\mu_{r_i^-}(j)) &= \left| \left\{ \mu_{r_i^-} : \mu_{r_i^-}(j) > S_p \right\} \right|, \\ N(\mu_{r_i^0}(j)) &= \left| \left\{ \mu_{r_i^0} : \mu_{r_i^0}(j) > S_p \right\} \right|, \\ N(\mu_{r_i^+}(j)) &= \left| \left\{ \mu_{r_i^+} : \mu_{r_i^+}(j) > S_p \right\} \right|, \end{aligned} \quad (4.42)$$

Les degrés d'appartenance qui prennent en compte la persistance des symptômes sont notés $\mu_{\mathcal{P}_{r_i^-}}(k)$, $\mu_{\mathcal{P}_{r_i^0}}(k)$ et $\mu_{\mathcal{P}_{r_i^+}}$. Ils s'explicitent :

$$\mu_{\mathcal{P}_{r_i^-}}(k) = \frac{p_i^-(k) + \mu_{r_i^-}(k)}{2}, \quad (4.43)$$

$$\mu_{\mathcal{P}_{r_i^0}}(k) = \frac{p_i^0(k) + \mu_{r_i^0}(k)}{2}, \quad (4.44)$$

$$\mu_{\mathcal{P}_{r_i^+}}(k) = \frac{p_i^+(k) + \mu_{r_i^+}(k)}{2}. \quad (4.45)$$

Les termes permettant le calcul de ces nouveaux degrés d'appartenance peuvent être pondérés. Dans notre cas le poids considéré pour chacun des termes donnés par les relations (4.43), (4.44) et (4.45) est égal à $\frac{1}{2}$.

De cette manière, plus les symptômes de défaut pour les valeurs d'un résidu sont persistants, plus la valeur de la fonction d'appartenance de ce résidu à la modalité associée à ce défaut est élevée.

Pour prendre en compte les symptômes des défauts, nous avons considéré l'approche de localisation floue définie dans la section (4.5.2) en remplaçant les degrés d'appartenance $\mu_{r_i^-}(k)$, $\mu_{r_i^0}(k)$ et $\mu_{r_i^+}(k)$ par les degrés d'appartenance $\mu_{\mathcal{P}_{r_i^-}}(k)$, $\mu_{\mathcal{P}_{r_i^0}}(k)$ et $\mu_{\mathcal{P}_{r_i^+}}(k)$ pour calculer les degrés de véracité correspondants à chaque défaut. Ainsi, les degrés de véracité décrits par la relation (4.30), sont exprimés comme suit :

$$\mu_{\delta_j}(k) = \min \left(\frac{\sum_{i \in \{i / g_{ij} \in \{0\}\}} \mu_{\mathcal{P}_{r_i^{g_{ij}}}}}{|\{i / g_{ij} \in \{0\}\}|}, \min(\mu_{\mathcal{P}_{r_i^{g_{ij}}}} / g_{ij} \in \{-, +\}) \right) \quad (4.46)$$

Nous avons considéré l'exemple 4.10 pour illustrer les deux cas : prise en compte de persistance temporelle de défaut comme il a été traité dans [Hamioud et al., 05a] et non prise en compte de celle-ci.

Exemple 4.10

Nous avons considéré les deux résidus $r_4(k)$ et $r_{37}(k)$ donnés dans l'exemple 4.1, avec introduction d'un défaut dans les valeurs de $h_4(k)$ à l'instant $k = 920000$ minutes. Ce défaut est défini par la relation (4.26), sur une durée de 400 minutes avec un taux faible fixé à 10%.

Les degrés de véracités de présence de défauts sur les variables h_1, h_4, h_{15} et h_{37} , sont calculés dans un premier cas par la relation (4.40) et dans un deuxième par la relation (4.46). Ils sont ensuite représentés respectivement dans les graphes des deux figures (4.14) et (4.15).

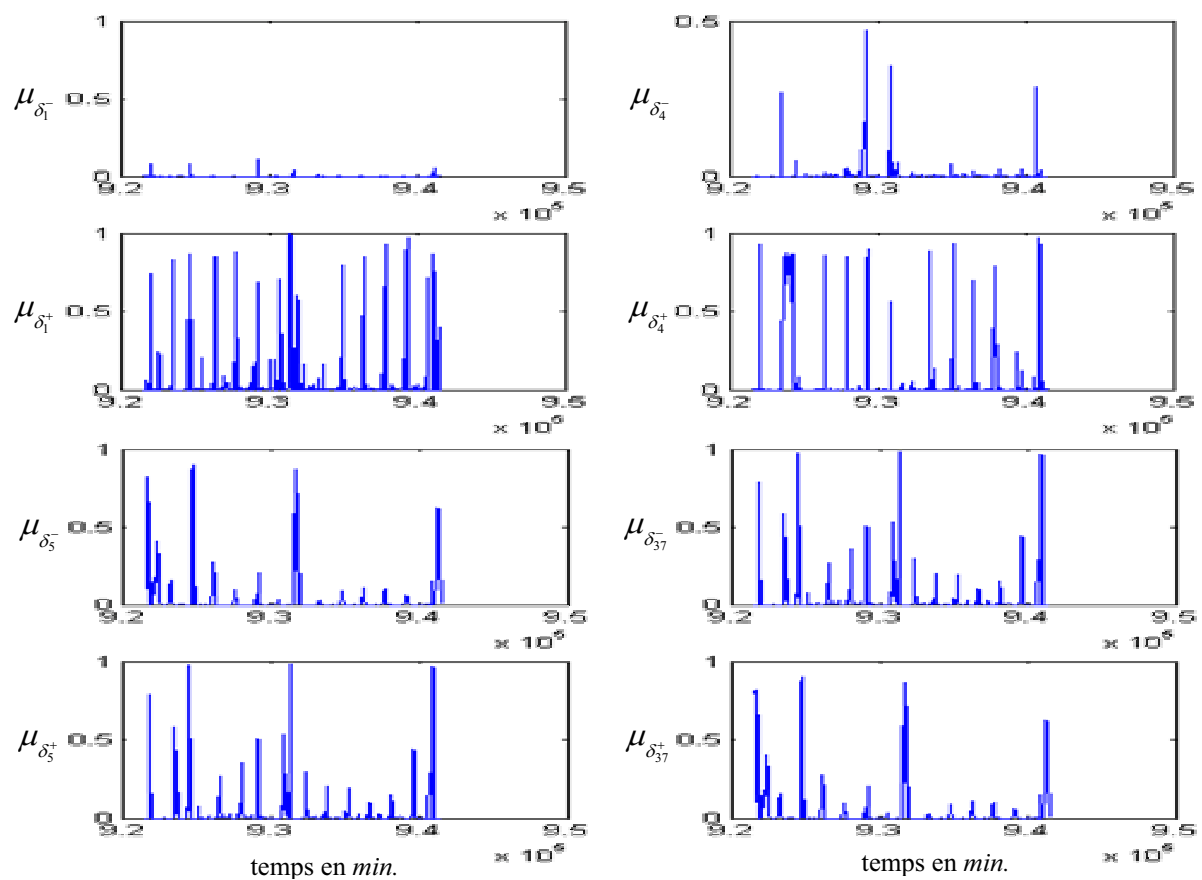


Figure 4.14. Degrés de véracité de chaque hypothèse sans prise en considération de la persistance de défaut

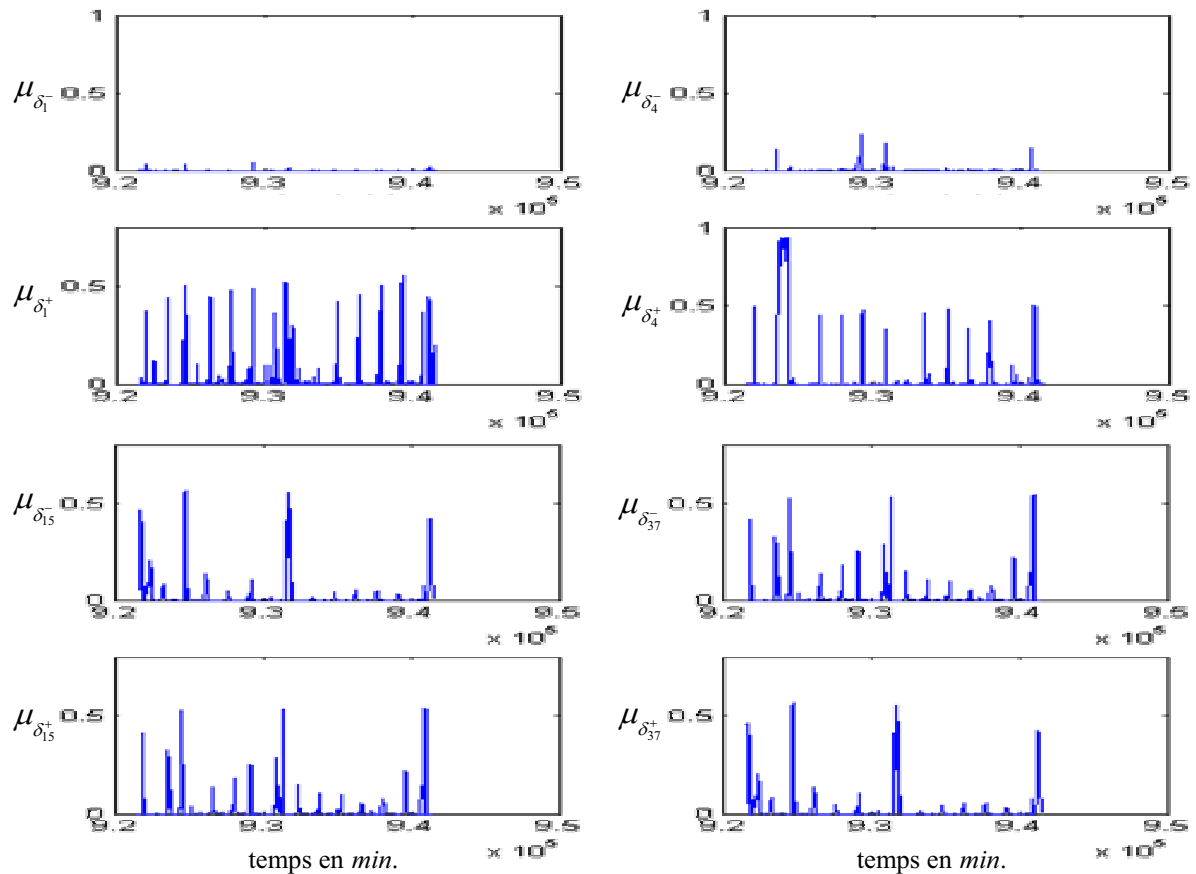


Figure 4.15. Degrés de véracité de chaque hypothèse avec prise en considération de la persistance de défaut

Nous remarquons d'après les deux figures 4.14 et 4.15 que le degré de véracité de présence de défaut positif sur la variable h_4 , indiqué par $\mu_{\delta_4^+}$, dépasse le seuil fixé à 0.6. Ce seuil a été fixé en se basant sur les résultats indiqués dans le tableau 4.13 où la probabilité de détection est supérieure à 50%.

Nous avons calculé la probabilité de détection et le taux de fausses alarmes avec les deux relations (4.21) et (4.23) pour les deux cas traités ci-dessus. La probabilité de détection et le taux de fausses alarmes sont de 94% et 88% pour le premier cas et de 94% et 16% pour le deuxième cas.

Nous avons repris le même exemple 4.5 pour l'étude d'un cas réel et nous avons appliqué cette fois-ci la méthode de localisation avec prise en compte de la persistance de défauts. Les résultats obtenus sont représentés par les graphes de la figure suivante :

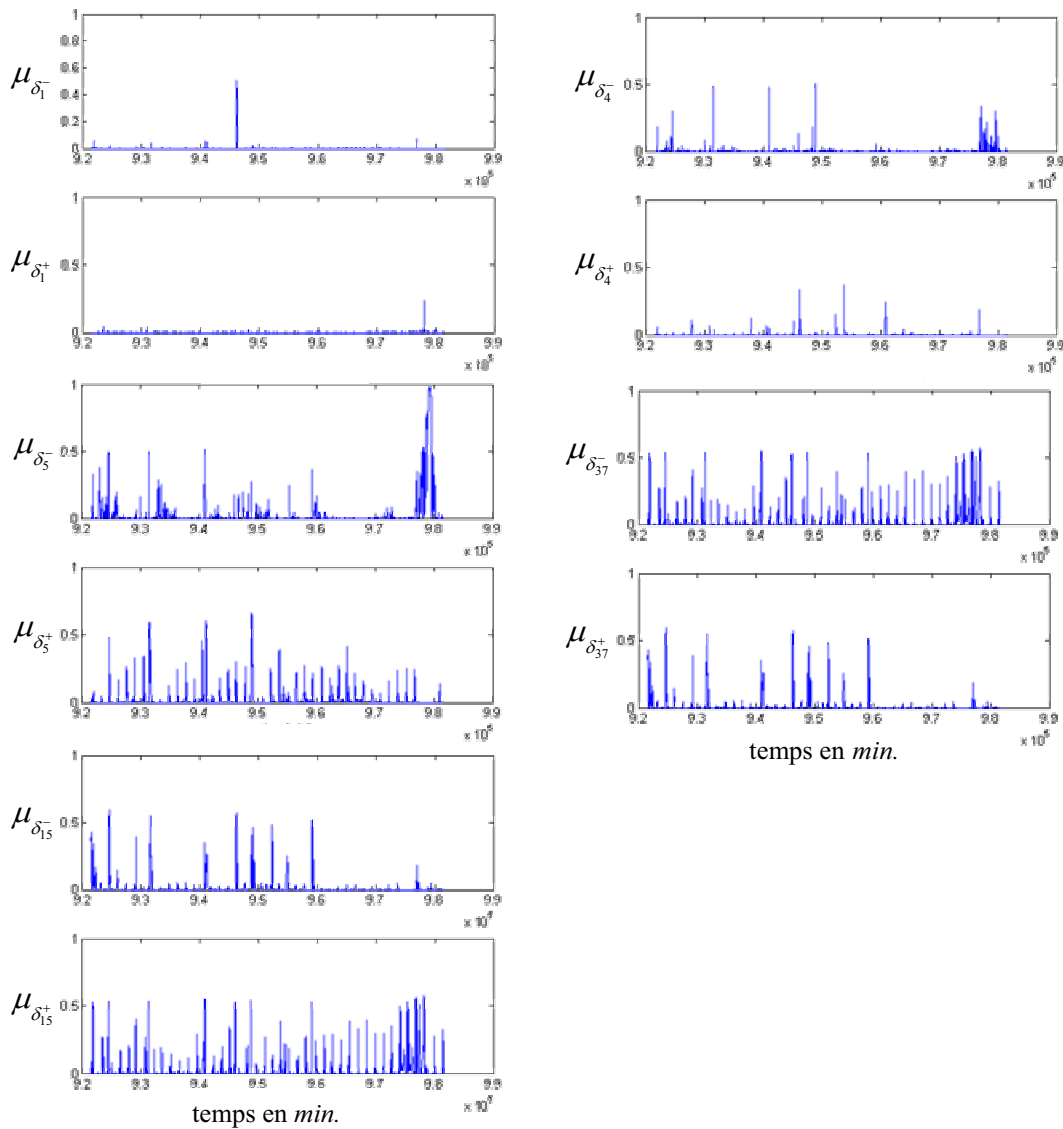


Figure 4.16. Degrés de véracité des différentes hypothèses concernant la présence des défauts

Nous avons calculé la probabilité de détection et le taux de fausses alarmes pour un seuil fixé à 0.6. La probabilité de détection obtenue est de 55% et le taux de fausses alarmes est de 17%. Hors dans le cas de non prise en compte de persistance de défaut pour un seuil de 0.6 (voir tableau 4.11), le taux de fausses alarmes est plus élevé, il est de 77%.

A partir de cet exemple de présence de défaut réel et de l'exemple 4.7 pour le cas de défaut simulé, nous pouvons conclure que la prise en compte de persistance de défaut diminue énormément le taux de fausses alarmes.

4.7 Comparaison des trois procédures de détection et de localisation considérées

Dans cette section, nous comparons les procédures de détection et de localisation : Celle basée sur le test de Page associée au calcul de distance, celle basée sur les concepts de la logique floue et celle dite de DMP.

Pour ce faire, nous avons repris le même exemple 4.5 qui correspond au cas de présence d'un défaut réel. Nous avons calculé la probabilité de détection, la le taux de fausses alarmes et la probabilité de fausses alarmes en utilisant les relations (4.22), (4.23) et (4.24) pour chaque cas.

En fait, pour chacune des trois méthodes, il est nécessaire de fixer certains paramètres. Ces paramètres sont différents d'une procédure à l'autre. Ainsi, pour que la comparaison ait un sens, nous avons défini l'intervalle dans lequel la probabilité de détection doit être obtenue. Les limites de cet intervalle sont fixées par une simple vérification des résultats obtenus pour chaque exemple considéré pour chacune des trois procédures. Les résultats obtenus sont récapitulés dans le tableau 4.15.

Méthodes	P_d	P_{fa}	τ_{fa}
Page-distance	46	3.2	65
A base de flou	55	0.32	17
A base de DMP	56	18	90

Tableau 4. 15. P_d , P_{fa} et τ_{fa} pour les trois procédures de détection et localisation d'un défaut réel

Nous remarquons d'après ces résultats que les probabilités de détection pour les procédures DMP et celle à base de concepts flous sont relativement significatives. Nous remarquons également que le taux de fausses alarmes pour la procédure à base des concepts flous est relativement très faible et vaut 17%.

Nous avons calculé le retard à la détection en utilisant la relation (4.27) pour les trois procédures citées précédemment. En ce qui concerne la procédure à base de test de Page-distance, une simple consultation du tableau 4.7, nous permet d'avoir : $k_r = 978310$ minutes. Sachant qu'un défaut réel est présent à l'instant $k_r = 978000$ minutes, nous avons donc, selon la relation (4.27), $\tau_{rd} = 310$ minutes. En ce qui concerne les deux autres procédures, ce retard de détection est presque nul.

En ce qui concerne la procédure de Page-distance, la détection des défauts ainsi que la localisation est possible mais avec un retard pour la détection. Elle est aussi intéressante pour des défauts de durées supérieures à 15 minutes sinon, le taux de fausses alarmes est trop élevé. Quant à la méthode à base de floue, elle est meilleure par rapport aux deux autres procédures car elle permet d'obtenir des taux de fausses alarmes plus faibles.

4.8 Conclusion

Nous avons appliqué trois procédures de détection-localisation des défauts.

La première méthode basée sur le test de Page pour la détection et sur le calcul de distance pour la localisation des défauts, permet une détection significative avec un faible taux de fausses alarmes pour des défauts caractérisés par des amplitudes définies avec un taux supérieur à 20% et de durées supérieures ou égales à 15 minutes. Cette procédure présente l'inconvénient de détection des défauts avec un retard. Cette méthode reste donc applicable en pratique car les défauts réels qui peuvent se présenter sur les mesures de notre système sont généralement caractérisés par des durées supérieures à 15 minutes.

La deuxième méthode basée sur les concepts de la logique floue est plus fiable, elle représente l'avantage de pouvoir détecter des défauts de faibles amplitudes avec de faible taux de fausses alarmes. Dans le cas des défauts de faibles amplitudes, cette méthode est intéressante en considérant le nouvel opérateur d'agrégation et en prenant en compte l'indicateur de persistance de défauts.

La troisième méthode dite DMP permet de fournir des bonnes détections mais avec un taux de fausses alarmes, élevé.

Nous pouvons dire que selon nos besoins, nous pouvons choisir l'une ou l'autre des trois procédures pour la détection et la localisation des défauts.

Conclusion

Le travail que nous avons effectué dans ce mémoire concernait l'étude du problème de diagnostic de fonctionnement des capteurs du réseau d'assainissement de Nantes. Cette étude a été l'objet d'une collaboration entre le Centre de Recherche en Automatique de Nancy, le Laboratoire Central des Ponts et Chaussées et le GEMCEA dans le cadre d'études et de recherches sur la validation automatisée des réseaux d'assainissement.

Les méthodes de détection et de localisation se basent sur le principe de vérification de la cohérence entre les résultats de mesure effectués pendant le fonctionnement réel du système et un ou plusieurs modèles de référence. La modélisation est donc indispensable quelque soit les techniques de détection et de localisation adoptée.

Dans le cas de notre application, nous nous sommes intéressés aux techniques de détection et de localisation à base de modèles analytiques. Plusieurs travaux ont été réalisés dans ce domaine mais la difficulté reste de pouvoir les appliquer sur les systèmes réels du fait de la sensibilité des méthodes aux incertitudes de mesure et de modélisation (structure et paramètres des modèles approximatifs) et la possibilité des défaillances multiples.

Notre travail a été orienté vers le développement d'une méthode de détection et de localisation hors ligne des défauts de mesures issues des réseaux d'assainissement. La méthode que nous avons choisie est basée sur l'utilisation de la redondance analytique.

Les résultats de mesures, que nous avons eu à traiter sont issus du réseau d'assainissement de Nantes. Ils représentent des données de différents types, mesurées ou évaluées de manières diverses et sont entachées des erreurs de mesures et perturbées de différentes manières. Avant d'appliquer la méthode de détection et de localisation que nous avons adoptée. Il nous a fallu passer par les étapes suivantes :

1. La prise de connaissance du réseau de mesure et des données de Nantes, y compris l'outil de gestion (MINAUTOR), dans lequel les données sont enregistrées, classées selon le type de capteur et sont horodatées, nous a permis de définir tous les points de mesure (nom du point

de mesure, la mesure en question, la méthode de mesure (type de capteur)), les relations physiques entre les points de mesure.

2. L'analyse visuelle de l'ensemble des données (visualisation graphique) nous a permis de déterminer des valeurs aberrantes et de les classer selon les deux types suivants :

- Des anomalies de fonctionnements qui sont généralement liées aux défaillances des capteurs de mesure installés dans le réseau : celles qui se maintiennent une fois apparues et celles qui se manifestent fugitivement. Elles sont le plus souvent dues à des défaillances de capteurs. Les causes de ces défaillances sont notamment le colmatage des capteurs (présence de déchets sur les capteurs qui peuvent les isoler de l'environnement de mesure) et de dérive de capteurs. Ceci nous a conduit à nous nous concentrer sur la détection et la localisation des défauts de mesure produits par ce type de défaillance plus particulièrement.

- Des anomalies sur le processus : les mesures effectuées sont justes dans le cas de non-présence de défaillances de capteurs mais elles ne sont pas représentatives du fonctionnement normal du système.

3. l'application d'une procédure de prétraitement que nous avons réalisée. Ce prétraitement basé sur une phase de rééchantillonnage, nous a permis la mise en forme des données brutes pour qu'elles soient exploitables ultérieurement et d'une phase de filtrage : filtre Passe-bas et filtre Ebarbeur. Ces filtres nous ont permis d'éliminer les bruits de mesure et les données aberrantes respectivement.

Une fois passé par les étapes citées ci-dessus, et donc après avoir effectué des traitements sur les données nous avons pu obtenir un ensemble de mesures validées ou « crédibles » que nous avons pu utiliser dans une étape de modélisation. Celle-ci a pour objectif de fournir des relations de redondance analytique permettant la détection et la localisation des défaillances des capteurs.

Nous avons constaté depuis les tables de corrélations établies entre les différentes variables et présentées dans le chapitre 4, des corrélations significatives (supérieur à 0.8) entre quelques variables. Ainsi des modèles de régression linéaire ont été pris en considération. Les structures des modèles ont été établies en se basant sur une analyse de ces corrélations. Par la suite, l'identification des paramètres des modèles a été réalisée par une technique de régression linéaire simple. Nous avons appliqué après avoir obtenu ces modèles, une

technique de validation croisée des modèles pour mesurer la qualité de ces modèles ainsi que leur validité.

Ces modèles nous ont permis de générer des résidus : signaux indicateurs de défauts, de définir les signatures des défauts expérimentales. Ces signatures sont des représentations simplifiées des symptômes de défauts.

Nous avons considéré par la suite une phase de détection et de localisation des défauts de mesure. Pour l'étape de détection, nous avons appliqué une procédure basée sur le test de Page sur l'ensemble des résidus obtenus, en réinitialisant le test après chaque détection. La définition des paramètres de cette procédure a été effectuée par apprentissage et en utilisant la tolérance associée à chaque résidu. En ce qui concerne l'étape de localisation, trois méthodes ont été considérées. Elles reposent toutes sur le même principe général, basé sur l'utilisation des signatures des défauts théoriques pour désigner la situation de défaut la plus vraisemblable. La première méthode de localisation de défaut consiste à calculer la distance Euclidienne entre les signatures des défauts théoriques et chaque signature de défaut expérimentale. Cette méthode est simple dans son application car elle n'a pas de paramètres à fixer ou à définir mais elle n'est pas très efficace. D'une part, elle ne fournit aucune indication sur le signe du défaut localisé et d'une autre part, elle est très sensible aux incertitudes de mesures et à celles des modèles. La deuxième méthode qui a été considérée est appelée DMP, fournit le signe de défaut localisé, mais ses performances sont toutefois encore limitées. Une troisième méthode a été considérée : elle est basée sur les concepts de la logique floue. Le raisonnement flou est par nature approprié à l'exploitation de données et de modèles incertains, et il est également adapté à la prise en compte des informations de natures diverses. Il nous a permis d'améliorer les performances de la localisation. Moyennant certaines adaptations comme la définition d'un nouvel opérateur d'agrégation pour le calcul des facteurs de vraisemblances de défauts ou encore la prise en considération d'un indicateur de persistance de défauts. Ces améliorations diminuent fortement la sensibilité de la méthode aux incertitudes. Les performances de ces méthodes ont été vérifiées par calcul des probabilités de détection et des taux de fausses alarmes et nous ont permis, d'établir une comparaison entre elles. L'exploitation des données réellement perturbées nous a permis de valider ces méthodes et les améliorations apportées.

Parmi les perspectives concernant cette application :

- L'exploitation des modèles établis à des pas de temps d'une heure et de la journée.
- L'application de ces méthodes de diagnostic sur d'autres banques de données pour vérifier leur efficacité et leur performance.
- La construction de modèles non linéaires pouvant expliquer les variables débits en fonction des hauteurs et des vitesses.

Références

- [AFNOR 88] AFNOR (1988), *Maintenance industrielle*, Recueil de normes française AFNOR, AFNOR-UTE.
- [AFNOR 94] AFNOR (1994), *Maintenance-concepts et définitions des activités de maintenance*, Norme NF X 60-010, Association française de normalisation, p. 28.
- [Alexandre 98] Alexandre O. et Grand d'Esnon A. (1998), “Le coût des services d'assainissement ruraux. Evaluation des coûts d'investissement et d'exploitation.” *Techniques Sciences Méthodes*, vol. 93, n°. 7, pp 19-31.
- [Arrêté 94] Arrêté interministériel du 22 décembre 1994, “Autosurveillance des systèmes d'assainissement de plus de 2000 EH”, mentionnées aux articles L. 372-1-1 et L.372-3, *Journal Officiel du 10 février 1995*.
- [Atanassov 99] Atanassov, K.T. (1999), *Intuitionistic fuzzy sets*, Phisica-Verlag, Heidelberg.
- [Aumond 01] Aumond M. et Joannis C. (2001), “Construction de profils journaliers types pour la validation de résultats de mesures de débits en réseaux d'assainissement”, *2^{ème} Colloque Automatique et Environnement*, St Etienne.
- [Aumond 03] Aumond M., Rufflé S. et Joannis C. (2003), “Validation de résultats de mesure pour l'autosurveillance des réseaux d'assainissement, méthodologie et exemples”, *Rapport Agence de l'Eau Loire*, Bretagne.
- [Basseville 86] Basseville M. (1986), “On-line detection of jumps in mean”, in Basseville M., Benveniste A., *Detection of abrupt changes in signals and dynamical systems*, Lecture notes in Control and Information Sciences, *Springer-Verlag*, vol. 77, pp. 12-26.
- [Basseville 88] Basseville M. (1988), “Detecting changes in signals and systems-a survey”, *Automatica*, vol. 24, n°. 3, pp. 309-326.
- [Basseville 93] Basseville M. et Nikifirov I.V. (1993), *Detection of abrupt changes : Theory and applications*, Prentice Hall., Information and System Sciences Serie, Englewood Cliffs, NJ.

- [Basseville 96] Basseville M. et Cordier M.O. (1996), “Surveillance et diagnostic de systèmes dynamiques - Approches complémentaires du traitement de signal et de l’intelligence artificielle”, Publication interne, n°. 1004, Institut de Recherche en Informatique et Systèmes Aléatoires, Rennes, p. 40.
- [Barré 71] Barré de Saint Venant (1871), “Théorie du mouvement non permanent des eaux”, *Comptes Rendus de l'Académie des Sciences*, Paris, vol. 73, pp. 147-237.
- [Ben Abdennour 01] Ben Abdennour R., Borne P. et Khsouri M. (2001), *Identification et commande numérique des procédés industriels*, Méthodes et pratiques de l’ingénieur, Automatique, Technip, Paris.
- [Bennis 95] Bennis S., Berrada F., et Kang N. (1995) “Improving single variable and multivariable techniques for estimating missing hydrological data”, *Journal of Hydrology*, vol. 191, n° 1-4, pp. 87-105, Elsevier Science.
- [Bennis 96] Bennis S., Côté S., et Kang N. (1996), “Validation des données hydrométriques par des techniques multivariées de filtrage”, *Journal Canadien de génie civil*, vol. 23, n° 1, pp. 218-230.
- [Bennis 97] Bennis S., Berrada, F., et Kang, N. (1997), “Improving single variable and Multivariable techniques for estimating Missing Hydrological Data”, *Journal of Hydrology*, vol. 191, pp. 87-105.
- [Bennis 00] Bennis S. et Kang N. (2000), “Multivariate technique for validating historical hydrometric data with redundant measurement”, *Nordic Hydrology*. vol. 31, n° 2, pp. 107-126.
- [Berrada 96] Berrada F., Bennis S. et Gagnon L. (1996), “Validation des données hydrométriques par des techniques univariées de filtrage”, *Journal Canadien de Génie Civil*, vol. 23, pp. 872-892.
- [Bertrand-Krajewski 00] Bertrand-Krajewski J.L, Laplace D., Joannais C. et Chebbo G. (2000), *Mesure en hydrologie urbaine et assainissement*, Lavoisier, Paris.
- [Boukhris 98] Boukhris A. (1998), “Identification de systèmes non-linéaires par une approche multi-modèle – application à la modélisation de la relation pluie-débit pour le diagnostic de fonctionnement de capteurs”. Thèse de doctorat, INPL, Nancy.
- [Chatain 93] Chatain J.N. (1993), *Diagnostic par système expert*, Traité des nouvelles technologies, série diagnostic et maintenance, Hermès, Paris.

-
- [Chen 91] Chen S., Cowan C.F.N. et Grant P.M. (1991), "Orthogonal least squares learning algorithm for radial basis functions networks", *IEEE Trans. On Neural Networks*, vol. 2, n°. 2, pp. 302-309.
- [Chen 99] Chen J. et Patton R.J. (1999), *Robust model-based fault diagnosis for dynamic systems*, Kluwer Academic Publishers, Massachusetts.
- [Chen 00] Chen G., Wie Q. et Kerre E. (2000), "Fuzzy data mining : discovery of fuzzy generalized association rules", *G. Bordogna and G. Pasi, Recent Research Issues on Management of Fuzziness in Databases*, Springer Physica-Verlag, pp. 45-66.
- [Chow 73] Chow V.T. (1973), *Open Channel Hydraulics*, Mc GrawHill, New York.
- [Circulaire 95] Circulaire du 12 mai 1995, "Systèmes d'assainissement de plus de 2000 EH", *Ministère de l'aménagement et de l'environnement - Recueil de textes sur l'assainissement - version 4-juin 2001*.
- [Combastel 02] Combastel C., Gentil S. et Rognon J.P. (2002), "Génération de résidus adaptative pour le diagnostic de la machine asynchrone", *Conférence Internationale Francophone d'Automatique, CIFA '02*, Nantes.
- [De Kleer 87] De Kleer J. et Williams B.C., (1987), "Diagnosing multiple faults", *Artificial Intelligence*, n°. 32, pp. 97-130.
- [Décret 94] Décret du 3 juin 1994 relatif à l'assainissement des eaux usées urbaines mentionnées articles L.372-1 et L.372-3 du code des communes, *Journal Officiel du 8 juin 1994*.
- [Directive 91] Directive européenne du 21 mai 1991, "Eaux résiduaires urbaines".
- [Directive 00] Directive européenne du 23 octobre 2000, "Directive cadre sur l'eau".
- [Dorr 95] Dorr R. (1995), "Détection de défauts de capteurs par redondances matérielle et analytique : Application aux centrales nucléaires", Thèse de Doctorat, INPL, Nancy.
- [Düştægör 06] Düştægör D., Frisk E., Cocquempot V., Krysander M. et Staroswiecki M. (2006), "Structural analysis of fault isolability in the DAMADICS benchmark", *Control Engineering Practice*, vol. 14, n°. 6, pp. 597-608.
- [Fantana 95] Fantana S. et Mizier M.-O. (1995), "La débitimétrie : délicate à mettre en œuvre", *Hydroplus*, n°.53, pp.36-46.

- [Frank 90] Frank P.M. (1990), "Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy-A survey and some new results", *Automatica*, vol. 26, pp. 459-474.
- [Frank 91] Frank P.M. (1991), "Enhancement in robustness in observer-based fault detection", *International Journal of Control*, vol. 59, pp. 955-981.
- [Frank 00] Frank P.M., Ding S.X. et Marcu T. (2000), "Model-based fault diagnosis in technical processes". *Transactions of the Institution of Measurement and Control*, vol. 22, n°. 1, pp. 57-101.
- [Gertler 85] Gertler J. et Singer D. (1985), "Augmented models for statistical fault isolation in complex dynamic systems", *American Control Conference, ACC'85*, Boston, MA.
- [Gertler 88] Gertler J. (1988), "survey of model-based failure detection and isolation in complex plants. *IEEE Control Systems Magazine*", vol. 8, n°. 6, pp. 663-670.
- [Gertler 92] Gertler J. et Anderson K.C. (1992), "An evidential reasoning extension to quantitative model-based failure diagnosis", *IEEE Transactions on Systems Man and Cybernetics*, vol. 22, n°. 2, pp. 275-285.
- [Giraud 98] Giraud D. (1998), "Diagnostic des systèmes complexes par agrégation des méthodes : Application à une station d'épuration", Thèse de Doctorat, Université Henri Poincaré, Nancy.
- [Giuliani 98] Giuliani S. (1998), "Détection et localisation des défaillances de capteurs dans un réseau de mesures pluviométriques - cas de l'agglomération nancéenne". Thèse de doctorat, INPL, Nancy.
- [Hamioud 03] Hamioud F., Joannais C. et Ragot J. (2003) "Statistical modelling for validation of hydrometric data issued from the sewer networks", *19th European Junior Scientist Workshop on Process Data and Integrated Urban Water Modelling, EJSW'03*, Meaux-la-Montagne.
- [Hamioud 05a] Hamioud F., Ragot J. et Joannis C. (2005), "Fault diagnosis based on fuzzy concepts : Application to data of monitoring system on sewer network" *3rd IEEE International Conference on Systems, Signals & Devices, SSD'05*, Sousse.
- [Hamioud 05b] Hamioud F., Ragot J. et Joannis C. (2005), "Fault diagnosis for validation of hydrometric data collected from sewer networks" *10th International Conference on Urban Drainage, ICUD'10*, Copenhagen.

- [Himmelblau 78] Himmelblau, D.M. (1978), "Fault detection and diagnosis in chemical and petrochemical processes", *Elsevier*, pp. 343-393.
- [Hinkley 71] Hinkley D.V. (1971), "Inference about the change-point from cumulative sum-tests", *Biometrika*, vol. 58, n°. 3, pp. 509-523.
- [IFEN 06] IFEN (2006), "L'assainissement en France en 1998 et 2001 ", *Les dossiers de l'IFEN*, n°. 3, p. 29.
- [Isermann 97] Isermann R. et Ballé P. (1997), "Trends in the application of model-based fault detection and diagnosis of technical processes", *Control Engineering Practise*, vol. 5, n°. 5, pp. 709-719.
- [Jay 88] Jay F. (1988), *IEEE Standard dictionary of Electrical and Electronics terms*, 4^{ème} éd., The institute of Electrical and Electronics engineers, NY.
- [Johnson 02] Johnson R.A. et Wichern D.W. (2002), *Applied Multivariate Statistical Analysis*, Prentice Hall, 5^{ème} éd., Upper Saddle River.
- [Kleinbaum 98] Kleinbaum D.G, Lawrence L.L., Kupper, Muller K.E. et Nizam A. (1998), *Applied regression analysis and other multivariate methods*, 3^{ème} éd., Duxbury press, Paris.
- [Koenigsberg 79] Koenigsberg W. (1975), *Recursive least squares parameter estimation and associated computer programs*, Charles Stark Draper Laboratory, Cambridge Massachusset.
- [Kramer 87] Kramer M.A. (1987), "Malfunction diagnosis using quantitative models with non Boolean reasoning in expert systems", *AIChE Journal*, vol. 33, n°. 1, pp. 130-140.
- [Kramer 93] Kramer M.A. et Mah, R.S.H. (1993), "Model-based monitoring", Rippin D. Hale J. et Davis J. éd., *Proceedings of the 2nd International Conference on Foundation of Computer-Aided Process Operations, FOCAPO*, pp.45-68.
- [Landau 01] Landau I.D. et Besançon-Voda A. (2001), *Identification des systèmes*, traité information, commande et communication. Hermès Science Publications. Paris.
- [Le Gauffre 06] Le gauffre P., Joannis C., Vasconcelos E., Gibello C., Breyse D. et Desmulliez J.J. (2006), "performance indicators and multi-criteria decision support for sewer asset management", *ASCE Journal of Infrastructure Systems*, accepté
- [MacKay 92] MacKay D.J.C. (1992), "Bayesian interpolation", *Neural Computation*, vol. 4, n°. 3, pp. 415-447.

- [Mandel 98] Mandel D., Maquin D. et Ragot J. (1998), "Process supervision based on fuzzy reasoning: Application to a Urban Supply Network", *3rd IFAC Workshop on Online Fault Detection and Supervision in the Chemical Process Industries*, Solaize.
- [Maquin 91] Maquin D. et Ragot J. (1991), "Comparison of gross errors detection methods in process data", *Proceedings of the 30th IEEE Conference on Decision and Control*, Brighton, vol. 3, pp. 2253-2261.
- [Maquin 00] Maquin D. et Ragot J. (2000), *Diagnostic des systèmes linéaires*, Hermès, Paris.
- [Mourad 01] Mourad M. et Bertrand-Krajewski J.L. (2001), "Procédure automatisée d'aide à la validation de données environnementales en hydrologie urbaine", *2^{ème} Colloque Automatique et Environnement*, Saint-Etienne.
- [Mourad 05] Mourad M., Bertrand-Krajewski J.-L., Hamioud F. et Joannis C. (2005), "Validation de données en hydrologie urbaine : principes généraux et exemples d'application", *Conférence sur l'Autosurveillance, Diagnostic Permanent et Modélisation des Flux Polluants en Réseaux d'Assainissement Urbains*, Marne La Vallée.
- [Patton 97] Patton R.J. (1997), "Fault-tolerant control: The 1997 situation". In *IFAC Fault Detection, Supervision and Safety for Technical Processes*, Kingston Upon Hull, UK, pp. 1029-1051.
- [Page 54] Page E.S. (1954), "Continuous inspection schemes", *Biometrika*, vol. 41, pp. 100-115.
- [Pau 81] Pau L.F. (1981), "Failure diagnostic and performance monitoring", *Control and System theory*, Marcel Decker, vol. 11, pp. 5-7.
- [Petti 90] Petti T.F., Klein J. et Djurjati P.S. (1990), "Diagnostic model processor: Using deep knowledge for process fault diagnosis", *AIChE Journal*, vol. 36, n°. 4, pp. 565-575.
- [Petti 92] Petti T.F. (1992), *Using mathematical models in knowledge-based control systems*. Thèse de Doctorat, University of Delaware.
- [Ragot 90] Ragot J., Darouach M., Maquin D. et Bloch G. (1990), *Validation de données et diagnostic*, traité des nouvelles technologies, série diagnostic et maintenance, Hermès, Paris.

- [Rengaswamy 01] Ranganaswamy R., Mylaraswamy D., Arzén K.E. et Venkatasubramanian V. (2001), "A comparison of model-based and neural network-based diagnosis methods", *Engineering Applications of Artificial Intelligence*, vol. 14, pp. 805-818.
- [Russel 99] Russel E.L., Chiang L.H. et Braatz R.D. (1999), "Data-driven techniques for fault detection and diagnosis in chemical processes. *Springer*, London.
- [Simani 02] Simani S. et Patton R.J. (2002), "Model-based data-driven approaches to robust fault diagnosis in chemical processes", *15th IFAC World Congress on Automatic Control, IFAC'02*, Barcelona, vol. 3, pp. 2551- 2556.
- [Surgenor 92] Surgenor B.W. et Jofriet P.J. (1992), "Thermal fault analysis and the diagnostic model processor", *Proceedings of the IFAC Symposium, On-Line Fault Detection and Supervision in the Chemical Process Industries*, Delaware, pp. 20-25.
- [Tomassone 92] Tomassone R., Audrain S., Lesquoy-de Turckheim E. et Millier C., (1992), *La régression nouveaux regards sur une ancienne méthode statistique*, 2^{ème} éd. Masson. Paris.
- [Walter 94] Walter E. et Pronzato L. (1994), *Identification de modèles paramétriques*. Modélisation, analyse, simulation, commande, Masson, Paris.
- [Walter 97] Walter E., Pronzato L. et Norton J. (1997), *Identification of parametric models from experimental data*, Communications and control engineering series, Masson, Paris.
- [Ulieru 94] Ulieru M. (1994), "A fuzzy logic based computer assisted fault diagnosis system", *European Journal of Diagnosis and Safety in Automation, Hermès*, vol. 4, n° 3, pp. 405-441.
- [Van Schrick 97] Van Schrick D. (1997), "Remarks on terminologie in the fields of supervision", *Proceedings of the IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes*, Hull, UK., 24-26 August, pp. 1291-1296.
- [Zwingelstein 95] Zwingelstein, G. (1995), *Diagnostic des défaillances, théorie et pratique pour les systèmes industriels*, Traité des nouvelles technologies, série diagnostic et maintenance, Hermès, Paris.

Annexe A Points de mesure

Dans cette annexe, nous présentons dans un tableau, les différents points de mesure, les types de mesure préconisés sur chaque point ainsi que le type d'équipement utilisé sur tout le système d'assainissement.

Point de mesure	Type de mesure	Equipement
Quai fosse Déversoir 27 Point de transfert 16	hauteurs vitesse	capteur hauteur piézo rejet capteur hauteur Ultrasonique à l'exutoire capteur hauteur vanne capteur hauteur ultrasonique transfert capteur vitesse Doppler
Saint Mihiel Déversoir mobile 25 Déversoir fixe 112 Déversoir mobile 31	hauteurs	capteur hauteur piézo capteur hauteur Ultrasonique capteur hauteur vanne capteur hauteur Ultrasonique transfert capteur hauteur piézo rejet capteur hauteur piézo rejet capteur hauteur vanne capteur hauteur Ultrasonique transfet
Broutelle Guai Robert Déversoir mobile 30 Pluviomètre	hauteurs vitesses pluie	capteur hauteur piézo capteur vanne capteur hauteur capteur vitesse temps transit (corde1) capteur vitesse temps transit (corde2) pluviomètre.

Tableau A. Différents points de mesure et équipements correspondants

Point de mesure	Type de mesure	Equipement
Cordon Bleu Déversoir 26	hauteurs	hauteur piézo capteur hauteur Ultrasonique à l'exutoire capteur vanne capteur hauteur piézo
Bouhier Point de transfert 17	hauteur vitesse	capteur hauteur piézo capteur vitesse Doppler
Route de Ste Luce Point de transfert 8	hauteur vitesse	capteur hauteur piézo capteur vitesse Doppler
Bd du Massacre Point de transfert 23	hauteur vitesse	capteur hauteur piézo capteur vitesse Doppler
Tougas Déversoir d'orage Station d'épuration pluviomètre	hauteur vitesses pluie	capteur hauteur piézo 2 capteurs vitesse Doppler pluviomètre
Chemin du Château Point de transfert 3	hauteur vitesse	capteur hauteur piézo capteur vitesse Doppler
Petite Californie Station d'épuration Pluviomètre Points entrées-sorties	hauteur vitesse pluie	capteur hauteur piézo capteur vitesse Doppler pluviomètre

Tableau A. (suite) Différents points de mesure et équipements correspondants

Point de mesure	Type de mesure	Equipement
Duchesse Anne Point de transfert 15 Déversoir fixe 104 Déversoir fixe 105	hauteurs vitesses	3 capteurs hauteur piézo capteur vitesse Doppler capteur vitesse Doppler
Dervallières Déversoir fixe 126 pluviomètre	hauteur pluie	capteur hauteur piézo pluviomètre
Bouchaud Déversoir fixe 123	hauteur	capteur hauteur piézo
Ste Anne pluviomètre	pluie	Pluviomètre
Fillée Pluviomètre	pluie	Pluviomètre
Chevalerie Pluviomètre	pluie	Pluviomètre
Canaris Pluviomètre	pluie	pluviomètre

Tableau A. (suite) Différents points de mesure et équipements correspondants

Annexe B Prétraitement des données

Cette annexe est consacrée à la description de la phase préliminaire de prétraitement des données brutes issues du réseau de mesure de Nantes.

B.1 Problématique

Les signaux issus du réseau de mesure sont acquis à des pas de temps variables. La problématique consiste à préparer un ensemble de signaux (sains) apte à l'analyse. Il est utilisé pour établir des modèles. Cette préparation des signaux est représentée par un prétraitement.

Un prétraitement des signaux consiste en une première étape de rééchantillonnage (interpolation linéaire), qui permet d'obtenir une période de rééchantillonnage fixe et identique pour tous les signaux considérés. Dans une deuxième étape, il consiste à identifier pour chaque signal, toutes les anomalies possibles, en se basant sur l'ensemble des anomalies défini dans la section 1.5 du premier chapitre. Enfin, Nous appliquons des techniques de filtrage permettant d'éliminer d'une part, ces anomalies et d'une autre part, les bruits de mesure dans le cas de nécessité.

B.2 Prétraitement des signaux

B.2.1 Rééchantillonnage

Le rééchantillonnage consiste à transformer un signal déjà échantillonné en un signal échantillonné à des instants différents du premier. Il permet plus particulièrement de passer d'un signal à pas de temps variable à un signal à pas de temps fixe.

Une période de rééchantillonnage élevée conduit à une perte d'information, par contre une période de rééchantillonnage faible augmente le coût des calculs (temps de calcul) sans augmenter la précision. Cette dernière est limitée par l'échantillonnage initial : les informations obtenues à partir du système sont limitées en fréquence par le critère de Shannon et ont été définies lors de la conception du système de mesure en fonction des objectifs métrologiques fixés par les concepteurs.

Dans le cas pratique, le choix de la période d'échantillonnage doit obéir à la contrainte [Ben Abdennour et al., 01] suivante :

$$T_e > T_a,$$

avec T_a le temps d'acquisition.

Dans le cas de notre application, cette condition est respectée car le temps d'acquisition est de l'ordre de quelques secondes.

Les signaux débits, hauteurs et vitesses considérés dans ce travail ont été interpolés par une procédure d'interpolation linéaire. En ce qui concerne le signal pluie, où les valeurs enregistrées représentent des cumuls sur une période donnée, nous utilisons une interpolation constante-avant.

Interpolation linéaire :

Elle est effectuée pour des signaux échantillonnés et enregistrés d'une manière instantanée. Elle est donc bien adaptée à ces signaux.

Soit le signal $y(k)$ défini sur un intervalle de temps de largeur n , où k représente le temps, un intervalle de temps $[j \ l]$ de largeur inférieure à n ,

i un nombre entier positive et $j < i < l$.

Quand nous ne disposons pas de $y(i)$, l'interpolation permet de calculer cette valeur comme suit :

$$y(i) = y(j) + (i - j) \frac{(y(l) - y(j))}{l - j}, \quad (\text{B.1})$$

Interpolation constante-avant :

Elle est effectuée pour des signaux où les valeurs représentent des cumuls enregistrés pour un pas de temps donné.

Soit le signal pluie (en intensité) $y(k)$, défini par un ensemble de n valeurs.

Pour une interpolation à la minute, les valeurs du signal sur la période d'enregistrement $[j \ l]$ où nous ne disposons que de la valeur $y(l)$, sont données à chaque instant i comme suit :

$$y(i) = y(l), \text{ avec } i = j, j + 1, j + 2, \dots, l, \quad (\text{B.2})$$

Cette interpolation constante-avant est illustrée par la figure B.1.

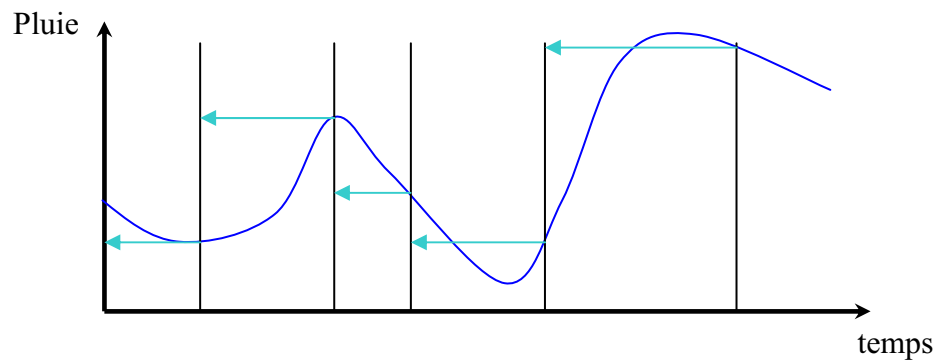


Figure B.1 Interpolation constante-avant

Une fois l'ensemble des signaux interpolés avec une période de rééchantillonnage d'une minute, est obtenu, nous procédons au filtrage.

Le rééchantillonnage est appliqué sur les données du réseau d'assainissement de Nantes donné par une description préalable. Les signaux bruts représentent des entrées avec des pas de temps variables et des durées d'enregistrement quasi-semblables sur une période d'un mois.

Pour illustration, nous nous contentons de présenter qu'un ensemble des graphes et des tableaux illustratifs de la méthode et des résultats obtenus. Pour lire aisément les graphes, il suffit de se référer au tableau 3.1 du chapitre 3 où chaque signal est indexé par un numéro.

Pour les signaux débits, vitesses et hauteurs, le rééchantillonnage est effectué avec un pas de temps minimal d'une minute. Les résultats obtenus sont représentés dans la figure B.2, où nous représentons le signal brut et le signal rééchantillonné. Ces signaux sont horodatés et définis sur une durée de trois heures et vingt minutes du 01 octobre 2001.

Dans les graphes de la figure B.2, nous remarquons, que le nombre d'échantillons sur la période définie avant l'application du rééchantillonnage, est différent d'un signal à un autre. Ce nombre devient le même après le rééchantillonnage.

B.2.2 Filtrage

Dans les graphes de la figure B.3, donnée ci-après, nous représentons d'une part les signaux bruts et d'une autre part, ceux obtenus après le filtrage (le filtrage consiste à appliquer les deux filtre exposés dans le chapitre 3 : le filtre "trim" puis le filtre passe-bas).

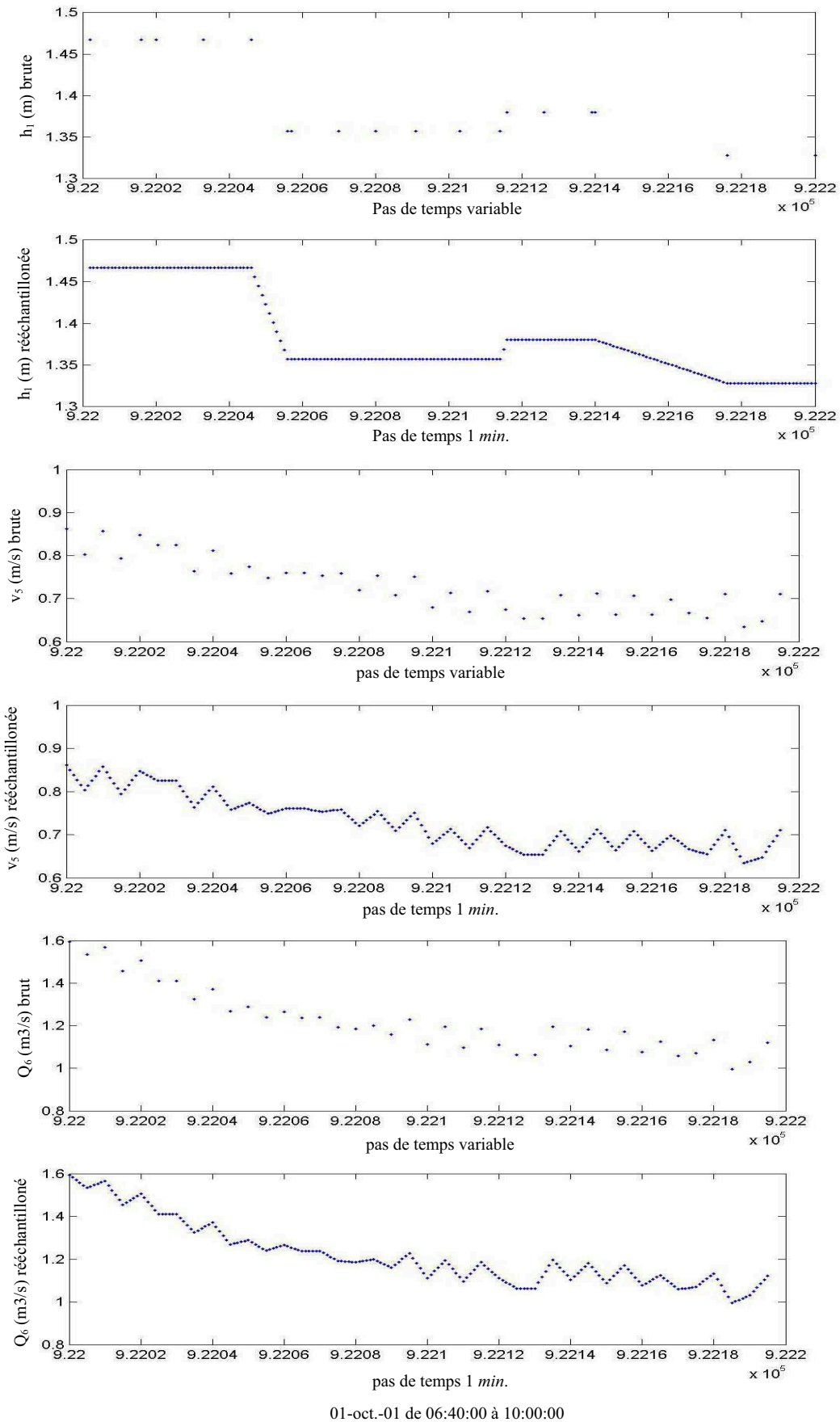


Figure.B.2 Signaux bruts et signaux rééchantillonnés, pas de temps d'une minute.

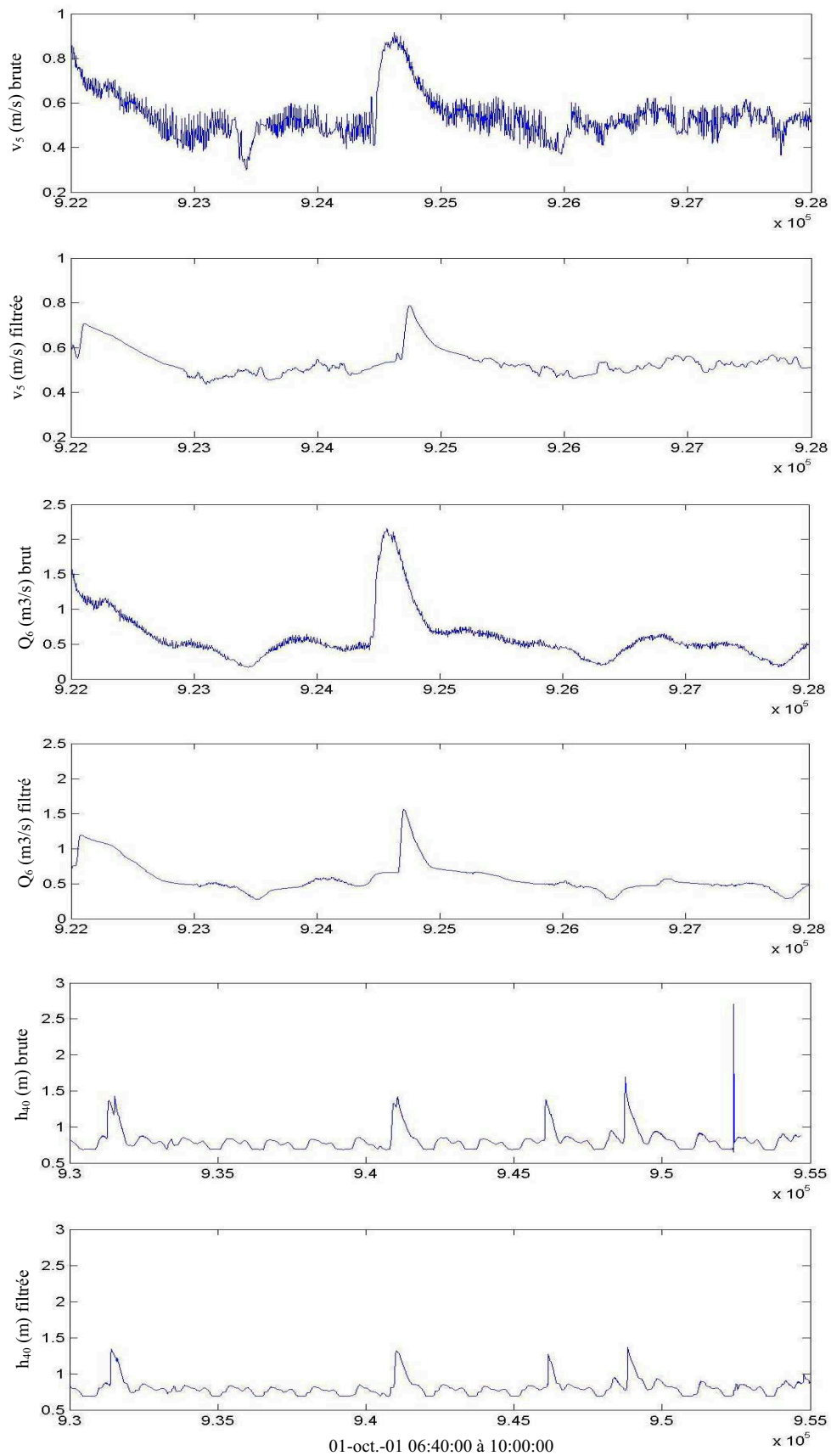


Figure B.3 Signaux filtrés avec un filtre "trim" et puis un filtre "passe-bas".

Nous remarquons d'après les deux premiers graphes de la figure B.3, que les signaux bruts sont très bruités et que le filtrage passe-bas est indispensable. Les valeurs aberrantes sont éliminées par l'application du filtre "trim" comme illustré par le troisième graphe de la figure B.3.

Annexe C Résultats de l'application de l'OLS

étape	Débit sélectionné	<i>AIC</i>	taux d'explication
1	Q_{30}	-7.80	0.92
2	Q_{21}	-8.20	0.95
3	C	-8.45	0.96
4	Q_6	-8.45	0.96
5	Q_{36}	-8.45	0.96

Tableau C.1. Résultats de l'application de la technique OLS pour le débit Q_{24}

étape	Débit sélectionné	<i>AIC</i>	taux d'explication
1	C	-9.27	0.97
2	Q_{24}	-10.31	0.99
4	Q_6	-10.31	0.99
5	Q_{21}	-10.31	0.99
6	Q_{36}	-10.51	0.99

Tableau C.2. Résultats de l'application de la technique OLS pour le débit Q_{30}

étape	Débit sélectionné	<i>AIC</i>	taux d'explication
1	Q_6	-4.91	0.94
2	Q_{21}	-5.38	0.96
3	Q_{30}	-5.67	0.97
4	C	-5.74	0.97
5	Q_{24}	-5.74	0.97

Tableau C.3. Résultats de l'application de la technique OLS pour le débit Q_{36}

Annexe D Validation croisée des modèles

Cette annexe est consacrée à exposer les résultats de l'application de la procédure de validation croisée sur l'ensemble des modèles établis.

D.1 Validation croisée des modèles

La procédure de validation mentionnée dans la section 3.6 du chapitre 3 a été appliquée pour l'ensemble des modèles établis. Nous avons considéré trois mois de données, que nous avons divisée en deux périodes ; la première période de 10500 mesures, une deuxième de 77336 mesures. Nous avons considéré les trois cas suivants :

Cas 1 : les données de la première période sont considérées pour identifier les modèles et sont notées, jeu-identification. Celles de la deuxième sont considérées pour valider les modèles et sont notées jeux-test ;

Cas 2 : les données de la deuxième période sont considérées comme jeu-identification et celles de la première sont considérée comme jeux-test.

Pour chaque variable à expliquer, nous avons donc identifier deux modèles (qui ne diffèrent que par les valeurs des paramètres). Les paramètres de ces deux modèles même avec peu de différence, doivent être unifiés pour obtenir un seul et unique modèle. Pour ce faire, nous avons appliqué la procédure des moindres carrées récursive [Königsberg, 79] pour vérifier la stabilité des paramètres de ce modèle. Ainsi pouvoir considéré un seul et unique modèle dont les paramètres sont stables.

Dans ce qui suit, nous représentons tout d'abord, tous les signaux utilisés. Ensuite, nous présentons chaque modèle établi. Puis, nous représentons l'estimation donnée par le modèle de chaque variable à expliquer en fonction de celle-ci pour le premier cas mentionné ci-dessus. Enfin, nous présentons les résultats de calcul des deux grandeurs r^2 et s_e pour les trois cas mentionnés ci-dessus, en les regroupant dans des tableaux.

a. Signaux utilisés pour la modélisation

Dans cette section, nous représentons chaque variable en fonction du temps sur une période de dix jours considérée du mois d'octobre 2001, avec un temps d'échantillonnage d'une minute.

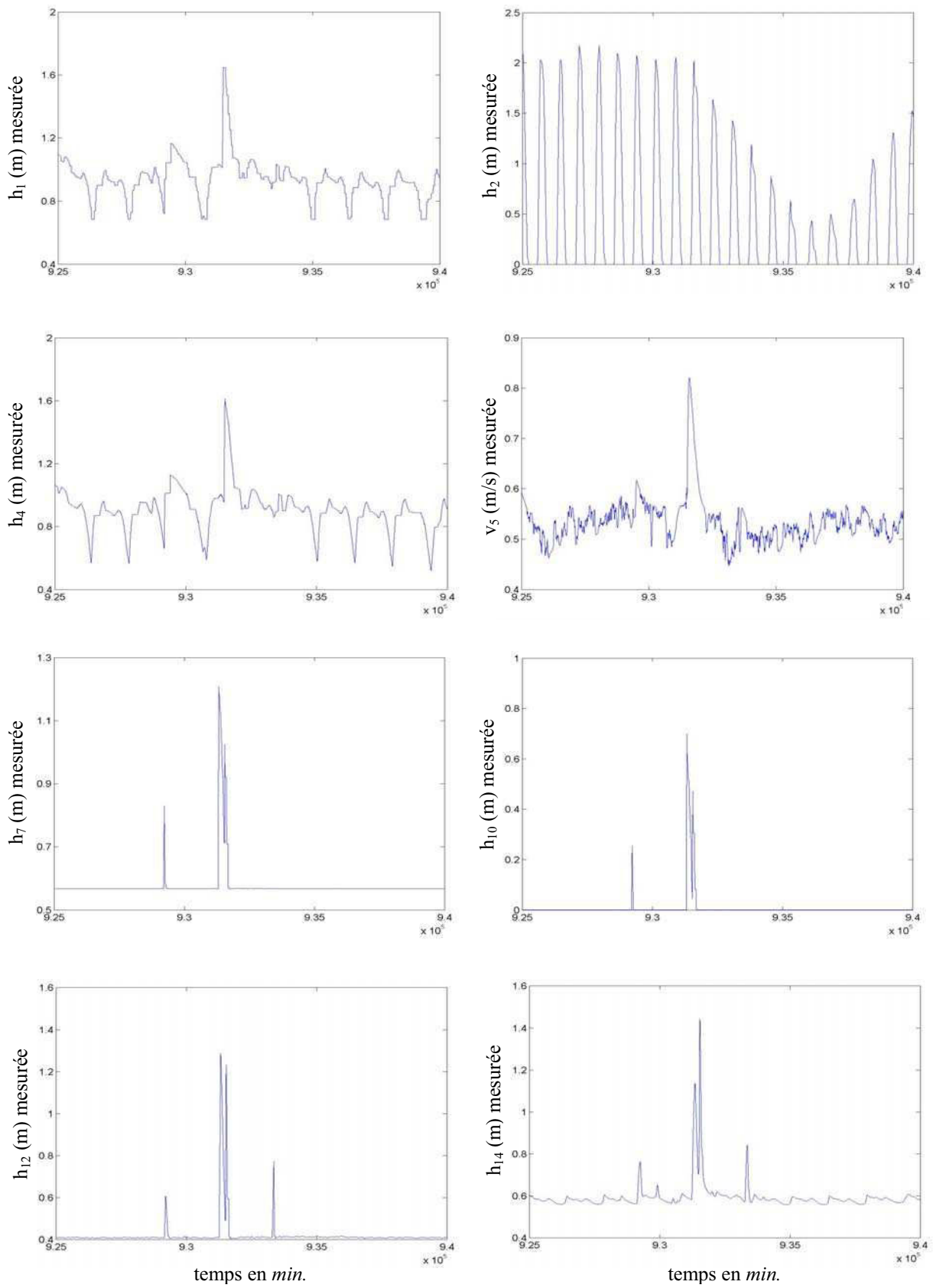


Figure D.1. Signaux $h_1, h_2, h_4, v_5, h_7, h_{10}, h_{11}, h_{12}$ en fonction du temps.

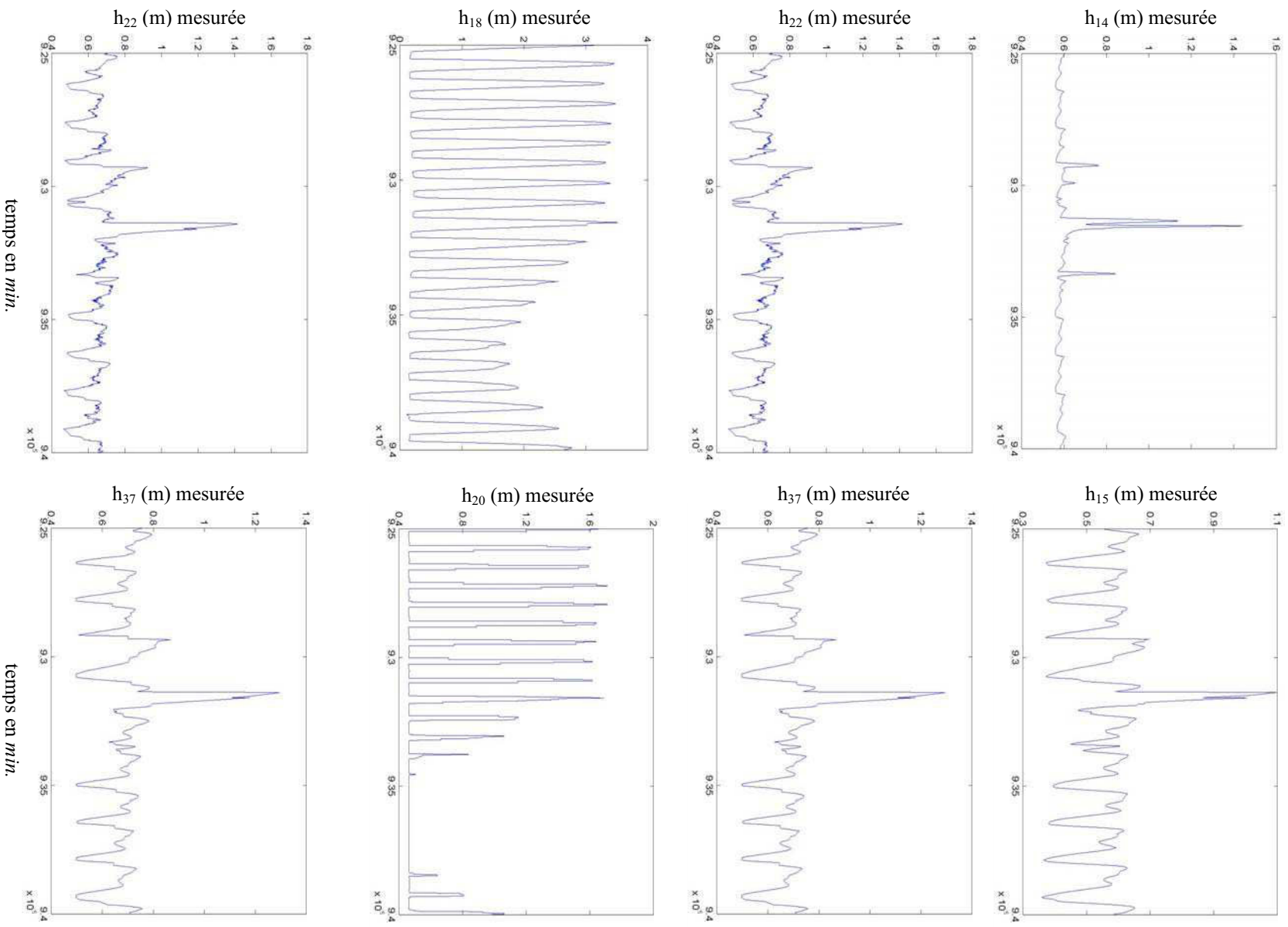


Figure D.2. Signaux h_{14} , h_{15} , h_{18} , h_{20} , h_{22} , h_{37} , h_{40} en fonction du temps.

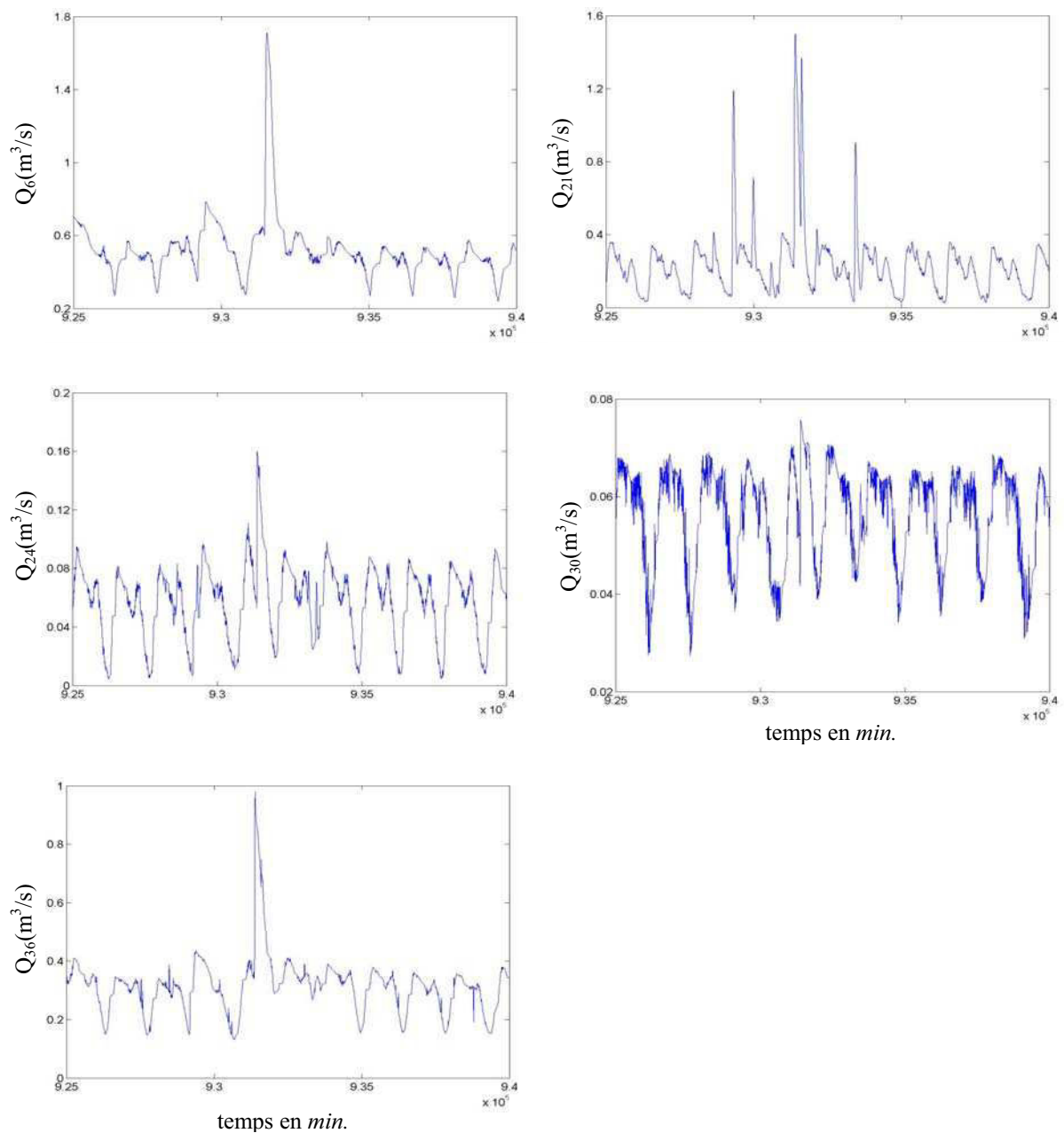


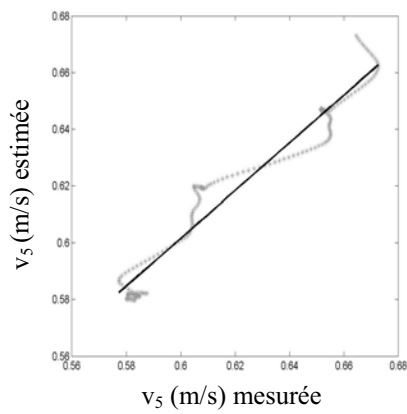
Figure D.3. Signaux débits.

- Modèles établis

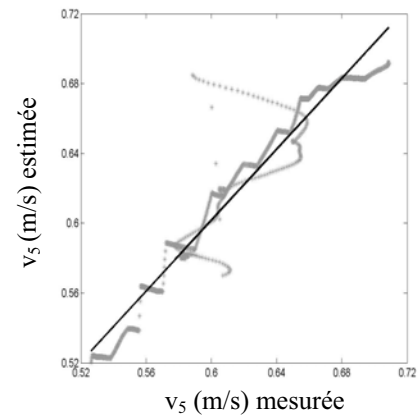
Dans cette section, nous donnons l'expression de chaque modèle établi. Nous présentons des courbes où nous traçons en superposition, l'estimation de chaque variable à expliquer en fonction de celle-ci et la droite de la régression linéaire correspondante au modèle pour les deux cas mentionnés ci-dessus et nous représentons les valeurs mesurées de la variable à expliquer $y(k)$ en superposition avec son estimée $\hat{y}(k)$, donnée par le modèle.

Modèle v_5

$$\hat{v}_5(k) = 0.29 + 0.50h_1(k) - 0.26h_4(k).$$



a. période 1 \equiv jeu-identification



b. période 2 \equiv jeu-test

Figure D.4. Identification du modèle v_5 sur période 1 et validation sur période 2.

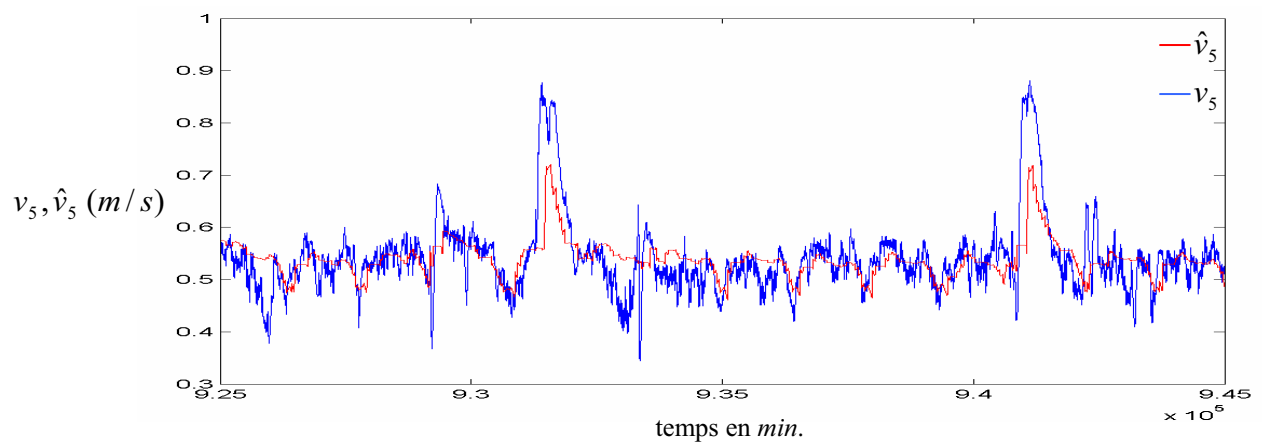
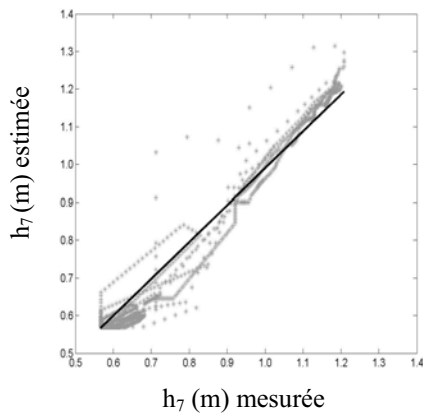


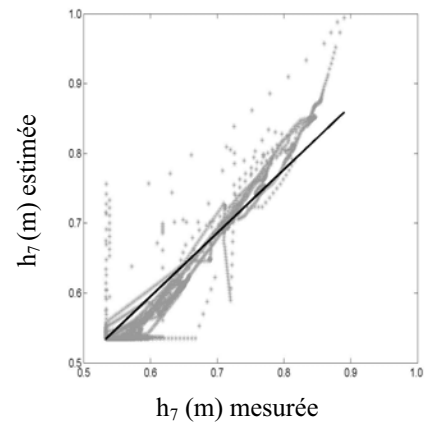
Figure D.5. La mesure v_5 et son estimée \hat{v}_5 en superposition, données en fonction du temps.

Modèle h_7

$$\hat{h}_7(k) = 0.57 + 1.06h_{10}(k).$$



a. jeu-identification : données période 1



a. jeu-test : données période 2

Figure D.6. Identification du modèle h_7 sur période 1 et validation sur période 2.

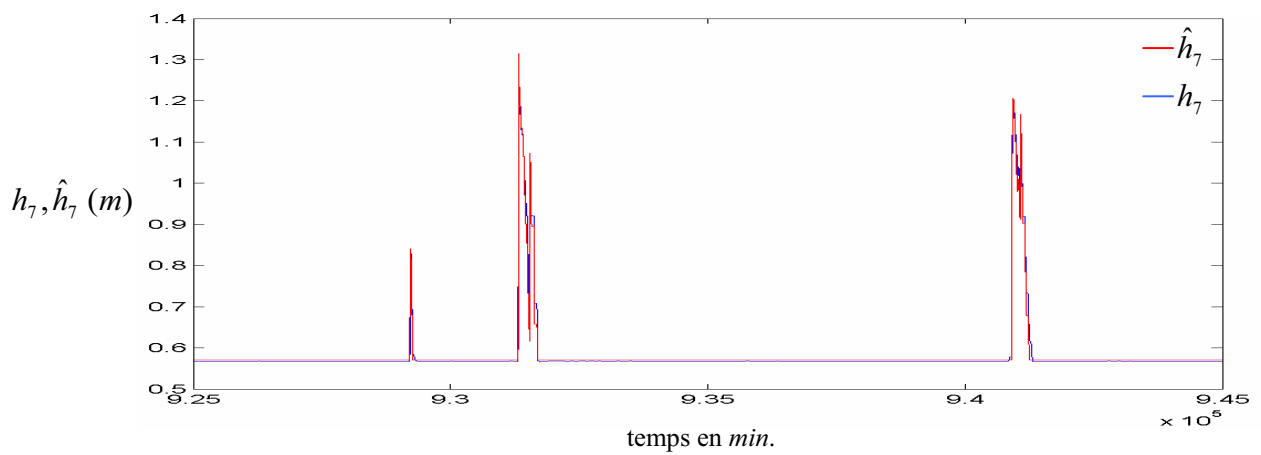
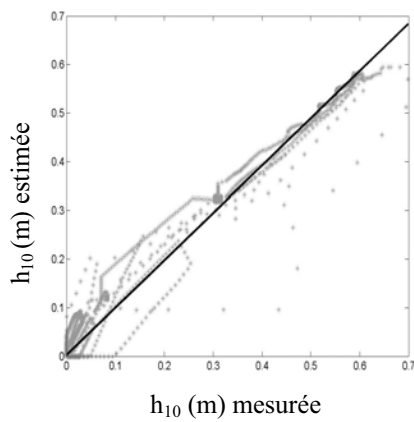


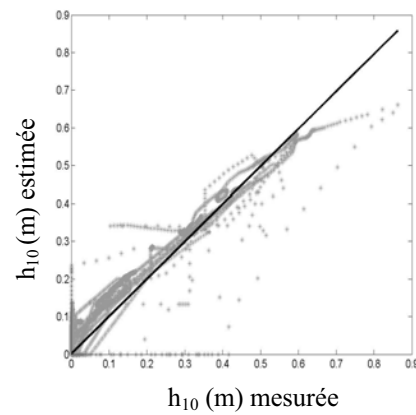
Figure D.7. La mesure h_7 et son estimée \hat{h}_7 en superposition, données en fonction du temps.

Modèle h_{10}

$$\hat{h}_{10}(k) = -0.55 + 0.98h_7(k) - 0.05h_{11}(k).$$



a. jeu-identification : données période 1



b. jeu-test : données période 2

Figure D.8. Identification du modèle h_{10} sur période 1 et validation sur période 2.

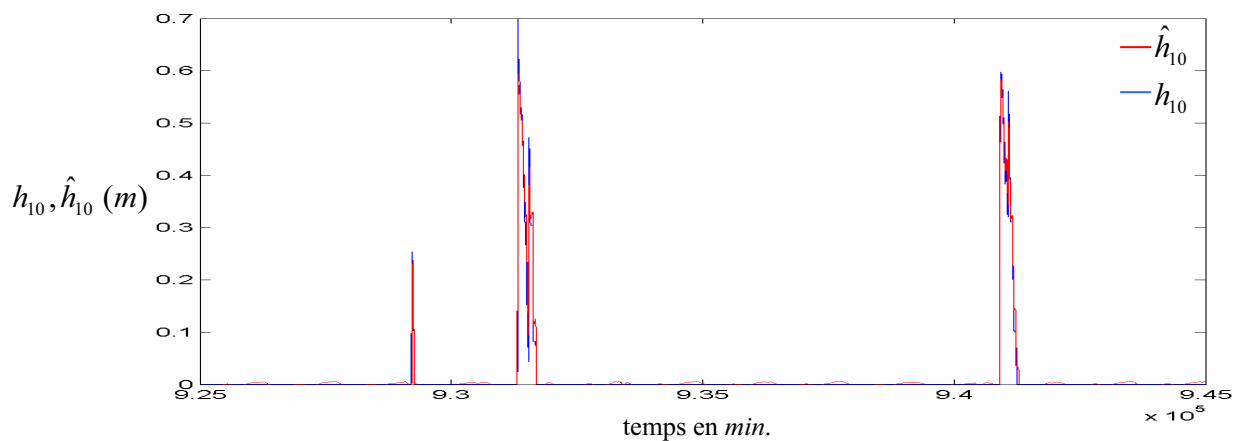
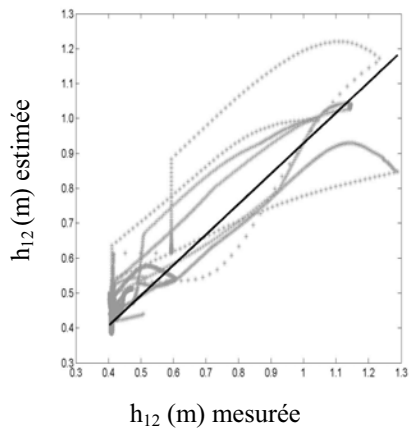


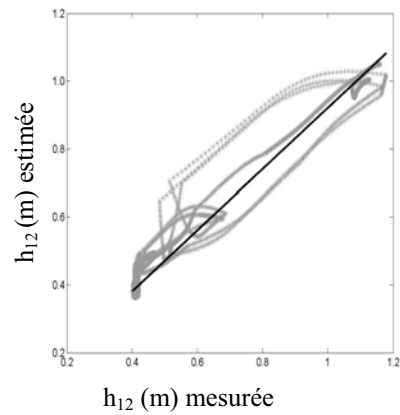
Figure D.9. La mesure h_{10} et son estimée \hat{h}_{10} en superposition, données en fonction du temps.

Modèle h_{12}

$$\hat{h}_{12}(k) = -0.10 + 0.87h_{14}(k).$$



a. jeu-identification : données période 1



b. jeu-test : données période 2

Figure D.10. Identification du modèle h_{12} sur période 1 et validation sur période 2.

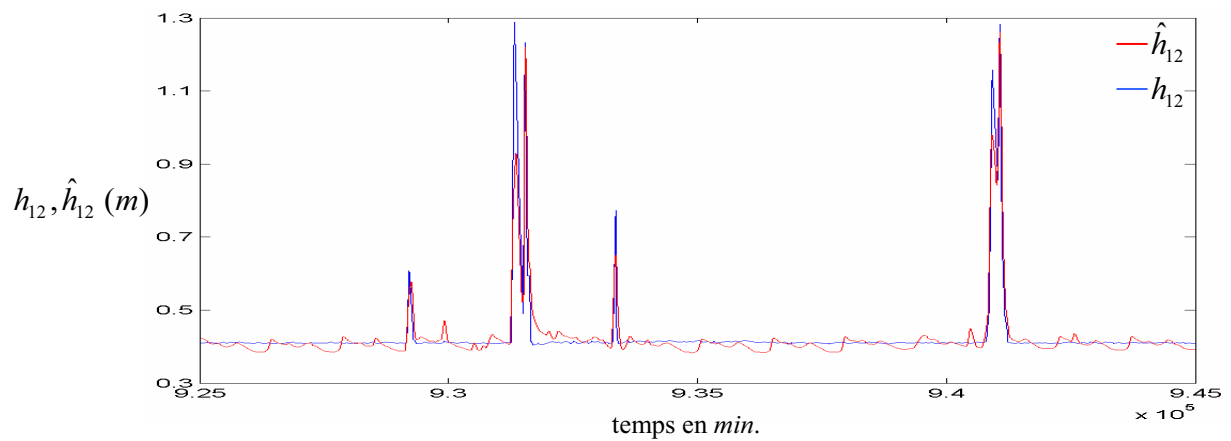
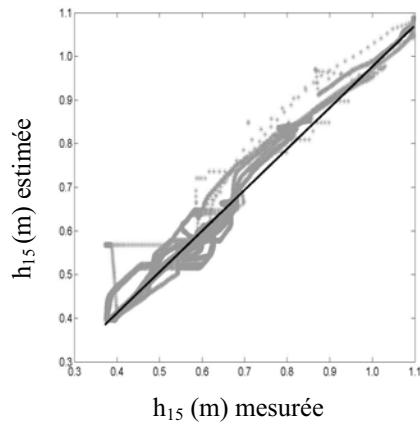


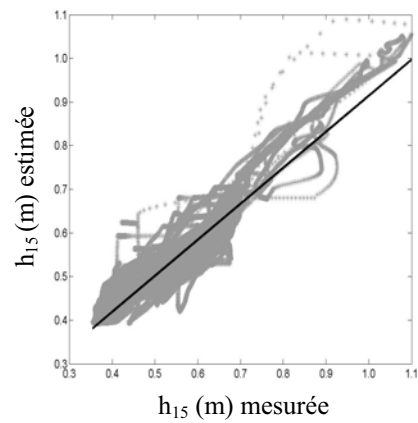
Figure D.11. La mesure h_{12} et son estimée \hat{h}_{12} en superposition, données en fonction du temps.

Modèle h_{15}

$$\hat{h}_{15}(k) = -0.04 + 0.88h_{37}(k).$$



a. jeu-identification : données période 1



b. jeu-test : données période 2

Figure D.12. Identification du modèle h_{15} sur période 1 et validation sur période 2.

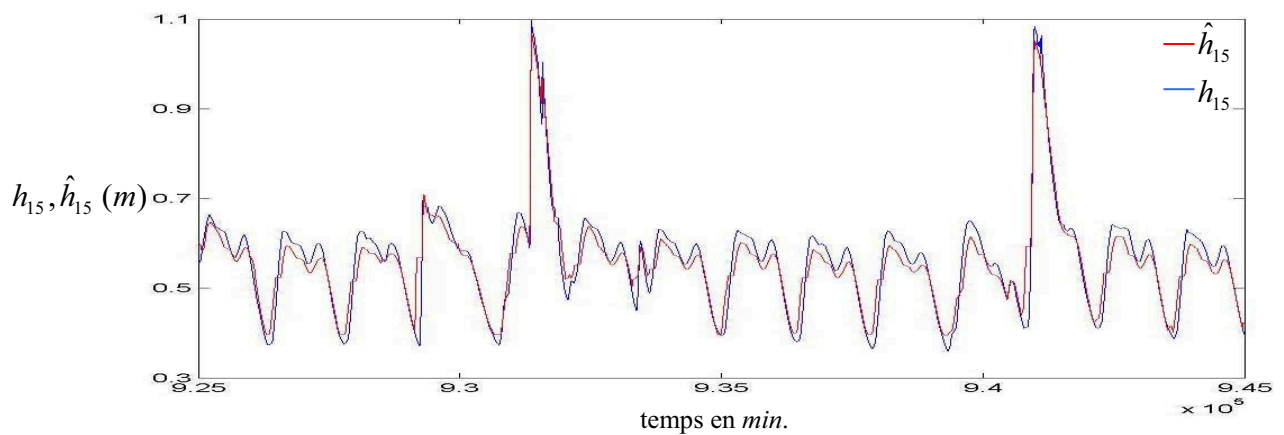
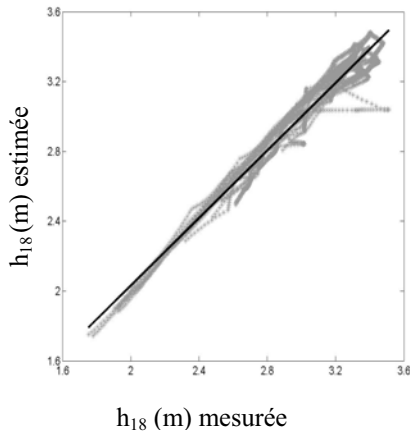


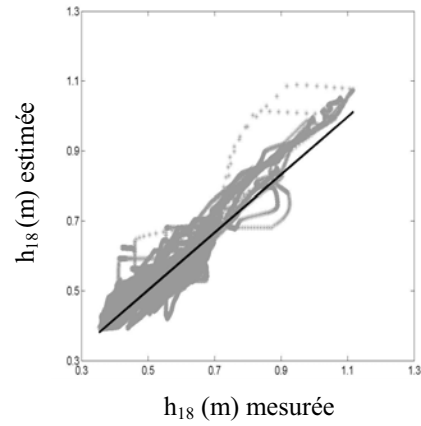
Figure D.13. La mesure h_{15} et son estimée \hat{h}_{15} en superposition, données en fonction du temps.

Modèle h_{18}

$$\hat{h}_{18}(k) = 1.22 + 1.03h_2(k + 30).$$



a. jeu-identification : données période 1



b. jeu-test : données période 2

Figure D.14. Identification du modèle h_{18} sur période 1 et validation sur période 2.

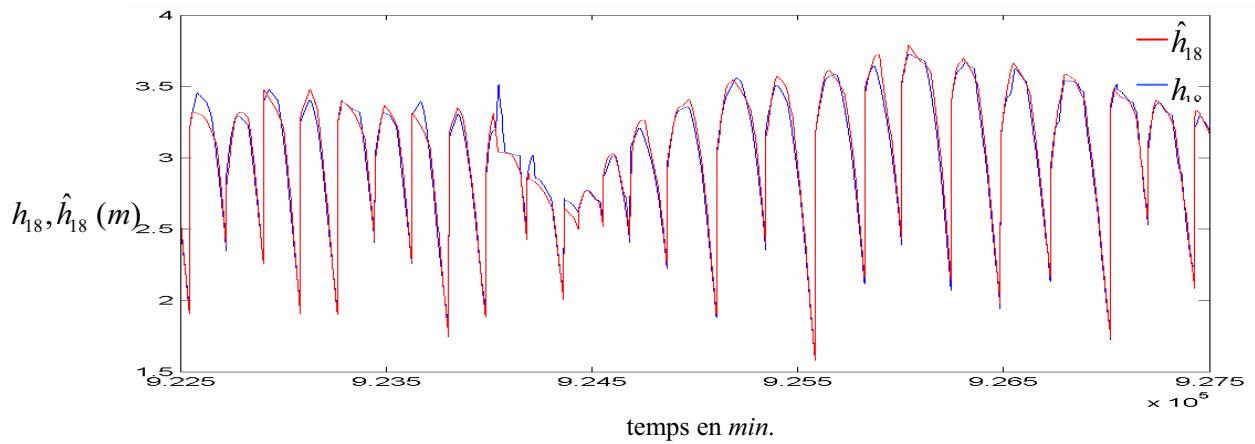


Figure D.15. La mesure h_{18} et son estimée \hat{h}_{18} en superposition, données en fonction du temps.

Modèle h_{20}

$$\hat{h}_{20}(k+30) = -0.09 + 0.79h_2(k).$$

Le signal h_{20} est défini sur une période de 32600 échantillons, la période d'identification est égale à 3000 où le signal est défini à partir d'un seuil égal à 0.92 m.

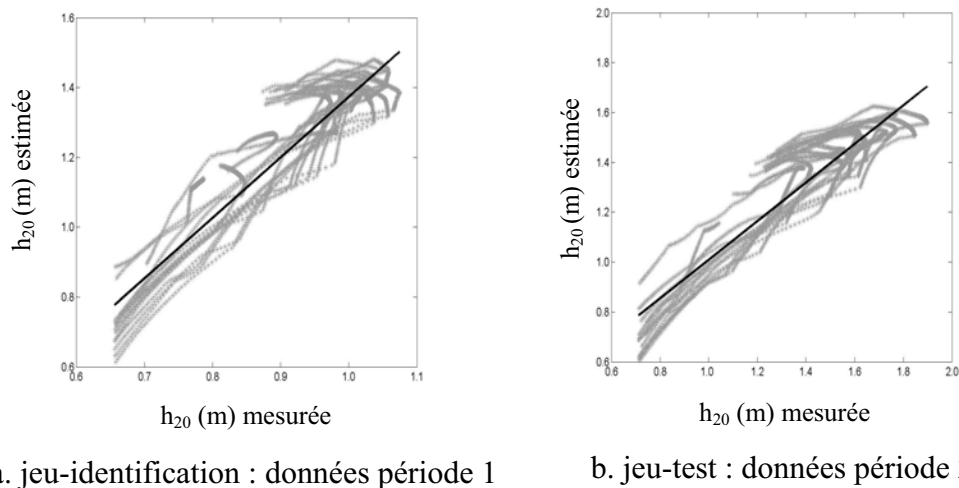


Figure D.16. Identification du modèle h_{20} sur période 1 et validation sur période 2.

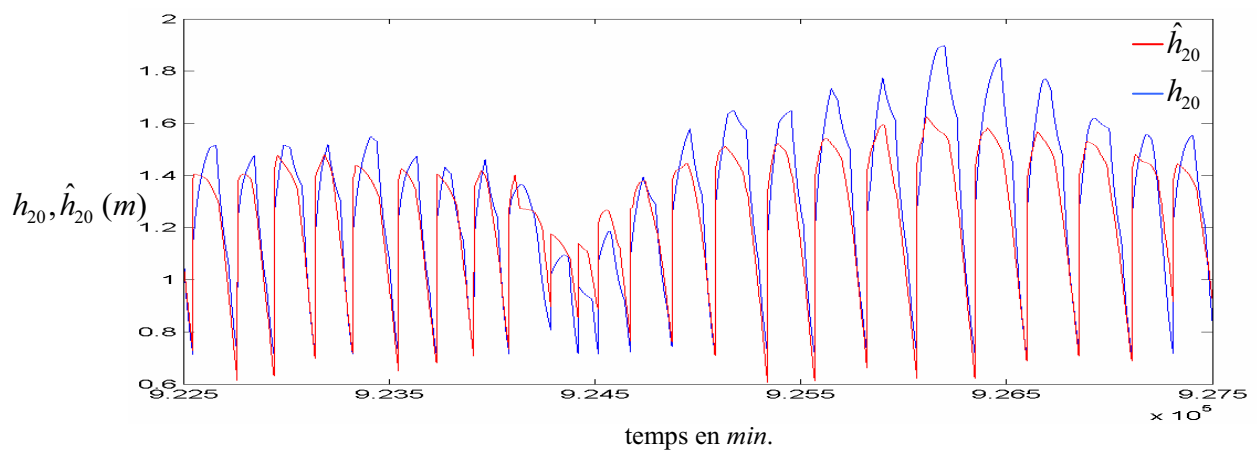
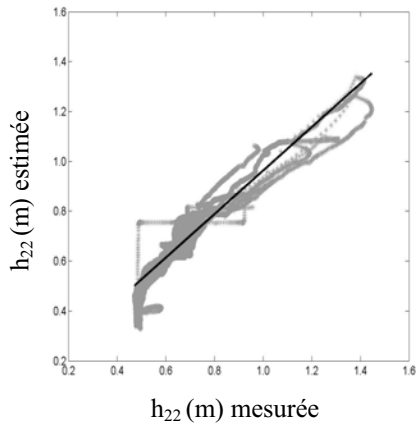


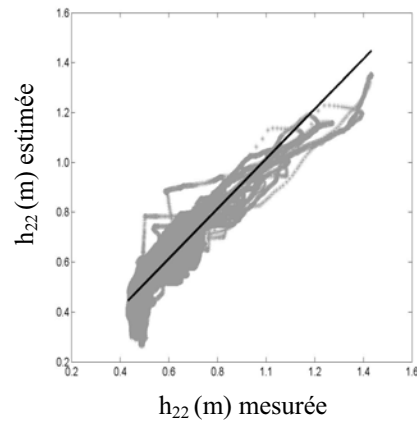
Figure D.17. La mesure h_{20} et son estimée \hat{h}_{20} en superposition, données en fonction du temps.

Modèle h_{22}

$$\hat{h}_{22}(k) = -0.21 + 0.20h_1(k + 140) + 0.75h_4(k + 100) + c$$



a. jeu-identification : données période 1



b. jeu-test : données période 2

Figure D.18. Identification du modèle h_{22} sur période 1 et validation sur période 2.

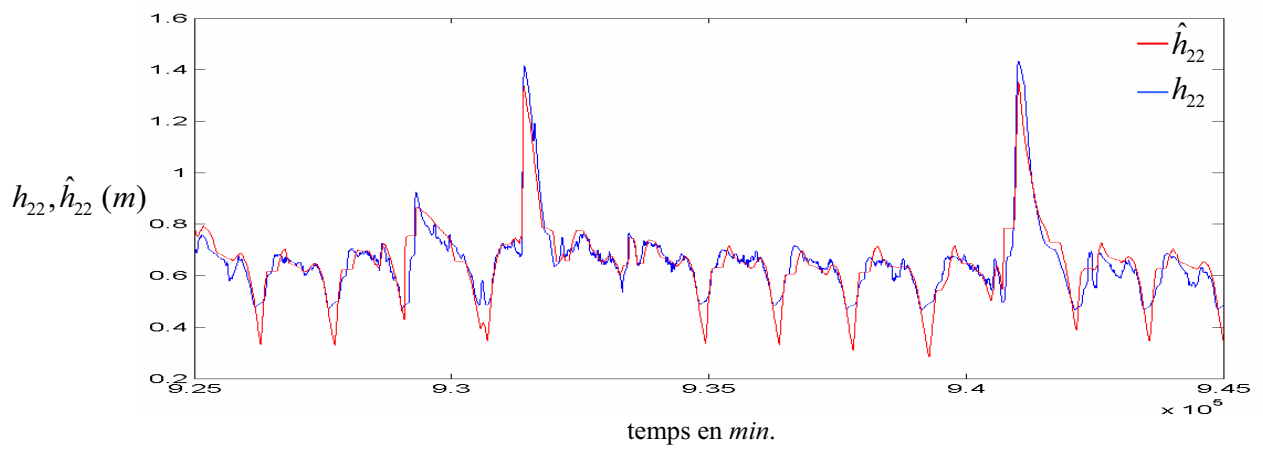
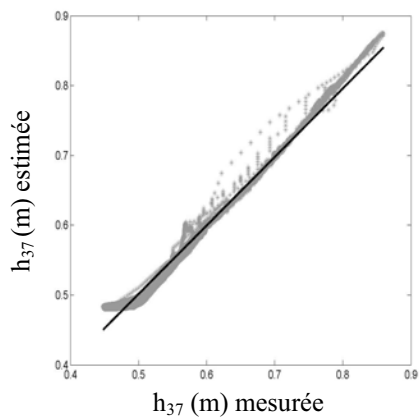


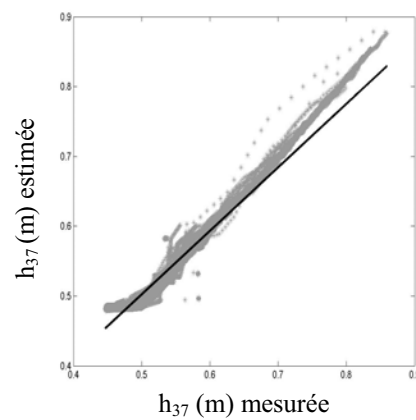
Figure D.19. La mesure h_{22} et son estimée \hat{h}_{22} en superposition, données en fonction du temps.

Modèle h_{37}

$$\hat{h}_{37}(k) = -0.019 + 0.37h_1(k) + 0.66h_{15}(k).$$



a. jeu-identification : données période 1



b. jeu-test : données période 2

Figure D.20. Identification du modèle h_{37} sur période 1 et validation sur période 2.

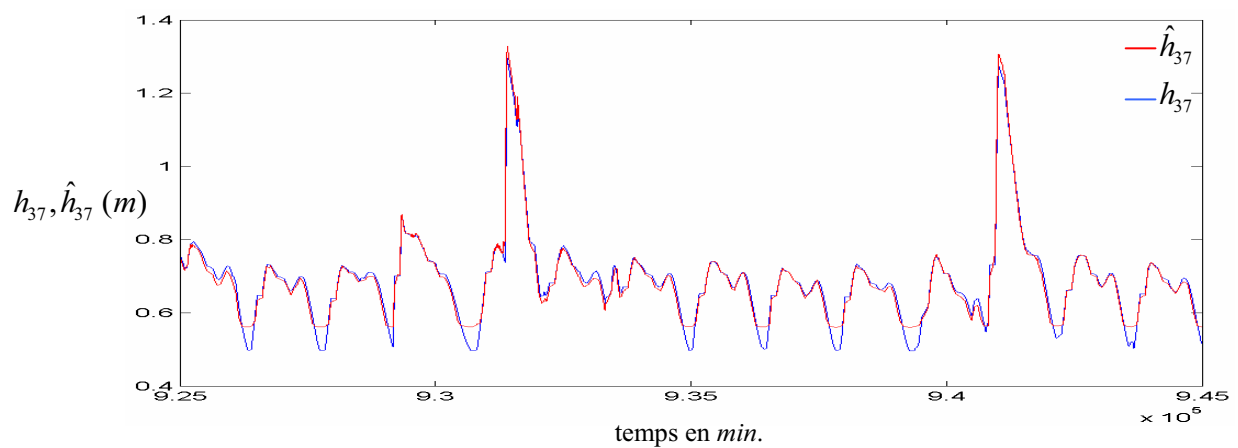
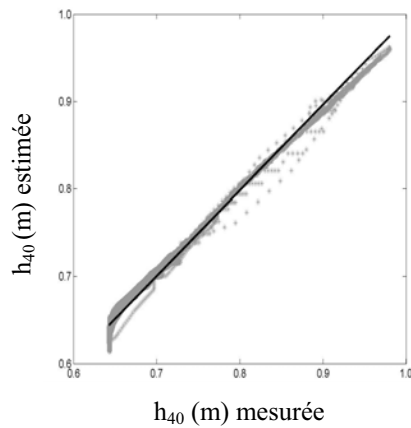


Figure D.21. La mesure h_{37} et son estimée \hat{h}_{37} en superposition, données en fonction du temps.

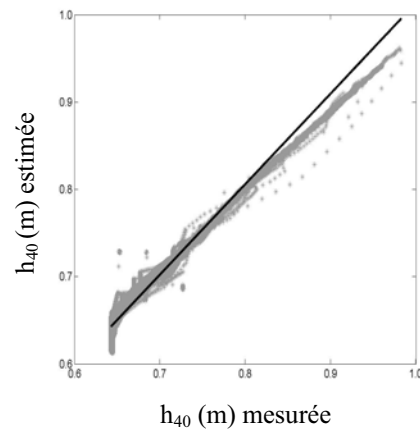
Modèle h_{40}

$$\hat{h}_{40}(k) = 0.20 + 0.85h_{37}(k).$$

Le signal h_{40} est défini à partir d'un seuil égal à 0.92 m .



a. jeu-identification : données période 1



b. jeu-test : données période 2

Figure D.22. Identification du modèle h_{40} sur période 1 et validation sur période 2.

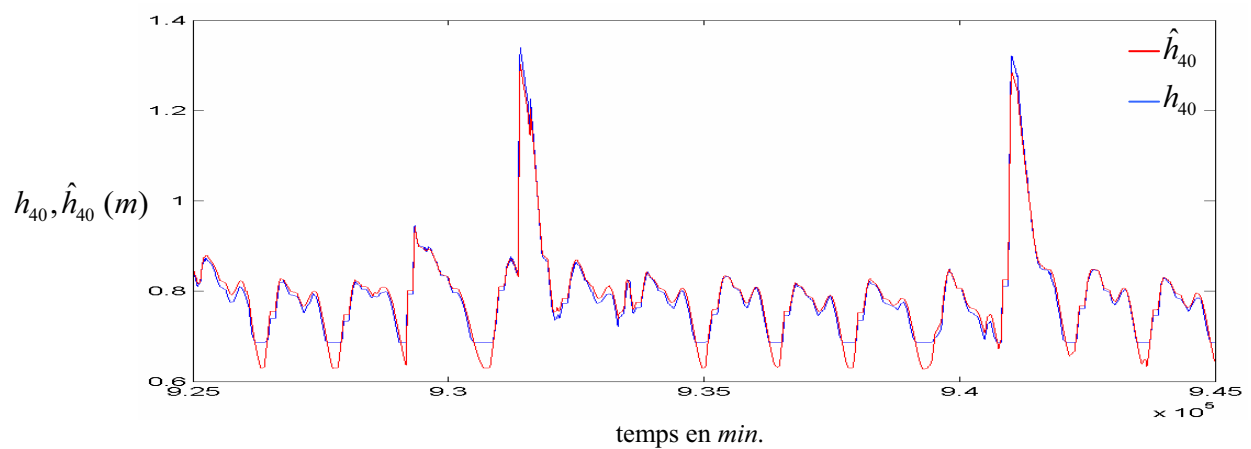


Figure D.23. La mesure h_{40} et son estimée \hat{h}_{40} en superposition, données en fonction du temps.

Les résultats de calcul des deux grandeurs r^2 et s_e pour les trois cas 1, 2, et 3, mentionnés ci-dessus, regroupés dans les tableaux D.1.a, b, c et les tableaux D.2.a, b, c et les tableaux D.3.a, b, c, respectives, donnés ci-dessous.

Dans les tableaux D.1.c et D.2.c et D.3.a, concernant la période 3, nous avons mis des croix dans quelques cases et cela est dû à l'absence de données pour cette période.

Dans tous les tableaux donnés par la suite, au lieu d'indiquer un signal en fonction du temps, nous l'avons noté sans la variable temps indiquée par k et cela pour avoir plus de souplesse dans la notation.

variable mesurée y_i	r^2 en %	s_e	10% de \bar{y}
h_4	98	0.024	0.097
v_5	94	0.010	0.062
h_7	98	0.017	0.060
h_{10}	98	0.015	0.0028
h_{12}	87	0.044	0.044
h_{15}	94	0.033	0.0599
h_{18}	96	0.070	0.296
h_{20}	87	0.083	0.1202
h_{22}	87	0.063	0.071
h_{37}	98	0.021	0.074
h_{40}	98	0.019	0.0832

Tableau D.1.a. Evaluation des modèles avec : période 1 considérée comme jeu-identification

variable mesurée y_i	r^2 en %	s_e	10% de \bar{y}
h_4	98	0.025	0.090
v_5	94	0.032	0.058
h_7	93	0.014	0.058
h_{10}	98	0.013	0.00072
h_{12}	75	0.033	0.042
h_{15}	94	0.035	0.056
h_{18}	96	0.057	0.307
h_{20}	87	0.136	0.136
h_{22}	80	0.060	0.063
h_{37}	98	0.023	0.068
h_{40}	94	0.022	0.078

Tableau D.1.b. Evaluation des modèles avec : période 2 considérée comme jeu-test

variable mesurée y_i	r^2 en %	s_e	10% de \bar{y}
h_4	×	×	×
v_5	×	×	×
h_7	98	0.011	0.0582
h_{10}	98	0.009	0.0001
h_{12}	92	0.030	0.042
h_{15}	94	0.032	0.056
h_{18}	×	×	×
h_{20}	90	0.105	0.1037
h_{22}	89	0.077	0.060
h_{37}	98	0.024	0.0694
h_{40}	96	0.022	0.079

Tableau D.1.c. Evaluation des modèles avec : période 3 considérée comme jeu-test

Les tableaux D.2.a, D.2.b et D.2.c donnés ci-après, sont établis pour le deuxième cas et pour les trois périodes : une pour identification et deux pour validation.

Variable mesurée y_i	r^2 en %	s_e	10% de \bar{y}
h_4	97	0.024	0.090
v_5	83	0.029	0.062
h_7	93	0.014	0.058
h_{10}	93	0.012	0.0007
h_{12}	76	0.033	0.042
h_{15}	91	0.033	0.056
h_{18}	99	0.051	0.31
h_{20}	87	0.011	0.14
h_{22}	81	0.051	0.063
h_{37}	93	0.027	0.067
h_{40}	93	0.020	0.077

Tableau D.2.a. Evaluation des modèles avec : période 2 considérée comme jeu-identification

Variable mesurée y_i	r^2 en %	s_e	10% de \bar{y}
h_4	98	0.026	0.097
v_5	90	0.016	0.061
h_7	98	0.017	0.060
h_{10}	98	0.017	0.0027
h_{12}	72	0.028	0.043
h_{15}	92	0.036	0.060
h_{18}	97	0.074	0.29
h_{20}	87	0.11	0.13
h_{22}	85	0.072	0.068
h_{37}	97	0.025	0.071
h_{40}	97	0.023	0.079

Tableau D.2.b. Evaluation des modèles avec : période 1 considérée comme jeu-test

variable mesurée y_i	r^2 en %	s_e	10% de \bar{y}
h_4	×	×	×
v_5	×	×	×
h_7	98	0.010	0.058
h_{10}	98	0.010	0.0012
h_{12}	92	0.028	0.041
h_{15}	94	0.032	0.057
h_{18}	×	×	×
h_{20}	91	0.15	0.12
h_{22}	82	0.061	0.061
h_{37}	93	0.027	0.069
h_{40}	96	0.023	0.079

Tableau D.2.c. Evaluation des modèles avec : période 3 considérée comme jeu-test

Les tableaux D.3.a, D.3.b et D.3.c donnés ci-après, sont établis pour le troisième cas et pour les trois périodes : une pour identification et deux pour validation.

variable mesurée y_i	r^2 en %	s_e	10% de \bar{y}
h_4	×	×	×
v_5	×	×	×
h_7	98	0.017	0.060
h_{10}	98	0.015	0.0028
h_{12}	87	0.044	0.044
h_{15}	94	0.033	0.0599
h_{18}	×	×	×
h_{20}	87	0.083	0.1202
h_{22}	87	0.063	0.071
h_{37}	98	0.021	0.074
h_{40}	98	0.019	0.0832

Tableau D.3.a. Evaluation des modèles avec : période 3 considérée comme jeu-identification

variable mesurée y_i	r^2 en %	s_e	10% de \bar{y}
h_4	×	×	×
v_5	×	×	×
h_7	93	0.014	0.0577
h_{10}	98	0.013	0.00072
h_{12}	75	0.033	0.0419
h_{15}	94	0.035	0.0556
h_{18}	96	0.057	0.307
h_{20}	87	0.136	0.136
h_{22}	80	0.060	0.063
h_{37}	98	0.023	0.068
h_{40}	94	0.022	0.078

Tableau D.3.b. Evaluation des modèles avec : période 1 considérée comme jeu-test

Variable mesurée y_i	r^2 en %	s_e	10% de \bar{y}
h_4	×	×	×
v_5	×	×	×
h_7	98	0.011	0.0582
h_{10}	98	0.009	0.0001
h_{12}	92	0.030	0.042
h_{15}	94	0.032	0.056
h_{18}	×	×	×
h_{20}	90	0.105	0.1037
h_{22}	89	0.077	0.060
h_{37}	98	0.024	0.0694
h_{40}	96	0.022	0.079

Tableau D.3.c. Evaluation des modèles avec : période 2 considérée comme jeu-test

A partir de ces tableaux, nous remarquons que :

- $r^2\% \geq 80\%$ pour toutes les variables considérées,
- les valeurs des s_e sont en majorité inférieures à 10% de \bar{y} .

Pour la variable $h_{10}(k)$, nous avons dans les trois tableaux une valeur de $s_e > 50\%\bar{y}$, cela signifie que le modèle correspondant ne sera pas valide malgré que $r^2\% > 80\%$ pour ce modèle. En ce qui concerne la variable $h_{12}(k)$, nous remarquons que $s_e \geq 13\%\bar{y}$ pendant la deuxième période de validation. Pour le modèle correspondant à la variable $h_{12}(k)$, nous pouvons dire qu'il est utile vu que nous ne sommes pas loin du cas idéal.

Dans la suite, nous avons noté par modèle Q_j , le modèle qui permet d'expliquer le *jème* débit (exprimé en fonction des hauteurs et des vitesses comme indiqué dans le chapitre 3), par des débits explicatifs (exprimés également en fonction des hauteurs et des vitesses).

Modèle Q_{24}

$$Q_{24}(k) = 1.56Q_{30}(k+15) + 0.043Q_{21}(k+8) - 0.044.$$

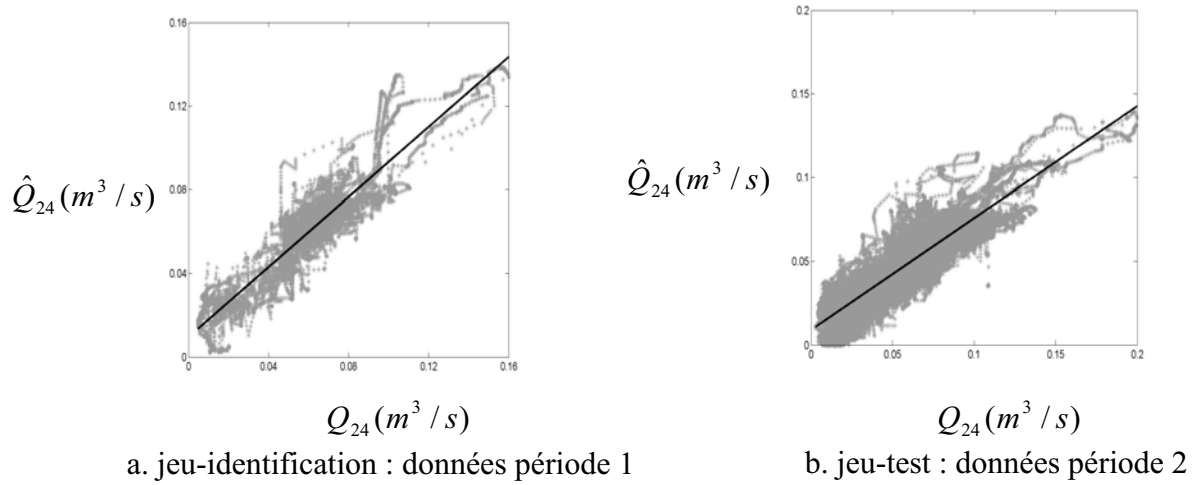


Figure D.24. Identification du modèle Q_{24} sur période 1 et validation sur période 2.

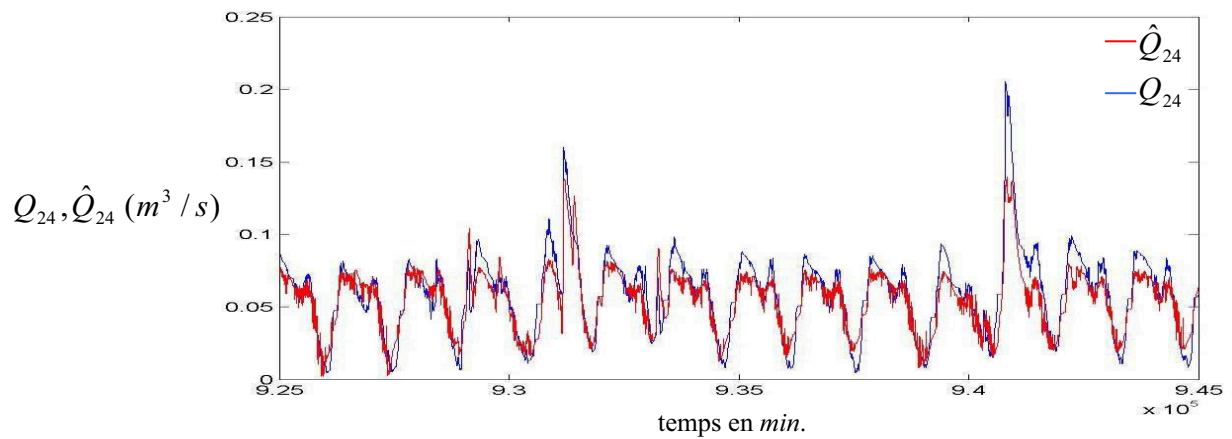


Figure D.25. Allure de Q_{24} et \hat{Q}_{24} en superposition en fonction du temps.

Modèle Q_{30}

$$Q_{30}(k) = 0.33Q_{24}(k-15) + 0.039.$$

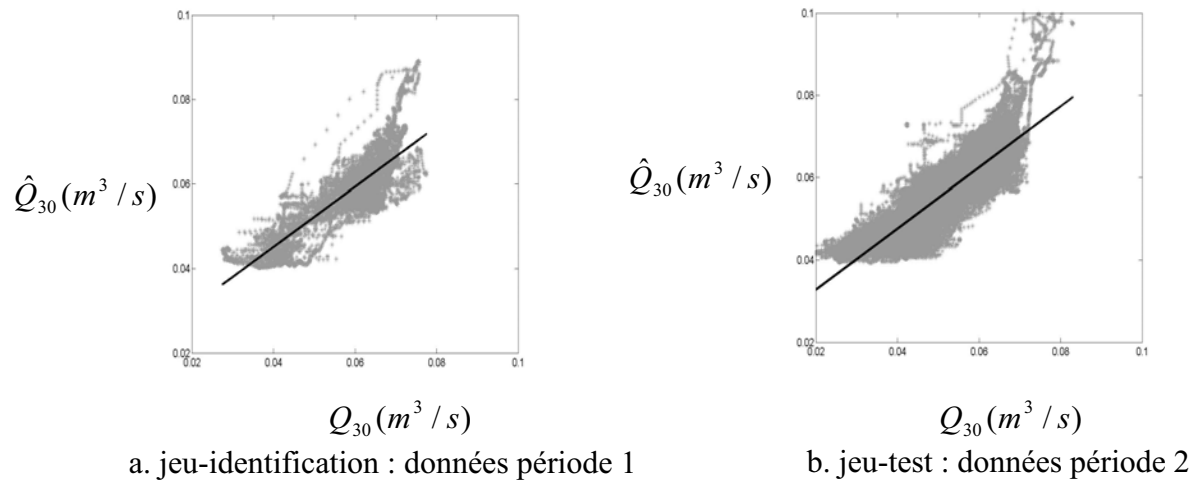


Figure D.26. Identification du modèle Q_{30} sur période 1 et validation sur période 2.

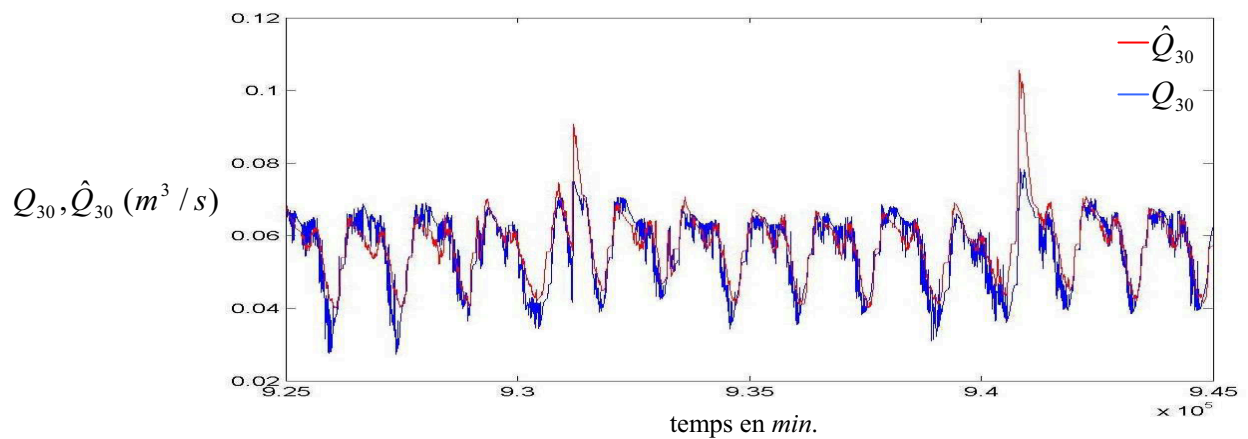


Figure D.27. Allure de Q_{30} et \hat{Q}_{30} en superposition en fonction du temps

Modèle Q_{36}

$$Q_{36}(k) = 0.19Q_6(k + 116) + 0.27Q_{21}(k - 12) + 2.39Q_{30}(k - 10) + 0.016.$$

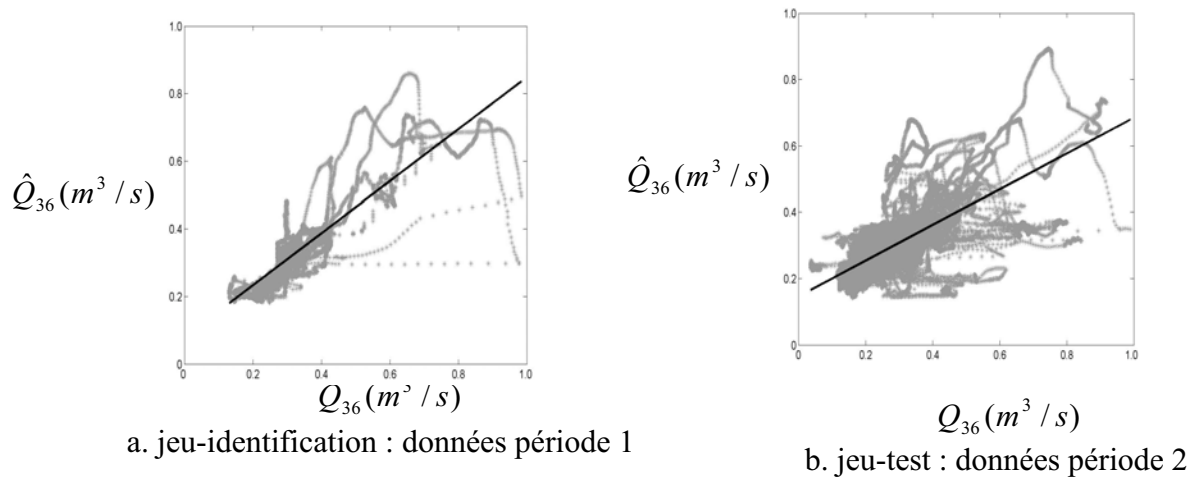


Figure D.28. Identification du modèle Q_{36} sur période 1 et validation sur période 2.

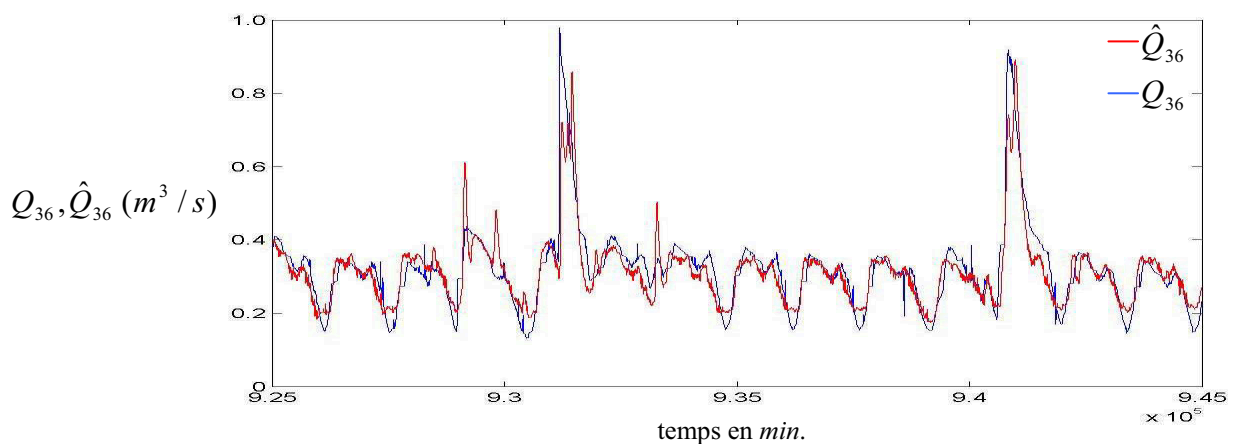


Figure D.29. Allure de Q_{36} et \hat{Q}_{36} en superposition en fonction du temps

Les résultats de calcul des deux grandeurs r^2 et s_e pour les modèles débits pour les deux premiers cas d'identification et de validation 1 et 2, mentionnés ci-dessus, sont regroupés dans les tableaux D. 4 et D. 5 respectives, donnés ci-dessous.

Variable mesurée y_i	r^2 en %	s_e	10% de \bar{y}
Q_6	90	0.074	0.060
Q_{24}	84	0.010	0.0059
Q_{30}	70	0.0058	0.0058
Q_{21}	72	0.14	0.029
Q_{36}	79	0.065	0.035

Tableau D.4. Evaluation des modèles débits avec : période 1 considérée comme jeu-identification

Variable mesurée y_i	r^2 en %	s_e	10% de \bar{y}
Q_6	84	0.071	0.054
Q_{24}	79	0.017	0.0063
Q_{30}	68	0.0062	0.0058
Q_{21}	66	0.14	0.023
Q_{36}	78	0.052	0.032

Tableau D.5. Evaluation des modèles débits avec : période 2 considérée comme jeu-test

Nous remarquons que pour les trois modèles \hat{Q}_6 , \hat{Q}_{24} et \hat{Q}_{36} , nous avons obtenu une valeur de s_e proche de 10% de \bar{Q}_6 , une valeur de s_e très proche de 20% de \bar{Q}_{24} et une valeur de s_e très proche de \bar{Q}_{36} . Nous remarquons aussi que pour le modèle \hat{Q}_{24} et le modèle \hat{Q}_{36} nous avons obtenu une valeur de r^2 proche de 80 mais supérieur à 80 pour le modèle \hat{Q}_6 . Ces modèles peuvent être considérés pour la suite de notre travail. Les deux modèles \hat{Q}_{21} et \hat{Q}_{30} ne seront pas considérés par la suite car ils ne vérifient pas les conditions de validation.

résidu \ variable	h_1	h_2	h_4	v_5	Q_6	h_7	h_{10}	h_{11}	h_{12}	h_{14}	h_{15}	h_{18}	h_{20}	Q_{21}	h_{22}	Q_{24}	Q_{30}	Q_{36}	h_{37}	h_{40}
r_4	-0.99	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r_5	-0.44	0	0.17	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r_6	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	-0.28		-1.60	0	0
r_7	0	0	0	0	0	1	-1.06	0	0	0	0	0	0	0	0	0	0	0	0	0
r_{10}	0	0	0	0	0	-0.98	1	0.05	0	0	0	0	0	0	0	0	0	0	0	0
r_{12}	0	0	0	0	0	0	0	0	1	-0.87	0	0	0	0	0	0	0	0	0	0
r_{15}	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	-0.88	0
r_{18}	0	-1.03	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
r_{20}	0	-0.79	0	0	0	0	0	0	0	0	0		1	0	0	0	0	0	0	0
r_{22}	-0.20	0	-0.75	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
r_{24}	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.04	0	1	-1.56		0	0
r_{36}	0	0	0	0	-1.89	0	0	0	0	0	0	0	0	-0.27	0	0	-2.39	1	0	0
r_{37}	-0.32	0	0	0	0	0	0	0	0	0	-0.68	0	0	0	0	0	0	0	0	0
r_{40}	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	-0.85	1

Tableau E.1. Les sensibilités $S_{i,j}$ pour l'ensemble de modèles établis.

résidu \ variable	h_1	h_2	h_4	v_5	Q_6	h_7	h_{10}	h_{11}	h_{12}	h_{14}	h_{15}	h_{18}	h_{20}	Q_{21}	h_{22}	Q_{24}	Q_{30}	Q_{36}	h_{37}	h_{40}
r_4	-1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r_5	-1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r_6	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	-1	0	-1	0	0
r_7	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0
r_{10}	0	0	0	0	0	-1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
r_{12}	0	0	0	0	0	0	0	0	1	-1	0	0	0	0	0	0	0	0	0	0
r_{15}	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	-1	0
r_{18}	0	-1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
r_{20}	0	-1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
r_{22}	-1	0	-1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
r_{24}	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	1	-1	0	0	0
r_{36}	0	0	0	0	-1	0	0	0	0	0	0	0	0	-1	0	0	-1	1	0	0
r_{37}	-1	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0
r_{40}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	1

Tableau E.2. Matrice des signatures des défauts théoriques Λ_{ij} , pour l'ensemble de modèles établis

AUTORISATION DE SOUTENANCE DE THESE
DU DOCTORAT DE L'INSTITUT NATIONAL
POLYTECHNIQUE DE LORRAINE

o0o

VU LES RAPPORTS ETABLIS PAR :

Monsieur Belkacem OULD BOUAMAMA, Professeur, USTL-LAGIS, Villeneuve d'Ascq

Monsieur Patrice CHATELLIER, Directeur de Recherche, LCPC, Paris

Le Président de l'Institut National Polytechnique de Lorraine, autorise :

Madame HAMIPOUD Farida

à soutenir devant un jury de l'INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE,
une thèse intitulée :

NANCY BRABOIS
2, AVENUE DE LA
FORET-DE-HAYE
BOITE POSTALE 3
F — 5 4 5 0 1
VANDŒUVRE CEDEX

"Validation de données débitmétriques issues de réseaux d'assainissement"

en vue de l'obtention du titre de :

DOCTEUR DE L'INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE

Spécialité : « **Automatique et traitement du signal** »

Fait à Vandoeuvre, le 08 janvier 2007

Le Président de l'IN.P.L.,

F. LAURENT



Résumé

L'objectif de ce mémoire de thèse est de valider les mesures *débimétriques* issues des réseaux d'assainissement. Des défaillances techniques dans les capteurs de mesure ou l'existence des phénomènes perturbateurs produisent des anomalies sur ces mesures. Notre travail consiste à proposer une procédure de diagnostic hors ligne capable de détecter et de localiser les défauts de mesures. Cette procédure est à base de modèles analytiques.

Nous avons établi dans un premier temps, un ensemble de modèles linéaires pour générer des résidus. Ces derniers ont permis d'établir la matrice des signatures des défauts expérimentales. Ensuite, une première technique de diagnostic a été considérée. Elle consiste à combiner le test de *Page* pour la détection des défauts et le calcul de distance Euclidienne entre les signatures des défauts théoriques et la signature expérimentale de chaque défaut pour la localisation des capteurs en défauts. Puis, deux autres techniques ont été proposées. Elles sont basées sur l'analyse conjointe des résidus générés et des symptômes des défauts. L'une est appelée DMP (Model Diagnostic Processor) et l'autre est à base de concepts de la logique floue. A la fin, les performances de ces techniques ont été calculées pour pouvoir les comparer. Nous avons amélioré les performances de la dernière technique en proposant un nouvel opérateur d'agrégation et en considérant la persistance des défauts.

Mots clés: défaut, détection, diagnostic, floue, localisation, redondance, réseau d'assainissement, résidu.

Abstract

The objective of this thesis is to validate flow rate measurements collected from the sewer networks. Technical failures in the sensors or the existence of the disturbing phenomena produce anomalies to these measures. Our work consists in proposing a procedure of diagnosis off line able to detect and isolate the defects of measurements. This procedure is based on analytical models.

We first established, a set of linear models used to generate faults indicators called residuals. These last ones allowed us to establish the experimental faults signatures matrix. Then, a first technique of diagnosis was considered. It consists in combining the test of Page for faults detection and on the calculation of Euclidean distance between the theoretical faults signatures and the experimental fault signature of each defect for the isolation of the faulty sensor. Two other techniques suggested, are based on the joint analysis of the generated residuals and the symptoms of the defects. One is called DMP (Model Diagnostic Processor) and the other is based on the fuzzy logic concepts. In the end, the performances of these techniques were calculated to be able to compare them. We improved the performances of the last considered technique by proposing a new operator of aggregation and by considering the persistence of the defects.

Keywords: fault, detection, diagnosis, fuzzy, isolation, redundancy, residual, sewer network.