



École Doctorale Mathématiques, Sciences de l'Information et de l'Ingénieur

---

UdS - INSA - ENGEES

## THESE

présentée pour obtenir le grade de

**Docteur de l'Université de Strasbourg**  
**Discipline : Mécanique et ingénierie des systèmes**  
**Spécialité : Mécanique des fluides**

par

**Damien KUSS**

# **Gestion intégrée de rejets d'assainissement : applicabilité de la mesure de pollution par spectrophotométrie UV/Visible et des techniques d'identification de système**

Soutenance publique prévue le 16 décembre 2010

### **Membres du jury :**

Directeur de thèse : M. Alain HEDUIT, Directeur de recherches, CEMAGREF

Rapporteur externe : Mme Sylvie BARRAUD, Maître de conférences, Université de Lyon 1

Rapporteur externe : Mme Marion GILSON, Maître de conférences, Université de Nancy 1

Examineur : M. Hugues GARNIER, Professeur des universités, Université de Nancy 1

Examineur : M. Bruno TASSIN, Directeur de recherches, Ecole des ponts ParisTech

Examineur : M. Philippe ACKERER, Directeur de recherches, LHYGES

Membre invité : M. Antoine-Georges SADOWSKI, Directeur de l'équipe HU ENGEES/IMFS

Membre invité : M. José VAZQUEZ, Maître de conférences, ENGEES/IMFS

---

**IMFS**

---

## Résumé

Depuis 1964 et la première loi sur l'eau, une attention toujours plus fine est portée à la préservation des milieux aquatiques. La stratégie la plus récente de lutte contre la pollution des milieux récepteurs est l'approche intégrée, consacrée par la directive cadre sur l'eau du 23 octobre 2000. Globale, dynamique et complexe, cette approche nécessite l'emploi de techniques permettant d'évaluer finement la dynamique des flux de pollutions rejetés dans les milieux récepteurs par l'ensemble du système. Dans ce cadre, la thèse a illustré l'intérêt de la spectrophotométrie UV/Visible et des techniques d'identification de systèmes.

La première application concerne la mesure en continu de concentrations en DCO et MES en entrée de station de traitement des eaux usées. Il a été montré que les performances des modèles statistiques PLS (Partial Least Squares) utilisés pour la mesure des concentrations pouvaient être améliorées par une sélection optimisée des longueurs d'onde. De plus, afin de circonscrire le domaine d'utilisation des modèles et de compléter de manière adéquate la collection d'étalonnage, nous avons proposé une démarche d'établissement de domaine de validité des modèles.

La détection de pics de pollution bactériologique par spectrophotométrie UV/Visible a ensuite été explorée. Des problèmes expérimentaux ont rendu impossible l'utilisation de méthodes d'analyse multivariées. Cependant, l'augmentation concomitante des valeurs d'absorbance et de teneurs en *Escherichia coli* par temps de pluie a pu être mise en exergue.

Enfin, des techniques d'identification de modèle à temps continu ont été appliquées dans le cadre d'une modélisation pluie-débit en contexte urbain. Elles permettent de formaliser le problème mathématique d'estimation paramétrique d'une manière à la fois élégante et performante. Elles présentent également l'avantage de ne pas fixer *a priori* la structure mathématique du modèle à identifier, tout en conservant la possibilité d'une interprétation hydrologique.

**Mots clés** : Gestion intégrée - Système d'assainissement - Mesure en continu - Identification de systèmes - Modélisation pluie-débit

---

## Abstract

Integrated wastewater system management is the most recent strategy developed in order to minimize wastewater discharges impacts on receiving waters. Global, dynamic and complex, this strategy requires to assess the loads of wastewater discharged by the whole system. Within this framework, the thesis has shown the interest of : a) continuous-time measurements of pollutants by UV/Visible spectroscopy ; b) system identification techniques.

The first study carried out concerns the real-time measurement of COD and SS concentrations at the inlet of a wastewater treatment plant. It has been shown that the performance of PLS models used for the measurement of concentrations can be improved by an optimized selection of wavelengths. Moreover, when the dataset used for calibration is small, it is crucial to restrict the use of models within a validity domain. We propose a methodology mainly based on normalised spectra classification and on comparison of calibration data with statutory measurements.

Detection of bacteriological peaks of pollution with UV/Visible spectroscopy has also been investigated. Experimental issues made impossible the calibration of statistical models. However, a concomitant increase of absorbance values and E. coli concentrations were observed during rainfall events.

Finally, system identification techniques have been applied to the rainfall-runoff modelling of an urban catchment. A continuous-time model taking into account infiltration processes and runoff transfer was identified. This method allows to formalize the mathematical problem of parameter estimation in a elegant and successful way. Another benefit is that the model structure does not need to be fixed *a priori*, while preserving the possibility of an hydrological interpretation of the identified model.

**Keywords** : Integrated wastewater system management - *in-situ* sensor - UV/Vis spectroscopy - system identification - rainfall-runoff modelling

---

## Avant-propos

Mes remerciements s'adressent en premier lieu à Antoine-Georges Sadowski qui m'a donné goût à la recherche et qui a été à l'initiative de cette thèse. Je le remercie profondément pour ses conseils avisés, pour sa gentillesse, son humanité et son éclectisme.

Je dois également beaucoup à Hugues Garnier qui m'a fait connaître l'identification de systèmes et avec lequel j'ai pris un très grand plaisir à travailler. Je le remercie aussi pour son soutien dans les moments de doute.

Je remercie ensuite tous les membres du jury de thèse, et plus particulièrement les deux rapporteurs, Marion Gilson et Sylvie Barraud, pour avoir accepté d'évaluer mon travail.

J'exprime mes sincères remerciements à Alain Héduit, mon directeur de thèse, pour avoir accepté de diriger jusqu'au bout ce travail, bien que cela n'ait pas été prévu de la sorte initialement.

Merci également à José Vazquez pour avoir trouvé un partenariat avec Véolia eau, sans lequel le financement de la thèse par le ministère de l'agriculture n'aurait été possible. Merci aussi pour son appui, son encadrement et pour les discussions scientifiques au sommet du Mauna Kea.

Je souhaiterais inclure dans mes remerciements Mathieu Zug, ingénieur à Véolia eau, pour m'avoir fourni des données expérimentales et m'avoir permis de découvrir le fonctionnement d'un projet de recherche piloté par une multinationale de l'eau.

Je remercie du fond du coeur Jonathan Wertel et Jean-Bernard Bardiaux, mes collègues et amis, pour leur soutien, leur relecture et toutes les discussions passées autour d'un café, d'un repas ou d'une table de ping-pong.

J'exprime aussi ma gratitude à toutes les personnes qui ont contribué d'une manière ou d'une autre à mon travail de recherche : Marie-Pierre Ottermatte, Carole Fillingier, Martin Fischer, Vincent Laurain, Stéphane Vuilleumier, Robert Mosé, Thierry Leviandier, Joseph Erhardt, Nadia Pfeiffer-Puh, Jean-Paul Schlepp.

Je voudrais par ailleurs remercier toute l'équipe enseignante et encadrante de l'ENGEES, qui m'a redonné goût aux études, et m'a permis de m'épanouir pendant ces six années. Merci notamment à Claude Chef Giordano, Bruno Chadebec, Sylvie Valentin et Dominique Montero.

Merci enfin à ma femme, Valérie, mon soutien des tous le instants, sans qui je n'aurais jamais eu le courage de reprendre mes études et de me lancer dans cette thèse.

*A ma femme Valérie et à Amaury*

*Le monde n'en a jamais autant su autant sur lui-même et  
sur sa nature qu'aujourd'hui, pourtant ça ne lui sert à rien.  
Arturo Perez Reverte - Le peintre de batailles*

# Table des matières

<b>Résumé</b>	<b>1</b>
<b>Abstract</b>	<b>2</b>
<b>Avant-Propos</b>	<b>3</b>
<b>Table des matières</b>	<b>5</b>
<b>Liste des tableaux</b>	<b>12</b>
<b>Table des figures</b>	<b>15</b>
<b>I Introduction</b>	<b>24</b>
<b>1 Analyse réglementaire vis à vis de la gestion intégrée</b>	<b>25</b>
1.1 L'approche intégrée : aboutissement d'une gestion séculaire des eaux résiduaires	25
1.1.1 Évolution des objectifs assignés à l'assainissement . . . . .	25
1.1.2 Prise en compte progressive de toutes les sources de dégradation des milieux récepteurs . . . . .	27
1.1.3 Multiplication des stratégies de lutte contre la dégradation des milieux récepteurs . . . . .	29
1.2 Déclinaison de l'approche intégrée au niveau opérationnel . . . . .	34
1.2.1 Cas général . . . . .	34

---

1.2.2	Contexte spécifique aux zones de baignade . . . . .	37
1.3	Conclusion . . . . .	40
<b>2</b>	<b>Contexte scientifique et objectifs de la thèse</b>	<b>42</b>
2.1	Projet GIRAC . . . . .	42
2.1.1	Présentation du projet . . . . .	42
2.1.2	Insertion de la thèse au sein du projet GIRAC . . . . .	43
2.1.3	Nature du partenariat avec Véolia eau . . . . .	43
2.2	Modélisation intégrée bassin d'orage - station de traitement d'eaux usées . .	44
2.2.1	Description de l'étude . . . . .	44
2.2.2	Utilité de la mesure de pollution par spectrophotométrie UV/Visible .	46
2.3	Objectifs de la thèse . . . . .	47
<b>II</b>	<b>Mesures en continu de pollution par spectrophotométrie UV/Visible</b>	
	<b>Application à la mesure de concentration en DCO et MES en entrée de</b>	
	<b>station de traitement des eaux usées</b>	<b>50</b>
<b>3</b>	<b>Synthèse bibliographique sur l'utilisation de la spectrophotométrie UV/Visible</b>	<b>52</b>
	<b>pour la mesure en continu de pollution en système d'assainissement</b>	
3.1	La spectrophotométrie UV/Visible . . . . .	53
3.1.1	Rayonnement lumineux . . . . .	53
3.1.2	Absorption moléculaire . . . . .	54
3.1.3	Diffusion optique . . . . .	55
3.1.4	Analyse quantitative : loi de Beer-Lambert . . . . .	55
3.2	Méthodes d'analyse multivariées . . . . .	57
3.2.1	Position du problème de quantification . . . . .	57
3.2.2	Méthodes statistiques . . . . .	58
3.2.3	Méthode déterministe : déconvolution de spectre . . . . .	62
3.3	Procédure de calage et de validation d'un modèle prédictif de concentrations	63
3.3.1	Collecte des données d'étalonnage ( <i>Étape 1</i> ) . . . . .	64

---

3.3.2	Analyse des données ( <i>Étape 2</i> ) . . . . .	64
3.3.3	Étalonnage et validation d'un modèle prédictif ( <i>Étape 3 et 4</i> ) . . . . .	64
3.3.4	Prédictions des concentrations et mesures de contrôle ( <i>Étape 5</i> ) . . . . .	66
3.4	Spectrophotomètres in-situ . . . . .	67
3.4.1	STIP-scan (Endress-Hauser) . . . . .	67
3.4.2	STAC (Secomam) . . . . .	68
3.4.3	scan spectrolyser (scan Messtechnik GmbH ) . . . . .	69
3.5	Applications de la spectrophotométrie UV/Visible pour la quantification de paramètres de pollution physico-chimique dans les eaux usées . . . . .	71
3.5.1	Paramètres de pollution étudiés . . . . .	71
3.5.2	Domaine d'application . . . . .	75
3.5.3	Collecte des données d'étalonnage et de validation . . . . .	76
3.5.4	Élimination de valeurs "aberrantes" . . . . .	77
3.5.5	Méthodes de régression testées . . . . .	77
3.5.6	Performances des méthodes d'étalonnage . . . . .	79
3.5.7	Bruit . . . . .	81
3.5.8	Problématique de mesure des concentrations par temps de pluie . . . . .	82
3.6	Conclusion . . . . .	83
<b>4</b>	<b>Mesure de pollution en entrée de station de traitement d'eaux usées</b>	<b>86</b>
4.1	Introduction . . . . .	86
4.2	Matériel et méthodes . . . . .	87
4.2.1	Étape 1 : description du site d'étude . . . . .	89
4.2.2	Étape 2 : collecte des données d'étalonnage . . . . .	91
4.2.3	Étape 3 : analyse des données collectées . . . . .	95
4.2.4	Étape 4 : étalonnage et validation des modèles prédictifs . . . . .	95
4.2.5	Description des algorithmes de sélection de longueurs d'onde utilisés . . . . .	96
4.2.6	Étape 5 : détermination du domaine de validité des modèles prédictifs . . . . .	101
4.2.7	Étape 6 : mesures in-situ et mesures de contrôle . . . . .	102

---

4.3	Résultats . . . . .	102
4.3.1	Qualité des mesures (étape 3) . . . . .	102
4.3.2	Étalonnage et validation croisée (étape 4) . . . . .	110
4.3.3	Domaine de validité des modèles (étape 5) . . . . .	129
4.3.4	Invalidité des mesures in-situ 2010 : description du dysfonctionnement	138
4.3.5	Utilisation des modèles sur la chronique de mesures in-situ de 2008 (étape 6) . . . . .	139
4.4	Conclusion . . . . .	150

### **III Applicabilité de la spectrophotométrie UV/Visible pour la détection de pollution bactériologique à court terme** **153**

#### **5 Étude bibliographique** **155**

5.1	Les indicateurs de contamination fécale . . . . .	156
5.1.1	Notion d'indicateur de contamination fécale . . . . .	156
5.1.2	Groupes de bactéries utilisés comme indicateur de contamination fécale	156
5.2	Dénombrement des germes indicateurs de contamination fécale . . . . .	159
5.2.1	Méthodes d'énumération basées sur la mise en culture . . . . .	159
5.2.2	Méthode enzymatique d'évaluation rapide de la teneur en E. coli . . .	161
5.2.3	Méthodes moléculaires . . . . .	161
5.3	Indicateurs chimiques de pollution fécale . . . . .	162
5.3.1	Stérols/Stanols . . . . .	163
5.3.2	Caféine . . . . .	163
5.3.3	Azurants optiques . . . . .	164
5.4	Variabilité spatiale des germes ICF au niveau des zones de baignade . . . . .	165
5.4.1	Variabilité spatiale perpendiculairement à la plage . . . . .	166
5.4.2	Variabilité spatiale parallèlement à la plage . . . . .	167
5.4.3	Variation de la teneur en ICF suivant la verticale . . . . .	167
5.5	Liens entre germes ICF et certains facteurs environnementaux . . . . .	168

---

5.5.1	Influence des marées sur la teneur en ICF . . . . .	168
5.5.2	Influence de l'ensoleillement sur la teneur en ICF . . . . .	170
5.5.3	Influence d'autres facteurs environnementaux . . . . .	171
5.5.4	Prédiction de la teneur en ICF à partir de paramètres environnementaux	173
<b>6</b>	<b>Applicabilité de la spectrophotométrie UV/Visible pour la détection de pics de pollution bactériologique</b>	<b>178</b>
6.1	Collecte des données de concentrations en E. coli et des spectres d'absorbance	180
6.1.1	Site étudié - Mesures . . . . .	180
6.1.2	Spectrophotomètre in-situ en ligne . . . . .	180
6.1.3	Prélèvement et conservation d'échantillons . . . . .	182
6.1.4	Dénombrement de E. coli . . . . .	183
6.1.5	Constitution de la base de données . . . . .	183
6.2	Analyse de la qualité des mesures . . . . .	184
6.2.1	Dérive des mesures spectrophotométriques . . . . .	184
6.2.2	Biais dans le protocole de prélèvement . . . . .	187
6.2.3	Conclusion . . . . .	188
6.3	Recherche d'une relation spectre d'absorbance - concentration en E. coli . . . . .	189
6.3.1	Analyse des mesures réalisées en 2007 . . . . .	189
6.3.2	Analyse des mesures réalisées en 2008 . . . . .	195
6.3.3	Analyse des mesures réalisées en 2009 . . . . .	196
6.3.4	Synthèse . . . . .	205
6.4	Propositions d'amélioration du protocole expérimental . . . . .	206
6.4.1	Découpage du protocole expérimental en deux phases . . . . .	206
6.4.2	Organisation des prélèvements . . . . .	207
6.4.3	Quantification des incertitudes de mesure . . . . .	207
6.4.4	Modification de la cuve de mesure ou mesure en by-pass . . . . .	208
6.4.5	Maintenance du spectrophotomètre . . . . .	210
6.4.6	Contrôle qualité . . . . .	210

---

6.5	Conclusion . . . . .	210
<b>IV Modélisation pluie-débit de bassins versants urbains à partir de données expérimentales</b>		<b>213</b>
<b>7</b>	<b>Présentation des approches usuelles de modélisation pluie-débit en milieu urbain</b>	<b>215</b>
7.1	Perception de la transformation pluie-débit . . . . .	215
7.2	Questions préalables à la représentation de la transformation pluie-débit . . .	217
7.2.1	Modèle global ou distribué . . . . .	217
7.2.2	Type de modèle . . . . .	219
7.3	Représentation de la transformation pluie-débit . . . . .	220
7.3.1	Fonction de production . . . . .	221
7.3.2	Fonction de transfert . . . . .	225
7.4	Conclusion . . . . .	228
<b>8</b>	<b>Proposition d'un modèle pluie-débit</b>	<b>229</b>
8.1	Apports des techniques d'identification de systèmes et insertion au sein du modèle pluie-débit . . . . .	230
8.1.1	Méthodologie d'identification de systèmes . . . . .	230
8.1.2	Approche de modélisation DBM . . . . .	233
8.1.3	Identification directe de modèles à temps continu . . . . .	234
8.1.4	Couplage de modèle conceptuel et de modèle "statistique" . . . . .	235
8.2	Modèle pluie-débit proposé . . . . .	236
8.2.1	Modèle de pertes initiales . . . . .	236
8.2.2	Modèle de pertes continues et transfert du ruissellement . . . . .	238
8.2.3	Méthodologie itérative de calage de l'ensemble du modèle pluie-débit	250
8.3	Conclusion . . . . .	253

---

---

<b>9</b>	<b>Modélisation pluie-débit sur le bassin versant du Stang Alar (Brest)</b>	<b>255</b>
9.1	Site d'étude . . . . .	255
9.2	Acquisition de mesures . . . . .	257
9.2.1	Mesures pluviométriques . . . . .	257
9.2.2	Mesures de débit ruisselé à l'exutoire du bassin versant . . . . .	260
9.3	Analyses des hyétogrammes et des hydrogrammes résultants . . . . .	262
9.4	Constitution des chroniques de calage et de validation . . . . .	265
9.5	Résultats . . . . .	266
9.5.1	Comparaison des modèles MISO et SISO . . . . .	267
9.5.2	Valeurs des paramètres du modèle de pertes initiales . . . . .	268
9.5.3	Équivalence avec les modèles de réservoirs et interprétation physique . . . . .	271
9.5.4	Estimation de valeurs optimisées pour le domaine de validité du modèle . . . . .	275
9.6	Conclusion . . . . .	281
	<b>Conclusion générales et perspectives</b>	<b>283</b>
	<b>Annexes</b>	<b>302</b>
<b>A</b>	<b>Validation du dosage de DCO par la méthode de reflux en système fermé suivi d'un dosage par colorimétrie avec le bichromate de potassium</b>	<b>304</b>
<b>B</b>	<b>Transfert de spectres</b>	<b>307</b>
<b>C</b>	<b>Analyse des signaux de concentration pour les modèles fbiPLS et mwPLS</b>	<b>309</b>
<b>D</b>	<b>Régression linéaire effectuée à partir des échantillons mesurés en 2007</b>	<b>313</b>
<b>E</b>	<b>Comparaison des concentrations en E. coli mesurées et simulées en 2007</b>	<b>315</b>

# Liste des tableaux

1.1	Norme guide et norme impérative définies par la directive n° 76/160 pour les paramètres microbiologiques . . . . .	38
1.2	Classes de qualité définies par la directive n° 2006/7/CE . . . . .	38
3.1	Etudes ayant porté sur la mesure de concentrations de paramètres de pollution par spectrophotométrie UV/Visible . . . . .	74
3.2	Emplacement du spectrophotomètre et intervalles de variation de la DCO et MES des échantillons prélevés pour l'étalonnage de la sonde . . . . .	76
3.3	Nombre d'échantillons utilisés pour l'étalonnage de la sonde et partition en jeu de données pour l'étalonnage et la validation des modèles . . . . .	77
3.4	Méthodes de régression utilisées pour l'étalonnage du spectrophotomètre . . . . .	78
4.1	Valeurs des différents paramètres analysés dans le cadre de l'autosurveillance de la STEU de Rosheim (2002 à 2009) . . . . .	90
4.2	Relations entre les paramètres de pollution en entrée de station . . . . .	91
4.3	Résultats des analyses réalisées sur les échantillons prélevés en entrée de la station de Rosheim . . . . .	94
4.4	Synthèse des résultats d'analyses effectués sur les échantillons prélevés en entrée de la station de Rosheim . . . . .	94
4.5	Résultats de l'étalonnage avec une régression linéaire simple . . . . .	111
4.6	Résultats de l'étalonnage avec calage global et local s::can . . . . .	112
4.7	Résultats de l'étalonnage avec le modèle PLS local . . . . .	113
4.8	Résultats de l'étalonnage avec la régression OPPPLS . . . . .	114

---

4.9	Résultats de l'étalonnage avec la régression fbiPLS . . . . .	116
4.10	Résultats de l'étalonnage avec la régression mwPLS . . . . .	118
4.11	Résultats de l'étalonnage avec une régression linéaire simple . . . . .	120
4.12	Résultats de l'étalonnage avec calage global et local s::can . . . . .	121
4.13	Résultats de l'étalonnage avec régression PLS locale utilisant l'intégralité du spectre d'absorbance . . . . .	122
4.14	Résultats de l'étalonnage avec la régression OPPPLS . . . . .	123
4.15	Résultats de l'étalonnage avec la régression fbiPLS . . . . .	125
4.16	Résultats de l'étalonnage avec la régression mwPLS . . . . .	126
4.17	Comparaison des concentrations moyennes 24h en DCO issues des modèles avec la concentration moyenne 24h issue de l'autosurveillance sur la journée du 16/04/2010 . . . . .	146
4.18	Comparaison des concentrations moyennes 24h en MES des modèles avec la concentration moyenne 24h issue de l'autosurveillance sur la journée du 16/04/2010 . . . . .	146
4.19	Erreurs relatives moyennes sur la prédiction des concentrations en DCO pour la campagne de mesure du 16/04/2008 . . . . .	147
4.20	Erreurs relatives moyennes sur la prédiction des concentrations en DCO pour la campagne de mesure du 18/04/2008 . . . . .	147
5.1	Variables incluses dans les modèles de prédiction d'ICF . . . . .	173
5.2	Coefficient de détermination $r^2$ obtenus pour les différents modèles de prédiction des concentrations en ICF . . . . .	174
6.1	Variabilité des spectres de temps sec à 250 et 730 nm . . . . .	184
6.2	Partition des 101 spectres d'absorbance de 2007 en classes de concentrations	192
6.3	Comparaison des valeurs de RMSE pour trois modèles possibles . . . . .	205
8.1	Correspondance des structures de modèles linéaires identifiées avec les modèles de réservoirs . . . . .	250
8.2	Résultats des 30 simulations de Monte-Carlo . . . . .	253
9.1	Caractéristiques des événements pluvieux mesurés . . . . .	259

---

---

9.2	Volumes ruisselés mesurés à l'exutoire du bassin versant pour chacun des événements pluvieux mesurés . . . . .	264
9.3	Meilleurs modèles pluie-débit obtenus suite aux simulations . . . . .	266
9.4	Valeurs de lag-time et de surface active calculées avec les modèles SISO et MISO . . . . .	274
9.5	Valeurs moyennes du coefficient de ruissellement et des pertes initiales calculés par régression linéaire à partir de chaque série de pluie . . . . .	274
9.6	Estimation des paramètres du modèle MISO dont une entrée prend en compte les déversements ou les erreurs de mesure sur les pics de débits élevés . . . .	277
A.1	Erreurs de mesure obtenues avec la méthode normalisée par comparaison avec des solutions étalon . . . . .	305
A.2	Erreurs de mesure obtenues avec la méthode normalisée par comparaison avec des solutions étalon . . . . .	306
B.1	Résultats du transfert pour les spectres d'entrée (220-735nm) . . . . .	308

# Table des figures

1.1	Pollution par les eaux d'égout des agglomérations urbaines de plus de 50.000 habitants, et pollutions de surfaces existantes en 1961. Source : (Colas, 1964)	28
1.2	Exemple de coordination de techniques alternatives sur deux bassins-versants. Source : (Azzout <i>et al.</i> , 1994)	32
1.3	L'approche intégrée : aboutissement d'une gestion séculaire des eaux résiduaires	33
2.1	Flux annuels de DCO biodégradable et N-NH <sub>4</sub> rejetés dans le milieu récepteur pour différentes valeurs du couple volume du bassin d'orage, débit admissible en entrée de STEU	46
2.2	Intégration des thèmes de recherche au sein d'une gestion intégrée de système d'assainissement	49
3.1	Structure du chapitre 3	52
3.2	Spectre électromagnétique	53
3.3	Méthodes mathématiques d'étalonnage de modèles prédictifs de concentrations	58
3.4	Étapes de développement d'un modèle prédictif de concentrations (inspiré de Bertrand (2005))	63
3.5	Procédure de validation croisée	65
3.6	Spectrophotomètre UV/Visible STIP-scan de ENDRESS-HAUSER	68
3.7	Station d'alerte compacte (STAC) de SECOMAM (source : Secomam)	69
3.8	Spectrophotomètre UV/Visible spectro :lyser de s :can Messtechnik GmbH	69
3.9	Schéma de fonctionnement du spectro::lyser (source : Langergraber <i>et al.</i> (2003))	70

---

3.10 Spectres d'absorbance d'un échantillon d'eau usée brute et filtrée à différentes porosités . . . . .	72
3.11 Spectres de MES . . . . .	73
3.12 Surparamétrisation du modèle PLS ou régression linéaire multiple ? (source : Hochedlinger et al. (2006)). . . . .	81
3.13 Bruit sur le signal de concentration en MES et DCO. (source : Maribas et al. (2008)). . . . .	82
3.14 Mesures de concentration de DCO et MES par temps de pluie (source : Maribas et al. (2008)) . . . . .	83
4.1 Méthodologie d'établissement de modèle prédictif proposée . . . . .	88
4.2 Photos du spectro::lyser installé en entrée de la station de traitement des eaux usées de Rosheim . . . . .	92
4.3 Spectres d'absorbance bruts . . . . .	95
4.4 Méthodes d'étalonnage testées pour l'établissement de modèles prédictifs de concentrations en MES et DCO . . . . .	96
4.5 Méthodes de sélection de longueurs d'onde présentées (seules les méthodes affectées d'un numéro ont été testées) . . . . .	97
4.6 Méthodologie de détermination d'un domaine de validité . . . . .	101
4.7 Absorbances à 255 nm mesurées avec les spectrophotomètres Perkin-Elmer et s::can . . . . .	104
4.8 Spectres d'absorbance bruts . . . . .	104
4.9 Spectres d'absorbance normalisés . . . . .	106
4.10 Analyse en composantes principales des spectres d'absorbance normalisés : représentation des individus sur le plan 1-2 . . . . .	107
4.11 Dendogramme des spectres d'absorbance normalisés . . . . .	108
4.12 Coefficient de détermination $r^2$ entre la DCO et l'absorbance pour chaque longueur d'onde . . . . .	109
4.13 Régression linéaire simple $DCO=f(Abs)$ réalisée sur la classe de spectres principale . . . . .	110
4.14 Valeur des fonctions de coût RMSEC et RMSECV en fonction de la longueur d'onde . . . . .	111

---

---

4.15	Concentration en DCO en fonction de l'absorbance à 565 nm . . . . .	112
4.16	Valeur des fonctions de coût RMSEC et RMSECV pour le modèle PLS local	113
4.17	Valeur des fonctions de coût RMSEC et RMSECV avec la régression OPPPLS	114
4.18	Longueurs d'onde sélectionnées dans le modèle OPPPLS1 . . . . .	115
4.19	Valeur des fonctions de coût RMSEC et RMSECV obtenues avec la sélection de longueurs d'onde du modèle fbiPLS3 . . . . .	116
4.20	Représentation graphique de la sélection d'intervalles effectuée pour le modèle fbiPLS3 . . . . .	117
4.21	Valeur des fonctions de coût RMSEC et RMSECV avec la régression mwPLS	118
4.22	Valeur des fonctions de coût RMSEC et RMSECV en fonction de la longueur d'onde . . . . .	120
4.23	Concentration en MES en fonction de l'absorbance à 550 nm . . . . .	121
4.24	Valeur des fonctions de coût RMSEC et RMSECV pour le modèle PLS local	122
4.25	Valeur des fonctions de coût RMSEC et RMSECV avec la régression OPPPLS	123
4.26	Longueurs d'onde sélectionnées avec le modèle OPPPLS1 . . . . .	124
4.27	Valeur des fonctions de coût RMSEC et RMSECV obtenues avec la sélection de longueurs d'onde du modèle fbiPLS2 . . . . .	125
4.28	Représentation graphique de la sélection d'intervalles effectué pour le modèle fbiPLS2 . . . . .	126
4.29	Valeur des fonctions de coût RMSEC et RMSECV avec la régression mwPLS	127
4.30	Classification des spectres mesurés in-situ sur la chronique de mesures d'ab- sorbance à 255 nm . . . . .	130
4.31	Représentation de huit spectres mesurés en 2008 sur le plan de projection 1-2 du modèle d'ACP des spectres normalisés d'étalonnage . . . . .	131
4.32	Représentation d'un panel de spectres mesurés en 2008 sur le plan de projec- tion 1-2 du modèle d'ACP des spectres normalisés d'étalonnage . . . . .	133
4.33	Percentile correspondant à chaque valeur de DCO . . . . .	134
4.34	Percentile correspondant à chaque valeur de MES . . . . .	135
4.35	Variabilité mensuelle des paramètres Débit, DCO, MES et N-NH <sub>4</sub> . . . . .	136
4.36	Variabilité des moyennes mensuelles des paramètres Débit ; DCO, MES et N-NH <sub>4</sub> . . . . .	137

---

---

4.37	Dérive du signal d'absorbance à 255 nm sur les mesures in-situ effectuées en 2010 . . . . .	138
4.38	Signal d'absorbance à 255 nm acquis lors de mesures effectuées in-situ en 2008	139
4.39	Absorbances à 255 et 565 nm mesurées pendant la journée du 18/04/2008 .	140
4.40	Concentrations en DCO obtenues avec six modèles différents pendant la journée du 18/04/2008 . . . . .	143
4.41	Concentrations en MES obtenues avec six modèles différents pendant la journée du 18/04/2008 . . . . .	145
4.42	Concentrations en DCO obtenues avec six modèles différents pendant la journée du 16/04/2008 . . . . .	148
4.43	Concentrations en DCO obtenues avec six modèles différents pendant la journée du 18/04/2008 . . . . .	149
4.44	Méthodologie de prévision à court terme de la concentration en E. coli dans les zones de baignade (les zones en pointillés correspondent aux axes de travail à développer durant la thèse). . . . .	154
5.1	Régression linéaire en coordonnées Log-Log entre le nombre de coliformes fécaux et le nombre de E. coli dans des rivières du bassin de la Seine (Garcia-Armisen, 2006) . . . . .	157
5.2	Diagramme simplifié du cycle de vie de E. coli dans l'environnement (Source : (Ishii et Sadowsky, 2008) . . . . .	159
5.3	Profil en travers type montrant les différents endroits d'échantillonnage étudiés dans l'étude USEPA. Source : (Wymer <i>et al.</i> , 2005) . . . . .	165
5.4	Grille d'échantillonnage type utilisée dans l'étude USEPA. Source : (Wymer <i>et al.</i> , 2005) . . . . .	166
5.5	Influence de la distance à la plage sur la teneur en ICF. Source : (Wymer <i>et al.</i> , 2005) . . . . .	167
5.6	Influence de la profondeur de prélèvement sur la teneur en ICF. Source : (Wymer <i>et al.</i> , 2005) . . . . .	168
5.7	Influence de la marée sur la concentration en EI. (Spring : hautes eaux, Neap : basses eaux, Ebb : déversement des rivières dans la mer, Flood : reflux de la mer dans les estuaires). Source : (Boehm et Weisberg, 2005) . . . . .	169
5.8	Influence du cycle lunaire sur la teneur en ICF. Source : (Boehm et Weisberg, 2005) . . . . .	170

---

5.9	Influence de la marée sur la concentration en EI. Source : (Wymer <i>et al.</i> , 2005)	170
5.10	Influence du soleil sur la teneur en ICF. Source : (Wymer <i>et al.</i> , 2005)	171
5.11	Influence de la pluviométrie, du vent, de la température de l'eau et de l'air sur la teneur en ICF. Source : (Wymer <i>et al.</i> , 2005)	172
6.1	Représentation schématique de la problématique de l'étude visant à établir une relation entre spectres d'absorbance et concentration en E. coli	179
6.2	Bassin versant du Stang ALar à Brest	181
6.3	Photo du dispositif expérimental installé dans la station de pompage de Palaren (Brest)	181
6.4	Photo du spectrophotomètre utilisé	182
6.5	Spectrophotomètre immergé dans la cuve de mesure	182
6.6	Évolution de l'absorbance à 250 nm en 2009	185
6.7	Dix spectres de temps sec mesurés en 2009	185
6.8	Évolution des spectres de temps sec en 2007	186
6.9	Valeurs des concentrations de E. coli e temps sec et par tranche horaire	188
6.10	Coefficient de détermination entre la concentration en E. coli et l'absorbance (2007)	190
6.11	Concentrations en E. coli en fonction de l'absorbance à 730 nm (2007)	191
6.12	Distribution des valeurs d'absorbance pour chacune des six classes de concentration de E. coli (2007)	193
6.13	Analyse en composantes principales des spectres de 2007. Représentation des individus sur le plan 1-2	194
6.14	Coefficient de détermination entre la concentration en E. coli et l'absorbance (2008)	195
6.15	Coefficient de détermination entre la concentration en E. coli et l'absorbance (2009)	196
6.16	Comparaison des concentrations mesurées et simulées pendant l'événement pluvieux du 10/06/2009	198
6.17	Comparaison des concentrations mesurées et simulées pendant la période du 18 au 25/06/2009	199
6.18	Ajustement d'une fonction E. coli=f(Abs <sub>250nm</sub> )	200

---

---

6.19	Concentrations mesurées et simulées à partir de l'ajustement d'une fonction non linéaire $E. coli=f(Abs_{250nm})$ entre le 21/06 et le 25/06/2009 . . . . .	201
6.20	Applicabilité d'un profil journalier moyen de concentration en <i>E. coli</i> sur la période allant du 18/06 au 25/06/2009 . . . . .	202
6.21	Comparaison des concentrations mesurées et simulées pendant la période du 01/07 au 15/07/2009 . . . . .	203
6.22	Comparaison des concentrations mesurées et simulées pendant la période du 15/07 au 25/07/2009 . . . . .	204
6.23	Schéma de modification de la cuve de mesure : vidange par siphon et ajout d'un agitateur hélicoïdal . . . . .	208
6.24	Insert permettant un mesure spectrophotométrique en by-pass . . . . .	209
6.25	Insert permettant un mesure spectrophotométrique en by-pass et nettoyage mécanique . . . . .	209
7.1	Représentation des processus physiques impliqués dans la transformation pluie-débit en milieu urbain . . . . .	216
7.2	Différences de représentation entre modèles globaux et modèles distribués au sein des logiciels d'hydrologie urbaine. (a) : modèle global (b) : modèle distribué, en rouge les exutoires de chaque sous bassin versant où sont injectés les débits pour la modélisation hydraulique . . . . .	219
7.3	Flux de modélisation intégrant les fonctions de production et de transfert . . . . .	220
7.4	Modélisation des pertes initiales par un modèle de réservoir . . . . .	223
7.5	3 représentations de l'infiltration . . . . .	225
7.6	Modèles de réservoirs pour le transfert du ruissellement au sein d'un bassin versant. (a) : réservoir linéaire simple. (b) : cascade de réservoirs . . . . .	226
8.1	Prise en compte des perturbations dans un système. (a) : Système <i>vrai</i> avec perturbations; (b) : Modèle avec représentation des perturbations sous la forme d'un bruit additif sur la sortie . . . . .	231
8.2	Insertion de la procédure d'identification du modèle linéaire à temps continu au sein de la procédure de calage du modèle pluie-débit . . . . .	233
8.3	Approche hybride de modélisation proposée . . . . .	236
8.4	Schéma de principe du modèle de pertes initiales . . . . .	237

---

---

8.5	Extrait de la chronique de pluies et de débits utilisée pour les simulations de Monte-Carlo (30 chroniques de débits comprenant 65475 observations) . . .	252
9.1	Vue aérienne du bassin versant du Stang Alar ( <i>en bleu : le Stang Alar ; en jaune : les branches principales du réseau d'assainissement pluvial</i> ) . . . . .	256
9.2	Situation des postes pluviométriques par rapport au bassin versant du Stang Alar . . . . .	258
9.3	Cumul des lames d'eau précipitées pendant la campagne de mesure sur les trois postes pluviométriques étudiés . . . . .	258
9.4	Loi hauteur-vitesse déterminée à l'exutoire du bassin versant . . . . .	260
9.5	Loi hauteur-débit déterminée à l'exutoire du bassin versant . . . . .	261
9.6	Hyétogrammes et hydrogrammes mesurés du 11/07 au 08/09/2008 . . . . .	263
9.7	Analyse en composantes principales intégrant les données de pluies et de débits	265
9.8	Débit mesuré et simulé (modèles n°1 et 6) le 02/08/2008 . . . . .	267
9.9	Débit mesuré et simulé (modèles n°1 et 6) le 04/09/2008 . . . . .	268
9.10	Hyétogrammes modifiés après application du modèle de pertes initiales . . . .	269
9.11	Évolution de $R_7^2$ en fonction du couple de paramètres du modèle de pertes initiales en calage. (Le modèle linéaire à temps continu étant fixé) . . . . .	270
9.12	Évolution de $R_7^2$ en fonction du couple de paramètres du modèle de pertes initiales en validation. (Le modèle linéaire à temps continu étant fixé) . . . .	270
9.13	Hydrogrammes mesurés et simulés (modèles n° 1 et 2) le 28/07/2008 (Validation) . . . . .	272
9.14	Hydrogrammes mesurés et simulés (modèles n° 1 et 2) le 16/08/2008 (Validation) . . . . .	272
9.15	Hydrogrammes mesurés et simulés (modèles n° 1 et 2) le 02/08/2008 (Validation) . . . . .	273
9.16	Hydrogrammes mesurés et simulés (modèles n° 1 et 2) le 05/09/2008 (Calage)	273
9.17	Hydrogrammes mesurés et simulés (modèle n° 2) le 12/08, 31/07 et 01/09 .	276
9.18	Interprétation schématique du modèle pluie-débit à deux entrées . . . . .	278
9.19	Hydrogrammes mesurés et simulés avec le modèle MISO (31/07 et 01/09/2008)	279
9.20	Décomposition des hydrogrammes simulés le 31/07 et 01/09/2008 en deux composantes . . . . .	280

---

---

A.1	Comparaison de la méthode normalisée ISO 6060 :1989 (référence) et de la méthode micro-DCO . . . . .	306
B.1	Comparaison de deux mesures de spectres d'absorbance réalisées avec le spectrophotomètre Perkin-elmer et s::can . . . . .	308
C.1	Concentrations en DCO obtenues avec les trois modèles fbiPLS pendant la journée du 18/04/2008 . . . . .	310
C.2	Concentrations en DCO obtenues avec les 3 modèles mwPLS pendant la journée du 18/04/2008 . . . . .	311
C.3	Concentrations en MES obtenues avec les trois modèles fbiPLS pendant la journée du 18/04/2008 . . . . .	312
D.1	Régression linéaire effectuée sur 98 individus . . . . .	314
E.1	Comparaison des concentrations en E. coli mesurées et simulées du 12 au 19/06/2007 . . . . .	316
E.2	Comparaison des concentrations en E. coli mesurées et simulées du 19 au 25/06/2007 . . . . .	316
E.3	Comparaison des concentrations en E. coli mesurées et simulées du 25/06 au 02/07/2007 . . . . .	317
E.4	Comparaison des concentrations en E. coli mesurées et simulées du 02 au 10/07/2007 . . . . .	317
E.5	Comparaison des concentrations en E. coli mesurées et simulées du 10 au 17/07/2007 . . . . .	318
E.6	Comparaison des concentrations en E. coli mesurées et simulées du 17 au 25/07/2007 . . . . .	318
E.7	Comparaison des concentrations en E. coli mesurées et simulées du 25/07 au 01/08/2007 . . . . .	319
E.8	Comparaison des concentrations en E. coli mesurées et simulées du 01 au 15/08/2007 . . . . .	319
E.9	Comparaison des concentrations en E. coli mesurées et simulées du 15 au 30/08/2007 . . . . .	320

**Première partie**

**Introduction**

# Chapitre 1

## Analyse réglementaire vis à vis de la gestion intégrée

L'étude du contexte réglementaire montre trois changements principaux dans la façon d'aborder la protection des milieux récepteurs et la limitation des impacts liés aux rejets urbains de temps de pluie. Ces évolutions permettront de mettre en relief, dans un contexte opérationnel, l'intérêt des thèmes de recherche abordés dans le cadre de la présente thèse.

### **1.1 L'approche intégrée : aboutissement d'une gestion séculaire des eaux résiduaires**

#### **1.1.1 Évolution des objectifs assignés à l'assainissement**

L'évacuation rapide des eaux usées et pluviales hors des zones urbanisées résulte du mouvement hygiéniste apparu au XIX<sup>ème</sup> siècle : le but était principalement la réduction des épidémies liées aux germes pathogènes d'origine fécale (Dupuy et Knaebel, 1982). L'objectif associé était la prévention des inondations. Depuis peu, la société impose au système d'assainissement de nouvelles contraintes comme la préservation des milieux récepteurs et de leur biodiversité. L'étude des différents textes législatifs relatifs à l'eau et l'assainissement est révélatrice de ce changement de paradigme.

---

La première loi sur l'eau du 16 décembre 1964 vise à lutter "*contre la pollution des eaux et leur régénération, dans le but de satisfaire ou concilier les exigences :*

- *de l'alimentation en eau potable des populations et de la santé publique ;*
- *de l'agriculture, de l'industrie, des transports et de toutes autres activités humaines d'intérêt général ;*
- *de la vie biologique du milieu récepteur et spécialement de la faune piscicole ainsi que des loisirs, des sports nautiques et de la protection des sites ;*
- *de la conservation et de l'écoulement des eaux."*

L'ordre dans lequel le législateur énumère les objectifs de la lutte contre la pollution des eaux n'est pas anodin. La préservation des usages et la protection sanitaire, cités dans les deux premiers alinéas, semblent revêtir davantage d'importance que la protection des milieux aquatiques et de leur biodiversité.

La seconde loi du 03 janvier 1992 marque une évolution dans le paradigme qui prévalait jusqu'alors. En effet, l'article 2 précise que la loi a pour objet "*une gestion équilibrée de la ressource en eau. Cette gestion équilibrée vise à assurer :*

- *la préservation des écosystèmes aquatiques, des sites et zones humides.*

...

- *la valorisation de l'eau comme ressource économique."*

La préservation des écosystèmes aquatiques est ici mise en relief avant la préservation des usages. L'ordre de l'énumération montre un rééquilibrage des objectifs au profit de la protection des milieux aquatiques et suggère une complémentarité des objectifs. L'évolution de la réglementation révèle une nouvelle façon de voir le monde, monde dans lequel l'homme n'est qu'un maillon d'un écosystème complexe qu'il cherche à protéger.

Ce glissement opéré entre 1964 et 1992 semble irréversible. En effet, plus proche de nous, la loi du 21 avril 2004 transposant la directive cadre sur l'eau du 23 octobre 2000 martèle l'urgence de restaurer la qualité des milieux aquatiques et fixe des objectifs de qualité. En 2015, toutes les masses d'eau devront atteindre le bon état.

---

### 1.1.2 Prise en compte progressive de toutes les sources de dégradation des milieux récepteurs

Tout au long du XX<sup>ème</sup> siècle, les pressions sur les milieux récepteurs aquatiques n'ont cessé de s'accroître du fait de la croissance économique et de l'expansion des agglomérations. Jusque dans les années 1970, il était acquis que la pollution des milieux récepteurs était essentiellement due à des rejets ponctuels d'eaux usées domestiques et d'eaux industrielles (Colas, 1964; Wolfrom, 1964). La figure 1.1 extraite de la publication de Colas (1964) montre que la plupart des agglomérations de plus de 50 000 habitants n'étaient pas (ou seulement partiellement) dotées de stations de traitement des eaux usées. On peut également remarquer sur cette carte l'impact des rejets en terme de pollution sur les milieux récepteurs. La première loi sur l'eau a permis de financer un grand nombre de stations de traitement d'eaux usées grâce à la création des six agences financières de bassin, plus tard renommées agences de l'eau. Selon le rapport du sénateur Miquel (2003), le nombre de stations de traitement des eaux usées est passé de 2 115 en 1970, à 7 542 en 1980, 11 500 en 1990, et 14 377 fin 2000. Un gros effort a ainsi été effectué concernant le traitement de la pollution de temps sec collectée dans les réseaux d'assainissement, ce qui a permis d'enrayer en partie la dégradation des principaux cours d'eau.

Cependant, suite à de nombreuses études américaines menées à partir des années 1960 aux États-Unis (Weibel *et al.*, 1964; Burm *et al.*, 1968; Bryan, 1972; Sartor et Boyd, 1972) puis en France dans les années 1980-1990 (Philippe et Ranchet, 1987; Deutsch et Hemain, 1984; Chebbo, 1992), nous nous sommes rendus compte que l'impact des rejets urbains de temps de pluie sur le milieu récepteur était très significatif. Les rejets urbains considérés dans ces études sont issus de réseaux d'assainissement unitaires collectant à la fois les eaux usées domestiques et les eaux pluviales.

Une des caractéristiques principales des rejets urbains de temps de pluie est leur teneur importante en matière en suspension liée au lessivage des bassins-versants et à la remise en suspension de dépôts accumulés dans le réseau d'assainissement. Par ailleurs, la majeure partie de la pollution est fixée à ces particules solides (DCO, éléments traces organiques et métalliques,...) (Chebbo, 1992; Chocat *et al.*, 2007; Chebbo et Gromaire, 2004). Enfin, même si les concentrations en polluants des rejets urbains de temps de pluie sont très variables, les flux déversés dans le milieu récepteur sans traitement préalable sont très conséquents.

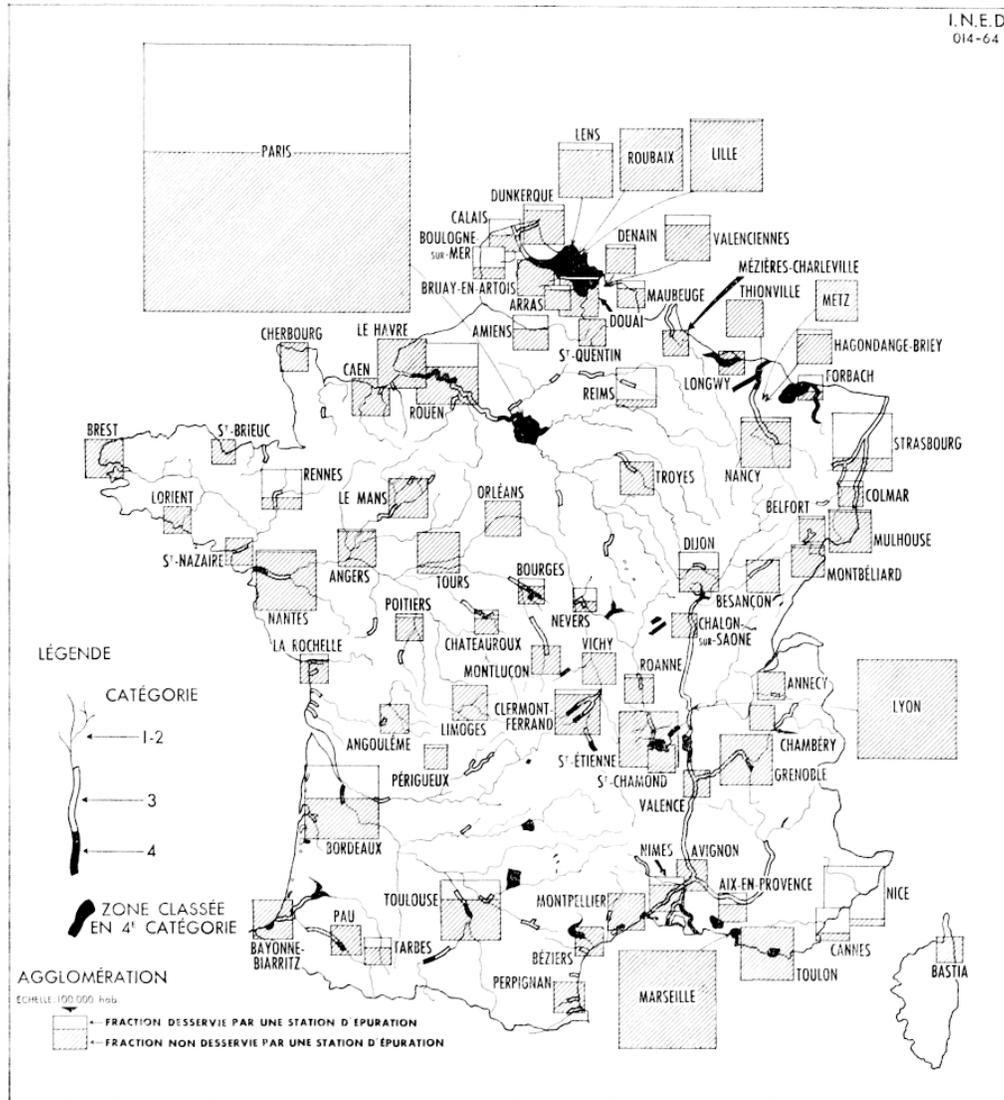


Fig. 1.1 – Pollution par les eaux d'égout des agglomérations urbaines de plus de 50.000 habitants, et pollutions de surfaces existantes en 1961. Source : (Colas, 1964)

---

La directive Eaux Résiduaires Urbaines (ERU) du 21 mai 1991 transposée en droit français par la loi sur l'eau du 03 janvier 1992 traduit le changement dans notre perception de l'origine de la dégradation des milieux récepteurs. Elle prend en compte la nécessité de surveiller les stations de traitement et les eaux réceptrices « pour faire en sorte que l'environnement soit protégé des effets négatifs du déversement des eaux résiduaires », y compris celles résultant d'événements pluvieux<sup>1</sup>.

### **1.1.3 Multiplication des stratégies de lutte contre la dégradation des milieux récepteurs**

L'évolution des objectifs de protection et la meilleure connaissance des sources de pollution ont été accompagnés par un affinement des stratégies de lutte contre la pollution des milieux récepteurs.

Les années 1960 marquent l'ère de la collecte des eaux usées et du traitement de la pollution de temps sec par la construction de stations de traitement d'eaux usées. Par exemple, sur l'agglomération Lilloise, les principales STEU ont été mises en service après 1964 (Marquette (1969) - Houplain-Ancoisne (1977) - Villeneuve d'Ascq (1975)).

Les années 1980 et 1990 voient le développement de trois stratégies complémentaires. La première est une stratégie d'interception et de traitement des rejets unitaires de temps de pluie, basée sur la notion de premier flot d'orage<sup>2</sup>, et qui a conduit à la construction d'ouvrages de rétention (ATV, 1977; FNDAE, 1988). Or quelques études ont montré que la pollution était le plus souvent distribuée de manière uniforme (Bertrand-Krajewski *et al.*, 1998; Herremans *et al.*, 1995) si bien qu'il est très difficile d'intercepter la majeure partie du flux de pollution de temps de pluie étant donné les volumes en jeu. La construction de bassins de rétention qui intercepteraient 80 % de la masse de polluants est de plus économiquement inacceptable.

La directive ERU du 21 mai 1991 officialise la stratégie d'interception des rejets unitaires de temps de pluie. En effet, elle impose aux États membres que « les eaux résiduaires qui pénètrent dans les systèmes de collecte soient soumises à un traitement secondaire ou à un traitement équivalent » (art. 4). Toutefois, dans l'annexe I.A. il est concédé qu'il n'est pas

---

<sup>1</sup>Les ERU sont définies comme composées des eaux ménagères usées ou comme le mélange de celles-ci avec les eaux industrielles et/ou les eaux de ruissellement

<sup>2</sup>Dans la publication de Bertrand-Krajewski *et al.* (1998), il est considéré qu'un premier flot d'orage existe si 80 % (ou plus) de la masse de polluants est contenue dans les premiers 30 % du volume écoulé.

---

possible de construire des systèmes de collecte et des stations d'épuration permettant de traiter toutes les eaux usées dans des situations telles que la survenance de précipitations exceptionnellement fortes. Pour des événements pluvieux de moindre intensité, il convient donc de stocker les rejets urbains dans des bassins d'orage avant de les traiter.

La seconde stratégie, qui a connu un très fort développement dans les années 1980 et 1990, est la collecte séparée des eaux usées domestiques et des eaux pluviales (réseaux séparatifs) (IFEN, 2004). Selon l'IFEN, le linéaire de canalisations d'eaux usées et d'eaux pluviales a augmenté respectivement de 17% et 5% entre 1998 et 2001. Les réseaux unitaires correspondent à un choix ancien, mais leur part ne cesse de diminuer au profit des réseaux séparatifs, et ce quelle que soit la taille de la commune. Dans un réseau séparatif, les eaux usées domestiques sont traitées par la station de traitement des eaux usées et les eaux pluviales sont rejetées après traitement primaire dans les milieux récepteurs. Cette stratégie présente en théorie de multiples avantages. Tout d'abord, le flux d'eaux usées domestiques n'est plus mélangé aux eaux pluviales et les rejets de ces eaux fortement chargées au niveau des déversoirs d'orage sont supprimés. Par ailleurs, le flux hydraulique acheminé à la station de traitement des eaux usées est plus régulier, ce qui constitue un avantage pour leur traitement. Malheureusement, l'efficacité de cette stratégie est entravée par les erreurs de branchements sur le réseau ainsi que par la forte charge des eaux de chaussée et de toiture en micropolluants métalliques.

La troisième stratégie concerne l'utilisation de techniques alternatives (pluviales) s'opposant au « tout au réseau ». Leur objectif est de réguler les débits et de favoriser l'infiltration par l'utilisation de méthodes variées telles que les puits d'infiltration, les chaussées à structure réservoir ou les bassins de retenue. Bien que la mise en place opérationnelle de cette stratégie intervienne le plus souvent dans un objectif de lutte contre les inondations, les bienfaits en terme de protection des milieux récepteurs sont réels. Outre la régulation des débits, les techniques alternatives limitent à *la source* le ruissellement et l'entraînement de particules chargées en polluants organiques ou minéraux vers les milieux récepteurs (Azzout *et al.*, 1994). Quelques villes françaises comme Bordeaux ou Lyon ont intégré dans leur document d'urbanisme des obligations explicites de limitation de l'imperméabilisation et imposent par voie de conséquence l'usage de techniques alternatives. A Bordeaux, depuis 1987, le plan d'occupation des sols impose à tout nouvel aménagement de ne rejeter que le débit correspondant à une imperméabilisation de 30%.

---

Enfin, la dernière stratégie apparue au début des années 2000 est **l'approche intégrée**, consacrée par la directive cadre sur l'eau du 23 octobre 2000 (transposée par la loi 21 avril 2004) et dont une définition partielle est suggérée dans l'article 2 de l'arrêté du 22 juin 2007 "relatif à la collecte, au transport et au traitement des eaux usées des agglomérations d'assainissement ainsi qu'à la surveillance de leur fonctionnement et de leur efficacité [...]":

*"les systèmes de collecte et les stations d'épuration d'une agglomération d'assainissement ainsi que les dispositifs d'assainissement non collectifs doivent être dimensionnés, conçus, réalisés, réhabilités, exploités comme des ensembles techniques cohérents. Les règles de dimensionnement, de réhabilitation et d'exploitation doivent tenir compte des effets cumulés de ces ensembles sur le milieu récepteur de manière à limiter les risques de contamination ou de pollution des eaux [...]".*

De manière plus générale, l'approche intégrée peut être définie comme une approche globale et multidimensionnelle qui :

- intègre l'ensemble des objectifs de protection : protection des milieux récepteurs (biocénose et biotope), protection des usages (AEP, santé publique, agriculture, industrie, pêche,...) mais aussi lutte contre les inondations ;
- prend en compte toutes les sources de pollutions : rejets de temps-sec, rejets de temps de pluie ;
- utilise et coordonne tout l'éventail des stratégies possibles à l'échelle d'une agglomération : traitement de la pollution de temps sec, interception et traitement de la pollution de temps de pluie, séparation des flux d'écoulement, techniques alternatives.

L'approche intégrée vise à minimiser l'impact cumulé de l'ensemble des rejets d'assainissement à l'échelle d'une agglomération. C'est donc une approche systémique qui revient à considérer que la station de traitement des eaux usées, le réseau d'assainissement, les ouvrages de stockage et de surverse forment un seul et unique ensemble dénommé système d'assainissement (CERTU, 2003).

L'approche intégrée est également une approche dynamique qui nécessite la coordination de l'ensemble des stratégies mises en oeuvre à l'échelle d'une agglomération (cf. FIG. 1.2). En effet, tous les ouvrages formant le système d'assainissement interagissent entre eux sur des intervalles de temps très courts. La modification d'un paramètre sur un des ouvrages a des conséquences sur le reste du système. Par exemple, la diminution ou l'augmentation du débit

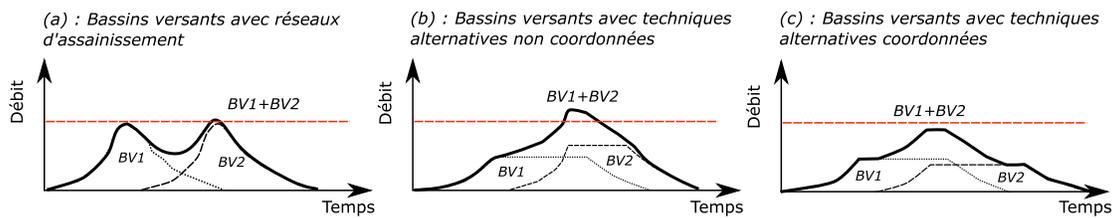


Fig. 1.2 – Exemple de coordination de techniques alternatives sur deux bassins-versants. Source : (Azzout *et al.*, 1994)

admissible d'une station de traitement des eaux usées a des répercussions sur les flux de pollution déversés par l'ensemble du système<sup>3</sup>.

Remarquons enfin que le succès de cette stratégie est fortement conditionné par le travail collaboratif de tous les acteurs de l'aménagement de l'espace urbain : élus, urbanistes, architectes, techniciens, citoyens,...

L'approche intégrée est donc l'aboutissement d'une gestion séculaire des eaux résiduaires. La FIG. 1.3 illustre le cheminement qui a vu l'émergence de cette stratégie.

<sup>3</sup>Dans le cas 'un réseau de collecte unitaire

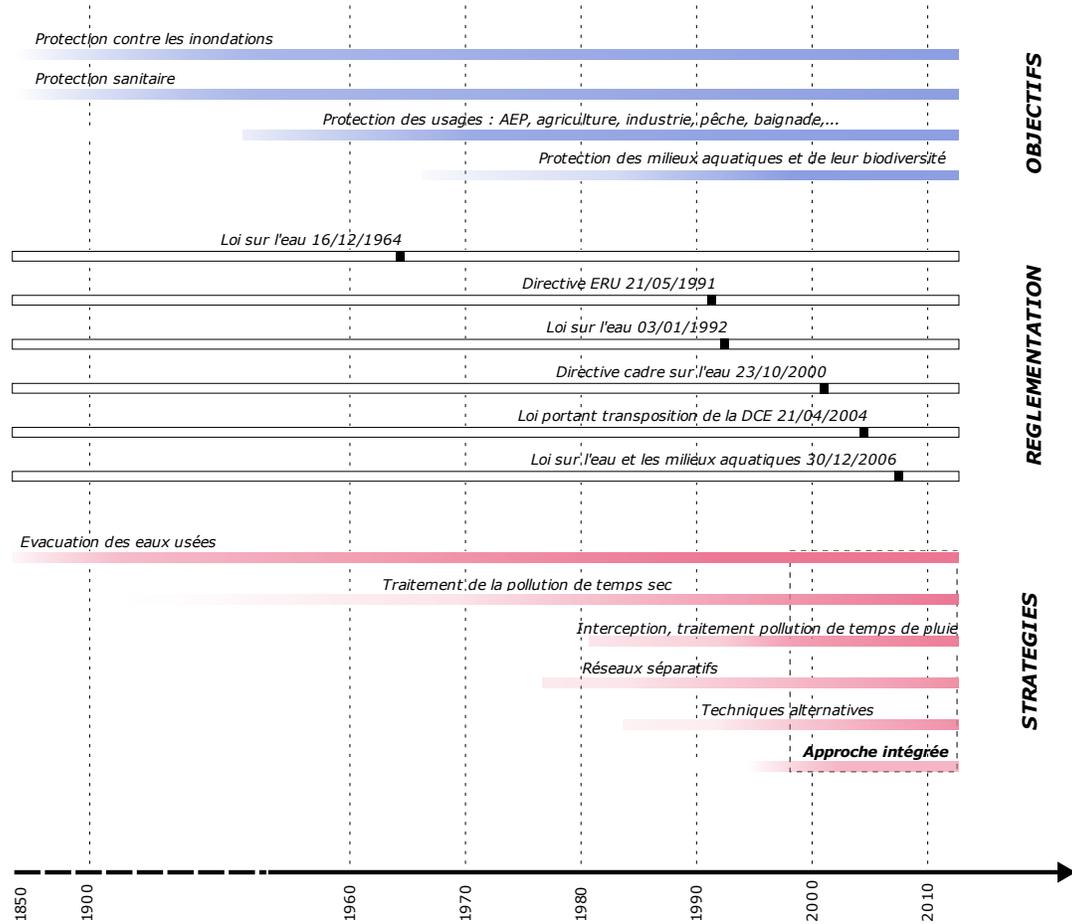


Fig. 1.3 – L’approche intégrée : aboutissement d’une gestion séculaire des eaux résiduaires

---

## 1.2 Déclinaison de l'approche intégrée au niveau opérationnel

### 1.2.1 Cas général

#### 1.2.1.1 L'autosurveillance : outil d'évaluation des flux de pollution rejetés à l'échelle d'un système d'assainissement

La mise en oeuvre opérationnelle d'approches intégrées passe par la connaissance des flux de pollution rejetés à l'échelle d'un système d'assainissement. L'acquisition de cette information est rendue obligatoire par l'arrêté du 22 juin 2007 qui fixe les règles d'autosurveillance au niveau de chaque point de rejet : déversoirs d'orage, bassins, stations de traitement des eaux usées.

L'article 18 de l'arrêté du 22 juin 2007 précise que les déversoirs d'orage "*situés sur un tronçon destiné à collecter une charge brute de pollution organique par temps sec supérieure à 600 kg/j de DBO<sub>5</sub> font l'objet d'une surveillance, permettant de mesurer en continu le débit et d'estimer la charge polluante (MES, DCO) déversée par temps de pluie ou par temps sec*".

Les stations de traitement des eaux usées sont également soumises à l'autosurveillance. L'article 19 de l'arrêté précise qu'un programme de surveillance doit être mis en place comprenant au minimum les paramètres suivants : pH, débit, DBO<sub>5</sub>, DCO, MES (cas le moins restrictif des stations recevant moins de 120 kg/j de DBO<sub>5</sub>).

La connaissance des flux de pollution déversés dans le milieu récepteur au niveau de chaque point de rejet a pour objectif d'évaluer les principales sources de pollution pour lesquelles des actions correctives devront être envisagées. Ces actions concernent aussi bien l'urbanisme du bassin versant (augmentation de la part des eaux infiltrées), que la modification de la gestion des eaux pluviales au sein du réseau d'assainissement, des ouvrages de stockage et de la station de traitement des eaux usées.

#### 1.2.1.2 Insuffisances de l'autosurveillance au niveau de l'estimation des flux de pollution

Selon le commentaire technique de l'arrêté du 22 juin 2007, l'autosurveillance "*s'inscrit dans une démarche qualité visant :*

- 
- pour l'exploitant<sup>4</sup>, à vérifier, en continu, l'adéquation entre les objectifs fixés<sup>5</sup> et les résultats obtenus ;
  - pour le service de police de l'eau et les agences de l'eau, à limiter les contrôles directs ;
  - à disposer de données fiables sur le fonctionnement des systèmes d'assainissement.

Or, la vérification en continu de l'adéquation entre les objectifs fixés et les résultats obtenus n'est pas réalisable pour plusieurs raisons. Tout d'abord, la pollution des rejets urbains de temps de pluie est sujette à une extrême variabilité : variabilité entre bassins versants, variabilité entre événements pluvieux et variabilité au sein même d'un événement pluvieux (Valiron et Tabuchi, 1992; Desbordes *et al.*, 1994; Chebbo *et al.*, 1995; Gromaire *et al.*, 2001).

Ainsi, il a été montré qu'à l'échelle annuelle, l'évaluation des flux de pollution à  $\pm 25$  % au niveau d'un point de rejet nécessite l'échantillonnage d'une trentaine d'échantillons. Or le nombre de campagnes de prélèvements dépasse rarement le nombre de cinq (Mourad *et al.*, 2005).

De plus, l'état du milieu récepteur est évalué à une fréquence moins élevée que les rejets de pollution, ce qui réduit les possibilités de lier les impacts aux sources de pollution. Par exemple, à Strasbourg, les mesures physico-chimiques sur les eaux de l'Ill sont réalisées une fois par mois alors que les rejets des déversoirs d'orage sont suivis en continu. Ceci est d'autant plus préjudiciable que les rejets urbains de temps de pluie sont caractérisés par un certain nombre d'effets immédiats ou chroniques (Brelot et Chocat, 1996) tels que :

- la désoxygénation, conséquence du rejet de matières organiques en quantité importante et dont la dégradation par les micro-organismes est consommatrice d'oxygène ;
- la toxicité de certains paramètres de pollution comme la forme non ionisée de l'azote ammoniacal.

---

<sup>4</sup>Selon la définition de l'approche intégrée donnée au paragraphe précédent, les données d'autosurveillance devraient guider l'action de l'ensemble des acteurs de l'aménagement de l'espace urbain, et pas seulement celle des techniciens.

<sup>5</sup>Selon ce même document, l'autosurveillance a pour objectif de vérifier que l'ensemble réseau de collecte-station de traitement des eaux usées est adapté au milieu récepteur et permet d'atteindre les objectifs de qualité.

---

### 1.2.1.3 Quelles sont les évolutions nécessaires ?

Le lien impact-source de pollution étant difficile à établir, les actions visant à réduire l'impact du système d'assainissement sur le milieu récepteur ne peuvent être prises que sur la base d'une évaluation fiable des flux de pollution déversés. Les méthodes optiques de mesure in-situ et en continu de la pollution (turbidimétrie et spectrophotométrie UV/Visible) constituent une alternative crédible aux méthodes conventionnelles avec des réductions importantes de l'incertitude sur l'estimation des flux de pollution aussi bien à l'échelle événementielle qu'annuelle (Lacour *et al.*, 2010; Langergraber *et al.*, 2003; Rieger *et al.*, 2006). Comme le souligne Barraud *et al.* (2008), l'échantillonnage traditionnel et les analyses normalisées en laboratoire ne permettent pas d'obtenir le même niveau d'information pour un coût acceptable. Toutefois, l'applicabilité de ces outils mérite encore d'être étudiée, notamment au regard de la variabilité des caractéristiques de la pollution en système d'assainissement.

De plus, jusqu'à maintenant, l'attention s'est portée sur la mise en place - souvent non coordonnée - des différentes stratégies de lutte contre la pollution des milieux récepteurs. Sur les agglomérations où toutes les stratégies ont été déployées, les gains sont dorénavant à trouver dans la mise en œuvre de stratégies de contrôle intégré en temps réel du système d'assainissement dont la pertinence ne pourra être évaluée que par une mesure en continu des flux de pollution déversés et une modélisation intégrée du système.

La modélisation intégrée permet de simuler sans risques environnementaux de nouvelles stratégies de contrôle intégré en temps réel du système. De nombreuses études ont déjà porté sur la modélisation intégrée de systèmes d'assainissement qui requiert le couplage de modèles "bassin versant - réseau de collecte - station de traitement" (Meirlaen *et al.*, 2002; Rauch *et al.*, 2002; Vanrolleghem *et al.*, 2005; Schutze *et al.*, 1999; Lau *et al.*, 2002). Le couplage de ces modèles est complexe du fait de la différence des variables, pas de temps de calcul, types d'équations utilisées dans chacun des modèles. Ainsi, des simplifications dans la représentation du système sont effectuées, notamment sur les parties où les influences aval sont limitées et où les possibilités de régulation sont réduites (Vanrolleghem *et al.*, 2005).

Dans l'objectif d'un contrôle intégré en temps-réel du système d'assainissement, la mesure en continu des flux de pollution a deux vocations principales. La première est de fournir des données fiables pour le calage et la validation des modèles intégrés. Une fois les stratégies de

---

contrôle intégrées opérationnelles, nous pouvons alors envisager d'utiliser ces mesures pour la mise en œuvre d'actionneurs susceptibles de modifier l'écoulement des flux.

## **1.2.2 Contexte spécifique aux zones de baignade**

Une partie de la thèse porte sur la gestion de rejets d'assainissement d'une commune du littoral où les enjeux touristiques et sanitaires liés à la baignade se superposent aux enjeux environnementaux.

### **1.2.2.1 Seuils de qualité microbiologiques réglementaires applicables aux eaux de baignade**

#### **1.2.2.1.a Directive n° 76/160**

Les seuils de qualité des eaux de baignade sont définis par la directive européenne n° 76/160 du 08/12/1975 transposée dans les articles L.1332-1 et suivants et D.1332-1 et suivants du Code de la Santé publique relatifs aux baignades et aux piscines.

Pour évaluer la qualité des eaux de baignade, cette directive prend en compte dix neuf paramètres dont quatre paramètres microbiologiques. Deux niveaux de qualité sont définis : la norme guide et la norme impérative. La norme guide correspond aux valeurs que le gestionnaire du site de baignade doit s'efforcer de respecter. La norme impérative correspond à la valeur qui ne peut être dépassée que dans 5% des cas. Les valeurs des normes guide et impérative sont données dans le TAB. 1.1.

#### **1.2.2.1.b Directive n° 2006/7/CE du 15/02/2006**

Une nouvelle directive sur les eaux de baignade a été adoptée le 15/02/2006 (directive n° 2006/7/CE). Cette directive remplacera progressivement la directive de 1975 jusqu'à son abrogation complète le 31 décembre 2014.

La transposition en droit français de cette directive a été effectuée avec la loi sur l'eau et les milieux aquatiques du 30 décembre 2006. Les dispositions réglementaires sont fixées par le

Paramètres	Norme guide	Norme Impérative <sup>1</sup>	Fréq. d'échantillonnage min.
Coliformes totaux /100 ml	500	10 000	bimensuelle
Coliformes fécaux /100 ml	100	2 000	bimensuelle
Streptocoques fécaux /100 ml	100	-	2
Salmonelles /1 l	-	0	2
Enterovirus PFU/ 10 l	-	0	2

(1) : percentile 95 %

(2) : Teneur à vérifier par les autorités compétentes lorsqu'une enquête effectuée dans la zone de baignade en révèle la présence possible ou une détérioration de la qualité des eaux.

Tab. 1.1 – Norme guide et norme impérative définies par la directive n° 76/160 pour les paramètres microbiologiques

décret 2008-990 du 18 septembre 2008 ainsi que par les deux arrêtés du 22 et 23 septembre 2008.

La nouvelle directive ne comprend que deux paramètres de pollution microbologique : *Escherichia coli* (EC) et Entérocoques intestinaux (EI). Les eaux de baignade sont classées en quatre classes de qualité : excellente, bonne, suffisante et insuffisante.

Paramètre	Type d'eau	Excellente qualité	Bonne qualité <sup>1</sup>	Qualité Suffisante <sup>2</sup>
EI (nb/100 ml)	Eau douce	200	400	330
	Eau de mer	100	200	185
EC (nb/100 ml)	Eau douce	500	1000	900
	Eau de mer	250	500	500

(1) : percentile 95 % (2) : percentile 90 %

Tab. 1.2 – Classes de qualité définies par la directive n° 2006/7/CE

---

### 1.2.2.2 Enjeu de détection de pollutions à court terme

La nouvelle directive sur les eaux de baignade incite les collectivités (gestionnaires de plages) à détecter de manière précoce les pollutions à court terme<sup>6</sup>.

Un classement du site de baignade est établi sur la base des résultats de dénombrement de l'ensemble des prélèvements effectués sur le site et prévus dans le calendrier de surveillance pour les quatre années précédentes. Les collectivités sont autorisées à écarter de la base de données utilisée pour le classement du site de baignade, les échantillons résultant d'une pollution à court terme, dans la limite de 15%, à condition que des mesures de gestion adéquates aient été prises (telles que l'information des baigneurs).

La détection précoce des pollutions à court terme revêt donc un enjeu capital, tant en terme de protection sanitaire des baigneurs, qu'en terme de classement du site de baignade. En effet, ne pas retenir les échantillons résultant d'une pollution à court terme dans la base de données servant au calcul du classement revient à exclure les valeurs les plus élevées (AFFSET, 2007b,a).

Toutefois, la directive ne précise pas les modalités de détection possible d'une pollution à court terme. Une pollution à court terme peut donc être considérée soit comme une pollution constatée au vu des résultats d'un prélèvement, soit comme une pollution prévisible au regard d'évènements météorologiques ou autres.

Le dénombrement de EC ou EI avec les méthodes normalisées actuelles ne permettent d'obtenir un résultat de prélèvement que 48h après le prélèvement. Il ne serait pas sérieux de fonder la mise en place d'une procédure d'alerte précoce dans ces conditions étant donné la forte variabilité temporelle des concentrations de EC et de EI dans les eaux de baignade. En effet, d'après une étude menée par Wymer *et al.* (2005), les concentrations mesurées diffèrent d'un facteur deux entre deux journées consécutives et il n'existe plus de corrélation significative entre les concentrations mesurées à trois-quatre jours d'intervalle.

Les besoins en terme d'outils de détection et de prédiction de flux de pollution bactériologiques rejetés au niveau des zones de baignade sont donc primordiaux afin de répondre aux exigences

---

<sup>6</sup>Selon la directive n°2007/7/CE, une pollution à court terme est "une contamination microbiologique (...) qui a des causes clairement identifiables, qui ne devrait normalement pas affecter la qualité des eaux de baignade pendant plus de soixante-douze heures environ à partir du moment où la qualité de ces eaux a commencé à être affectée et pour laquelle l'autorité compétente a établi des procédures afin de prévenir et de gérer de telles pollutions à court terme".

---

de la directive. De tels outils permettraient de diminuer l'exposition de baigneurs à des eaux chargées en germes pathogènes. De manière tout à fait secondaire - bien que cela représente un enjeu financier énorme - les collectivités pourraient améliorer le classement de leurs zones de baignade par non prise en compte dans le classement d'échantillons prélevés pendant des pollutions à court terme.

### 1.3 Conclusion

Depuis 1964 et la première loi sur l'eau, une attention toujours plus fine est portée à la préservation des milieux aquatiques. Plusieurs stratégies de lutte contre la pollution ont ainsi été mises en place afin d'enrayer la pollution des milieux récepteurs par les rejets d'eaux résiduares urbaines. Elles ont eu pour but le traitement de la pollution de temps sec, l'interception de la pollution de temps de pluie en réseau unitaire, la construction de réseaux séparatifs, et le développement de techniques alternatives.

La dernière stratégie en date est l'approche intégrée, consacrée par la directive cadre sur l'eau du 23 octobre 2000 (transposée par la loi 21 avril 2004). C'est une approche globale qui prend en compte l'ensemble des objectifs de protection et les différentes sources de pollution afin de coordonner à l'échelle d'une agglomération tout l'éventail possible des stratégies de lutte contre la pollution des milieux récepteurs.

Le succès de cette approche est conditionné par la connaissance fine des flux de pollution rejetés par l'ensemble du système d'assainissement, et par le partage de cette information dans le cadre d'un travail collaboratif de l'ensemble des acteurs de l'aménagement de l'espace urbain.

La mesure des flux de pollution au niveau de chaque point de rejet constitue la pierre angulaire de ce dispositif. Toutefois, la fréquence des analyses prévues par la réglementation ne permet pas de tenir compte des effets immédiats ou chroniques des rejets sur les milieux récepteurs. De plus, la fréquence moins élevée des analyses physico-chimiques réalisées dans les milieux récepteurs réduit la possibilité d'établir un lien entre impact et source de pollution.

De nombreuses études ont montré le potentiel des méthodes optiques de mesure en continu des flux de pollution (turbidimétrie et spectrophotométrie UV/Visible). Cependant des recherches doivent encore être menées afin d'améliorer les méthodes d'étalonnage et de définir

---

leur domaine de validité, au regard de la variabilité de la qualité des eaux usées notamment par temps de pluie.

Par ailleurs, les systèmes d'assainissement des grandes agglomérations étant dynamiques et complexes, les marges de progrès se trouvent désormais au niveau de la modélisation et du contrôle intégré en temps réel de l'ensemble du système. La mise en œuvre de ces outils n'est possible qu'avec une évaluation fiable des flux de pollution à pas de temps très fin au niveau de chaque point de rejet. De plus, les modélisations intégrées sont délicates du fait du couplage des modèles de chaque composante du système et requièrent donc des simplifications dans la représentation de parties du système où les influences aval et les possibilités de régulation sont faibles. L'identification de modèles globaux pluie-débit à partir de données échantillonnées prend ici tout son sens. Elle permet de simuler très simplement les débits rejetés à l'exutoire de bassins-versants urbains et réduit la complexité des modélisations menées à l'échelle de l'ensemble du système d'assainissement.

Pour les communes du littoral, les enjeux sanitaires liés aux eaux de baignade s'ajoutent aux enjeux environnementaux. La dernière directive sur les eaux de baignade incite les collectivités à détecter les pollutions à court terme. Or, les méthodes normalisées de dénombrement des germes indicateurs de contamination fécale ne permettent d'obtenir un résultat que sous 48h et ne sont donc pas adaptées à la mise en place de procédures d'alertes précoces.

Bien qu'aucun lien spécifique n'ait été établi entre spectrophotométrie UV/Visible et concentration en germes ICF, les mesures en continu de pollution par spectrophotométrie UV/Visible et la modélisation pluie-débit pourraient peut-être permettre une détection des pics de pollution bactériologiques rejetés dans le milieu marin.

## Chapitre 2

# Contexte scientifique et objectifs de la thèse

Les thèmes de recherche abordés dans la thèse s'inscrivent également dans le cadre de deux études scientifiques.

### 2.1 Projet GIRAC

La première étude est le projet Gestion Intégrée des Rejets d'Assainissement Côtiers (GIRAC) piloté par Véolia eau sur les communes de Brest et Saint-Malo et qui associe également l'Ifremer, Littoralis composé de NKE, Hocer, Idehesa, et Météo France.

#### 2.1.1 Présentation du projet

Le but de ce projet est la mise en œuvre d'une approche intégrée permettant de prévenir et minimiser la pollution des eaux de baignade par les eaux usées du système d'assainissement, et ce particulièrement par temps de pluie.

Il est motivé par une évolution de la réglementation sur les eaux de baignade (cf. §1.2.2.2) ainsi que par des enjeux touristiques et sanitaires importants. La ville de Saint-Malo voit sa population passer de 50 à 120 000 habitants en période estivale, et la baignade constitue

---

un attrait touristique majeur pour les nombreux vacanciers. Toutefois, l'augmentation de la population génère des contraintes d'exploitation des systèmes d'assainissement. Tout rejet d'eaux usées dans le milieu récepteur doit être évité afin de ne pas générer des risques sanitaires pour les baigneurs.

Les mesures de gestion ne pourront être prises qu'après une évaluation des flux de pollution bactériologique rejetés dans le milieu récepteur. A court terme, le but du projet est donc de quantifier les flux de pollution bactériologique provenant des principales sources ponctuelles (rivières, rejets d'assainissement) et de les injecter dans un modèle hydrodynamique qui simulera en continu la qualité des eaux de baignade.

### **2.1.2 Insertion de la thèse au sein du projet GIRAC**

La thèse a été financée par le Ministère de l'Alimentation, de l'Agriculture et de la Pêche sur la base d'un dossier monté conjointement par le laboratoire Hydraulique Urbaine (IMFS-ENGEES) et Véolia eau.

Selon ce dossier, l'objectif de la thèse était le développement d'un modèle intégré afin de tester différentes stratégies de minimisation des flux de pollution rejetés dans le milieu naturel par l'ensemble du système d'assainissement, poursuivant ainsi le thème de recherche initié par le laboratoire avec les études de Le Roy (2004) et Kuss (2007).

Toutefois, le principal verrou scientifique de ce projet réside dans l'estimation en temps réel des flux de pollution bactériologique rejetés dans le milieu marin. Il a donc été décidé, sur proposition de Véolia, de concentrer notre effort de recherche sur ce thème en réalisant :

- l'estimation des concentrations en germes indicateurs de contamination fécale par spectrophotométrie UV/Visible ;
- et l'estimation des débits rejetés dans le milieu marin par identification d'un modèle pluie-débit.

### **2.1.3 Nature du partenariat avec Véolia eau**

La direction du projet de recherche a été assumée par Véolia. Ceci inclut les choix de site d'étude et de protocoles expérimentaux ainsi que la réalisation de toutes les tâches de terrain (prélèvement d'échantillons, maintenance du site expérimental,...).

---

Les données acquises sur le site d'étude nous ont été confiées dans le but de rechercher :

- la possibilité de détecter des pics de pollution bactériologique par spectrophotométrie UV/Visible ;
- et la possibilité de mise au point d'un modèle pluie-débit-pollution dont les données de sortie pourraient être utilisées comme données d'entrée d'un modèle de dispersion de la pollution en milieu marin.

Les campagnes de mesures réalisées par notre partenaire étaient concentrées sur les saisons de baignade (juin à septembre). Il n'a pas été possible de réaliser des mesures complémentaires, ni de nous investir au delà du cadre défini. Nous avons donc fait le choix de poursuivre les recherches localement, en évaluant l'applicabilité de la spectrophotométrie UV/Visible pour la mesure en continu de en entrée de station de traitement des eaux usées.

## **2.2 Modélisation intégrée bassin d'orage - station de traitement d'eaux usées**

Ces travaux de recherche réalisés à Strasbourg s'inscrivent dans le prolongement d'une étude réalisée en 2007 et 2008 sur la modélisation intégrée du couple bassin d'orage - station de traitement des eaux usées dans l'objectif de réduire les impacts des ERU rejetées dans le milieu récepteur, notamment par temps de pluie (Kuss *et al.*, 2008).

### **2.2.1 Description de l'étude**

L'étude propose une nouvelle méthodologie de dimensionnement de bassin d'orage (BO) dans le cadre de la gestion intégrée de systèmes d'assainissement.

La méthodologie suivie dans l'étude est constituée de huit étapes principales :

1. définition des objectifs et définition du système,
2. choix d'un critère d'évaluation,
3. choix des variables d'optimisation,
4. définition de stratégies de contrôle,
5. construction du modèle intégré,

- 
6. construction d'une chronique de débit et de pollution pour les simulations,
  7. calage du modèle,
  8. test des stratégies de contrôle de type émission.

La performance du système d'assainissement a été évaluée sur la base des flux de pollution rejetés par le système à l'échelle annuelle. Nous avons pour cela au préalable identifié les paramètres physico-chimiques les plus significatifs en terme d'impact sur le milieu récepteur (azote ammoniacal et DCO biodégradable).

Le volume du bassin d'orage et le débit maximum admissible ont été choisis comme variable d'optimisation du système. Pour optimiser le fonctionnement du système, nous avons donc balayé une partie de l'espace de ces deux variables pour identifier le couple (sous) "optimal" permettant de minimiser les flux de pollution rejetés dans le milieu récepteur.

L'optimum "économique" du système est atteint pour un volume de BO de 2520 m<sup>3</sup> avec un débit admissible par la STEU de 300 m<sup>3</sup>/h, ce qui correspond à la situation actuelle (cf. FIG. 2.1). En augmentant le débit admis par la STEU lorsque le bassin d'orage (BO) est plein, et en retardant la vidange du BO, les rejets d'azote ammoniacal sont diminués de 20 % pendant le temps de pluie. A l'échelle annuelle, cette stratégie utilisant un contrôle en temps réel du système permet une diminution du flux d'azote ammoniacal rejeté de 7,5 % par rapport à la situation actuelle. En d'autres termes, pour un volume de BO de 2520 m<sup>3</sup>, cette stratégie est aussi performante qu'une stratégie "statique" avec un volume de BO de 3000 m<sup>3</sup>.

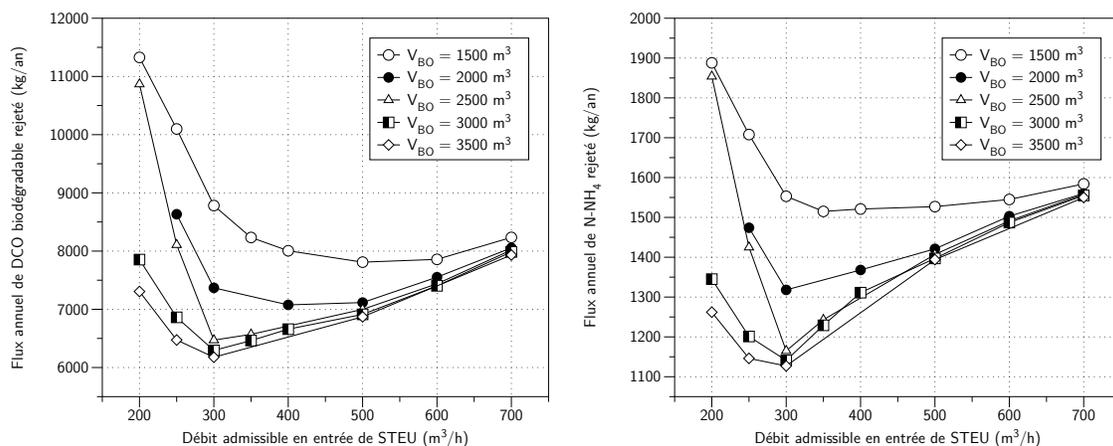


Fig. 2.1 – Flux annuels de DCO biodégradable et N-NH<sub>4</sub> rejetés dans le milieu récepteur pour différentes valeurs du couple volume du bassin d'orage, débit admissible en entrée de STEU

## 2.2.2 Utilité de la mesure de pollution par spectrophotométrie UV/Visible

La validité de l'étude menée en début de thèse est fortement conditionnée par la qualité de la description de la pollution en entrée du modèle. Dans notre cas, nous avons reconstitué une chronique de pollution sur une année civile en nous basant sur des relations entre paramètres hydrauliques et paramètres de pollution établies avec les données d'autosurveillance. Les relations obtenues sont basées sur des moyennes journalières. Il nous a donc fallu dans un second temps répartir la pollution journalière obtenue sur chaque pas de temps de calcul.

La spectrophotométrie UV/Visible offre la possibilité d'acquérir des mesures de pollution à pas de temps très fin sur le long terme et ouvre donc de nouvelles perspectives. Elle permettrait de réaliser la modélisation du couplage bassin d'orage - station de traitement des eaux usées directement à partir de chroniques de pollution mesurées, réduisant ainsi les incertitudes liées à la description de la pollution utilisée comme entrée du modèle.

---

## 2.3 Objectifs de la thèse

La thèse s'insère donc dans un contexte réglementaire et scientifique relatif à la gestion intégrée des systèmes d'assainissement. La figure 2.2 synthétise le concept d'approche intégrée et montre la manière dont les thèmes de recherches traités s'insèrent dans l'ensemble du dispositif.

Les objectifs de la thèse sont de tester l'applicabilité de deux outils dont l'utilisation conjointe pourrait permettre une évaluation en continu des flux de pollution rejetés dans les milieux récepteurs. Le premier est la mesure de concentrations de paramètres de pollution par spectrophotométrie UV/Visible. Le second est la prédiction des débits rejetés par temps de pluie à l'exutoire de bassins-versants urbains grâce à l'identification de modèles pluie-débit à partir de données expérimentales. Les travaux réalisés s'intègrent en partie dans le cadre du projet Gestion Intégrée des Rejets d'Assainissement Côtiers (GIRAC) piloté par Veolia eau dont le but est l'évaluation et la minimisation des flux de pollution bactériologiques rejetés par toutes les composantes du système d'assainissement.

La thèse est structurée en trois parties, chacune d'entre elle correspondant à un objectif particulier.

### Objectif n°1

Le premier objectif porte sur l'évaluation de la spectrophotométrie UV/Visible pour la mesure en continu de concentrations en paramètres globaux de pollution (DCO, MES) en entrée de station de traitement des eaux usées. Cette technique de mesure en continu de pollution en réseau d'assainissement ou au niveau de station de traitement des eaux usées a déjà fait l'objet de nombreuses publications. Nous nous attacherons donc à compléter ces études en portant notre attention notamment sur :

- l'amélioration de la performance des modèles prédictifs<sup>1</sup> de concentrations en DCO et MES par utilisation d'algorithmes de sélection optimisée de longueurs d'onde ;
- la détermination d'un domaine de validité associé à chaque procédure d'étalonnage tenant compte de la variabilité des caractéristiques des eaux usées mesurées in-situ.

---

<sup>1</sup>de type Partial Least Squares

---

### **Objectif n°2**

Par la suite, nous rechercherons si la spectrophotométrie UV/Visible permet de détecter et/ou quantifier des pics de pollution bactériologiques au niveau de milieux récepteurs. Notre étude se basera sur des mesures effectuées par Véolia eau à l'exutoire du bassin versant du Stang Alar (situé à brest) pendant les saisons de baignade 2007, 2008 et 2009.

### **Objectif n°3**

La dernière partie sera consacrée à l'identification d'un modèle pluie-débit sur le bassin versant du Stang Alar afin de simuler en continu les débits des rejets d'assainissement pluviaux déversés dans le milieu marin. Nous testerons une technique d'identification de modèles linéaires à temps continu qui n'a pas encore été testée sur des bassins versants urbains et nous évaluerons les apports de cette technique.

On remarquera enfin qu'en plus d'une unité thématique, la thèse a fait le choix d'une unité sur les outils mathématiques utilisés. En effet, les mesures de concentration par spectrophotométrie UV/Visible et la modélisation pluie-débit font intervenir des modèles linéaires statistiques issus de l'identification de systèmes à partir de données expérimentales.

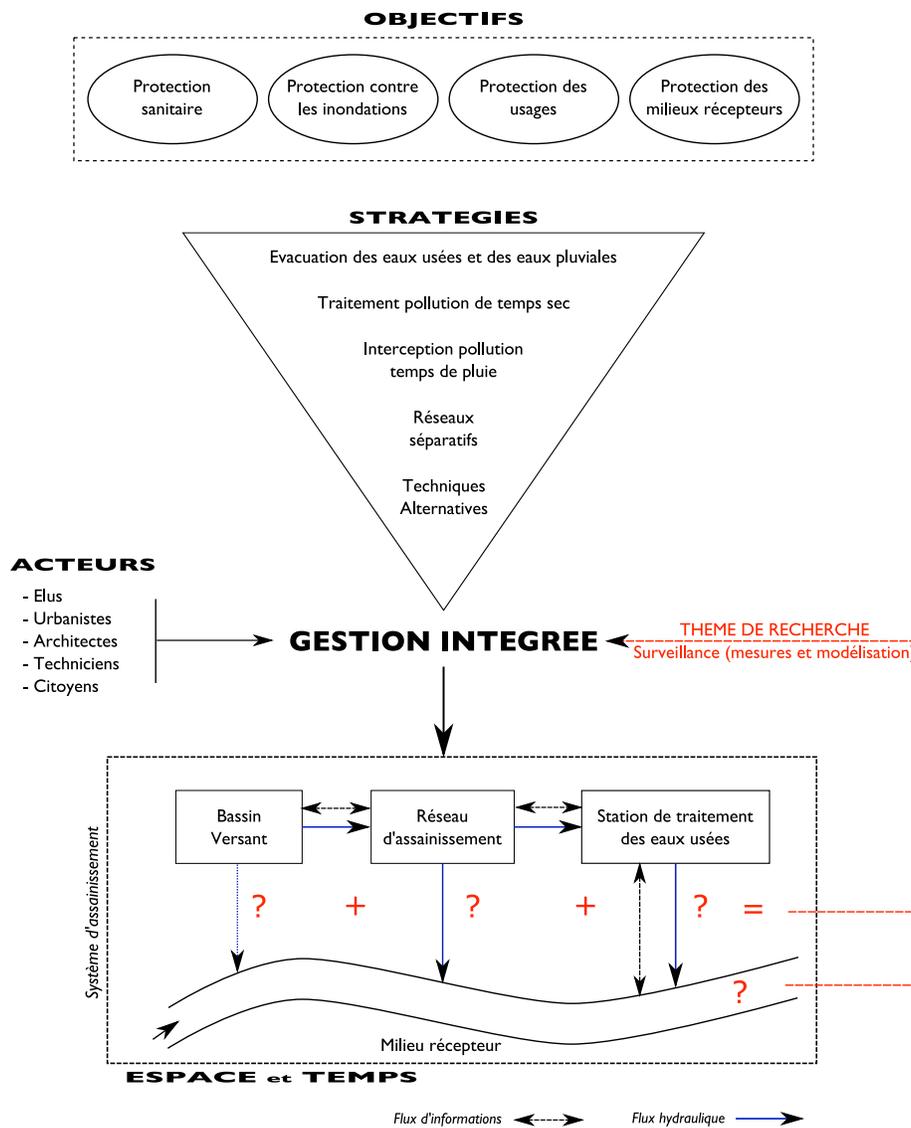


Fig. 2.2 – Intégration des thèmes de recherche au sein d'une gestion intégrée de système d'assainissement

## **Deuxième partie**

# **Mesures en continu de pollution par spectrophotométrie UV/Visible Application à la mesure de concentration en DCO et MES en entrée de station de traitement des eaux usées**

---

Cette partie a pour objectif l'évaluation de la spectrophotométrie UV/Visible comme moyen d'acquisition de mesures en continu de pollution au niveau de station de traitement des eaux usées sur de longues séries temporelles.

Elle est divisée en deux chapitres. Le premier est une synthèse bibliographique dédiée à l'utilisation de la spectrophotométrie UV/Visible pour la mesure de pollution en système d'assainissement. Ce chapitre servira également de fondement aux méthodes utilisées dans la partie suivante concernant la détection de pics de pollution bactériologique.

Le second chapitre porte sur l'amélioration de techniques d'étalonnage et sur la définition de conditions d'utilisation des modèles prédictifs utilisés pour la prédiction des concentrations en DCO et MES.

## Chapitre 3

# Synthèse bibliographique sur l'utilisation de la spectrophotométrie UV/Visible pour la mesure en continu de pollution en système d'assainissement

La structure de ce chapitre est illustrée en FIG. 3.1.

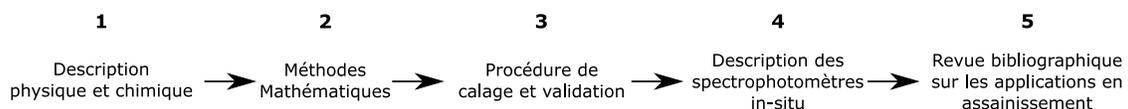


Fig. 3.1 – Structure du chapitre 3

Ce chapitre commencera par aborder la spectrophotométrie UV/Visible sous un angle physique en présentant les phénomènes d'absorption et de diffusion du rayonnement électromagnétique. Les méthodes mathématiques d'analyse multivariable permettant l'établissement de modèles prédictifs de concentrations seront ensuite décrites. Elles seront alors replacées dans le cadre d'une procédure générale de calage et de validation de modèles prédictifs. Le chapitre se

---

terminera par une description succincte des spectrophotomètres utilisables in-situ ainsi que par une revue bibliographique des différentes études portant sur la mesure en continu de paramètres de pollution globaux.

La revue bibliographique permettra d'identifier les apports et limites des études ayant porté sur la mesure en continu de paramètres globaux de pollution en système d'assainissement. Ses conclusions permettront de dégager les thèmes de recherche à approfondir.

### 3.1 La spectrophotométrie UV/Visible

#### 3.1.1 Rayonnement lumineux

Le spectre électromagnétique est le domaine où existent des rayonnements électromagnétiques. Il s'étend sur un domaine très vaste allant des rayons  $\gamma$  à très courte longueur d'onde ( $\lambda=10^{-13}$  m) aux ondes radios ( $\lambda=103$  m).

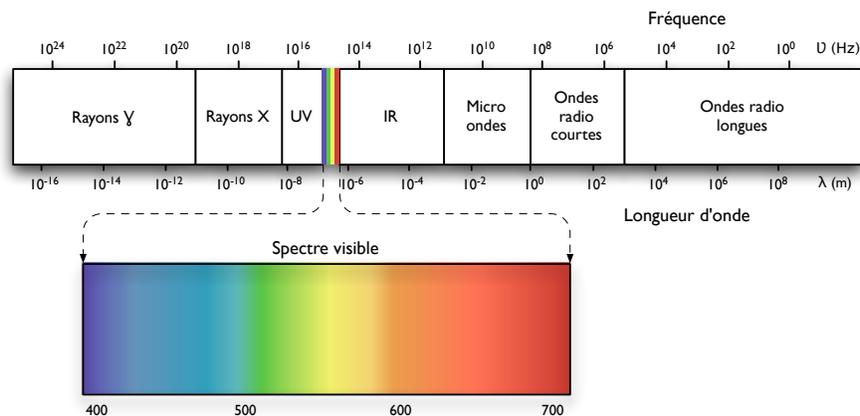


Fig. 3.2 – Spectre électromagnétique

Le domaine visible, de 400 à 800 nm correspond à la seule partie du spectre visible par l'homme. Le domaine ultraviolet correspond à une gamme de longueurs d'ondes légèrement plus courtes allant de 185 à 400 nm.

---

### 3.1.2 Absorption moléculaire

Le principe de la spectrophotométrie UV/Visible repose sur l'absorption du rayonnement UV/Visible par les molécules. En effet, lorsqu'une molécule est soumise à un rayonnement électromagnétique, les électrons absorbent une quantité finie d'énergie pour passer de l'état fondamental à un état excité.

L'excitation d'électrons correspond à des transitions électroniques entre différents niveaux d'énergie (niveaux électroniques, niveaux de vibration et niveaux de rotation). A température ambiante, la plupart des molécules se trouvent dans leur état électronique et vibrationnel fondamental.

Le rayonnement électromagnétique UV/Visible est une onde électromagnétique dont l'énergie  $E$  est reliée à la longueur d'onde par la relation :

$$E = \frac{h c}{\lambda} \quad (3.1)$$

avec :

- $h$  constante de Planck ( $h = 6,63 \cdot 10^{-34}$  J.s) ;
- $c$  célérité du rayonnement dans le milieu où il se propage ( $\text{ms}^{-1}$ ) ;
- $\lambda$  longueur d'onde du rayonnement électromagnétique (nm) ;

D'après la relation (3.1), le domaine de l'UV/Visible correspond à des énergies de 6,5 à 1,5 eV. Les écarts d'énergie entre niveaux électroniques sont du même ordre de grandeur et correspondent donc à des absorptions dans le domaine UV/Visible. En revanche, les écarts d'énergie entre niveaux vibrationnels et rotationnels sont beaucoup plus faibles et conduisent à des absorptions dans le domaine de l'infrarouge moyen et lointain.

Toutefois, lorsqu'une molécule est soumise à un rayonnement dans le domaine de l'UV/Visible, les transitions électroniques, vibrationnelles et rotationnelles se superposent, ce qui explique la complexité des spectres d'absorption. Pour toutes les molécules mises en solution, le solvant vient masquer la structure fine du spectre et celui-ci se présente alors sous la forme de bandes plus ou moins larges.

---

### 3.1.3 Diffusion optique

Lorsque la spectrophotométrie UV/Visible est utilisée sur des eaux naturelles ou des eaux usées contenant des matières colloïdales et des matières en suspension, la diminution de l'intensité du rayonnement lumineux mesurée après traversée de la cellule résulte de la conjonction de la diffusion du rayonnement par les particules et de l'absorption moléculaire du rayonnement par les molécules.

La diffusion se présente sous la forme de réfraction, diffraction ou réflexion. L'intensité et la direction du rayonnement diffusé dépend à la fois de la taille des particules et de la longueur d'onde du rayonnement lumineux. On peut circonscrire trois domaines en utilisant un paramètre adimensionnel  $\alpha$  (Azema *et al.*, 2002) :

$$\alpha = \frac{\pi d}{\lambda} \quad (3.2)$$

où  $d$  est le diamètre des particules et  $\lambda$  la longueur d'onde du rayonnement lumineux.

Lorsque  $\alpha < 0,3$  (ou  $d < \frac{\lambda}{10}$ ), le rayonnement est diffusé dans toutes les directions de l'espace, aussi bien derrière que devant la particule. Nous sommes dans le domaine de la *diffusion de Rayleigh*. La diffusion est proportionnelle à  $(\lambda^{-1})^4$ . Les longueurs d'onde courtes sont donc davantage diffusées.

Si  $\alpha > 30$ , les lois d'optique géométrique et de diffraction s'appliquent. La diffraction du rayonnement a principalement lieu devant la particule.

Enfin, le domaine où  $0,3 < \alpha < 30$  correspond à la limite de validité des modèles évoqués précédemment. La diffusion est influencée par les phénomènes de réfraction, de réflexion et d'absorption.

### 3.1.4 Analyse quantitative : loi de Beer-Lambert

La spectrophotométrie UV/Visible permet de mesurer la concentration de molécules qui absorbent et/ou diffusent à une longueur d'onde donnée. En effet, il existe une relation entre quantité de rayonnement transmis par le milieu et concentration des molécules qui absorbent et/ou diffusent : la loi de Beer-Lambert.

---

Considérons un milieu homogène traversé par un rayonnement lumineux. Selon la loi de Beer-Lambert, la variation d'intensité du rayonnement lumineux dans le milieu est proportionnelle à l'intensité du rayonnement, soit :

$$\frac{dI}{dx} = -\Gamma' I = -(\tau + K) I \quad (3.3)$$

où  $\Gamma'$  est le coefficient linéaire d'extinction ( $\text{m}^{-1}$ ),  $\tau$  le coefficient linéaire de diffusion et  $K$  le coefficient linéaire d'absorption. Par intégration, la relation précédente s'écrit :

$$A_\lambda = \log \left( \frac{I_\lambda^0}{I_\lambda} \right) = \Gamma I \quad (3.4)$$

avec  $A_\lambda$  l'absorbance (sans unité),  $\Gamma = \frac{\Gamma'}{2,303}$ , et  $I$  l'épaisseur du milieu traversé par le rayonnement lumineux. Le coefficient linéaire d'extinction s'exprime souvent en fonction de la concentration des molécules qui absorbent et diffusent le rayonnement :

$$\Gamma = \varepsilon_\lambda c \quad (3.5)$$

avec :

$\varepsilon_\lambda$             coefficient spécifique d'extinction molaire ( $\text{L mol}^{-1} \text{cm}^{-1}$ )  
 $c$                 concentration des molécules ( $\text{mol L}^{-1}$ )

On retrouve alors la forme connue de la relation de Beer-Lambert, à savoir :

$$A_\lambda = \log \left( \frac{I_\lambda^0}{I_\lambda} \right) = \varepsilon_\lambda I c \quad (3.6)$$

avec :

$A_\lambda$             absorbance du milieu à la longueur d'onde  $\lambda$  (sans unité) ;  
 $I_\lambda^0$  et  $I_\lambda$     intensité du rayonnement lumineux avant et après traversée de la cellule ;  
 $\varepsilon_\lambda$             coefficient spécifique d'extinction molaire ( $\text{L mol}^{-1} \text{cm}^{-1}$ ) ;  
 $I$                 trajet optique de la cellule (cm) ;  
 $c$                 concentration des molécules ( $\text{mol L}^{-1}$ ).

La loi de Beer-Lambert s'applique dans les conditions suivantes :

- 
- le milieu traversé par le rayonnement lumineux doit être homogène ;
  - la solution ne doit pas être trop concentrée. Dans le cas contraire, il peut y avoir des interactions soluté-solvant voire soluté-soluté. La loi n'est vérifiée que dans un domaine limité de l'absorbance, entre 0 et 2 ;
  - le rayonnement lumineux doit être monochromatique.

Cette loi constitue la base théorique de tout problème de quantification. Elle est surtout utilisée en laboratoire sur des solutions ne comprenant qu'une espèce absorbant à une longueur d'onde donnée. Le coefficient spécifique d'extinction molaire est déterminé par régression linéaire simple sur une collection d'échantillons pour laquelle on dispose des valeurs conjointes d'absorbance à la longueur d'onde où l'espèce absorbe, et des concentrations de l'espèce chimique déterminées avec la méthode analytique de référence. Une fois ce coefficient calculé, la loi de Beer-Lambert peut-être utilisée pour la prédiction de la concentration de l'espèce considérée dans de nouveaux échantillons.

Lorsque plusieurs espèces absorbent à une longueur d'onde donnée, et qu'elles n'interfèrent pas entre elles, l'absorbance mesurée est égale à la somme des absorbances de chacune des espèces. On recourt alors à des méthodes d'analyses multivariées pour la détermination des concentrations de chaque espèce.

## 3.2 Méthodes d'analyse multivariées

### 3.2.1 Position du problème de quantification

Le but est l'établissement d'un modèle prédictif de concentration d'une ou plusieurs espèces à partir d'une collection d'échantillons servant à l'étalonnage du modèle. Cette collection comprend pour chaque échantillon :

- la mesure du spectre d'absorbance réalisée sur un plage spectrale étendue ;
- et les dosages des différentes espèces pour lesquelles un modèle doit être étalonné.

Les spectres d'absorbance enregistrés peuvent être regroupés dans une matrice **A**, comprenant  $m$  lignes et  $n$  colonnes  $[m \times n]$ . Les lignes de la matrice ( $\mathbf{s}_i^T, i \in \{1, \dots, m\}$ ) correspondent à

une observation (ou spectre) et les colonnes ( $\mathbf{abs}_j, j \in \{1, \dots, n\}$ ) correspondent aux valeurs d'absorbance pour chaque longueur d'onde de la plage spectrale de mesure.

Pour chacune des observations (i.e. pour chacun des échantillons) correspond un spectre d'absorbance et les concentrations de chaque composant analysé en laboratoire. Les concentrations des  $l$  composants quantifiés sont regroupées dans une matrice  $\mathbf{C}$ , de dimension  $[m \times l]$ . L'ensemble des concentrations correspondant à un composé sera notée  $\mathbf{c}_j$ , avec  $j \in \{1, \dots, l\}$ .

Les prochains paragraphes sont dédiés à la description rapide de quelques méthodes d'analyse multivariées qui peuvent être classées en deux catégories : les méthodes statistiques et les méthodes déterministes (cf. FIG 3.3). Nous détaillerons principalement les méthodes statistiques. Pour une description complète de ces méthodes, le lecteur pourra se référer aux articles de Haaland et Thomas (1988), Bertrand (2005), Phatak et De Jong (1997) ainsi qu'aux ouvrages de Thomas (2007) et (Varmuza et Filzmoser, 2009).

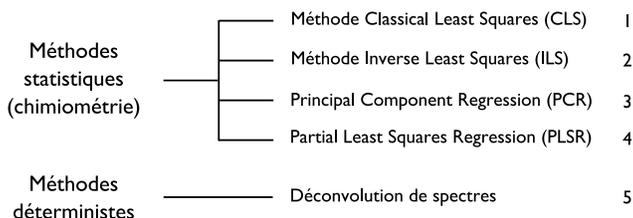


Fig. 3.3 – Méthodes mathématiques d'étalonnage de modèles prédictifs de concentrations

## 3.2.2 Méthodes statistiques

### 3.2.2.1 Méthode CLS (Classical Least Squares)

La méthode CLS est une extension multivariable de la loi de Beer-Lambert. Elle considère que chaque spectre d'absorbance est égal à la combinaison linéaire de chaque "composant pur" présent dans la solution. En notation matricielle, le modèle à résoudre s'écrit :

$$\mathbf{A} = \mathbf{CK} + \mathbf{E}_A \quad (3.7)$$

Dans cette équation  $\mathbf{E}_A$  est l'erreur réalisée sur la mesure de l'absorbance. Notons donc que l'erreur du modèle est supposée n'être due qu'à la mesure de l'absorbance, ce qui n'est

---

pas réaliste.  $\mathbf{K}$  est la matrice regroupant les spectres d'absorbance des composants purs à concentration et parcours optique unitaire  $[l \times n]$ . Cette méthode s'applique uniquement si tous les composants purs de la solution sont connus, ce qui n'est pas le cas d'une eau usée ou d'une eau naturelle.

### 3.2.2.2 Méthode ILS (Inverse Least Squares)

Dans la méthode ILS, la concentration d'un composé est considérée comme une fonction de l'absorbance. Il s'agit donc d'une formulation inverse par rapport à la méthode CLS. Le modèle associé s'écrit :

$$\mathbf{C} = \mathbf{A}\mathbf{P} + \mathbf{E}_C \quad (3.8)$$

Où  $\mathbf{P}$  est une matrice de coefficients inconnus. Dans ce modèle les erreurs ( $\mathbf{E}_C$ ) sont supposées n'être dues qu'aux mesures de concentrations. Ce modèle apparaît donc plus réaliste que le modèle CLS puisque les erreurs sont davantage imputables aux mesures analytiques de concentration qu'aux mesures de spectres d'absorbance.

Par ailleurs, le principal avantage de la méthode ILS est qu'elle peut s'appliquer même si on ne connaît pas la composition chimique de la solution.

Si nous supposons que les erreurs de mesure sur les concentrations de chaque composant sont indépendantes, alors le modèle général décrit par l'équation (3.8) peut s'écrire pour chaque composant :

$$\mathbf{c}_j = \mathbf{A}\mathbf{p}_j + \mathbf{e}_{c_j} \quad (3.9)$$

où  $\mathbf{c}_j [m \times 1]$  est le vecteur colonne de concentrations comprenant toutes les observations pour un même composant ;  $\mathbf{p}_j [n \times 1]$  est le vecteur de régression inconnu. L'équation (3.9) illustre que la méthode peut-être appliquée même si tous les composants ne sont pas connus, ce qui est le cas pour des eaux usées ou des eaux de rivière.

Les coefficients du vecteur de régression  $\mathbf{p}_j$  sont obtenus par résolution de l'équation (3.9) par moindres carrés :

$$\hat{\mathbf{p}}_j = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{c}_j \quad (3.10)$$

---

La prédiction de la concentration pour de nouvelles observations s'effectue en remplaçant  $\mathbf{p}_j$  par son estimation  $\hat{\mathbf{p}}_j$  dans l'équation (3.9) :

$$\hat{\mathbf{c}}_j = \mathbf{A}\hat{\mathbf{p}}_j \quad (3.11)$$

Malheureusement cette méthode ne peut être appliquée que sur un nombre très restreint de longueurs d'ondes. En effet, cette méthode requiert l'inversion de la matrice  $(\mathbf{A}^T\mathbf{A})$  de dimension  $[n \times n]$  et l'inversion n'est possible que si  $n \leq m$ , où  $m$  est le nombre d'échantillons disponibles pour l'étalonnage. Comme dans la plupart des cas le nombre d'échantillons est bien inférieur au nombre de longueurs d'onde, cette méthode ne peut être appliquée dans l'état sur l'intégralité de la matrice  $\mathbf{A}$ . Une sélection des longueurs d'ondes les plus pertinentes pour la prédiction du composant considéré s'avère alors nécessaire.

Par ailleurs, lorsque le nombre de longueurs d'onde disponible est grand, ce qui est toujours le cas avec la résolution actuelle des appareils de mesure, un problème de colinéarité des colonnes de  $\mathbf{A}$  va rendre difficile l'inversion de la matrice  $(\mathbf{A}^T\mathbf{A})$  et la solution instable. La colinéarité signifie que des colonnes de  $\mathbf{A}$  sont le résultat de combinaisons linéaires d'autres colonnes de  $\mathbf{A}$ . Ces colonnes n'apportent donc aucune information supplémentaire quant à la prédiction de  $\mathbf{c}$ .

Différents types de méthodes ont été développés pour s'affranchir des problèmes listés précédemment. Leur but est d'utiliser une nouvelle matrice d'observations spectrales dérivée de  $\mathbf{A}$  et dont le rang est inférieur ou égal au nombre d'échantillons. Cette réduction de dimension peut être réalisée de deux manières différentes :

- par sélection des longueurs d'ondes de  $\mathbf{A}$  les plus pertinentes ;
- par projection du nuage d'observations dans un sous espace vectoriel de dimension inférieure ou égale à  $m$ .

Ces deux stratégies peuvent également être combinées et s'avèrent très performantes (Leardi et Norgaard, 2004; Norgaard *et al.*, 2000; Abrahamsson *et al.*, 2003; Hóskuldsson, 2001; Jiang *et al.*, 2002; Zou *et al.*, 2007). Intéressons nous prioritairement aux méthodes utilisant des projections.

---

### 3.2.2.3 Régression PCR et PLS

Les régressions en composantes principales (PCR) et les régressions des moindres carrés partiels (PLSR) sont deux méthodes voisines qui utilisent l'information contenue dans l'intégralité du spectre d'absorbance. Dans ces méthodes factorielles, la matrice contenant les  $m$  spectres d'absorbance  $\mathbf{A}$  est décomposée en un produit de deux matrices plus petites  $\mathbf{T}$  [ $m \times h$ ] et  $\mathbf{B}$  [ $h \times n$ ] :

$$\mathbf{A} = \mathbf{T}\mathbf{B} + \mathbf{E}_A \quad (3.12)$$

L'équation (3.12) montre un changement de système de coordonnées avec  $\mathbf{B}$  la matrice contenant les vecteurs de la nouvelle base, et  $\mathbf{T}$  la matrice contenant les coordonnées (intensités ou scores) dans le nouveau système. Le but de ces méthodes est d'effectuer une projection d'un nuage de points de dimension  $n$  (un spectre) dans un sous espace vectoriel de dimension  $h$ , avec  $h < m$ . On utilise alors la projection du nuage de points, i.e.  $\mathbf{T}$  pour effectuer la régression. L'équation à résoudre devient alors :

$$\mathbf{c} = \mathbf{T}\mathbf{q} + \mathbf{e}_c \quad (3.13)$$

où  $\mathbf{q}$  est le vecteur de régression contenant les coefficients reliant les intensités aux concentrations. Dans les méthodes PCR et PLS, les colonnes de  $\mathbf{T}$  sont orthogonales et on a également  $h < n$ . Par conséquent, l'inversion de la matrice  $(\mathbf{T}^T\mathbf{T})$  devient triviale. Notons que la régression effectuée avec  $\mathbf{T}$  permet de supprimer les problèmes d'inversion de matrice liés à la colinéarité des colonnes de  $\mathbf{A}$  et au nombre réduit d'observations par rapport au nombre de longueurs d'ondes ( $m < n$ ).

Dans la méthode PCR le calcul de  $\mathbf{T}$  est indépendant des concentrations et est réalisé par une analyse en composantes principales. La régression est ensuite réalisée à partir des scores et non des absorbances. Cette seconde étape est donc équivalente à une régression de type ILS.

Dans la régression PLS, le calcul des composantes  $\mathbf{T}$  se fait en tenant compte de la concentration à prédire. Les équations (3.12) et (3.13) sont donc résolues simultanément.

---

### 3.2.3 Méthode déterministe : déconvolution de spectre

Les approches exposées dans les paragraphes précédents font intervenir des méthodes statistiques permettant la résolution d'un problème de quantification chimique. Les projections réalisées dans les approches PCR et PLS aboutissent au calcul d'une base de vecteurs (**B**) ayant la dimension des spectres d'absorbance mais n'ayant aucune signification chimique.

La méthode de déconvolution de spectre adopte quant à elle une approche déterministe. Elle consiste à exprimer un spectre mesuré comme la somme de plusieurs spectres de référence correspondant à des paramètres de pollution dont les quantités sont connues. Cette méthode s'approche donc de la régression CLS à la différence près que la matrice **K** de spectres de référence est fixée à l'avance. En notation matricielle, cette approche vise à établir pour chaque mesure de spectre  $\mathbf{s}^T$  [1x $\eta$ ] la relation :

$$\mathbf{s}^T = \mathbf{K}^T \mathbf{b} + \mathbf{e}_a \quad (3.14)$$

où **b** est le vecteur de régression. Cette méthode, utilisée par Thomas (2007) pour la détermination de paramètres de pollution dans les eaux usées et les eaux naturelles, considère que tout spectre est une combinaison linéaire de spectres de référence.

Pour établir la collection de spectres de référence, il est nécessaire de disposer d'une base de données très importante de spectres adaptés aux suspensions étudiées. Les spectres retenus sont classés en deux groupes :

- Un premier groupe de spectres correspondant à des composés spécifiques. Il comprend un spectre de référence pour les nitrates et les détergents. Ces spectres peuvent directement être établis par des mesures spectroscopiques de ces composés isolés.
- Le second groupe de spectres comprend un spectre de référence pour les MES, les composés solubles ainsi que les colloïdes. Ces spectres ne peuvent être obtenus directement et sont établis de façon expérimentale ou mathématique (par exemple en retranchant le spectre d'un échantillon filtré à  $1\mu m$  au spectre brut pour les MES)

A chaque spectre de référence ainsi obtenu est associée une concentration. Les concentrations des spectres de référence sont regroupées dans un vecteur colonne  $\mathbf{c}_{ref}$ .

Les concentrations du paramètre de pollution souhaité sont alors obtenus par la relation suivante :

$$\mathbf{c} = \mathbf{b}^T \mathbf{c}_{ref} \quad (3.15)$$

---

La déconvolution des spectres peut être appliquée pour tout type d'eau (eau de mer, eau souterraine) sous réserve d'un choix cohérent des spectres de référence à utiliser. Elle donne d'excellents résultats mais nécessite une base de spectres assez importante avant de pouvoir être opérationnelle (Thomas, 2007).

### 3.3 Procédure de calage et de validation d'un modèle prédictif de concentrations

La procédure de calage et de validation d'un modèle prédictif de concentrations comporte cinq étapes principales représentées sur la FIG 3.4.

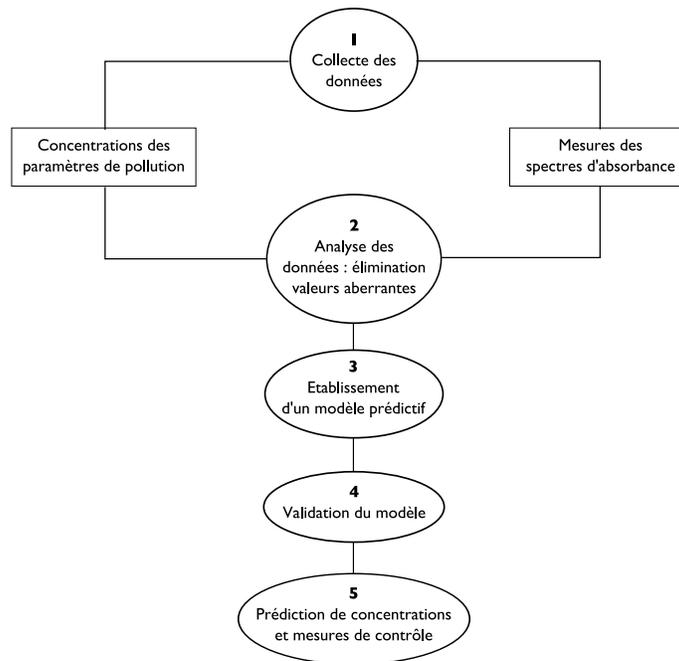


Fig. 3.4 – Étapes de développement d'un modèle prédictif de concentrations (inspiré de Bertrand (2005))

---

### **3.3.1 Collecte des données d'étalonnage (Étape 1)**

Cette première étape consiste à constituer une base de données d'échantillons qui seront utilisés pour l'étalonnage et la validation du modèle prédictif de concentrations. Ce jeu de données comprend pour chaque échantillon :

- les concentrations des paramètres de pollution pour lesquels on cherche à établir un modèle prédictif. Les concentrations sont déterminées à partir de méthodes de référence ;
- les spectres d'absorbance correspondant strictement au même échantillon.

### **3.3.2 Analyse des données (Étape 2)**

La seconde étape consiste à analyser la qualité des mesures de concentrations et de spectres d'absorbance afin d'éliminer d'éventuelles valeurs aberrantes.

### **3.3.3 Étalonnage et validation d'un modèle prédictif (Étape 3 et 4)**

Avant de commencer la procédure de calage et validation, il convient en premier lieu de choisir un modèle mathématique parmi ceux décrits au §3.2. Nous ne nous intéresserons ici qu'à l'étalonnage et à la validation de modèles statistiques.

L'étalonnage et la validation du modèle prédictif peuvent être réalisés à partir de deux jeux de données indépendants (validation simple). Cette procédure n'est pas la plus performante. En effet, elle conduit à perdre de l'information, puisque seulement une partie des échantillons est utilisée pour l'étalonnage du modèle prédictif. On recourt de préférence à la procédure de validation croisée dans un double objectif (Bertrand, 2005; Martens et Dardenne, 1998) :

1. valider le modèle ;
2. choisir la dimension optimale du modèle, *i.e.* le nombre adéquat de composantes.

#### **3.3.3.1 Description de la procédure de validation croisée**

Lorsque qu'une validation croisée est effectuée, l'étalonnage du modèle prédictif est réalisé sur l'intégralité des données de la collection d'étalonnage.

La procédure de validation croisée consiste à recommencer plusieurs fois la validation du modèle. La FIG. 3.5 illustre la manière dont est conduite cette procédure. Le jeu de données est tout d'abord divisé en  $p$  partitions. Les partitions peuvent être générées de manière aléatoire, ou de façon chronologique. On forme alors toutes les combinaisons possibles de  $p-1$  partitions dans l'ensemble des  $p$  partitions. Un modèle prédictif est étalonné pour chacune de ces combinaisons.

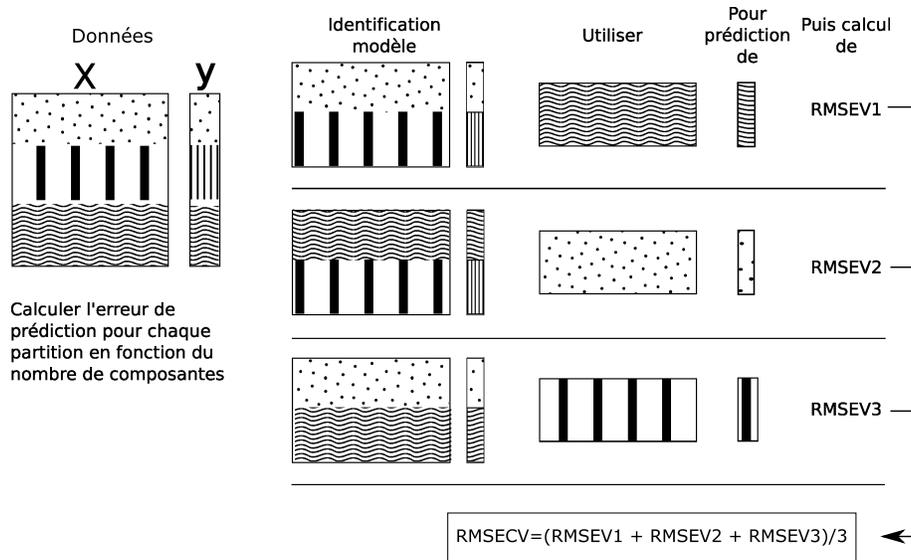


Fig. 3.5 – Procédure de validation croisée

Chacun des  $p$  modèles est alors appliqué sur la partition restante réservée à la validation du modèle. L'erreur de validation RMSEV est calculée pour chacune de ces partitions :

$$RMSEV_i = \sqrt{\frac{\sum_{j=1}^{np_i} (\hat{y}_j - y_j)^2}{np_i}} \quad (3.16)$$

où  $np_i$  est le nombre d'échantillons contenu dans la  $i^{\text{ème}}$  partition non utilisée pour la formation du modèle,  $\hat{y}_j$  est la concentration prédite du  $j^{\text{ème}}$  échantillon de la partition, et  $y_j$  est la concentration mesurée du  $j^{\text{ème}}$  échantillon de la partition.

---

L'erreur de validation croisée est la moyenne de toutes les erreurs de validation obtenues pour chaque combinaison :

$$RMSECV = \frac{\sum_{i=1}^p RMSEV_i}{p} \quad (3.17)$$

De la même manière, l'erreur de calage (ou étalonnage) RMSEC est calculée pour chaque partition incluse dans le modèle, puis moyennée.

### 3.3.3.2 Choix de la dimension du modèle prédictif

Les méthodes de régression multivariable de type PCR ou PLS projettent le nuage de points correspondant aux différentes observations dans un sous-espace vectoriel de dimension réduite. Le choix de la dimension de ce sous-espace vectoriel est un paramètre du modèle qui doit être déterminé. On parle également de détermination du nombre de composantes du modèle.

Plus la dimension du sous-espace vectoriel choisi est grande, moindre est la perte d'information générée par la projection. Cependant, l'introduction d'un nombre trop élevé de composantes conduit à la surparamétrisation du modèle aboutissant à de mauvaises prédictions réalisées sur de nouveaux échantillons.

La fonction RMSEC, fonction du nombre de composantes retenu dans le modèle, est strictement décroissante, chaque composante supplémentaire ajoutant un degré de liberté supplémentaire.

On utilise donc la fonction RMSECV pour évaluer le nombre optimal de composantes. Pour chaque combinaison de  $p-1$  partitions, un modèle est étalonné avec  $1, 2, 3, \dots, n_c$  composantes,  $n_c$  étant strictement inférieur au nombre d'échantillons. La fonction d'erreur RMSECV est alors calculée pour chaque nombre de composantes. Le nombre optimal de composantes est celui qui minimise la fonction RMSECV.

### 3.3.4 Prédictions des concentrations et mesures de contrôle (Étape 5)

Une fois le modèle prédictif calé et validé, il peut être utilisé pour la prédiction de concentrations sur de nouveaux échantillons. Les prédictions devront être effectuées au sein du

---

domaine de validité du modèle prédictif, *i.e.* pour des valeurs de concentrations et de spectres d'absorbance observées pendant l'étalonnage.

La validité de l'étalonnage devra être vérifiée régulièrement au moyen de mesures de contrôle portant sur des échantillons dont les concentrations sont connues ou sur de l'eau distillée afin de contrôler l'obstruction des optiques.

## **3.4 Spectrophotomètres in-situ**

De nombreux spectrophotomètres permettant une mesure in-situ en temps-réel ont été développés ces dernières années. La présentation des appareils qui va suivre n'est pas exhaustive mais montre bien les différences de fonctionnement que nous pouvons rencontrer. Nous parlerons plus en détail du scan spectrolyser utilisé dans le cadre de la thèse.

### **3.4.1 STIP-scan (Endress-Hauser)**

Le STIP-scan un spectrophotomètre UV/Visible dont la plage de mesure spectrale s'étend de 190 à 720 nm. La sonde est prévue pour être directement placée dans le flux liquide. Un piston aspire l'échantillon d'eau à analyser dans une cuve de mesure.

La mesures sont réalisées par cycle de remplissage - mesures spectrophotométriques - vidange qui durent de une à cinq minutes.

Le spectrophotomètre utilise la décantation de l'eau usée au sein de la cellule de mesure pour compenser la part de l'absorbance liée à la diffusion du rayonnement par les particules. Des mesures optiques sont effectuées à intervalles très rapprochés afin de mesurer le spectre de l'eau brute et le spectre de l'eau décantée. En suivant l'évolution de l'absorbance au cours du cycle de mesure, les algorithmes permettent de calculer l'indice de boues lorsque la sonde est placée dans un bassin d'aération.

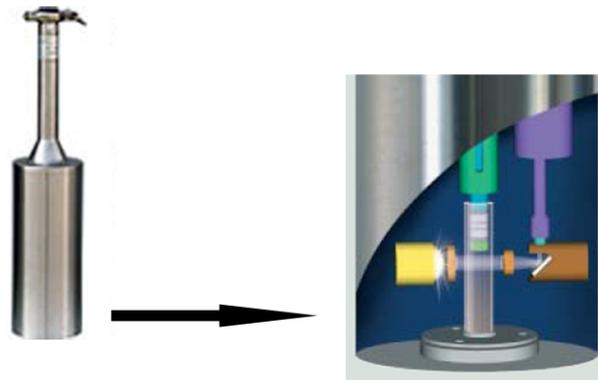


Fig. 3.6 – Spectrophotomètre UV/Visible STIP-scan de ENDRESS-HAUSER

Le STIP-can permet de mesurer en continu la DCO, le COT, les MES et les nitrates en entrée de station de traitement des eaux usées.

Enfin, aucune information n'a été communiquée concernant la nature des algorithmes permettant d'établir la relation entre spectre d'absorbance et concentration des différents paramètres de pollution. Selon la notice technique, il est vraisemblable que des régressions linéaires mono-variables soient utilisées.

### 3.4.2 STAC (Secomam)

La STAC (pour station d'alerte compacte) est un spectrophotomètre qui mesure le spectre d'absorbance de l'échantillon entre 200 et 320 nm. Le spectrophotomètre n'est pas placé dans le flux hydraulique. L'échantillon est prélevé par aspiration. La mesure optique se fait dans une cuve de mesure à alimentation continue.

Un blanc est réalisé avant chaque mesure avec de l'eau déminéralisée. Un nettoyage de la cuve de mesure avec de l'eau chlorée est également possible. Ce système devrait donc être a priori peu sensible à l'encrassement de la cuve de mesure. Toutefois, contrairement aux autres appareils, il génère une consommation d'eau déminéralisée et chlore.

Le calcul des concentrations est réalisé grâce à un algorithme de déconvolution de spectres.

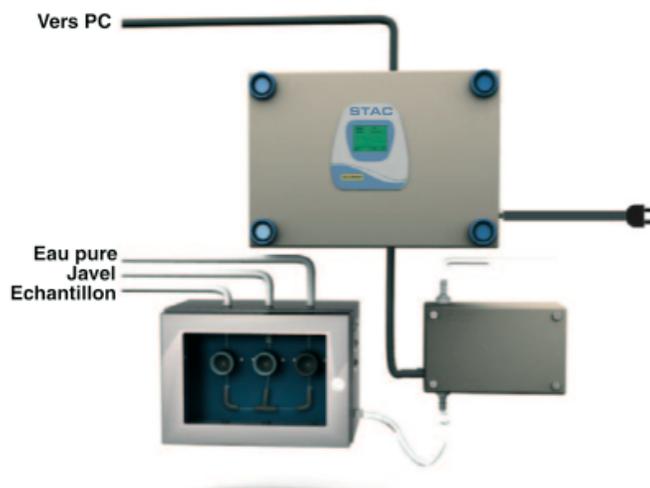


Fig. 3.7 – Station d'alerte compacte (STAC) de SECOMAM (source : Secomam)

### 3.4.3 s::can spectro::lyser (s::can Messtechnik GmbH )

Le spectro::lyser est un spectrophotomètre in-situ fabriqué par s::can Messtechnik GmbH (Autriche). Ce spectrophotomètre mesure l'atténuation de la lumière dans l'UV et le visible dans le domaine [200nm-750nm]. Il s'agit d'un spectrophotomètre à double faisceau dont la source de lumière est une lampe à xénon. La sonde est submersible et est conçue pour être placée directement dans l'effluent. Contrairement aux deux appareils cités précédemment, la mesure n'a pas lieu dans une cellule mais directement dans le liquide sans échantillonnage. La mesure d'un spectre d'absorbance est rapide (environ 15 secondes), ce qui permet un suivi en quasi temps réel de la qualité de l'effluent.



Fig. 3.8 – Spectrophotomètre UV/Visible spectro :lyser de s :can Messtechnik GmbH

---

L'appareil est muni d'un système de nettoyage pneumatique pour éviter l'encrassement de la fenêtre de mesure. Le chemin optique est fixe. Il existe différentes largeurs de chemin optique suivant l'application dans laquelle le spectrophotomètre est utilisé (eau potable, eau usée,...).

Cette sonde permet la mesure des principaux paramètres de pollution qui nous intéressent en entrée de station de traitement des eaux usées, à savoir la DCO, les MES, les  $\text{NO}_3^-$ , le COT. Les concentrations sont calculées grâce à des modèles PLS étalonnés sur une base de données d'échantillons établie par le constructeur (calage global). Cet étalonnage global peut être amélioré par une mesure conjointe sur quelques échantillons de spectres d'absorbance et de concentrations par méthodes normalisées. L'étalonnage local réalisé par ces mesures complémentaires consiste en un ajustement linéaire ou polynomial du calage global réalisé par le constructeur.

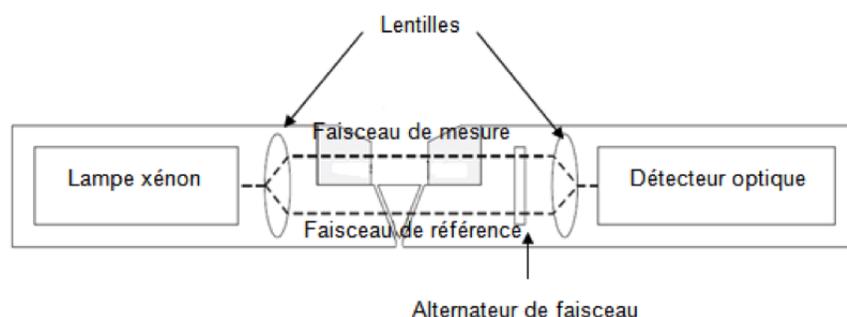


Fig. 3.9 – Schéma de fonctionnement du spectrolyser (source : Langergraber *et al.* (2003))

Un algorithme de compensation de la turbidité permet d'éliminer la part d'atténuation de la lumière liée à la diffusion du rayonnement par les particules et permet donc de calculer plus finement la concentration de paramètres dissous (ex : DCO filtrée).

Par ailleurs, la sonde peut être alimentée en courant alternatif 220V/50Hz ou en courant continu 12-24V. Sa consommation électrique assez faible est idéale pour être utilisée sur le terrain et être alimentée à l'aide de batterie ou par système solaire. Un système d'acquisition de données dans la sonde permet de stocker les spectres mesurés (la sonde peut stocker un mois de mesures à raison d'une mesure toutes les 30 min). La figure 3.9 montre le principe de fonctionnement de la sonde.

---

Notons que les absorbances mesurées avec cet appareil sont ramenées à une unité de longueur fixe permettant ainsi de comparer les valeurs acquises avec des appareils ayant des chemins optiques différents. Elles sont donc exprimées en  $m^{-1}$ .

### **3.5 Applications de la spectrophotométrie UV/Visible pour la quantification de paramètres de pollution physico-chimique dans les eaux usées**

Les prochains paragraphes ont pour objet une synthèse bibliographique des différentes études portant sur la mesure en temps-réel de paramètres de pollution physico-chimiques par spectrophotométrie UV/Visible en système d'assainissement. Nous nous intéresserons :

- aux paramètres de pollution pouvant être quantifiés par spectrophotométrie UV/Visible ;
- aux domaines d'application des études portant sur la quantification de ces paramètres en système d'assainissement ;
- aux collections d'étalonnage acquises et au choix de méthode de validation des modèles (validation simple ou validation croisée) ;
- aux méthodes d'élimination des valeurs aberrantes ;
- aux méthodes mathématiques de régression utilisées et à leurs performances respectives ;
- et enfin aux problématiques de bruit de signal et de performance des modèles prédictifs en temps de pluie.

#### **3.5.1 Paramètres de pollution étudiés**

Les nitrates et de nombreuses molécules organiques dont les groupements fonctionnels possèdent des liaisons chimiques insaturées absorbent le rayonnement électromagnétique dans l'ultraviolet. Ces propriétés rendent donc possible la détermination de relations entre l'absorbance mesurée et la concentration de paramètres de pollution globaux tels que la Demande Chimique en Oxygène (DCO), la Demande Biochimique en Oxygène en cinq jours ( $DBO_5$ ) et le Carbone Organique Total (COT) (Bertrand-Krajewski *et al.*, 2000; Thomas, 2007).

La FIG. 3.10 illustre l'influence respective des matières en suspensions et des composés organiques dissous sur la mesure du spectre d'absorbance d'un échantillon d'eau usée brute.

---

L'expérience menée au laboratoire d'analyses de l'ENGEES a consisté à mesurer le spectre d'absorbance d'un échantillon d'eau usée brute et du même échantillon filtré à différentes porosités : 500, 100, 53, 11, 1,2, 0,45, et 0,2  $\mu\text{m}$ . On considérera par simplification que le filtrat à 0,2  $\mu\text{m}$  ne comprend que des composés dissous. La FIG. 3.10 montre que les composés organiques dissous absorbent le rayonnement électromagnétique dans l'UV, et que leur absorbance au delà de 400 nm est négligeable. Les valeurs d'absorbance observées dans le domaine visible du spectre sont donc imputables aux MES. Les spectres correspondant aux MES peuvent être obtenus par différence des spectres des différents filtrats (cf. FIG. 3.11).

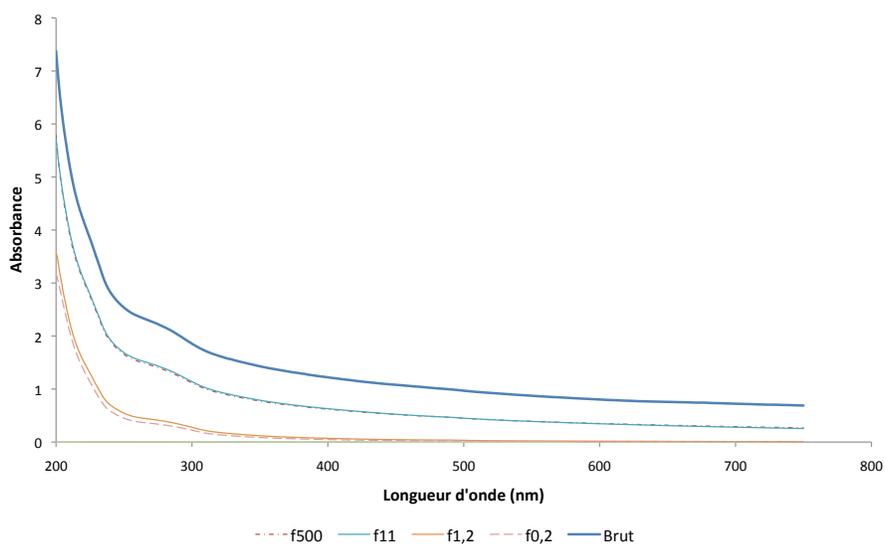


Fig. 3.10 – Spectres d'absorbance d'un échantillon d'eau usée brute et filtrée à différentes porosités

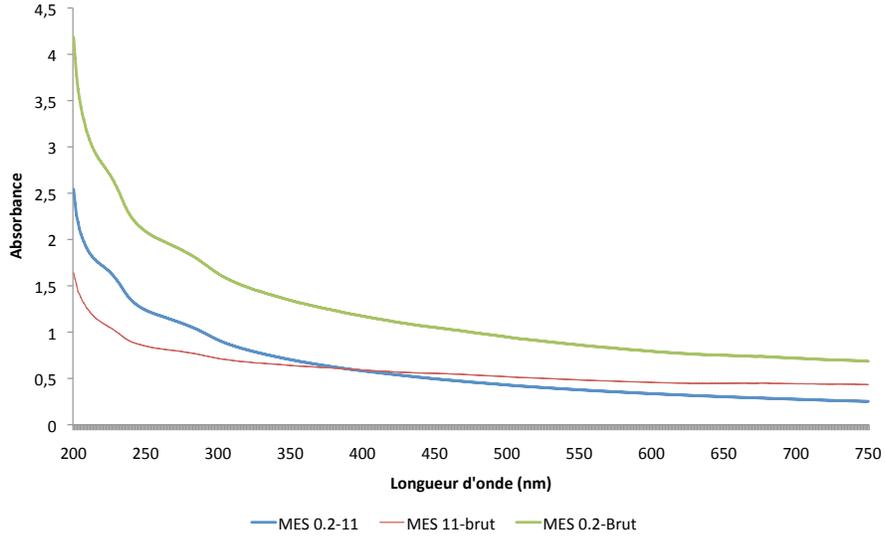


Fig. 3.11 – Spectres de MES

Le TAB. 3.5.1 récapitule les principales études et méthodes utilisées pour la détermination des concentrations de paramètres globaux dans les eaux usées par spectrophotométrie UV/Visible. On constate que toutes premières études visaient à établir des relations avec les paramètres de pollution sur la base d'une ou deux longueurs d'ondes. Le développement conjoint des spectrophotomètres permettant de balayer un ensemble de longueurs d'ondes et des méthodes mathématiques de régression multivariées (PLS notamment) ont permis l'identification de modèles plus complexes et performants.

Nous ne nous intéresserons par la suite qu'aux publications où un spectrophotomètre in-situ a été testé pour la quantification de la DCO et des MES sur des eaux résiduaires brutes (Langergraber *et al.*, 2003; Rieger *et al.*, 2006; Hochedlinger *et al.*, 2006; Torres et Bertrand-Krajewski, 2008; Maribas *et al.*, 2008) avec utilisation de modèles prédictifs statistiques de type PLS.

Le spectrophotomètre utilisé dans toutes ces études est le spectro::lyser. Leur objet est la validation de l'utilisation in-situ de cette sonde en comparant la performance de différentes méthodes de régression.

Méthode	Longueurs d'onde (nm)	Paramètre	Application	Auteurs
RLS	254	COT	Eaux usées domestiques brutes et traitées	Dobbs <i>et al.</i> (1972)
	254	DCO	Eaux usées domestiques et industrielles	Mrkva (1975)
	280	DBO	Effluents agricoles	Brookman (1997)
ILS 2 l.o.	254 et 350	DCO, COT	Eaux usées domestiques	Matsche <i>et Stumwóhrer</i> (1996)
Déconvolution	205-330	COT, DCO, DBO, MES, NO <sub>3</sub> <sup>-</sup>	Eaux usées domestiques et industrielles	Thomas <i>et al.</i> (1996); El Khorasani <i>et al.</i> (1999); Thomas (2007)
PLS	200-730	DCO, MES, NO <sub>3</sub> <sup>-</sup>	Eaux usées domestiques brutes et traitées	Langergraber <i>et al.</i> (2003), Rieger <i>et al.</i> (2004), Hochedlinger <i>et al.</i> (2005), Rieger <i>et al.</i> (2006), Hochedlinger <i>et al.</i> (2006), Torres et Bertrand-Krajewski (2008), Rieger <i>et al.</i> (2008), Maribas <i>et al.</i> (2008)

Tab. 3.1 – Etudes ayant porté sur la mesure de concentrations de paramètres de pollution par spectrophotométrie UV/Visible

---

L'étude de Maribas *et al.* (2008) aborde les difficultés rencontrées pour l'étalonnage de la sonde avec des influents de temps de pluie ainsi que des problèmes de bruit sur le signal de concentration.

### 3.5.2 Domaine d'application

La sonde a été étudiée à différents emplacements du système d'assainissement : déversoir d'orage (Hochedlinger *et al.*, 2006), sortie de décanteur primaire (Langergraber *et al.*, 2003; Maribas *et al.*, 2008), entrée de bassin de décantation (Torres et Bertrand-Krajewski, 2008), et entrée de station de traitement des eaux usées (Rieger *et al.*, 2006; Maribas *et al.*, 2008) (cf. TAB. 3.5.2).

La publication de Rieger *et al.* (2006) porte sur les influents de six stations de traitement des eaux usées suisses mais ne précise pas l'emplacement exact d'installation du spectrophotomètre au sein de la station de traitement des eaux usées. Par ailleurs, dans l'étude menée par Maribas *et al.* (2008), la sonde a été évaluée sur trois emplacements différents :

- dans une cuve alimentée par les effluents du décanteur primaire (lieu où l'étalonnage a été réalisé) ;
- dans le canal d'alimentation du dessableur (canal aéré) ;
- et enfin dans le décanteur primaire.

La plupart de ces articles décrivent en détail les méthodes d'étalonnage testées sans préciser la nature des effluents (domestiques, industriels, agro-alimentaires,...), le type de réseau d'assainissement, les conditions environnementales pendant lesquelles les échantillons ont été prélevés (temps sec, temps de pluie, pointe journalière,...) voire l'intervalle des concentrations mesurées.

L'intervalle de variation des concentrations mesurées est dans la plupart des cas relativement faible. Ceci peut-être lié à plusieurs facteurs. Tout d'abord, la décantation primaire réalise un abattement de la pollution particulaire, et tend par conséquent à limiter l'intervalle de variation des concentrations mesurées. De plus lorsque les mesures sont réalisées en réseau pluvial, les concentrations en DCO et MES sont plus faibles, surtout en temps sec où subsistent uniquement les eaux claires parasites. Enfin, le troisième facteur explicatif est le prélèvement d'échantillons uniquement pendant des périodes de temps sec.

Étude	Emplacement	Variation DCO (mg/l)	Variation MES (mg/l)
Hochedlinger et al. (2006)	Déversoir d'orage	np*	np*
Maribas et al. (2008)	STEU : décanteur primaire	250-1000	np*
Rieger et al. (2006)	Entrée STEU	50-900 <sup>1</sup>	40-250
Torres et Bertrand-Krajewski (2008)	Entrée de bassin de décan-tation (réseau pluvial)	4-364	2-390
Langergraber et al. (2003)	STEU : décanteur primaire	50-480	25-160

(\*) : np = non précisé – (1) : 4 échantillons sur 108 ont des concentrations supérieures à 600 mg/l

Tab. 3.2 – Emplacement du spectrophotomètre et intervalles de variation de la DCO et MES des échantillons prélevés pour l'étalonnage de la sonde

### 3.5.3 Collecte des données d'étalonnage et de validation

Le nombre minimal d'échantillons collectés pour l'étalonnage de la sonde est de seize (cf. TAB. 3.5.3). Plus le jeu de données est important et plus les erreurs de tout type (dosage, mesure de spectre, prélèvement) seront moyennées et n'influenceront pas la régression.

Toutes les études effectuent des validations simples en réservant 50 % environ du jeu de données à la validation du modèle. Seules deux d'entre elles précisent la façon dont est effectuée la partition. Langergraber *et al.* (2003) ont choisi d'effectuer une partition séquentielle. Par ailleurs, une validation croisée a été effectuée uniquement dans le but de choisir le nombre de composantes du modèle.

Torres et Bertrand-Krajewski (2008) adoptent une stratégie différente en incluant dans chaque partition des échantillons provenant de condition de temps sec et de temps de pluie, et couvrant toute l'étendue de l'intervalle de variation des concentrations. Il est étonnant qu'ils n'aient pas utilisé une procédure de validation croisée. En effet, le jeu de données est limité et comprend seize échantillons dont huit sont utilisés pour la validation.

Étude	Nbre d'échantillons	dont étalonnage	dont validation
Hochedlinger et al. (2006)	72	np*	np*
Maribas et al. (2008)	25	10	15
Rieger et al. (2006)	17 à 85	np*	np*
Torres et Bertrand-Krajewski (2008)	16	8	8
Langergraber et al. (2003)	52	28	24

\* : np = non précisé

Tab. 3.3 – Nombre d'échantillons utilisés pour l'étalonnage de la sonde et partition en jeu de données pour l'étalonnage et la validation des modèles

### 3.5.4 Élimination de valeurs "aberrantes"

Les travaux de Langergraber *et al.* (2003) et Rieger *et al.* (2006) incluent dans la procédure d'étalonnage une étape de détection et élimination des "*valeurs aberrantes*".

Dans la première étude, les échantillons dont les valeurs de concentration calculées ne sont pas dans l'intervalle de confiance à 95 % ont été éliminés de la base de données. Dans la seconde étude, huit valeurs de concentration semblent aberrantes du fait de spectres compensés en turbidité ayant des caractéristiques différentes de l'ensemble des autres spectres. Ces échantillons ont également été écartés. Nous pouvons supposer que les mauvaises prédictions de concentration obtenues pour quelques échantillons dans la première étude sont dues à des formes de spectres d'absorbance différentes.

Le choix de l'élimination de "*valeurs aberrantes*" est discutable. En effet, nous ne pouvons pas valider les modèles identifiés de la sorte pour une utilisation en continu, n'ayant aucun élément pour relier des formes de spectres à des conditions environnementales particulières (temps de pluie, nuit) et ne connaissant pas leur fréquence d'apparition.

### 3.5.5 Méthodes de régression testées

Le TAB. 3.4 récapitule les méthodes de régression utilisées pour l'étalonnage de la sonde. Dans la plupart des articles, les méthodes d'étalonnage fournies par le constructeur (Calage

global et local s::can) sont comparées à des régressions linéaires simples ou à des régressions PLS établies à partir d'échantillons prélevés sur le site d'application.

La publication de Hochedlinger *et al.* (2006) teste le plus grand nombre de méthodes d'étalonnage incluant des méthodes de régression linéaires et non-linéaires.

Les modèles PLS identifiés localement dans ces articles n'utilisent pas l'intégralité de l'information contenue dans le spectre d'absorbance. Des sélections de longueurs d'onde sont effectuées, le plus souvent de manière empirique (Langergraber *et al.*, 2003; Hochedlinger *et al.*, 2006; Rieger *et al.*, 2006) et une fois seulement avec un algorithme de sélection optimisée de longueurs d'onde (Torres et Bertrand-Krajewski, 2008). Les sélections empiriques utilisent des gammes de longueurs d'onde connues historiquement pour leur corrélation avec les paramètres de pollution considérés (cf. TAB. 3.5.1). L'algorithme OPPLS de sélection de variables utilisé par Torres et Bertrand-Krajewski (2008) a pour objectif de sélectionner le jeu de longueurs d'onde optimal qui minimise l'erreur de prédiction. Cette méthode s'avère donc plus pertinente que les sélections réalisées *a priori*. Nous pouvons toutefois regretter que les longueurs sélectionnées avec l'algorithme OPPLS n'aient pas été comparées aux longueurs d'ondes habituellement utilisées pour la DCO et les MES.

Étude	Méthodes de régression	Sélection de variables	Méthode de sélection
Hochedlinger et al.(2006)	Glob. Cal, Local PLS, SVM, RLS, M5, LMS	O	Empirique (7 l.o., 242,5 à 267,5 nm)
Maribas et al. (2008)	Glob. Cal, Local Cal.	N	-
Rieger et al. (2006)	Glob. Cal., Swiss Cal., Local PLS	O	Empirique (3 zones - env. 10 l.o.)
Torres et Bertrand-Krajewski (2008)	Glob. Cal., Local Cal., local PLS	O	OPPLS
Langergraber et al. (2003)	RLS, Local PLS	O	Empirique (5 à 10 l.o.)

Glob. cal : Étalonnage global constructeur - Local Cal. : Étalonnage constructeur ajusté localement - SVM : machines vecteurs support - Local PLS : modèle PLS identifié à partir d'échantillons prélevés localement - RLS : régression linéaire simple - M5 : algorithme utilisant des arbres de classification - LMS : "Linear Median Squares" régression

Tab. 3.4 – Méthodes de régression utilisées pour l'étalonnage du spectrophotomètre

---

## **3.5.6 Performances des méthodes d'étalonnage**

### **3.5.6.1 Comparaison des performances des modèles PLS établis localement avec les calages globaux et locaux**

Trois des cinq études (Rieger *et al.*, 2006; Hochedlinger *et al.*, 2006; Torres et Bertrand-Krajewski, 2008) montrent que les modèles PLS établis à partir de données échantillonnées localement permettent d'améliorer la qualité des prédictions comparativement au calage global ou local du constructeur.

Dans l'article de Rieger *et al.* (2006), la performance des différentes méthodes est évaluée par l'étendue des intervalles de confiance à 95 % sur la prédiction d'une nouvelle concentration de DCO. Celui-ci varie de 99,9 à 342,5 mg/l pour un calage global et de 66 à 285,1 mg/l pour un modèle PLS local. Le gain est par contre à peine perceptible pour les MES : l'intervalle varie de 38,6 à 81,7 mg/l pour un calage global contre 27,6 à 83,5 mg/l pour un modèle PLS local.

L'étude de Torres et Bertrand-Krajewski (2008) utilise le coefficient de détermination  $r^2$  de la régression et la fonction RMSEP pour comparer les différentes méthodes. L'identification d'un modèle PLS avec l'algorithme OPPLS apporte une réelle plus-value pour la prédiction des concentrations de DCO. La fonction d'erreur RMSEP passe de 39 mg/l dans le cas d'un calage local à 14 mg/l pour le modèle PLS local. Pour les MES, les résultats obtenus avec une calibration locale et avec un modèle PLS local sont équivalents (RMSEP de 13 et 11 mg/l).

Enfin, dans la publication de Hochedlinger *et al.* (2006), les performances des modèles sont comparées avec l'erreur relative de prédiction correspondant à une fréquence cumulée de 80 %. Le modèle PLS local donne de bien meilleurs résultats pour la DCO avec une erreur de relative de 16 % contre 32 % pour la calibration globale.

### **3.5.6.2 Comparaison des performances des modèles multivariables avec les performances des modèles de régression linéaire simple**

Les modèles PLS locaux sont également comparés aux régressions linéaires simples (Langergraber *et al.*, 2003; Hochedlinger *et al.*, 2006; Torres et Bertrand-Krajewski, 2008). Les

---

deux méthodes donnent des résultats assez proches avec toutefois un léger avantage pour les modèles PLS locaux.

Dans l'article de Langergraber *et al.* (2003), le coefficient de détermination  $r^2$  calculé à partir des prédictions de DCO est de 0,9 pour le modèle PLS local contre 0,88 pour la régression linéaire simple. L'écart de performance est plus important pour les MES, avec un  $r^2$  de 0,95 pour le modèle PLS local contre 0,83 pour la régression linéaire simple. Nous ne disposons malheureusement pas d'informations concernant les domaines spectraux utilisés.

Dans l'étude de Torres et Bertrand-Krajewski (2008), une régression linéaire simple à partir de mesures de turbidité a été réalisée. Pour la DCO, le modèle PLS local est bien plus performant avec une valeur de RMSEP égale à 14 mg/l contre 49 mg/l pour la régression linéaire simple. Dans le cas de la détermination des concentrations en MES, les méthodes donnent des résultats proches (RMSEP de 21 mg/l pour la régression linéaire et 8 mg/l pour le modèle PLS). Une fois encore, les domaines spectraux utilisés ne sont pas mentionnés. Nous pouvons supposer que les meilleurs résultats obtenus pour la DCO avec le modèle PLS local sont dus à la prise en compte de longueurs d'onde dans l'UV et le visible. Pour les MES, la proximité des résultats obtenus avec les deux méthodes est vraisemblablement due à la sélection de longueurs d'ondes appartenant aux mêmes domaines spectraux.

Enfin, l'étude de Hochedlinger *et al.* (2006) attribue de meilleurs résultats à la régression linéaire simple avec une erreur relative égale 15 % contre 16,3 % pour le modèle PLS. Ces résultats sont surprenants. Nous tenterons de les expliquer dans le prochain paragraphe.

### **3.5.6.3 Nombre de composantes des modèles PLS identifiés**

Seuls Langergraber *et al.* (2003) indiquent le nombre de composantes retenu lors de l'identification du modèle PLS (9 composantes pour le modèle de DCO). Il s'agit pourtant d'un paramètre essentiel conditionnant la capacité de prédiction du modèle. Un nombre trop élevé de composantes entraîne une surparamétrisation du modèle se caractérisant par un bon étalement et de mauvaises prédictions de concentration pour de nouveaux échantillons.

L'étude de Hochedlinger *et al.* (2006) mentionne une surparamétrisation du modèle PLS local (cf. FIG. 3.12). Ceci semble difficilement compréhensible alors que celui-ci est établi sur la base de sept longueurs d'onde. Les longueurs d'onde sont toutes très proches et très corrélées. La réduction de la dimension du nuage de points devrait donc aboutir à l'identification d'un

---

modèle à une composante qui ne peut donc pas être surparamétrisé. La figure reproduite dans l'article montrant bien une surparamétrisation, nous pouvons supposer qu'une régression linéaire multiple de type ILS a été effectuée à la place d'une régression PLS.

### 3.5.7 Bruit

Les origines du bruit sur le signal de concentration sont diverses :

- variabilité réelle des concentrations au sein de l'influent ;
- perturbation de la mesure des spectres d'absorbance par des particules de taille importante ou par des bulles d'air ;
- surparamétrisation du modèle de régression ;

Le bruit est pénalisant car il altère la lecture des concentrations et réduit voire anéantit les possibilités de contrôle en temps-réel du système à partir des mesures de concentration. Lorsqu'il prend des proportions trop importantes, un post-traitement du signal ne permet même plus d'évaluer les concentrations.

L'article de Hochedlinger *et al.* (2006) met en évidence de manière certaine une surparamétrisation du modèle PLS (ou ILS ?). La FIG. 3.12 représentant les concentrations calculées avec la régression linéaire simple (SLR) montre une autre source de bruit due à la variabilité réelle des concentrations ou à la perturbation de la mesure des spectres d'absorbance par des particules ou des bulles d'air. La diminution du bruit peut être obtenue dans ce cas en adoptant un pas de temps de mesure plus grand ou en filtrant le signal (filtre passe bas par exemple).

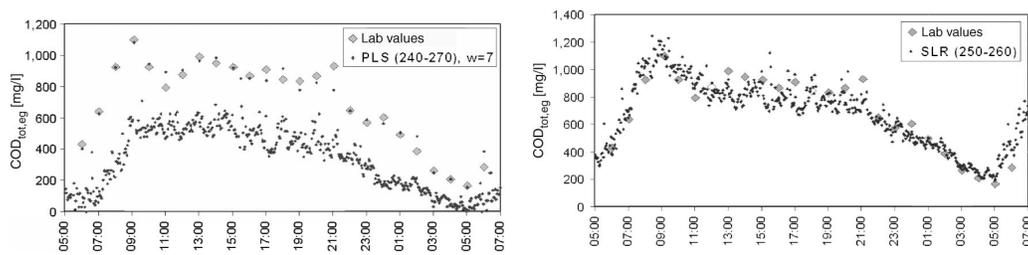


Fig. 3.12 – Surparamétrisation du modèle PLS ou régression linéaire multiple? (source : Hochedlinger et al. (2006)).

---

Le bruit observé dans la publication de Maribas *et al.* (2008) et visible sur la FIG. 3.13, a pour origine, selon les auteurs, la présence de particules et de bulles d'air. Comme on peut l'observer sur ce cas précis, le post-traitement du signal ne permettra probablement pas d'éliminer le bruit du signal provenant des perturbations. Dans ce cas, la meilleure façon de limiter le bruit est donc de choisir un emplacement de mesure adéquat limitant la présence de grosses particules (sables) et de bulles d'air.

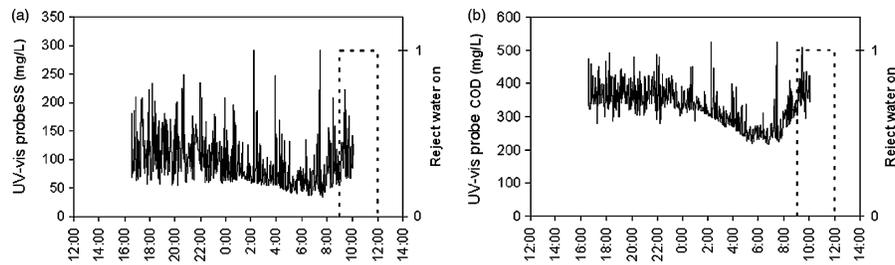


Fig. 3.13 – Bruit sur le signal de concentration en MES et DCO. (source : Maribas *et al.* (2008)).

### 3.5.8 Problématique de mesure des concentrations par temps de pluie

Seul Maribas *et al.* (2008) a étudié la validation de l'étalonnage sur des mesures en temps-réel de concentrations en DCO et MES effectuées par temps de pluie.

Pour que la sonde spectrophotométrique réalise une bonne évaluation des concentrations par temps de pluie, il est nécessaire que le jeu de données comprenne suffisamment de données reflétant la variabilité de la matrice des eaux usées pendant ces périodes. L'étude de Maribas *et al.* (2008) illustre la difficulté à prendre en compte dans l'étalonnage l'ensemble des changements de la matrice provoqués par des événements pluvieux. Le jeu de données de validation contient trois échantillons aux concentrations en DCO supérieures à 800 mg/l et ayant été probablement prélevés par temps de pluie. Les concentrations estimées en calage local sont très proches des concentrations mesurées. Toutefois, lors d'une campagne de mesure supplémentaire réalisée en temps de pluie, les pics de concentrations en DCO et MES estimés par spectrophotométrie UV/Visible sont sous estimés par rapport aux pics de concentrations mesurés en laboratoire avec les méthodes analytiques de référence (cf. FIG. 3.14).

---

Ceci montre que la variabilité de la matrice des eaux est très grande par temps de pluie et qu'il est très difficile de la prendre en compte.

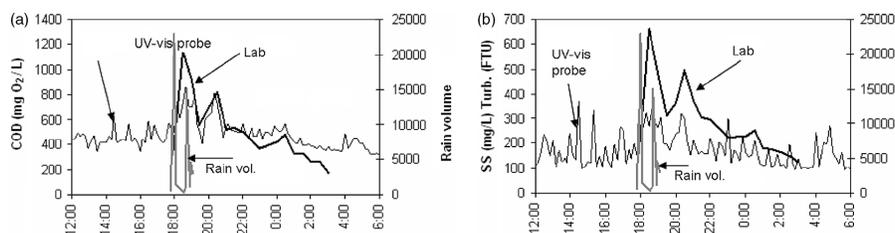


Fig. 3.14 – Mesures de concentration de DCO et MES par temps de pluie (source : Maribas et al. (2008))

### 3.6 Conclusion

La spectrophotométrie UV/Visible est une technique de mesure qui devrait permettre l'évaluation des concentrations de paramètres globaux de pollution tels que la DCO, la DBO<sub>5</sub>, le COT. En effet de nombreuses molécules organiques absorbent le rayonnement électromagnétique dans l'UV. Il est ainsi possible d'établir des relations entre absorbance et concentrations des paramètres de pollution.

La présence de MES au sein d'une solution est responsable de phénomènes de diffusion du rayonnement électromagnétique qui sont fonction de la taille des particules. Comme pour l'absorption moléculaire, il est possible de quantifier les MES par étalonnage d'une relation entre quantité du rayonnement diffusé et concentration en MES. Les longueurs d'onde considérées sont situées dans le domaine du visible étant donné que les composés organiques dissous présents n'interfèrent pas dans ce domaine spectral (au delà de 500 nm).

La détermination de la concentration de paramètres de pollution globaux tels que la DCO dans des eaux usées est problématique. En effet, la DCO est sous forme dissoute et particulaire. De plus, la proportion de ces deux fractions est variable et le pourcentage de matières volatiles en suspension n'est pas constant. Ainsi, la détermination de relations entre concentration et absorbance est le plus souvent réalisée par application de méthodes d'analyse multivariées

---

permettant d'exploiter l'ensemble de l'information contenue dans la mesure des spectres d'absorbance. Les méthodes les plus utilisées sont la déconvolution de spectres et les régressions PLS.

Ces dix dernières années ont vu le développement de spectrophotomètres permettant une mesure in-situ et en continu des concentrations de paramètres globaux de pollution. L'un de ces spectrophotomètres (spectro::lyser, s::can Messtechnik GmbH) a fait l'objet de très nombreuses études visant :

- à tester son usage en différents points du système d'assainissement : déversoir d'orage, entrée et sortie de station de traitement des eaux usées,...
- et à comparer les performances du modèle prédictif fourni par le constructeur - et ajusté localement - avec de nouveaux modèles, de type PLS la plupart du temps, étalonnés à partir de données échantillonnées sur le site d'étude.

Les apports de ces études sont très nombreux. Il a été montré la pertinence de cette technique pour la mesure in-situ et en continu de la concentration des paramètres de pollution tels la DCO ou les MES. Les modèles PLS établis localement ont permis d'améliorer la qualité de la prédiction des concentrations, surtout pour la DCO. Le gain de performance obtenu pour la prédiction des concentrations en MES semble moins évident. Ces publications ont également mis en évidence la nécessité de maîtriser le bruit sur le signal de prédiction de concentrations. Les perturbations du signal observées ont pour origine l'interférence des MES et des bulles d'air mais peuvent aussi être liées à une surparamétrisation des modèles prédictifs.

Les limites de ces études concernent tout d'abord la non définition des conditions d'utilisation des modèles prédictifs. L'étalonnage des modèles PLS a été réalisé dans la plupart des études sur un domaine restreint de concentrations. Par ailleurs les spectres acquis lors de l'étalonnage n'intègrent pas toutes les variations possibles de la matrice des eaux usées. Les modèles ne sont donc pas utilisables sur des spectres aux caractéristiques quantitatives et qualitatives qui n'ont pas été observées pendant l'étalonnage. L'étude de Maribas *et al.* (2008) soulève d'ailleurs la difficulté de prédire les concentrations par temps de pluie en raison d'une variation importante de la matrice des eaux usées. Or, la mesure des concentrations par temps de pluie constitue justement un des intérêts majeurs de l'utilisation de cette technique de mesure.

Par ailleurs, il semble que les performances des modèles prédictifs de concentration en DCO et MES puissent encore être améliorées. En effet, ils sont la plupart du temps basés sur une sélection empirique de longueurs d'ondes. Il a été montré sur de nombreuses applications spec-

---

troscopiques infrarouge et proche infrarouge qu'une sélection optimisée de longueurs d'ondes permettait d'améliorer les prédictions des modèles PLS. Sur les études réalisées en assainissement et utilisant la spectrophotométrie UV/Visible pour l'évaluation des concentrations en DCO et MES, seuls PLS Torres et Bertrand-Krajewski (2008) ont développé un algorithme de sélection de longueurs d'ondes. Les résultats obtenus sont très encourageants et montrent l'intérêt de telles procédures de sélection pour la mesure des concentrations de paramètres globaux de pollution.

## Chapitre 4

# Mesure de pollution en entrée de station de traitement d'eaux usées

### 4.1 Introduction

Le chapitre précédent a permis de montrer les apports et limites des études portant sur les mesures en continu de paramètres globaux de pollution en système d'assainissement. L'objectif de ce chapitre est de compléter les précédents travaux de recherche.

#### **Objectif n°1**

Le premier objectif de ce chapitre concerne l'amélioration des performances des modèles prédictifs de concentrations par l'utilisation d'algorithmes de sélection de longueurs d'onde.

Les méthodes de régression multivariable comme la régression PLS permettent de prendre en compte l'intégralité du spectre sans avoir à effectuer auparavant une sélection de variables. Toutefois, aussi bien au niveau théorique qu'expérimental, il a été mis en évidence que la sélection de variables est importante et nécessaire pour obtenir des modèles parcimonieux et robustes (Norgaard *et al.*, 2000; Tan et Li, 2008).

Une procédure exhaustive de sélection variables (OPPPLS) a été testée avec succès par Torres et Bertrand-Krajewski (2008). Même si les résultats obtenus sont très satisfaisants, cette méthode ne tient pas compte des synergies entre différentes régions spectrales. Il est donc possible que l'ensemble de longueurs d'ondes sélectionné soit sous-optimal.

---

Nous proposons l'utilisation de deux algorithmes basés sur la sélection d'un ou plusieurs intervalles de longueurs d'ondes. Le premier est l'algorithme *forward-backward interval PLS* (fbiPLS) qui constitue un nouveau développement de la méthode de sélection de longueurs d'ondes *interval PLS* créée par Norgaard *et al.* (2000). Le second algorithme proposé est l'algorithme *moving window PLS* (mwPLS) qui a été développé par Jiang *et al.* (2002). Ces méthodes ont été testées dans nombre d'applications en spectroscopie infrarouge ou proche infrarouge dans le cadre de la quantification de composés chimiques. Toutefois il n'a pas encore été montré leur plus-value en spectrophotométrie UV/Visible pour la sélection des longueurs d'ondes les plus adaptées à la prédiction de la concentration en DCO et MES.

Les performances des modèles établis avec les procédures fbiPLS et mwPLS seront comparées avec celles des modèles issus de l'utilisation de l'algorithme proposé par Torres et Bertrand-Krajewski (2008), et avec les méthodes d'étalonnage les plus couramment utilisées (calage global et local *s::can*, régression linéaire simple et modèles PLS sans sélection de longueurs d'ondes).

### **Objectif n°2**

Le second objectif est la détermination des conditions d'utilisation des modèles prédictifs. En effet, les modèles ne peuvent être appliqués que sur de nouveaux spectres d'absorbance correspondant à des caractéristiques quantitatives et qualitatives observées pendant l'étalonnage. Si la collection d'étalonnage n'intègre pas toutes les variations possibles de la matrice des eaux usées, il convient alors de déterminer le domaine de validité des modèles prédictifs.

Nous proposons une méthode de détermination d'un domaine de validité basée sur la comparaison de la collection d'étalonnage avec les données d'autosurveillance ainsi que sur une étude qualitative des spectres normalisés.

## **4.2 Matériel et méthodes**

La méthodologie d'établissement de modèles prédictifs de concentrations en DCO et MES adoptée dans le cadre la thèse s'inspire de la méthodologie générale présentée au paragraphe 3.3. Elle comporte six étapes illustrée en FIG. 4.1. La description de chaque étape sera détaillée dans les prochains paragraphes.

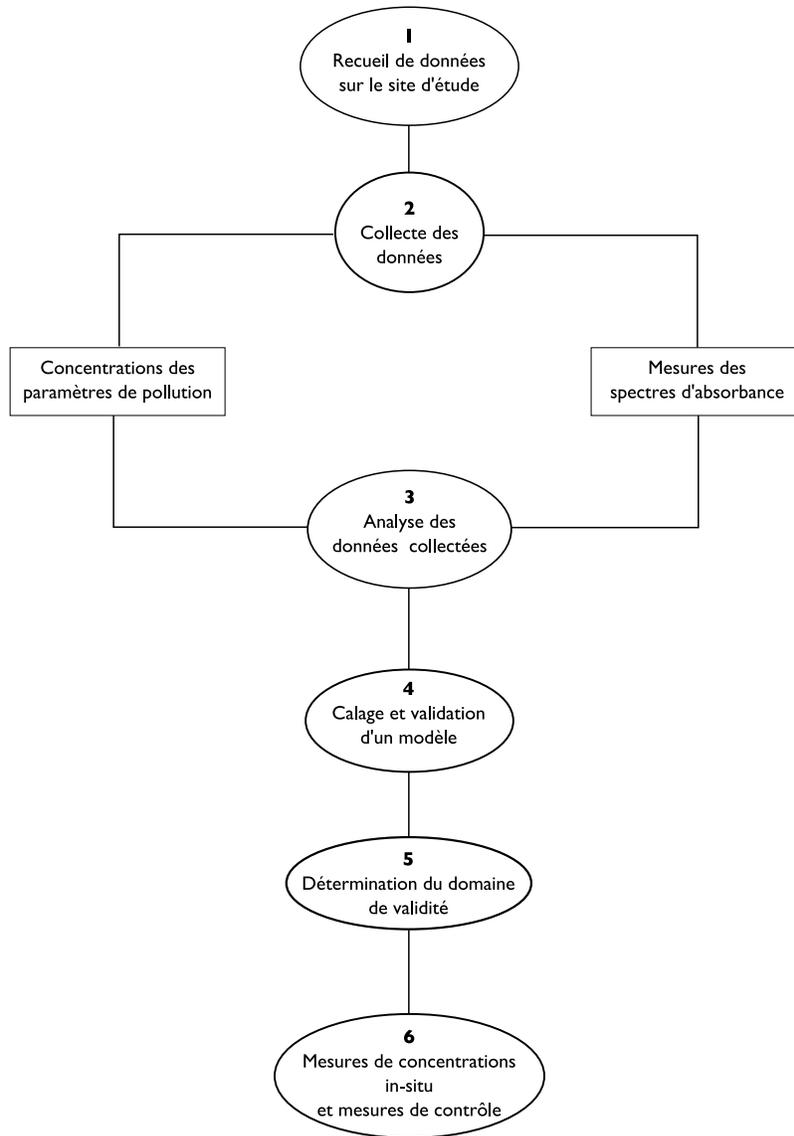


Fig. 4.1 – Méthodologie d'établissement de modèle prédictif proposée

---

## **4.2.1 Étape 1 : description du site d'étude**

L'étalonnage de la sonde spectrophotométrique a été réalisé hors-ligne (en laboratoire) sur des eaux résiduaires de la station de traitement des eaux usées de Rosheim (67). Après étalonnage, le spectrophotomètre a été placé en entrée de station afin d'acquérir des mesures en temps-réel.

### **4.2.1.1 Description du système d'assainissement**

Le syndicat du Rosenmeer regroupe trois communes, pour une population d'environ environ 7800 habitants :

- Rosheim(4 500 habitants),
- Bischoffsheim (2 700 habitants),
- et Rosenwiller (600 habitants).

Ces communes sont situées dans le département du Bas-Rhin, à une trentaine de kilomètres au Sud-Ouest de Strasbourg, entre Molsheim et Obernai.

Au niveau topographique, ces communes sont situées sur une zone de transition entre la plaine d'Alsace et le piémont des Vosges. Il s'agit d'une zone viticole où les collines exposées au sud sont majoritairement occupées par les vignes (Rosheim et Rosenwiller). Cette activité induit une surcharge de pollution organique saisonnière estimée à 5 500 équivalent habitant (EH). Par ailleurs, à la pollution d'origine domestique et viticole s'ajoute une pollution d'origine industrielle liée aux entreprises installées sur les Communes de Rosheim et de Bischoffsheim.

Les eaux usées collectées sur ces trois communes sont traitées à la station de traitement des eaux usées située sur le territoire communal de Rosheim. Un bassin d'orage en dérivation, situé à l'amont immédiat de l'installation de traitement, permet de stocker et de traiter une partie des effluents de temps de pluie avant leur rejet dans le Rosenmeer.

Le réseau d'assainissement est majoritairement unitaire (57 km) avec quelques branches nouvelles en séparatif (3km). Il comporte neuf déversoirs d'orages, deux bassins de pollution dont un est situé juste en amont de la station de traitement des eaux usées, un poste de refoulement et un poste de relevage.

---

La station de traitement des eaux usées du syndicat du Rosenmeer est de type boues activées à très faible charge avec nitrification et dénitrification associée à un traitement physico-chimique du phosphore.

Le bassin d'orage situé juste en amont de la station de traitement des eaux usées est de type transit à connexion latérale avec une capacité de stockage maximale de 2520 m<sup>3</sup>. Son alimentation et sa vidange sont réalisées au niveau de la bêche de relevage de la station de traitement des eaux usées.

#### 4.2.1.2 Synthèse des données d'autosurveillance

L'étude des analyses issues de l'autosurveillance a pour but de caractériser la pollution des échantillons prélevés pour l'étalonnage du spectrophotomètre par rapport à la pollution admise en entrée de station de traitement.

Nous avons recueilli les données d'autosurveillance couvrant les années civiles 2002 à 2009. Les valeurs des différents paramètres de pollution analysés en entrée de station sont récapitulées dans le tableau 4.1. Les valeurs moyennes de DCO, MES, DBO<sub>5</sub>, NH<sub>4</sub> et NTK sont caractéristiques d'effluents domestiques urbains, ce qui est confirmé par les valeurs moyennes de ratios entre les différents paramètres de pollution (cf. TAB. 4.2).

Le tableau 4.1 nous renseigne également sur les intervalles de variation de chacun des paramètres de pollution. Les valeurs de DCO fluctuent entre 75 mg/l et 1708 mg/l. Les valeurs de MES sont quant à elles comprises entre 35 et 1734 mg/l. L'objectif du recueil d'échantillons est de couvrir l'étendue de la variation de chaque paramètre.

Paramètre	Débit (m <sup>3</sup> /j)	Pluv. (mm)	DCO (mg/l)	DBO <sub>5</sub> (mg/l)	MES (mg/l)	N-NH <sub>4</sub> (mg/l)	NTK (mg/l)
Moyenne	2535	1,5	540	244	282	32	48
Percentile 90	4349	4,8	870	428	470	47	74
Écart type	1114	3,4	248	146	191	13	19
Minimum	961	0,0	75	27	35	3	7
Maximum	6519	29,0	1708	733	1734	67	102
Nb Valeurs	525	520	420	100	418	387	100

Tab. 4.1 – Valeurs des différents paramètres analysés dans le cadre de l'autosurveillance de la STEU de Rosheim (2002 à 2009)

<b>Ratio</b>	<b>Valeurs calculées</b>	<b>Valeurs usuellement rencontrées*</b>
DCO/DBO <sub>5</sub>	2,2	2,2 - 2,4
DCO/MES	2.0	1.8 - 2.5
DBO <sub>5</sub> /NTK	5,1	4 - 5
DCO/NTK	10.9	8,8 - 12

(\*) : source Sadowski (2002)

Tab. 4.2 – Relations entre les paramètres de pollution en entrée de station

#### 4.2.1.3 Installation du spectrophotomètre in-situ

Après étalonnage, le spectrophotomètre a été placé dans le canal d'entrée de la station de traitement des eaux usées, en aval de la bêche de relevage et juste à l'amont du dégrilleur (cf. FIG. 4.2).

Les mesures effectuées in-situ reflètent la qualité des eaux usées à l'aval du réseau d'assainissement sauf pendant les périodes de vidange du bassin d'orage. En effet, les effluents de vidange du bassin contiennent beaucoup moins de pollution sous forme particulaire (Leclaire, 1997).

Le canal où a été installé le spectrophotomètre est alimenté de manière discontinue par les pompes de la bêche de relevage. Un seuil rectangulaire a été mis en place dans le canal afin d'augmenter le tirant d'eau pendant les périodes de fonctionnement des pompes et ainsi permettre les mesures spectrophotométriques.

Deux campagnes de mesures ont été réalisées in-situ. La première a eu lieu en avril 2008 pendant huit jours avec un spectro::lyser prêté par Véolia eau. La seconde s'est déroulée en mai-juin 2010 après l'acquisition d'un spectro::lyser par l'équipe HU ENGEES-IMFS.

#### 4.2.2 Étape 2 : collecte des données d'étalonnage

Nous avons constitué une base de données de quinze échantillons prélevés entre le 02 février 2010 et le 28 avril 2010.

La base de données contient :



Fig. 4.2 – Photos du spectro::lyser installé en entrée de la station de traitement des eaux usées de Rosheim

- les concentrations en DCO et MES des échantillons prélevés ;
- et les spectres d'absorbance des mêmes échantillons mesurés avec le spectrophotomètre utilisé in-situ et/ou avec un spectrophotomètre de laboratoire.

Afin de caractériser plus finement la pollution des échantillons prélevés, nous avons réalisé des analyses complémentaires de DCO filtrée à  $0,45 \mu\text{m}$ , d'azote ammoniacal et de conductivité.

#### 4.2.2.1 Prélèvement et conservation des échantillons

Les prélèvements ont porté sur des échantillons instantanés (11/15) ainsi que sur des échantillons moyens 24h (4/15). Les échantillons instantanés correspondent aux conditions de mesure dans lesquelles le spectrophotomètre a été installé. Les échantillons moyens 24h permettent quant à eux d'agglomérer la variabilité de la pollution écoulée sur une journée.

Les échantillons instantanés ont été prélevés manuellement à l'aide d'un seau dont le contenu a été versé après brassage dans un flacon en polyéthylène de 2 litres. Les échantillons moyens

---

24h ont été prélevés avec le préleveur réfrigéré de la station de traitement des eaux usées qui est situé 80 cm en aval du spectrophotomètre.

Pendant le transport jusqu'au laboratoire d'analyses, tous les échantillons ont été conservés dans une glacière. Les analyses ont été effectuées dans les 4h suivant le recueil des échantillons à la station de traitement des eaux usées.

#### **4.2.2.2 Mesures analytiques de la DCO et des MES**

Les dosages de DCO ont été réalisés suivant la méthode normalisée de reflux en tube fermé ISO 15705 : 2002. Dans cette méthode, la matière oxydable contenue dans un échantillon est oxydée par chauffage à reflux en milieu fortement acide avec une quantité connue de bichromate de potassium dans une éprouvette fermée. La consommation d'oxygène par l'échantillon provoque un changement de couleur dont l'absorbance est proportionnelle à la quantité de bichromate de potassium réduit et se mesure en équivalent d'oxygène. Cette méthode n'ayant jamais été utilisée auparavant au laboratoire, elle a été validée et comparée à la méthode normalisée ISO 6060 : 1989 avant d'effectuer les dosages sur les échantillons d'eau usée (cf. annexe A). Avant analyse, les échantillons ont été mixés dans un blender afin de garantir une meilleure homogénéité de la prise d'essai. De plus, pour réaliser le dosage de la DCO filtrée, les échantillons ont été filtrés sur un filtre en nitrate de cellulose de porosité 0,45  $\mu\text{m}$ .

Le dosage des MES a été effectué par filtration sur filtre en fibre de verre (norme NF EN 872). Les filtres utilisés sont des filtres Whatman type GF/C. Enfin, l'azote ammoniacal a été dosé avec la méthode de Nessler.

Les résultats des analyses sont présentés dans les tableaux 4.3 et 4.4. Notons que trois des quinze échantillons ont été prélevés par temps de pluie (05/02, 18/02 et 01/03).

#### **4.2.2.3 Mesures spectrophotométriques**

Durant le premier mois, les spectres ont été mesurés avec un spectrophotomètre de laboratoire Perkin-Elmer. En effet, le spectrophotomètre in-situ (spectro::lyser) n'a été livré qu'à la fin du mois de février.

A la réception du spectro::lyser, les spectres ont alors été mesurés avec les deux appareils. Un calage d'une loi de transfert des spectres mesurés avec le spectrophotomètre Perkin-Elmer vers

Date	Type	DCO (mg/l)		DCO <sub>f</sub> (mg/l)		MES (mg/l)		N-NH <sub>4</sub> (mg/l)		Cond. μS/cm
		$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	
02/02	inst	610	11,4	302	4,2	243	5,8	32	1,0	1144
05/02	m24h	701	14,3	196	5,9	380	13,4	46	0,6	1122
09/02	inst	640	8,2	160	25,0	352	14,7	62	2,4	1167
16/02	inst	336	11,1	122	33,2	154	11,7	49	1,2	999
18/02	m24h	359	5,9	177	0,5	145	1,4	21	0,3	1111
22/02	inst	373	3,6	149	0,6	166	5,2	46	0,7	916
01/03	m24h	237	1,6	59	0,6	149	13,4	18	0,1	596
15/03	inst	616	19,7	235	17,6	284	21,7	75	8,4	1195
19/03	inst	598	8,7	215	2,2	266	19,2	53	1,8	1273
25/03	inst	758	107,8	160	2,7	376	18,0	72	0,3	1193
31/03	inst	518	2,7	189	3,4	247	29,6	65	0,9	1116
09/04	inst	648	6,9	160	2,0	309	32,7	69	0,3	1402
15/04	inst	467	5,0	152	2,6	249	4,4	65	0,8	1113
22/04	inst	778	18,3	192	12,0	453	12,4	83	6,4	1264
28/04	inst	751	31,4	166	4,7	631	71,0	73	1,8	1171

Tab. 4.3 – Résultats des analyses réalisées sur les échantillons prélevés en entrée de la station de Rosheim

	DCO	DCO <sub>f</sub>	MES	N-NH <sub>4</sub>	Cond.
Nb. Valeurs	15	15	15	15	15
moyenne	559	176	294	55	1119
E. Type	170	54	132	20	183
P 10 %	345	133	151	25	949
P 90%	755	227	424	74	1269

Tab. 4.4 – Synthèse des résultats d'analyses effectués sur les échantillons prélevés en entrée de la station de Rosheim

le spectro::lyser a été réalisé de manière à pouvoir constituer une base de données complète de spectres d'absorbance issus du spectro::lyser (cf. Annexe B). Les spectres d'absorbance mesurés avec le spectro::lyser ont fait l'objet de trois réplicats. La valeur moyenne des trois spectres a été intégrée dans la base de données.

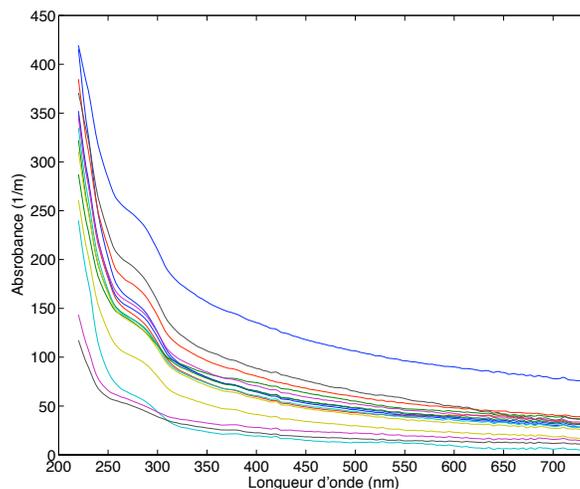


Fig. 4.3 – Spectres d'absorbance bruts

### 4.2.3 Étape 3 : analyse des données collectées

Cette étape a deux objectifs. Le premier est l'élimination de valeurs aberrantes résultant d'erreurs de mesure. Le second est de classifier qualitativement les spectres d'absorbance à partir de leur forme. En effet, il a été montré que ce critère intègre nombre d'informations sur les caractéristiques des eaux usées (proportion respective de la fraction dissoute et particulaire notamment) (Thomas, 2007). Les outils utilisés sont la normalisation de spectres, qui permet de s'affranchir de la composante quantitative, ainsi que des outils d'analyse de données tels que l'analyse en composantes principales et la classification par méthode des k plus proches voisins (Sharaf *et al.*, 1986). Dans cette méthode les observations (ici des spectres) sont classifiées en fonction de leur distance respective avec les k plus proches voisins.

### 4.2.4 Étape 4 : étalonnage et validation des modèles prédictifs

Différentes méthodes d'étalonnage de modèles prédictifs ont été testées (cf. FIG. 4.4) :

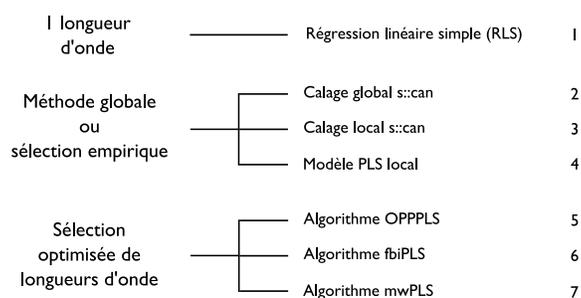


Fig. 4.4 – Méthodes d'étalonnage testées pour l'établissement de modèles prédictifs de concentrations en MES et DCO

L'algorithme utilisé pour effectuer les régressions PLS est l'algorithme SIMPLS créé par de Jong (1993) et implémenté dans la boîte à outils PLS toolbox 5.0. pour MATLAB (Eigenvector Research).

Ne disposant que de quinze échantillons, nous avons opté pour une validation croisée des modèles prédictifs de type *Leave one out*. Cette méthode consiste à retirer successivement chaque échantillon disponible dans le jeu de validation.

#### 4.2.5 Description des algorithmes de sélection de longueurs d'onde utilisés

Les spectrophotomètres actuels comportent une résolution très fine<sup>1</sup> ce qui génère des matrices d'observations avec un nombre élevé de colonnes. Une des propriétés des régressions PCR et PLS est de réduire la dimension de l'espace des spectres par une projection des spectres dans un sous espace vectoriel de dimension inférieure. Cependant, l'ensemble des longueurs d'ondes ne concourent pas à la prédiction de la concentration du paramètre de pollution recherché. Une procédure de sélection de variables permet souvent d'améliorer les résultats de la régression (Norgaard *et al.*, 2000; Tan et Li, 2008; Zou *et al.*, 2007).

Les prochains paragraphes présentent les méthodes de sélection de longueurs d'ondes testées (cf. FIG. 4.5).

<sup>1</sup>Les spectrophotomètres de laboratoire peuvent mesurer les spectres d'absorbance avec une résolution inférieure à 1nm. Pour une mesure dans l'UV et le Visible, les matrices d'observations peuvent donc comporter plus de 500 colonnes.

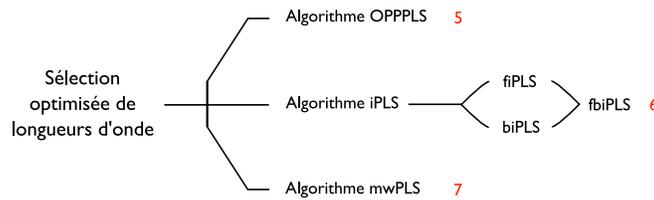


Fig. 4.5 – Méthodes de sélection de longueurs d’onde présentées (seules les méthodes affectées d’un numéro ont été testées)

#### 4.2.5.1 Algorithme OPPPLS

De nombreuses méthodes de sélections de longueurs d’onde existent, telle que la méthode OPPPLS créée par Torres et Bertrand-Krajewski (2008). Il s’agit d’une procédure exhaustive de sélection basée sur le coefficient de détermination  $r^2$  entre chaque variable spectrale et les concentrations. Les variables sont rangées par ordre décroissant de  $r^2$ . Un modèle est identifié à partir des  $i$  premières variables,  $\forall i \in 1, 2, \dots, n$ ,  $n$  étant le nombre de longueurs d’ondes mesurées. Le modèle qui minimise la fonction d’erreur RMSECV ou RMSEP<sup>2</sup> est retenu.

Cette méthode a donné d’excellents résultats lors de son application en assainissement pluvial. Toutefois, elle ne tient pas compte des synergies entre différentes régions spectrales. Elle peut donc conduire à l’identification d’un ensemble de variables sous-optimal. Cette méthode fait le postulat que l’ajout successif de variables spectrales ayant un coefficient de détermination  $r^2$  élevé avec la concentration à prédire améliorera les résultats du modèle. Or, deux variables voisines ayant une coefficient  $r^2$  élevé sont souvent colinéaires, et n’apportent qu’une information redondante. Il est donc possible que cette méthode conduise à identifier des domaines spectraux très étendus dont seulement une partie de l’information est utile.

Différentes méthodes de sélection de variables permettent de prendre en compte les synergies entre différentes régions spectrales, et devraient permettre d’éliminer plus facilement toute longueur d’onde n’apportant aucune information pour la prédiction de la concentration des paramètres de pollution. Citons notamment les méthodes interval PLS (iPLS), backward iPLS, forward iPLS ou moving window PLS, testées avec succès dans nombre d’applications de spectroscopie proche infrarouge ou infrarouge (Jiang *et al.*, 2002; Leardi et Norgaard, 2004; Norgaard *et al.*, 2000; Tan et Li, 2008).

<sup>2</sup>la fonction RMSEP correspond à la fonction de coût calculée en validation simple

---

#### 4.2.5.2 Algorithme forward-backward iPLS (fbiPLS)

L'algorithme forward-backward interval PLS (fbiPLS) constitue une nouvelle adaptation de l'algorithme interval PLS (iPLS) créé par Norgaard *et al.* (2000) et testé dans sa version originale en spectroscopie proche infrarouge.

##### 4.2.5.2.a Algorithme iPLS

Avant d'aborder la procédure fbiPLS, il est nécessaire de revenir sur la méthode interval PLS (iPLS) introduite par Norgaard *et al.* (2000). Dans l'algorithme iPLS, le spectre est divisé en un nombre défini d'intervalles. Un modèle PLS local est calculé pour chaque intervalle. Le nombre de composantes sélectionné est celui qui minimise l'erreur de validation croisée RMSECV ou de prédiction RMSEP. L'intérêt de cette méthode réside dans sa présentation graphique qui permet de visualiser les intervalles où le modèle local est plus performant que le modèle global utilisant l'ensemble de l'information spectrale.

Par ailleurs, même si cette méthode tient compte des synergies entre les variables spectrales au sein d'un même intervalle, elle ne tient pas compte des synergies entre différentes régions spectrales. Elle permet de sélectionner le meilleur intervalle dans une partition donnée, mais ne permet pas d'identifier le meilleur ensemble d'intervalles qui minimisera l'erreur de validation croisée ou de prédiction (Leardi et Norgaard, 2004).

Une solution consisterait à former toutes les combinaisons possibles de 2, 3, 4, ...intervalles et de calculer l'erreur de validation croisée ou l'erreur de prédiction. Cependant cette solution trouve ses limites au niveau des temps de calculs. Pour l'exemple, une partition d'un spectre en 15 intervalles, entraîne 455 combinaisons de 3 intervalles, 1365 de 4 intervalles, 3003 partitions de 5 intervalles...

##### 4.2.5.2.b Algorithme forward iPLS (fiPLS)

La procédure forward iPLS est une procédure itérative dérivée de la procédure iPLS et décrite dans Zou *et al.* (2007). Comme dans la procédure iPLS, le spectre est scindé en  $n$  intervalles,  $n$  étant déterminé par l'utilisateur. La matrice contenant tous les spectres d'absorbance  $\mathbf{A}$  est donc scindée en plusieurs matrices  $\mathbf{A} = \{\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_n\}$ . Pour initialiser la procédure, on

---

forme  $n$  modèles PLS pour chaque intervalle, en faisant varier le nombre de composantes. L'intervalle qui minimise RMSECV est sélectionné :

$$\mathbf{A}_{S_1} = \underset{A_j}{\operatorname{argmin}} \{ \operatorname{RMSECV}(\mathbf{A}_j) \}, 1 \leq j \leq n \quad (4.1)$$

Une fois le premier intervalle sélectionné, la procédure consiste à sélectionner le second intervalle parmi les intervalles restants  $\mathbf{A}_j, j \neq S_1$ . Les erreurs de validation sont calculées pour chaque modèle. L'intervalle qui minimise l'erreur de prédiction est sélectionné si l'erreur de prédiction calculée à cette étape est inférieure à l'erreur calculée à l'étape précédente :

$$\begin{aligned} \mathbf{A}_{S_2} &= \underset{A_j}{\operatorname{argmin}} \{ \operatorname{RMSECV}(\{\mathbf{A}_{S_1}, \mathbf{A}_j\}) \}, 1 \leq j \leq n, j \neq S_1 \\ &\text{si } \operatorname{RMSEP}(\{\mathbf{A}_{S_1}, \mathbf{A}_j\}) < \operatorname{RMSEP}(\mathbf{A}_j) \end{aligned} \quad (4.2)$$

Les intervalles suivants sont sélectionnés suivant le même principe. Si lors de la  $i^{\text{ème}}$  étape, l'ajout d'un intervalle supplémentaire ne permet plus de diminuer l'erreur de validation croisée ou de prédiction, alors l'algorithme s'arrête. La condition d'arrêt s'écrit donc de la manière suivante :

$$\begin{aligned} &\text{Si } \operatorname{RMSECV}(\{\mathbf{A}_{S_1}, \mathbf{A}_{S_2}, \dots, \mathbf{A}_{S_{i-1}}, \mathbf{A}_j\}) \geq \operatorname{RMSECV}(\{\mathbf{A}_{S_1}, \mathbf{A}_{S_2}, \dots, \mathbf{A}_{S_{i-1}}\}) \\ &\forall j \in \{1, \dots, n\}, j \neq \{S_1, S_2, \dots, S_{i-1}\} \\ &\Rightarrow \text{FIN} \end{aligned} \quad (4.3)$$

#### 4.2.5.2.c Algorithme backward iPLS (biPLS)

La procédure backward est la procédure itérative inverse de la procédure fiPLS (Zou *et al.*, 2007). Elle consiste à retrancher successivement les intervalles qui pénalisent le plus la valeur de RMSECV. Comme dans la procédure iPLS, le spectre d'absorbance est partitionné en  $n$  intervalles. La procédure est alors la suivante :

1. L'initialisation de l'algorithme est effectuée en calculant la valeur de RMSECV pour un modèle PLS basé sur l'intégralité du spectre d'absorbance, *i.e.* à partir des  $n$  intervalles.
2. Lors de la seconde itération, un modèle prédictif est établi pour chaque combinaison possible de  $n-1$  intervalles. Si le meilleur de ces modèles diminue l'erreur de validation croisée, alors l'intervalle non utilisé pour l'établissement du modèle prédictif est écarté.

- 
3. Les itérations suivantes se succèdent suivant le même principe. L'algorithme s'arrête lorsque le retrait d'un intervalle ne permet plus de diminuer la fonction RMSECV.

#### 4.2.5.2.d Algorithme fbiPLS

L'algorithme forward-backward interval PLS est inspiré de la procédure de sélection de longueurs d'onde forward-backward utilisée par Rossi *et al.* (2006) et basée sur le calcul de l'information mutuelle. Les auteurs de cet article avancent que l'utilisation stricte d'une procédure itérative *forward* aboutit souvent à la sélection d'un ensemble de variable sous-optimal. L'inclusion à une itération donnée d'une variable inadéquate peut en effet modifier considérablement le choix des variables suivantes.

Il est donc proposé de coupler les procédures *forward* et *backward* de la manière suivante :

- La procédure *forward* est effectuée de manière plus souple en ajoutant à chaque itération l'intervalle qui minimise la fonction de coût, même si toutefois la valeur de la fonction de coût l'itération  $i$  est supérieure à l'erreur de prédiction calculée à l'itération  $i - 1$ .
- Cette procédure est suivie de la procédure *backward* qui permettra d'éliminer tous les intervalles qui ont été ajoutés à tort dans la première sélection. Le but est d'éliminer tous les intervalles qui ne concourent pas à la prédiction de la concentration du paramètre de pollution.

#### 4.2.5.3 Algorithme "moving window PLS" (mwPLS)

Le troisième algorithme de sélection de longueurs d'onde utilisé dans cette thèse est l'algorithme *moving window PLS* créé par Jiang *et al.* (2002).

Cette méthode consiste à former des modèles prédictifs basés sur une fenêtre glissante de longueurs d'onde dont l'étendue est déterminée *a priori* par l'utilisateur. A chaque position de la fenêtre glissante, un modèle prédictif est étalonné avec  $1, 2, 3, \dots, n_c$  composantes. Les fonctions d'erreur RMSEC et RMSECV sont calculées pour chaque position de la fenêtre glissante et pour chaque dimension du modèle. On choisit alors la position de la fenêtre glissante pour laquelle le minimum de RMSECV est atteint.

Quelque soit la procédure de sélection de longueurs d'onde, le choix d'un modèle prédictif est le résultat d'un compromis entre qualité de prédiction et simplicité du modèle. Le risque est de privilégier des modèles comprenant un nombre trop important de dimensions et apportant un gain limité en terme de prédiction. En effet, l'ajout d'une dimension supplémentaire aura alors de fortes probabilités de se traduire par un bruit sur le signal de concentrations. Il est ainsi prudent de choisir un modèle prédictif plus simple que le modèle apparemment optimal (Bertrand, 2005).

#### 4.2.6 Étape 5 : détermination du domaine de validité des modèles prédictifs

Les modèles prédictifs ne peuvent être utilisés pour la prédiction de la concentration de nouveaux échantillons que dans un domaine de validité défini. Celui-ci a été basé sur trois aspects (cf. FIG. 4.6).

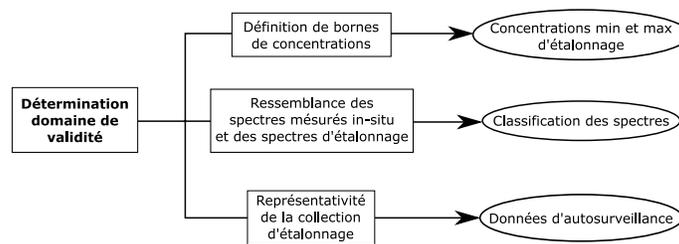


Fig. 4.6 – Méthodologie de détermination d'un domaine de validité

Tout d'abord, l'utilisation des modèles a été restreinte à un domaine de concentrations correspondant aux valeurs extrêmes rencontrées dans le jeu de données d'étalonnage.

Nous avons par ailleurs défini une classification basée sur la forme de spectre d'absorbance afin de repérer les spectres ayant des formes atypiques par rapport aux spectres d'étalonnage. En effet, les formes de spectre sont représentatives de la matrice des eaux usées. Un changement de composition des eaux, en temps de pluie par exemple, modifie considérablement la forme des spectres et rend les prédictions aléatoires. La classification a été réalisée par la méthode des k plus proches voisins (k=1 dans notre cas). Les spectres mesurés in-situ ont ensuite été classés en assignant à chaque observation la classe du spectre d'étalonnage le plus proche.

---

Enfin, nous avons étudié les données d'autosurveillance acquises sur le site d'étude entre 2002 et 2009 dans un double objectif. Le premier est la critique de l'étendue des concentrations du jeu de données d'étalonnage par rapport à l'étendue des concentrations mesurées sur le site d'étude. Ceci nous permettra en outre de compléter le plus judicieusement le jeu de données d'étalonnage et ainsi étendre le domaine de validité des modèles. Le second objectif est l'évaluation de la représentativité de la collection d'étalonnage par rapport à la variabilité mensuelle des influents.

#### **4.2.7 Étape 6 : mesures in-situ et mesures de contrôle**

Cette étape concerne l'analyse des mesures de concentrations effectuées in-situ et en continu. Elle comporte tout d'abord une analyse qualitative des signaux de concentrations obtenus avec chaque méthode d'étalonnage testée. Les concentrations calculées avec les modèles prédictifs ont ensuite été comparées à des mesures analytiques de concentrations réalisées sur des prélèvements instantanés ainsi que sur un prélèvement moyen journalier (16/04/2010).

### **4.3 Résultats**

#### **4.3.1 Qualité des mesures (étape 3)**

##### **4.3.1.1 Dosage des paramètres de pollution**

La réplicabilité des analyses est satisfaisante. Les coefficients de variation sont respectivement de 2,6, 4,8, 5,9 et 4,7 % pour la DCO, la  $DCO_f$ , les MES et le  $N-NH_4$ . Le coefficient de variation maximum est de 14 % et est obtenu pour la mesure de DCO du 25/03/2010. Les rapports entre les différents paramètres semblent confirmer la cohérence de la valeur de DCO pour cet échantillon.

Plusieurs échantillons présentent des valeurs de concentrations atypiques dont l'origine doit être mise au crédit de conditions extérieures particulières. Le premier est l'échantillon du 28/04/2010 prélevé en temps sec, et caractérisé par des concentrations en DCO et MES élevées et un rapport DCO/MES de 1,19 qui est très faible comparé au rapport observé sur

---

l'ensemble du jeu de données (min : 1,19 ; moy : 2 ; max ; 2,5). La concentration en N-NH<sub>4</sub> et la conductivité sont également très élevées.

L'échantillon du 01/03/2010 est prélevé en temps de pluie et présente des concentrations en DCO, MES, N-NH<sub>4</sub>, et conductivité très faibles.

#### **4.3.1.2 Mesure des spectres d'absorbance**

Les six premiers spectres d'absorbance ont été mesurés avec le spectrophotomètre de laboratoire Perkin-Elmer, puis convertis en "équivalent spectro::lyser" au moyen d'une loi de transfert établie à partir des mesures des spectres réalisées sur les neuf échantillons suivants. La procédure de transfert détaillée en Annexe B présente d'excellents résultats. La FIG. 4.7 montre les mesures d'absorbance à 255 nm obtenues avec les deux spectrophotomètres pour chacun des quinze échantillons. L'échantillon du 28/04/2010 est très largement en dehors de l'intervalle de confiance à 95 % de la régression effectuée sur les quatorze premières observations. La différence de mesure entre les deux appareils est probablement due à un dysfonctionnement du spectro::lyser qui a été identifié lors des mesures effectuées in-situ (cf. §4.3.4). L'échantillon du 28/04/2010 a donc été exclu de la collection d'étalonnage.

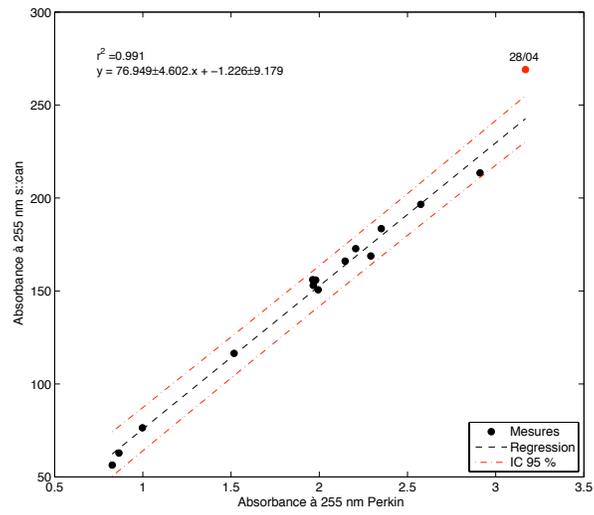


Fig. 4.7 – Absorbances à 255 nm mesurées avec les spectrophotomètres Perkin-Elmer et s::can

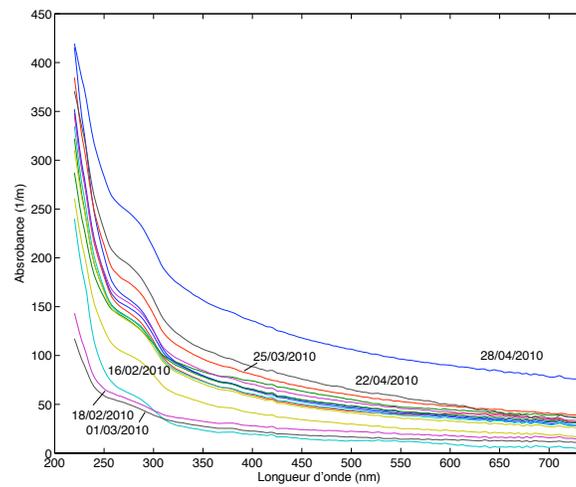


Fig. 4.8 – Spectres d'absorbance bruts

---

La figure 4.8 montre les spectres d'absorbance des quinze échantillons. On observe que les spectres couvrent une plage étendue de valeurs d'absorbance, et ce quel que soit la longueur d'onde considérée. Les parties de spectres mesurées dans le domaine spectral du visible sont bruitées. Le bruit est vraisemblablement dû à l'interaction du rayonnement avec les MES.

#### **4.3.1.2.a Classification des spectres d'absorbance**

L'analyse vise à classer les spectres d'absorbance suivant leur forme. En effet, les formes de spectre caractérisent la matrice des eaux usées d'un point de vue qualitatif Thomas (2007); Vaillant *et al.* (2002).

L'observation des spectres d'absorbance bruts (cf. FIG. 4.8) permet d'identifier trois spectres d'absorbance atypiques (05/02,16/02 et 25/03) qui croisent les spectres voisins.

L'étude qualitative des spectres d'absorbance est facilitée par la normalisation des spectres qui a pour objet d'éliminer l'effet d'échelle apporté par les concentrations. Plusieurs normalisations sont possibles. La normalisation choisie consiste pour chaque spectre à diviser les valeurs d'absorbance par l'aire totale du spectre. Tous les spectres ont ainsi une aire identique. Les spectres normalisés sont présentés en FIG. 4.9. L'échantillon du 16/02 se distingue nettement des autres échantillons.

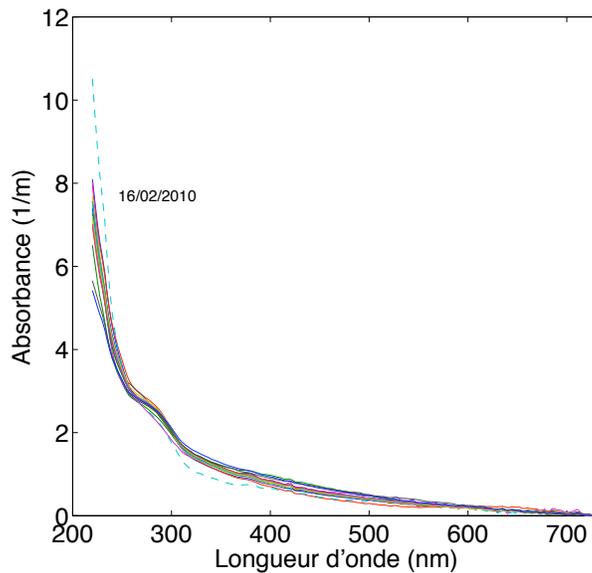


Fig. 4.9 – Spectres d'absorbance normalisés

Toutefois, la superposition de l'ensemble des autres spectres ne facilite pas l'analyse de données. Nous avons donc réalisé une analyse en composantes principales menée sur les spectres normalisés. Le pourcentage de l'inertie capté par les premiers axes principaux est de 99,1%. La FIG. 4.10 montre la projection des échantillons sur les deux premiers axes principaux. L'ACP a été associée à une classification effectuée par recherche des k plus proches voisins (cf. FIG. 4.11). Cette analyse permet de regrouper les spectres normalisés au sein de classes uniformes et fait ressortir les spectres atypiques.

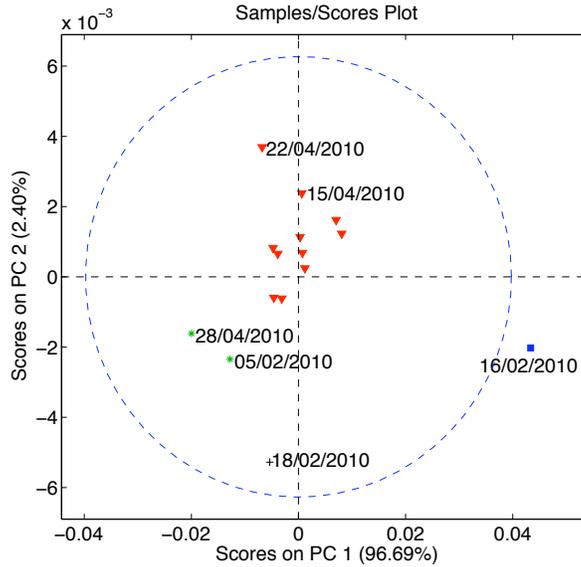


Fig. 4.10 – Analyse en composantes principales des spectres d'absorbance normalisés : représentation des individus sur le plan 1-2

La figure 4.10 illustre la variabilité des formes de spectres. Quatre classes de spectres peuvent ainsi être circonscrites. La classe principale comporte onze échantillons. Les quatre échantillons restants sont regroupés au sein de trois classes. L'échantillon du 16/02/2010 se distingue encore, ce qui est normal puisque les analyses précédentes nous avaient déjà permis de le repérer. Les spectres du 28/04/2010 et du 05/02/2010 ont été regroupés au sein d'une même classe. Rappelons que l'échantillon du 28/04/2010 a été exclu de la collection d'échantillonnage en raison d'une incohérence avec la mesure spectrophotométrique de contrôle. Enfin, le spectre du 18/02/2010 est identifié comme faisant partie d'une classe différente.

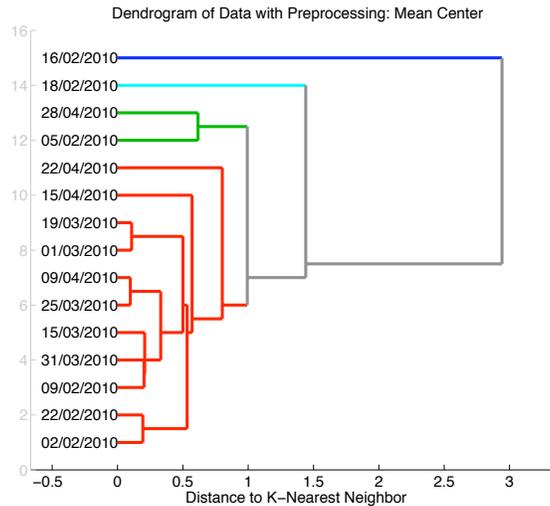


Fig. 4.11 – Dendrogramme des spectres d'absorbance normalisés

#### 4.3.1.2.b Lien entre classification des spectres et les paramètres de pollution

L'analyse de données effectuée aux paragraphes précédents a permis de classifier les échantillons suivant la forme de leur spectre, sans tenir compte des valeurs des paramètres de pollution.

L'objectif de ce paragraphe est de montrer qu'un étalonnage portant sur des spectres aux formes similaires conduit à de meilleurs résultats mais ne prend pas en compte la variabilité des spectres et des conditions environnementales.

Pour illustrer ce propos, nous avons éliminé de la base de données les échantillons du 05/02/2010, 16/02/2010, 18/02/2010 et 28/04/2010 et établi une régression linéaire simple  $DCO=f(Abs)$ , à partir de longueur d'onde présentant le coefficient de détermination le plus élevé avec la DCO. Les échantillons du 05/02/2010 et du 18/02/2010 correspondent à des échantillons de temps de pluie qui ont été exclus de la classe dominante.

Les résultats sont présentés en FIG. 4.12 et 4.13. La régression a été effectuée à partir de longueur d'onde de 607,5 nm avec un coefficient de détermination  $r^2$  de 0,977. Lorsque la régression linéaire est effectuée sur l'ensemble des échantillons, le coefficient  $r^2$  n'est plus que de 0,812 (cf. paragraphe 4.3.2.1.a).

---

Cette démarche d'élimination d'échantillons atypiques n'est acceptable que :

- s'il est possible d'effectuer un étalonnage pour chaque classe de spectres ;
- ou si le domaine de validité du modèle prédictif est précisé ;

La logique d'élimination d'échantillons atypiques émanant des articles de Langergraber *et al.* (2003); Rieger *et al.* (2006) conduit à améliorer les résultats de l'étalonnage sur un jeu de données restreint, mais n'apporte pas de précision sur la validité des concentrations prédites pour de nouveaux échantillons présentant des caractéristiques différentes du jeu d'étalonnage.

La solution idéale consisterait donc à disposer d'une base de données de taille suffisante permettant de réaliser un étalonnage pour chaque classe de spectres. Dans notre cas, ne disposant que de quatorze échantillons, nous avons choisi de réaliser l'étalonnage sur l'ensemble du jeu de données.

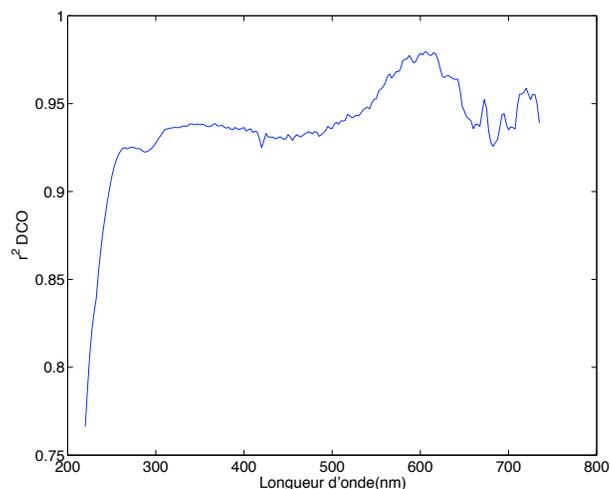


Fig. 4.12 – Coefficient de détermination  $r^2$  entre la DCO et l'absorbance pour chaque longueur d'onde

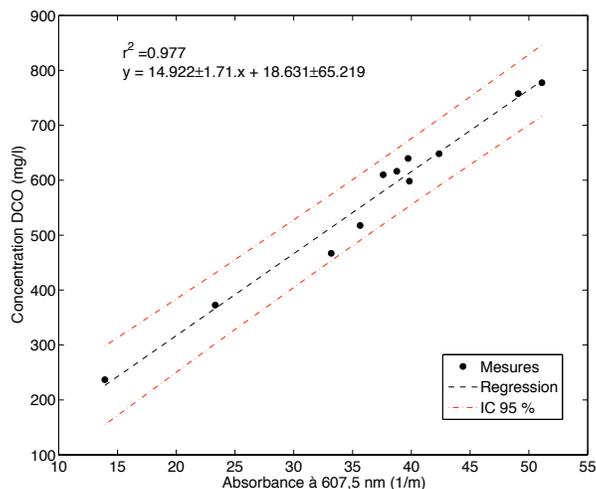


Fig. 4.13 – Régression linéaire simple DCO=f(Abs) réalisée sur la classe de spectres principale

### 4.3.2 Étalonnage et validation croisée (étape 4)

Ce paragraphe présente les résultats de l'étalonnage du spectro::lyser pour les mesures de concentrations de DCO et MES obtenues avec sept méthodes différentes.

#### 4.3.2.1 DCO

##### 4.3.2.1.a Régression linéaire simple (1)

Les résultats de l'étalonnage sont présentés dans le tableau 4.5. Les minimum des fonctions de coût RMSEC et RMSECV sont obtenus à 565 nm dans la partie visible du rayonnement électromagnétique (cf. FIG. 4.14). Ces résultats sont surprenants compte-tenu que la DCO est habituellement prédite à partir de l'absorbance à la longueur d'onde de 254 nm. Les formes des fonctions RMSEC et RMSECV (cf. FIG. 4.14) semblent indiquer que la prédiction de la DCO est davantage influencée par la diffusion du rayonnement électromagnétique par des particules en suspension que par l'absorption moléculaire du rayonnement électromagnétique dans l'UV. En effet, les molécules organiques n'absorbent pas le rayonnement dans ce domaine

spectral. Cette hypothèse est corroborée par une corrélation très importante entre la DCO et les MES ( $r^2=0,882$ ).

La régression  $DCO=f(Abs_{252,5\text{ nm}})$  est présentée en FIG. 4.15. Le coefficient de détermination  $r^2$  de la régression est de 0,938, ce qui est excellent.

Modèle	Longueur d'onde	RMSEC (mg/l)	RMSECV (mg/l)
RLS1	565 nm	40	51

Tab. 4.5 – Résultats de l'étalonnage avec une régression linéaire simple

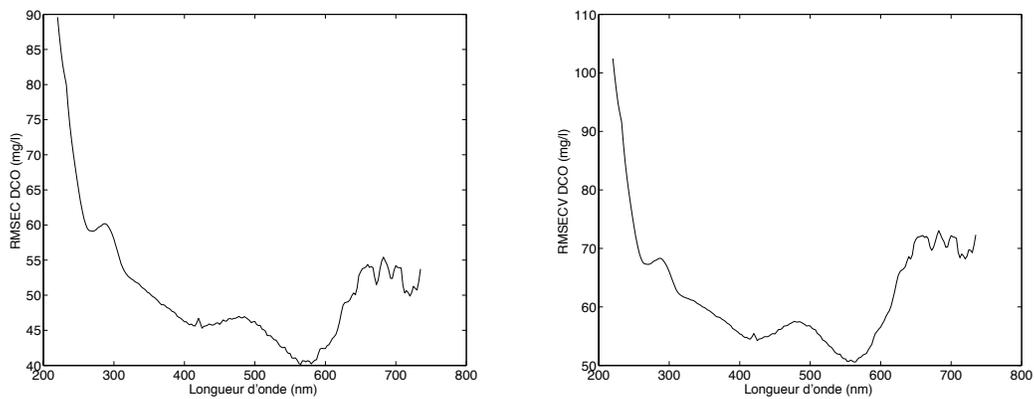


Fig. 4.14 – Valeur des fonctions de coût RMSEC et RMSECV en fonction de la longueur d'onde

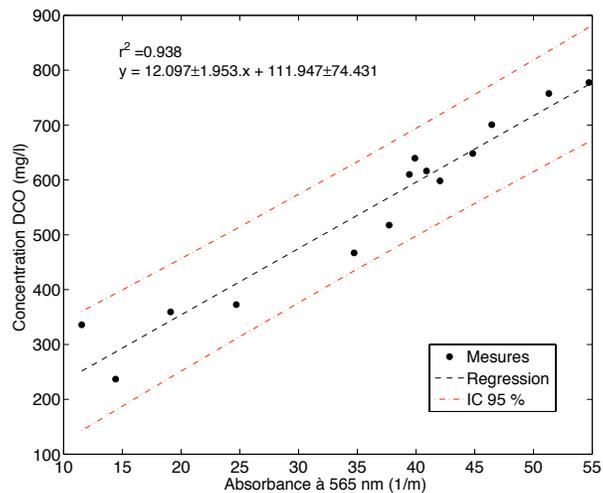


Fig. 4.15 – Concentration en DCO en fonction de l’absorbance à 565 nm

#### 4.3.2.1.b Calage global et local s::scan (2 et 3)

Le calage global fourni par la société s::scan s’avère inadapté pour la mesure des concentrations en DCO. En effet, la fonction RMSEC est de 156 mg/l soit plus du triple de la valeur obtenue avec une régression linéaire simple. Le calage local permet d’améliorer la qualité de la prédiction, mais les résultats sont toujours en léger retrait par rapport à la régression linéaire simple, les fonctions RMSEC et RMSECV étant égales à 60 et 69 mg/l.

Modèle	Longueur d’onde	RMSEC (mg/l)	RMSECV (mg/l)
C. global	inconnu	156	-
C. local	inconnu	60	69

Tab. 4.6 – Résultats de l’étalonnage avec calage global et local s::scan

#### 4.3.2.1.c Modèle PLS local (4)

Le modèle PLS local identifié utilise l’intégralité du domaine spectral mesuré (220 à 735 nm). Le minimum de la fonction RMSECV (29 mg/l) est obtenu avec un modèle à cinq

composantes. La prédiction des concentrations en DCO est meilleure qu'avec une régression linéaire simple, mais ceci se fait au prix d'une augmentation importante de la dimension du modèle.

Modèle	Longueur d'onde (nm)	Nbre composantes	RMSEC (mg/l)	RMSECV (mg/l)
PLS local	220-735	5	10	28

Tab. 4.7 – Résultats de l'étalonnage avec le modèle PLS local

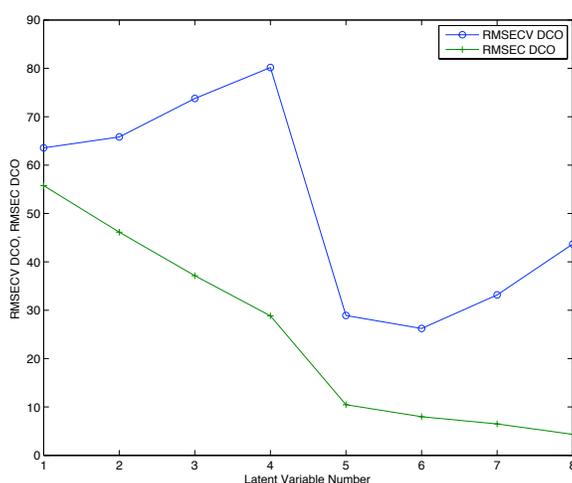


Fig. 4.16 – Valeur des fonctions de coût RMSEC et RMSECV pour le modèle PLS local

#### 4.3.2.1.d Modèle OPPPLS (5)

Un modèle OPPPLS a été identifié sur la base de l'observation de la variation de la fonction RMSECV en fonction du nombre de variables incluses dans le modèle (cf. FIG. 4.17). Les caractéristiques du modèle sont récapitulées dans le TAB. 4.14.

Les erreurs de prédiction obtenues avec ce modèle sont très faibles. Cinq composantes ont été retenues, le gain obtenu par l'ajout d'une sixième composante étant négligeable. Par ailleurs, les erreurs de prédiction par rapport à des modèles simplifiés ne comprenant qu'une ou deux

composantes sont beaucoup plus faibles. La fonction RMSECV vaut respectivement 57 et 62 mg/l pour des modèles à une et deux composantes et ne vaut plus que 21 mg/l pour un modèle à cinq composantes. Nous devons toutefois vérifier que ces bons résultats ne se manifestent pas en partie par une augmentation du bruit sur le signal de concentrations mesuré in-situ.

Les longueurs d'ondes sélectionnées semblent pertinentes. La plage spectrale est étendue (cf. FIG. 4.18) et inclut les longueurs d'ondes proches de 254 nm où des corrélations élevées entre DCO et absorbance sont habituellement constatées. Elle reprend par ailleurs le domaine spectral où les erreurs de calage et de validation croisée calculées avec la régression linéaire simple étaient les plus faibles (cf. FIG. 4.14).

Modèle	Longueur d'onde (nm)	Nbre de composantes	RMSEC (mg/l)	RMSECV (mg/l)
OPPPLS1	192 l.o.. - cf. fig. 4.18	5	13	21

Tab. 4.8 – Résultats de l'étalonnage avec la régression OPPPLS

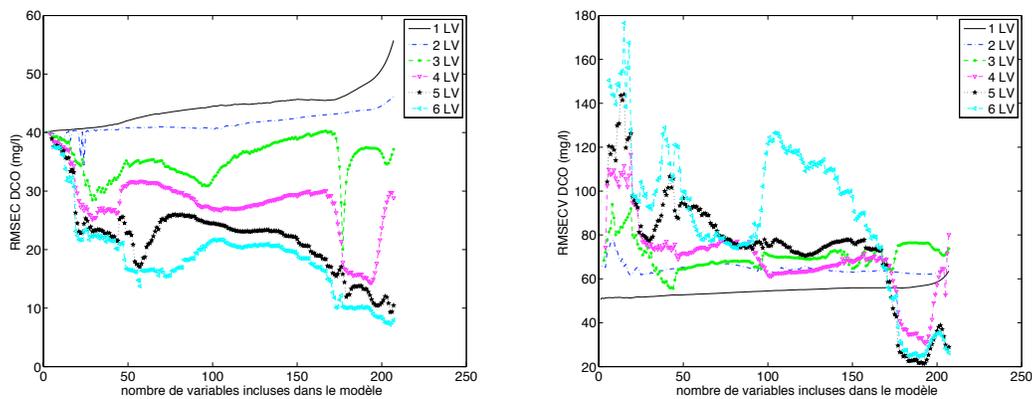


Fig. 4.17 – Valeur des fonctions de coût RMSEC et RMSECV avec la régression OPPPLS

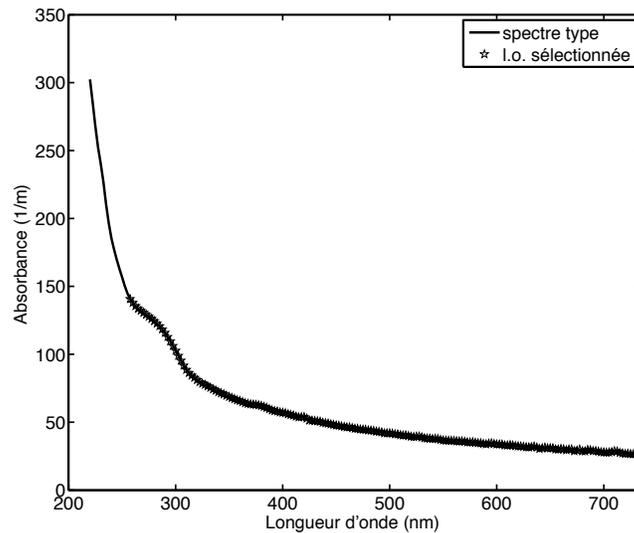


Fig. 4.18 – Longueurs d'onde sélectionnées dans le modèle OPPPLS1

#### 4.3.2.1.e Modèle fbiPLS (6)

Trois modèles fbiPLS ont été identifiés. Leurs caractéristiques sont détaillées dans le TAB. 4.9. Les erreurs de prédiction en calage et validation croisée sont inférieures à celles obtenues avec le modèle OPPPLS1.

Les modèles sélectionnés permettent d'obtenir des prédictions de meilleure qualité avec des dimensions de modèle inférieures. Le modèle fbiPLS3 est à ce titre intéressant : il ne comprend que trois composantes et les erreurs de prédiction sont du même ordre de grandeur que celles obtenues avec le modèle OPPPLS1. Il constitue le meilleur compromis entre qualité de prédiction et simplicité de modèle.

La figure 4.20 illustre les valeurs de la fonction RMSECV obtenues pour chacun des intervalles. Le graphique de gauche correspond aux résultats de l'algorithme forward iPLS et celui de droite à l'algorithme backward iPLS. Les intervalles en vert correspondent aux intervalles sélectionnés. Les droites en pointillés correspondent aux valeurs de RMSECV obtenues avec des modèles PLS utilisant l'ensemble de longueurs d'ondes. Cette figure montre le gain de

performance obtenu avec une sélection de variables adéquate par rapport à l'utilisation de modèles PLS basés sur l'ensemble du domaine spectral.

Les longueurs d'ondes sélectionnées sont aussi bien situées dans l'UV que dans le visible. Elles permettent de prendre en compte l'absorption moléculaire des composés organiques ainsi que la diffusion du rayonnement électromagnétique par des particules "titrant" en DCO.

Modèle	Taille intervalles	Long. d'onde (nm)	Nbre Compos.	RMSEC (mg/l)	RMSECV (mg/l)
fbiPLS1	5	(*)	5	6	12
fbiPLS2	2	(**)	4	6	9
<b>fbiPLS3</b>	2	(**)	3	15	19

(\*) : 220-242,5 ; 295-305 ; 357,5-367,5 ; 445-455 ; 545-567,5 – (\*\*): 225-232,5 ; 300 ; 302,5 ; 335 ; 337,5 ; 560-567,5 – (\*\*\*) : 240 ; 242,5 ; 330-342,5 ; 425 ; 427,5 ; 545-552,5 ; 560 ; 562,5

Tab. 4.9 – Résultats de l'étalonnage avec la régression fbiPLS

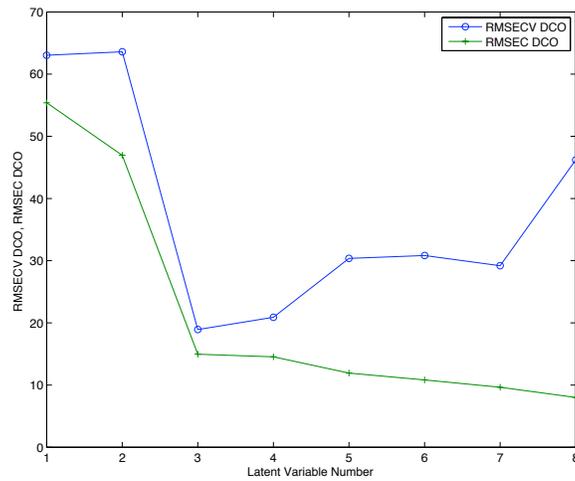


Fig. 4.19 – Valeur des fonctions de coût RMSEC et RMSECV obtenues avec la sélection de longueurs d'onde du modèle fbiPLS3

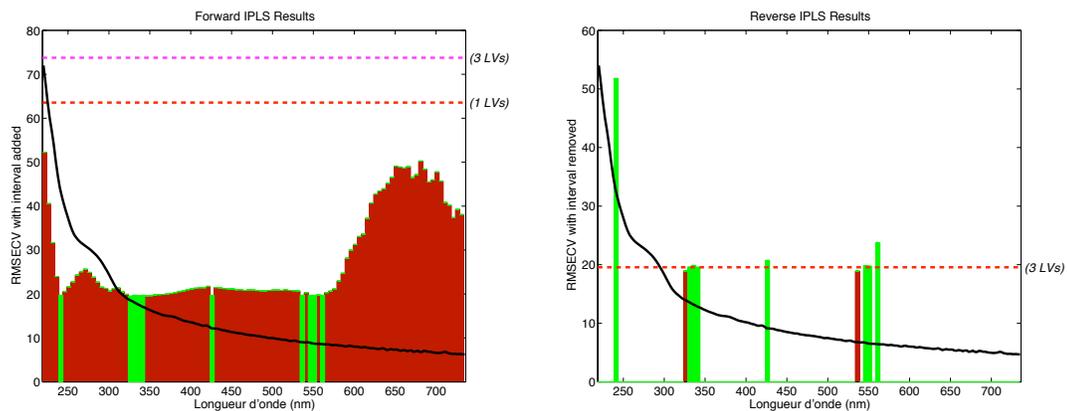


Fig. 4.20 – Représentation graphique de la sélection d’intervalles effectuée pour le modèle fbiPLS3

#### 4.3.2.1.f Modèle mwPLS (7)

Deux modèles ont été identifiés avec la méthode mwPLS (cf. TAB. 4.16). Les résultats sont en retrait par rapport à ceux obtenus avec les méthodes fbiPLS et OPP PLS.

Toutefois, les deux modèles sont parcimonieux et ne comprennent respectivement que deux et trois composantes. Ceci est dû au fait que les longueurs d’ondes sélectionnées sont contiguës et sont de ce fait très corrélées ce qui rend la réduction de la dimension du modèle plus aisée.

Le couplage des deux modèles, et par conséquent des deux intervalles, permet de réduire l’erreur de de calage à 27 mg/l. Ce couplage ne pourra être validé que lorsque nous disposerons d’échantillons supplémentaires permettant d’effectuer une validation simple du modèle. Cependant, les plages de longueurs d’onde du modèle  $mwpls_{1+2}$  sont proches de celles du modèle fbiPLS3 et semblent donc pertinentes.

Modèle	Taille intervalles (l.o.)	Long. d'onde (nm)	Nbre Compos.	RMSEC (mg/l)	RMSECV (mg/l)
mwPLS1	40	262,5-360	3	30	47
mwPLS2	40	472,5-570	2	35	47
mwPLS1+2*	40	262,5-360 472,5-570	-	27	-

(\*) :  $DCO = 0,5578.DCO_{mwpls1} + 0,4422.DCO_{mwpls2}$

Tab. 4.10 – Résultats de l'étalonnage avec la régression mwPLS

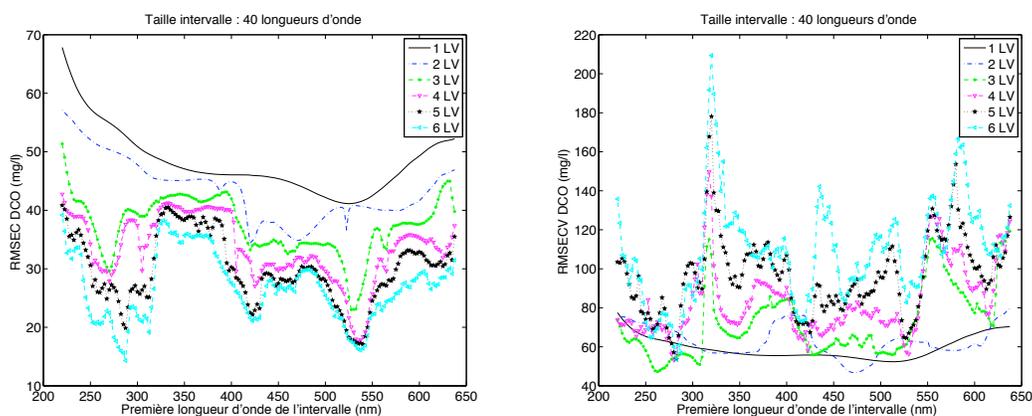


Fig. 4.21 – Valeur des fonctions de coût RMSEC et RMSECV avec la régression mwPLS

#### 4.3.2.1.g Conclusion sur les prédictions DCO

Les principales conclusions résultant des sept procédures d'étalonnage du spectrophotomètre pour la prédiction des concentrations en DCO sont les suivantes :

- les modèles donnant les meilleures prédictions sont les modèles fbiPLS (1 à 3) et OPPPLS1. La sélection de longueurs d'ondes apporte donc une réelle plus value avec des erreurs de validation croisée inférieures de 58 à 82 % à l'erreur obtenue avec le modèle RLS. Les longueurs d'ondes sélectionnées sont situées dans l'UV (entre 240 et 360 nm) et dans le visible (entre 425 et 570 nm). Elles correspondent à l'absorption de molécules organiques dans l'UV et à la diffusion du rayonnement par des particules titrant en DCO.

- 
- Les modèles mwpls sont moins performants que les modèles fbiPLS et OPPPLS avec des erreurs de validation croisée du même ordre de grandeur que le modèle RLS. Les longueurs d'ondes sélectionnées sont cette fois situées dans l'UV ou le visible. Le couplage des deux modèles mwPLS est plus performant en calage. Sa validation ne pourra être effectuée qu'ultérieurement sur des données complémentaires.
  - Les algorithmes fbiPLS et mwPLS conduisent à performance égale, à des modèles plus parcimonieux que ceux identifiés avec l'algorithme OPPPLS avec seulement deux ou trois composantes ;
  - Le modèle PLS local sans sélection de longueurs d'ondes est également performant avec des erreurs de validation croisée légèrement supérieures à celles obtenues avec les modèles fbiPLS et OPPPLS.
  - Les résultats obtenus avec les modèles "calage local s::can" et RLS sont nettement en retrait.
  - Enfin, le calage global s::can ne peut pas être utilisé pour la mesure des concentrations en DCO, les erreurs de calage étant trop importantes (RMSEC = 156 mg/l) ;

#### **4.3.2.2 MES**

La méthodologie a été reconduite pour l'étalonnage d'un loi  $MES=f(\text{Absorbance})$ .

##### **4.3.2.2.a Régression linéaire simple (1)**

Les résultats de l'étalonnage par régression linéaire simple sont présentés dans le TAB. 4.11. La longueur d'onde de 550 nm a été sélectionnée. Les régressions linéaires simples effectuées pour la prédiction des concentrations en DCO et MES sont basées sur des longueurs d'onde voisines. Ceci semble confirmer que la DCO est majoritairement sous forme particulaire.

Le coefficient de détermination de la régression RLS1 est égal à 0,849 (cf. FIG. 4.23). Cette valeur est satisfaisante. Elle est du même ordre de grandeur que celle obtenue dans la publication de Langergraber *et al.* (2003) ( $r^2=0,83$ ).

Modèle	Longueur d'onde	RMSEC (mg/l)	RMSECV (mg/l)
RLS	550 nm	36	43

Tab. 4.11 – Résultats de l'étalonnage avec une régression linéaire simple

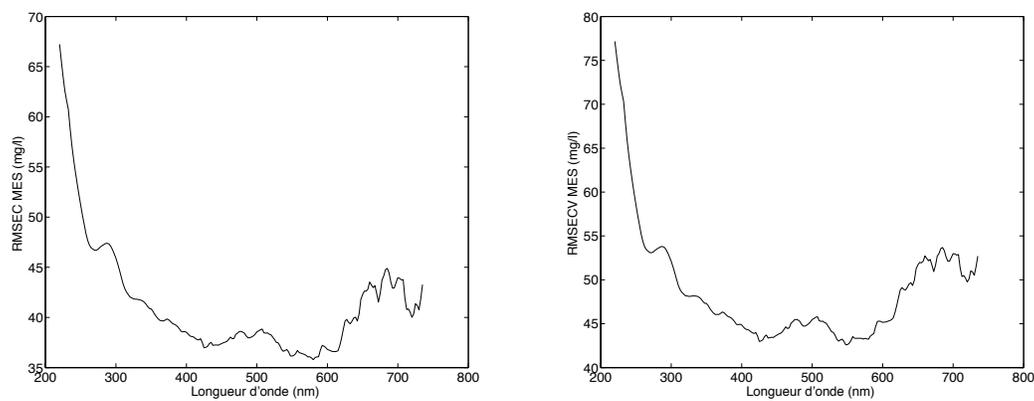


Fig. 4.22 – Valeur des fonctions de coût RMSEC et RMSECV en fonction de la longueur d'onde

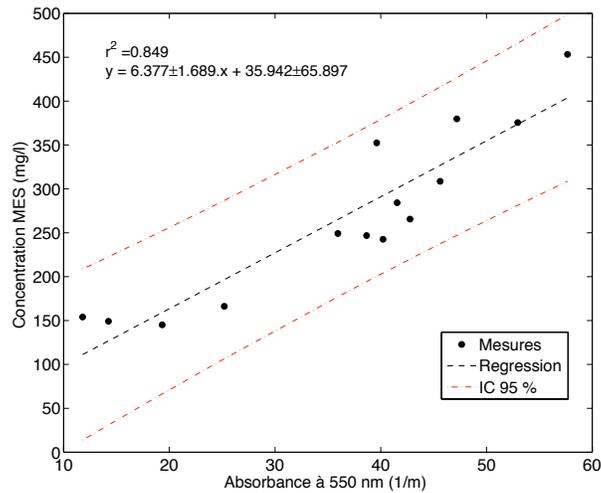


Fig. 4.23 – Concentration en MES en fonction de l’absorbance à 550 nm

#### 4.3.2.2.b Calage global et local scan (2 et 3)

Les résultats de l’étalonnage sont présentés dans le TAB. 4.12. Les valeurs de RMSEC et RMSECV sont légèrement supérieures à celles obtenues avec la régression linéaire simple.

Notons que l’échantillon du 16/02/2010 contribue à hauteur de 36 et 32 % des erreurs de calage et de validation croisée. La mauvaise prédiction de la concentration en MES est due à la forme particulière du spectre qui est atypique par rapport à celle des autres spectres. On constate en effet que ce spectre a l’absorbance dans le visible la plus faible alors que les échantillons du 18/02/2010 et du 01/03/2010 ont des concentrations en MES plus faibles.

Modèle	Longueur d’onde	RMSEC (mg/l)	RMSECV (mg/l)
C. globale	inconnu	115	-
C. locale	inconnu	41	44

Tab. 4.12 – Résultats de l’étalonnage avec calage global et local scan

#### 4.3.2.2.c Modèle PLS local (4)

Les caractéristiques du modèle PLS local identifié pour l'étalonnage d'une loi  $MES=f(Abs)$  sont présentées dans le TAB. 4.13. Les valeurs de RMSEC et RMSECV sont du même ordre de grandeur que celles obtenues avec le calage local ou la régression linéaire simple.

Modèle	Longueur d'onde (nm)	Nbre composantes	RMSEC (mg/l)	RMSECV (mg/l)
PLS local	220-735	2	37	43

Tab. 4.13 – Résultats de l'étalonnage avec régression PLS locale utilisant l'intégralité du spectre d'absorbance

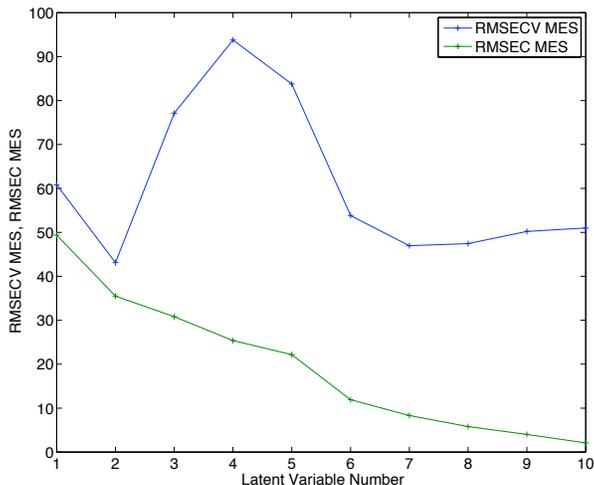


Fig. 4.24 – Valeur des fonctions de coût RMSEC et RMSECV pour le modèle PLS local

#### 4.3.2.2.d Modèle OPPPLS (5)

Deux modèles PLS ont été identifiés en utilisant l'algorithme OPPPLS. Leurs caractéristiques sont détaillées dans le TAB. 4.14. Les valeurs de RMSEC et RMSECV sont bien inférieures à celles obtenues avec les autres méthodes avec cependant un nombre de composantes supérieur ou égal à cinq. Il semble plus prudent d'utiliser le modèle OPPPLS1, le second modèle

apportant des gains de performance limités et comportant en outre une composante supplémentaire.

Le modèle OPPPLS1 est basé sur la sélection de 94 longueurs d'ondes (cf. FIG. 4.26) situées dans le domaine visible du rayonnement électromagnétique qui correspondent à des couleurs allant du violet (400 nm) à l'orange (600 nm). Les résultats de la sélection sont cohérents avec la longueur d'onde sélectionnée pour la régression linéaire simple.

Modèle	Longueur d'onde (nm)	Nbre de composantes	RMSEC (mg/l)	RMSECV (mg/l)
OPPPLS1	94 l.o.*	5	11	30
OPPPLS2	100 l.o.	6	9	24

(\*) : cf. FIG. 4.26

Tab. 4.14 – Résultats de l'étalonnage avec la régression OPPPLS

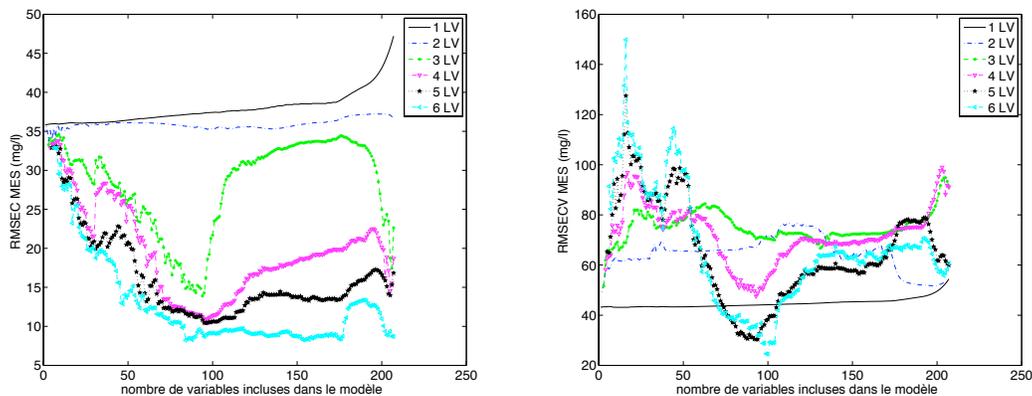


Fig. 4.25 – Valeur des fonctions de coût RMSEC et RMSECV avec la régression OPPPLS

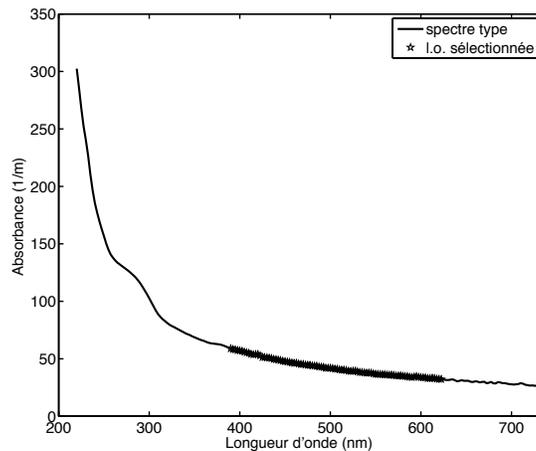


Fig. 4.26 – Longueurs d'onde sélectionnées avec le modèle OPPPLS1

#### 4.3.2.2.e Modèle fbiPLS (6)

Trois modèles fbiPLS ont été identifiés. Leurs caractéristiques sont récapitulées dans le TAB. 4.15.

Les erreurs de prédiction des deux premiers modèles sont légèrement supérieures à celles obtenues avec les modèles OPPLS1 et OPPPLS2. Cependant, ces deux modèles ne comprennent respectivement que trois et deux composantes. Ils seront donc préférés aux modèles identifiés avec l'algorithme OPPPLS. Le modèle fbiPLS2 constitue un très bon compromis entre qualité de prédiction et simplicité du modèle. Sa plage de longueur d'onde est en outre cohérente au regard de celle sélectionnée avec le modèle OPPPLS1 (cf. FIG. 4.28).

Le modèle fbiPLS3 donne d'excellents résultats avec des erreurs de prédiction RMSEC et RMSECV égales à 7 et 10 mg/l. Le nombre de composantes est égal à cinq comme pour le modèle OPPPLS1. La sélection de longueurs d'ondes effectuée nous laisse toutefois perplexes. En effet, les longueurs d'ondes sont réparties sur l'ensemble du spectre, aussi bien dans le domaine de l'UV que celui du visible. Une interprétation physique de cette sélection est difficile. Par ailleurs, l'erreur de validation croisée est inférieure à la moyenne des écarts types des répliqués de pesage des MES ( $s_{moy}=14,6$  mg/l). Les bons résultats sont donc davantage liés à une surparamétrisation du modèle qu'à une sélection adéquate d'intervalles.

Cette surparamétrisation pourra être vérifiée en observant les mesures de concentrations effectuées in-situ avec le spectro::lyser.

Modèle	Taille intervalles	Long. d'onde (nm)	Nbre Compos.	RMSEC (mg/l)	RMSECV (mg/l)
fbiPLS1	5	(*)	3	28	33
fbiPLS2	5	(**)	2	32	36
fbiPLS3	3	(***)	5	7	10

(\*) : 432,5-467,5 – (\*\*): 445-480 – (\*\*\*) : 242,5-247,5; 340-345; 437,5-442,5; 467,5-480; 580-585

Tab. 4.15 – Résultats de l'étalonnage avec la régression fbiPLS

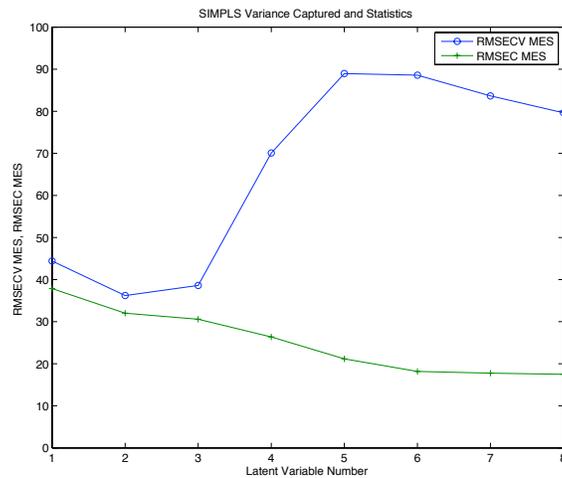


Fig. 4.27 – Valeur des fonctions de coût RMSEC et RMSECV obtenues avec la sélection de longueurs d'onde du modèle fbiPLS2

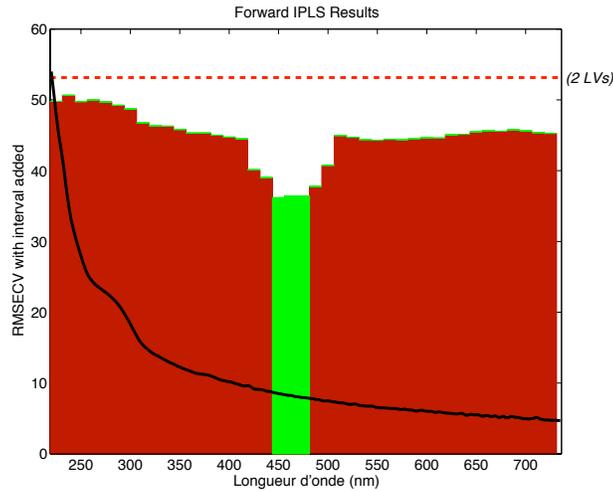


Fig. 4.28 – Représentation graphique de la sélection d'intervalles effectué pour le modèle fbiPLS2

#### 4.3.2.2.f Modèle mwPLS (7)

Un modèle PLS a été identifié avec la méthode de sélection de longueurs d'ondes mwPLS. Ses caractéristiques sont détaillées dans le TAB. 4.16.

Ce modèle est très proche du modèle fbiPLS2 : il comprend deux composantes et la plage de longueurs d'onde sélectionnée est légèrement moins étendue.

Modèle	Taille intervalles (l.o.)	Long. d'onde (nm)	Nbre Compos.	RMSEC (mg/l)	RMSECV (mg/l)
mwPLS1	10	450-472,5	2	31	36

Tab. 4.16 – Résultats de l'étalonnage avec la régression mwPLS

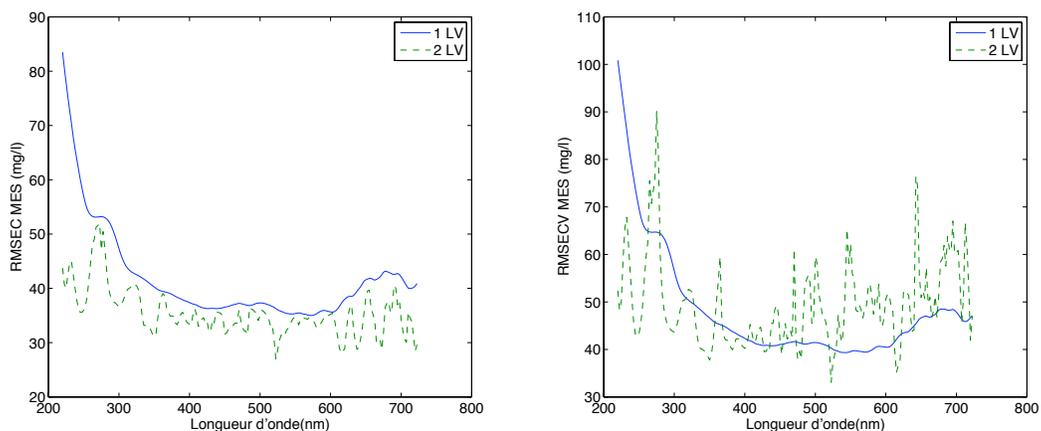


Fig. 4.29 – Valeur des fonctions de coût RMSEC et RMSECV avec la régression mwPLS

#### 4.3.2.2.g Conclusion sur les prédictions de MES

Comme pour la prédiction des concentrations en DCO, les modèles PLS intégrant un algorithme de sélection de longueurs d'onde sont les plus performants : les erreurs de validation croisée sont inférieures à 36 mg/l, ce qui est satisfaisant. Les trois meilleurs modèles sont les modèles mwPLS1, fbiPLS2 et OPPPLS.

Les algorithmes fbiPLS et mwPLS permettent l'identification de modèles plus simples, i.e. comprenant un nombre inférieur de dimensions, que l'algorithme OPPPLS. Ceci est lié à la sélection d'une plage de longueurs d'onde plus réduite ne contenant pas d'information inutile pour la prédiction des concentrations en MES. Les longueurs d'ondes sélectionnées sont situées dans le domaine du visible entre 430 et 630 nm.

Les prédictions réalisées avec les trois autres modèles (calage local scan, modèle RLS et PLS local) sont légèrement en retrait (d'environ 10 mg/l) et donnent des résultats équivalents.

#### 4.3.2.3 Conclusion

Sept méthodes d'étalonnage ont été comparées pour la mesure de concentrations de DCO et MES par spectrophotométrie UV/Visible, trois d'entre elles comprenant un algorithme de sélection de variables.

---

Nous avons montré que les algorithmes de sélection de variables permettent d'améliorer considérablement les résultats des modèles prédictifs de concentration en DCO et MES. La diminution de l'erreur de validation croisée atteint respectivement 82 et 43 % pour la DCO et les MES par rapport à une régression linéaire simple. En effet, bien que la plus value des méthodes de sélection de variables ait été mise en valeur en spectroscopie infrarouge ou proche infrarouge pour la détermination de concentrations de composés chimiques (Jiang *et al.*, 2002; Abrahamsson *et al.*, 2003; Leardi et Norgaard, 2004; Norgaard *et al.*, 2000; Zou *et al.*, 2007), elle n'avait fait l'objet jusque là que d'une étude réalisée par Torres et Bertrand-Krajewski (2008) dans le domaine de la détermination des concentrations de paramètres de pollution globaux dans les eaux usées.

Nous avons en outre proposé l'utilisation de deux algorithmes de sélection de longueurs d'ondes, dont un est nouveau :

- l'algorithme forward-backward iPLS ;
- et l'algorithme moving-window PLS.

Dans notre application, les modèles fbiPLS sont aussi performants que les modèles OPPPLS, aussi bien pour la prédiction des concentrations en DCO que des concentrations en MES. Les modèles mwPLS donnent d'excellents résultats pour les MES mais s'avèrent moins adaptés à la prédiction des concentrations en DCO. Les meilleurs modèles sont en effet basés sur l'utilisation de plages spectrales situées dans l'UV **et** le visible, permettant ainsi de mieux prendre en compte la contribution des MES aux concentrations en DCO. Les performances des modèles mwPLS devraient toutefois être améliorées par couplage de modèles établis sur plages spectrales différentes.

Par ailleurs, les méthodes fbiPLS et mwPLS conduisent à l'identification de modèles plus parcimonieux que l'algorithme OPPPLS. Cette différence provient vraisemblablement d'une sélection de longueurs d'ondes plus restreinte permettant de davantage éliminer l'information qui ne contribue pas à la prédiction des concentrations.

La parcimonie des modèles identifiés est un point clef pour le suivi en continu des concentrations en DCO et MES, et l'utilisation éventuelle des signaux de concentrations pour une gestion en temps-réel de composantes du système d'assainissement. En effet, il convient d'éviter toute surparamétrisation qui se caractériserait par des oscillations à haute fréquence au niveau du signal de concentrations.

---

### **4.3.3 Domaine de validité des modèles (étape 5)**

Cette partie a pour objet de déterminer les conditions d'utilisation des modèles prédictifs établis à partir de la collection d'étalonnage.

#### **4.3.3.1 Domaine de validité des concentrations prédites**

Les valeurs de DCO du jeu de données d'étalonnage sont comprises entre 237 et 778 mg/l. Les concentrations en MES sont quant à elles comprises entre 141 et 631 mg/l. Ces intervalles constituent le domaine de validité des concentrations prédites par les modèles. En dehors de ces intervalles il est possible que les modèles donnent des valeurs de concentrations aberrantes.

#### **4.3.3.2 Domaine de validité des modèles basé sur la forme des spectres d'absorbance**

Les modèles ne peuvent être utilisés que sur des formes de spectres possédant les mêmes caractéristiques que celles du jeu de données d'étalonnage. La forme des spectres d'absorbance est un très bon indicateur de la composition des eaux usées. En effet elle intègre la plupart des informations concernant les proportions relatives des différents constituants solubles ou particuliers. L'objectif de ce paragraphe est donc de déterminer si les formes de spectres mesurés in-situ en continu ont des caractéristiques voisines des spectres utilisés pour l'étalonnage des modèles.

La classification établie au paragraphe 4.3.1.2.a a été appliquée aux spectres d'étalonnage. Les classes des spectres mesurés in-situ ont alors été déterminées par recherche du spectre d'étalonnage plus proche voisin. Les résultats sont présentés en FIG. 4.30. Le graphique supérieur montre l'évolution de l'absorbance à 255 nm. Les couleurs correspondent à la classe de spectre identifiée. Le graphique inférieur montre l'évolution du débit moyen horaire sur la même période.

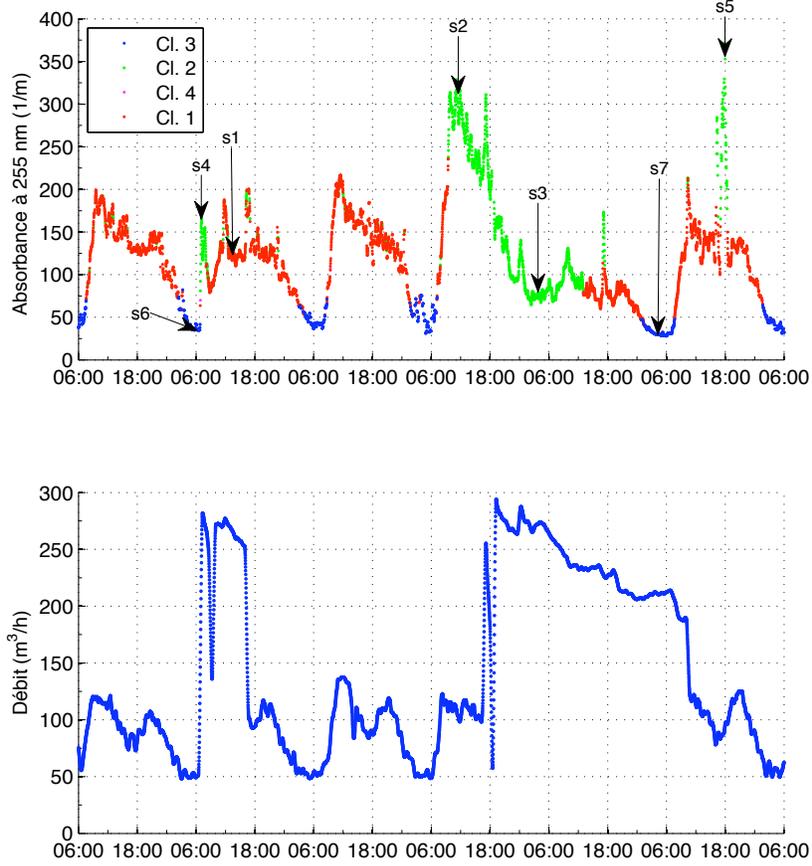


Fig. 4.30 – Classification des spectres mesurés in-situ sur la chronique de mesures d'absorbance à 255 nm

---

La FIG. 4.30 montre que les classes de spectres peuvent être rattachées à des conditions environnementales. La classe 1 représentée en rouge correspond à la classe dominante rencontrée en étalonnage. Les concentrations prédites avec cette classe de spectre devraient donc être prédites avec précision.

Les spectres de la classe 2 (représentée en vert) correspondent à des conditions de temps de pluie comme l'atteste la chronique de débit. Les formes de spectres sont particulièrement différentes de la forme de spectre dominante utilisée pour l'étalonnage des modèles. Les prédictions sont donc sujettes à caution.

Les spectres de classe 3 (représentée en bleu) correspondent aux périodes nocturnes où les absorbances les plus faibles sont mesurées. La classe 4 (représentée en vert) est observée sur les mêmes périodes mais est très peu représentée. Comme pour la classe 2, les classes 3 et 4 sont sous représentées dans la collection d'étalonnage. Les prédictions effectuées sur de nouveaux spectres appartenant à ces classes doivent donc être considérées avec précaution.

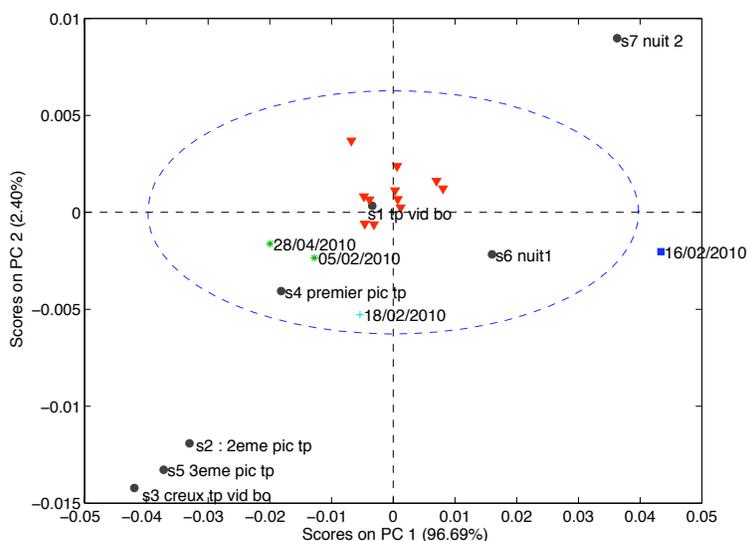


Fig. 4.31 – Représentation de huit spectres mesurés en 2008 sur le plan de projection 1-2 du modèle d'ACP des spectres normalisés d'étalonnage

Nous avons choisi sept spectres sur la chronique de mesures de 2008 (cf. FIG. 4.30, s1 à s7) et nous les avons comparés aux spectres d'étalonnage en les projetant sur le plan de

---

projection 1-2 du modèle ACP établi avec les spectres normalisés d'étalonnage. Les résultats sont présentés en FIG. 4.31. Nous observons un regroupement des individus correspondant à des conditions de temps de pluie dans le quart inférieur gauche du plan. L'individu s4 est le plus proche des individus 05/02/2010 et 18/02/2010 qui ont été prélevés par temps de pluie lors de l'étalonnage. L'individu s1 correspond à une mesure effectuée lors de la vidange du bassin d'orage. L'ACP nous montre cependant qu'il possède les mêmes caractéristiques que les échantillons de temps sec de la collection d'étalonnage. Les individus s6 et s7 correspondent à des mesures effectuées la nuit au moment où les minimums d'absorbance sont mesurés. L'ACP montre une discrimination de ces deux spectres avec des scores positifs importants sur le premier axe.

Nous avons également intégré un panel plus complet d'individus sur le plan de projection 1-2 (cf. FIG. 4.32). La différenciation des classes par projection sur le plan 1-2 est bien visible. Les spectres de temps sec, correspondant à la classe dominante de la collection d'étalonnage, sont situés dans la partie centrale du plan. Les spectres de temps de pluie sont quant à eux situés dans la partie inférieure gauche du plan. Les individus de ces deux classes sont distribués le long d'un axe oblique. Les spectres mesurés en période nocturne et correspondant à une proportion importante d'eaux claires parasites sont situés dans la partie supérieure droite du plan.

Cette approche de classification pour la validation de modèles prédictifs de concentrations utilisés in-situ est doublement intéressante. En effet, elle permet de circonscrire sur les données mesurées, les zones pour lesquelles les modèles sont applicables. Par ailleurs, elle facilite la réflexion sur la manière dont doit être complété le jeu de données d'étalonnage afin qu'il soit le plus représentatif des données rencontrées in-situ.

#### **4.3.3.3 Représentativité des concentrations du jeu d'étalonnage**

16 % des valeurs d'autosurveillance de concentrations en DCO sont supérieures à la concentration maximale de la collection d'étalonnage (cf. FIG. 4.33). Nous ne disposons par ailleurs que d'un échantillon à la concentration inférieure au percentile 20%. Le jeu de données d'étalonnage n'est pas suffisamment consistant et devra être complété avec des échantillons à faibles et fortes concentrations.

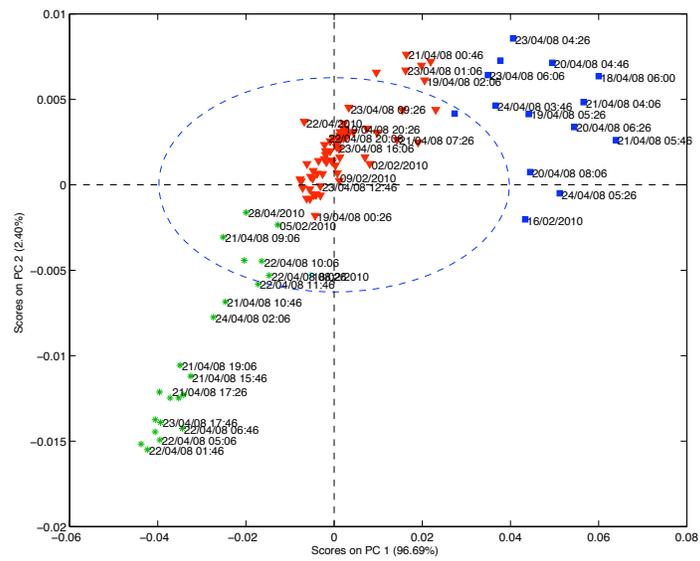


Fig. 4.32 – Représentation d'un panel de spectres mesurés en 2008 sur le plan de projection 1-2 du modèle d'ACP des spectres normalisés d'étalonnage

---

La situation est similaire pour les MES : la concentration maximale du jeu d'étalonnage correspond au percentile 84% et la concentration minimale au percentile 17%.

Les modèles prédictifs sont donc utilisables pour les conditions les plus courantes. Toutefois, un des objectifs de la mesure en continu de pollution est de pouvoir quantifier les concentrations associées à des événements extrêmes. Cet objectif ne pourra être réalisé qu'en complétant de manière adéquate le jeu de données d'étalonnage.

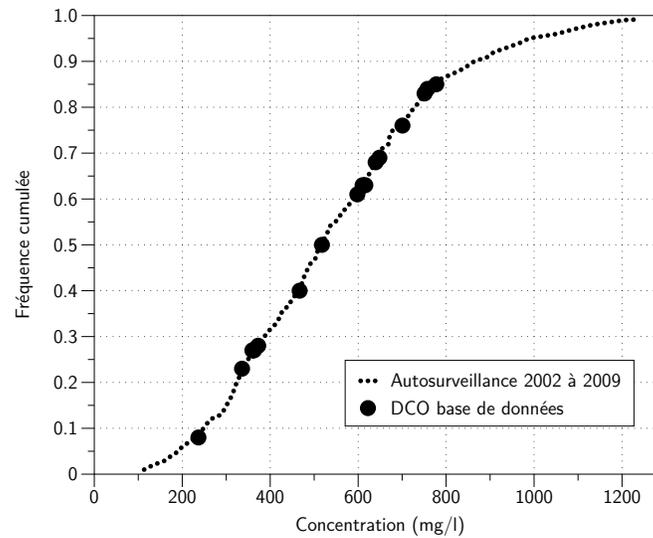


Fig. 4.33 – Percentile correspondant à chaque valeur de DCO

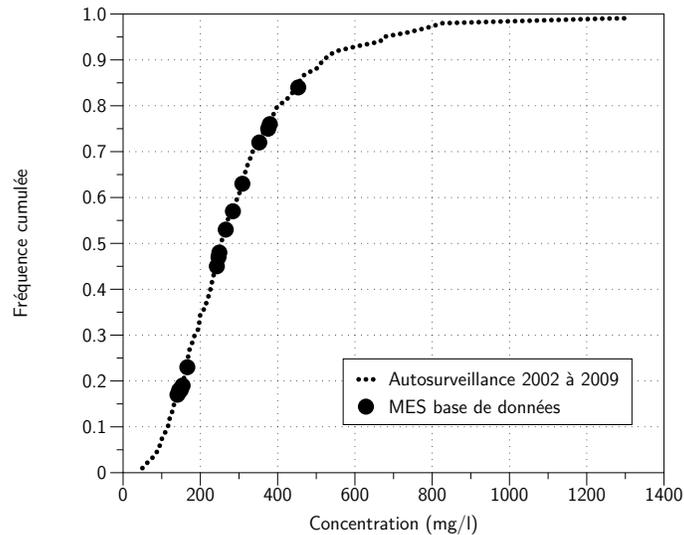


Fig. 4.34 – Percentile correspondant à chaque valeur de MES

#### 4.3.3.4 Représentativité de la base de données par rapport à la variabilité mensuelle de la qualité des influents

L'objectif de ce paragraphe est de critiquer la représentativité de la période d'échantillonnage par rapport à la variabilité de pollution observée tout au long de l'année.

Les figures 4.35 et 4.36 illustrent la variabilité mensuelle des paramètres de pollution. De janvier à avril, période correspondant aux mois où les échantillons ont été prélevés, les valeurs de DCO et MES sont sensiblement constantes. A partir de mai, ces deux paramètres connaissent des variations importantes et le rapport DCO/MES chute en dessous de 2 jusqu'au mois d'août. La moyenne mensuelle de DCO augmente à partir du mois d'août jusqu'à 800 mg/l au mois d'octobre, mois pendant lequel sont effectuées les vendanges. La figure 4.35 montre par ailleurs que les concentrations les plus élevées en DCO sont mesurées en période de vendanges avec des concentrations moyennes journalières allant jusqu'à 1700 mg/l. Un pic de MES à 380 mg/l est également observé pendant ce même mois et un deuxième pic de MES a lieu au mois de juin.

La période pendant laquelle les échantillons ont été prélevés n'est donc pas représentative de la variabilité de la pollution observée au cours d'une année. Les variations mensuelles

du rapport DCO/MES avec un pic de MES en été et une augmentation de la DCO en période de vendanges se traduisent vraisemblablement par des formes de spectres d'absorbance différentes qui ne sont pas prises en compte dans l'étalonnage. La collection d'étalonnage devra donc être complétée avec des échantillons prélevés pendant le second semestre.

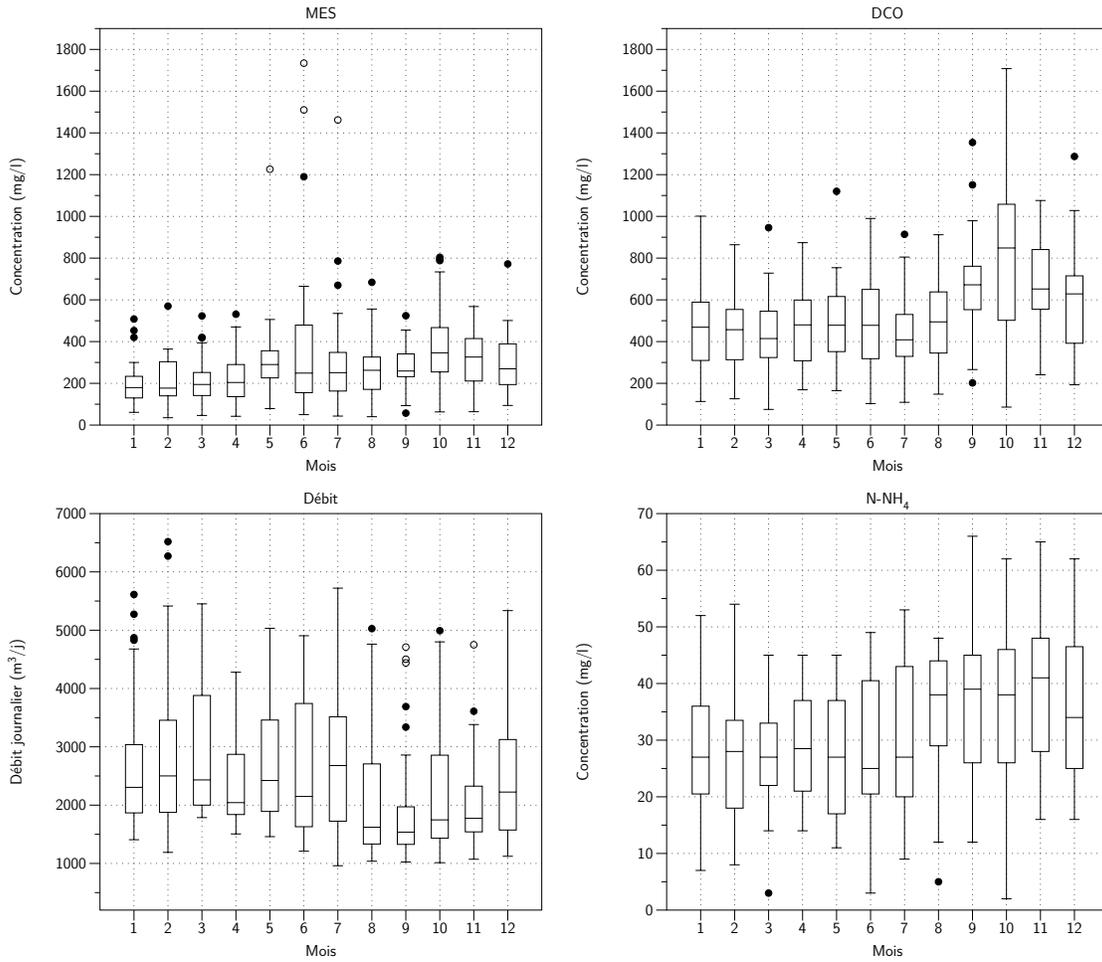


Fig. 4.35 – Variabilité mensuelle des paramètres Débit, DCO, MES et N-NH<sub>4</sub>

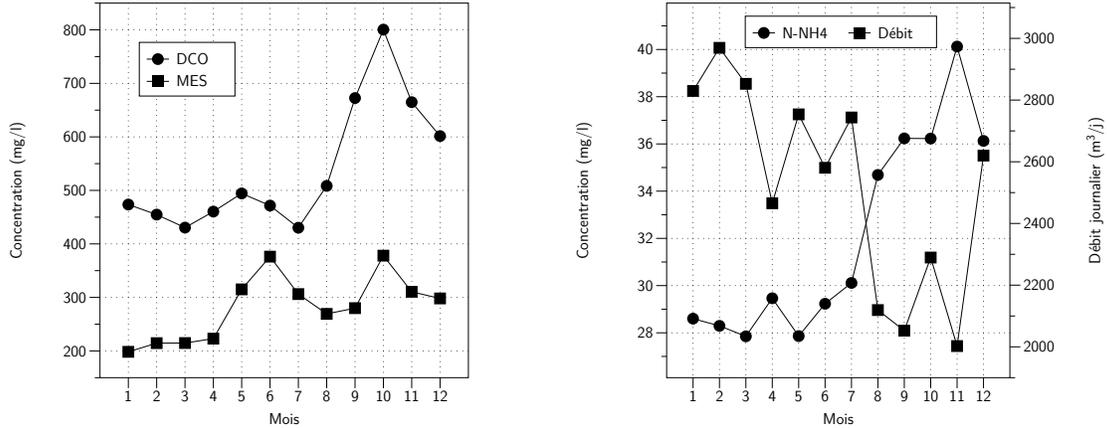


Fig. 4.36 – Variabilité des moyennes mensuelles des paramètres Débit ; DCO, MES et N-NH<sub>4</sub>

#### 4.3.3.5 Conclusion

L'étude du domaine de validité des modèles montre que le jeu de données d'étalonnage n'est pas consistant et doit être complété de manière adéquate. L'étendue des concentrations de la collection d'étalonnage est trop réduite aussi bien pour les MES que pour la DCO. Des échantillons à faibles et fortes concentrations devront donc être prélevés. La variabilité mensuelle de la qualité des eaux usées rend également nécessaire d'effectuer de nouveaux prélèvements pendant le second semestre, et surtout en période de vendanges.

Enfin, des informations très intéressantes sont apportées par la classification qualitative des spectres mesurés in-situ. Elle permet tout d'abord d'identifier les spectres correspondant à des classes sous représentées dans la collection d'étalonnage et pour lesquelles les prédictions doivent être considérées avec précaution.

Le second intérêt réside dans le lien effectué entre classe de spectre et conditions environnementales. Il apparaît ainsi que les classes sous représentées dans les données d'étalonnage correspondent à des mesures effectuées en période nocturne et par temps de pluie. La nuit, les eaux usées comprennent une proportion importante d'eaux claires parasites, modifiant ainsi la matrice des eaux usées. En temps de pluie, les spectres observés sont beaucoup plus "plats" traduisant une proportion beaucoup plus importante de MES liée à un lessivage du bassin-versant et du réseau d'assainissement.

---

Les modèles ne sont donc pas utilisables sur ces périodes. Les prochains prélèvements visant à compléter la collection d'étalonnage devront donc porter en priorité sur ces deux périodes.

#### 4.3.4 Invalidité des mesures in-situ 2010 : description du dysfonctionnement

Un dysfonctionnement du spectro::lyser a été identifié lors de son installation in-situ. Les mesures effectuées in-situ en 2010 ne sont pas exploitables. Le dysfonctionnement a vraisemblablement affecté la mesure de spectre effectué en laboratoire le 28/04/2010, comme l'atteste les mesures de contrôles effectuées avec le spectrophotomètre de laboratoire Perkin-Elmer. Cet échantillon a été enlevé de la base de données d'étalonnage (cf. §4.3.1.2).

Nous proposons de détailler la méthodologie d'identification de ce dysfonctionnement. Le spectro::lyser a été installé sur site le 18 mai 2010. Une dérive du signal affectant toutes les longueurs d'ondes a rapidement été identifiée. La FIG 4.37 montre l'évolution de l'absorbance à 255 nm sur la période allant du 20/05/2010 au 15/06/2010. La pente de la droite symbolisant la dérive du signal est de  $2,6 \text{ m}^{-1} \cdot \text{j}^{-1}$ . Le nettoyage manuel des optiques effectué le 04/06/2010 n'a eu qu'un effet très limité. La dérive ne peut donc pas être expliquée par l'obstruction des optiques du spectrophotomètre.

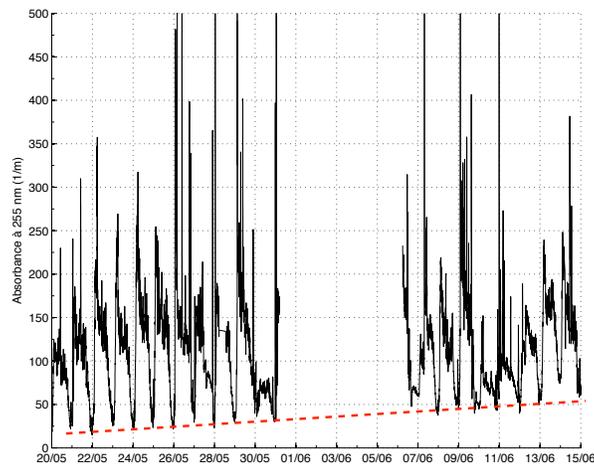


Fig. 4.37 – Dérive du signal d'absorbance à 255 nm sur les mesures in-situ effectuées en 2010

---

La FIG. 4.38 montre l'évolution de l'absorbance à 255 nm de mesures réalisées sur le même site, dans les mêmes conditions expérimentales, avec un spectro::lyser prêté par Véolia eau. Aucune dérive sur le signal d'absorbance n'est ici constatée.

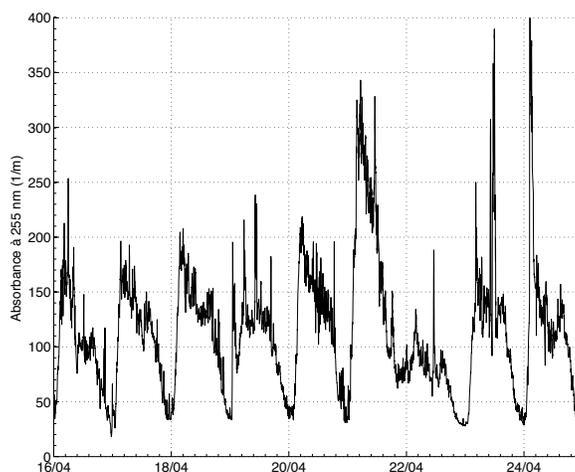


Fig. 4.38 – Signal d'absorbance à 255 nm acquis lors de mesures effectuées in-situ en 2008

Le paramétrage du nettoyage pneumatique était le même dans les deux cas de figure, avec un nettoyage effectué toutes les trois mesures.

Des mesures complémentaires en laboratoire ont permis de montrer que la dérive était due à un défaut d'étanchéité des joints entourant les optiques. Nous avons placé le spectrophotomètre dans de l'eau distillée, le nettoyage pneumatique étant désactivé, et avons constaté une augmentation de l'absorbance.

#### **4.3.5 Utilisation des modèles sur la chronique de mesures in-situ de 2008 (étape 6)**

Les mesures in-situ effectuées en 2010 n'étant pas exploitables, nous avons utilisé des spectres d'absorbance acquis lors d'une campagne de mesure effectuée en 2008 sur le même site d'étude et dans les mêmes conditions expérimentales. Le spectrophotomètre était un spectro::lyser prêté par Véolia eau et possédant les mêmes caractéristiques.

---

Même si les deux appareils possèdent les mêmes caractéristiques techniques, il est possible qu'il subsiste des différences de mesures entre les deux appareils.

#### 4.3.5.1 Exploitabilité des modèles sur des journées de temps sec : variation des concentrations et bruit des signaux

##### 4.3.5.1.a Variations journalières des absorbances

Les six modèles étudiés pour la mesure des concentrations en DCO et MES utilisent des longueurs d'ondes situées dans le domaine de l'UV et du visible. Les absorbances à 255 et 565 nm mesurées pendant la journée du 18/04/2008 ont donc été tracées en FIG. 4.39 afin d'observer l'évolution des signaux mesurés.

Nous constatons que le bruit sur le signal d'absorbance à 565 nm est bien plus important que le bruit sur le signal d'absorbance à 265 nm. Ceci est probablement dû aux interactions du rayonnement lumineux avec les limons et sables contenus dans les eaux usées.

Le bruit observé dans les longueurs d'onde du domaine visible affectera donc plus particulièrement tous les modèles utilisant ces longueurs d'ondes. Un filtrage des spectres d'absorbance sera vraisemblablement nécessaire afin de rendre les signaux de concentration exploitables.

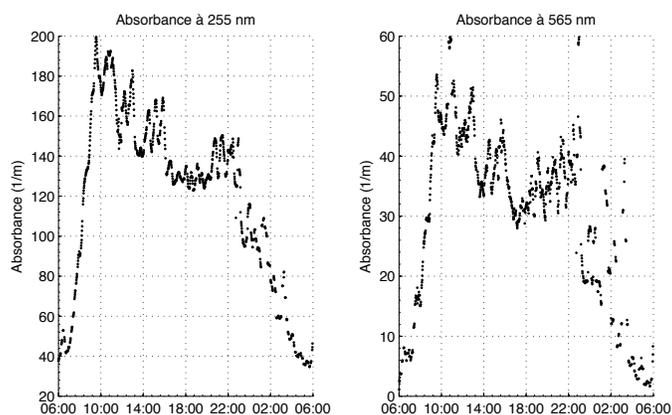


Fig. 4.39 – Absorbances à 255 et 565 nm mesurées pendant la journée du 18/04/2008

---

#### **4.3.5.1.b Exploitabilité des modèles de prédiction de la concentration en DCO**

##### **modèles fbiPLS**

Les signaux de concentration en DCO calculés avec les trois modèles fbiPLS et les modèles mwPLS pour la journée du 18/04/2008 sont présentés en annexe C. Les trois modèles fbiPLS donnent des valeurs négatives en période nocturne. Ils s'avèrent donc totalement inadaptés à la prédiction de concentration en DCO pour des spectres d'absorbance ne correspondant pas au domaine de validité de l'étalonnage, surtout pour les faibles valeurs d'absorbance. Notons que les valeurs négatives sont obtenues sur les plages où les spectres sont de classe 3. Cela confirme donc que les modèles ne doivent pas être utilisés sur cette classe de spectres. De plus, ils sont fortement influencés par le bruit affectant les mesures d'absorbance dans le domaine du visible.

Afin de valider l'un des trois modèles pour une utilisation in-situ, il conviendra donc au préalable de :

- compléter la base de données d'étalonnage avec des échantillons prélevés pendant la nuit ;
- et éventuellement effectuer un filtrage des spectres d'absorbance dans le visible.

Il pourra être également judicieux de recueillir des échantillons très chargés en DCO afin d'éviter toute dérive du modèle en borne haute du domaine de validité de l'étalonnage.

##### **modèles mwPLS**

Le signal de concentration calculé avec le modèle mwPLS1 est le moins bruité. Il utilise en effet les absorbances mesurées entre 262,5 et 360 nm. En période nocturne, les concentrations calculées avec les trois modèles sont positives. Les modèles mwPLS sont donc moins sensibles que les modèles fbiPLS au changement de classe de spectre.

##### **Comparaison de toutes les méthodes d'étalonnage (sauf calage global)**

Les concentrations en DCO calculées avec six méthodes d'étalonnage testées sont tracées en FIG. 4.42.

Les modèles OPPPLS et PLS local aboutissent également à des concentrations négatives en période nocturne et présentent donc les mêmes défauts que les modèles fbiPLS.

---

Le signal de concentration issu du calage local s::can est le moins bruité. Il s'agit vraisemblablement d'un modèle PLS à une composante établi à partir d'une plage spectrale dans l'UV. Pendant la phase d'étalonnage, ce modèle s'avérait moins performant en terme de qualité de prédiction. Il a cependant l'avantage d'être moins sensible à des conditions différentes de celles observées pendant l'étalonnage. Le calage effectué par s::can est réalisé à partir d'une base de données importante comprenant des échantillons dans toutes les gammes de concentrations rencontrées. Ceci explique donc qu'il soit moins sensible aux faibles valeurs d'absorbance rencontrées en période nocturne.

Le signal de concentration issu de la régression linéaire simple donne également des valeurs positives en période nocturne. Il présente l'avantage, par rapport au calage s::can d'être directement établi à partir des longueurs d'ondes les plus pertinentes sur le site d'étude.

#### **4.3.5.1.c Exploitabilité des modèles de prédiction de la concentration en MES**

##### **modèles fbiPLS**

Les signaux de concentration en MES calculés avec les trois modèles fbiPLS pour la journée du 18/04/2008 sont présentés en annexe C. Des différences locales affectent les signaux au niveau d'un pic de concentration. Le signal obtenu avec le modèle fbiPLS2 sont les plus plausibles au vu des résultats donnés par les autres modèles (cf. FIG. 4.41). Il présente en outre l'avantage d'être très peu bruité.

##### **Comparaison de toutes les méthodes d'étalonnage (sauf calage global)**

Les concentrations calculées avec les méthodes d'étalonnage testées sont tracées en FIG. 4.41.

Le modèle PLS local, la régression linéaire simple et le calage local s::can donnent des signaux de concentration très proches, aussi bien au niveau des valeurs de concentration prédites qu'en terme de bruit.

Le modèle OPPPLS aurait pu être rattaché à ce premier regroupement de modèles. Cependant, nous observons qu'il a vraisemblablement tendance à surestimer des pics de concentrations pour des spectres présentant des valeurs d'absorbance élevées.

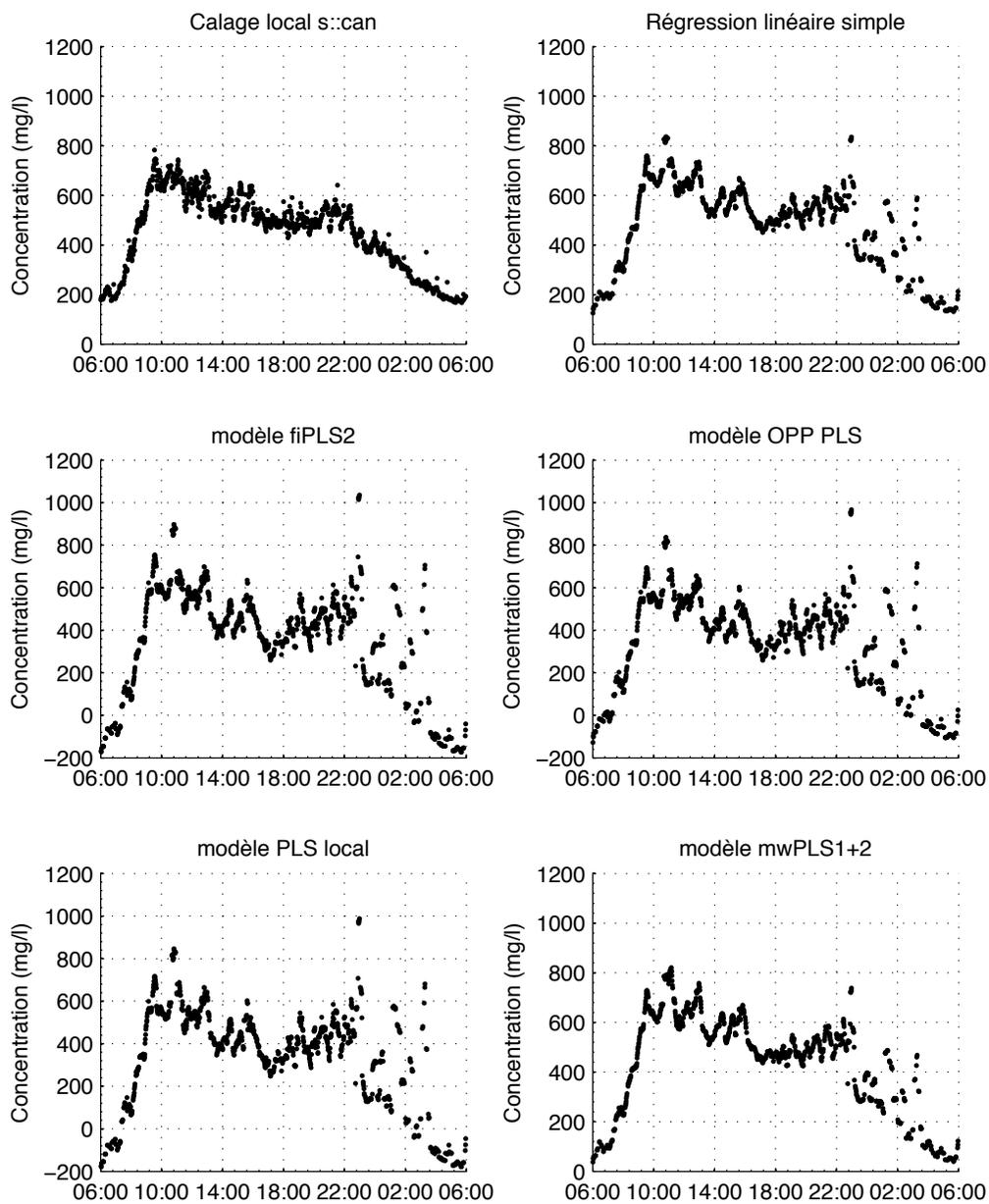


Fig. 4.40 – Concentrations en DCO obtenues avec six modèles différents pendant la journée du 18/04/2008

---

Les modèles mwPLS et fbiPLS2 intègrent le même nombre de composantes et sont basés sur des plages spectrales voisines. Les concentrations calculées sont donc similaires. Le signal de concentration obtenu est le moins bruité, et par conséquent le plus exploitable pour une utilisation in-situ, d'autant plus que les résultats d'étalonnage de ces deux modèles sont excellents. Le "lissage" du signal est lié à la nature même du modèle et aux valeurs des coefficients du vecteur de régression. En effet, les absorbances aux longueurs d'onde proches de 450 nm sont aussi bruitées que celle mesurée à 565 nm.

#### **4.3.5.2 Comparaison des résultats des modèles avec des mesures analytiques**

##### **4.3.5.2.a Comparaison avec les concentrations moyennes 24h d'autosurveillance du 16/04/2010**

Les concentrations moyennes 24h en DCO et MES calculées avec chacun des modèles ont été comparées avec les concentrations moyennes 24h d'autosurveillance sur la journée du 16/04/2010. Les résultats sont récapitulés dans les TAB. 4.17 et 4.18.

Les concentrations instantanées des modèles prédictifs de DCO fbiPLS, OPPPLS et PLS local ont été limitées à zéro afin de ne pas prendre en compte les valeurs négatives de concentrations calculées en période nocturne. Le modèle mwPLS<sub>1+2</sub> donne les meilleurs résultats avec un écart de 3 % par rapport à la valeur d'autosurveillance. Les écarts obtenus avec les modèles "calage local s::can" et RLS sont en retrait mais restent acceptables avec des écarts respectifs de 18 et 25 % par rapport à la valeur d'autosurveillance. Les trois autres modèles sous estiment la valeur de concentration en raison d'une mauvaise prédiction des faibles valeurs de concentration.

Concernant les MES, tous les modèles réalisent d'excellentes prédictions avec des écarts compris entre 0,5 et 13%. Les deux meilleurs modèles sont les modèles RLS, mwPLS et fbiPLS2.

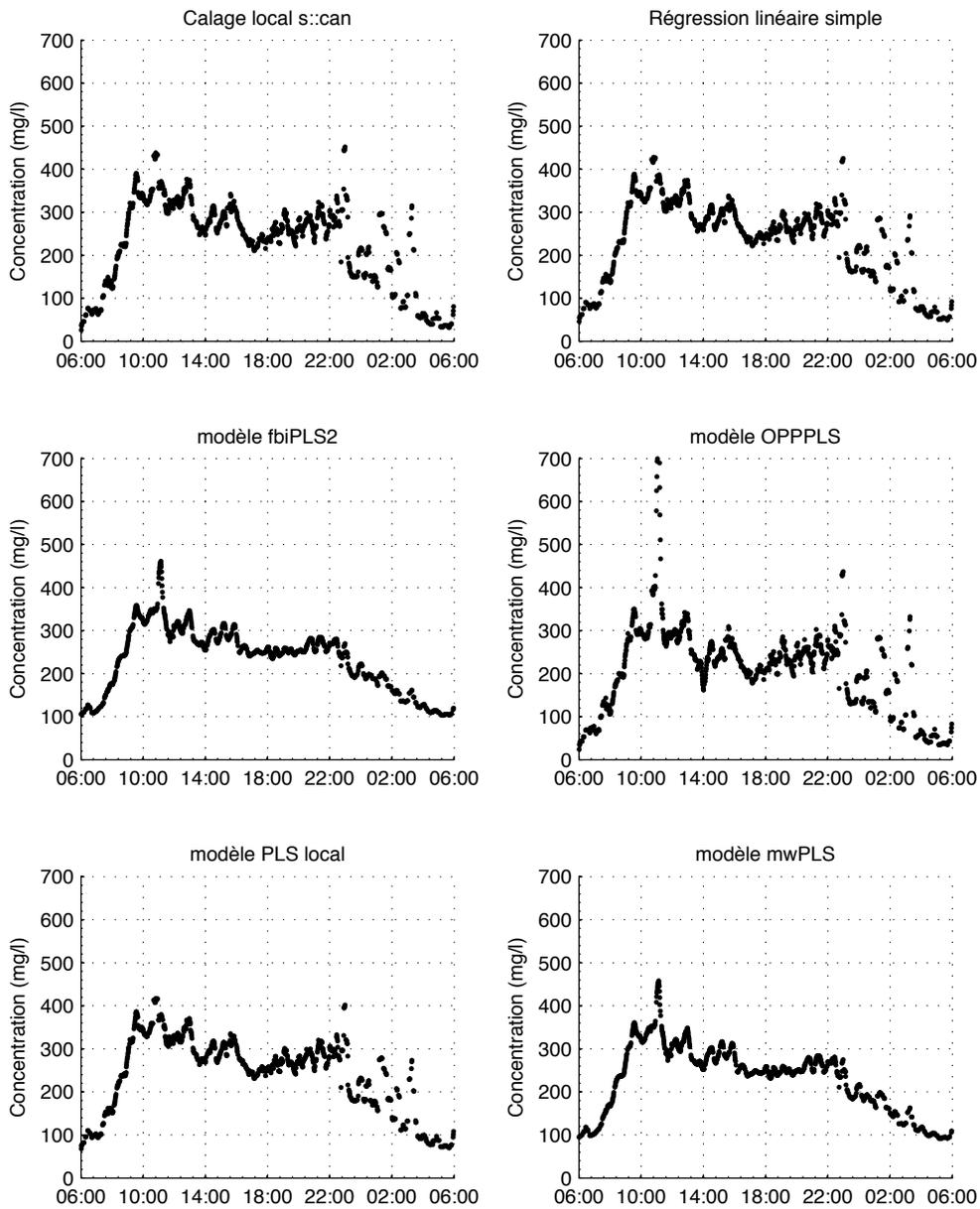


Fig. 4.41 – Concentrations en MES obtenues avec six modèles différents pendant la journée du 18/04/2008

<b>Autos.</b> (mg/l)	<b>Cal. local</b> (mg/l)	<b>mwPLS</b> (mg/l)	<b>fbiPLS</b> (mg/l)	<b>OPPPLS</b> (mg/l)	<b>PLS local</b> (mg/l)	<b>RLS</b> (mg/l)
342	402	(1) : 297 (2) : 423 (1+2) : 353	(1) : 218 (2) : 267 (3) : 293	273	245	427

Tab. 4.17 – Comparaison des concentrations moyennes 24h en DCO issues des modèles avec la concentration moyenne 24h issue de l'autosurveillance sur la journée du 16/04/2010

<b>Autos.</b> (mg/l)	<b>Cal. local</b> (mg/l)	<b>mwPLS</b> (mg/l)	<b>fbiPLS</b> (mg/l)	<b>OPPPLS</b> (mg/l)	<b>PLS local</b> (mg/l)	<b>RLS</b> (mg/l)
207	201	204	210*	180	217	208

(\*) : concentration calculée avec le modèle fbiPLS<sub>2</sub>. Les concentrations moyennes 24h des modèles fbipls<sub>1</sub> et 3 sont égales à 238 et 161 mg/l

Tab. 4.18 – Comparaison des concentrations moyennes 24h en MES des modèles avec la concentration moyenne 24h issue de l'autosurveillance sur la journée du 16/04/2010

#### 4.3.5.2.b Comparaison avec les concentrations mesurées pendant deux campagnes de prélèvements

Les concentrations des six modèles prédictifs de concentration en DCO ont été comparées à des concentrations mesurées pendant deux campagnes de prélèvement (16/04/2008 et 18/04/2008). Les résultats sont présentés en FIG. 4.42 et 4.43.

Sur la première campagne du 16/04/2008, les modèles calage local s::can et RLS sont visiblement moins performants, confirmant ainsi les résultats obtenus lors de l'étalonnage. Les quatre autres modèles donnent visuellement de meilleurs résultats. Les erreurs relatives moyennes ont été calculées pour chacun des modèles et sont présentées dans le TAB. 4.19. Les erreurs relatives les plus faibles sont obtenues avec le modèle mwPLS<sub>1</sub> suivi des modèles OPPPLS, PLS local et fbiPLS<sub>2</sub>.

<b>Cal. local</b> (Er. %)	<b>. mwPLS</b> (Er. %)	<b>fbiPLS</b> (Er. %)	<b>OPPPLS</b> (Er. %)	<b>PLS local</b> (Er. %)	<b>RLS</b> (Er. %)
58	(1) : 21 (2) : 61 (1+2) : 37	(1) : 30 (2) : 29 (3) : 31	26	27	58

Tab. 4.19 – Erreurs relatives moyennes sur la prédiction des concentrations en DCO pour la campagne de mesure du 16/04/2008

Les erreurs relatives sont élevées. Elles s'expliquent en partie par l'hétérogénéité de l'influent dans le canal de mesure. Les prélèvements ont été réalisés avec un seau et les mesures ont été horodatées avec l'horloge de la centrale d'acquisition s::can. Nous ne pouvons toutefois garantir l'exacte similarité des échantillons mesurés en laboratoire et de l'échantillon mesuré in-situ par le spectro::lyser.

Les résultats de la campagne du 18/04/2010 sont présentés en FIG. 4.43 et TAB. 4.20. Le modèle RLS donne les plus mauvaises prédictions avec une erreur relative moyenne de 22%. Les modèles mwpls1 et calage local s::can sont légèrement meilleurs avec des erreurs relatives respectives de 16 et 19 %. Les trois meilleurs modèles sont les modèles PLS local, OPPPLS et fbiPLS<sub>1</sub> avec des erreurs relatives moyennes inférieures à 11 % ce qui est excellent.

<b>Cal. local</b> (Er. %)	<b>mwPLS</b> (Er. %)	<b>fbiPLS</b> (Er. %)	<b>OPPPLS</b> (Er. %)	<b>PLS local</b> (Er. %)	<b>RLS</b> (Er. %)
16	(1) : 19 (2) : 25 (1+2) : 21	(1) : 11 (2) : 12 (3) : 11	10	9	22

Tab. 4.20 – Erreurs relatives moyennes sur la prédiction des concentrations en DCO pour la campagne de mesure du 18/04/2008

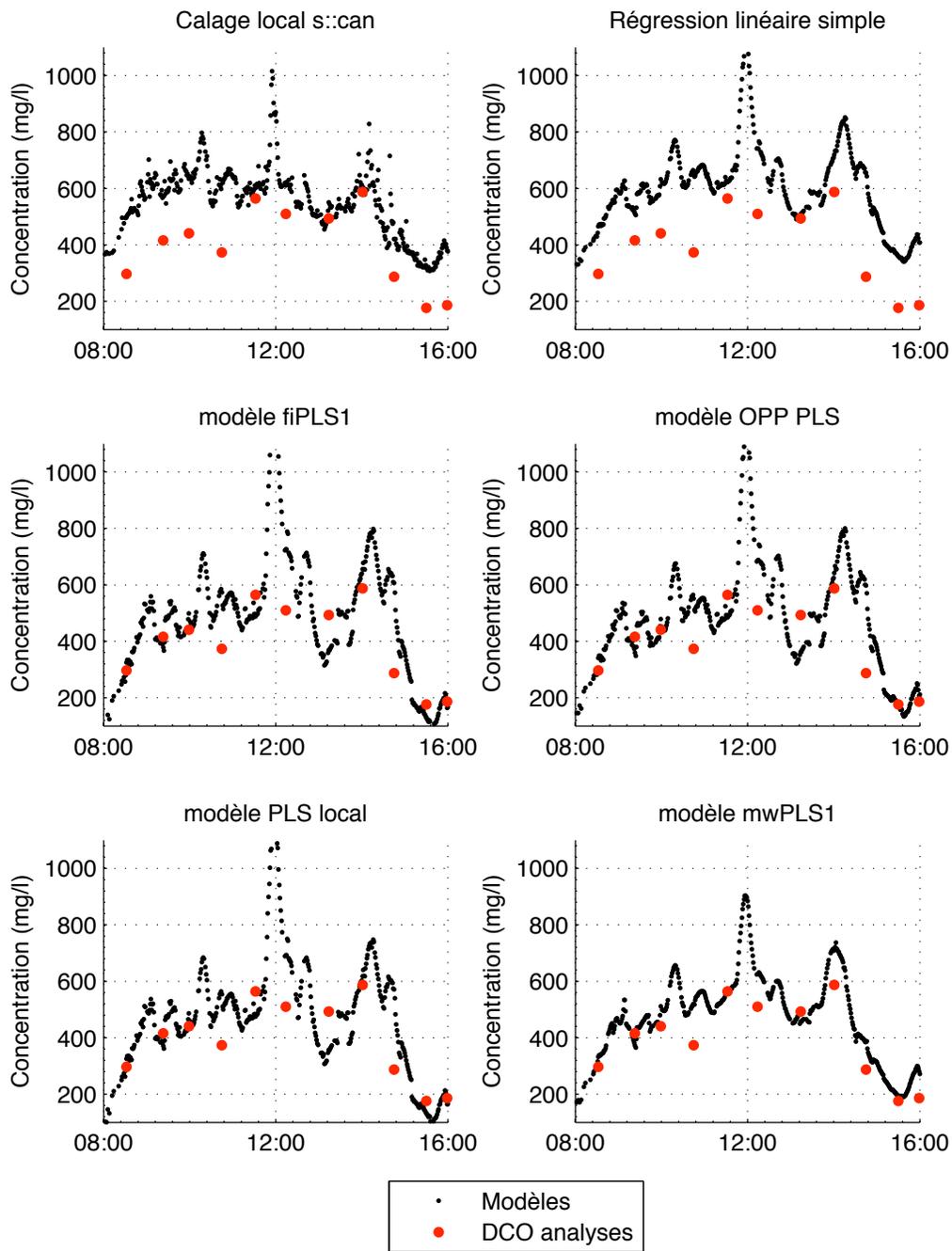


Fig. 4.42 – Concentrations en DCO obtenues avec six modèles différents pendant la journée du 16/04/2008

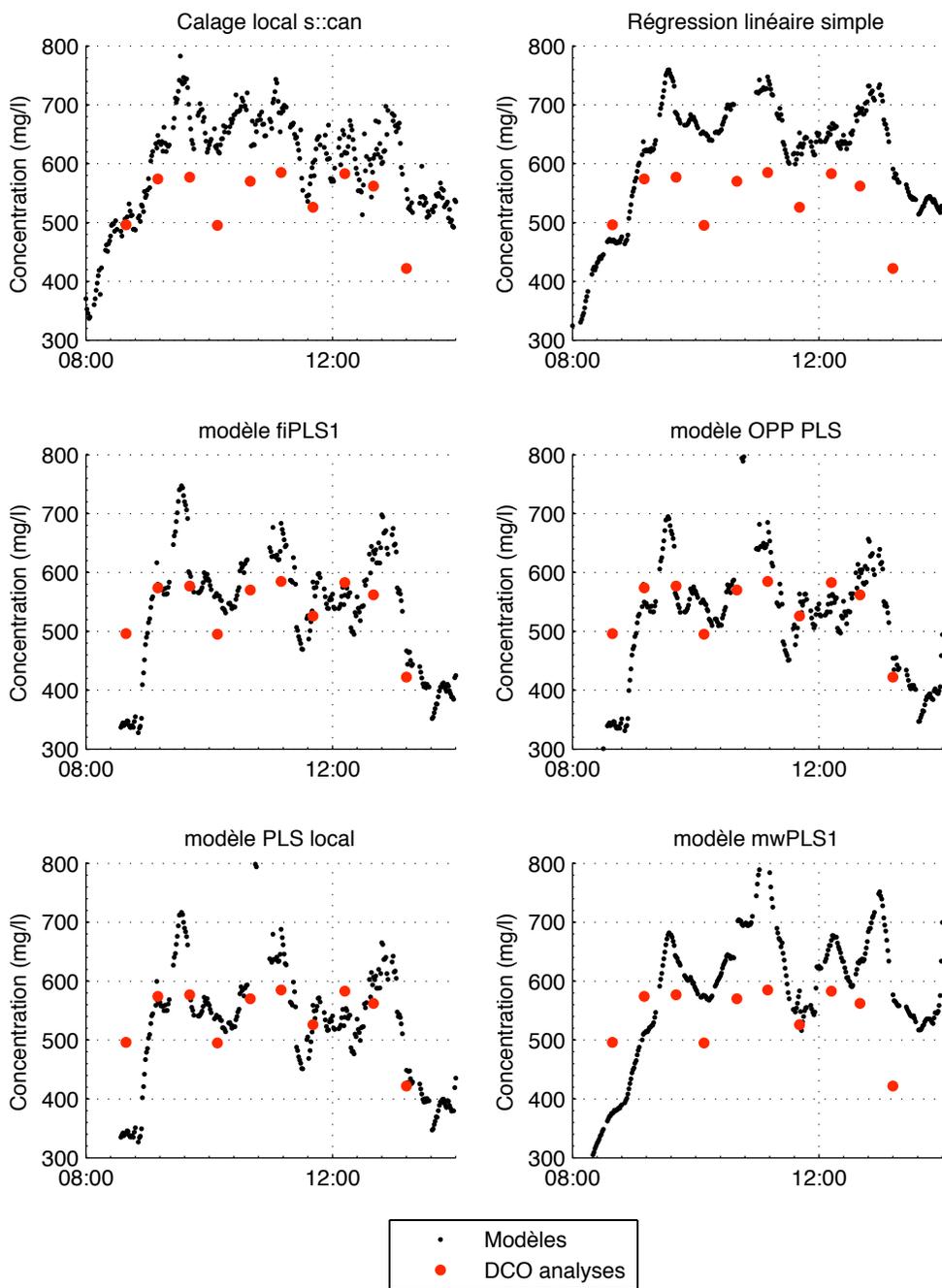


Fig. 4.43 – Concentrations en DCO obtenues avec six modèles différents pendant la journée du 18/04/2008

---

#### 4.3.5.3 Conclusion

Les prédictions de concentrations en DCO sur des mesures effectuées in-situ montrent la sensibilité de trois modèles à une utilisation en période nocturne. Les modèles fbiPLS, OPPPLS et PLS local conduisent en effet à des prédictions de concentrations négatives. Les trois autres modèles sont par conséquent davantage adaptés à l'évaluation des concentrations moyennes journalières par temps sec. Lorsque les modèles sont utilisés au sein de leur domaine de validité, les résultats de l'étalonnage sont confirmés avec d'excellentes prédictions de concentrations réalisées avec les modèles fbiPLS, OPPPLS, mwPLS et PLS local. Aucune différence de performance à l'avantage des modèles fbiPLS et OPPPLS n'a été constatée lors des deux campagnes de mesures par rapport au modèle PLS local. Par ailleurs, l'observation des signaux de concentrations sur une journée de temps-sec révèle très peu de différences entre les trois modèles. Enfin, le bruit des signaux de concentrations de tous les modèles est bien maîtrisé. Ils sont donc exploitables tels quels.

L'exploitation des mesures in-situ montre une très faible sensibilité des modèles prédictifs de concentration en MES à une utilisation en période nocturne. La mesure de contrôle d'autosurveillance aboutit à des écarts entre concentration mesurée et simulée très faibles pour l'ensemble des modèles. L'écart relatif le plus élevé est obtenu avec le modèle OPPPLS (13%). Le bruit des signaux de concentrations est particulièrement bien maîtrisé pour les modèles fbiPLS et mwPLS. Leur utilisation en temps sec est donc recommandée. Le bruit du signal de concentrations en MES du modèle OPPPLS est en revanche très présent. Il a vraisemblablement pour origine un nombre de composantes élevé (6). En effet, tous les autres modèles intègrent au maximum deux composantes et ont des signaux plus exploitables. La parcimonie des modèles fbiPLS et mwPLS est donc dans ce cas particulièrement appréciable.

### 4.4 Conclusion

Le travail effectué complète les travaux dédiés à la mesure en temps réel in-situ (et simultanée) de la DCO, des MES et par spectrophotométrie UV/Visible (Langergraber *et al.*, 2003; Rieger *et al.*, 2004, 2006; Torres et Bertrand-Krajewski, 2008; Maribas *et al.*, 2008).

Cette partie de thèse montre que les performances des modèles PLS utilisés en spectrophotométrie UV/Visible pour la prédiction des concentration en DCO et MES peuvent être

---

améliorées par une sélection pertinente de longueurs d'ondes. Lors de l'étalonnage et de la validation croisée des modèles, les meilleures prédictions de concentrations en DCO ont été obtenues avec les méthodes fbiPLS et OPPLS. Pour les MES, les meilleurs résultats sont obtenus avec les méthodes fbiPLS, OPPPLS et mwPLS.

Par ailleurs, les deux méthodes proposées (fbiPLS et mwPLS) conduisent à l'identification de modèles plus parcimonieux, ce qui constitue un sérieux avantage pour l'exploitabilité des signaux de concentrations mesurés en continu. Ceci est particulièrement observable sur les prédictions de concentrations en MES où les modèles fbiPLS et mwPLS intègrent deux composantes alors que le modèle OPPPLS contient trois dimensions supplémentaires.

Le second intérêt du travail réalisé réside dans la définition des conditions d'utilisation des modèles prédictifs. Nous avons tout d'abord restreint la validité des prédictions aux bornes minimales et maximales de concentrations de la collection d'étalonnage. Toutefois ceci n'est pas suffisant. Nous avons donc critiqué la représentativité des données de concentrations en exploitant les données d'autosurveillance. Enfin, nous avons proposé une méthode qualitative visant à comparer les formes des spectres mesurés in-situ par rapport aux formes de spectres de la collection d'étalonnage.

Nous avons ainsi montré que notre collection d'étalonnage n'est pas consistante et n'intègre pas l'ensemble de la variabilité horaire et mensuelle des eaux usées. De plus, la classification qualitative des spectres d'absorbance permet d'identifier des spectres atypiques, sous représentés dans la collection d'étalonnage, et pour lesquels les prédictions de concentrations doivent être considérées avec précaution. Ceci est particulièrement valable pour les prédictions de DCO qui sont établies à partir de mesures d'absorbance dans l'UV et le visible. Cette méthode permet en outre d'établir un lien entre classe de spectres et conditions environnementales. Ainsi, les modèles prédictifs ne peuvent être utilisés en période nocturne et par temps de pluie sans compléter au préalable la collection d'étalonnage avec des échantillons appartenant à ces classes.

La variabilité de la matrice des eaux usées étant très importante, il semble difficile d'établir un modèle unique valable pour toutes les conditions. La meilleure stratégie consisterait par conséquent en l'identification d'un modèle PLS pour chaque classe de spectres d'absorbance ou en l'étalonnage d'un modèle unique correspondant à des conditions d'utilisation restreintes.

Un étalonnage aussi fin des spectrophotomètres in-situ n'est pas envisageable pour un gestionnaire de système d'assainissement. La démarche de calage local proposée par s::can est

---

certes moins précise que toutes les méthodes d'étalonnage testées, mais a l'avantage d'être la plus simple et la plus rapide à mettre en oeuvre. Toutefois, n'ayant pas accès aux données d'étalonnage, il n'est pas possible de circonscrire son domaine d'utilisation. Ceci rend donc improbable l'utilisation du calage local s::can lorsque la variation de la matrice des eaux usées est importante (réseau unitaire) et lorsque l'objectif est la prédiction des concentrations instantanées.

## **Troisième partie**

# **Applicabilité de la spectrophotométrie UV/Visible pour la détection de pollution bactériologique à court terme**

# Introduction

Cette partie s'insère dans le cadre du projet GIRAC (Gestion Intégrée des Rejets d'Assainissement Côtiers). Le but du projet est d'identifier les flux de pollution bactériologique provenant des principales sources (rivières, rejets d'assainissement) et de les injecter dans un modèle hydrodynamique qui simulera la dispersion de la pollution au sein du milieu marin (cf. FIG. 4.44).

La présente étude se concentre uniquement sur l'évaluation en temps réel des pics de pollution bactériologique à l'exutoire de la rivière Stang Alar située sur la commune de Brest. Après une revue bibliographique s'intéressant aux différentes techniques de mesure et de modélisation de la pollution d'origine fécale, nous évaluerons la possibilité de détecter des pics de pollution bactériologique à l'aide de mesures spectrophotométriques.

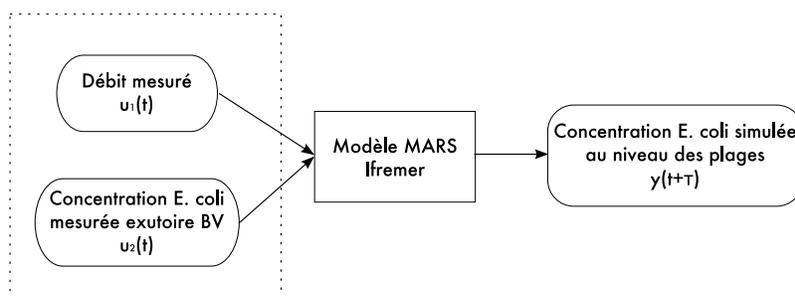


Fig. 4.44 – Méthodologie de prévision à court terme de la concentration en E. coli dans les zones de baignade (les zones en pointillés correspondent aux axes de travail à développer durant la thèse).

## Chapitre 5

# Étude bibliographique

Ce chapitre a pour vocation, grâce à une étude bibliographique, de nous faire acquérir des connaissances sur :

1. la notion d'indicateur de contamination fécale (ICF) et sa pertinence vis à vis des risques sanitaires ;
2. les techniques de dénombrement de germes ICF ;
3. les autres voies de recherche permettant d'identifier un risque sanitaire pour les baigneurs ;
4. la dynamique des concentrations en germes ICF au niveau des sites de baignade ;
5. les possibilités de modéliser en continu les teneurs en germes ICF au niveau des sites de baignade ;

Cette étude permettra d'identifier si des travaux de recherche ont déjà porté sur l'évaluation de la pollution bactériologique par spectrophotométrie UV/Visible. Nous pourrons ainsi évaluer la crédibilité de cette piste de recherche, notamment au regard des facteurs environnementaux influençant les teneurs en germes ICF au niveau des sites de baignade.

---

## 5.1 Les indicateurs de contamination fécale

### 5.1.1 Notion d'indicateur de contamination fécale

La détection de tous les pathogènes dans l'eau est impossible. En effet, il existe une diversité telle de bactéries, virus, protozoaires qu'il n'est pas envisageable de tous les quantifier. Par ailleurs, la faible abondance de chaque pathogène nécessiterait de concentrer les échantillons. Enfin, pour nombre de micro-organismes, il n'existe pas de méthode normalisée et rapide de quantification.

Toutes ces raisons plaident pour l'évaluation de la qualité microbiologique des eaux selon le concept de *germe indicateur de contamination fécale* (ICF). Un germe ICF peut être défini de la manière suivante (Bitton, 2005; Garcia-Armisen, 2006) :

- Un indicateur n'est pas nécessairement pathogène.<sup>1</sup>
- Sa présence indique l'existence d'une contamination par des matières fécales.
- Son absence indique l'inexistence d'une contamination d'origine fécale.
- Son taux de survie est similaire à celui des pathogènes.
- Il ne se reproduit pas dans les eaux naturelles.
- Son unique origine doit être la contamination fécale.
- Enfin, les méthodes analytiques mises en oeuvre pour sa détermination doivent être peu onéreuses.

### 5.1.2 Groupes de bactéries utilisés comme indicateur de contamination fécale

Historiquement, les coliformes totaux et les coliformes fécaux sont les principaux indicateurs de contamination fécale utilisés (USEPA, 2003). Toutefois, ces indicateurs sont de plus en plus remplacés par *Escherichia coli* (EC) et les Entérocoques intestinaux (EI), qui semblent plus adaptés pour évaluer le risque sanitaire associé aux différentes utilisations de l'eau (Cabelli, 1983; Dufour, 1984; EPA, 1986; Edberg *et al.*, 2000; Fewtrell et Bartram, 2001).

---

<sup>1</sup>Un bon indicateur de contamination fécale n'est d'ailleurs pas pathogène.

---

### 5.1.2.1 Coliformes totaux

La pertinence de ce groupe comme indicateur de contamination fécale est de plus en plus remise en cause. En effet, toutes les espèces de ce groupe ne sont pas spécifiques de la flore intestinale des animaux à sang chaud. D'autre part, certaines espèces d'origine tellurique ou aquatique peuvent croître dans un environnement aquatique (Tallon *et al.*, 2005).

### 5.1.2.2 Coliformes fécaux

Les coliformes fécaux sont un sous-groupe des coliformes totaux capables de se développer à 44°C. Ce groupe semble a priori plus approprié que les coliformes totaux en tant qu'indicateur de contamination fécale. Il est principalement constitué de *Escherichia coli*, *Klebsiella*, *Enterobacter* et *Citrobacter*.

*Escherichia coli* constitue la majeure partie des coliformes fécaux :

- 77% des coliformes fécaux selon une première étude de Garcia-Armisen (2006) (cf. FIG. 5.1) ;
- 63% selon une autre étude de l'USGS (Rasmussen et Ziegler, 2003).

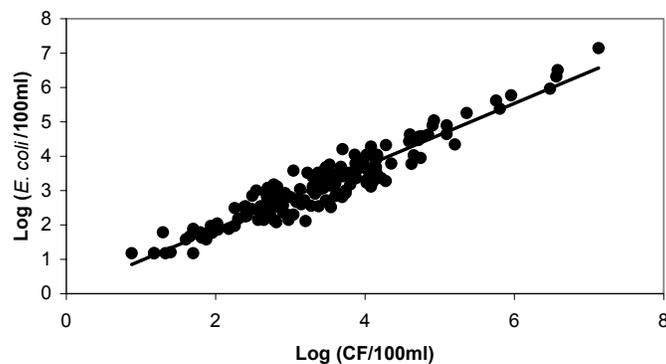


Fig. 5.1 – Régression linéaire en coordonnées Log-Log entre le nombre de coliformes fécaux et le nombre de *E. coli* dans des rivières du bassin de la Seine (Garcia-Armisen, 2006)

Toutefois, certains auteurs rapportent la présence de certaines espèces du groupe sans contamination fécale (McLellan *et al.*, 2001; Gauthier et Archibald, 2001).

---

### 5.1.2.3 *Escherichia coli*

*Escherichia coli* est aujourd'hui considéré comme le meilleur indicateur de contamination fécale (Baudiová, 1997; Edberg *et al.*, 2000). Il semble par ailleurs mieux corrélé au risque sanitaire. Cependant, de nombreuses études pointent la non spécificité de ce germe ICF (Solo-Gabriele *et al.*, 2000; Hardina et Fujioka, 2006; Ishii et Sadowsky, 2008).

Une description très détaillée des différentes sources de *E. coli* dans l'environnement est donnée dans la revue bibliographique effectuée par Ishii et Sadowsky (2008). Le premier habitat de *E. coli* est le bas intestin des animaux à sang chaud, les humains y compris. Une quantité importante de ces bactéries est rejetée dans l'environnement (qui constitue leur second habitat) via les matières fécales. Jusqu'à très récemment, il était admis que *E. coli* était incapable de survivre dans l'environnement, et n'était pas capable de croître dans des habitats secondaires comme, l'eau, les sédiments ou le sol. Les bactéries doivent en effet faire face à de nombreux stress tels que la température (faible ou élevée), une faible humidité, des variations de textures de sol, une faible abondance de matières organiques, le rayonnement solaire et la prédation. Des études récentes ont au contraire montré que *E. coli* peut survivre pendant de longues périodes dans l'environnement, et peut potentiellement se multiplier dans l'eau, sur les algues et dans les sols de régions subtropicales et tempérées (voir Ishii et Sadowsky 2008, p. 2). Les facteurs expliquant les concentrations importantes de *E. coli* trouvées dans l'environnement sont la présence de nutriments et une température élevée. Un diagramme simplifié du cycle de vie de *E. coli* dans l'environnement est représenté en FIG. 5.2.

### 5.1.2.4 Entérocoques intestinaux

Les Entérocoques intestinaux sont un bon indicateur spécifique de la contamination fécale. En effet, il existe une bonne corrélation entre abondance d'entérocoques intestinaux et apparitions de maladies gastro-intestinales sur les baigneurs (Cabelli *et al.*, 1982; Cabelli, 1989). Par ailleurs, les Entérocoques intestinaux survivent plus longtemps dans le milieu naturel que les *Escherichia coli*. Leur utilisation comme indicateur permet donc de détecter des contaminations plus anciennes (Edberg *et al.*, 1997). Toutefois, comme pour *E. coli*, des réserves peuvent être formulées quant à la spécificité des Entérocoques intestinaux comme source de pollution fécale humaine (Ferguson *et al.*, 2005; Anderson *et al.*, 1997; Solo-Gabriele *et al.*, 2000).

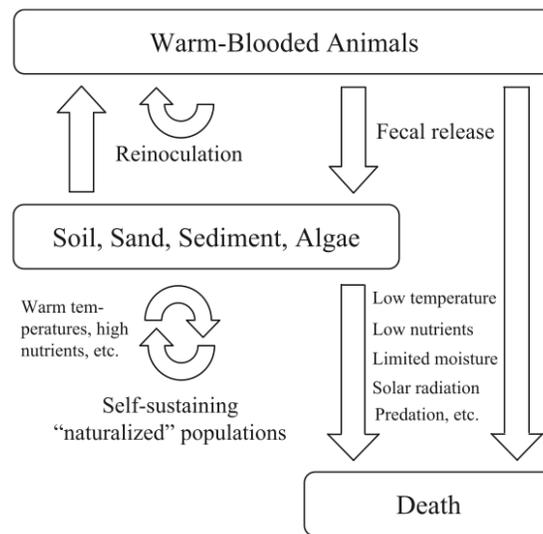


Fig. 5.2 – Diagramme simplifié du cycle de vie de *E. coli* dans l’environnement (Source : (Ishii et Sadowsky, 2008))

## 5.2 Dénombrement des germes indicateurs de contamination fécale

Il existe trois grands types de méthodes pour le dénombrement de germes indicateurs de contamination fécale :

- les méthodes d’énumération basées sur la mise en culture ;
- les méthodes enzymatiques d’évaluation rapide ;
- les méthodes moléculaires.

### 5.2.1 Méthodes d’énumération basées sur la mise en culture

#### 5.2.1.1 Principe des méthodes

Les méthodes d’énumération basées sur la mise en culture sont utilisées en routine pour l’évaluation de la qualité microbiologique des eaux de surface.

---

Une première famille de méthode est basée sur la détermination du nombre le plus probable (NPP). Des dilutions décimales de l'échantillon sont inoculées dans une série de tubes contenant un milieu de culture spécifique. Une loi statistique permet de calculer le NPP sur la base de tubes positifs dans chaque dilution.

La seconde famille de méthode est la filtration sur membrane. Un volume défini de l'échantillon est filtré et la membrane est incubée sur un milieu gélosé spécifique. Différentes conditions d'incubation peuvent être utilisées selon les bactéries recherchées. Les résultats sont exprimés en Unité Formant Colonie (UFC/ 100 ml).

Des améliorations aux méthodes basées sur la mise en culture ont été apportées ces dernières années. On utilise les propriétés enzymatiques des coliformes et de *E. coli* pour augmenter la spécificité et diminuer le temps de réponse. Par exemple, pour mettre en évidence la présence de coliformes fécaux ou *E. coli*, on cherche à détecter l'activité de la  $\beta$ -D-galactosidase (enzyme spécifique des coliformes fécaux) et de la  $\beta$ -D-glucuronidase (enzyme spécifique des *E. coli*). Dans ce but, on introduit des substrats chromogéniques et fluorogéniques dans les milieux de culture. L'hydrolyse de ces substrats par les enzymes donnent des produits colorés ou fluorescents qui permettent une détection de l'activité enzymatique ciblée.

Ainsi, nombre d'analyses sont basées sur l'approche *defined substrate technique* (Edberg *et al.*, 2000). Dans cette approche, le seul substrat disponible est un composé nécessitant l'activité enzymatique ciblée pour être utilisé : seul le groupe de bactéries ciblé peut donc se développer. La méthode normalisée ISO 9308-3 permettant le dénombrement des *E. coli* dans les eaux de surface est basée sur ce principe.

Les techniques de mise en culture présentent quelques inconvénients. Tout d'abord, les analyses sont longues (de 18 à 48h). De plus, elles ne permettent pas de détecter les bactéries "viables mais non cultivables" (VBNC) ce qui les rend certainement inadaptées au dénombrement des coliformes dans les eaux naturelles.

### **5.2.1.2 Problématique des bactéries viables mais non cultivables**

L'existence de bactéries viables mais non cultivables est connue depuis les années 1980 (Colwell *et al.*, 1985; Grimes et Colwell, 1986). Différentes méthodes ont été utilisées pour mesurer l'activité ou la viabilité des bactéries :

- mesure de l'intégrité cellulaire,

- 
- mesure d'activités métaboliques (utilisation de substrat, synthèse de protéines, activité respiratoire, activité enzymatique, synthèse d'ARN,...).

L'état VBNC a été mis en évidence aussi bien pour des bactéries autochtones du milieu aquatique que pour des bactéries entériques. Les conditions favorables au développement des bactéries entériques sont celles de leur habitat naturel : température de 37°C, présence de beaucoup de sucres et acides aminés (Savageau, 1983). Dans le milieu naturel, les conditions sont défavorables à la survie de ces bactéries : température plus faible, variation de pH, prédation et limitation en nutriments. Une partie de ces bactéries peut survivre en entrant dans l'état VBNC. De nombreuses publications ont montré que les bactéries perdent beaucoup plus rapidement leur capacité de croître sur des milieux de cultures spécifiques que leur activité métabolique (Colwell *et al.*, 1985; Oliver, 2005).

### **5.2.2 Méthode enzymatique d'évaluation rapide de la teneur en E. coli**

Des méthodes enzymatiques sans mise en culture ont été développées ces dernières années. Elles permettent le dénombrement de E. coli dans les eaux côtières, les eaux de surface et les eaux usées (Fiksdal *et al.*, 1994; George *et al.*, 2000, 2001; Lebaron *et al.*, 2005).

Le principe de la mesure est le suivant :

1. Filtration de 100 ml d'échantillon.
2. Incubation du filtre à 44°C en présence de substrat fluorogène (MUGlu) en concentration saturante.
3. Mesure de l'apparition de fluorescence au cours du temps.

Cette technique permet de mesurer l'activité  $\beta$ -D-glucuronidase (GLU) des E. coli en trente minutes. Elle a également l'avantage de détecter l'activité GLU des E. coli VBNC.

### **5.2.3 Méthodes moléculaires**

Cette famille de méthodes englobe les méthodes immunologiques, les méthodes Polymérase Chain Reaction (PCR), les méthodes Fluorescent In Situ Hybridization (FISH),... Toutes ces méthodes sont rapides et spécifiques mais ne font pas toujours la distinction entre bactéries

---

vivantes et mortes. Par ailleurs, elles sont coûteuses et ne sont pas applicables pour une utilisation in-situ en continu. Elles ne seront pas détaillées dans ce mémoire.

### **5.3 Indicateurs chimiques de pollution fécale**

Comme nous l'avons vu précédemment, l'énumération de EC et EI par les méthodes normalisées basées sur la mise en culture prennent un temps considérable pour l'obtention d'un résultat (18 à 48h). Pendant ce temps, des individus peuvent consommer ou être simplement en contact avec de l'eau contaminée.

Les méthodes PCR et les méthodes basées sur la mesure de l'activité enzymatique par fluorescence permettent de diminuer le temps d'analyse. Toutefois des problèmes de robustesse ont été constatés dans les deux types de méthodes (Glassmeyer *et al.*, 2005).

Par ailleurs, les indicateurs microbiologiques de contamination fécale (EC, EI ainsi que les coliformes) manquent de spécificité : ces bactéries ne proviennent pas exclusivement de pollution fécale humaine. On les retrouve également dans les fèces d'animaux et elles peuvent être également autochtones de milieux aquatiques. Ceci est particulièrement gênant, car il a été montré que les principaux risques de maladies gastro-entériques sont liés à des germes pathogènes d'origine anthropique (Cabelli, 1983; Dufour, 1984; EPA, 1986).

De très nombreuses études se sont donc intéressées à l'applicabilité de composés chimiques comme indicateurs de pollution fécale d'origine humaine. Les indicateurs chimiques peuvent être scindés en plusieurs classes :

- ceux qui sont produits par notre métabolisme ;
- ceux que nous consommons et qui traversent le corps ;
- ceux qui sont associés aux eaux grises.

Un grand nombre de traceurs ont été suggérés. Ceux qui reviennent le plus souvent dans la littérature sont le coprostanol (famille des stérols), la caféine, les produits pharmaceutiques, les azurants optiques, les surfactants et les fragrances. Les prochains paragraphes reprennent quelques éléments abordés dans la synthèse bibliographique de Hagedorn et Weisberg (2009) et dans l'étude de Glassmeyer *et al.* (2005).

---

### 5.3.1 Stérols/Stanols

Les stanols fécaux sont produits dans les intestins des animaux par biodégradation du cholestérol. La concentration des différents produits métaboliques issus de la dégradation du cholestérol varient selon le régime alimentaire et la flore intestinale.

Chez les humains, le cholestérol est principalement métabolisé en coprostanol ( $5\beta$ -stanol) et représente environ 60% des stanols. La proportion de coprostanol chez les animaux est moins importante ce qui permet l'identification de la source de pollution fécale.

Le ratio coprostanol/(coprostanol +  $5\beta$ stigmastanol) est souvent utilisé pour identifier l'origine de la pollution fécale. Si le ratio est inférieur à 0,23, l'origine de la pollution est 100 % animale (herbivores). En revanche, si le ratio est supérieur à 0,73, on peut garantir que la pollution fécale est 100 % d'origine humaine.

Des corrélations importantes ont été trouvées entre le coprostanol et les différents indicateurs de contamination fécale. Par contre, le lien entre coprostanol et germes pathogènes reste toujours à établir.

Le coprostanol est dans la plupart des études dosé par chromatographie en phase gazeuse couplée à la spectrométrie de masse (GC-MS). Cette technique de pointe est principalement utilisée dans le domaine pharmaceutique. Elle permet la détection de toutes sortes de substance avec des seuils de détection très faibles (substances à l'état de traces).

### 5.3.2 Caféine

De très nombreuses études ont montré la possibilité d'utiliser la caféine comme traceur de pollution fécale humaine. La caféine est présente dans nombre de boissons que nous buvons quotidiennement : café, thé, boisson énergisante, soda, chocolat. On estime qu'une personne consomme de 70 à 440 mg/j de caféine (Buerge *et al.*, 2003). Environ 3% de la caféine que nous ingérons n'est pas métabolisée est rejetée dans l'urine (Siegener et Chen, 2002). Cette quantité s'ajoute aux rejets directs dans les égouts des boissons précédemment citées.

La caféine est principalement métabolisée en paraxanthine, et ce composé n'est retrouvé que chez les humains. La recherche de ce métabolite comme indicateur de pollution fécale humaine a été très peu étudiée.

---

La caféine est bien biodégradée dans les stations de traitement d'eaux usées. Lors d'une étude menée au niveau de la station de traitement des eaux usées de Goteborg (Suède), des concentrations en caféine de 37  $\mu\text{g/l}$  et 4  $\mu\text{g/l}$  ont été respectivement mesurées dans les eaux usées brutes et les eaux traitées (Paxeus et Schroder, 1996).

La caféine est dosée avec des techniques coûteuses comme la chromatographie en phase gazeuse ou liquide couplée à la spectrométrie de masse (GC-MS ou HPLC). Quelques études rapportent la difficulté d'obtenir de bonnes corrélations avec les indicateurs de contamination fécale. Enfin, il est également montré que la caféine peut être détectée dans des milieux vraisemblablement non contaminés par l'homme.

### 5.3.3 Azurants optiques

Les azurants optiques sont des composés organiques présents dans les détergents ménagers pour "blanchir" les vêtements. Ces composés ont la propriété d'absorber la lumière ultraviolet à 365 nm et de réémettre de la lumière visible dans la portion bleue du spectre (411-435 nm). La réémission de couleur bleue contrebalance la couleur naturelle jaune du coton, et le fait apparaître d'un blanc éclatant.

5 à 75% de la quantité d'azurants optiques utilisés lors du lavage de nos vêtements est rejetée dans les égouts. Les azurants optiques sont également utilisés dans le papier toilette.

Les azurants optiques peuvent être quantifiés par plusieurs méthodes analytiques. La méthode la plus précise est la chromatographie en phase liquide couplée à la spectrométrie de masse. Cette méthode est coûteuse et requiert un personnel très qualifié. Les azurants peuvent également être quantifiés par fluorométrie. Cette méthode s'avère suffisamment précise, robuste, et est également beaucoup moins coûteuse. Cependant, la quantification par fluorométrie n'est pas spécifique. D'autres composés contenus dans la matière organique ou certains composés aromatiques absorbent et réémettent de la lumière aux mêmes longueurs d'ondes que les azurants. Pour éliminer ces interférences, les dernières études ont utilisé une propriété intéressante : les azurants se dégradent bien plus vite que les autres composants lorsqu'ils sont exposés à un rayonnement ultraviolet (Hartel *et al.*, 2007; Cao *et al.*, 2009). En exposant un échantillon sous une lumière UV et en effectuant des mesures de fluorescence à différents instants d'exposition (de  $t=0$  à  $t=30$  min), il est ainsi possible de déterminer la part de fluorescence attribuable aux azurants.

---

## 5.4 Variabilité spatiale des germes ICF au niveau des zones de baignade

La variabilité spatiale des germes ICF a été étudiée lors d'une campagne de mesure réalisée par l'USEPA dans le cadre du projet EMPACT (Environmental Monitoring for Public Access and Community Tracking) (Wymer *et al.*, 2005).

Cette campagne de mesure a porté sur cinq sites de baignade aux Etats-Unis dont :

- un situé au bord d'un lac (West Beach, Indiana) ;
- un au bord d'une rivière (Belle Isle Park, Detroit, Michigan) ;
- deux au bord de la mer (Wollaston Beach, Massachusetts et Imperial Beach, California) ;
- et un au niveau d'un estuaire (Miami Beach Park, Maryland).

Les EC ont été dénombrés sur les sites d'eau douce (rivière et lac) et les EI en eau de mer et estuarienne. Près de 2 500 prélèvements ont été effectués pour chaque site durant les mois de juillet et août. Sur chaque site, des échantillons ont été prélevés à trois distances respectives de la plage (cf. FIG. 5.3) de manière à caractériser la variabilité de la teneur en ICF perpendiculairement à la rive :

- au niveau de la zone Ankle (Cheville) caractérisée par une hauteur d'eau de 15 cm ;
- au niveau de la zone Knee (Genou) caractérisée par une hauteur d'eau de 50 cm ;
- au niveau de la zone Chest (Poitrine), caractérisée par une hauteur d'eau de 1,3 m.

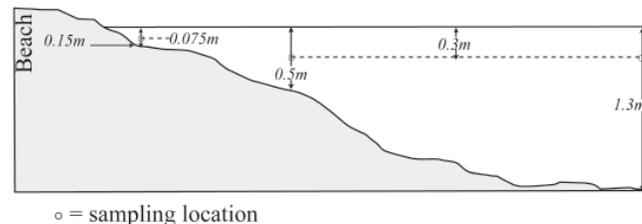


Fig. 5.3 – Profil en travers type montrant les différents endroits d'échantillonnage étudiés dans l'étude USEPA. Source : (Wymer *et al.*, 2005)

---

Par ailleurs, des prélèvements ont également été réalisés à 20 m d'intervalle sur chacune des zones de manière à étudier la variabilité spatiale de la teneur en *E. coli* parallèlement à la rive (cf. FIG. 5.4).

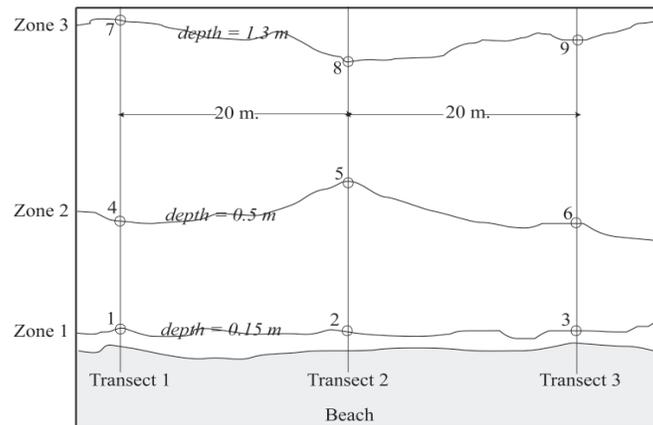


Fig. 5.4 – Grille d'échantillonnage type utilisée dans l'étude USEPA. Source : (Wymer *et al.*, 2005)

#### 5.4.1 Variabilité spatiale perpendiculairement à la plage

Sur tous les sites où des prélèvements ont été réalisés avec le protocole décrit au paragraphe précédent, on constate que les teneurs en ICF diminuent en s'éloignant de la plage (cf. FIG 5.5). Au niveau de la zone *cheville*, 44 % des échantillons dépassent le seuil de qualité. Au niveau des zones genou et poitrine, le pourcentage de dépassement de seuil de qualité n'est plus que de 34 et 6 %.

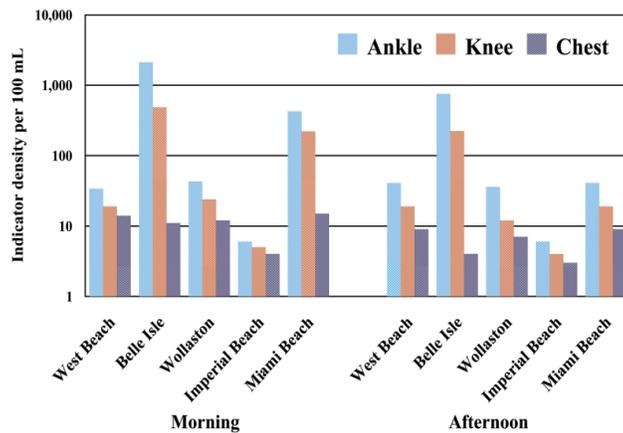


Fig. 5.5 – Influence de la distance à la plage sur la teneur en ICF. Source : (Wymer *et al.*, 2005)

#### 5.4.2 Variabilité spatiale parallèlement à la plage

Il n'a été noté aucune différence statistique significative entre les teneurs en ICF obtenues au niveau des différents de points de prélèvements au sein d'un même transect. Toutefois, l'étude fait remarquer, qu'en faisant l'hypothèse d'une distribution log-normale, il y a une chance sur deux pour qu'un échantillon prélevé à un instant  $t$  le long d'un transect ait une concentration qui diffère de plus de 0,28 log de la concentration d'un autre échantillon prélevé au même instant le long du même transect. Autrement dit, si on se réfère à un unique échantillon, il y a une chance sur deux pour qu'un second échantillon ait une concentration supérieure à deux fois la valeur du premier échantillon ou inférieure à la moitié de la valeur du premier.

#### 5.4.3 Variation de la teneur en ICF suivant la verticale

L'étude de l'USEPA montre que la teneur en ICF ne varie guère selon la profondeur de prélèvement comme le montre la FIG. 5.6.

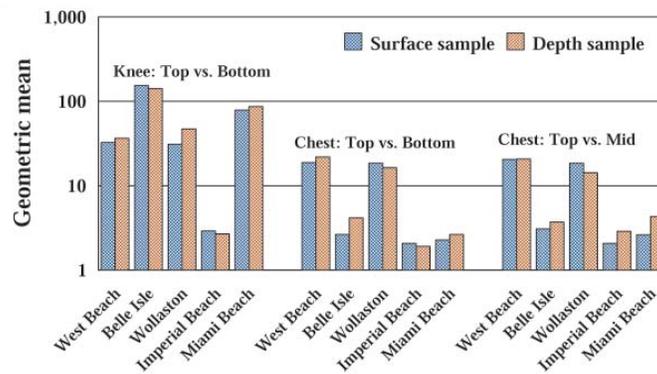


Fig. 5.6 – Influence de la profondeur de prélèvement sur la teneur en ICF. Source : (Wymer *et al.*, 2005)

## 5.5 Liens entre germes ICF et certains facteurs environnementaux

### 5.5.1 Influence des marées sur la teneur en ICF

L'influence de la marée sur la teneur en ICF a été analysée par Boehm et Weisberg (2005). Les données de leur étude sont issues de mesures effectuées sur soixante plages en Californie (USA), sur plusieurs années (au minimum trois ans), et avec un échantillonnage réalisé plusieurs fois par semaine. L'indicateur de contamination fécale utilisé dans le cadre de cette analyse est EI.

Il ressort que les concentrations en EI pendant les périodes de hautes eaux sont significativement plus grandes que les concentrations mesurées en basses eaux (cf. FIG. 5.7). L'autre facteur explicatif des teneurs en germes ICF est le reflux de la mer dans les terres ou le déversement des rivières dans la mer. Les plus fortes teneurs en germes ICF sont observées lorsque les conditions marée haute - déversement des rivières sont réunies.

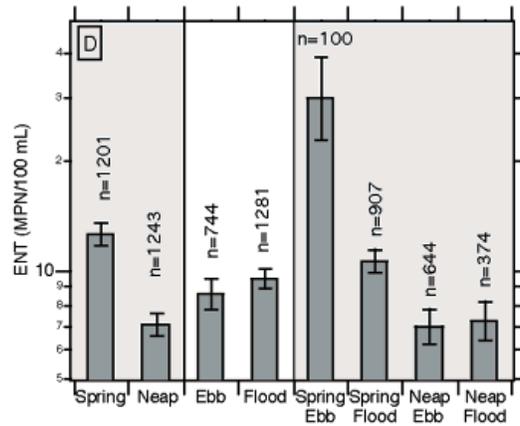


Fig. 5.7 – Influence de la marée sur la concentration en EI. (Spring : hautes eaux, Neap : basses eaux, Ebb : déversement des rivières dans la mer, Flood : reflux de la mer dans les estuaires). Source : (Boehm et Weisberg, 2005)

La FIG. 5.8 montre l'influence des marées sur la teneur en EI. Les concentrations maximales sont obtenues lors de marées de hautes eaux (14 à 15 jours après la pleine lune) alors que les concentrations les plus faibles sont obtenues lors de marées de basses eaux.

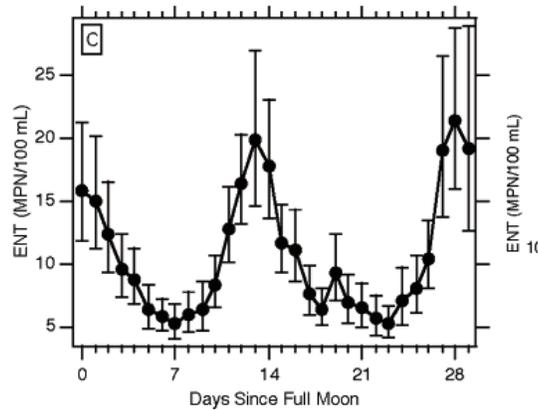


Fig. 5.8 – Influence du cycle lunaire sur la teneur en ICF. Source : (Boehm et Weisberg, 2005)

La campagne de mesure de l'USEPA de 2005 confirme le lien très important entre le niveau de la marée et la concentration en ICF comme le montre la FIG. 5.9.

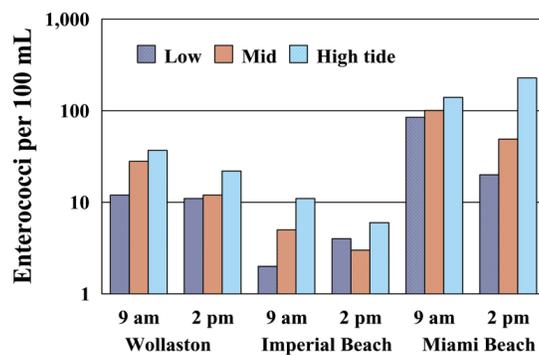


Fig. 5.9 – Influence de la marée sur la concentration en EI. Source : (Wymer *et al.*, 2005)

### 5.5.2 Influence de l'ensoleillement sur la teneur en ICF

La campagne de mesure menée par l'USEPA en 2005 (Wymer *et al.*, 2005) a montré l'influence du soleil sur la teneur en germes ICF mesurée dans les eaux de baignade. Les concentrations les plus élevées sont constatées les jours nuageux (cf. FIG. 5.10). Un jour nuageux est défini comme un jour où la couverture nuageuse excède 50%. Par ailleurs, les concentrations

---

en ICF sont plus élevées le matin que l'après-midi (+17% les jours nuageux, et +11% les jours ensoleillés).

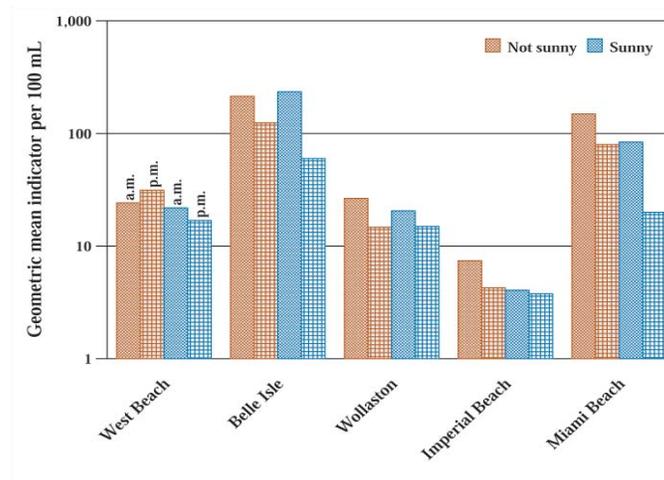


Fig. 5.10 – Influence du soleil sur la teneur en ICF. Source : (Wymer *et al.*, 2005)

### 5.5.3 Influence d'autres facteurs environnementaux

D'autres facteurs environnementaux sont également corrélés à la teneur en germes ICF : la pluviométrie précédant le prélèvement, le vent (vitesse et direction), la température de l'air, la température de l'eau, la turbidité, la hauteur des vagues (Wymer *et al.*, 2005; Olyphant et Whitman, 2004; Olyphant, 2005; Nevers et Whitman, 2005, 2008).

Les résultats de l'étude de l'USEPA (Wymer *et al.*, 2005) sont présentés en FIG. 5.11. Dans la majorité des sites d'étude, les concentrations en ICF sont plus importantes lorsque le vent souffle dans la direction plage->mer. Une hypothèse serait que le vent transporte des sédiments chargés en pollution fécale vers les zones de baignade. L'influence de la pluviométrie sur la teneur en germes ICF est également très marquée. Enfin, chose surprenante, les concentrations élevées en germes ICF ne sont pas liées à des températures élevées de l'air et de l'eau. Les températures de l'air et de l'eau sont corrélées à l'ensoleillement qui a un effet bactéricide.

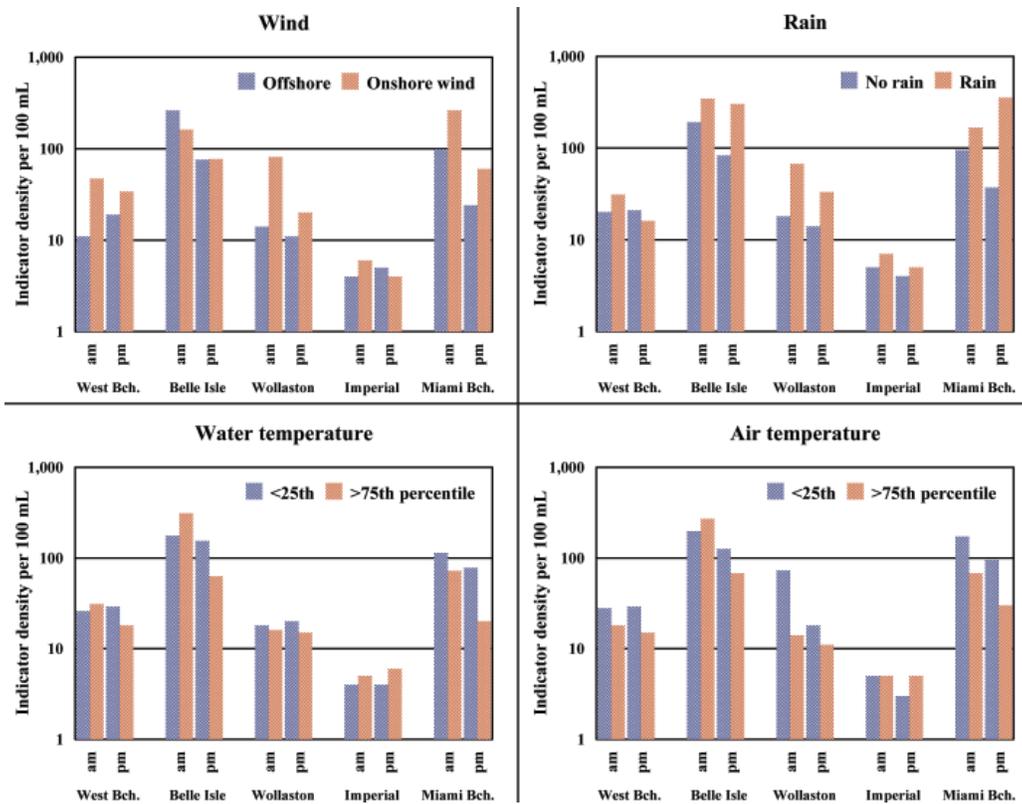


Fig. 5.11 – Influence de la pluviométrie, du vent, de la température de l'eau et de l'air sur la teneur en ICF. Source : (Wymer *et al.*, 2005)

## 5.5.4 Prédiction de la teneur en ICF à partir de paramètres environnementaux

Depuis moins de dix ans, des études s'intéressent à la prédiction en temps réel de la teneur en ICF des eaux de baignade à partir de mesures de paramètres environnementaux divers comme la température de l'eau, la turbidité, l'ensoleillement, la vitesse du vent, le niveau d'eau,... (Olyphant et Whitman, 2004; Olyphant, 2005; He et He, 2008). Les variables incluses dans chacun des modèles sont indiquées dans le TAB. 5.1. Le niveau de l'eau (lac ou mer), la hauteur des vagues et la turbidité sont les variables les plus utilisées dans les modèles de prédiction d'ICF.

Variables	Etudes*					
	1	2	3a	3b	4	5
Direction vent					X	
Vitesse vent	X	X				
Pluviométrie	X	X	X	X	X	X
Niveau d'eau	X	X	X	X		X
Hauteur des vagues		X	X		X	X
Ensoleillement	X	X				
Température eau	X		X	X		
Température air		X				
Turbidité	X			X	X	X
Turbidité source ponctuelle					X	X
Conductivité			X	X		
pH			X			
Chlorophylle						X

(\*) – (1) : Olyphant (2005) - (2) : Olyphant & Whitman (2004) - (3a) : He et al. (2008) (modèle ANN-4) - (3b) : He et al. (2008) (modèle ANN-1) - (4) : Nevers & Whitman (2008) - (5) : Nevers & Whitman (2005)

Tab. 5.1 – Variables incluses dans les modèles de prédiction d'ICF

La plupart des modèles de prédiction de la teneur en ICF sont des modèles statistiques de régression linéaire multivariés (MLR). He et He (2008) ont quant à eux utilisé des modèles non-linéaires basés sur des réseaux de neurones. Les résultats obtenus dans cette étude sont bons et semblent meilleurs que ceux obtenus avec des régressions linéaires multivariés (cf. TAB. 5.2). Le modèle testé donne peu de faux positifs et de faux négatifs (<10%). Boehm *et al.* (2007) ont comparé les résultats issus de nombreux modèles linéaires et non linéaires établis sur des données recueillies à Huntington State Beach (USA). Il ressort que les meilleurs modèles sont les modèles Partial Least Squares (PLS).

Toutes les études précédemment citées utilisent différentes techniques de sélection de variables :

- Nevers et Whitman (2005) : choix empirique et test de signification statistique des coefficients de régression ;
- Nevers et Whitman (2008) : sélection en deux étapes. 1- Sélection de variables qui n'ont pas de corrélation entre elles > à 0,6. 2- Affinement de la sélection basé le critère d'information d'Akaike (AIC). Ce critère permet de comparer plusieurs modèles entre eux, en pénalisant les modèles comportant trop de variables.
- Olyphant et Whitman (2004); Olyphant (2005) : sélection de variables suivant la valeur de la corrélation entre chaque variable et la variable à prédire.
- He et He (2008) sélection en deux étapes : 1- tests statistiques. 2- algorithmes de sélection génétiques.

$r^2$	<b>Olyphant (2005)</b>	<b>Olyphant &amp; Whitman (2004)</b>	<b>He et al. (2008)</b>	<b>Nevers &amp; Whitman (2008)</b>	<b>Nevers &amp; Whitman (2005)</b>
$r^2$ calibration	0,71	0,58	0,62-0,88	0,35	0,32-0,64
$r^2$ validation	-	-	0,69-0,77	0,48	-

Tab. 5.2 – Coefficient de détermination  $r^2$  obtenus pour les différents modèles de prédiction des concentrations en ICF

---

## Conclusion

### Considérations générales

La connaissance des facteurs influençant la probabilité d'une forte teneur en germes ICF d'une eau de baignade est primordiale pour diminuer les risques sanitaires liés à la baignade. Il a été montré que la pollution d'origine fécale a une variabilité spatiale importante, principalement perpendiculairement à la plage. Les teneurs en ICF les plus fortes sont rencontrées au bord des plages. En milieu marin, les marées ont également une forte influence sur les concentrations. Les teneurs les plus élevées sont observées en période de hautes eaux. L'influence d'apports de pollution fécale par les rivières semble secondaire par rapport à l'influence des marées. D'autres facteurs environnementaux sont également liés à la présence et la persistance d'ICF dans l'eau : l'ensoleillement, le vent, la pluie (qui est l'origine de rejets d'eaux usées),... Notons enfin que la variation temporelle des teneurs en ICF dans les eaux de baignade est considérable. Il existe un facteur deux entre les concentrations mesurées le jour  $j$  et le jour  $j+1$ . Par ailleurs, il n'existe plus de corrélation significative entre la teneur en ICF mesurée sur deux échantillons prélevés à trois ou quatre jours d'intervalle. Tous ces points sont autant de facteurs à prendre en compte pour l'identification de situations probables de pollution à court terme.

L'identification la plus sûre d'une pollution à court terme consiste à dénombrer en laboratoire les germes ICF. Toutefois, les méthodes normalisées basées sur des mises en culture d'échantillons nécessitent une durée d'incubation de 36h. Ce temps d'analyse couplé à la variabilité temporelle des germes ICF dans les eaux de baignade rend impossible toute mesure de prévention. De nouvelles méthodes de mesure, comme les méthodes enzymatiques rapides ou les méthodes PCR, permettent d'obtenir un résultat de dénombrement en moins de 2h. Cependant, ces techniques restent chères, demandent un personnel très spécialisé, et ne peuvent pas être implémentées in-situ. Certaines études pointent également le manque de robustesse de ces méthodes. Il est donc difficilement envisageable pour toutes les collectivités de mettre en oeuvre quotidiennement de telles techniques.

Des approches prédictives simples basées sur une étude statistique du lien entre germes ICF et paramètres environnementaux permettraient de compléter un dispositif d'alerte basé sur des mesures en laboratoire et d'alléger le programme de mesures. De nombreuses études, principalement américaines, ont établi une relation permettant de prédire la teneur en germes

---

ICF à partir de mesures de facteurs environnementaux effectuées in-situ et en temps réel. Les paramètres environnementaux le plus souvent inclus dans ces relations sont la pluviométrie, la hauteur d'eau, la hauteur des vagues, la turbidité, le vent (vitesse et direction). Ces approches sont loin d'être aussi performantes que des mesures effectuées en laboratoire. Toutefois, elles sont un outil d'aide à la décision supplémentaire pour les gestionnaires de sites de baignade et permettent d'optimiser le programme d'échantillonnage en fonction du risque prédit.

Notons enfin que les mesures en laboratoire et la prédiction de la teneur en ICF font le postulat -fait également par la réglementation- que le risque sanitaire est corrélé avec la présence de germes ICF. Or des études récentes montrent que E. coli et EI sont capables de croître dans des milieux aquatiques, dans des sédiments ou sur des algues. Les sources de ces micro-organismes ne sont pas que d'origine fécale comme on pouvait le croire jusque récemment. Ainsi, d'autres approches de détection plus spécifique d'une pollution fécale d'origine humaine se développent rapidement. Elles sont basées sur la détection de traceurs chimiques de pollution fécale. Les produits recherchés sont soit consommés par l'homme (caféine, produits pharmaceutiques), soit issus de notre métabolisme (stérols, stanols), soit liés à nos activités (azurants optiques). Ces traceurs sont généralement détectés par des techniques de pointe (GC-MS, HPLC), chères et nécessitant un personnel très qualifié. Dans l'ensemble de ces substances, la recherche d'azurants optiques semble prometteuse. En effet, elle peut être effectuée par spectrofluorométrie. Bien que non spécifique, cette méthode peut-être mise en place in-situ avec des spectrofluoromètres de terrain.

### **Apports de l'étude bibliographique dans le cadre du projet GIRAC**

Cette étude bibliographique ne nous a pas permis d'identifier des travaux de recherche portant sur la détection de pics de pollution bactériologique par spectrophotométrie UV/Visible. Cependant, les modèles statistiques de prédiction des concentration en E. coli évoqués précédemment intègrent souvent la turbidité comme variable explicative. La turbidimétrie étant une méthode voisine, mais plus grossière que la spectrophotométrie UV/Visible, il ne semble donc pas aberrant de chercher à identifier un lien potentiel entre spectres d'absorbance et teneur en germes ICF.

De plus, au regard de la revue bibliographique, la spectrofluorométrie semble être une technique pertinente pour le suivi in-situ en continu de traceurs chimiques de pollution fécale comme les azurants optiques. Dans l'hypothèse où l'évaluation des concentrations en germes

---

ICF par spectrophotométrie UV/Visible s'avérait impossible, cette piste de recherche mériterait d'être considérée.

Enfin, la complexité de la dynamique des germes ICF au niveau des zones de baignade a été mise en avant. L'évolution des concentrations est liée à de multiples facteurs environnementaux : marée, hauteur des vagues, vent,...Il est donc peu probable qu'un modèle hydrodynamique simulant une dispersion et une mortalité dans la mer des germes ICF provenant de sources ponctuelles, réussisse seul à prédire de manière adéquate les concentrations en germes ICF au niveau des sites de baignade.

## **Chapitre 6**

# **Applicabilité de la spectrophotométrie UV/Visible pour la détection de pics de pollution bactériologique**

### **Introduction**

Le projet GIRAC a pour objectif d'apporter aux collectivités des dispositifs de détection de pollution à court terme complémentaires aux mesures microbiologiques effectuées en laboratoire. La revue bibliographique a permis d'établir le lien entre facteurs environnementaux et germes ICF. Il a notamment été montré la corrélation entre turbidité et teneur en germes ICF sur de nombreux sites de baignade. La turbidité est souvent la conséquence de rejets d'eaux résiduares par temps de pluie.

Le projet GIRAC explore une piste analogue : il a été décidé d'étudier la possibilité de détecter des pollutions microbiologiques en rivière par spectrophotométrie UV/Visible. Ce choix est fondé sur deux hypothèses :

1. la qualité microbiologique des eaux de rivière étudiées est fortement conditionnée par l'absence ou la présence de rejets d'assainissement ;

- 
2. Il existe une corrélation entre paramètres de pollution physico-chimiques associés aux rejets d'eaux résiduaires urbaines (DCO, MES, Turbidité) et germes ICF.

La spectrophotométrie UV/Visible permettant la mesure in-situ en continu de paramètres globaux de pollution tels que la DCO ou les MES, nous recherchons donc dans ce projet à établir un lien indirect entre spectre d'absorbance et pollution d'origine fécale (cf. FIG. 6.1).

La recherche de corrélations entre teneur en *E. coli* et absorbances a été menée sur les eaux de la rivière Stang Alar qui se rejette dans la mer au niveau de la plage du "Moulin Blanc" dans la commune de Brest. L'objectif est d'évaluer en continu les apports de contamination fécale dans le milieu marin.

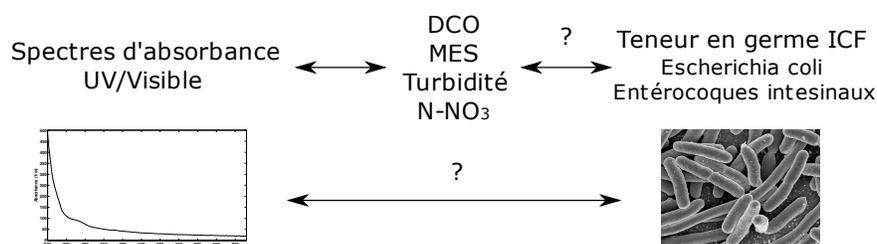


Fig. 6.1 – Représentation schématique de la problématique de l'étude visant à établir une relation entre spectres d'absorbance et concentration en *E. coli*

Cette méthode n'est bien entendu pas spécifique et des difficultés sont attendues. Les sources des MES et de DCO mesurées en rivière ne sont pas forcément liées au rejet d'eaux résiduaires urbaines. L'augmentation du débit constatée lors d'évènements pluvieux génère une remise en suspension de matières organiques. Les épisodes pluvieux lessivent en outre le bassin versant entraînant ainsi des matières organiques et minérales vers la rivière.

La méthodologie appliquée est inspirée de la procédure d'établissement de modèle prédictif décrite au §3.3 et comporte les étapes suivantes :

1. Collecte conjointe des données de concentration en *E. coli* et des spectres d'absorbance ;
2. Analyse de la qualité des données ;
3. Étude du lien possible entre concentration en *E. coli* et mesures d'absorbances.

---

## **6.1 Collecte des données de concentrations en E. coli et des spectres d'absorbance**

### **6.1.1 Site étudié - Mesures**

Le dispositif expérimental est situé à l'exutoire du bassin versant du Stang Alar sur la commune de Brest (cf. Fig. 6.2). La rivière Stang Alar qui traverse le bassin versant -et qui lui a donné son nom- se rejette dans la mer à proximité immédiate de la plage dite du "Moulin Blanc".

Le bassin versant a une surface de 580 ha. Sa partie Ouest est très urbanisée. Le réseau d'assainissement est principalement séparatif : les eaux pluviales sont rejetées directement dans le Stang Alar sans traitement préalable. Des rejets d'eaux usées dans les canalisations d'eaux pluviales sont constatés par temps sec. Les eaux pluviales rejetées dans le milieu récepteur contiennent donc une part d'eaux usées fortement chargées en pollution d'origine fécale. Pour limiter les rejets d'eaux usées par temps sec, des petits seuils ont été installés à l'exutoire des canalisations d'eaux pluviales afin de rediriger le débit d'eaux usées de temps sec vers les canalisations d'eaux usées.

La partie Est du bassin versant est principalement rurale. Elle est donc moins sujette à des apports de contamination fécale d'origine anthropique.

Le dispositif expérimental est installé 500 m en amont de l'embouchure du Stang Alar. Les eaux du Stang Alar sont pompées avec un débit de 14 m<sup>3</sup>/h dans une cuve de mesure de 3 m<sup>3</sup> où est placé un spectrophotomètre UV/Visible (cf. FIG. 6.23 et 6.5). Le temps de séjour théorique des eaux dans la cuve est donc de 13 minutes.

Des mesures spectrophotométriques et une quantification de E. coli ont été effectuées durant les trois dernières saisons de baignade par Véolia eau. En 2009, en plus de E. coli, d'autres paramètres de pollution ont été quantifiés en laboratoire (MES, NO<sub>3</sub><sup>-</sup>, Conductivité, Entérocoques intestinaux (EI)).

### **6.1.2 Spectrophotomètre in-situ en ligne**

La sonde utilisée pour l'étude est un spectrophotomètre in-situ spectro::lyser fabriqué par spectro::can Messtechnik GmbH (Autriche). Une photo du spectrophotomètre est visible en FIG.

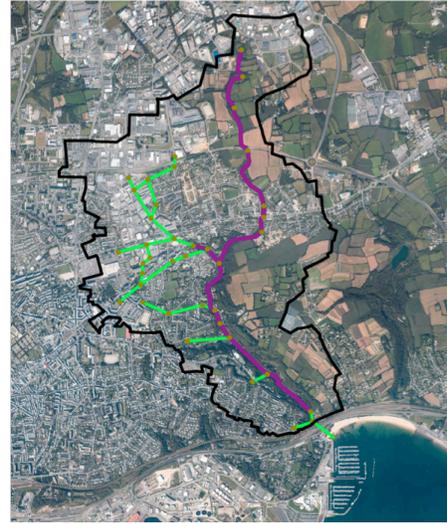


Fig. 6.2 – Bassin versant du Stang ALar à Brest



Fig. 6.3 – Photo du dispositif expérimental installé dans la station de pompage de Palaren (Brest)

---

6.4. Le chemin optique du modèle utilisé est de 35 mm. Le spectrophotomètre a été placé à la verticale au centre de la cuve de mesure, 40 cm en dessous de la surface libre (cf. FIG. 6.5). Une mesure de spectre d'absorbance a été effectuée toutes les trois minutes (200 nm à 730 nm) avec un nettoyage pneumatique des optiques avant chaque mesure.



Fig. 6.4 – Photo du spectrophotomètre utilisé



Fig. 6.5 – Spectrophotomètre immergé dans la cuve de mesure

### 6.1.3 Prélèvement et conservation d'échantillons

Afin d'établir un lien entre spectre d'absorbance et concentration en *E. coli*, des prélèvements d'échantillons ont été réalisés dans la cuve de mesure simultanément à une mesure de spectre.

En 2007, un opérateur de Véolia a prélevé chaque jour un échantillon d'eau dans la cuve de mesure. En 2008 et 2009, un préleveur réfrigéré automatique a été mis en place de manière à permettre une fréquence d'échantillonnage plus élevée. En 2008, le préleveur devait se déclencher si une valeur seuil d'absorbance était dépassée. Malheureusement, l'automatisa-

---

tion de ce système n'a pas fonctionné et seuls quelques échantillons prélevés ponctuellement ont pu être mis en correspondance avec les spectres d'absorbance mesurés aux moments des prélèvements. En 2009, des campagnes de mesures ont été organisées avec prélèvement automatique d'échantillons toutes les 01h30. Des échantillons ont également été prélevés lorsque des valeurs seuils d'absorbance étaient dépassées.

Les échantillons prélevés ont été transvasés dans des flacons stériles de 500 ml, puis conservés dans des glacières pendant leur transport vers le laboratoire. Tous les dénombrements de *E. coli* ont été effectués moins de 36 h après le prélèvement.

#### **6.1.4 Dénombrement de *E. coli***

Le dénombrement de *E. coli* a été effectué suivant la méthode NPP<sup>1</sup> miniaturisée ISO 9308-3. Cette méthode est basée sur l'approche *defined substrate technique* (Edberg *et al.*, 1991) qui consiste à ne fournir qu'un seul substrat dans le milieu de culture. Le substrat nécessite une activité enzymatique ciblée pour être hydrolysé et ne peut donc être utilisé que par le groupe de bactéries associé à cette activité enzymatique. Dans la méthode NPP miniaturisée, 200  $\mu$ l de plusieurs dilutions (1/2, 1/20,...) d'un échantillon sont placés dans 96 puits d'une microplaque contenant le substrat 4-méthylumbelliféryl- $\beta$ -D-glucuronide (MUGlu). La microplaque est incubée dans une étuve à 44°C pendant 36 à 48 h. Pendant l'incubation, l'hydrolyse du substrat par l'enzyme  $\beta$ -D-glucuronidase spécifique de *E. coli* donne lieu à un composé fluorescent, le méthylumbelliférol (MUF), qui peut être détecté sous une lampe à rayons ultraviolets. Le nombre de puits positifs, *i.e.* fluorescents, permettent la quantification de *E. coli* en utilisant une analyse statistique basée sur la loi de Poisson.

#### **6.1.5 Constitution de la base de données**

Après trois années de mesures, Véolia eau dispose d'une base de données de 226 mesures de spectres d'absorbance avec les concentrations en *E. coli* correspondantes (108 en 2007, 10 en 2008 et 108 en 2009).

---

<sup>1</sup>NPP : nombre le plus probable

---

## 6.2 Analyse de la qualité des mesures

### 6.2.1 Dérive des mesures spectrophotométriques

L'étude de la dérive du signal n'a pas été incluse directement dans le protocole de mesure expérimental. Une évaluation en temps-réel aurait été permise par des mesures de contrôle réalisées sur de l'eau distillée tout au long de la période de mesure et/ou par une observation plus fine des signaux d'absorbance au cours de chaque saison de baignade.

Afin d'identifier *a posteriori* une possible dérive du signal, nous avons cherché à comparer tout au long de la période de mesure des spectres correspondant à des conditions environnementales similaires (temps-sec).

#### 6.2.1.1 Dérive du signal en 2009

Dix spectres obtenus par temps sec<sup>2</sup> ont été sélectionnés sur l'ensemble de la période de mesure (cf. FIG. 6.6 et 6.7).

La cuve de mesure et le spectrophotomètre ont été nettoyés début août et les mesures ont été interrompues pendant 15 jours. L'effet du nettoyage de la sonde spectrophotométrique et de la cuve est visible : après le 15 août les spectres de temps sec ont des valeurs moins élevées que lors de la période de mesure juin-juillet (cf. aussi TAB. 6.1).

Par ailleurs, on remarque qu'au fur et à mesure des jours écoulés, les valeurs d'absorbance ont tendance à augmenter, et ce particulièrement dans le domaine de l'UV. La variation des valeurs d'absorbance pour les longueurs d'onde du domaine visible est minime.

Période	250 nm		700 nm	
	Moy.	E. Type	Moy.	E. Type
Juin-Juillet 2009	7,98	1,08	0,65	0,05
Après le 15 août 2009	5,81	0,97	0,26	0,06

Tab. 6.1 – Variabilité des spectres de temps sec à 250 et 730 nm

---

<sup>2</sup>Aucune précipitation n'a été observée pendant les 8 heures qui précèdent la mesure du spectre

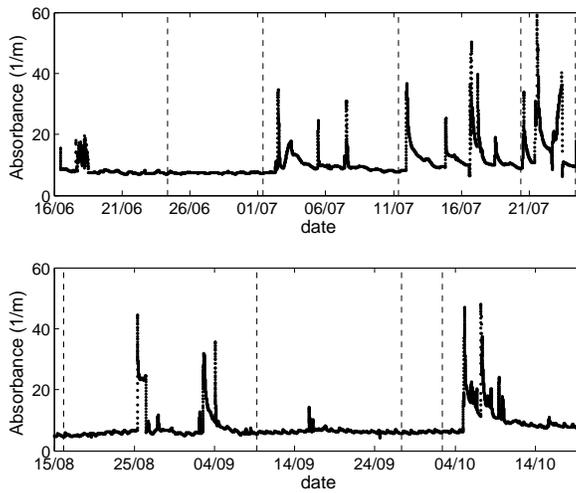


Fig. 6.6 – Évolution de l'absorbance à 250 nm en 2009

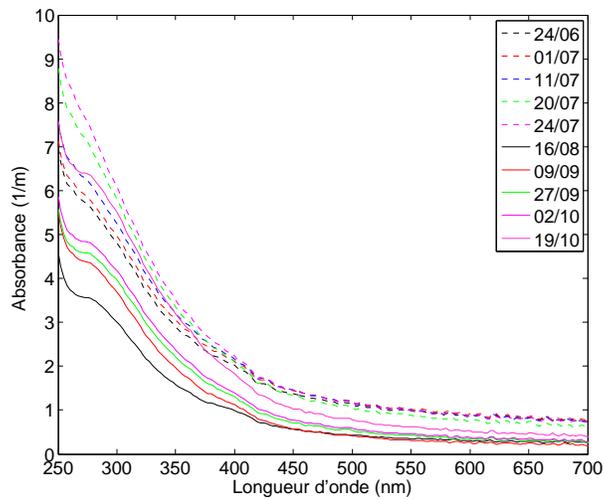


Fig. 6.7 – Dix spectres de temps sec mesurés en 2009

---

### 6.2.1.2 Dérive du signal en 2007

En 2007, une dérive des spectres de temps sec a également été constatée (cf. FIG. 6.8). Comme en 2009, la dérive affecte principalement le domaine de l'UV.

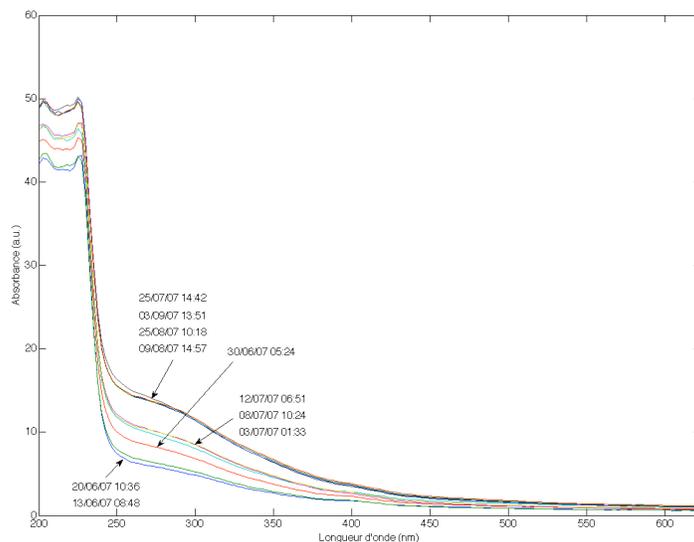


Fig. 6.8 – Évolution des spectres de temps sec en 2007

### 6.2.1.3 Origine de la dérive

La dérive est particulièrement observable sur le domaine [200 nm ; 550 nm] et en particulier dans l'ultraviolet (cf. FIG. 6.8). Ce domaine permet habituellement la détermination de matières organiques dissoutes ((Mrkva, 1983) et cf. §3.5.1). Plusieurs hypothèses peuvent être émises quand à son origine :

1. La dérive provient d'un encrassement des optiques par des matières colloïdales malgré le nettoyage à l'air comprimé.
2. La dérive a pour origine le piégeage de matières organiques dans la cuve de mesure.

La première hypothèse est vraisemblable. Les optiques sont certes nettoyées avec de l'air comprimé à chaque cycle de mesure, c'est à dire toutes les 3 minutes. Toutefois le nettoyage

---

pneumatique n'est pas totalement efficace. En effet, la sonde a été nettoyée manuellement une fois au cours de la saison de baignade 2009 et a permis de rabaisser la ligne de base du spectre d'absorbance à son niveau initial. Après étude de la dynamique de la dérive, une fréquence de nettoyage de 15 jours est recommandée.

La seconde hypothèse est également plausible. En effet, lors d'une visite ponctuelle à Brest en Juillet 2009, un piégeage des matières en suspension en fond de cuve a été constaté. La vidange de la cuve se fait par surverse ; elle fonctionne donc comme un décanteur. Les matières en suspension sont donc piégées. Même si la sonde spectrophotométrique est placée dans la partie supérieure de la cuve, il est probable que la configuration de la cuve entraîne l'augmentation des concentrations en MES sur l'intégralité du profil vertical.

#### **6.2.1.4 Conséquence sur la recherche de corrélations avec les concentrations en E. coli**

La dérive des spectres d'absorbance dans le domaine de l'UV est préjudiciable pour la recherche de corrélations entre absorbances et concentration en E. coli. En effet, ce domaine spectral est utilisé pour sa corrélation avec les concentrations en matières organiques et en DCO (Mrkva, 1983). Les rejets d'assainissement étant caractérisés par des concentrations fortes en DCO et par la présence de matières fécales, une corrélation indirecte entre concentration en E. coli et absorbance dans l'UV était attendue.

Malheureusement, du fait de la dérive, les corrélations obtenues sur de longues séries temporelles entre E. coli et absorbance dans l'UV ont de fortes chances d'être médiocres.

L'absorbance dans le visible, en *queue* de spectre, sera beaucoup moins sujette à la dérive du signal. Même si la recherche de corrélations sur ce domaine spectral semble moins pertinente, il est vraisemblable qu'elle donne de meilleurs résultats pour des études statistiques concernant des périodes d'un mois ou plus.

### **6.2.2 Biais dans le protocole de prélèvement**

Le protocole de prélèvement utilisé en 2007 et 2009 comporte plusieurs biais.

Tout d'abord, les échantillons ont majoritairement été prélevés en matinée, ce qui rend difficile l'étude de la variation journalière des concentrations, particulièrement en temps sec (cf. FIG.

---

6.9). Ceci concerne uniquement les périodes où la fréquence de prélèvement était journalière (intégralité de la saison de baignade 2007 et une partie de la saison 2009).

De plus, lors de l'utilisation de préleveur automatique réfrigéré en 2009, les flacons du préleveur n'étaient pas stériles. Les concentrations en *E. coli* peuvent donc être faussées. Les flacons étaient rincés à l'eau potable entre deux utilisations successives du préleveur.

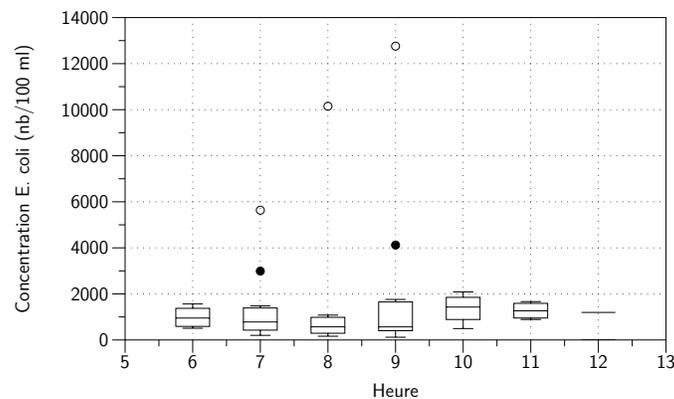


Fig. 6.9 – Valeurs des concentrations de *E. coli* e temps sec et par tranche horaire

### 6.2.3 Conclusion

Les trois principales observations sur le protocole expérimental portent donc sur :

1. la dérive du signal dans l'UV rendant problématique la recherche de corrélations avec les concentrations en *E. coli* sur ce domaine spectral ;
2. la vidange de la cuve de mesure par surverse. Celle-ci favorise en effet le piégeage des matières en suspension ;
3. le prélèvement systématique d'échantillons entre 6h00 et 12h00. Cette organisation ne permet pas d'étudier les variations journalières des concentrations en *E. coli* par temps sec.

Des propositions d'améliorations du protocole, basées sur l'expérience acquise au cours de ces trois dernières années ainsi que sur l'étude bibliographique, seront formulées au §6.4.

---

## **6.3 Recherche d'une relation spectre d'absorbance - concentration en E. coli**

Le but de cette étude est de rechercher une relation entre spectre d'absorbance et concentration en E. coli. à partir des mesures effectuées entre 2007 et 2009. Nous réaliserons des études statistiques à l'échelle de chaque saison de baignade ainsi que des observations à l'échelle événementielle. Pour des raisons de dérive du signal dans l'UV, les signaux d'absorbance seront uniquement étudiés dans le visible.

### **6.3.1 Analyse des mesures réalisées en 2007**

#### **6.3.1.1 Étude statistique sur l'intégralité de la saison de baignade**

##### **6.3.1.1.a Recherche d'une relation linéaire entre concentration en E. coli et absorbance à 730 nm**

Cent huit mesures conjointes de spectres d'absorbance et de concentrations en E. coli ont été effectuées en 2007. Sept des cent huit spectres d'absorbances ont été écartés. Deux d'entre eux sont saturés, très bruités et correspondent à la limite d'utilisation du spectrophotomètre. Un spectre d'absorbance est associé à une valeur aberrante de dénombrement. Enfin, les quatre autres spectres ont été éliminés en raison d'incohérences de mesures d'absorbance par rapport aux échantillons ayant des concentrations en E. coli voisines.

Le coefficient de détermination  $r^2$  entre la concentration en E. coli et l'absorbance a été calculé pour chacune des longueurs d'onde. Les résultats sont présentés en FIG. 6.10. La valeur maximum de  $r^2$  est égale à 0,77 ( $r=0,876$ ). Elle est atteinte à la longueur d'onde de 730 nm. Les maximums correspondent au domaine spectral du visible où sont quantifiés les MES et la turbidité<sup>3</sup>. Ce résultat est donc cohérent avec ceux obtenus dans les études citées au chapitre 5 où la turbidité avait été sélectionnée pour l'élaboration de modèles prédictifs de concentration en ICF. Toutefois, dans notre cas la corrélation entre la concentration en E. coli et la turbidité est plus importante que celle observée dans les études de Olyphant (2005)

---

<sup>3</sup>Selon la méthode 180.1 de l'US EPA, la turbidité est mesurée par néphélométrie dans la plage longueur d'onde 400-600 nm. La norme européenne préconise quant à elle une mesure par transmission dans le proche infrarouge à 860 nm

( $r_{max}=0,39$ ), Nevers et Whitman (2005) ( $r_{max}=0,4$ ) et He et He (2008) ( $r_{max}=0,66$ ). Ceci est certainement lié au fait que nous recherchons des corrélations directement après le rejet et non au niveau du site baignade. Enfin, notons que le coefficient de détermination  $r^2$  est très faible dans l'UV, très certainement pour les raisons évoquées au paragraphe précédent (dérive du signal).

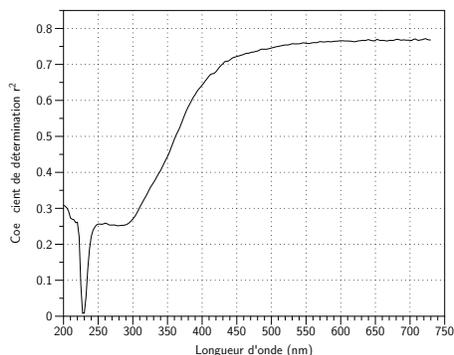


Fig. 6.10 – Coefficient de détermination entre la concentration en *E. coli* et l'absorbance (2007)

La FIG. 6.11 représente les concentrations en *E. coli* en fonction de l'absorbance à 730 nm sur l'ensemble de la saison de baignade. L'équation de la régression linéaire entre les deux variables est donnée par :

$$E. coli (nb/100ml) = 3154,7 (\pm 228,1) \cdot A_{730} (1/m) - 1259,1 (\pm 833,9) \quad (6.1)$$

où les chiffres entre parenthèses sont les intervalles de confiance à 95 % des coefficients de régression.

L'intervalle de confiance à 95 % d'une valeur prédite est donné par  $\hat{y}_0 \pm 7600$  *E. coli*/100 ml. La qualité de la prédiction n'est donc pas satisfaisante.

Quelques remarques méritent néanmoins d'être formulées quant à la régression effectuée. Tout d'abord, la valeur très importante de l'intervalle de confiance à 95 % des valeurs prédites est principalement due à deux individus qui génèrent plus de 70 % des erreurs quadratiques (cf. FIG. 6.11). D'autre part, le nombre d'individus dont l'absorbance à 730 nm dépasse  $6 \text{ m}^{-1}$  est seulement de trois et à tendance à déséquilibrer la régression. Les deux individus dont

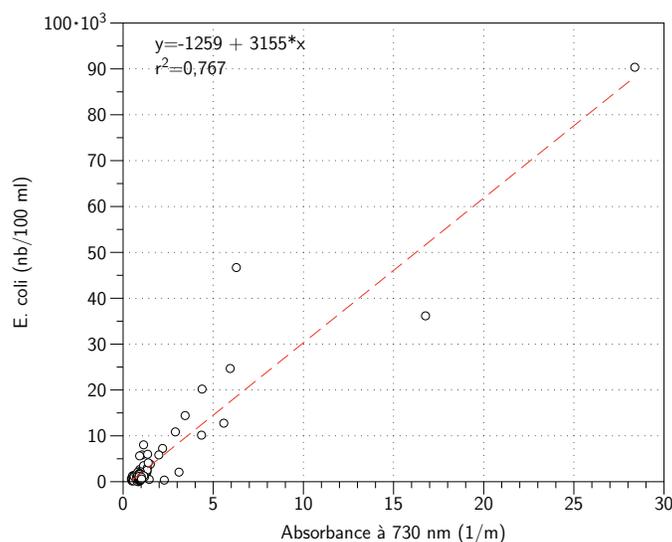


Fig. 6.11 – Concentrations en *E. coli* en fonction de l'absorbance à 730 nm (2007)

l'erreur de prédiction est la plus importante font partie de ces trois mesures. Leur inclusion dans la régression est donc discutable. Elle a néanmoins l'avantage d'étendre la capacité de prédiction pour des valeurs d'absorbance supérieures à  $6 \text{ m}^{-1}$ .

Une régression sans ces trois points a été effectuée. Le graphique de la régression est présenté en annexe D. L'équation de régression devient :

$$E. coli (nb/100ml) = 3619,7 (\pm 403,1) \cdot A_{730} - 1955,4 (\pm 608,7) \quad (6.2)$$

L'intervalle de confiance à 95 % d'un valeur prédite diminue logiquement et vaut  $\hat{y}_0 \pm 3820$  *E. coli*/100 ml. Les deux équations aboutissent à des prédictions très proches pour les faibles valeurs d'absorbance ( $A < 6 \text{ m}^{-1}$ ). La différence porte essentiellement sur la prédiction des pics de concentration de *E. Coli*. La seconde équation apparait à ce titre plus adaptée dans un objectif *sécuritaire*. Il serait intéressant de déterminer avec des simulations hydrodynamiques en milieu marin si la différence de prédiction à l'exutoire du bassin versant a une influence sur les concentrations en *E. coli* au niveau des sites de baignade. Nous pouvons toutefois supposer, que compte-tenu de la dilution des eaux de rivière dans la mer et de l'influence conjuguée d'autres facteurs, comme la marée, la hauteur des vagues ou l'ensoleillement, les

écarts de prédiction obtenus avec les deux équations auront un impact négligeable sur les concentrations en E. coli au niveau des sites de baignade.

### 6.3.1.1.b Sensibilité des mesures spectrophotométriques aux concentrations en E. Coli.

L'objectif de ce paragraphe est d'étudier la sensibilité des mesures spectrophotométriques aux concentrations en E. Coli. Nous analyserons également l'applicabilité de méthodes de régression multivariées.

Les cent un spectres d'absorbance ont été partitionnés en six classes de concentrations présentées dans le TAB. 6.2 et la FIG. 6.12.

On observe que la sensibilité des mesures d'absorbance à 730 nm est médiocre pour les faibles valeurs de concentration de E. coli. En effet, les valeurs moyennes des classes 1-2 et 3-4 sont similaires de même que leur distribution. La FIG. 6.12 illustre qu'aucune discrimination n'est possible entre les classes 1 et 2 et dans une moindre mesure entre les classes 3 et 4. Le test statistique non-paramétrique de Mann-Whitney confirme en outre que les populations dont sont issus les échantillons de ces classes ont des moyennes d'absorbance à 730 nm identiques<sup>4</sup>.

Classe	E. coli (nb/100 ml)	Effectif	E. coli (nb/100 ml)		Abs(730 nm)	
			moy.	s	moy.	s
1	<1000	48	515	241	0,85	0,27
2	1000-2000	26	1364	268	0,88	0,19
3	2000-5000	12	2743	706	1,34	0,60
4	5000-10000	6	6432	979	1,43	0,53
5	10000-20000	4	12047	1928	4,07	1,17
6	>20000	5	43618	28094	12,35	10,22

Tab. 6.2 – Partition des 101 spectres d'absorbance de 2007 en classes de concentrations

Le TAB. 6.2 ne permet de comparer les individus que pour une seule longueur d'onde. Pour prendre en compte l'intégralité du spectre<sup>5</sup>, une analyse en composantes principales a été

<sup>4</sup>Les p-values obtenues pour tester l'équivalence des moyennes des populations sont de 0,39 pour les classes 1-2 et de 0,75 pour les classes 3-4

<sup>5</sup>L'ACP a en fait été restreinte à la plage de longueurs d'ondes 240-730 nm. Les premières longueurs d'ondes sont en effet sujettes à beaucoup d'erreurs de mesure lorsque les valeurs d'absorbance sont élevées

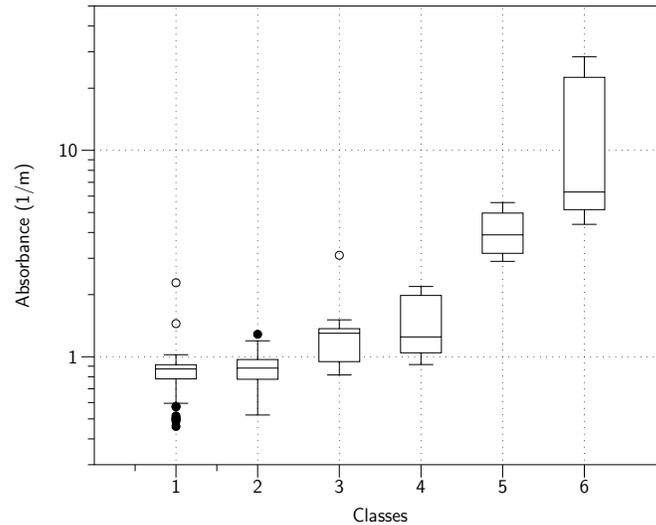


Fig. 6.12 – Distribution des valeurs d'absorbance pour chacune des six classes de concentration de E. coli (2007)

réalisée en incluant les cent un échantillons. La projection des spectres sur le plan 1-2 est représentée en FIG. 6.13. Le pourcentage de l'inertie capté par les deux premiers axes principaux est de 99,9 %, ce qui est très satisfaisant.

La FIG. 6.13 suscite de nombreuses remarques. Tout d'abord, l'étendue spatiale des individus au sein d'une même classe est très importante. Les classes 1 à 4 sont caractérisées par un nuage de points orienté en diagonale descendante. L'axe 2 sur lequel la dispersion des individus est majoritairement constatée correspond à un gradient de dérive du signal. L'axe 1 porte l'information concernant la concentration des individus. Il comporte également, dans une moindre mesure, une information sur la dérive du signal.

Par ailleurs, la projection des individus sur le plan 1-2 ne permet pas de réaliser une classification des individus par classe de concentration. Les classes 1,2,3 voire 4 sont indifférenciables. Il en est de même pour les classes 5 et 6. Les individus 35 et 97 appartenant à la classe 5 ont des scores sur le premier axe supérieurs aux individus 69 et 119 appartenant à la classe 6.

Ces observations montrent que l'emploi de méthodes de régression multivariées de type PCR ou PLS portant sur l'intégralité du spectre n'apportera aucune plus-value par rapport

---

une régression linéaire simple. De plus, elles confirment la faible sensibilité des mesures spectrophotométriques pour des faibles valeurs de concentration de E. coli.

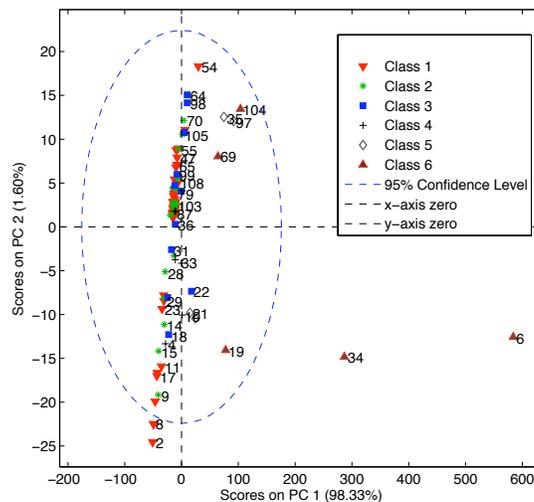


Fig. 6.13 – Analyse en composantes principales des spectres de 2007. Représentation des individus sur le plan 1-2

### 6.3.1.2 Observations à l'échelle événementielle

L'objectif de ce paragraphe est de comparer les concentrations en E. coli simulées et mesurées à l'échelle événementielle dans le domaine temporel.

L'équation de régression utilisée est l'équation (6.1) limitée par une valeur maximum de  $9,3 \cdot 10^4$  E. coli/100 ml.

Les graphiques illustrant les résultats de la régression sont présentés en annexe E. Les principales conclusions sont les suivantes :

- Les pics de concentrations simulés sont corrélés aux intensités de pluie, traduisant le fait que le spectrophotomètre capte la modification de la qualité des eaux du Stang-Alar suite aux événements pluvieux.

- 
- Le nombre de mesures effectuées pendant les événements pluvieux est insuffisant : 70 % des pics de concentration simulés sont supérieurs à 20 000 E. coli/100 ml alors que seulement 5 % des échantillons dépassent cette valeur.
  - De même, le pas de temps d'échantillonnage en temps sec ne permet pas de déterminer si le système est uniquement régi par une dynamique pluviale.

### 6.3.2 Analyse des mesures réalisées en 2008

En 2008, le protocole de prélèvement a été modifié. Véolia a souhaité prélever des échantillons au cours d'évènements pluvieux à l'aide d'un préleveur automatique asservi à un seuil d'absorbance. Ce dispositif n'a pas fonctionné ou très mal, si bien qu'il n'a pas été possible d'horodater les échantillons prélevés. C'est la raison pour laquelle nous disposons uniquement de dix échantillons (spectres et concentrations en E. coli).

La validité de cette base de données mérite d'être remise en cause. En effet, les corrélations entre les absorbances et les concentrations en E. coli sont cette année là insignifiantes (cf. FIG. 6.14). Nous ne réaliserons donc aucun traitement de données supplémentaire pour 2008.

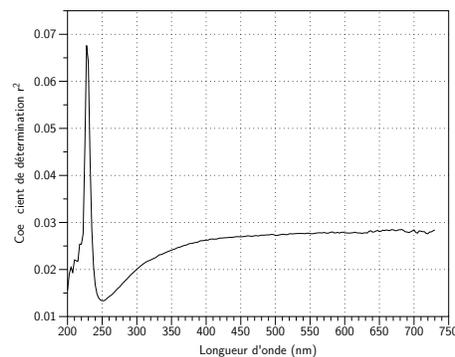


Fig. 6.14 – Coefficient de détermination entre la concentration en E. coli et l'absorbance (2008)

---

### 6.3.3 Analyse des mesures réalisées en 2009

#### 6.3.3.1 Étude statistique sur l'intégralité de la saison de baignade

Cent huit mesures conjointes de spectres d'absorbance et de concentrations en *E. coli* ont été acquises pendant la saison de baignade 2009.

Comme en 2008, les corrélations entre les concentrations en *E. coli* et les absorbances sont très faibles (cf. FIG. 6.15).

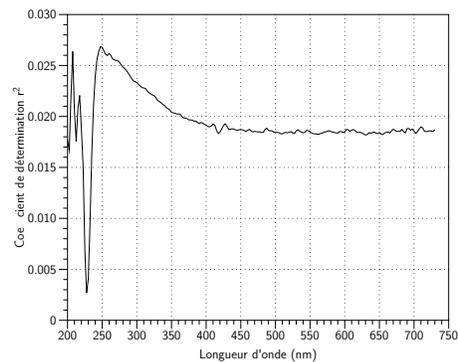


Fig. 6.15 – Coefficient de détermination entre la concentration en *E. coli* et l'absorbance (2009)

Les très faibles valeurs de  $r^2$  obtenues peuvent avoir des origines diverses comme :

- une modification radicale de la dynamique du système entre 2007 et 2009 due à des travaux sur le réseau d'assainissement ;
- un changement dans le protocole de prélèvement des échantillons. En 2007, un échantillon par jour était prélevé pendant la saison de baignade. En 2009, en plus des prélèvements journaliers, des campagnes de prélèvements ont été effectuées de manière à réduire le pas de temps d'échantillonnage notamment pendant les événements pluvieux ;
- des erreurs d'horodatage des échantillons prélevés et/ou des mesures spectrophotométriques.

Aucune information concernant les travaux effectués sur le réseau d'assainissement entre 2007 et 2009 n'a pu être collectée. Cette piste ne peut donc être investiguée.

---

L'influence du changement de protocole sur la valeur du coefficient  $r^2$  sera recherchée lors de l'observation des mesures à l'échelle événementielle. Nous disposons d'échantillons acquis en temps sec et en temps de pluie avec un pas de temps d'échantillonnage beaucoup plus fin (01h30 ou 2h). Ceci nous permettra d'étudier si un profil journalier de temps sec existe.

Enfin, les erreurs d'horodatage des mesures spectrophotométriques sont très peu probables. L'horloge de la centrale d'acquisition du spectrophotomètre a été vérifiée. Par contre, les erreurs d'horodatage des échantillons prélevés sont envisageables, les saisies des horaires de prélèvement étant manuelles. L'étude des chroniques de pluie et d'absorbance devrait nous permettre de repérer les erreurs d'horodatage les plus flagrantes.

Compte-tenu de l'absence de corrélation significative entre les absorbances et les concentrations en *E. coli*, aucune équation de régression ne sera recherchée.

### **6.3.3.2 Observations à l'échelle événementielle**

L'objectif de ce paragraphe est de comparer à l'échelle événementielle les concentrations mesurées et les concentrations simulées à l'aide de l'équation (6.1). Nous savons d'ores et déjà que cette relation ne pourra pas être validée sur tout ou partie de la chronique temporelle. Nous rechercherons alors si les concentrations mesurées peuvent être expliquées par d'autres facteurs qu'une dynamique pluviale.

#### **6.3.3.2.a Évènement pluvieux du 10/06/2009**

Sur cette série de données, le coefficient de détermination  $r^2$  vaut 0,51 à la longueur d'onde de 272,5 nm. A 730 nm, il ne vaut plus que 0,32. La dérive du signal dans l'UV est insignifiante sur une aussi courte période. Ceci explique très certainement la bonne valeur de  $r^2$  obtenue de l'UV.

La comparaison des concentrations en *E. coli* mesurées et simulées est présentée en FIG. 6.16. L'application du modèle de régression établi en 2007 reproduit convenablement la dynamique des concentrations mesurées. Toutefois, nous ne disposons pas de point de mesure permettant de valider l'intensité du pic de concentration simulé. La dynamique pluviale du système semble néanmoins plausible.

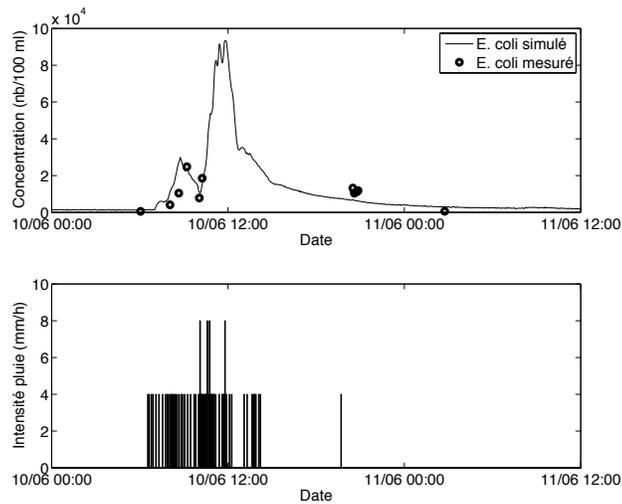


Fig. 6.16 – Comparaison des concentrations mesurées et simulées pendant l'événement pluvieux du 10/06/2009

### 6.3.3.2.b Période du 18/06/2009 au 25/06/2009

Ce second jeu de données correspond à une période de temps sec. Les données brutes d'absorbance étaient particulièrement bruitées pendant cette période. Ce phénomène est constaté lorsqu'un corps étranger vient obstruer périodiquement les optiques. Nous avons reconstitué le signal de temps sec égal correspondant à la ligne de base en filtrant le signal.

Le coefficient de détermination  $r^2$  ne dépasse pas 0,06 sur cette période. Pourtant, la FIG. 6.17 suggère qu'il existe un lien entre l'absorbance à 250 nm et les concentrations mesurées. On observe en effet entre le 21/06 et le 24/04 une dynamique similaire entre les concentrations mesurées et l'absorbance à 250 nm.

La FIG. 6.18 représente l'ajustement réalisé à partir des concentrations en E. coli et de l'absorbance à 250 nm. La fonction trouvée est égale à :

$$E.coli \text{ (nb/100 ml)} = \frac{3,2 \cdot 10^4}{1 + e^{-6 \cdot (A_{250} + 7,51)}} \quad (6.3)$$

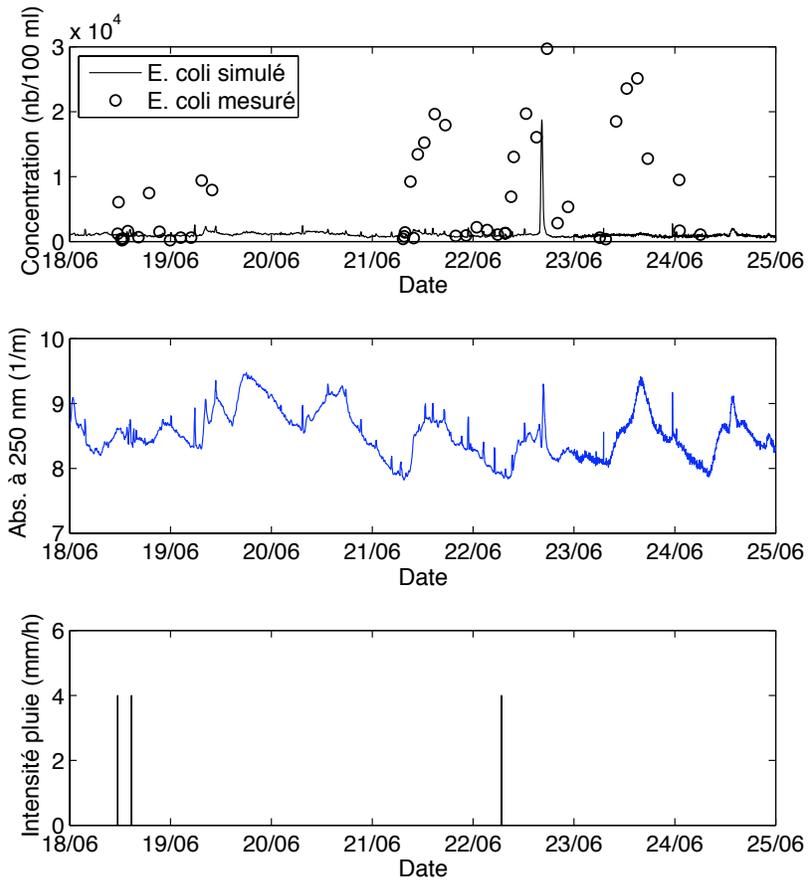


Fig. 6.17 – Comparaison des concentrations mesurées et simulées pendant la période du 18 au 25/06/2009

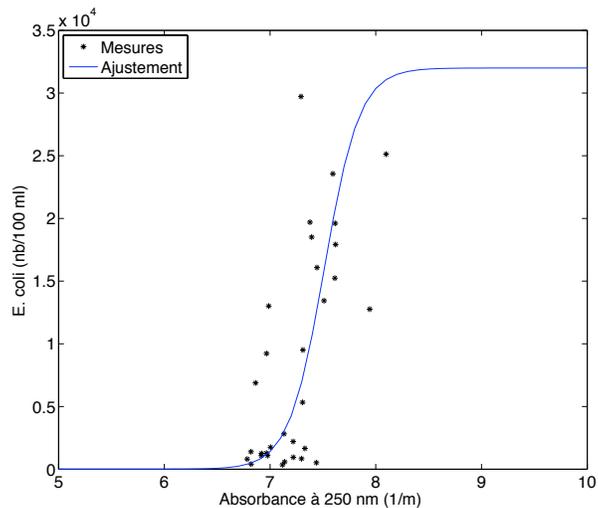


Fig. 6.18 – Ajustement d'une fonction  $E. coli = f(Abs_{250nm})$

Cette fonction a été tracée dans le domaine temporel et comparée aux mesures de concentration de *E. coli* (FIG. 6.19). Cette figure accrédite la thèse de l'existence d'un profil journalier de temps sec.

Toutefois, cet ajustement n'est pas pertinent pour les échantillons acquis entre le 18/06 et le 21/06. Nous avons alors tenté de répliquer un profil journalier moyen de temps sec. Le profil du 21/06 a été choisi. Cette méthode donne de meilleurs résultats pour la période du 21 au 25/06 comme le montre la FIG. 6.20 et le TAB. 6.3. Le modèle de régression linéaire établi sur les données de 2007 donne quant à lui de meilleurs résultats pour les échantillons prélevés le 18 et 19/06.

### 6.3.3.2.c Période du 01/07/2009 au 15/07/2009

Sur la troisième période analysée, les échantillons ont été prélevés au rythme de un par jour. La FIG. 6.21 illustre le biais dans le protocole de prélèvement où les échantillons sont systématiquement prélevés aux mêmes tranches horaires. Nous observons que le profil journalier de temps sec et le modèle basé sur la régression linéaire sont tous deux plausibles.



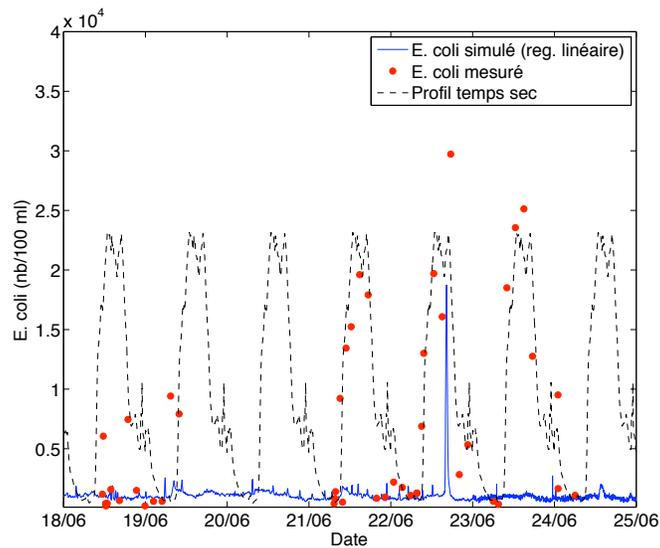


Fig. 6.20 – Applicabilité d'un profil journalier moyen de concentration en E. coli sur la période allant du 18/06 au 25/06/2009

#### 6.3.3.2.d Période du 15/07/2009 au 25/07/2009

Le dernier jeu de données concerne la période du 15 au 25 juillet 2009. L'horodatage de vingt six échantillons a été modifié. En effet, un décalage temporel de -1 jour, permet de retrouver une bonne correspondance entre concentrations mesurées et concentrations simulées. Les résultats sont présentés en FIG. 6.22.

La modification de l'horodatage des échantillons, qui peut apparaître comme scientifiquement discutable, mérite donc de considérer ces données avec une extrême précaution.

Le profil journalier de temps sec semble expliquer convenablement les concentrations mesurées le 15/07, le 16/07 et dans une moindre mesure le 17/07. En temps de pluie, la régression linéaire donne de meilleurs résultats comme nous pouvons le voir sur l'événement pluvieux du 21/07. Le 23/07, les mesures d'absorbance sont incohérentes. On observe en effet une montée et une descente brutale du signal de concentration simulée suite à deux pics de précipitations. Nous pouvons donc supposer qu'un corps étranger a obstrué les optiques du capteur pendant cette journée.

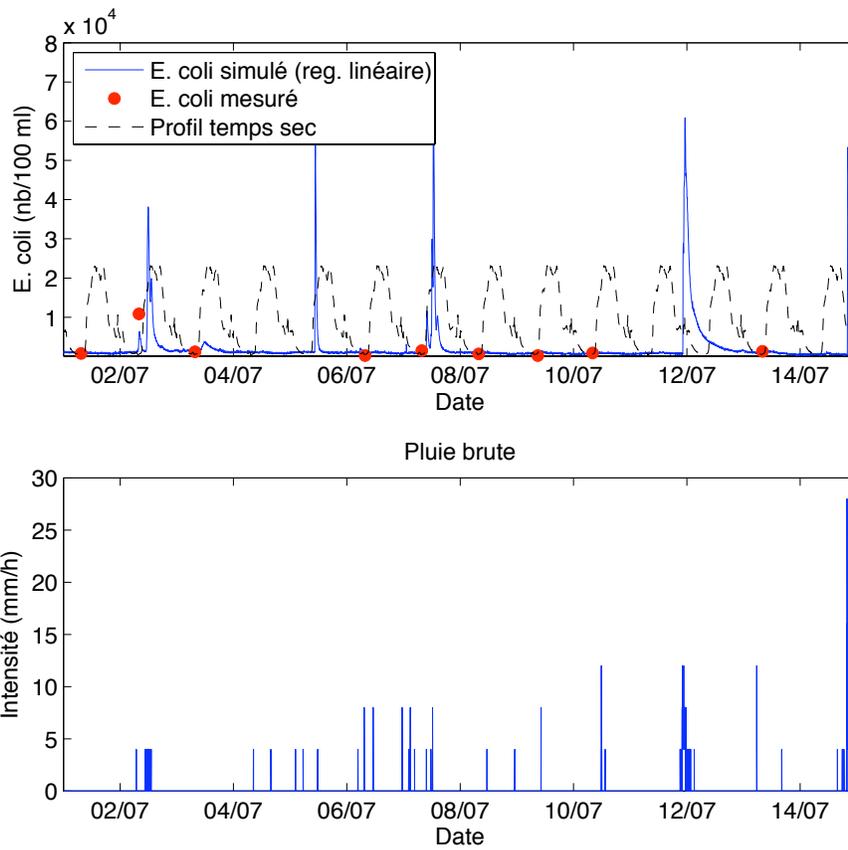


Fig. 6.21 – Comparaison des concentrations mesurées et simulées pendant la période du 01/07 au 15/07/2009

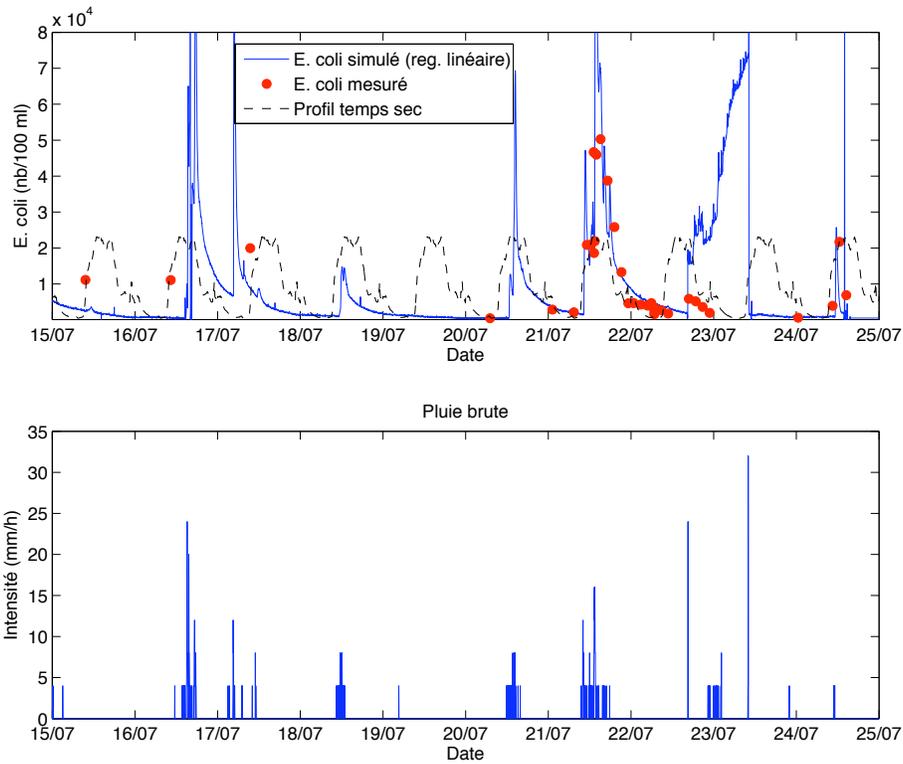


Fig. 6.22 – Comparaison des concentrations mesurées et simulées pendant la période du 15/07 au 25/07/2009

---

### 6.3.4 Synthèse

D'après les observations réalisées à l'échelle événementielle, le système est potentiellement régi par deux dynamiques distinctes :

- une dynamique de temps-sec où un profil journalier de concentration en *E. coli* est plausible. Ce profil indiquerait des rejets d'eaux usées de manière permanente dans le Stang Alar ;
- une dynamique de temps de pluie où les concentrations en *E. coli* sont corrélées à l'absorbance à 730 nm. L'augmentation d'absorbance n'est pas directement liée à l'apport de *E. coli* mais avant tout à une augmentation de la concentration en MES qui sont le vecteur d'*E. coli*.

Afin de vérifier l'applicabilité d'un profil journalier de concentration de *E. coli* ou de la régression linéaire, nous avons calculé la fonction d'erreur RMSE<sup>6</sup> pour chacune des périodes. Les résultats sont présentés dans le tableau 6.3. Nous avons également testé un modèle combinant à la fois le profil journalier et la régression linéaire, et correspondant à la valeur maximum de concentration obtenue avec l'application du profil ou de l'équation de régression :

$$E.coli (nb/100ml) = \max(E.coli_{profil}, E.coli_{régression}) \quad (6.4)$$

Période	RMSE ( <i>E. coli</i> /100 ml)		
	Profil journalier	Régression linéaire	Modèle
10/06/10	8,42 10 <sup>3</sup>	6,56 10 <sup>3</sup>	6,43 10 <sup>3</sup>
18 et 19/06	1,40 10 <sup>4</sup>	3,62 10 <sup>3</sup>	1,40 10 <sup>4</sup>
21 au 25/06	4,46 10 <sup>3</sup>	1,14 10 <sup>4</sup>	4,57 10 <sup>3</sup>
25/06 au 18/07	4,95 10 <sup>3</sup>	5,60 10 <sup>3</sup>	3,60 10 <sup>3</sup>
20 et 21/07	1,25 10 <sup>4</sup>	1,66 10 <sup>4</sup>	1,65 10 <sup>4</sup>

Tab. 6.3 – Comparaison des valeurs de RMSE pour trois modèles possibles

Les résultats obtenus avec le modèle combinant la régression et le profil journalier donnent de meilleurs résultats que ceux obtenus uniquement avec la régression linéaire, sauf pour les journées du 18 et 19/06. L'existence d'un profil journalier est donc à considérer avec attention. Toutefois, afin de valider définitivement cette hypothèse, il conviendrait de renouveler des prélèvements au pas de temps horaire sur plusieurs journées de temps sec.

---

<sup>6</sup>RMSE : Root Mean Squared Error

---

Par ailleurs, le modèle de régression établi à partir des données de 2007 ne peut être utilisé dans l'état. En effet, il comporte trop peu de données échantillonnées par temps de pluie avec de fortes concentrations. Rappelons que seules trois observations sur cent un ont des concentrations supérieures à  $3 \cdot 10^5$  E. coli/100 ml (et des absorbances à 730 nm supérieures à  $6 \text{ m}^{-1}$ ).

## **6.4 Propositions d'amélioration du protocole expérimental**

Le but de cette partie du mémoire est de suggérer des améliorations au protocole expérimental ainsi que des pistes de recherche complémentaires. Les principales préconisations sont les suivantes.

### **6.4.1 Découpage du protocole expérimental en deux phases**

L'installation d'un spectrophotomètre à l'exutoire du bassin versant du Stang Alar a permis d'acquérir une expérience sur la mesure en continu de pollution par spectrophotométrie UV/Visible sur de longues périodes. Toutefois, il aurait été préférable de découper le protocole expérimental en deux phases distinctes de manière à réduire au maximum les sources potentielles d'incertitude.

Lors de la première phase, une base de données robuste de spectres d'absorbance et de résultats de dénombrement devrait être constituée de manière à établir un étalonnage d'une loi absorbance/concentration en E. coli. Les mesures spectrophotométriques devraient être réalisées en laboratoire avec le spectrophotomètre de terrain de manière à s'affranchir :

- de la dérive du spectrophotomètre ;
- de la représentativité des eaux mesurées dans la cuve par rapport aux eaux de la rivière ;
- des problèmes de maintenance du spectrophotomètre in-situ (filasses ou autres corps étranger dans les optiques)

La validation sur site de la loi d'étalonnage devrait être réalisée lors de la deuxième phase en transférant le spectrophotomètre in-situ.

---

## 6.4.2 Organisation des prélèvements

Les paragraphes précédents ont mis en exergue plusieurs problèmes relatifs au prélèvement d'échantillons. Tout d'abord, lorsque le prélèvement est manuel, les échantillons sont systématiquement prélevés aux mêmes tranches horaires. De plus, la base de données comporte très peu d'échantillons avec de fortes concentrations en *E. coli*. Enfin, le protocole de prélèvement n'a pas permis d'établir la variabilité horaire et journalière des concentrations en temps sec.

Nous préconisons donc de modifier le programme de prélèvements d'échantillons en réalisant des prélèvements manuels quotidiens ainsi que des campagnes de prélèvements avec préleveur automatique aussi bien en temps sec qu'en temps de pluie.

L'étude de la dynamique de variation des concentrations en *E. coli* serait réalisée à l'aide des campagnes de prélèvements. Les prélèvements quotidiens permettraient quant à eux de recueillir des données sur l'intégralité de la saison de baignade sans alourdir inutilement le programme de prélèvements. Ceux-ci seraient aléatoirement répartis dans les tranches horaires des journées de travail.

En temps-sec, la variabilité horaire des concentrations devrait être étudiée. Le pas de temps d'échantillonnage serait fixé en fonction du nombre de flacons (quatorze en général pour les préleveurs réfrigérés). Les flacons de prélèvement en plastique seraient remplacés par des flacons en verre qui seront autoclavés entre chaque utilisation de manière à limiter les risques de contamination.

En temps de pluie, l'objectif est de prélever des échantillons pendant les pics de concentration. Le démarrage du préleveur serait asservi au dépassement d'un seuil d'absorbance. Une valeur de  $6 \text{ m}^{-1}$  à la longueur d'onde de 730 nm pourrait être retenue. Les flacons seront remplis en mode horaire avec un pas de temps d'échantillonnage réduit (10 à 15 min par exemple). Ce mode de prélèvement pourrait être couplé avec des prélèvements manuels par temps de pluie. La présence d'un opérateur sur une ou deux journées de temps de pluie permettrait également de recueillir des échantillons aux instants opportuns par suivi de la courbe d'absorbance.

## 6.4.3 Quantification des incertitudes de mesure

La méthode normalisée de dénombrement des *E. coli* est une méthode statistique permettant l'évaluation du nombre le plus probable d'*E. coli* ainsi que des intervalles de confiance à 95%.

---

Or, ces derniers ne sont pas communiqués par le laboratoire d'analyse. Leur prise en compte permettrait de mieux évaluer la qualité de la régression linéaire. De même, il serait judicieux lors de la phase 1, de réaliser les mesures spectrophotométriques en triplicat.

#### 6.4.4 Modification de la cuve de mesure ou mesure en by-pass

Deux nouveaux systèmes de mesure in-situ pourraient être envisagés. Ils sont présentés dans les prochains paragraphes.

##### 6.4.4.1 Modification de la cuve de mesure

La cuve de mesure fonctionne comme un décanteur, la vidange étant réalisée par surverse. Nous proposons donc de modifier la cuve de mesure en réalisant une vidange en fond de cuve et en ajoutant un agitateur mécanique (cf. FIG 6.23).

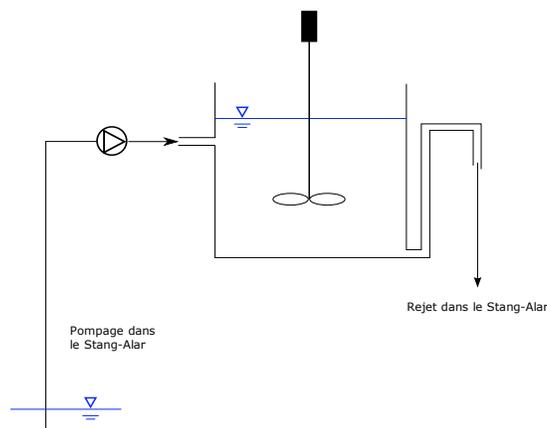


Fig. 6.23 – Schéma de modification de la cuve de mesure : vidange par siphon et ajout d'un agitateur hélicoïdal

Ces modifications devraient limiter le stockage de matières en suspension et favoriser l'homogénéisation des eaux au sein de la cuve.

---

#### 6.4.4.2 Mesure en by-pass

La cuve de mesure pourrait également être abandonnée au profit d'une mesure des eaux par by-pass au sein d'un insert (cf. FIG 6.24). Cette solution nécessite une pompe péristaltique qui effectue le prélèvement en continu des eaux du Stang Alar vers l'insert de mesure.



Fig. 6.24 – Insert permettant un mesure spectrophotométrique en by-pass

Un procédé de nettoyage mécanique composé d'une brosse peut-être adjoint au système en remplacement du nettoyage pneumatique afin de diminuer les problèmes de dérives constatés dans l'UV (cf. FIG. 6.25).

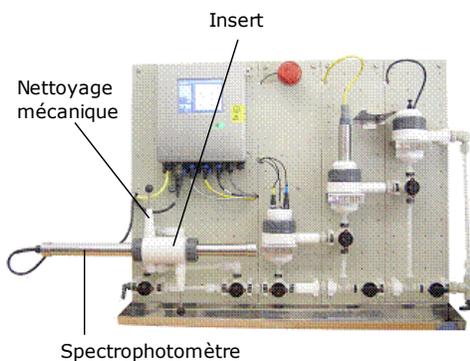


Fig. 6.25 – Insert permettant un mesure spectrophotométrique en by-pass et nettoyage mécanique

---

#### **6.4.5 Maintenance du spectrophotomètre**

Le nettoyage pneumatique du spectrophotomètre n'a pas permis d'éviter une dérive du signal, essentiellement constatée dans l'UV. L'étude de la dynamique des signaux d'absorbance montre qu'un nettoyage hebdomadaire à bi-mensuel permettrait d'éviter toute dérive. A cette occasion, la cuve de mesure pourrait également être vidangée et nettoyée.

#### **6.4.6 Contrôle qualité**

Afin de vérifier régulièrement la pertinence des mesures spectrophotométriques réalisées in-situ, nous proposons de réaliser des mesures de contrôle sur des échantillons strictement identiques. Les mesures de contrôle seraient effectuées en laboratoire avec un second spectrophotomètre.

### **6.5 Conclusion**

L'objectif de cette partie était de rechercher une relation entre concentration en E. coli et spectres d'absorbance sur des eaux de rivière échantillonnées à l'exutoire d'un bassin versant mixte urbain-rural. En effet, la qualité de la rivière Stang Alar où les mesures ont été réalisées, est influencée par les rejets du réseau d'assainissement pluvial, occasionnant ainsi des pics d'absorbance par temps de pluie. Par ailleurs, la présence de mauvais branchements sur le réseau, laisse présager des apports de contamination fécale en quantité importante dans le milieu récepteur par temps de pluie. Il semblait donc envisageable d'établir une relation entre absorbance(s) et concentration en E. coli sur la base de la dynamique pluviale de ce système.

Les mesures effectuées au cours trois saisons de baignade écoulées confirment la dynamique de temps de pluie du système. Ceci a pu être vérifié de manière globale sur l'ensemble de la saison de baignade 2007, et de manière événementielle sur deux épisodes pluvieux en 2009. Toutefois, il semble hasardeux d'établir une relation basée sur une régression linéaire simple ou multivariable. En effet, le nombre de données échantillonnées par temps de pluie est insuffisant et ne permet pas d'établir une relation statistiquement consistante. Il convient donc de compléter le jeu de données avec des échantillons à fortes concentrations en E. coli et fortes valeurs d'absorbance.

---

Les mesures effectuées sur la saison de baignade 2009 montrent également l'existence possible d'un profil journalier de concentrations en *E. coli*, celui-ci ayant été observé au mois de juin 2009 sur trois journées consécutives. Cette hypothèse doit être vérifiée en effectuant des campagnes de mesures par temps sec avec un pas de temps de prélèvement horaire. Si des rejets de pollution d'origine fécale existent, les mesures d'absorbance ne permettront pas à elles seules de détecter conjointement les pollutions de temps sec et de temps de pluie. En effet, lorsque le profil de concentrations en *E. coli* a été observé avec une étendue journalière de concentrations importante ( $4 \cdot 10^4$  *E. coli*/100 ml), les variations d'absorbance étaient très faibles ( $1 \text{ m}^{-1}$ ). Par temps de pluie l'étendue des concentrations de *E. coli* augmente environ d'un facteur deux alors que l'étendue des valeurs d'absorbance est multipliée par 40.

L'impossibilité de pouvoir tirer des conclusions définitives après trois saisons de baignade est principalement liée à des biais expérimentaux. Notons tout d'abord que les prélèvements ont majoritairement été effectués manuellement à des tranches horaires fixes. Ceci explique qu'une dynamique de temps-sec n'ait pas pu être détectée avant 2009. Cette procédure de prélèvements a par ailleurs généré beaucoup de données avec des faibles concentrations et au contraire un nombre insuffisant de données avec de fortes concentrations.

Par ailleurs, la réalisation de mesures spectrophotométriques in-situ a rendu complexe l'identification d'une relation entre absorbance et pollution fécale. En effet, Véolia a été confronté à de multiples problèmes expérimentaux tels que la dérive du signal d'absorbance qui empêche l'exploitation des longueurs d'onde de l'UV. Nous recommandons donc, pour des études similaires, d'établir la collection d'étalonnage hors-ligne, en laboratoire et en effectuant des mesures de contrôle avec un second spectrophotomètre. Cette séparation entre mesures in-situ des spectres d'absorbance et dénombrement de *E. coli* en laboratoire d'analyses a par ailleurs entraîné des erreurs d'horodatage d'échantillons.

De plus, notons que les mesures spectrophotométriques réalisées ne sont pas spécifiques. Les meilleures corrélations sont trouvées à partir de mesures d'absorbance dans le domaine du visible. Les valeurs d'absorbance dans ce domaine spectral sont liées à la présence de matières en suspension. Si la dérive du signal spectrophotométrique était maîtrisée, les longueurs d'ondes de l'UV pourraient certainement être exploitées et nous pourrions alors tenter d'établir des modèles de régression multivariables de type PLS. Une autre voie à explorer concerne la détection de traceurs chimiques de pollution fécale comme les azurants optiques (cf. §5.5.4).

---

Enfin, l'objectif du projet GIRAC étant la prédiction en continu des concentrations au niveau des sites de baignade, il serait pertinent d'établir leur sensibilité à des apports ponctuels (dans le temps et l'espace) de pollution fécale liés aux rejets d'assainissement. En effet, la synthèse bibliographique menée au paragraphe précédent a permis de montrer la variabilité spatiale et temporelle importante de la pollution d'origine fécale au niveau des sites de baignade et l'influence de facteurs extérieurs comme la marée, la hauteur d'eau ou l'ensoleillement. Des études statistiques basées sur le recueil d'un ensemble de données au niveau des sites de baignade et des sources ponctuelles auraient ainsi permis de mieux appréhender l'influence respective de chaque facteur. Des simulations hydrodynamiques simulant uniquement la dispersion de la pollution fécale apporteraient également de précieux renseignements. Si l'influence des apports ponctuels est faible, il est ainsi évident qu'une évaluation sommaire des flux de pollution d'origine fécale serait suffisante. L'évaluation des concentrations en germes indicateurs de pollution d'origine fécale n'apporterait dans ce cas précis aucune plus-value, et ce quelque soit la méthode employée.

## **Quatrième partie**

# **Modélisation pluie-débit de bassins versants urbains à partir de données expérimentales**

# Introduction

Cette partie s'inscrit également dans le cadre du projet GIRAC. Le but est de développer un modèle pluie-débit, adapté à des simulations en continu, dont le débit simulé pourra être utilisé comme condition aux limites d'un modèle hydrodynamique de convection-diffusion de la pollution en milieu marin.

Bien qu'en hydrologie urbaine l'usage de modèles distribués soit coutumier, les objectifs de cette modélisation ne requièrent que la détermination d'un modèle global à l'échelle du bassin versant. Le développement du modèle sera basé sur le couplage :

- de méthodes de modélisation habituellement utilisées en hydrologie urbaine ;
- de techniques d'identification de systèmes.

Cette approche de modélisation résulte en premier lieu de notre volonté d'identifier un modèle dont la structure rend possible une interprétation hydrologique de la transformation pluie-débit. Le choix des techniques d'identification de système utilisées est guidé par ce principe.

Dans le premier chapitre, nous présenterons les approches usuelles de modélisation en hydrologie urbaine afin de circonscrire une structure de modèle pour la modélisation pluie-débit sur le bassin versant du Stang Alar. Le second chapitre sera consacré à la présentation de l'approche de modélisation proposée qui est basée en grande partie sur des techniques d'identification de système. Enfin le troisième chapitre sera dédié à son application sur le bassin versant du Stang Alar.

## Chapitre 7

# Présentation des approches usuelles de modélisation pluie-débit en milieu urbain

### 7.1 Perception de la transformation pluie-débit

Toute représentation de la réalité dépend en premier lieu de la perception que nous en avons. La manière dont nous allons modéliser la transformation pluie-débit ne déroge pas à ce principe. En effet, comme nous le verrons par la suite, la plupart des processus physiques intervenant dans cette transformation, comme l'infiltration de l'eau de pluie dans le sol, sont difficilement quantifiables tant au niveau local qu'à l'échelle d'un bassin versant. Les incertitudes inhérentes à la quantification de chaque processus aboutissent à plusieurs représentations possibles, toutes subjectives et aussi pertinentes les unes que les autres.

Intéressons-nous donc dans un premier temps à la description de la transformation pluie-débit sur un bassin versant urbanisé, de même type que celui du Stang Alar, comprenant de la végétation, des surfaces imperméables (toitures, voirie,...), des surfaces perméables (pelouses, allées), un réseau d'assainissement séparatif et un cours d'eau comme exutoire. Une vision,

---

parmi d'autres, des processus impliqués dans la transformation pluie-débit est représentée sur la FIG. 7.1.

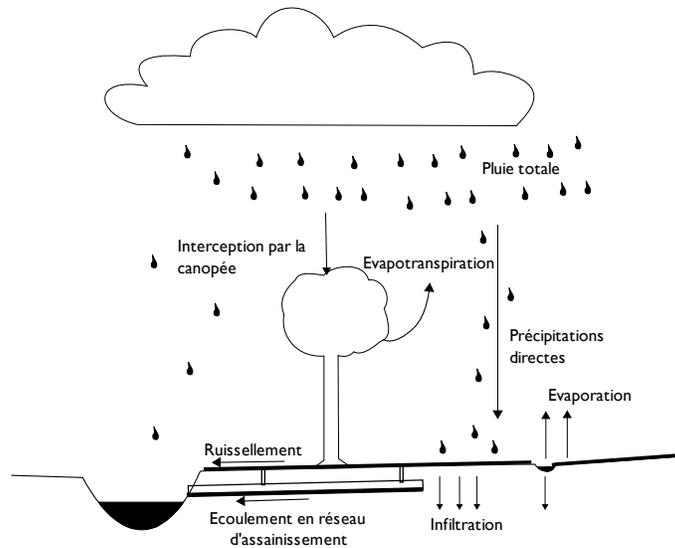


Fig. 7.1 – Représentation des processus physiques impliqués dans la transformation pluie-débit en milieu urbain

Toute la pluie qui s'abat sur un bassin versant ne contribue pas au ruissellement. Une grande partie est interceptée. Les phénomènes physiques d'interception comprennent l'interception par le feuillage de la végétation, la rétention de l'eau dans des dépressions à la surface du sol, l'évaporation et l'évapotranspiration, ainsi que l'infiltration de l'eau dans le sol.

La pluie qui contribue au ruissellement est appelée pluie effective (Beven, 2003). Son volume est égal au volume de l'hydrogramme résultant. Le ruissellement peut avoir lieu en surface ou dans les premiers horizons du sol. On considère que le ruissellement de surface a lieu quand l'épaisseur de la lame d'eau à la surface du sol est suffisante pour que les forces de gravité soient supérieures aux tensions de surface (Bertrand-Krajewski, 1992). Les écoulements de subsurface sont influencés par la présence ou l'absence d'une nappe affleurante. Les ruissellements de surface et de subsurface participent à l'augmentation du débit du cours d'eau à l'exutoire du bassin versant.

En milieu urbain, cette perception de la transformation pluie-débit est quelque peu différente. Tout d'abord, une grande partie des surfaces urbanisées sont imperméabilisées. L'infiltration

---

sur ces surfaces n'est cependant pas nulle et varie de 0,2 à 2 mm/h CERTU (2003). La contribution au ruissellement de ces surfaces est prépondérante. Desbordes (1974) propose de négliger la contribution des surfaces perméables lorsque le coefficient d'imperméabilisation est supérieur à 20 %.

Par ailleurs, la présence de végétation étant moindre, les phénomènes de pertes liés à l'évapotranspiration et à la rétention d'eau par le feuillage ou à la surface du sol sont moins importants qu'en milieu rural. Selon Desbordes (1974), la rétention d'eau à la surface du sol sur les surfaces imperméables varie entre 0,2 et 3 mm et atteint jusqu'à 15 mm sur des surfaces perméables végétalisées. Les rétentions d'eau par le feuillage atteindraient quant à elles jusqu'à 1,5 mm.

Enfin, le ruissellement de surface est collecté au sein du bassin versant par le réseau d'assainissement (canalisations pluviales). Il existe donc un écoulement en réseau d'assainissement en plus de l'écoulement de surface et de subsurface. Parallèlement à la distinction surfaces perméables/surfaces imperméables, il est donc pertinent de différencier les surfaces connectées au réseau d'assainissement des surfaces déconnectées. Le réseau d'assainissement a vocation à collecter le ruissellement de surface pour éviter les inondations (une grande partie des surfaces étant imperméabilisées). Il a pour conséquence d'accélérer les temps de transfert du ruissellement vers l'exutoire du bassin versant.

## **7.2 Questions préalables à la représentation de la transformation pluie-débit**

Avant de construire ou d'utiliser un modèle de transformation pluie-débit, quelques questions méritent d'être considérées afin de définir une structure générale de modèle adéquate.

### **7.2.1 Modèle global ou distribué**

Les phénomènes physiques décrits au paragraphe précédent peuvent être appréhendés à une échelle locale (quelques m<sup>2</sup>), à l'échelle d'un quartier ou à l'échelle d'un bassin versant (quelques hectares à quelques kilomètres carrés).

---

L'échelle à laquelle nous allons décider de décrire la transformation pluie-débit dépend des objectifs que nous assignons à la modélisation et des informations que nous voulons recueillir.

Il existe deux grandes familles de modèles : les modèles globaux et les modèles distribués. Les modèles globaux "*considèrent le bassin versant comme unité de base dans laquelle les variables d'état sont des moyennes établies sur l'ensemble du bassin versant*" (Beven, 2003). Dans ce type de modèle, aucune information ne sera disponible sur la répartition du débit dans les différentes branches du réseau d'assainissement au sein du bassin versant.

Le premier modèle global développé en hydrologie est celui de Mulvaney (1851), également connu sous le nom de *méthode rationnelle*. Il établit une relation linéaire entre le débit de pointe  $Q_p$  à l'exutoire d'un bassin versant et l'intensité moyenne de la pluie sur une durée correspondant au temps de concentration  $\bar{T}_{tc}$  :

$$Q_p = C A \bar{T}_{tc} \quad (7.1)$$

où  $C$  est le coefficient de ruissellement et  $A$  l'aire du bassin versant.

Les modèles distribués découpent le bassin versant en unités élémentaires possédant des caractéristiques similaires. Ils permettent d'effectuer des prédictions qui seront distribuées dans l'espace au sein du bassin versant.

Les logiciels d'hydrologie urbaine tels que CANOE<sup>®</sup>, INFOWORKS CS<sup>®</sup> ou MIKE URBAN<sup>®</sup> permettent de réaliser des modélisations distribuées (cf. FIG. 7.2(b)). Le bassin versant est découpé en sous bassins versants homogènes dans chacun desquels un modèle pluie-débit est appliqué. Ces modèles simulent la production et le transfert du ruissellement (de surface et en réseau d'assainissement) jusqu'à l'exutoire de chaque sous bassin versant. Les débits modélisés sont alors injectés dans un modèle hydraulique qui simule la propagation du flux hydraulique au sein du réseau d'assainissement et/ou des cours d'eau. Ces logiciels associent donc deux types de modèles de manière à fournir une information distribuée sur le bassin versant concernant les débits générés et les hauteurs d'eau au sein des canalisations.

Notons qu'il est également possible avec ces logiciels de concevoir un modèle global pour l'ensemble du bassin versant étudié (cf. FIG. 7.2(a)). Le modèle ne comprend alors pas de simulation hydraulique en canalisation. La seule information produite par la modélisation est l'hydrogramme à l'exutoire du bassin versant.

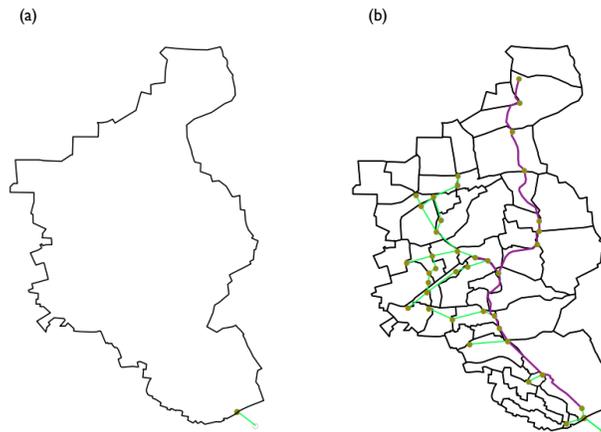


Fig. 7.2 – Différences de représentation entre modèles globaux et modèles distribués au sein des logiciels d'hydrologie urbaine. (a) : modèle global (b) : modèle distribué, en rouge les exutoires de chaque sous bassin versant où sont injectés les débits pour la modélisation hydraulique

### 7.2.2 Type de modèle

Le second choix porte sur le type de modèle à utiliser. Les modèles hydrologiques peuvent être regroupés en trois grandes catégories :

- Les modèles mécanistes qui sont basés sur les lois de la physique (mécanique, hydraulique...). Ces modèles nécessitent une description très fine du système à modéliser pour intégrer l'ensemble des paramètres nécessaires à la résolution des équations. Par exemple, le ruissellement de surface peut-être modélisé à partir d'équations basées sur la quantité de mouvement et nécessitant de connaître la pente, l'état de surface,... Pour plus de précisions sur ces modèles, le lecteur se référera à la thèse de doctorat de Bertrand-Krajewski (1992).
- Les modèles conceptuels, qui sont certainement les plus utilisés, et qui ont vocation à simuler le phénomène physique par une représentation simplifiée et *conceptuelle* des processus physiques en jeu. L'exemple le plus courant est le modèle de réservoir qui assimile le transfert de l'écoulement au sein d'un bassin versant à l'écoulement au sein d'un réservoir, sous la contrainte que le bilan de masse soit respecté.

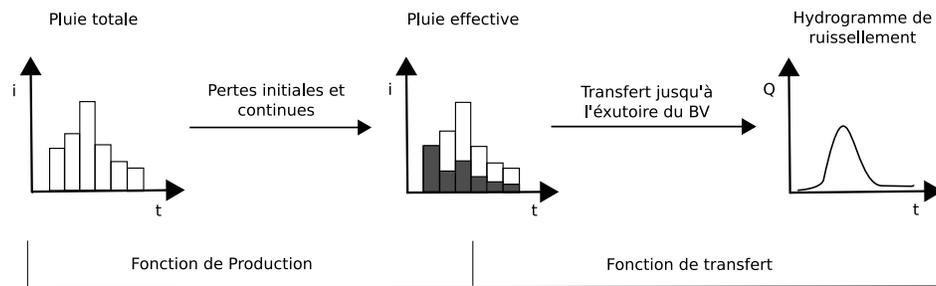


Fig. 7.3 – Flux de modélisation intégrant les fonctions de production et de transfert

- les modèles empiriques<sup>1</sup> qui établissent un lien entre pluie et débit au moyen de régressions linéaires ou non linéaires impliquant parfois d'autres variables (ex : pente, aire du bassin versant,...). Les modèles sont établis suite à une étude statistique de séries de mesures réalisées sur le bassin versant. Notons que les coefficients de régression ne permettent pas dans la majorité des cas de donner un sens physique à la relation obtenue. Des exemples de modèles empiriques sont donnés dans l'ouvrage de Beven (2003).

Cette classification n'est pas la seule à pouvoir être utilisée. Nous présenterons une autre approche résultant de l'identification de systèmes dans le chapitre suivant (cf. par. 8.1.2).

### 7.3 Représentation de la transformation pluie-débit

Comme cela a été évoqué dans le paragraphe précédent, les possibilités de représentation de la transformation pluie-débit au moyen de modèles numériques sont nombreuses. Toutefois, la modélisation en continu d'hydrogrammes est souvent effectuée en décomposant la transformation pluie-débit en deux composantes (Beven, 2003; Desbordes *et al.*, 1994) :

- la fonction de production qui détermine la quantité de pluie qui sera effectivement transformée en débit, dénommée pluie effective ;
- la fonction de transfert qui détermine quant à elle la distribution et le temps de transfert du ruissellement jusqu'à l'exutoire du bassin versant.

<sup>1</sup>également appelés modèles statistiques

---

Une représentation conceptuelle de ces deux composantes est schématisée en FIG. 7.3. Cette décomposition de la transformation pluie-débit est d'abord historique. Elle traduit la volonté des hydrologues, dès les premières études réalisées au début du 20<sup>ème</sup> siècle, de séparer les effets résultant de la *production* du ruissellement de ceux résultant de son *transfert*. Cela a permis de construire des modèles complexes mais ne nécessitant pas d'ordinateurs pour effectuer les calculs (Beven, 2003).

En effet, la transformation pluie-débit est sur une grande majorité de bassins versants non linéaire et la décomposition de cette transformation s'appuyant sur une fonction non linéaire (la fonction de production) et une fonction linéaire (la fonction de transfert) permet de simplifier le problème.

Ainsi, dès les années 1920 les études hydrologiques réalisées font l'hypothèse de linéarité du transfert du ruissellement vers l'exutoire du bassin versant (Ross, 1921; Zoch, 1934). Ces études découpent le bassin versant en zones caractérisées par un temps de transfert homogène. Il suffit donc de calculer la fonction de production sur chacune des zones puis de sommer les hydrogrammes en tenant compte des différents retards dans le transfert. L'hypothèse de linéarité est bien entendue une approximation car il est connu que les vitesses d'écoulement changent non linéairement avec le débit. Cependant, comme nous venons de le voir, cette simplification permet de réaliser un calcul simple pour le transfert de l'ensemble de l'écoulement sur le bassin versant et reste très souvent acceptable.

Le coeur du problème de toute identification d'un modèle de transformation pluie-débit est l'identification de la fonction de production qui intègre la majeure partie des non linéarités du système.

### 7.3.1 Fonction de production

La fonction de production établit la transformation entre l'intensité précipitée totale ( $I_T$ , [ $LT^{-1}$ ]) et la pluie effective ( $I_{Eff}$ , [ $LT^{-1}$ ]). La différence entre pluie totale et pluie effective est due à des pertes ( $P$ ), c'est à dire les processus physiques qui interceptent une partie de la pluie qui ne contribuera donc pas au ruissellement :

$$I_{Eff} = I_T - P \quad (7.2)$$

---

Les pertes résultent des différents processus physiques évoqués au paragraphe 7.1 tels que l'évaporation, l'évapotranspiration, la rétention d'eau dans des dépressions à la surface du sol ou par le couvert végétal ainsi que l'infiltration de l'eau de pluie dans le sol.

En hydrologie urbaine, les pertes sont généralement classées en deux catégories : les pertes initiales et les pertes continues. Les pertes initiales font intervenir les processus qui surviennent au début de chaque événement pluvieux, comme l'interception d'eau par la canopée (Chocat, 1981; Falk et Niemczynowicz, 1978). Les pertes continues traduisent quant à elles les processus ayant lieu tout au long de l'événement pluvieux. On considère généralement que les pertes continues sont assimilables à l'infiltration de l'eau dans le sol.

La transformation pluie totale-pluie effective est un problème non linéaire très difficile à évaluer. En effet, les pertes initiales et continues sont très dépendantes des conditions antécédentes à un événement pluvieux. De plus, chacun des processus physiques est très difficilement quantifiable. Enfin, la pluie effective est une grandeur fictive qui n'est pas mesurable. Il s'agit donc d'un problème récurrent dans toute identification d'un nouveau modèle pluie-débit.

Nous ne traiterons par la suite que les modèles conceptuels de quantification de la fonction production qui sont les plus utilisés en hydrologie urbaine.

### **7.3.1.1 Pertes initiales**

Intéressons-nous dans un premier temps à la modélisation des pertes initiales. Dans les logiciels d'hydrologie urbaine (CANOE<sup>©</sup>, INFOWORKS CS<sup>©</sup>, MIKE URBAN<sup>©</sup>), des pertes initiales forfaitaires sont appliquées au début de chaque épisode pluvieux.

L'étude bibliographique de Bertrand-Krajewski (1992) montre que les pertes initiales prises en compte dans la plupart des modèles sont comprises entre 0,25 et 3 mm.

L'utilisation d'un modèle de pertes initiales nécessite de repérer le commencement de chaque événement pluvieux. Ceci ne pose pas de problème lorsque les simulations sont réalisées sur des événements pluvieux indépendants mais s'avère très compliqué dans le cas de chroniques de pluies continues. Pour contourner cette difficulté, les pertes initiales peuvent être modélisées avec un modèle de réservoir (cf. FIG. 7.4). Le réservoir a une hauteur correspondant à la lame d'eau maximum pouvant être interceptée et un orifice permettant de prendre en compte les effets conjugués de l'évaporation et l'évapotranspiration. La surverse de ce réservoir constitue la part de pluie pouvant générer un ruissellement (avant infiltration).

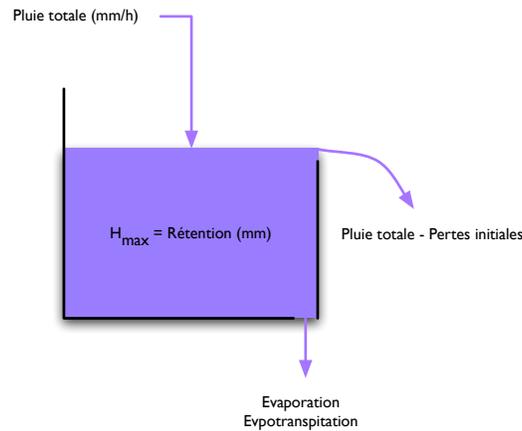


Fig. 7.4 – Modélisation des pertes initiales par un modèle de réservoir

### 7.3.1.2 Pertes continues

Les pertes continues sont également représentés dans la plupart des cas au moyen de modèles conceptuels. Trois grands types de modèles se dégagent (Beven, 2003; CERTU, 2003; Jovanovic, 1986).

Le modèle le plus simple considère que la capacité d'infiltration d'un sol est constante au cours du temps (modèle du  $\Phi$ -index, cf. FIG. 7.5(a)). Bien que simplissime et non réaliste, ce modèle donne parfois de bons résultats. Il est recommandé d'appliquer ce modèle uniquement pour les surfaces à faible infiltration (CERTU, 2003).

D'autres modèles décrivent les pertes continues comme étant proportionnelles à l'intensité de l'événement pluvieux (cf. FIG. 7.5(b)). Le ruissellement est donc lui aussi proportionnel à l'intensité de la pluie et nous pouvons écrire :

$$I_{Eff}(t) = C_R \cdot I_{Eff*}(t) \quad (7.3)$$

où  $C_R$  est le coefficient de ruissellement,  $I_{Eff*}$  l'intensité de la pluie totale diminuée des pertes initiales (cf. FIG. 7.4), et  $I_{Eff}$  l'intensité de la pluie effective. Dans l'équation 7.3 le coefficient de ruissellement est une constante exprimée par :

$$C_R = \frac{V_R}{A \cdot (L_T - PI)} \quad (7.4)$$

---

avec  $V_R$ , le volume d'eau ruisselée à l'exutoire du bassin versant ( $[LT^{-3}]$ ),  $A$ , l'aire du bassin versant ( $[L^2]$ ),  $L_T$  la lame d'eau précipitée totale ( $[L]$ ), et  $PI$ , les pertes initiales ( $[L]$ ).

Le coefficient de ruissellement  $C_R$  et les pertes initiales  $PI$  peuvent être déterminées par régression linéaire si l'on dispose de mesures conjuguées de pluie et de débit à l'exutoire du bassin versant. Lors de cette régression, il peut être judicieux de définir plusieurs domaines de lames d'eau précipitées afin de tenir compte de l'augmentation du coefficient de ruissellement expérimentalement constaté pour des événements pluvieux intenses (CERTU, 2003). En effet, conserver une valeur de coefficient de ruissellement identique quelle que soit la lame d'eau précipitée revient à dire que la capacité d'infiltration du sol est illimitée et uniquement fonction de l'intensité des précipitations, ce qui est évidemment faux.

Le modèle décrit par l'équation (7.3) est très simple à mettre en oeuvre et présente l'avantage d'être linéaire avec l'intensité de la pluie (diminuée des pertes initiales) ce qui simplifie le problème d'identification du modèle pluie-débit. Il constitue le modèle standard de tous les logiciels d'hydrologie urbaine et peut être utilisé pour des bassins versants urbains équipés de réseaux d'assainissement (CERTU, 2003). En contexte urbain, l'application de modèles plus complexes où le coefficient de ruissellement varie au cours d'un événement pluvieux ne s'avère pas justifié. En effet, les incertitudes sur les mesures de pluies et de débits sont généralement trop élevées pour évaluer l'amélioration réelle des résultats liés au changement de modèle (CERTU, 2003; Bertrand-Krajewski *et al.*, 2000).

Enfin, le troisième type de modèle décrit l'infiltration comme une fonction décroissante au cours de l'événement pluvieux (cf. FIG. 7.5(c)). Les modèles établis sur cette base sont en général dérivés du modèle établi par Horton (1933) qui a utilisé une fonction empirique, basée sur des expérimentations, pour décrire la diminution de la capacité d'infiltration d'un sol au cours du temps :

$$F_p = f_c + (f_0 - f_c) e^{-\beta t} \quad (7.5)$$

avec  $F_p$  la capacité d'infiltration du sol,  $f_0$  la capacité d'infiltration maximum du sol,  $f_c$  et  $\beta$  deux paramètres du modèle. Il est ainsi considéré que durant un événement pluvieux, si l'intensité de la pluie dépasse la capacité d'infiltration du sol, alors de l'eau commence à s'accumuler à la surface du sol et un ruissellement peut avoir lieu.

Ce modèle décrit d'une manière plus réaliste l'évolution de l'infiltration au cours du temps. Son utilisation doit être réservée aux bassins versants ruraux fortement végétalisés ou avec

---

des sols saturés sans leur horizon superficiel. Il n'est pas recommandé de les utiliser pour des sols nus et secs où les problèmes d'interface air/eau dans la zone superficielle non saturée sont importants CERTU (2003).

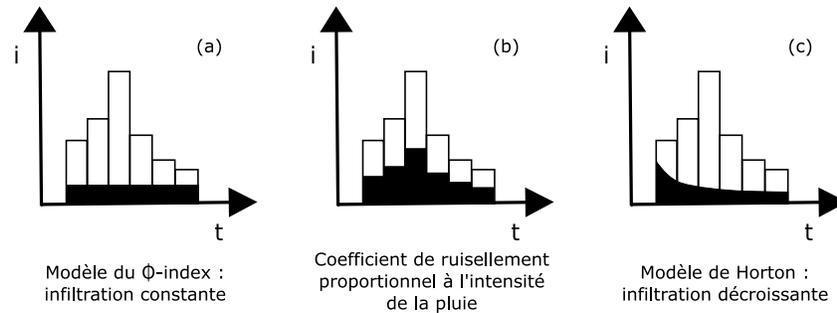


Fig. 7.5 – 3 représentations de l'infiltration

### 7.3.2 Fonction de transfert

La fonction de transfert établit la distribution et le temps de transfert du ruissellement jusqu'à l'exutoire du bassin versant. C'est une fonction conservatrice : le volume de l'hydrogramme à l'exutoire du bassin versant est égal au volume de la pluie effective. De plus, la majorité des fonctions de transfert font l'hypothèse de linéarité du transfert du ruissellement.

Les modèles mécanistes sont inapplicables à l'échelle de grands bassins versants. En effet, ils requièrent un nombre trop important de données descriptives et leur temps de calcul est prohibitif (Bertrand-Krajewski, 1992).

Les modèles conceptuels leur sont préférés (O'Loughlin *et al.*, 1996). Une des premières fonctions de transfert conceptuelle utilisée en hydrologie est l'hydrogramme unitaire. L'hydrogramme unitaire définit la relation entre une averse unitaire et l'hydrogramme résultant. Il est défini par Wilser and Brater (1949) comme "*l'hydrogramme d'un ruissellement de surface résultant d'une précipitation, intense, courte, appelée averse unitaire*". L'hydrogramme unitaire est établi suite à des séries de mesures effectuées sur un bassin versant étudié en sélectionnant les hydrogrammes résultant de précipitations se rapprochant des caractéristiques de l'averse unitaire. La problématique dans l'utilisation de cette fonction réside dans la détermination de la forme de l'hydrogramme (Beven, 2003).

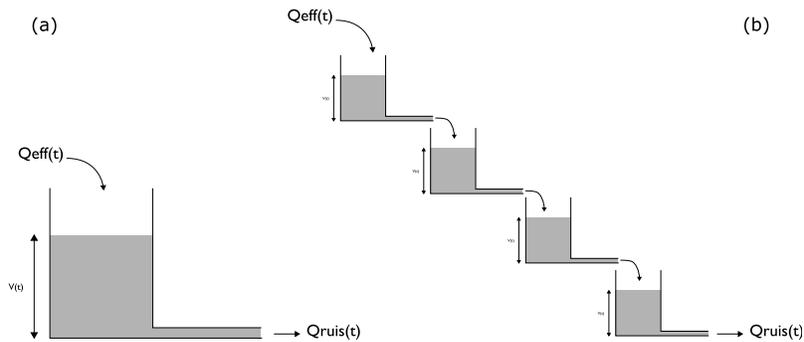


Fig. 7.6 – Modèles de réservoirs pour le transfert du ruissellement au sein d'un bassin versant. (a) : réservoir linéaire simple. (b) : cascade de réservoirs

En contexte urbain, les modèles de réservoirs sont certainement les fonctions de transfert du ruissellement les plus utilisées (Bertrand-Krajewski, 1992; O'Loughlin *et al.*, 1996; CERTU, 2003). Elles assimilent le ruissellement dans un bassin versant à l'écoulement au sein d'un réservoir ou plusieurs réservoirs associés en série ou en parallèle (cf. FIG. 7.6).

Le réservoir linéaire est la façon la plus simple, mais aussi la plus utilisée de représenter le transfert du ruissellement au sein d'un bassin versant urbain. Ce modèle est constitué de deux équations :

$$\frac{dV(t)}{dt} = Q_{eff}(t) - Q_{ruis}(t) \quad (7.6)$$

$$V(t) = KQ_{ruis}(t) \quad (7.7)$$

avec :

- $V(t)$  Volume du réservoir [ $L^3$ ];
- $Q_{eff}(t)$  Débit généré par la pluie effective [ $LT^{-3}$ ] :  $Q_{eff}(t) = C_R(t) A I_{eff}(t)$ ;
- $Q_{ruis}(t)$  Débit ruisselé à l'exutoire du bassin versant [ $LT^{-3}$ ];
- $K$  Lag-time [T].

La première équation (eq. 7.6) traduit un bilan de matière au sein du réservoir. La seconde (eq. 7.7) formule l'hypothèse que le volume d'eau stockée dans le réservoir  $V(t)$  est proportionnel au débit de sortie du réservoir  $Q_{ruis}(t)$ . Le coefficient de proportionnalité  $K$  est appelé lag-time et correspond à un temps de résidence moyen de l'écoulement au sein du réservoir. Il peut également être interprété comme un décalage temporel entre le centre de gravité de

---

l'hydrogramme de pluie effective ( $Q_{eff}(t)$ ) et le centre de gravité de l'hydrogramme ruisselé à l'exutoire du bassin versant ( $Q_{ruis}(t)$ ).

Ce modèle peut être réécrit sous la forme suivante :

$$\frac{dQ_{ruis}(t)}{dt} + \frac{1}{K} Q_{ruis}(t) = \frac{1}{K} Q_{eff}(t) \quad (7.8)$$

Il s'agit d'une équation différentielle du premier ordre dont la solution analytique s'écrit :

$$Q_{ruis}(t) = \frac{1}{K} \int_{t_0}^t Q_{eff}(t-u) e^{-\frac{(t-u)}{K}} du + Q_{ruis}(t_0) e^{-\frac{(t-t_0)}{K}} \quad (7.9)$$

Le modèle du réservoir linéaire est un modèle dérivé du modèle d'hydrogramme unitaire dans le sens où il est facile de déterminer analytiquement la réponse du modèle à une averse unitaire (fonction de dirac). L'hydrogramme unitaire<sup>2</sup> est déterminé en calculant la transformée de Laplace inverse de la fonction de transfert et s'écrit :

$$h(t) = \frac{1}{K} e^{-t/K} \quad (7.10)$$

Dans ce modèle la transmission du signal est instantanée. Le maximum de l'hydrogramme unitaire est obtenu pour  $t = 0$  et vaut  $1/K$ . Dans le cas de grands bassins versants, ou dans le cas de bassins versants urbains où les surfaces perméables sont non négligeables, la réponse du modèle peut être plus rapide que ce qui est effectivement observé dans la réalité. Pour prendre en compte ces décalages temporels, il est possible d'introduire un terme de retard  $\tau$  dans la formulation de l'hydrogramme unitaire (CERTU, 2003).

Le modèle du réservoir linéaire peut être généralisé en considérant un écoulement à travers une cascade de réservoirs (cf. FIG. 7.6b ; Nash, 1959). Ce type de fonction de transfert semble particulièrement adapté à des écoulements en bassins versants ruraux (Desbordes *et al.*, 1994). L'hydrogramme unitaire d'une cascade de réservoirs s'écrit :

$$h(t) = \left(\frac{t}{K}\right)^{N-1} \frac{e^{-t/K}}{K \Gamma(N)} \quad (7.11)$$

---

<sup>2</sup>Egalement appelé réponse impulsionnelle.

---

où  $\Gamma(N)$  est la fonction gamma définie par :  $\Gamma(N) = (N - 1)!$ , et où  $N$  est le nombre de réservoirs.

Comme pour la fonction de production, le transfert du ruissellement est donc régulièrement simulé par des modèles de réservoirs qui sont parfaitement adaptés à des simulations en continu.

## 7.4 Conclusion

Cette rapide étude bibliographique nous permet donc d'avoir des éléments de réflexion liés au choix d'une structure générale de modèle pluie-débit pour le bassin versant du Stang Alar.

La modélisation peut être abordée de manière globale ou distribuée. Le choix entre ces deux approches dépend des objectifs de la modélisation. Dans notre cas la modélisation a pour unique but de simuler en continu les débits de rejets d'assainissement déversés par temps de pluie dans le milieu marin. Nous appliquerons donc une approche de modélisation globale. Par ailleurs, un consensus se dégage pour l'utilisation de modèles conceptuels, particulièrement adaptés à la simulation sur des chroniques de pluies continues. Ces modèles présentent également l'avantage de fournir une interprétation hydrologique à la transformation pluie-débit.

Les pluies simulées étant d'intensité très variables, il est nécessaire de prendre en compte les pertes initiales avec un modèle de réservoir ou un modèle équivalent. Les pertes continues seront prises en compte par application d'un coefficient de ruissellement constant. Enfin, un modèle linéaire (modèle de réservoir par exemple) sera utilisé pour le transfert du ruissellement jusqu'à l'exutoire du bassin versant.

Sur la base de cette structure générale de modèle, il reste donc à effectuer un choix pertinent :

- de modèle de pertes initiales ;
- de modèle linéaire de transfert du ruissellement ;
- de méthode d'estimation paramétrique (calage) ;
- de critère(s) de sélection des modèles.

Ces choix seront abordés au chapitre suivant.

## Chapitre 8

# Proposition d'un modèle pluie-débit

### Introduction

Dans le chapitre 1, les approches de modélisation pluie-débit habituellement utilisées en hydrologie urbaine ont été rappelées et une structure générale de modèle a été définie pour la représentation de la transformation pluie-débit sur le bassin versant du Stang Alar.

Ce chapitre a pour objectif de proposer :

1. une approche de modélisation simple mais performante, basée sur la structure générale de modèle définie précédemment. Elle repose sur le couplage d'un modèle conceptuel non-linéaire de pertes initiales et d'un modèle linéaire à temps continu<sup>1</sup> incluant le transfert du ruissellement et les pertes continues.
2. une méthode de calage de l'ensemble du modèle pluie-débit basée en grande partie sur l'application de techniques d'identification de systèmes.

Les points clef de cette approche sont :

- la formulation d'un modèle original de pertes initiales ;
- l'identification optimisée d'un modèle linéaire à temps continu intégrant les pertes continues et le transfert du ruissellement ;
- le transfert du modèle linéaire à temps continu sous forme de modèle conceptuel de réservoir ;

---

<sup>1</sup>un modèle linéaire à temps continu se présente sous la forme d'une équation différentielle

- 
- l'utilisation de l'approche de modélisation *Data-Based Mechanistic modelling* (DBM) consistant à ne valider un modèle que s'il peut être interprété d'un point de vue physique (Young et Beven, 1994; Young, 1998);
  - et enfin l'estimation des paramètres de l'ensemble du modèle au moyen d'un algorithme itératif.

Le chapitre est découpé en deux sections. La première est consacrée à la présentation des apports des techniques d'identification de systèmes et à leur insertion au sein de l'approche de modélisation proposée. La présentation des modèles et des techniques d'identification utilisées seront abordés dans un deuxième temps.

## **8.1 Apports des techniques d'identification de systèmes et insertion au sein du modèle pluie-débit**

La principale originalité de l'approche de modélisation réside dans l'utilisation de techniques d'identification de systèmes pour l'établissement d'un modèle linéaire à temps continu intégrant les pertes continues et le transfert du ruissellement. Ce paragraphe aura donc pour vocation de présenter la méthodologie d'identification ainsi que l'intérêt des techniques utilisées.

### **8.1.1 Méthodologie d'identification de systèmes**

L'identification de systèmes est une discipline apparue dans les années 1960 et dont une revue historique peut être trouvée dans l'article de Gevers (2006). Elle vise à établir le modèle mathématique d'un système dynamique à partir d'observations expérimentales et d'une connaissance du système acquise *a priori*. Les modèles sont caractérisés par une structure et par des paramètres qu'il convient de choisir et d'ajuster afin de reproduire au mieux le comportement entrée-sortie du système.

Les méthodes d'identification de systèmes sont traditionnellement utilisées par les automaticiens pour déterminer des modèles permettant la synthèse de lois de commande. Cependant, ces techniques sont dorénavant employées dans d'autres domaines d'application comme l'hydrologie.

---

Seuls seront considérés les modèles paramétriques de type entrée-sortie qui sont les plus utilisés en hydrologie (Beven, 2003; Lovera, 2010; Young, 1998; Lees, 2000; Young et Beven, 1994; Lovera, 2010; Laurain *et al.*, 2010; Previdi *et al.*, 1999; Previdi et Lovera, 2009). Ces modèles consistent en des relations mathématiques faisant intervenir :

- la ou les entrées du système<sup>2</sup> ;
- la sortie du système ;
- et des paramètres *a priori* inconnus qu'il conviendra d'estimer au moyen de données expérimentales.

Ils prennent également en compte des perturbations, correspondant aux entrées non mesurées, aux erreurs de mesure et aux erreurs du modèle, qui sont généralement représentées sous la forme d'un bruit additif sur la sortie (cf. FIG 8.1).

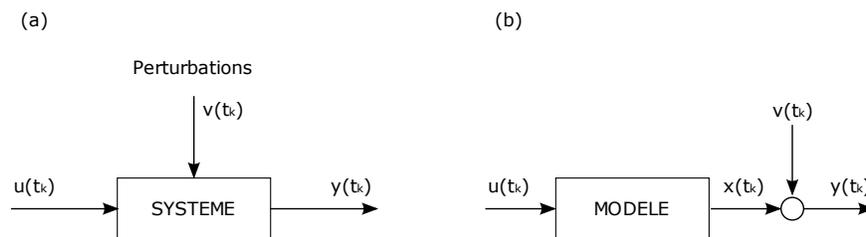


Fig. 8.1 – Prise en compte des perturbations dans un système. (a) : Système *vrai* avec perturbations ; (b) : Modèle avec représentation des perturbations sous la forme d'un bruit additif sur la sortie

On notera que la plupart des études portent sur l'identification de modèles paramétriques pluie-débit intégrant l'ensemble des non linéarités du système. L'approche proposée est différente. Les méthodes d'identification de système sont ici dédiées à l'identification d'une composante de la transformation pluie-débit et non de son intégralité. L'objectif est en effet d'identifier un modèle linéaire à temps continu intégrant les pertes continues et le transfert du ruissellement qui pourra être transféré sous forme de modèle de réservoir.

Il s'agit d'une approche voisine de Previdi *et al.* (1999) qui utilise les techniques d'identification dans le but de caler les paramètres de modèles conceptuels utilisés pour le transfert du

---

<sup>2</sup>un modèle à une entrée et une sortie est appelé modèle SISO (Single Input, Single Output. Les modèles à plusieurs entrées et une sortie sont quant à eux appelés modèles MISO (Multiple Input, Single Output)

---

ruissellement (modèles de réservoirs, méthode des courbes isochrones). Cependant, dans cette étude les fonctions de transfert avaient été calées sur des systèmes simples où l'ensemble des non linéarités pouvaient être négligées. Notre approche s'inscrit dans une démarche d'identification d'un modèle pluie-débit plus complexe, puisque l'entrée du modèle linéaire à temps continu est inconnue et constitue également la sortie du modèle non linéaire de pertes initiales (cf. FIG. 8.2 et §8.1.4).

En identification de systèmes, la méthodologie d'identification d'un modèle comporte cinq étapes principales :

1. Conception du protocole expérimental.
2. Acquisition des entrées/sorties sous un protocole d'expérimentation.
3. Identification (choix) de la structure du modèle.
4. Calage : estimation des paramètres du modèle.
5. Validation du modèle identifié (structure et valeur des paramètres).

Les deux premières étapes permettent d'acquérir les données expérimentales. Elles sont de la plus grande importance. En effet, elles conditionnent en grande partie le succès de la procédure générale d'identification. Les entrées du modèle pluie-débit doivent être choisies de manière à *exciter* suffisamment le système. Ainsi, en hydrologie urbaine, il convient de porter la plus grande attention à l'emplacement et au nombre de pluviomètres enregistreurs. En effet, la pluie est un phénomène naturel d'une grande hétérogénéité spatiale. Ainsi, pour assurer dans 75 % des cas une incertitude sur la lame d'eau précipitée inférieure à 20 %, Paturel *et al.* (1986) proposent pour des bassins versants de surface comprises entre cinq cent et mille hectares la mise en place de six à sept pluviographes.

De plus, le protocole expérimental doit être conçu de manière à limiter autant que possible les erreurs de mesures sur les entrées et sorties de systèmes.

L'étape n° 3 concerne le choix de la structure de modèle. En identification de systèmes, la structure de modèle est "l'ensemble de modèles candidats qui diffèrent entre eux par les valeurs et la dimension d'un vecteur de paramètres" (Garnier *et al.*, 2007). Elle correspond donc à une formulation mathématique générale dépendant :

- de la linéarité ou la non linéarité du système à modéliser ;

- de la variabilité du système dans le temps ;
- de la représentation du bruit additif ;
- de la représentation à temps discret ou à temps continu du système <sup>3</sup>.

La quatrième étape a pour but d'estimer les valeurs des paramètres en utilisant des algorithmes basés sur l'analyse des données expérimentales. Enfin, la dernière étape permet de vérifier que le modèle identifié est représentatif des données d'entrée et de sortie.

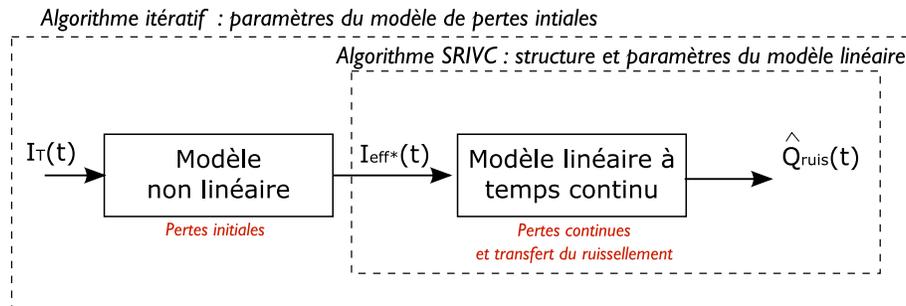


Fig. 8.2 – Insertion de la procédure d'identification du modèle linéaire à temps continu au sein de la procédure de calage du modèle pluie-débit

### 8.1.2 Approche de modélisation DBM

Les modèles identifiés n'apportent pas toujours d'information sur le fonctionnement interne du système dynamique et ne sont pas forcément interprétables d'un point de vue physique. Il s'agit d'un inconvénient majeur dans le cas de systèmes dont les caractéristiques internes sont amenées à évoluer. En effet, le moindre changement d'un paramètre du système nécessitera une nouvelle identification à partir de données expérimentales supplémentaires. De tels modèles sont appelés modèles *boite noire*. C'est le cas des modèles à temps discret (et plus particulièrement des modèles non-linéaires à temps discret) dont les équations aux différences ne permettent pas d'interpréter la dynamique du système.

Les modèles *boite grise* sont l'opposé des modèles *boite noire*. Leur formulation mathématique est issue d'une connaissance *a priori* de la dynamique du système. Ces modèles sont à

<sup>3</sup>La première famille consiste à caractériser la dynamique d'un système directement à partir des données échantillonnées. Les relations trouvées sont des équations aux différences. La seconde famille décrit les systèmes dynamiques en temps continu avec des équations différentielles.

---

rapprocher des modèles mécanistes décrits dans le chapitre 1. Ici, puisque la structure mathématique du modèle est connue *a priori*, le problème d'identification consiste uniquement en l'estimation des paramètres du modèle.

L'identification du modèle linéaire à temps continu s'insère dans une approche de modélisation de type DBM qui contraste aussi bien avec les modèles de type *boite noire* que ceux de type *boite grise*. La structure optimale du modèle est déterminée *a posteriori* comme pour les modèles de type *boite noire*. Toutefois, celle-ci ne sera validée que si une interprétation physique peut-être donnée au modèle identifié. L'interprétation physique passera, dans notre cas, par le transfert des modèles identifiés sous forme de modèles de réservoirs ainsi que par une validation de la valeur des paramètres estimés.

Cette approche présente deux avantages. Tout d'abord, elle apporte les mêmes degrés de liberté que pour l'identification de modèles *boite noire* dans le sens où l'identification porte sur une famille de formulation mathématique donnée et non sur une formulation mathématique précise. De plus, le modèle identifié a un sens physique. La modification d'un paramètre physique du système peut donc être reliée à la modification d'un paramètre du modèle. Ce second point constitue un sérieux avantage dans le cas de la gestion d'un système d'assainissement.

### **8.1.3 Identification directe de modèles à temps continu**

La seconde originalité de notre approche réside dans l'identification de modèles à temps continu. En effet, la plupart des articles relatifs à l'application de ces techniques en hydrologie portent sur l'identification de modèles à temps discret (Lovera, 2010; Young, 1998; Lees, 2000; Young et Beven, 1994; Previdi *et al.*, 1999; Previdi et Lovera, 2009). On remarquera d'ailleurs qu'un nombre très restreint d'études portent sur l'applicabilité de ces techniques d'identification en hydrologie urbaine (Lovera, 2010; Previdi *et al.*, 1999; Previdi et Lovera, 2009).

Les principaux avantages des techniques d'identification de modèles à temps continu peuvent être résumés de la manière suivante (Garnier et Wang, 2008; Rao et Garnier, 2002) :

- les paramètres du modèle sont uniques : contrairement aux modèles à temps discret, ils ne sont pas fonction du pas du temps d'échantillonnage ;

- 
- le modèle est exprimé sous la forme d'une équation différentielle. Il peut donc être interprété de manière physique comme les modèles dérivés de lois de la physique tels que la conservation de la masse ou de l'énergie ;
  - l'estimation des paramètres est meilleure pour un grand nombre de pas de temps d'échantillonnage, en particulier pour les pas de temps d'échantillonnage très courts (*i.e.* 5 min ou moins pour les mesures de pluie et de débit) ;

La formulation de modèles à temps continu peut être décrite de manière *naturelle*, les lois physiques étant généralement exprimées sous cette forme. Elle facilite donc l'interprétabilité physique des modèles identifiés. Les modèles linéaires à temps continu sont exprimés sous forme d'équations différentielles et pourront dans notre cas être transposés en modèles de réservoirs. La déduction des paramètres tels que le lag-time et le coefficient de ruissellement du bassin versant seront ainsi facilités.

De plus, les techniques d'identification rendent possible l'identification de modèles parcimonieux, c'est à dire des modèles comprenant un nombre optimum de paramètres. Dans le cas d'une transposition sous forme de modèles de réservoirs, le choix optimal d'association de réservoirs pourra être ainsi réalisé sur la base de critères statistiques indicateurs de la performance et de la parcimonie des modèles.

#### **8.1.4 Couplage de modèle conceptuel et de modèle "statistique"**

La troisième originalité de ce travail réside dans l'identification automatisée d'un modèle global pluie-débit couplant un modèle conceptuel de pertes initiales et un modèle linéaire à temps continu pour les pertes continues et le transfert du ruissellement (cf. FIG 8.3).

La structure générale du modèle pluie-débit proposé est voisine de celle utilisée dans la formulation du modèle *Hydromax* développé pour la prédiction des crues en Belgique (Bastin *et al.*, 2009). Ce modèle associe en effet une fonction de production basée sur des modèles conceptuels et un modèle linéaire à temps discret de type ARX<sup>4</sup> pour le transfert du ruissellement.

---

<sup>4</sup>Auto Régressive model with eXternal inputs. La prédiction du signal de sortie à un instant  $t$  est effectuée sur la base des mesures d'entrée(s) et de sortie du système acquises aux instants précédents

---

Sous réserve de la validité de l'hypothèse de linéarité des pertes continues avec l'intensité de la pluie, l'approche automatisée d'identification du modèle pluie-débit cumule de nombreux avantages :

- la structure du modèle linéaire à temps continu n'est pas déterminée *a priori* ;
- le modèle linéaire à temps continu identifié est interprétable en terme physique par transfert sous forme de modèle réservoir ;
- la procédure d'identification est simplifiée : les pertes continues sont estimées en bloc avec le transfert du ruissellement à partir des données échantillonnées ;
- l'estimation des paramètres du modèle à temps continu est proche de l'optimum ;
- les pertes initiales sont estimées simplement grâce à une procédure itérative ;

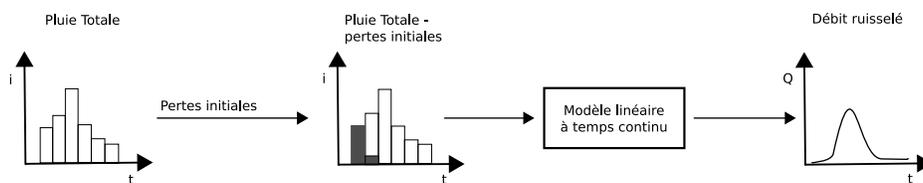


Fig. 8.3 – Approche hybride de modélisation proposée

## 8.2 Modèle pluie-débit proposé

### 8.2.1 Modèle de pertes initiales

Les pertes initiales introduisent des non linéarités dans la transformation pluie-débit. Leur importance dépend à la fois des caractéristiques du bassin versant (présence de végétation, pente,...) et des conditions antécédentes à chaque événement pluvieux.

Nous proposons l'emploi d'un nouveau modèle de pertes, compatible avec la simulation sur des chroniques de pluies continues, qui retranche des pertes initiales à l'intensité de la pluie totale. Les modèles de pertes initiales nécessitent par définition d'identifier le début de chaque événement pluvieux. Ceci ne pose pas de problème particulier lorsque les simulations sont menées sur des événements pluvieux indépendants.

Lorsque le modèle est utilisé sur des séries de pluie continues, il devient au contraire indispensable de repérer le commencement de chaque précipitation, ce qui s'avère en pratique très compliqué. Pour contourner ce problème, nous avons utilisé un modèle reposant sur le calcul de la somme des précipitations sur une fenêtre glissante de durée  $d$  précédant le pas de temps de mesure de l'intensité de la pluie (cf FIG. 8.4) :

$$S(j) = \sum_{k=j-d+1}^{k=j} I_T(k)\Delta t \quad (8.1)$$

où  $\Delta t$  est le pas de temps d'échantillonnage. L'idée de cette fenêtre glissante est de tenir compte des conditions antécédentes à l'événement pluvieux. Le modèle contient un second paramètre : la lame d'eau de référence de pertes initiales  $PI$ . A chaque pas de temps  $j$ ,  $j \in 1, \dots, n$ , la fonction somme  $S(j)$  est comparée à la valeur de référence de pertes initiales  $PI$ . Si  $S(j)$  est inférieur à  $PI$ , le bassin versant a encore une capacité de rétention de la lame d'eau précipitée, et l'intensité résultante  $I_{eff*}$  est nulle. Par contre, si la somme  $S(j)$  est supérieure à la valeur de référence des pertes initiales  $PI$ , alors l'intensité de pluie résultante est égale à l'intensité de pluie totale  $I_T(j)$ .

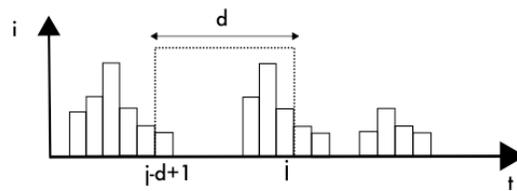


Fig. 8.4 – Schéma de principe du modèle de pertes initiales

Le modèle peut être mathématiquement décrit de la manière suivante :

$$\left\{ \begin{array}{l} \text{si } S(j) < PI \text{ et } S(j+1) < PI \Rightarrow I_{eff*}(j) = 0 \\ \text{si } S(j) < PI \text{ et } S(j+1) \geq PI \\ \quad \Rightarrow I_{eff*}(j+1) = \frac{(S(j+1) - PI)}{\Delta t} \\ \text{sinon } I_{eff*}(j) = I_T(j) \end{array} \right. \quad (8.2)$$

Dans ce système d'équations,  $I_{eff*}$  représente l'intensité de la pluie totale retranchée des pertes initiales. On remarquera que le modèle fait l'hypothèse que le bassin versant est sec après une durée  $d$  consécutive à la dernière précipitation. Ce modèle est voisin du modèle de réservoir décrit au §7.3.1.1. Il comporte toutefois un taux d'évaporation variable<sup>5</sup> qui est nul lorsque les précipitations sont importantes. De plus, ce modèle ne nécessite pas d'estimation de l'évaporation et de l'évapotranspiration, ce qui constitue un avantage. En effet, l'évaporation et l'évapotranspiration ne sont pas constantes et dépendent de paramètres extérieurs non mesurés comme la température ou l'ensoleillement.

## 8.2.2 Modèle de pertes continues et transfert du ruissellement

### 8.2.2.1 Position du problème d'identification

Après avoir pris en compte les pertes initiales, nous cherchons à :

- identifier la structure optimale d'un modèle linéaire à temps-continu incluant les pertes continues et le transfert du ruissellement ;
- et estimer ses paramètres depuis des données échantillonnées (i.e. discrètes) de pluie et débit.

L'hypothèse principale du problème d'identification est que le système associant les pertes continues au transfert du ruissellement est un système mono-entrée, mono-sortie, linéaire à temps continu, causal et invariant dans le temps. Ce système est alors décrit par l'équation différentielle à coefficients constants suivante :

$$a_0^o q(t) + a_1^o \frac{dq(t)}{dt} + \dots + \frac{d^{n_a} q(t)}{dt^{n_a}} = b_0^o I_{eff*}(t - \tau) + b_1^o \frac{dI_{eff*}(t - \tau)}{dt} + \dots + b_{n_b}^o \frac{d^{n_b} I_{eff*}(t - \tau)}{dt^{n_b}} \quad (8.3)$$

$I_{eff*}(t)$  représente le signal d'entrée, ici la pluie totale retranchée des pertes initiales.  $q(t)$  correspond à la réponse non bruitée du système à  $I_{eff*}(t)$ , i.e le débit ruisselé non bruité à l'exutoire du bassin versant.  $\tau$  est un retard représentant le décalage temporel entre le signal d'entrée et de sortie.

<sup>5</sup>Le taux d'évaporation vaut :  $\max(I_T(j) - I_T(j - d), 0)$

L'équation (8.3) peut également s'écrire sous la forme compacte suivante :

$$q(t) = G_o(p) I_{eff*}(t) = \frac{B_o(p)}{A_o(p)} I_{eff*}(t - \tau) \quad (8.4)$$

où  $G_o(p)$  est l'opérateur de transfert du système vrai.  $p$  représente l'opérateur de dérivée ( $p^n x(t) = \frac{d^n x(t)}{dt^n}$ ). Enfin,  $A_o(p)$  et  $B_o(p)$  sont des polynômes, supposés premiers entre eux, qui peuvent être écrits sous la forme suivante :

$$A_o(p) = p^{n_a} + a_{n_a-1}^o p^{n_a-1} + \dots + a_1^o p + a_0^o \quad (8.5a)$$

$$B_o(p) = b_{n_b}^o p^{n_b} + b_{n_b-1}^o p^{n_b-1} + \dots + b_1^o p + b_0^o, \quad n_a \geq n_b \quad (8.5b)$$

Par ailleurs, on suppose que les perturbations agissent sur la sortie non bruitée sous la forme d'un bruit additif et qu'elles sont non corrélées avec le signal d'entrée  $I_{eff*}(t - \tau)$ . Le système vrai est alors décrit par :

$$S : Q(t) = G_o(s) I_{eff*}(t - \tau) + H_o(s) e_o(t) \quad (8.6)$$

avec :

- $Q(t)$  : sortie bruitée du système, *i.e.* le débit mesuré,
- $H_o(s)$  : fonction rationnelle qui décrit le bruit additif sur la sortie,
- $e_o(t)$  : bruit blanc à temps continu à moyenne nulle et variance égale à  $\lambda_o \delta(t)$
- $\delta(t)$  : impulsion de dirac

Les modèles recherchés sont de la forme :

$$\mathcal{M} : \begin{cases} q(t_k) = G(p, \boldsymbol{\theta}) u(t_k - \tau) \\ Q(t_k) = q(t_k) + v(t_k) \end{cases} \quad (8.7)$$

avec :

$$\boldsymbol{\theta} = [a_{n_a-1}, \dots, a_0, b_{n_b}, \dots, b_0]^T \quad (8.8)$$

Le remplacement de  $t$  par  $t_k$  illustre que le modèle est recherché à partir de données échantillonnées, avec  $t_k = kT_e$ , où  $T_e$  est la période d'échantillonnage. Le terme perturbation est modélisé comme un bruit à temps discret à moyenne nulle  $v(t_k)$ . D'autre part,  $G(p, \boldsymbol{\theta})$  est

---

l'opérateur de transfert défini par :

$$\mathcal{G} : G(p, \theta) = \frac{B(s)}{A(s)} = \frac{b_{n_b}^o p^{n_b} + b_{n_b-1}^o p^{n_b-1} + \dots + b_1^o p + b_0^o}{p^{n_a} + a_{n_a-1}^o p^{n_a-1} + \dots + a_1^o p + a_0^o} \quad (8.9)$$

Le problème d'identification consiste donc :

- à identifier la meilleure structure mathématique du modèle, représentée par le trio  $[n_b \ n_a \ \tau]$  pour l'équation (8.3) ;
- et à estimer le vecteur de paramètres  $\theta$  qui caractérise cette structure, à partir de  $N$  échantillons de signaux d'entrée et sortie  $E_N = \{I_{eff*}(t_k) ; Q(t_k)\}_{k=1}^N$ .

### 8.2.2.2 Algorithme SRIVC (Simplified Refined Instrumental Variable for Continuous-time models)

Nous considérerons par la suite que  $\tau=0$  afin d'alléger les notations.

L'identification de la structure du modèle et l'estimation des paramètres associés seront effectués avec la méthode de l'estimateur optimal de la variable instrumentale SRIVC. Avant d'aborder cette méthode d'identification, nous présenterons les méthodes de base que sont :

- la méthode des filtres de variables d'état (FVE),
- et la méthode de la variable instrumentale / FVE.

La description de ces méthodes est inspirée des ouvrages et articles de Garnier et Wang (2008) et Garnier *et al.* (2007).

#### 8.2.2.2.a Méthode des filtres de variables d'état

La méthode des filtres de variables d'état est une des approches classiques d'identification de modèles paramétriques à temps continu. Elle est fondée sur la minimisation d'une fonction de coût basée sur l'erreur d'équation et comprend deux étapes :

- la première consiste à reconstruire les dérivées successives des signaux d'entrée et de sortie en appliquant un filtrage linéaire aux données échantillonnées ;

- 
- la seconde étape est dédiée à l'estimation paramétrique avec la technique d'estimation des moindres carrés.

Le filtrage linéaire est une étape spécifique aux approches d'identification de modèles à temps continu. En effet, la résolution du système d'équations par moindres carrés nécessite l'estimation des dérivées successives des données échantillonnées d'entrée et de sortie, celles-ci n'étant pas mesurées.

### Filtrage linéaire

Considérons l'équation différentielle (8.7) du modèle recherché dans un contexte non bruité. L'application d'un filtre linéaire de forme  $F(p) = 1/E(p)$ , aux deux membres de l'équation précédente donne :

$$\frac{p^{n_a}}{E(p)} q(t) + \sum_{i=0}^{n_a-1} a_i \frac{p^i}{E(p)} q(t) = \sum_{i=0}^{n_b} b_i \frac{p^i}{E(p)} I_{eff*}(t) \quad (8.10)$$

Dans la méthode des filtres de variables d'état, le filtre prend la forme suivante :

$$F(s) = \frac{1}{E(s)} = \frac{1}{(s + \lambda)^{n_a}} \quad (8.11)$$

où  $s$  est la variable de Laplace et  $\lambda$  est la pulsation de coupure du filtre. Définissons alors un ensemble de filtres  $L_j(s)$  pour  $j = 0, 1, \dots, n_a$  ayant pour forme :

$$L_j(s) = \frac{s^j}{E(s)} = \frac{(\lambda)^{n_a} s^j}{(s + \lambda)^{n_a}} \quad (8.12)$$

et soit  $I_j(t)$  leur réponse impulsionnelle. L'équation (8.10) peut alors être réécrite de la manière suivante :

$$q_f^{(n_a)}(t) + a_{n_a-1} q_f^{(n_a-1)}(t) + \dots + a_0 q_f(t) = b_{n_b} I_{eff*,f}^{(n_b)}(t) + b_{n_b-1} I_{eff*,f}^{(n_b-1)}(t) + \dots + b_0 I_{eff*,f}(t) \quad (8.13)$$

où  $q_f(t)$  et  $l_{eff*,f}(t)$  sont les sorties et entrées préfiltrées du système :

$$\begin{cases} q_f^{(i)}(t) = [l_i * q](t) \\ l_{eff*,f}^{(i)}(t) = [l_i * l_{eff*}](t) \end{cases} \quad (8.14)$$

avec :

$l_i(t)$  : réponse impulsionnelle des filtres  $L_j(s)$  ;  
 $*$  : opérateur de convolution

Les sorties des filtres  $q_f^{(i)}(t)$  et  $l_{eff*,f}^{(i)}(t)$  fournissent ainsi une estimation des dérivées des signaux d'entrée et de sortie dans la bande fréquentielle d'intérêt, *i.e.* dans la plage de fréquence dans laquelle on recherche l'adéquation entre le système et le modèle. Ces estimées peuvent ensuite être utilisées au sein de techniques de régression linéaire.

A l'instant  $t = t_k$ , et en prenant en compte un bruit additif sur la sortie, l'équation (8.13) peut être réécrite sous forme de régression linéaire :

$$q_f^{(n_a)}(t_k) = \boldsymbol{\varphi}^T(t_k)\boldsymbol{\theta} + \varepsilon_{EE}(t_k, \theta) \quad (8.15)$$

où  $\boldsymbol{\varphi}^T(t_k)$  est le vecteur de régression qui prend la forme :

$$\boldsymbol{\varphi}^T(t_k) = [-Q_f^{(n_a-1)}(t_k) \cdots - Q_f(t_k) l_{eff*,f}^{(n_b)}(t_k) \cdots l_{eff*,f}(t_k)] \quad (8.16)$$

et avec :

$\boldsymbol{\theta}$  : vecteur de paramètres défini par l'équation (8.8) ;  
 $\varepsilon_{EE}(t_k, \theta)$  : erreur d'équation.

### Estimateur des moindres carrés / FVE

Disposant de N échantillons mesurés aux instants  $t_1, t_2, \dots, t_N$ , l'équation (8.15) peut être réécrite sous forme matricielle de la manière suivante :

$$\mathbf{Q}_N = \boldsymbol{\Phi}_N \boldsymbol{\theta} + \mathbf{V}_N \quad (8.17)$$

---

avec :

- $\mathbf{Q}_N$  : vecteur colonne  $[N \times 1]$  regroupant les mesures de débit aux instants  $t_1, t_2, \dots, t_N$  ;
- $\Phi_N$  : matrice de dimension  $[N \times (n_a + n_b + 1)]$  regroupant les vecteurs de régression aux instants  $t_1, t_2, \dots, t_N$  ;
- $\mathbf{V}_N$  : vecteur de dimension  $[N \times 1]$  regroupant les erreurs d'équation aux instants  $t_1, t_2, \dots, t_N$ .

L'estimée  $\hat{\theta}$  du vecteur de paramètres est le vecteur minimisant :

$$\hat{\theta} = \arg \min_{\theta} \mathbf{V}_N(\theta, E_N) = \arg \min_{\theta} \sum_{k=1}^N \varepsilon_{EE}^2(t_k, \theta) \quad (8.18)$$

L'estimateur des moindres carrés est donc donné par :

$$\hat{\theta}_{mc} = [\Phi_N^T \Phi_N]^{-1} \Phi_N^T \mathbf{Q}_N \quad (8.19)$$

En présence de bruit blanc ou coloré, l'estimateur des moindres carrés associé au filtre de variable d'état est asymptotiquement biaisé (Garnier et Wang, 2008; Soderstrom et Stoica, 2002). Il ne peut donc être utilisé sur des données mesurées de débit qui sont la plupart du temps très bruitées.

### 8.2.2.2.b Estimateur de la variable instrumentale / FVE

L'estimateur des moindres carrés ne peut pas être utilisé sur des données bruitées en raison de la corrélation du bruit avec les dérivées successives du signal de sortie. Pour contourner ce problème, un autre estimateur, comme celui de la variable instrumentale doit être utilisé. L'estimateur de la variable instrumentale est une variante de la méthode des moindres carrés qui présente l'avantage de reposer sur les techniques de régression linéaire.

Cette méthode consiste à introduire un vecteur  $\mathbf{z}(t_k)$  appelé instrument ou variable instrumentale, dont les composantes également appelées instruments ou variables instrumentales soient suffisamment corrélées avec les composantes du vecteur de régression  $\varphi(t_k)$  mais non corrélées avec le bruit additif sur la sortie  $v(t_k)$ .

Plusieurs choix pour l'instrument sont possibles. Parmi ceux-ci l'estimateur de la variable instrumentale à modèle auxiliaire est particulièrement facile à mettre en oeuvre. Le vecteur de variables instrumentales est dans ce cas construit de la manière suivante :

$$\mathbf{z}^T(t_k) = F(p) \left[ -q^{(n_a-1)}(t_k, \hat{\boldsymbol{\theta}}_{mc}) \cdots - q(t_k, \hat{\boldsymbol{\theta}}_{mc}) l_{eff*}^{(n_b)}(t_k) \cdots l_{eff*}(t_k) \right] \quad (8.20)$$

avec :

$F(p)$  : filtre de variables d'état défini par l'équation (8.11) ;

$q(t_k, \hat{\boldsymbol{\theta}}_{mc})$  : sortie d'un modèle auxiliaire obtenu à partir de l'estimation par moindres carrés / FVE :

$$q_f^{(n_a-1)}(t_k, \hat{\boldsymbol{\theta}}_{mc}) = G(p, \hat{\boldsymbol{\theta}}_{mc}) l_{eff*}(t_k)$$

Disposant de  $N$  échantillons mesurés aux instants  $t_1, t_2, \dots, t_N$ , l'estimateur de la variable instrumentale à modèle auxiliaire associé au filtre de variables d'état est alors donné par :

$$\hat{\boldsymbol{\theta}}_{vi} = [\mathbf{Z}_N^T \boldsymbol{\Phi}_N]^{-1} \mathbf{Z}_N^T \mathbf{Q}_N \quad (8.21)$$

où  $\mathbf{Z}_N$  est la matrice de dimension  $[N \times (n_a + n_b + 1)]$  regroupant l'ensemble des vecteurs de variables instrumentales pour chaque instant d'échantillonnage  $t_1, t_2, \dots, t_N$ .

L'estimateur de la variable instrumentale/FVE est asymptotiquement non biaisé mais n'est pas à variance minimale (Garnier et Wang, 2008). Un choix plus adéquat d'instrument doit donc être recherché afin de diminuer la variance de l'estimation des paramètres. Cette méthode mérite cependant d'être considérée avec attention. En effet, elle permet de fournir une première estimation des paramètres pour des algorithmes itératifs comme l'algorithme SRIVC.

---

### 8.2.2.2.c Estimateur optimal de la variable instrumentale SRIVC

Il a été montré que l'estimateur optimal de la variable instrumentale est obtenu pour le choix d'instrument suivant (Garnier et Wang, 2008; Soderstrom et Stoica, 2002) :

$$\begin{cases} \mathbf{z}_{opt}^T(t_k) = F_{opt}(\rho) \left[ -q^{(n_a-1)}(t_k) \cdots - q(t_k) l_{eff*}^{(n_b)}(t_k) \cdots l_{eff*}(t_k) \right] \\ F_{opt}(\rho) = \frac{1}{H_o(\rho)A_o(\rho)} \end{cases} \quad (8.22)$$

$H_o(\rho)$  : "vrai" modèle de bruit (cf. eq. 8.6) ;

$A_o(\rho)$  : dénominateur de la fonction de transfert du système vrai ;

$q(t_k)$  : sortie non bruitée du système à l'instant  $t_k$  ;

L'estimateur optimal nécessite la connaissance de caractéristiques du système qui ne sont pas accessibles dans des applications réelles. Pour contourner ce problème, on recourt soit à des procédures itératives, soit à des algorithmes comprenant plusieurs étapes.

L'algorithme SRIVC (*Simplified Refined Instrumental Variable method for Continuous-time systems*) est un algorithme itératif proposé par (Young, 2002) et qui constitue une simplification de l'algorithme RIVC développé par le même auteur (Young et Jakeman, 1980). Il a été testé avec succès dans de nombreuses applications pratiques, dont des modélisations pluie-débit (Young et Garnier, 2006; Young, 2004). Le code de calcul est implémenté dans les boîtes à outils CAPTAIN (cf. <http://www.es.lancs.ac.uk/cres/captain/>) et CONTSID (cf. <http://www.cran.uhp-nancy.fr/>) développées pour Matlab.

Cette méthode représente une extension logique de la méthode de la variable instrumentale optimale/FVE. Elle fait l'hypothèse que le bruit additif est blanc gaussien. Le filtre optimal  $F_{opt}(\rho)$  défini dans l'équation (8.22) devient alors égal à  $1/A(\rho)$ . Or,  $A(\rho)$  est *a priori* inconnu. On utilise alors un algorithme itératif qui va ajuster progressivement l'estimée de la variable instrumentale  $\hat{\theta}^i$  à partir d'une estimée initiale  $\hat{\theta}^0$  jusqu'à convergence. L'algorithme SRIVC comporte les étapes suivantes :

Etape 1. Initialisation : calcul d'une première estimation du vecteur de paramètre  $\hat{\theta}^0$  et du filtre à temps continu  $F_{opt}(\rho, \hat{\theta}^0)$  avec l'algorithme des moindres carrés.

Etape 2. Estimation itérative du vecteur de paramètres  $\hat{\theta}_{SRIVC}$

---

Pour  $j = 1$  à convergence

- (a) Générer la série de sortie non bruitée  $\hat{q}(t_k, \hat{\theta}^{j-1})$  en utilisant le modèle auxiliaire construit avec les polynômes  $A(p, \hat{\theta}^{j-1})$  et  $B(p, \hat{\theta}^{j-1})$  basés sur le vecteur de paramètre  $\hat{\theta}^{j-1}$  obtenu à l'itération  $j - 1$ .
- (b) Préfiltrer les séries d'entrée  $I_{eff*}(t_k)$ , de sortie non bruitée  $\hat{q}(t_k, \hat{\theta}^{j-1})$  et de sortie mesurée  $Q(t_k)$  avec le filtre  $F(p, \hat{\theta}^{j-1})$  et former les matrices  $\mathbf{Z}_N(\theta^{j-1})$  et  $\Phi_N(\theta^{j-1})$ .
- (c) Calculer le vecteur de paramètres  $\hat{\theta}^j$  d'après l'équation (8.21).

La convergence est atteinte lorsque la variation relative sur les paramètres devient inférieure à un seuil de tolérance donné.

L'estimateur SRIVC est asymptotiquement non biaisé quelle que soit la nature du bruit additif, et à variance minimale lorsque le bruit additif est blanc. Lorsque le bruit est coloré, comme dans la plupart des applications pratiques en hydrologie, l'estimateur est toujours asymptotiquement sans biais, mais sa variance n'est plus minimale. Les nombreuses applications sur lesquelles l'algorithme a été testé ont néanmoins montré que les paramètres étaient estimés avec une variance relativement faible (Garnier et Wang, 2008). On retiendra que l'algorithme SRIVC est une méthode itérative, rapidement convergente, robuste au bruit, et qui donne des estimations des paramètres proches des estimations optimales d'un point de vue statistique, même si le bruit associé au modèle est coloré.

### 8.2.2.3 Identification de la structure du modèle

L'identification de la structure du modèle, *i.e.*  $[n_b \ n_a \ \tau]$ , est effectuée grâce à deux critères statistiques. Le premier est le coefficient de détermination  $R_T^2$ , basé sur l'erreur entre la sortie mesurée et la sortie simulée par le modèle :

$$R_T^2 = 1 - \frac{\sigma_{\hat{\varepsilon}}^2}{\sigma_Q^2} \quad (8.23)$$

où  $\sigma_{\hat{\varepsilon}}^2$  est la variance du bruit estimé  $\hat{\varepsilon}(t_k)$  et  $\sigma_Q^2$  est la variance de la sortie mesurée  $Q(t_k)$ . Le second critère, *YIC*, mesure de l'identifiabilité du modèle *i.e.* la manière dont les paramètres

estimés sont définis statistiquement (voir *i.e.* Young et Garnier (2006); Garnier et Wang (2008)).

$$YIC = \log_e \frac{\hat{\sigma}_{\hat{\epsilon}}^2}{\sigma_Q^2} + \log_e \{NEVN\}; \quad NEVN = \frac{1}{n_\theta} \sum_{i=1}^{n_\theta} \frac{\hat{p}_{ii}}{\hat{\theta}_i^2} \quad (8.24)$$

Dans cette relation,  $n_\theta$  est le nombre de paramètres estimés,  $\hat{p}_{ii}$  est le  $i^{\text{ème}}$  élément de la diagonale de la matrice de covariance  $\mathbf{P}_\theta$ .  $\hat{p}_{ii}$  est donc une mesure de la variance de l'incertitude sur le  $i^{\text{ème}}$  paramètre estimé.  $\hat{\theta}_i^2$  est le carré du  $i^{\text{ème}}$  paramètre estimé dans le vecteur  $\theta$ .  $\hat{p}_{ii}/\hat{\theta}_i^2$  est donc une mesure normalisée de l'incertitude sur le  $i^{\text{ème}}$  paramètre estimé.

Plus la valeur de  $R_T^2$  est proche de 1 et plus l'écart entre la sortie estimée et la sortie mesurée est faible. De la même manière, plus  $YIC$  est négatif et plus le modèle est parcimonieux. La structure optimale de modèle résulte donc d'un compromis entre précision et parcimonie.

La méthode d'identification de la structure du modèle à temps continu présente deux avantages principaux :

1. la structure du modèle est choisie *a posteriori* sur la base de deux critères statistiques indicateurs de la précision et de la parcimonie des modèles ;
2. l'identification et l'estimation des paramètres peuvent être automatisés en définissant une fonction de coût basée sur les critères  $R_T^2$  et  $YIC$ .

### 8.2.2.4 Interprétation physique : équivalence des structures identifiées avec les modèles de réservoirs

#### 8.2.2.4.a Méthodologie

L'approche DBM consiste à ne valider un modèle que si une interprétation physique peut être donnée à la structure identifiée ainsi qu'aux paramètres estimés. La validation des modèles a été effectuée en deux étapes :

1. transfert du modèle linéaire identifié sous forme de modèle de réservoir de manière à en faciliter l'interprétation physique ;
2. validation du modèle par comparaison des caractéristiques du modèle et des caractéristiques du bassin versant.

---

Tous les modèles de réservoirs <sup>6</sup> sont des modèles linéaires à temps continu. Cependant, tous les modèles linéaires à temps continu ne peuvent pas être transférés sous forme de modèle de réservoirs. Les racines des polynômes des fonctions de transfert doivent en effet être réelles et positives.

### 8.2.2.5 Exemples

Un modèle du première ordre caractérisé par le trio [ $n_b = 0$   $n_a = 1$   $\tau = 0$ ] correspond au modèle du réservoir linéaire largement utilisé en hydrologie pour le transfert du ruissellement (voir Beven (2003) ou CERTU (2003)). En effet, un modèle du premier ordre identifié selon la procédure décrite dans les paragraphes précédents s'écrit :

$$\frac{dQ_{ruis}(t)}{dt} + a_0 Q_{ruis}(t) = b_0 I_{eff*}(t) \quad (8.25)$$

Le modèle du réservoir linéaire s'exprime quant à lui par :

$$\frac{dQ_{ruis}(t)}{dt} + \frac{1}{K} Q_{ruis}(t) = \frac{1}{K} Q_{eff}(t) \quad (8.26)$$

La différence entre les deux systèmes provient du terme source. Si nous considérons que l'infiltration est proportionnelle à l'intensité de la pluie, nous pouvons écrire :

$$Q_{eff}(t) = C_R A I_{eff*}(t) = A_{eff} I_{eff*}(t) \quad (8.27)$$

où  $C_R$  est le coefficient de ruissellement, considéré comme constant pendant un événement pluvieux et quelque soit l'événement pluvieux.  $A_{eff}$  exprime la surface effective<sup>7</sup> du bassin versant qui contribue au ruissellement.

Il est donc dorénavant possible de relier les paramètres de l'équation (8.25) à ceux des équations (8.26) et (8.27) :

$$\begin{cases} a_0 = \frac{1}{K}, \\ b_0 = \frac{A_{eff}}{K}. \end{cases} \quad (8.28)$$

---

<sup>6</sup>modèle du réservoir linéaire, cascade de réservoirs, associations de réservoirs en parallèle

<sup>7</sup>également appelée surface active

---

L'équation (8.28) montre que les paramètres physiques du réservoir linéaire ( $K$ ,  $C_R$ ) peuvent être déterminés de manière automatique et optimale depuis  $a_0$  and  $b_0$  pour une structure du premier ordre.

Considérons dorénavant pour l'exemple le cas de deux réservoirs associés en parallèle. La fonction de transfert associée à un réservoir linéaire s'exprime par :

$$H(s) = \frac{B(s)}{A(s)} = \frac{\beta_0}{s + \alpha_0} \quad (8.29)$$

avec  $\alpha_1 = 1/K$  et  $\beta_0 = A_{eff} * \alpha_0$ .

La fonction de transfert  $H(s)$  de 2 réservoirs en parallèle s'écrit donc :

$$H(s) = \frac{\beta_0}{s + \alpha_0} + \frac{\beta'_0}{s + \alpha'_0} = \frac{s(\beta_0 + \beta'_0) + (\alpha_0\beta'_0 + \alpha'_0\beta_0)}{s^2 + s(\alpha_0 + \alpha'_0) + (\alpha_0\alpha'_0)} \quad (8.30)$$

L'identification d'un modèle linéaire à temps continu de deuxième ordre où  $n_b = 1$  et  $n_a = 2$  revient à estimer les paramètres de la fonction de transfert  $H_1(s)$  :

$$H_1(s) = \frac{b_1 s + b_0}{s^2 + a_1 s + a_0} \quad (8.31)$$

Par identification entre les équations (8.30) et (8.31) il est donc possible de relier les paramètres  $b_0$ ,  $b_1$ ,  $a_1$ ,  $a_0$  aux paramètres du réservoir linéaire  $K$  et  $A_{eff}$ . Toutefois, la résolution du système d'équations, n'aboutit pas forcément à des valeurs de  $K$  ou de  $A_{eff}$  ayant un sens physique (valeurs négatives ou complexes).

De la même façon, on peut montrer que tout système décrit par un modèle de réservoirs associés en parallèle, en série ou par une combinaison de réservoirs en parallèle et en série est équivalent à un modèle linéaire à temps continu. Le tableau 8.1 récapitule les différentes associations de réservoirs correspondant aux trio  $[n_b \ n_a \ \tau]$ .

### 8.2.2.6 Cas des modèles MISO (Multiple Input Single Output)

Si un modèle linéaire à temps continu est identifié avec un décalage temporel  $\tau$ , le décalage temporel des modèles de réservoirs résultant de la procédure de décomposition décrite au paragraphe précédent sera aussi égal à  $\tau$ .

Association de réservoirs	Correspondance $[n_b \ n_a \ \tau]$
1 réservoir	[0 1 0]
2 réservoirs en série	[0 2 0]
3 réservoirs en série	[0 3 0]
2 réservoirs en parallèle	[1 2 0]
2 réservoirs en parallèle et 1 réservoir en série	[1 3 0]
3 réservoirs en parallèle	[2 3 0]

Tab. 8.1 – Correspondance des structures de modèles linéaires identifiées avec les modèles de réservoirs

L'identification d'une fonction de transfert composée de deux modèles de réservoirs associés en parallèle avec des décalages temporels différents nécessite d'effectuer la procédure d'identification avec des modèles à plusieurs entrées. Chaque entrée du modèle intègre alors un décalage temporel différent. Il convient toutefois de vérifier que les deux entrées soient suffisamment décorréelées.

## 8.2.3 Méthodologie itérative de calage de l'ensemble du modèle pluie-débit

### 8.2.3.1 Présentation de la méthodologie

Un algorithme itératif est proposé pour identifier la structure optimale du modèle linéaire à temps continu et estimer l'ensemble des paramètres du modèle pluie-débit. Il comporte les étapes suivantes :

1. Initialisation avec l'identification de la structure mathématique du modèle linéaire à temps continu depuis la pluie totale. La meilleure structure mathématique du modèle, *i.e.* le meilleur trio  $[n_b \ n_a \ \tau]$  est choisi sur la base des critères  $R_T^2$  et  $YIC$ .
2. Le modèle linéaire à temps continu identifié à l'étape 1 est fixé. Une recherche exhaustive du couple de paramètres  $(PI, d)$ <sup>8</sup> qui maximise le coefficient  $R_T^2$  entre la sortie simulée et mesurée est effectuée.
3. Les valeurs des paramètres du modèle de pertes initiales sont fixées. Il est alors vérifié que la structure  $[n_b \ n_a \ \tau]$  choisie depuis la pluie totale à l'étape 1 est toujours la même

<sup>8</sup>couple de paramètres du modèle de pertes initiales

---

en utilisant comme entrée du modèle linéaire à temps continu la pluie totale défalquée des pertes initiales. Si tel n'est pas le cas, l'algorithme recommence à l'étape 2.

On remarquera que cet algorithme est simple mais n'aboutit pas forcément à une solution optimale. En effet, le meilleur couple  $(PI, d)$  n'est pas estimé pour toutes les structures possibles  $[n_b, n_a, \tau]$ . Cependant, l'importance relative des pertes initiales en milieu urbain ne devraient pas trop entraver les performances de cet algorithme qui constitue un bon compromis entre performance et temps de calcul.

### 8.2.3.2 Validation de la méthodologie sur des données de simulation

L'objectif de ce paragraphe est de valider la méthodologie d'identification en effectuant des simulations de Monte-Carlo. Une chronique de débits a été créée à partir d'un modèle pluie-débit de même structure que celui que nous cherchons à identifier sur le bassin versant du Stang Alar. Les mesures de pluie ont été acquises à Saint-Malo. Un bruit additif coloré a été ajouté sur la sortie du modèle de la manière à générer trente chroniques de débits. Les débits bruités sont exprimés de la manière suivante :

$$Q(t) = q(t)(1 + 0,06e(t)) + 0,006e(t) \quad (8.32)$$

avec  $Q(t)$  la sortie bruitée du modèle,  $q(t)$  la sortie du système vrai, et  $e(t)$  un bruit caractérisé par une distribution de probabilité normale ( $\bar{x}=0$ ;  $\sigma=1$ ). Les résultats des simulations sont présentées dans le tableau 8.2.

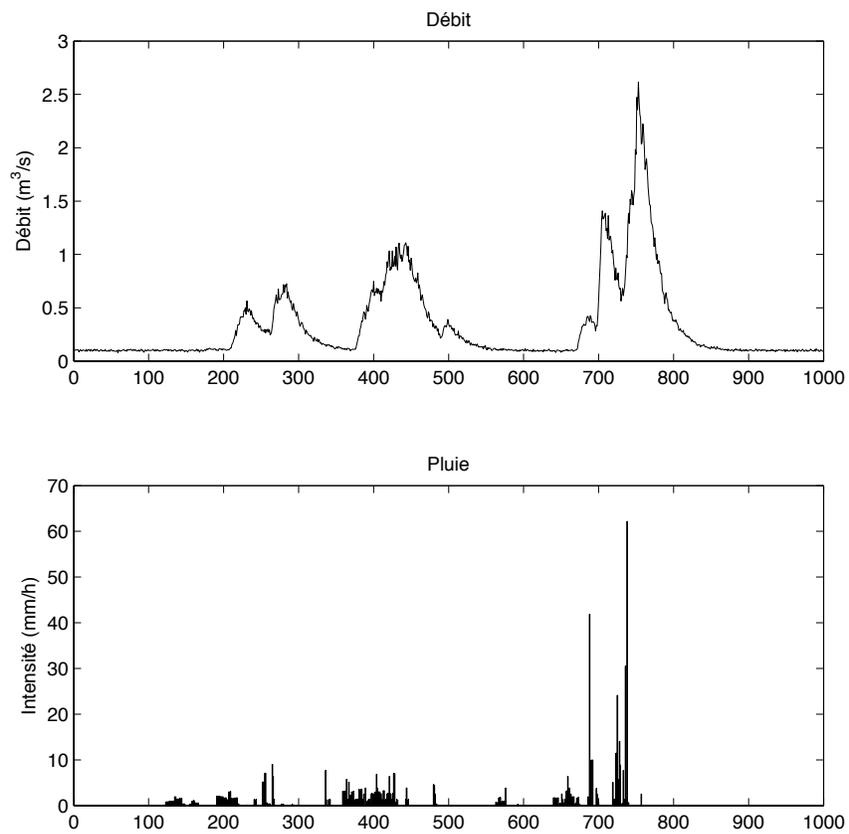


Fig. 8.5 – Extrait de la chronique de pluies et de débits utilisée pour les simulations de Monte-Carlo (30 chroniques de débits comprenant 65475 observations)

	<b>PI</b> (mm)	<b>durée</b> (mm/h)	<b>[n<sub>b</sub> n<sub>a</sub> τ]</b> -	<b>R<sub>T</sub><sup>2</sup></b> -	<b>YIC</b> -
Moyenne	1,6	3	[0 1 3]	0,997	-21,25
Écart type	0	0	-	3 10 <sup>-4</sup>	0,21
Valeur vraie	1,6	3	[0 1 3]	-	-
Moyenne	1,6	3	[0 2 10]	0,992	-17,14
Écart type	0	0	-	3,5 10 <sup>-4</sup>	0,098
Valeur vraie	1,6	3	[0 2 10]	-	-

Tab. 8.2 – Résultats des 30 simulations de Monte-Carlo

L'algorithme itératif utilisé permet donc :

- d'identifier la structure optimale de modèle à temps continu avec une estimation non biaisée de la valeur des paramètres<sup>a</sup> ;
- d'estimer les valeurs vraies de paramètres du modèle de pertes initiales. Les valeurs de pertes initiales sont recherchées de manière exhaustive. Ceci explique que l'écart type sur l'estimation de la valeur des paramètres soit nulle.

L'approche de modélisation proposée va dorénavant être testée sur le bassin versant du Stang Alar à Brest.

<sup>a</sup>les estimations de la valeur des paramètres n'ont pas été intégrées de manière à ne pas alourdir le tableau.

### 8.3 Conclusion

Le modèle pluie-débit proposé couple un modèle conceptuel de pertes initiales et un modèle linéaire à temps continu intégrant les pertes continues et le transfert du ruissellement. Le choix de cette structure de modèle résulte de la volonté d'interpréter la transformation pluie débit avec les approches habituellement utilisées en hydrologie urbaine. La modélisation des pertes continues et du transfert du ruissellement avec un modèle linéaire à temps continu répond à cette volonté. En effet, les modèles linéaires à temps continu ont des formulations mathématiques voisines des modèles conceptuels de réservoir qui faciliteront par voie de conséquence leur interprétation.

---

Le modèle pluie-débit sera calé directement à partir de données échantillonnées en faisant appel à des techniques d'identification de systèmes. Les méthodes choisies permettent de formaliser le problème mathématique d'estimation paramétrique d'une manière à la fois *élégante* et performante. Elles présentent également l'avantage de ne pas fixer *a priori* la structure mathématique de modèle à identifier. Le choix de la structure est effectué *a posteriori* sur la base des valeurs de deux critères statistiques indicateurs de la performance et de la parcimonie du modèle.

Enfin, la validation du modèle est effectuée avec une approche dite *data-based mechanistic modelling* consistant à ne retenir un modèle que si une interprétation physique cohérente peut être donnée au modèle ainsi qu'aux paramètres identifiés. Le choix de la structure du modèle facilitera donc l'interprétation et la validation du modèle.

Une procédure itérative a été proposée afin d'identifier la structure optimale du modèle linéaire à temps continu et estimer l'ensemble des paramètres du modèle pluie-débit. Elle a été testée avec succès sur un jeu de données de simulation dans le cadre de simulations de Monte-Carlo.

Nous proposons dorénavant son application sur un jeu de données pluie-débit acquis sur le bassin versant du Stang Alar.

## Chapitre 9

# Modélisation pluie-débit sur le bassin versant du Stang Alar (Brest)

Après une présentation succincte du site d'étude, nous décrivons les méthodes d'acquisition de données de pluies et de débits. Les données mesurées seront alors analysées, puis partitionnées en deux jeux indépendants de calage et validation. La dernière partie du chapitre sera dédiée à la présentation des résultats de l'identification du modèle pluie-débit fondée sur la méthodologie présentée au chapitre précédent.

### 9.1 Site d'étude

Le site d'étude est le bassin versant du Stang Alar situé sur la commune de Brest dans le Finistère (29). Une description succincte de ce bassin versant a déjà été effectuée dans la partie précédente. Le bassin versant a une superficie totale de 580 ha. Sa partie Ouest est très urbanisée. La partie Est du bassin versant est recouverte majoritairement de cultures et de prairies. Notons également la présence de quelques lotissements.

Le bassin est traversé par la rivière Stang Alar qui prend sa source au nord du bassin versant. Le Stang Alar se rejette dans la mer au niveau de la plage du "Moulin Blanc".

Le bassin versant est également caractérisé par de fortes pentes. La pente moyenne du cours d'eau est de 2,3 %. Les pentes dans les directions Est-Ouest et Ouest-Est sont également

---

importantes avec des valeurs moyennes de 6,3 % et 7 %. Nous devons donc nous attendre à des transferts rapides du ruissellement vers l'exutoire du bassin versant.

Sur les parties urbanisées, le réseau d'assainissement est majoritairement de type séparatif. Les conduites pluviales se rejettent dans le Stang Alar. Des rejets d'eaux usées dans les canalisations d'eaux pluviales sont constatés par temps sec. Pour limiter l'impact de ces rejets sur le milieu récepteur, des petits seuils ont été installés à l'exutoire des canalisations d'eaux pluviales afin de rediriger le débit d'eaux usées de temps sec vers les canalisations d'eaux usées.

Des déversements d'eaux usées sont également possibles au niveau de la station de pompage de Palaren située dans la partie aval du bassin versant ainsi qu'au niveau d'un bassin de stockage d'eaux usées situé lui au centre du bassin versant.

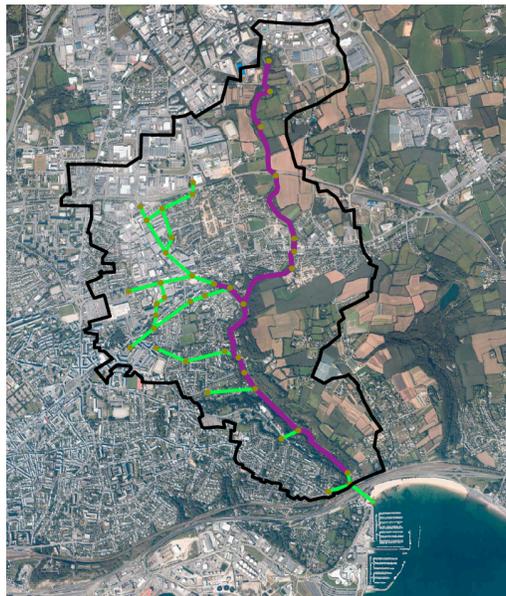


Fig. 9.1 – Vue aérienne du bassin versant du Stang Alar (*en bleu : le Stang Alar ; en jaune : les branches principales du réseau d'assainissement pluvial*)

---

## 9.2 Acquisition de mesures

L'acquisition des mesures a été réalisée directement par Véolia eau pendant une période de deux mois (11/07/2008 au 10/09/2008).

### 9.2.1 Mesures pluviométriques

Les données pluviométriques sont issues de postes pluviométriques de la Communauté Urbaine de Brest Métropole. Cinq postes pluviométriques sont disposés à proximité du bassin versant mais aucun n'est recensé en son sein. Les postes pluviométriques les plus proches sont dénommés Hôtel communautaire, Pen Ar Chleuz et Kerkaredec<sup>1</sup> (voir FIG. 9.2).

Les pluviomètres sont à auget basculeur. La lame d'eau unitaire recueillie par les augets est de 0,2 mm. Le pas de temps d'échantillonnage est de 3 min.

Pendant la campagne de mesure, vingt-cinq événements pluvieux ont été recensés (cf. TAB. 9.1). La lame d'eau précipitée pendant cette période est respectivement de 221, 224,4 et 210,8 mm au niveau des postes Kerkaredec, Hôtel communautaire et Pen ar Chleuz. L'écart des précipitations entre les différents postes sur l'ensemble de la période est de 6,3 %. Les écarts de précipitations à l'échelle d'un événement pluvieux peuvent atteindre 15,5 mm et plus de 110 %. La figure 9.3 illustre également l'hétérogénéité spatiale des pluies constatée pendant la campagne de mesure sur les trois postes pluviométriques. Les courbes de lames d'eau cumulées ont globalement la même allure, mais on constate localement des différences significatives.

Les mesures de précipitations effectuées sur les trois postes pluviométriques entourant le bassin versant mettent en exergue l'hétérogénéité spatiale des pluies. Il est donc primordial de choisir le pluviomètre ou la combinaison de pluviomètre la plus représentative du débit généré à l'exutoire du bassin versant. Ainsi, les modélisations seront effectuées en prenant successivement comme entrée du modèle l'un des trois postes pluviométriques ainsi qu'une combinaison linéaire des postes.

---

<sup>1</sup>Le pluviomètre zone portuaire a été écarté ; en effet, il ne fonctionnait pas pendant la période de mesure

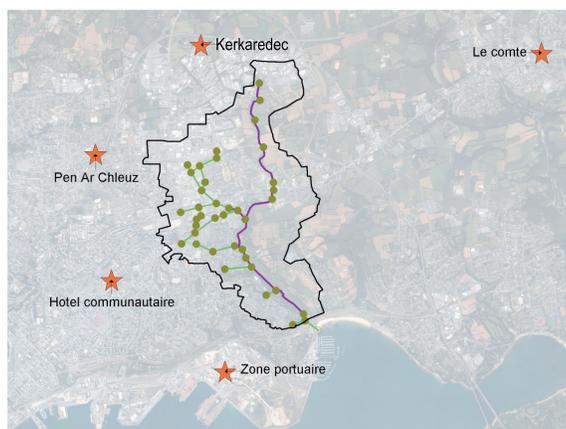


Fig. 9.2 – Situation des postes pluviométriques par rapport au bassin versant du Stang Alar

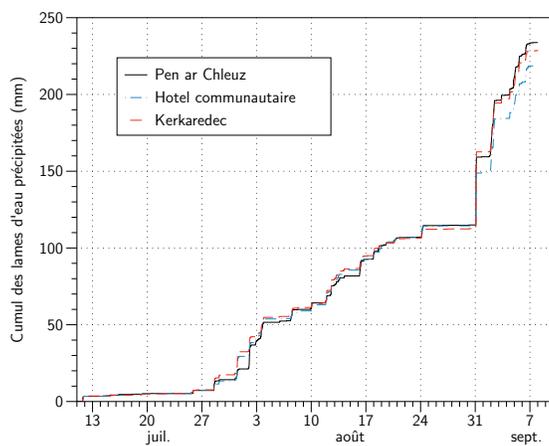


Fig. 9.3 – Cumul des lames d'eau précipitées pendant la campagne de mesure sur les trois postes pluviométriques étudiés

N <sup>o</sup>	Date	Durée	Kerkaredec			Hôtel comm.			Pen ar chleuz		
		d	L	I <sub>max</sub>	$\bar{I}$	L	I <sub>max</sub>	$\bar{I}$	L	I <sub>max</sub>	$\bar{I}$
		(min)	(mm)	(mm/h)	(mm/h)	(mm)	(mm/h)	(mm/h)	(mm)	(mm/h)	(mm/h)
1	11/07 14 :56	210	3,4	20,0	1,0	3,4	12,0	1,0	3,4	8,0	1,0
2	25/07 13 :56	1125	2,4	4,0	0,7	2,2	4,0	0,6	2,0	8,0	0,6
3	28/07 11 :56	399	7,8	20,0	2,2	4,0	24,0	1,1	6,2	16,0	1,8
4	29/07 02 :56	309	2,0	8,0	0,6	2,0	8,0	0,6	0,8	4,0	0,2
5	31/07 02 :26	939	15,0	12,0	4,3	16,0	12,0	4,6	7,0	8,0	2,0
6	01/08 20 :56	636	9,8	12,0	2,8	9,0	12,0	2,6	15,6	16,0	4,5
7	02/08 19 :26	519	2,2	4,0	0,6	4,2	4,0	1,2	3,8	4,0	1,1
8	03/08 10 :26	717	10,2	12,0	2,9	10,8	12,0	3,1	10,6	12,0	3,0
9	07/08 01 :55	915	5,4	16,0	1,5	4,0	16,0	1,1	7,2	20,0	2,1
10	09/08 21 :25	324	2,4	4,0	0,7	4,0	8,0	1,1	4,2	8,0	1,2
11	11/08 10 :55	1023	8,6	32,0	2,5	8,4	40,0	2,4	4,8	16,0	1,4
12	12/08 11 :55	153	6,8	40,0	1,9	5,4	20,0	1,5	5,6	20,0	1,6
13	12/08 19 :28	1128	5,8	20,0	1,7	5,8	16,0	1,7	5,2	20,0	1,5
14	14/08 01 :25	213	1,4	4,0	0,4	2,6	16,0	0,7	1,2	12,0	0,3
15	16/08 03 :25	825	8,2	4,0	2,3	7,0	8,0	2,0	10,8	12,0	3,1
16	17/08 21 :55	228	4,8	8,0	1,4	4,6	8,0	1,3	5,0	8,0	1,4
17	18/08 12 :55	567	2,4	8,0	0,7	4,8	20,0	1,4	3,6	12,0	1,0
18	19/08 03 :55	627	1,0	4,0	0,3	1,8	4,0	0,5	1,8	8,0	0,5
19	24/08 01 :25	381	5,8	4,0	1,7	7,2	4,0	2,1	7,6	8,0	2,2
20	31/08 00 :54	156	50,0	116,0	14,3	34,6	128,0	9,9	44,4	104,0	12,7
21	01/09 18 :45	1011	31,8	16,0	9,1	35,0	20,0	10,0	36,8	12,0	10,5
22	03/09 08 :15	12	2,6	32,0	0,7	0,0	0,0	0,0	3,0	44,0	0,9
23	04/09 08 :24	339	4,4	20,0	1,3	4,0	20,0	1,1	4,2	16,0	1,2
24	04/09 16 :09	2247	20,2	36,0	5,8	19,6	36,0	5,6	22,4	24,0	6,4
25	06/09 08 :39	1008	6,6	12,0	1,9	10,4	16,0	3,0	7,2	12,0	2,1

Tab. 9.1 – Caractéristiques des événements pluvieux mesurés

---

## 9.2.2 Mesures de débit ruisselé à l'exutoire du bassin versant

Les débits ont été évalués à l'exutoire du bassin versant du Stang Alar, avant le rejet dans le milieu marin, dans une canalisation circulaire de diamètre 2000 mm. Le calcul du débit est basé sur des mesures de hauteur d'eau couplées et à une loi hauteur-vitesse établie pendant une campagne de mesure temporaire (cf. FIG. 9.4). Pendant cette campagne de mesure, les vitesses moyennées sur la section mouillée de la canalisation ont été mesurées avec un vélocimètre à effet Doppler. Les hauteurs ont quant à elles été mesurées avec une sonde à ultrason en suspension au dessus de la surface libre. Les résultats des mesures et la loi de régression identifiés sont présentés en FIG. 9.4 et 9.5.

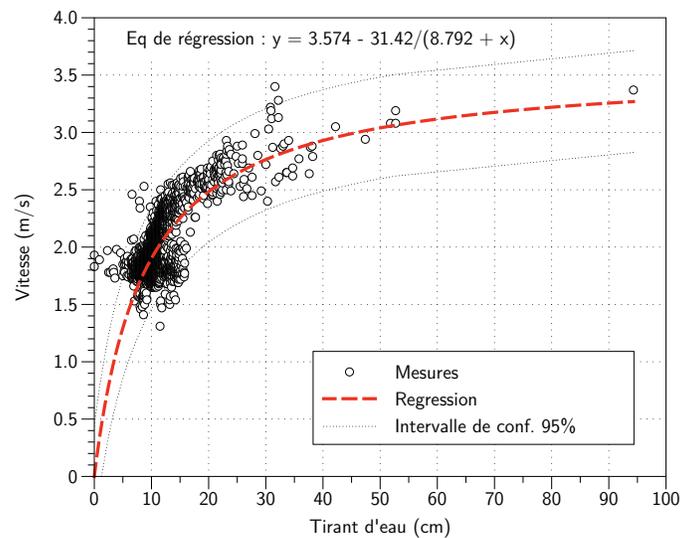


Fig. 9.4 – Loi hauteur-vitesse déterminée à l'exutoire du bassin versant

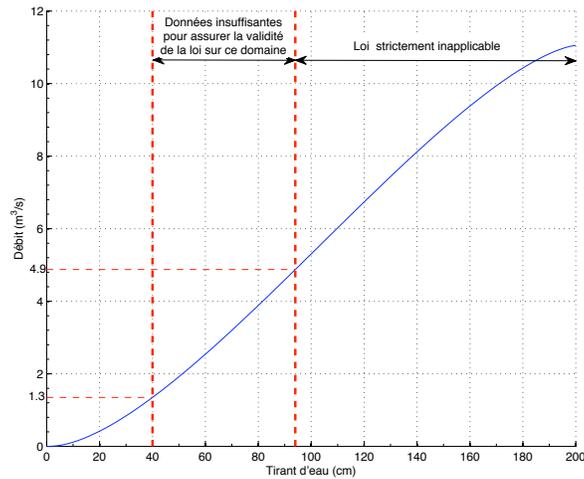


Fig. 9.5 – Loi hauteur-débit déterminée à l'exutoire du bassin versant

Trois domaines peuvent être circonscrits :

1.  $h \leq 40$  cm ( $Q < 1,35$  m<sup>3</sup>/s) : domaine de validité de la régression effectuée ;
2.  $40$  cm  $< h \leq 94$ cm ( $1,35$  m<sup>3</sup>/s  $< Q \leq 4,87$  m<sup>3</sup>/s) : seules six observations hauteur-  
vitesse sont disponibles. La loi hauteur-débit ne devrait donc pas être appliquée sur ce  
domaine ;
3.  $h > 94$ cm ( $Q > 4,87$  m<sup>3</sup>/s) : aucune donnée n'a été mesurée. La loi hauteur-débit  
est donc strictement inapplicable sur ce domaine.

Malgré l'invalidité de la loi hauteur-débit pour des débits supérieurs à 1,35 m<sup>3</sup>/s, il a été décidé d'utiliser la loi hauteur-débit sans restrictions. L'identification d'un modèle pluie-débit a été effectuée à partir de ces données entachées d'erreurs. Le domaine de validité des modèles identifiés sera par contre limité à 1,35 m<sup>3</sup>/s.

Sur la période de mesure allant du 11/07 au 10/09, en utilisant sans restrictions la loi hauteur-débit, cinq événements pluvieux sur vingt-cinq ont généré un débit maximum supérieur à 1,35

---

m<sup>3</sup>/s dont trois ont un débit supérieur à 4,87m<sup>3</sup>/s (cf. TAB. 9.2). Des débordements dans le lit majeur du cours d'eau ont par ailleurs été constatés par l'exploitant du réseau d'assainissement lors de l'événement du 31/08/08. La mesure de hauteur effectuée dans la canalisation ne prend donc pas en compte la totalité de l'écoulement.

L'événement pluvieux du 31/08/2008 ne sera donc pas pris en compte pour l'identification du modèle pluie-débit.

### **9.3 Analyses des hyétogrammes et des hydrogrammes résultants**

Une analyse des données de pluies et débits a été réalisée dans le but de repérer d'éventuelles erreurs de mesure ou des données atypiques ne correspondant pas au comportement attendu du système pluie-débit. L'analyse a consisté en un examen visuel des hyétogrammes et hydrogrammes ainsi qu'en une analyse en composantes principales (ACP).

Les hyétogrammes et hydrogrammes mesurés pendant la campagne de mesures sont tracés en FIG. 9.6. On peut observer la présence de deux pics de débit supérieurs à 1,35 m<sup>3</sup>/s et qui ne trouvent probablement pas leur origine au niveau du hyétogramme mesuré. Ces pics ont été observés le 31/07 (n°5), et le 01/09 (n°21).

Les pluies du 31/07, du 01/08 et du 03/08 ont des caractéristiques similaires en terme de lame d'eau précipitée et d'intensité maximale, mais ne génèrent pas de pics de débit du même ordre de grandeur (cf. TAB 9.2). Il en est de même pour la pluie du 01/09 dont l'intensité maximale de 16 mm/h (au niveau du poste de Kerkaredec) ne laisse pas présager d'un pic de débit aussi important (5,41 m<sup>3</sup>/s).

L'ACP (normée) permet de compléter l'examen visuel des données (cf. FIG. 9.7). Elle a été réalisée sur les données événementielles de pluie et de débit incluses dans les tableaux 9.1 et 9.2. Seule la chronique de pluie Kerkaredec a été prise en compte. La FIG. 9.7 fait ressortir les individus n° 20, 21 et 5.

L'événement n° 20 a un score élevé sur le premier axe correspondant à des valeurs élevées de débit, volume ruisselé et intensité maximum,... Rappelons qu'il ne sera pas pris en compte pour l'identification du modèle pluie-débit en raison des observations formulées au paragraphe précédent.

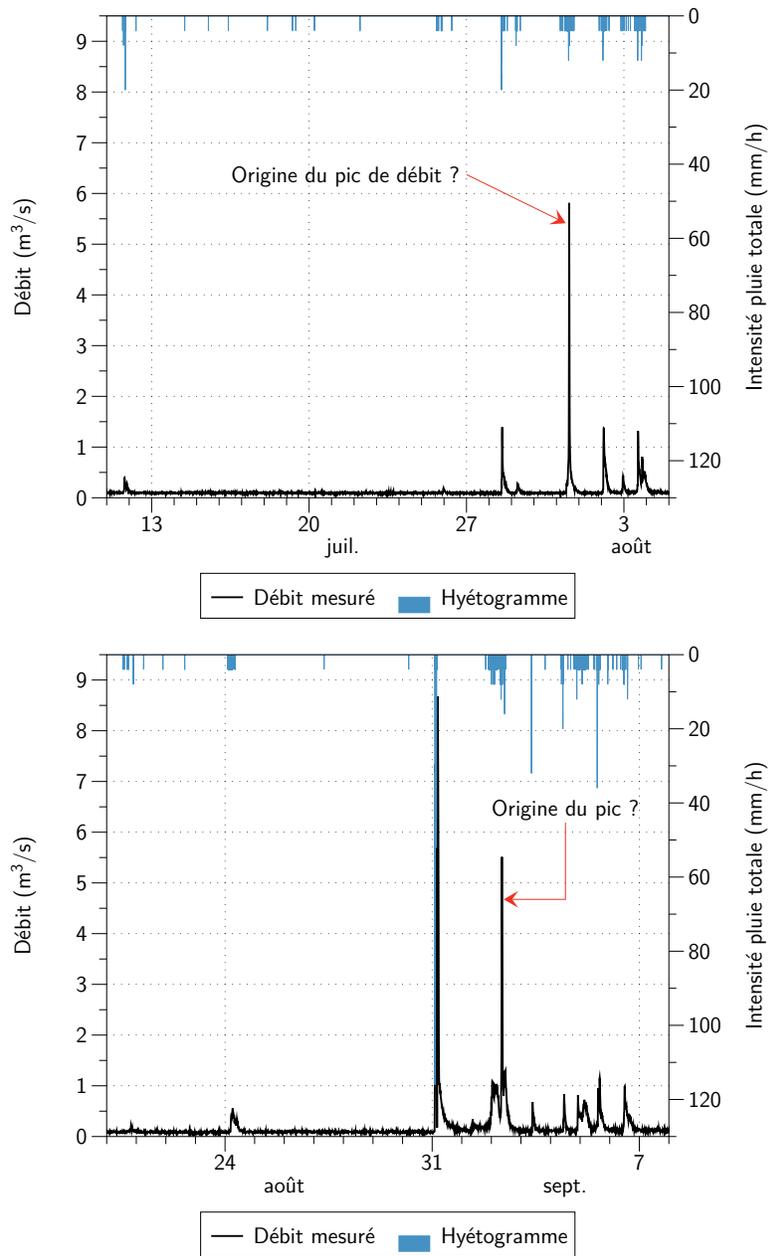


Fig. 9.6 – Hyétogrammes et hydrogrammes mesurés du 11/07 au 08/09/2008

N°	Date	Débit exutoire BV	
		V (m <sup>3</sup> )	Q <sub>max</sub> (m <sup>3</sup> /h)
1	11/07/08 14 :56	2599	0,30
2	25/07/08 13 :56	503	0,09
3	28/07/08 11 :56	7185	1,29
4	29/07/08 02 :56	1399	0,18
5	31/07/08 02 :26	21798	5,71
6	01/08/08 20 :56	10839	1,27
7	02/08/08 19 :26	2515	0,32
8	03/08/08 10 :26	15068	1,21
9	07/08/08 01 :55	4086	0,42
10	09/08/08 21 :25	1996	0,39
11	11/08/08 10 :55	8931	3,42
12	12/08/08 11 :55	6935	2,39
13	12/08/08 19 :28	2377	0,36
14	14/08/08 01 :25	343	0,11
15	16/08/08 03 :25	7221	0,50
16	17/08/08 21 :55	4918	0,49
17	18/08/08 12 :55	1535	0,21
18	19/08/08 03 :55	600	0,14
19	24/08/08 01 :25	5785	0,45
20	31/08/08 00 :54	74765	8,57
21	01/09/08 18 :45	51762	5,41
22	03/09/08 08 :15	2889	0,58
23	04/09/08 08 :24	3502	0,74
24	04/09/08 16 :09	25767	1,06
25	06/09/08 08 :39	10717	0,89

Tab. 9.2 – Volumes ruisselés mesurés à l'exutoire du bassin versant pour chacun des événements pluvieux mesurés

La FIG. 9.7 montre que les individus n° 5 et 21 sont caractérisés par des débits maximum élevés mais des intensités maximum faibles. Plusieurs hypothèses peuvent être formulées :

1. les pics sont dûs à l'invalidité de la loi hauteur-débit sur ce domaine ;
2. il existe des déversements au niveau du bassin de stockage d'eaux usées ;

---

Selon Véolia, les mesures d'autosurveillance du système d'assainissement ne révèlent aucun déversement au niveau du bassin de stockage pendant la période de mesure. Les pics de débits des événements n°5 et 21 sont donc vraisemblablement dus à l'invalidité de la loi hauteur débit au delà de  $1,35 \text{ m}^3/\text{s}^a$ .

<sup>a</sup>Il est également possible que le capteur de surverse du bassin de stockage n'ait pas fonctionné

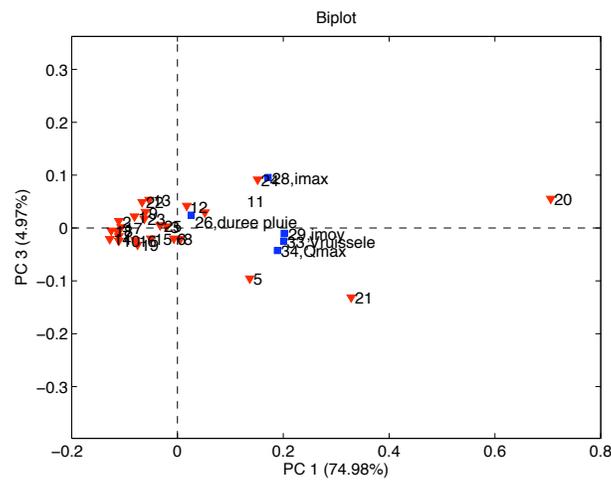


Fig. 9.7 – Analyse en composantes principales intégrant les données de pluies et de débits

## 9.4 Constitution des chroniques de calage et de validation

Le but de ce paragraphe est de définir des chroniques temporelles de calage et de validation en tenant compte de l'analyse de données effectuées précédemment. Afin d'être cohérent avec l'utilisation en continu du modèle de transformation pluie-débit, nous avons choisi de constituer des chroniques temporelles continues de calage et de validation.

La chronique temporelle de calage commence le 18 août et se termine le 10 septembre (20 jours, 7 événements pluvieux). La chronique de validation s'étend quant à elle du 11 juillet au 18 août (38 jours, 17 événements pluvieux). L'événement pluvieux n°20 a été écarté de la chronique de calage du modèle (cf. §9.2.2).

## 9.5 Résultats

Les résultats des modélisations sont synthétisés dans le tableau 9.3. Seules sont présentées les structures mathématiques de modèle qui peuvent donner lieu à une interprétation physique, *i.e.* celles dont les valeurs calculées d'aire effective et de lag-time sont réelles et positives.

n°	Type	$[n_b \ n_a \ \tau]$	Poste	PI	d	$R_{T_{cal}}^2$	YIC	$R_{T_{val}}^2$
1	SISO <sup>1</sup>	0 1 8	Hôtel C. + Kerkar.*	2,8	3,6	0,80	-10,1	0,75
2	SISO <sup>1</sup>	0 1 8	Hôtel C. + Kerkar.*	1,4	3,8	0,78	-10,0	0,74
3	SISO <sup>1</sup>	0 1 8	Hôtel C.	2,9	3,8	0,82	-10,4	0,71
4	SISO <sup>1</sup>	0 1 9	Kerkaredec	2,5	3,5	0,76	-9,7	0,68
5	SISO <sup>1</sup>	0 1 10	Pen ar Chleuz	2,6	3,2	0,73	-9,3	0,37
6	MISO <sup>2</sup>	0 1 2 1 1 8	Hôtel C. + Kerkar.*	2,8	3,4	0,81	-7,5	0,76
7	MISO <sup>2</sup>	0 1 2 0 1 8	Hôtel C. + Kerkar.*	1,4	3,8	0,81	-7,5	0,74

(1) : SISO : modèle à une entrée et une sortie – (2) : MISO : modèle à plusieurs entrées et une sortie  
 (\*) :  $I$  (mm/h) = 0,5  $I_{\text{Hôtel C.}}$  + 0,5  $I_{\text{Kerkaredec}}$

Tab. 9.3 – Meilleurs modèles pluie-débit obtenus suite aux simulations

Les meilleurs modèles sont les modèles n° 1,2, 6 et 7. Ils sont obtenus à partir d'une combinaison linéaire des chroniques de pluie Kerkaredec et Hôtel communautaire. L'utilisation des données de pluies de deux postes pluviométriques permet de moyennner les erreurs de mesure liées à l'hétérogénéité spatiale de la pluie et améliore ainsi les résultats en validation.

Tous les modèles linéaires à temps continu identifiés sont des modèles du premier ordre avec un décalage temporel compris entre huit et dix pas de temps (24 à 30 minutes). Malgré les fortes pentes du bassin versant, la réponse hydrologique n'est pas immédiate. L'influence des

---

espaces végétalisés se fait ressentir. Par ailleurs, on notera que la structure de ces modèles correspond à des réservoirs linéaires simples.

### 9.5.1 Comparaison des modèles MISO et SISO

Les modèles n° 6 et 7 comprennent deux entrées permettant de prendre en compte un décalage temporel différent pour chaque signal. Les valeurs de YIC obtenues pour ces deux modèles sont en retrait par rapport à celles obtenues avec les modèles SISO. De plus, le gain obtenu sur les valeurs de  $R^2_7$  en calage et en validation n'excédant pas 0,01, les modèles n° 1 ou 2 seront préférés.

Les hydrogrammes simulés avec les modèles n°1 (SISO) et 6 (MISO) sont présentés en FIG. 9.8 et 9.9. Les hydrogrammes simulés avec les deux modèles sont très proches. La différence est uniquement visible lors du retour au débit de base. A la vue de ces figures, l'utilisation de modèles MISO ne semble pas justifiée.

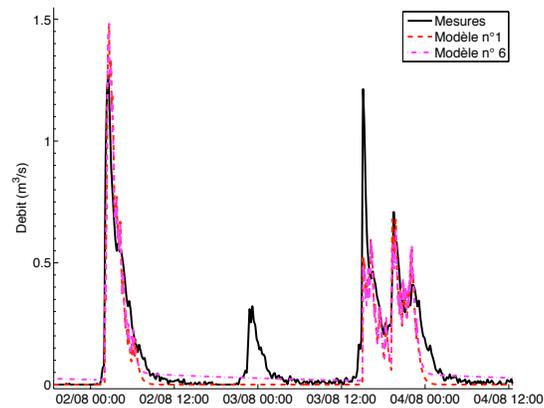


Fig. 9.8 – Débit mesuré et simulé (modèles n°1 et 6) le 02/08/2008

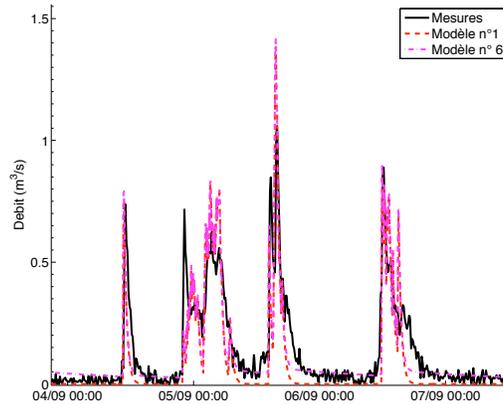


Fig. 9.9 – Débit mesuré et simulé (modèles n°1 et 6) le 04/09/2008

## 9.5.2 Valeurs des paramètres du modèle de pertes initiales

La figure 9.10 montre la pertinence du modèle de pertes initiales. Il affecte les précipitations sporadiques qui ne génèrent aucun ruissellement ainsi que le début de chaque "bloc" de précipitations. Le paramètre  $d$  du modèle permet de réinitialiser plus ou moins vite les pertes initiales.

### 9.5.2.1 Valeurs optimum des paramètres du modèle de pertes initiales

Lors du calage du modèle n°1, l'optimum de  $R_T^2$  est atteint avec le couple de paramètres ( $PI = 2,8 \text{ mm}$  ;  $d=3,6 \text{ h}$ )<sup>2</sup> (cf. FIG 9.11) . En validation, l'optimum de  $R_T^2$  est obtenu pour des valeurs de  $PI$  et  $d$  voisines ( $PI=2,4$  ;  $d=3,6$  - cf. FIG. 9.12).

<sup>2</sup>Ces résultats ont été obtenus avec une combinaison linéaire des hyétogrammes mesurés à Kerkaredec et Hôtel communautaire

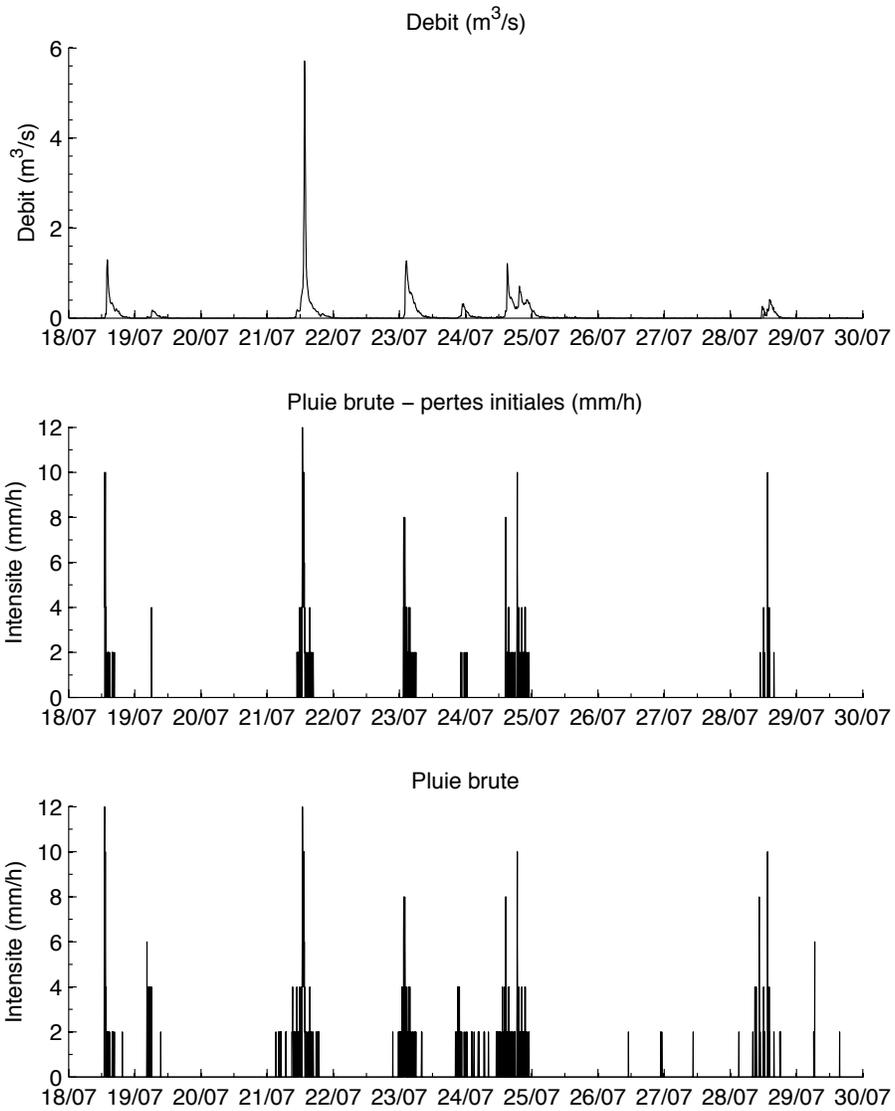


Fig. 9.10 – Hyétogrammes modifiés après application du modèle de pertes initiales

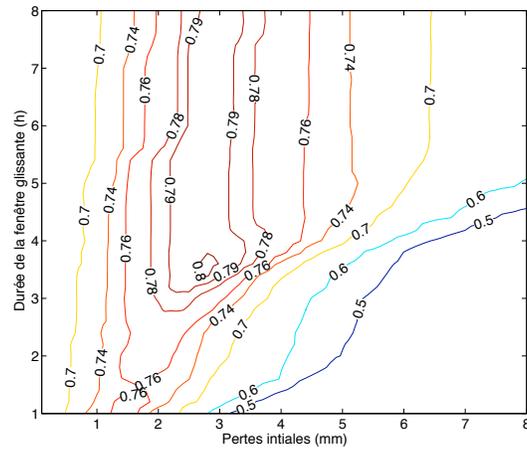


Fig. 9.11 – Évolution de  $R_T^2$  en fonction du couple de paramètres du modèle de pertes initiales en calage. (Le modèle linéaire à temps continu étant fixé)

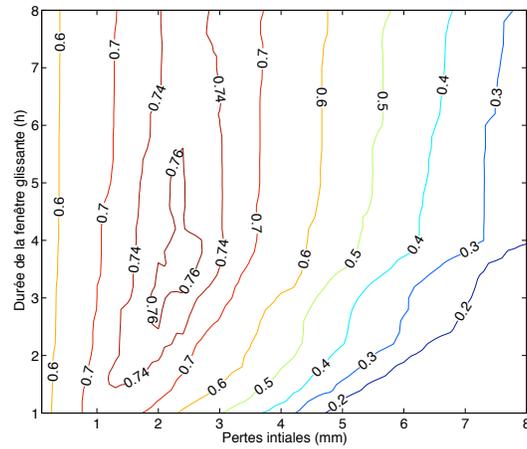


Fig. 9.12 – Évolution de  $R_T^2$  en fonction du couple de paramètres du modèle de pertes initiales en validation. (Le modèle linéaire à temps continu étant fixé)

---

### 9.5.2.2 Pertinence des valeurs calculées

Les valeurs des pertes initiales des modèles n° 1, 3, 4, 5 et 6 ont été déterminées automatiquement avec l'algorithme itératif proposé au §8.2.3.1. Les valeurs de pertes initiales des modèles n° 2 et 7 ont quant à elles été ajustées manuellement à partir de l'observation des hydrogrammes simulés et mesurés. Les FIG. 9.13, 9.14, 9.15 et 9.16 illustrent la variabilité de la valeur "vraie" des pertes initiales qui dépend fortement des conditions antécédentes à chaque événement pluvieux. Le modèle n°1 fournit les meilleures valeurs de  $R_T^2$  mais le modèle n°2 est plus performant sur quelques événements.

La FIG. 9.16 montre les hydrogrammes mesurés et simulés sur plusieurs jours consécutifs. Sur le premier événement pluvieux, le bassin versant est vraisemblablement sec, et des pertes initiales de 2,8 mm semblent adaptées. Le second événement pluvieux survient moins de dix heures après le premier. Les pertes de 1,4 mm semblent ici plus pertinentes. La FIG. 9.16 illustre donc l'importance des conditions antécédentes à chaque événement pluvieux. Elle montre par ailleurs que le modèle de pertes initiales est trop simple pour prendre en compte la variation des conditions antécédentes. Les résultats pourraient être améliorés en faisant varier le paramètre PI en fonction de la lame d'eau précipitée pendant les heures précédant l'événement pluvieux.

Les estimations des paramètres du modèle de pertes initiales calculées avec l'algorithme itératif (modèle n°1) permettent de maximiser le coefficient  $R_T^2$ . Cependant, l'observation des hydrogrammes mesurés et simulés révèle une variabilité importante de la valeur vraie de pertes initiales. Ainsi, le modèle n°2 présentant des valeurs de pertes initiales inférieures s'avère plus pertinent quelques événements pluvieux.

Une représentation optimale des pertes initiales nécessiterait donc de faire varier le paramètre PI en fonction de la lame d'eau précipitée pendant les heures précédant l'événement pluvieux.

### 9.5.3 Équivalence avec les modèles de réservoirs et interprétation physique

Les valeurs de lag-time et de surface actives ont été calculées pour chaque modèle identifié (cf. TAB. 9.4). Le coefficient de ruissellement obtenu avec l'identification du modèle n° 1 est égal

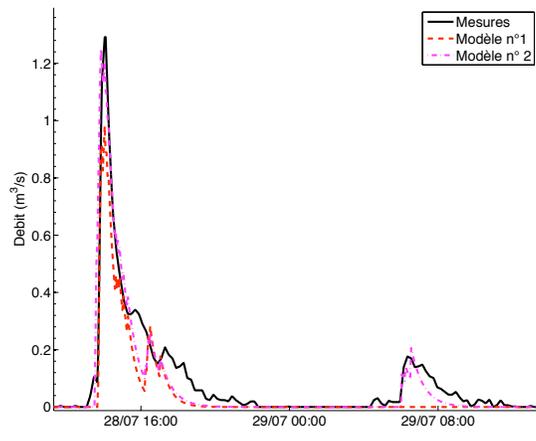


Fig. 9.13 – Hydrogrammes mesurés et simulés (modèles n° 1 et 2) le 28/07/2008 (Validation)

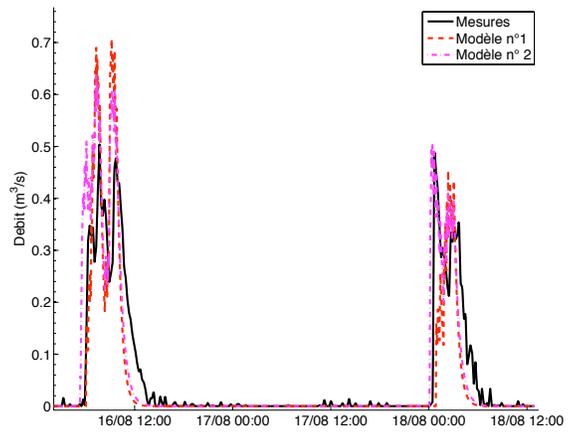


Fig. 9.14 – Hydrogrammes mesurés et simulés (modèles n° 1 et 2) le 16/08/2008 (Validation)

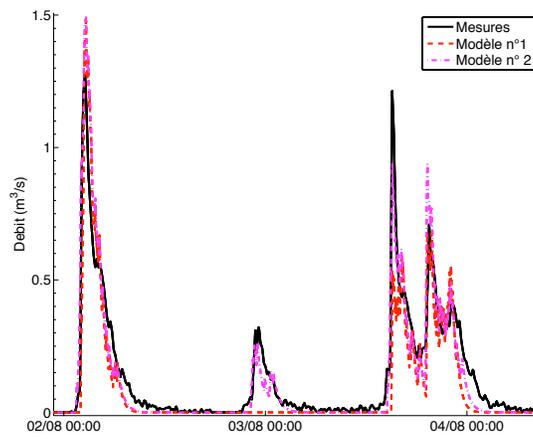


Fig. 9.15 – Hydrogrammes mesurés et simulés (modèles n° 1 et 2) le 02/08/2008 (Validation)

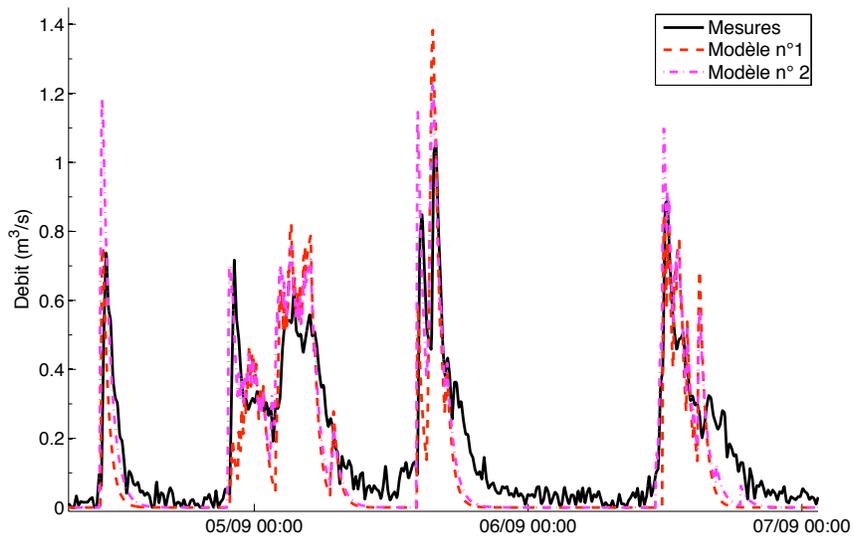


Fig. 9.16 – Hydrogrammes mesurés et simulés (modèles n° 1 et 2) le 05/09/2008 (Calage)

à 26,6%. Les valeurs de coefficient de ruissellement ont également été calculées en effectuant une régression linéaire à partir des volumes ruisselés et des volumes totaux (cf. TAB. 9.5). Le coefficient de ruissellement moyen des séries Kerkaredec et Hôtel communautaire est égal 26,7%. L'estimation réalisée avec l'algorithme SRIVC est donc pertinente.

Modèle n°1	Modèle n° 2	Modèle n° 7
$\begin{cases} K = 31 \text{ min} \\ Sa = 154,3 \text{ ha} \end{cases}$	$\begin{cases} K = 40 \text{ min} \\ Sa = 144,3 \text{ ha} \end{cases}$	$\begin{cases} K_1 = 573 \text{ min} ; K_2 = 34 \text{ min} \\ Sa_1 = 46,8 \text{ ha} ; Sa_2 = 122,7 \text{ ha} \end{cases}$

Tab. 9.4 – Valeurs de lag-time et de surface active calculées avec les modèles SISO et MISO

Poste	Paramètre	moyenne $\pm$ IC <sub>95%</sub>
Kerkaredec	$C_R$	0,277 $\pm$ 0,023
	PI (mm)	1,1 $\pm$ 0,5
Hôtel communautaire	$C_R$	0,258 $\pm$ 0,021
	PI (mm)	1,0 $\pm$ 0,5
Pen ar Chleuz	$C_R$	0,239 $\pm$ 0,040
	PI (mm)	0,7 $\pm$ 1,1

Tab. 9.5 – Valeurs moyennes du coefficient de ruissellement et des pertes initiales calculés par régression linéaire à partir de chaque série de pluie

Les valeurs de lag-time estimées des modèle n° 1 et 2 sont respectivement de 31 et 40 min. Elles sont cohérentes par rapport à la valeur calculée avec la formule de Desbordes (Desbordes, 1974) (20 à 25 min suivant la lame d'eau prise en compte). La procédure d'identification permet donc d'évaluer facilement une valeur de lag-time optimisée à partir des données expérimentales. Cependant, les valeurs calculées sont inférieures au décalage temporel moyen entre les centres de gravité des hyétogrammes et des hydrogrammes (119 min). Elles sont également inférieures au décalage temporel moyen entre les intensités et débits maximum (60 min). On remarquera par ailleurs, que les retours au débit de base sont systématiquement mal reproduits et sont toujours plus rapides que ceux réellement observés. Ces deux points laissent penser que les valeurs de lag-time calculées avec l'algorithme SRIVC sont sous-estimées. Les valeurs de lag-time et de surface active calculées correspondent au

---

meilleur compromis permettant de respecter la dynamique du signal de débit et de minimiser les erreurs de prédiction.

De plus, le coefficient de ruissellement global du modèle MISO n° 7 vaut 29 % et semble surestimé à la vue des coefficients de ruissellements calculés par régression linéaire. Enfin les valeurs de pertes initiales des modèles n° 2 et 7 sont plus proches des valeurs de pertes initiales obtenues par régression linéaire. Des pertes initiales de 2,8 mm sont donc vraisemblablement irréalistes et correspondraient uniquement à des valeurs plausibles lorsque le bassin versant est sec.

La phase d'interprétation physique du modèle met en exergue la sensibilité des méthodes d'identification à l'inclusion de données non validées. L'estimation paramétrique a pour objectif de maximiser  $R_7^2$ . Or, ce paramètre tient particulièrement compte des valeurs élevées de débits pour lesquelles les incertitudes sont les plus importantes.

Il est donc vraisemblable que les valeurs de lag-time estimées soient biaisées. En effet, des faibles valeurs de lag-time, comme celles obtenues avec les modèles n°1 et 2 ont comme conséquence, à valeur de coefficient de ruissellement constante, d'augmenter l'estimation du débit de pointe.

Les débits de pointe étant entachés d'erreurs, il convient de trouver une méthode permettant d'améliorer la prédiction des débits dans le domaine de validité retenu ( $Q < 1,35 \text{ m}^3/\text{s}$ ).

#### **9.5.4 Estimation de valeurs optimisées pour le domaine de validité du modèle**

Même si les faibles valeurs de lag-time identifiées ont pour conséquence de diminuer les erreurs de prédictions sur les débits de pointe, celles-ci restent très importantes (cf. FIG. 9.17). Elles atteignent plus de  $3 \text{ m}^3/\text{s}$  pour les événements pluvieux du 31/07 et du 01/09/2008 (cf. FIG. 9.17).

L'inclusion de l'hydrogramme de l'événement n° 21 au sein de la chronique de calage déséquilibré l'identification et aboutit à des valeurs de lag-time et de surface active vraisemblablement biaisées. Pour s'affranchir de ce problème, trois solutions sont possibles :

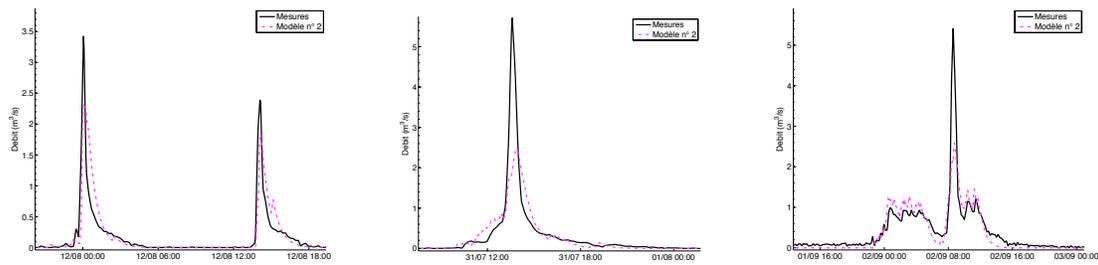


Fig. 9.17 – Hydrogrammes mesurés et simulés (modèle n° 2) le 12/08, 31/07 et 01/09

1. Calage du modèle sur une chronique continue n'incluant pas les événements n° 5 et 21. Cette solution est sécuritaire mais exclut complètement deux événements pluvieux. Elle est pénalisante étant donné la taille réduite de notre jeu de données.
2. Limitation des débits de l'hydrogramme de l'événement n°21 à  $1,35 \text{ m}^3/\text{s}$ . Cette solution présente l'avantage de ne modifier qu'une toute petite partie des données et de conserver dans l'état la chronique de calage.
3. Identification d'un modèle à plusieurs entrées prenant en compte les pics de débits correspondant à des déversements possibles d'eaux usées ou des erreurs de mesure. Cette solution est plus complexe à mettre en oeuvre. Toutefois, dans l'hypothèse où les déversements d'eaux usées seraient avérés, elle présente l'avantage de pouvoir prédire à l'avenir les futurs déversements.

Nous avons choisi de mettre en oeuvre la troisième solution. La nouvelle identification de modèle fait l'hypothèse que les pics de débit des événements n° 5 et 21 (et éventuellement 11 et 12) sont dûs soit à des erreurs de mesure, soit à des déversements d'eaux usées au niveau du bassin de stockage. Nous avons donc cherché à identifier un modèle MISO à deux entrées :

- La première composante du modèle correspond aux apports pluviaux (réseau d'assainissement pluvial + apports directs). L'entrée du modèle  $u_1(t)$  correspond à la pluie totale défalquée des pertes initiales.
- La seconde composante du modèle intègre les pics de débit correspondant soit à des déversements d'eaux usées, soit à des erreurs de mesure. Afin, de simuler des éventuels déversements ou des erreurs de mesure associées aux débits élevés, il convient de trouver

un signal d'entrée représentatif de cette dynamique.

Nous proposons une seconde entrée  $u_2(t)$  définie comme suit :

$$\begin{cases} \text{si } \hat{Q}_{ruis}(t) - Q_{vid} > 0 \\ \text{alors : } u_2(t) = (\hat{Q}_{ruis}(t) - Q_{vid}) \\ \text{sinon : } u_2(t) = 0 \end{cases} \quad (9.1)$$

où  $\hat{Q}_{ruis}(t)$  est le débit ruisselé estimé avec le modèle linéaire utilisé dans les paragraphes précédents.  $Q_{vid}$  représente un débit de vidange. Si le débit ruisselé est inférieur au débit de vidange alors il n'y aura pas de déversements. Si le débit ruisselé est supérieur au débit de vidange, un déversement dans le Stang Alar sera simulé. Le débit de vidange est un hyper-paramètre qu'il convient d'ajuster au moyen d'un algorithme itératif. Une première estimation de ce débit pourra être donnée visuellement en comparant les hydrogrammes mesurés et simulés obtenus au paragraphe précédent. En effet, le but est de retrancher au débit simulé un débit constant, de telle sorte que l'hydrogramme résultant ne présente que des pics pour les événements où le débit de pointe est largement sous-estimé.

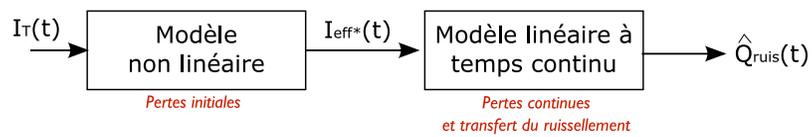
Une représentation schématique de la procédure d'identification du modèle MISO est donnée en FIG. 9.18. Les résultats obtenus sont récapitulés dans le TAB. 9.6.

$[n_b \ n_a \ \tau]$	Poste	PI	d	$Q_{vid}$	$R^2_{Tcal}$	YIC	$R^2_{Tval}$
0 1 6 - 0 1 0	2 postes pluvio.	1,4	3,8	1,5	0,95	-8,32	0,87

Tab. 9.6 – Estimation des paramètres du modèle MISO dont une entrée prend en compte les déversements ou les erreurs de mesure sur les pics de débits élevés

Les valeurs de  $R^2_T$  obtenues lors du calage et de la validation sont excellentes. Elles illustrent que le modèle reproduit cette fois beaucoup plus fidèlement la dynamique du signal de débit. La FIG. 9.5.4 montre les résultats obtenus sur les événements pluvieux n° 5 et 21. La FIG. 9.5.4 permet de visualiser la contribution de chaque composante du modèle sur ces deux événements.

(a) : simulation d'un modèle pluie-débit - Générer  $I_{eff}^*(t)$  et  $\hat{Q}_{ruis}(t)$



(b) : identifier un modèle MISO à 2 entrées

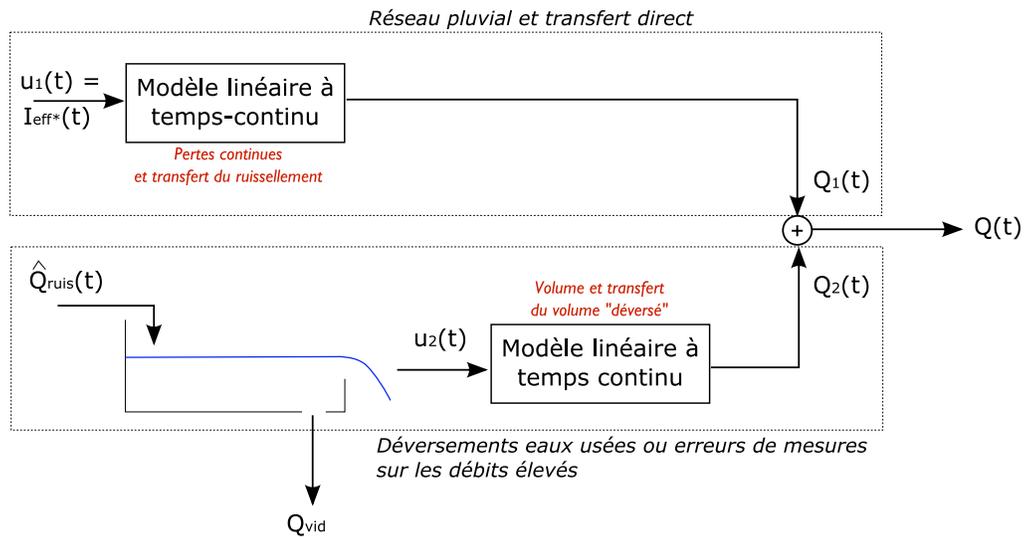


Fig. 9.18 – Interprétation schématique du modèle pluie-débit à deux entrées

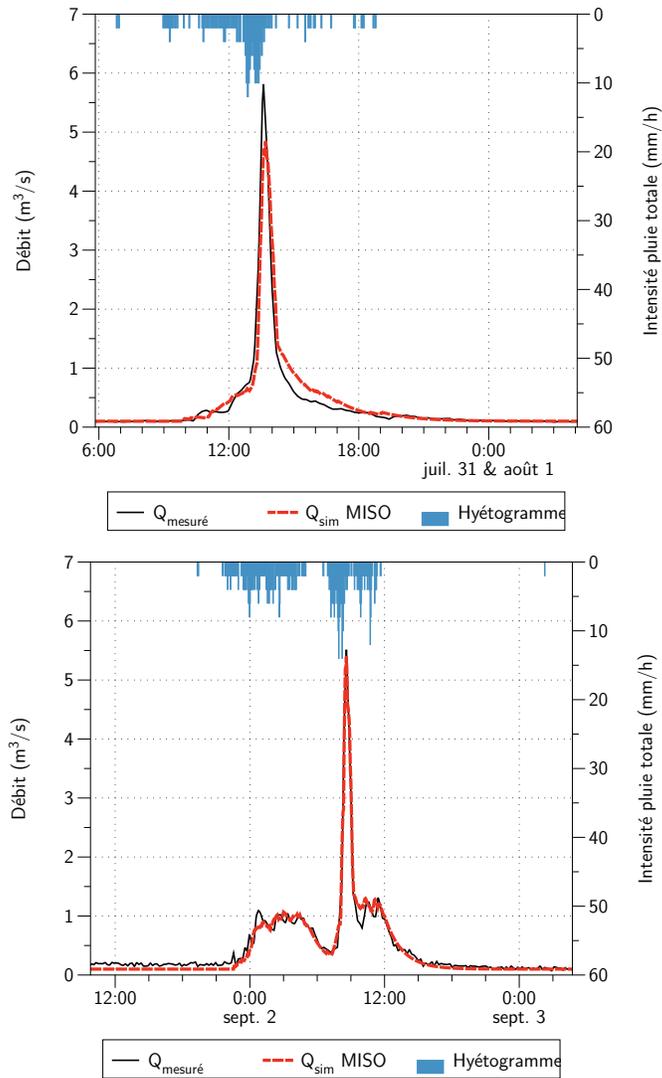


Fig. 9.19 – Hydrogrammes mesurés et simulés avec le modèle MISO (31/07 et 01/09/2008)

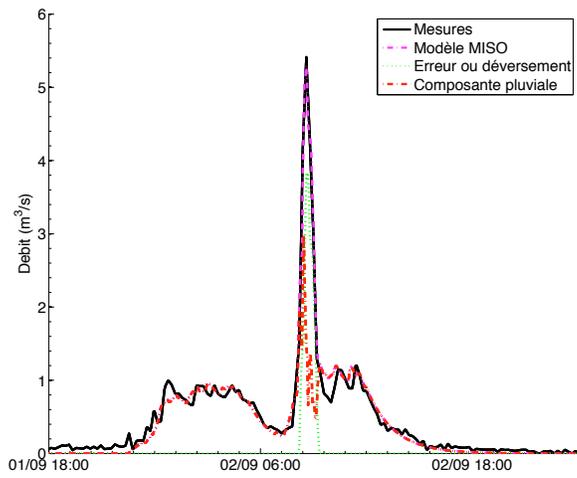
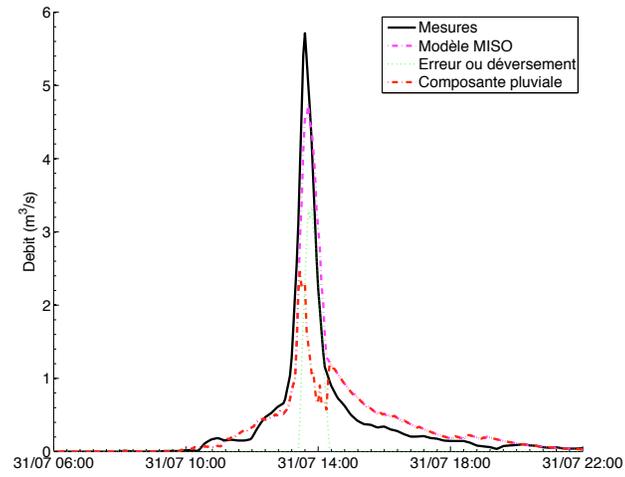


Fig. 9.20 – Décomposition des hydrogrammes simulés le 31/07 et 01/09/2008 en deux composantes

---

La valeur de lag-time de la composante pluviale est dorénavant de 91 min et correspond davantage aux décalages temporels observés entre centre de gravité des hyétogrammes et des hydrogrammes. Le coefficient de ruissellement correspondant à la composante pluviale du modèle vaut 22,2%, ce qui est tout à fait cohérent. En effet, le coefficient de ruissellement calculé en enlevant les événements n°5 et 21 (pour lesquels un déversement d'eaux usées est plausible) est de 23%.

La procédure d'identification d'un modèle MISO permet de décomposer l'hydrogramme simulé en deux composantes. La première correspond à la dynamique vérifiée du système. La seconde correspond soit à des déversements qui n'ont pas pu être mesurés soit à des erreurs de mesures.

Si les déversements sont avérés, les deux composantes du modèle pourront être utilisées, sous réserve de la validation ultérieure de la loi hauteur-débit sur l'intégralité de la plage de débit observable.

Si les campagnes de mesures ultérieures montrent au contraire que les pics de débits sont imputables à des erreurs de mesure, seule la première composante du modèle pourra être utilisée.

## 9.6 Conclusion

Ce chapitre avait pour objectif l'identification d'un modèle pluie-débit global sur le bassin versant du Stang Alar permettant de simuler en continu les débits rejetés dans le milieu marin.

L'objectif n'est que partiellement atteint. En effet, l'application de la procédure d'identification a été entravée par des problèmes expérimentaux concernant l'acquisition des débits. Ceux-ci sont mesurés sur la base d'une loi hauteur-débit établie pendant une campagne de mesure temporaire. Or, les observations correspondant à des débits supérieurs à 1,35 m<sup>3</sup>/s sont trop peu nombreuses pour valider cette loi au delà de ce seuil de débit. Par ailleurs, l'analyse des hyétogrammes et hydrogrammes fait ressortir des pics de débits (> à 4 m<sup>3</sup>/s) ne pouvant être expliqués que par des erreurs de mesure ou des déversements d'eaux usées au niveau d'un bassin situé dans la partie centrale du bassin versant.

---

Malgré ce problème, la pertinence de l'approche de modélisation et l'intérêt des techniques d'identification de modèles à temps continu ont tout de même pu être mis en avant. Un modèle a pu être identifié au sein du domaine de validité restreint de débits.

L'application de la procédure d'identification de modèle pluie-débit sur le bassin versant du Stang Alar révèle ainsi l'importance capitale de la phase expérimentale d'acquisition de données. Le modèle pluie-débit ne pouvant être validé que pour une plage limitée de débits, nous préconisons de recommencer la procédure entière d'identification en apportant une attention particulière à la mesure des débits.

# Conclusion générale et perspectives

Les objectifs de cette thèse étaient de tester l'applicabilité :

- de la spectrophotométrie UV/Visible :
  - pour mesurer en continu les concentrations en DCO et MES en système d'assainissement sur de longues séries temporelles incluant une grande diversité de conditions environnementales,
  - pour estimer en continu la concentration en E. coli dans des eaux de rivière dont la qualité est influencée par des rejets d'assainissement.
- de techniques d'identification de systèmes pour développer un modèle pluie-débit simplifié en bassin versant mixte rural-urbain.

## Mesures en continu de concentrations en DCO et MES en système d'assainissement

L'étalonnage de modèles PLS pour la prédiction de concentrations en DCO et MES par spectrophotométrie UV/Visible a été réalisé. Il a été montré que les performances des modèles PLS peuvent être améliorées par une sélection optimisée de longueurs d'onde. Trois algorithmes de sélection ont été testés : l'algorithme OPPPLS créé par Torres et Bertrand-Krajewski (2008), une nouvelle adaptation de l'algorithme interval PLS, et l'algorithme moving window PLS. Les trois méthodes permettent d'établir des modèles PLS aux performances équivalentes, sauf dans le cas de la prédiction de la DCO où les modèles mwPLS sont en retrait. De plus, les modèles fbiPLS et mwPLS sont plus parcimonieux que les modèles OPPPLS et constituent

---

un avantage pour un suivi en continu des concentrations. La parcimonie des modèles identifiés est liée à l'exclusion d'un plus grand nombre de longueurs d'onde ne concourant pas directement à la prédiction du paramètre de pollution.

Il a également été montré qu'une collection d'étalonnage réduite ne permet pas de prendre en compte toutes les variations possibles de la matrice des eaux usées. Les modèles PLS établis à partir de ces données ne peuvent donc pas être utilisés sur de longues séries temporelles incluant des conditions environnementales variées. Afin de circonscrire le domaine d'utilisation des modèles et de compléter de manière adéquate la collection d'étalonnage, nous avons proposé une démarche qualitative de classification des spectres d'absorbance. Cette démarche rend possible l'identification sur des mesures in-situ de spectres dont les caractéristiques sont sous-représentées dans les données d'étalonnage.

Les résultats obtenus ouvrent de nouvelles perspectives pour la mesure en continu des paramètres de pollution globaux, et plus particulièrement pour la DCO. La classification qualitative des spectres d'absorbance pourra être utilisée afin de compléter les collections d'étalonnage et pour établir des modèles prédictifs adaptés à chaque classe de spectre. On notera toutefois que cette démarche est très lourde à mettre en oeuvre. Elle semble donc difficilement applicable pour une majorité de collectivités.

## **Projet GIRAC**

### **Estimation de la teneur en E. coli d'eaux de rivières**

La possibilité d'utiliser la spectrophotométrie UV/Visible pour effectuer des prédictions de concentrations en E. coli sur des eaux de rivière a été explorée. La recherche de corrélations a été entravée par de nombreux problèmes expérimentaux. Le premier d'entre eux est la dérive du signal d'absorbance, essentiellement dans le domaine UV, imputable à la conception de la cuve de mesure et à un nettoyage insuffisant. Cette dérive ne rend pas possible l'étude de corrélations entre E. coli et absorbances dans l'UV sur de longues séries temporelles. Le deuxième problème est le prélèvement systématique manuel des échantillons aux mêmes tranches horaires, rendant ainsi impossible l'étude de la dynamique horaire des concentrations par temps sec. Le troisième problème concerne des erreurs d'horodatage d'échantillons constatées en 2008 et en 2009.

---

Malgré toutes ces difficultés, l'augmentation concomitante des valeurs d'absorbance et de teneurs en *E. coli* par temps de pluie a été mise en exergue. L'insuffisance de données correspondant à des concentrations et valeurs d'absorbance élevées ne permet pas d'établir une relation fiable et utilisable par temps de pluie. En temps sec, des campagnes de prélèvements horaires ont révélé la plausibilité de l'existence d'un profil journalier de concentration en *E. coli*. Des modélisations effectuées sur l'ensemble de la saison de baignade 2009 confirment cette hypothèse.

Des conclusions définitives ne peuvent donc pas être tirées au bout de trois années de mesures. Par conséquent, des propositions de modification de protocole expérimental ont été exprimées afin d'améliorer le déroulement futur de cette étude. Nous proposons en premier lieu de réaliser la phase d'étalonnage d'un modèle prédictif hors-ligne afin de réduire les erreurs de mesure spectrophotométriques et les erreurs d'horodatage. Les mesures spectrophotométriques pourraient ainsi être réalisées en triplicat et permettraient de garantir l'exacte similarité avec l'échantillon utilisé pour le dénombrement de *E. coli*. De plus, l'organisation des prélèvements devra être modifiée de manière à étudier la variabilité horaire des concentrations en *E. coli* et à compléter les données avec de fortes concentrations en *E. coli*. Une fois le modèle prédictif établi, nous proposons de modifier le système de mesure de manière à limiter les possibilités de dérive des signaux d'absorbance. Des mesures de contrôle devront également être régulièrement effectuées afin de vérifier l'inexistence d'une dérive ou d'un dysfonctionnement du spectrophotomètre.

### **Développement d'un modèle pluie-débit**

Parallèlement à l'estimation des concentrations en *E. coli* par spectrophotométrie, un modèle pluie-débit a été développé en proposant l'utilisation de techniques d'identification de modèle à temps-continu et de l'approche *data-based mechanistic modelling*.

Le modèle pluie-débit couple un modèle conceptuel de pertes initiales à un modèle linéaire à temps continu regroupant les pertes continues et le transfert du ruissellement. Le choix de cette structure de modèle résulte de la volonté d'interpréter la transformation pluie-débit avec les approches habituellement utilisées en hydrologie urbaine. En effet, les modèles linéaires à temps continu ont des formulations mathématiques voisines des modèles conceptuels de réservoir. L'interprétation de ces modèles en sera donc facilitée.

---

Les techniques d'identification choisies permettent de formaliser le problème mathématique d'estimation paramétrique d'une manière à la fois *élégante* et performante. Elles présentent également l'avantage de ne pas fixer *a priori* la structure mathématique de modèle à identifier. Le choix de la structure est effectué *a posteriori* sur la base des valeurs de deux critères statistiques indicateurs de la performance et de la parcimonie du modèle.

La pertinence de cette approche a été démontrée par des simulations réalisées sur des chroniques de pluie et de débit fictives. Elle a ensuite été appliquée sur le bassin versant du Stang Alar avec des données de pluie et de débit mesurées pendant la saison de baignade 2008. Les chroniques expérimentales de débits ne peuvent être validées pour des débits supérieurs à 1,35 m<sup>3</sup>/s par manque de données. Un modèle pluie-débit a tout de même pu être identifié sur un domaine de validité des débits restreint, montrant ainsi l'intérêt des techniques d'identification utilisées.

L'analyse des hyétogrammes et des hydrogrammes a par ailleurs révélé la présence de quelques pics de débits très élevés (> 4 m<sup>3</sup>/s) imputables aux erreurs de mesure sur les débits ou à des déversement d'eaux usées au niveau d'un bassin de stockage. L'étude doit donc être complétée par une nouvelle campagne de mesures de hauteurs et de vitesses permettant de fiabiliser la mesure des débits et valider ou non l'hypothèse de déversements d'eaux usées dans le Stang Alar.

Enfin, d'un point de vue méthodologique, il pourrait être intéressant de simuler la transformation pluie-débit avec un modèle linéaire à paramètres variants. Ces modèles ont connu de récents développements et ont été appliqués avec succès pour la modélisation pluie-débit de bassins-versants viticoles (Laurain, 2010). Le choix de cette structure permettrait d'identifier un modèle intégrant l'ensemble de la transformation pluie-débit, tout en conservant une possibilité d'interprétation hydrologique du modèle. De plus, son intérêt réside dans la prise en compte des non linéarités du système et de l'évolution des paramètres d'état du bassin versant au cours du temps (tels que la saturation des sols, la rétention d'eau dans les dépressions du sol, ... ).

Cette thèse a donc montré l'intérêt de la spectrophotométrie UV/Visible et des techniques d'identification de systèmes pour l'évaluation en continu des flux de pollution rejetés dans les milieux récepteurs. Par ailleurs, elle présente une unité thématique au niveau des outils mathématiques utilisés. En effet, les modèles prédictifs de concentrations et les modèles pluie-

---

débit sont tous deux des modèles statistiques identifiés à partir de données expérimentales. L'intérêt de la thèse nous a donc sembler résider également dans l'appropriation de méthodes statistiques d'identification dans le champ de l'assainissement.

On notera cependant que, contrairement aux modèles directement issus de lois de la physique, l'établissement de tels modèles requiert un nombre très important de données afin qu'ils puissent être utilisés sur de longues séries temporelles, incluant des conditions environnementales variées. L'utilisation de chroniques d'étalonnage restreintes nécessite alors de préciser leurs conditions d'utilisation.

# Bibliographie

- C. Abrahamsson, J. Johansson, A. Sparen et F. Lindgren : Comparison of different variable selection methods conducted on nir transmission measurements on intact tablets. *Chemo-metrics and Intelligent Laboratory Systems*, 69:3–12, 11 2003.
- AFFSET : Classement de la qualité des eaux de baignade à l'échelon national par la méthode de la nouvelle directive européenne 2006/7/ce. Rap. tech., 9 2007a.
- AFFSET : Valeurs seuils échantillon unique pour les eaux de baignade : étude de faisabilité méthodologique. Rap. tech., 2007b.
- S. A. Anderson, S. J. Turner et G. D. Lewis : Enterococci in the new zealand environment : Implications for water quality monitoring. *Water Science and Technology*, 35(11-12):325–331, 1997.
- ATV : Arbeitsblatt a 118 : Richtlinien für die hydraulische berechnung von schmutz, regen und mischwasserkanalen. *Gesellschaft zur Förderung der Abwassertechnik (GFA)*, 1977.
- N. Azema, M. F. Pouet, C. Berho et O. Thomas : Wastewater suspended solids study by optical methods. *Colloids and Surfaces A : Physicochemical and Engineering Aspects*, 204 (1-3):131–140, 2002.
- Y. Azzout, S. Barraud, F. N. Cres et E. ALFAKIH : Techniques alternatives en assainissement pluvial : Choix, conception, réalisation et entretien. *Ed. Tec & Doc*, 1994.
- S. Barraud, J.-L. L. Bertrand-Krajewski, G. Lipeme Kouyi, P. Breil et P. Namour : Les flux de polluants dans le bassin du rhône : leur caractérisation dans différents contextes géographiques et fonctionnels. *4ème journée thématique de la ZABR*, 10 p., 2008.

- 
- G. Bastin, L. Moens et P. Dierickx : On-line river ow forecasting with hydromax : successes and challenges after twelve years of experience. *In 15th IFAC Symposium on System Identification Saint-Malo, France, July 6-8, 2009*, 2009.
- D. Baudiová : Evaluation of escherichia coli as the main indicator of faecal pollution. *Water Science and Technology*, 35(11-12):333 – 336, 1997.
- D. Bertrand : Etalonnage multidimensionnel : application aux données spectrales. *Techniques de l'ingénieur*, vol. SL1(P264):P264.1–P264.21, 2005.
- J. L. Bertrand-Krajewski, D. Laplace, C. Joannis et G. Chebbo : *Mesures en hydrologie urbaine et assainissement*. Ed. Tec & Doc, 2000.
- J.-L. L. Bertrand-Krajewski : *Modélisation conceptuelle du transport solide en réseau d'assainissement*. Thèse de doctorat, Université Louis Pasteur, Strasbourg, p. 206 p., 1992.
- J. L. Bertrand-Krajewski, G. Chebbo et A. Saget : Distribution of pollutant mass vs volume in stormwater discharges and the first flush phenomenon. *Water Research*, 32(8):2341–2356, 1998.
- K. J. Beven : *Rainfall-Runoff Modelling : The Primer*. John Wiley & Sons, Chichester, 2003.
- G. Bitton : *Wastewater microbiology*. John Wiley & Sons, Inc, 2005.
- A. B. Boehm, R. L. Whitman, M. B. Nevers, D. Hou et S. B. Weisberg : Now-casting recreational water. *in Statistical Framework for Recreational Water Quality Criteria and Monitoring (ed L. J. Wymer)*, John Wiley & Sons, Ltd, Chichester, UK., 2007.
- A. B. Boehm et S. B. Weisberg : Tidal forcing of enterococci at marine recreational beaches at fortnightly and semidiurnal frequencies. *Environmental Science and Technology*, 39 (15):5575–5583, 2005.
- E. Brelot et B. Chocat : Impact des rejets sur les milieux récepteurs. *La Houille Blanche*, 1-2:16–21, 1996.
- S. K. E. Brookman : Estimation of biochemical oxygen demand in slurry and effluents using ultra-violet spectrophotometry. *Water Research*, 31(2):372–374, 1997.
- E. H. Bryan : Quality of stormwater drainage from urban land. *Journal of the American Water Resources Association*, 8(3):578–588, 1972.

- 
- I. J. Buerge, H. R. Buser, M. D. Muller et T. Poiger : Behavior of the polycyclic musks hhcb and ahtn in lakes, two potential anthropogenic markers for domestic wastewater in surface waters. *Environmental Modelling and Software*, 37(24):5636–5644, 2003.
- R. J. Burm, D. F. Krawczyk et G. L. Harlow : Chemical and physical comparison of combined and separate sewer discharges. *Journal of the Water Pollution Control Federation*, 40 (1):112–126, 1968.
- V. J. Cabelli : Health effects criteria for marine recreational waters. *EPA Document Number EPA-600/1-80-031, Health Effects Research Laboratory, Office of Research and Development, United States Environmental Protection Agency, Research Triangle Park, North Carolina, USA*, 1983.
- V. J. Cabelli : Swimming-associated illness and recreational water quality criteria. *Water Science and Technology*, 21(2):13–21, 1989.
- V. J. Cabelli, A. P. Dufour, L. J. McCabe et M. Levin : Swimming-associated gastroenteritis and water quality. *American journal of epidemiology*, 115(4):606, 1982.
- Y. Cao, J. F. Griffith et S. B. Weisberg : Evaluation of optical brightener photodecay characteristics for detection of human fecal contamination. *Water Research*, 43(8):2273 – 2279, 2009.
- CERTU : La ville et son assainissement : principes, méthodes et outils pour une meilleure intégration dans le cycle de l'eau. *Guide technique*, 503 p., 2003.
- G. Chebbo : *Solides des rejets urbains par temps de pluie : caractérisation et traitabilité*. Thèse de doctorat, ENPC Paris, 410 p., 1992.
- G. Chebbo et M. C. Gromaire : The experimental urban catchment ['le marais' in paris : what lessons can be learned from it? *Journal of Hydrology*, 299(3-4):312 – 323, 2004.
- G. Chebbo, J. M. Mouchel, A. Saget et M. Gousailles : La pollution des rejets urbains par temps de pluie : flux, nature et impacts. *Technique - Sciences - Méthodes*, 11:796–806, 1995.
- B. Chocat, J. L. Bertrand-Krajewski et S. Barraud : Eaux pluviales urbaines et rejets urbains par temps de pluie. *Techniques de l'ingénieur. Technologies de l'eau*, 2(W6800), 2007.

- 
- B. Chocat : *Un système d'aide à la gestion, la prévision et la conception des équipements d'assainissement*. Thèse de doctorat, INSA Lyon and University Claude Bernard of Lyon (France), 131 p., 1981.
- R. Colas : Le problème de l'eau. *Population*, 19eme annee(1):31–54, 1964.
- R. R. Colwell, P. R. Brayton, D. J. Grimes, D. B. Roszak, S. A. Huq et L. M. Palmer : Viable but non-culturable vibrio cholerae and related pathogens in the environment : implications for release of genetically engineered microorganisms. *Nature Biotechnology*, 3(9):817–820, 1985.
- S. de Jong : Simpls : an alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18(3):251–263, 1993.
- M. Desbordes : *Réflexions sur les méthodes de calcul des réseaux urbains d'assainissement pluvial*. Thèse de doctorat, USTL, Montpellier, 224 p., 1974.
- M. Desbordes, A. Bachoc, J. P. Tabuchi, G. Chebbo et J. P. Philippe : La pollution des rejets urbains par temps de pluie : quantité, origine et nature. *La Houille Blanche*, 1-2:21–33, 1994.
- J. C. Deutsch et J. C. Hemain : Main results of the french national programme of urban runoff quality measurement. In *Proceedings of the 3rd International Conference on Urban Storm Drainage*, Chalmers University, Gothenburg, Sweden, p. 939–946, 1984.
- R. A. Dobbs, R. H. Wise et R. B. Dean : The use of ultra-violet absorbance for monitoring the total organic carbon content of water and wastewater. *Water Research*, 6(10):1173 – 1180, 1972.
- A. P. Dufour : Health effects criteria for fresh recreational waters. *EPA Document Number EPA-600/1-84-004. Health Effects Research Laboratory, Office of Research and Development, USEPA, Research Triangle Park, North Carolina, USA*, 1984.
- G. Dupuy et G. Knaebel : *Assainir la ville hier et aujourd'hui*. Dunod, 1982.
- S. C. Edberg, M. J. Allen et D. B. Smith : Defined substrate technology method for rapid and specific simultaneous enumeration of total coliforms and escherichia coli from water : collaborative study. *Journal of the Association of Official Analytical Chemists*, 74(3):526–529, 1991.

- 
- S. C. Edberg, H. Leclerc et J. Robertson : Natural protection of spring and well drinking water against surface microbial contamination. ii. indicators and monitoring. parameters for parasites. *Critical Reviews in Microbiology*, 23(2):179–206, 1997.
- S. C. Edberg, E. W. Rice, R. J. Karlin et M. J. Allen : Escherichia coli : the best biological drinking water indicator for public health protection. *In Symposium Series (Society for Applied Microbiology)*, num. 29, p. 106S, 2000.
- H. El Khorassani, P. Trebuchon, H. Bitar et O. Thomas : A simple uv spectrophotometric procedure for the survey of industrial sewage system. *Water Science and Technology*, 39 (10-11):77–82, 1999.
- EPA : Ambient water quality criteria for bacteria -1986. *EPA/440/5-84-002. Office of Water Regulations and Standards, USEPA, Washington, DC*, 1986.
- J. Falk et K. Niemczynowicz : Characteristics of the above-ground runoff in sewer catchments. *Urban Storm Drainage*, p. 159–171, 1978.
- D. M. Ferguson, D. F. Moore, M. A. Getrich et M. H. Zhouandai : Enumeration and speciation of enterococci found in marine and intertidal sediments and coastal water in southern california. *Journal of applied microbiology*, 99(3):598–608, 2005.
- L. Fewtrell et J. Bartram : *Water quality : Guidelines, standards, and health : Assessment of risk and risk management for water-related infectious disease*. IWA Publishing, 2001.
- L. Fiksdal, M. Pommepuy, M. P. Caprais et I. Midttun : Monitoring of fecal pollution in coastal waters by use of rapid enzymatic techniques. *Applied and environmental microbiology*, 60 (5):1581, 1994.
- FNDAE : Les bassins dorage sur les réseaux d'assainissement. *Document technique n°6*, p. 61, 1988.
- T. Garcia-Armisen : *Etude de la dynamique des Escherichia coli dans les rivières du bassin de la Seine*. Thèse de doctorat, Université Libre de Bruxelles, 173 p., 2006.
- H. Garnier, M. Gilson, T. Bastogne et A. Richard : Identification de modèles paramétriques à temps continu. *Techniques de l'ingénieur*, S 7 140:17 p., 2007.

- 
- H. Garnier et L. Wang, éd. *Identification of continuous-time models from sampled data*. Springer-Verlag, London, 2008.
- F. Gauthier et F. Archibald : The ecology of "fecal indicator" bacteria commonly found in pulp and paper mill water systems. *Water Research*, 35(9):2207 – 2218, 2001.
- I. George, P. Crop et P. Servais : Use of  $\beta$ -d-galactosidase and  $\beta$ -d-glucuronidase activities for quantitative detection of total and fecal coliforms in wastewater. *Canadian Journal of Microbiology*, 47(7):670–675, 2001.
- I. George, M. Petit et P. Servais : Use of enzymatic methods for rapid enumeration of coliforms in freshwaters. *Journal of applied microbiology*, 88(3):404–413, 2000.
- M. Gevers : A personal view of the development of system identification - a 30-year journey through an exciting field. *IEEE Control Systems Magazine*, 26(6):93–105, 2006.
- S. T. Glassmeyer, E. T. Furlong, D. W. Kolpin, J. D. Cahill, S. D. Zaugg, S. L. Werner, M. T. Meyer et D. D. Kryak : Transport of chemical and microbial compounds from known wastewater discharges : potential for use as indicators of human fecal contamination. *Environmental Science and Technology*, 39(14):5157–5169, 2005.
- D. J. Grimes et R. R. Colwell : Viability and virulence of escherichia coli suspended by membrane chamber in semitropical ocean water. *FEMS Microbiology Letters*, 34(2):161 – 165, 1986.
- M. C. Gromaire, S. Garnaud, M. Saad et G. Chebbo : Contribution of different sources to the pollution of wet weather flows in combined sewers. *Water Research*, 35(2):521–33, 2001.
- D. M. Haaland et E. V. Thomas : Partial least-squares methods for spectral analyses. 1. relation to other quantitative calibration methods and the extraction of qualitative information. *Analytical Chemistry*, 60(11):1193–1202, 1988.
- C. Hagedorn et S. B. Weisberg : Chemical-based fecal source tracking methods : current status and guidelines for evaluation. *Reviews in Environmental Science and Biotechnology*, 8(3):275–287, 2009.

- 
- C. M. Hardina et R. S. Fujioka : Soil : the environmental source of escherichia coli and enterococci in hawaii's streams. *Environmental toxicology and water quality*, 6(2):185–195, 2006.
- P. G. Hartel, C. Hagedorn, J. L. McDonald, J. A. Fisher, M. A. Saluta, J. W. Dickerson, L. C. Gentit, S. L. Smith, N. S. Mantripragada, K. J. Ritter et C. N. Belcher : Exposing water samples to ultraviolet light improves fluorometry for detecting human fecal contamination. *Water Research*, 41(16):3629–42, 8 2007.
- L. M. He et Z. L. He : Water quality prediction of marine recreational beaches receiving watershed baseflow and stormwater runoff in southern california, usa. *Water Research*, 42 (10-11):2563–73, 2008.
- L. Herremans, J. Daub et T. Striebel : Inexistence du premier flot d'orage - exemple d'un petit bassin autoroutier. *Documents de travail pour le groupe "Rejets urbains de temps de pluie " de l'AGHTM*, p. 2, 1995.
- A. Hóskuldsson : Variable and subset selection in pls regression. *Chemometrics and Intelligent Laboratory Systems*, 55(1-2):23–38, 2001.
- M. Hochedlinger, G. Gruber et H. Kainz : Assessment of spill flow emissions on the basis of measured precipitation and waste water data. *Atmospheric Research*, 77(1-4):74–87, 2005.
- M. Hochedlinger, H. Kainz et W. Rauch : Assessment of cso loads–based on uv-vis-spectroscopy by means of different regression methods. *Water Science and Technology*, 54(6-7):239–246, 2006.
- R. Horton : The role of infiltration in the hydrologic cycle. *Transactions of the American Geophysical Union*, 14:446–460, 1933.
- IFEN : Les progrès de la collecte des eaux usées et pluviales. *Les données de l' environnement*, 93:4, 2004.
- S. Ishii et M. J. Sadowsky : Escherichia coli in the environment : Implications for water quality and human health. *Microbes and environments*, 23(2):101–108, 2008.
- J. H. Jiang, R. J. Berry, H. W. Siesler et Y. Ozaki : Wavelength interval selection in multicomponent spectral analysis by moving window partial least-squares regression with applications

- 
- to mid-infrared and near-infrared spectroscopic data. *Analytical Chemistry*, 74(14):3555–3565, 7 2002.
- S. Jovanovic : Hydrologic approaches in urban drainage system modelling. urban drainage modelling. *Dubrovnik, Pergamon Press editor, London, UK*, p. 185–208, 1986.
- D. Kuss : Gestion intégrée du système d'assainissement du rosenmeer. Mémoire de fin d'études d'ingénieur, ENGEES, 96 p., 2007.
- D. Kuss, A. G. Sadowski, J. Vazquez et C. Beck : Minimization of urban wastewater system impacts on receiving waters by the optimization of the couple storm tank/waste water treatment plant. In R. W. Babcock, Jr et R. Walton, eds : *ASCE Conference Proceedings*, vol. 316, p. 662, 2008.
- C. Lacour, G. Chebbo et C. Joannis : Évaluation de flux de polluants dans un réseau unitaire à partir de mesures en continu de turbidité. *Technique - Sciences - Méthodes*, 1-2:47–53, 2010.
- G. Langergraber, N. Fleischmann et F. Hofstadter : A multivariate calibration procedure for uv/vis spectrometric quantification of organic matter and nitrate in wastewater. *Water Science and Technology*, 47(2):63–71, 2003.
- J. Lau, D. Butler et M. Schütze : Is combined sewer overflow spill frequency/volume a good indicator of receiving water quality impact? *Urban water*, 4(2):181–189, 2002.
- V. Laurain : *Contributions à l'identification de modèles paramétriques non linéaires. Application à la modélisation de bassins versants ruraux*. Thèse de doctorat, Université Henri Poincaré, Nancy 1, 142 p., 2010.
- V. Laurain, M. Gilson, S. Payraudeau, C. Grégoire et H. Garnier : Identification de modèles lpv : application à la modélisation pluie-débit d'un bassin versant viticole. *Sixième Conférence Internationale Francophone d'Automatique, CIFA 2010, Nancy : France*, 2010.
- K. Le Roy : Optimisation du traitement des rejets unitaires de temps de pluie par un couplage de modèles réseau et ouvrages de traitement. *Mémoire de fin d'études d'ingénieur - ENGEES*, 80 p., 2004.

- 
- R. Leardi et L. Norgaard : Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions. *Journal of Chemometrics*, 18(11):486–497, 2004.
- P. Lebaron, A. Henry, A. S. Lepeuple, G. Pena et P. Servais : An operational method for the real-time monitoring of e. coli numbers in bathing waters. *Marine pollution bulletin*, 50(6):652–659, 2005.
- C. Leclaire : Etude du fonctionnement dun bassin de pollution en tête dune station dépuration ; propositions de gestion. *Mémoire de Mastère spécialisé Eau potable et assainissement, ENGEES*, 67 p., 1997.
- M. J. Lees : Data-based mechanistic modelling and forecasting of hydrological systems. *Journal of Hydroinformatics*, 2(1):15–34, 2000.
- M. Lovera : Black-box modelling for the rainfall-runoff relationship : an overview. *In Proceedings of the Conférence Internationale Francophone dAutomatique*, 2010.
- A. Maribas, N. Laurent, P. Battaglia, M. do Carmo Lourenço da Silva, M.-N. N. Pons et B. Loison : Monitoring of rain events with a submersible uv/vis spectrophotometer. *Water Science and Technology*, 57(10):1587–93, 2008.
- H. A. Martens et P. Dardenne : Validation and verification of regression in small data sets. *Chemometrics and Intelligent Laboratory Systems*, 44(1-2):99 – 121, 1998.
- N. Matsche et K. Stumwóhrer : Uv absorption as control-parameter for biological treatment plants. *Water Science and Technology*, 33(12):211–218, 1996.
- S. L. McLellan, A. D. Daniels et A. K. Salmore : Clonal populations of thermotolerant enterobacteriaceae in recreational water and their potential interference with fecal escherichia coli counts. *Applied and environmental microbiology*, 67(10):4934, 2001.
- J. Meirlaen, J. Van Assel et P. A. Vanrollegem : Real time control of the integrated urban wastewater system using simultaneously simulating surrogate models. *Water Science and Technology*, 45(3):109–116, 2002.
- G. Miquel : Rapport sur la qualité de l'eau et de l'assainissement en france. *Office parlementaire d'évaluation des choix scientifiques et technologiques*, 195 p., 2003.

- 
- M. Mourad, J. L. Bertrand-Krajewski et G. Chebbo : Sensitivity to experimental data of pollutant site mean concentration in stormwater runoff. *Water Science and Technology*, 51(2):155–162, 2005.
- M. Mrkva : Automatic uv-control system for relative evaluation of organic water pollution. *Water Research*, 9(5-6):587–589, 1975.
- M. Mrkva : Evaluation of correlations between absorbance at 254 nm and cod of river waters. *Water Research*, 17(2):231 – 235, 1983.
- T. J. Mulvaney : On the use of self-registering rain and flood gauges in making observations of the relations of rainfall and flood discharges in a given catchment. *Proceedings of the institution of Civil Engineers of Ireland*, 4:19–31, 1851.
- J. E. Nash : Systematic determination of unit hydrograph parameters. *Journal of Geophysical Research*, 64(1):111–115, 1959.
- M. B. Nevers et R. L. Whitman : Coastal strategies to predict escherichia coli concentrations for beaches along a 35 km stretch of southern lake michigan. *Environmental Science and Technology*, 42(12):4454–4460, 2008.
- M. B. Nevers et R. L. Whitman : Nowcast modeling of escherichia coli concentrations at multiple urban beaches of southern lake michigan. *Water Research*, 39(20):5250 – 5260, 2005.
- L. Norgaard, A. Saudland, J. Wagner, J. P. Nielsen, L. Munck et S. B. Engelsen : Interval partial least-squares regression (ipls) : A comparative chemometric study with an example from near-infrared spectroscopy. *Applied spectroscopy*, 54(3):413–419, 3 2000.
- J. D. Oliver : The viable but nonculturable state in bacteria. *Journal of Microbiology*, 43 (1):93–100, 2005.
- G. O’Loughlin, W. Huber et B. Chocat : Rainfall-runoff processes and modelling. *Journal of Hydraulic Research*, 34(6):733–746, 1996.
- G. A. Olyphant : Statistical basis for predicting the need for bacterially induced beach closures : Emergence of a paradigm ? *Water Research*, 39(20):4953 – 4960, 2005.

- 
- G. A. Olyphant et R. L. Whitman : Elements of a predictive model for determining beach closures on a real time basis : the case of 63rd street beach chicago. *Environmental monitoring and assessment*, 98(1-3):175–90, 11 2004.
- E. Paturel, M. Desbordes et J. M. Masson : Evaluation de l'influence de la densité des réseaux pluviométriques sur la détermination des lames d'eau précipitées. *Laboratoire d'Hydrologie Mathématique, Université Montpellier II*, note interne LHM 14:44 p., 1986.
- N. Paxeus et H. F. Schroder : Screening for non-regulated organic compounds in municipal wastewater in goteborg, sweden. *Water Science and Technology*, 33(6):9–15, 1996.
- A. Phatak et S. De Jong : The geometry of partial least squares. *Journal of Chemometrics*, 11(4):311–338, 1997.
- J. P. Philippe et J. Ranchet : Pollution des eaux de ruissellement pluvial en zone urbaine. *Synthèse des mesures sur dix bassins versants en région parisienne. Rapport de recherche LPC*, 142, 1987.
- F. Previdi et M. Lovera : Identification of parametrically-varying models for the rainfall-runoff relationship in urban drainage networks. *15th IFAC Symposium on System Identification, Saint-Malo : France*, p. 1768–1773, 2009.
- F. Previdi, M. Lovera et S. Mambretti : Identification of the rainfall-runoff relationship in urban drainage networks. *Control Engineering Practice*, 7(12):1489 – 1504, 1999.
- G. P. Rao et H. Garnier : Numerical illustrations of the relevance of direct continuous-time model identification. *In 15th Triennial IFAC World Congress on Automatic Control, Barcelona (Spain)*, 2002.
- P. P. Rasmussen et A. C. Ziegler : *Comparison and continuous estimates of fecal coliform and escherichia coli bacteria in selected Kansas streams, May 1999 through April 2002*. US Geological Survey, 2003.
- W. Rauch, J. L. Bertrand-Krajewski, P. Krebs, O. Mark, W. Schilling, M. Schütze et P. A. Vanrolleghem : Mathematical modelling of integrated urban drainage systems. *Water Science and Technology*, 45(3):81–94, 2002.

- 
- L. Rieger, G. Langergraber, D. Kaelin, H. Siegrist et P. A. Vanrolleghem : Long-term evaluation of a spectral sensor for nitrite and nitrate. *Water Science and Technology*, 57 (10):1563–1569, 2008.
- L. Rieger, G. Langergraber et H. Siegrist : Uncertainties of spectral in situ measurements in wastewater using different calibration approaches. *Water Science and Technology*, 53 (12):187–197, 2006.
- L. Rieger, G. Langergraber, M. Thomann, N. Fleischmann et H. Siegrist : Spectral in-situ analysis of no<sub>2</sub>, no<sub>3</sub>, cod, doc and tss in the effluent of a wwtp. *Water Science and Technology*, 50(11):143–152, 2004.
- C. N. Ross : The calculation of flood discharge by the use of time contour plan isochrones. *Transactions of the institue of Engineers, Australia*, 2:85–92, 1921.
- F. Rossi, A. Lendasse, D. Francois, V. Wertz et M. Verleysen : Mutual information for the selection of relevant variables in spectrometric nonlinear modelling. *Chemometrics and Intelligent Laboratory Systems*, 80(2):215–226, 2 2006.
- A. G. Sadowski : Traitement des eaux usées urbaines. *Cours ENGEES*, 523 p., 2002.
- J. D. Sartor et a. Boyd, Gail B. : *Water pollution aspects of street surface contaminants. Environmental protection technology series, EPA-R2-72-081*. Office of Research and Monitoring, U.S. Environmental Protection Agency, Washington, 1972.
- M. A. Savageau : Escherichia coli habitats, cell types, and molecular mechanisms of gene control. *American Naturalist*, 122(6):732–744, 1983.
- M. Schutze, D. Butler et M. B. Beck : Synopsis—a tool for the development and simulation of real-time control strategies for the urban wastewater system. *In Eighth International Conference on Urban Storm Drainage*, p. 1847–1854. Sydney, 1999.
- M. A. Sharaf, D. L. Illman et B. R. Kowalski : *Chemometrics*. John wiley & sons, new york édn, 1986.
- R. Siegener et R. F. Chen : Caffeine in boston harbor seawater. *Marine pollution bulletin*, 44 (5):383–387, 5 2002.

- 
- T. Soderstrom et P. Stoica : Instrumental variable methods for system identification. *Circuits Systems and Signal Processing*, 21(1):1–9, 1 2002.
- H. M. Solo-Gabriele, M. A. Wolfert, T. R. Desmarais et C. J. Palmer : Sources of escherichia coli in a coastal subtropical environment. *Applied and Environmental Microbiology*, 66 (1):230, 2000.
- P. Tallon, B. Magajna, C. Lofranco et K. T. Leung : Microbial indicators of faecal contamination in water : a current perspective. *Water, Air, and Soil Pollution*, 166(1):139–166, 2005.
- C. Tan et M. L. Li : Mutual information-induced interval selection combined with kernel partial least squares for near-infrared spectral calibration. *Spectrochimica Acta*, 71(4):1266–1273, 12 2008.
- O. Thomas, F. Theraulaz, C. Agnel et S. Suryani : Advanced uv examination of wastewater. *Environmental technology*, 17(3):251–261, 1996.
- O. Thomas : *UV-visible Spectrophotometry of Water and Wastewater*. 2007.
- A. Torres et J. L. Bertrand-Krajewski : Partial least squares local calibration of a uv-visible spectrometer used for in situ measurements of cod and tss concentrations in urban drainage systems. *Water Science and Technology*, 57(4):581–588, 2008.
- USEPA : Bacterial water quality standards for recreational waters (freshwater and marine waters). Rap. tech., 6 2003.
- S. Vaillant, M. F. Pouet et O. Thomas : Basic handling of uv spectra for urban water quality monitoring. *Urban water*, 4(3):273–281, 2002.
- F. Valiron et J. P. Tabuchi : *Maitrise de la pollution urbaine par temps de pluie. Etat de l'art*. Tec et doc, 1992.
- P. A. Vanrolleghem, L. Benedetti et J. Meirlaen : Modelling and real-time control of the integrated urban wastewater system. *Environmental Modelling and Software*, 20(4):427–442, 2005.
- K. Varmuza et P. Filzmoser : *Introduction to multivariate statistical analysis in chemometrics*. CRC Press, Boca Raton, 2009.

- 
- Y. Wang, M. J. Lysaght et B. R. Kowalski : Improvement of multivariate calibration through instrument standardization. *Analytical Chemistry*, 64(5):562–564, 1992.
- Y. Wang, D. J. Veltkamp et B. R. Kowalski : Multivariate instrument standardization. *Analytical Chemistry*, 63(23):2750–2756, 1991.
- S. R. Weibel, R. J. Anderson et R. L. Woodward : Urban land runoff as a factor in stream pollution. *Journal of the Water Pollution Control Federation*, 36(7):914–924, 1964.
- M. Wolfrom : La pollution des eaux du rhin. *Annuaire français de droit international*, 10:737–763, 1964.
- L. Wymer, A. Dufour, K. Brenner, J. Martinson, W. Stutts et S. Schaub : The impact beaches project. Results from a study on microbiological monitoring in recreational waters. *Office of Research and Development, U.S. Environmental Protection Agency*, EPA 600/R-04/023:83 p., 8 2005.
- P. Young : Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. *Environmental Modelling and Software*, 13(2):105–122, 1998.
- P. C. Young : Optimal iv identification and estimation of continuous-time tf models. *In 15th Triennial IFAC World Congress on Automatic Control*, 2002.
- P. C. Young : Identification and estimation of continuous-time hydrological models from discrete-time data. *In Hydrology : Science and Practice for the 21st Century*, vol. 1, p. 406–413, London, 2004. British Hydrological Society.
- P. C. Young et K. J. Beven : Data-based mechanistic modelling and the rainfall-flow non-linearity. *Environmetrics*, 5(3):335–363, 1994.
- P. C. Young et H. Garnier : Identification and estimation of continuous-time, data-based mechanistic (dbm) models for environmental systems. *Environmental Modelling and Software*, 21(8):1055–1072, 8 2006.
- P. C. Young et A. Jakeman : Refined instrumental variable methods of recursive time-series analysis - part iii. extensions. *International Journal of Control*, 31(4):741–764, 1980.
- R. T. Zoch : On the relation between rainfall and streamflow. *Monthly Weather review*, 62 (9):315–322, 1934.

---

X. Zou, J. Zhao et Y. Li : Selection of the efficient wavelength regions in ft-nir spectroscopy for determination of ssc of "fuji" apple based on bpls and fipls models. *Vibrational Spectroscopy*, 44(2):220–227, 2007.

---

# Annexes

## **Annexe A**

# **Validation du dosage de DCO par la méthode de reflux en système fermé suivi d'un dosage par colorimétrie avec le bichromate de potassium**

Le but de cette annexe est de vérifier la validité de la méthode normalisée de reflux en tube fermé ISO 15705 : 2002. Nous avons dosé la DCO de solutions étalon de concentrations connues allant de 10 mg/l à 900 mg/l avec la méthode normalisée ISO 6060 :1989 et la méthode de reflux en tube fermé. Toutes les analyses ont été réalisées en triplicat.

### **A.1 Méthode normalisée ISO 6060 :1989**

Les résultats d'analyse sont présentés dans le tableau A.1. L'écart relatif maximum est de 58% et est obtenu pour une concentration de 30 mg/l. Cette concentration constitue la limite de validité de la norme. La fonction d'erreur RMSE est égale à 39,9 mg/l, ce qui est satisfaisant. Par ailleurs, la réplicabilité de la méthode est excellente avec des valeurs de coefficients de variation n'excédant pas 12 %.

<b>Concentration solution étalon (mg/l)</b>	<b>Concentration mesurée (mg/l)</b>	<b>Ecart relatif</b>	<b>Ecart Type</b>	<b>CV</b>
10	10	0%	1,2	0,12
30	47	58%	0,4	0,01
70	80	14%	8,2	0,12
100	104	5%	6,8	0,07
300	381	27%	2,1	0,01
500	547	9%	4,7	0,01
700	736	5%	26,1	0,04
900	948	5%	0,8	0,00

**RMSE            39,9 mg/l**

Tab. A.1 – Erreurs de mesure obtenues avec la méthode normalisée par comparaison avec des solutions étalon

## **A.2 Méthode de reflux en système fermé**

Les résultats d'analyse sont présentés dans le tableau A.2. La précision de la méthode est excellente : la fonction d'erreur vaut 20,4 mg/l et est inférieure à celle obtenue avec la méthode de référence. Toutefois, la réplicabilité est moins bonne avec des valeurs de coefficients de variation légèrement supérieures.

Les résultats d'analyse des deux méthodes sont représentés en FIG. A.1

Concentration solution étalon (mg/l)	Concentration mesurée (mg/l)	Ecart relatif	Ecart Type	CV
10	17,5	75%	29,9	2,99
30	29,5	2%	15,8	0,53
70	67,2	4%	11,6	0,17
100	100,5	0%	7,2	0,07
300	348,1	16%	85,1	0,28
500	478,0	4%	39,0	0,08
700	712,6	2%	48,3	0,07
900	882,3	2%	35,2	0,04

**RMSE**                      **20,4**

Tab. A.2 – Erreurs de mesure obtenues avec la méthode normalisée par comparaison avec des solutions étalon

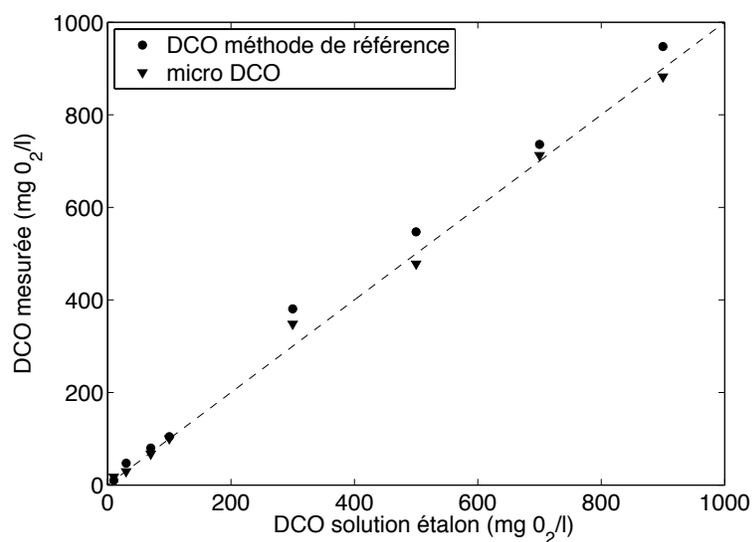


Fig. A.1 – Comparaison de la méthode normalisée ISO 6060 :1989 (référence) et de la méthode micro-DCO

## Annexe B

# Transfert de spectres

La mesure de spectres d'absorbance de deux échantillons d'eau identiques sur deux spectrophotomètres différents aboutit à des résultats divergents comme l'illustre la FIG. B.1. Les mesures effectuées avec le spectrophotomètre de laboratoire sont réalisées avec des cuves en quartz de dimensions interne 1cm x 1cm. Le chemin optique du spectro :lyser est de 2 mm et les spectres d'absorbance sont données en  $m^{-1}$ . La conversion des spectres d'absorbance d'un appareil à l'autre devrait donc être d'un facteur 100. La différence entre les deux spectres présentés sur la FIG. B.1 montre que le coefficient de conversion n'est pas de 100. On constate également un léger décalage des spectres en longueur d'onde.

Pour effectuer le transfert de spectres mesurés uniquement avec le spectrophotomètre perkin-elmer, nous avons utilisé l'ensemble des spectres acquis simultanément avec les deux appareils durant les mois de mars et avril et cherché une relation (en gardant des spectres pour valider le modèle choisi).

Les trois méthodes de calage testées ont été :

- l'utilisation d'une relation linéaire du type  $\mathbf{A}_{s::can} = b_1 \mathbf{A}_{perkin}$ , avec  $b_1$  coefficient de proportionnalité ;
- l'utilisation d'une fonction affine du type  $\mathbf{A}_{s::can} = b_1 \mathbf{A}_{perkin} + b_0$  ;
- l'utilisation d'un algorithme de transfert de calage de spectres via un programme existant sous le logiciel MATLAB. La méthode utilisé est la méthode PDS (Piece-wise Direct Standardization) (Wang *et al.*, 1991, 1992)

La performance des méthodes a été évaluée en calculant la somme des erreurs quadratiques obtenues sur l'ensemble des longueurs d'onde. Les résultats sont présentés dans le tableau B.1. Sans grande surprise, la méthode PDS de réajustement direct de la réponse physique du spectrophotomètre s::can connaissant celle du spectrophotomètre perkin donne les meilleurs résultats.

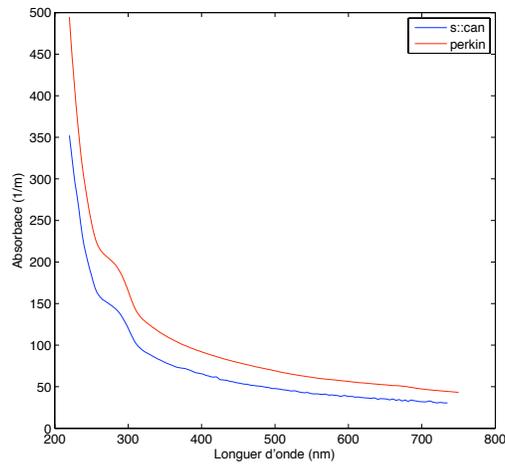


Fig. B.1 – Comparaison de deux mesures de spectres d'absorbance réalisées avec le spectrophotomètre Perkin-elmer et s::can

Méthode utilisée	Coefficients obtenus	SSE
$\mathbf{A}_{s::can} = b_1 \mathbf{A}_{perkin}$	$b_1=76.0$	18180
$\mathbf{A}_{s::can} = b_1 \mathbf{A}_{perkin} + b_0$	$b_1=77.5$ et $b_0=-2.07$	16040
Méthode PDS	-	14073

Tab. B.1 – Résultats du transfert pour les spectres d'entrée (220-735nm)

## Annexe C

# Analyse des signaux de concentration pour les modèles fbiPLS et mwPLS

### C.1 Mesure des concentrations en DCO

#### C.1.1 modèles fbiPLS

Les signaux de concentration en DCO calculés avec les trois modèles fbiPLS ont été tracés en FIG. C.3 pour la journée du 18/04/2008. Ils ont été comparés au modèle de régression linéaire simple établi à partir de la longueur d'onde de 565 nm.

Comme nous pouvions le prévoir, les trois modèles sont influencés par le bruitage des spectres d'absorbance dans le domaine du visible. Cependant, aucun des modèles ne présente les caractéristiques d'une surparamétrisation observable par des oscillations intempestives du signal.

En période nocturne, les trois modèles donnent des concentrations négatives. Les absorbances les plus faibles observées pendant l'étalonnage étaient respectivement égales à 56,4 et 11,6  $\text{m}^{-1}$ . Pendant la journée du 18/04/2010 les absorbances minimales à ces deux longueurs d'onde étaient bien plus faibles avec des valeurs de 34,7 et 1,6  $\text{m}^{-1}$ . Ces trois modèles s'avèrent donc totalement inadaptés à la prédiction de concentration en DCO pour des spectres d'absorbance ne correspondant pas au domaine de validité de l'étalonnage, surtout pour les faibles valeurs d'absorbance. Notons que les valeurs négatives sont obtenues sur les

plages où les spectres sont de classe 3. Cela confirme donc que les modèles ne doivent pas être utilisés sur cette classe de spectres.

Au sein du domaine de validité de l'étalonnage, les modèles sont pertinents. En effet, pour des concentrations prédites avec le modèle RLS1 supérieures à 350 mg/l, les trois modèles fbiPLS donnent des prédictions voisines et légèrement inférieures aux prédictions données avec le modèle RLS1.

Afin de valider l'un des trois modèles pour une utilisation in-situ, il conviendra donc au préalable de :

- compléter la base de données d'étalonnage avec des échantillons prélevés pendant la nuit ;
- et éventuellement effectuer un filtrage des spectres d'absorbance dans le visible.

Il pourra être également judicieux de recueillir des échantillons très chargés en DCO afin d'éviter toute dérive du modèle en borne haute du domaine de validité de l'étalonnage.

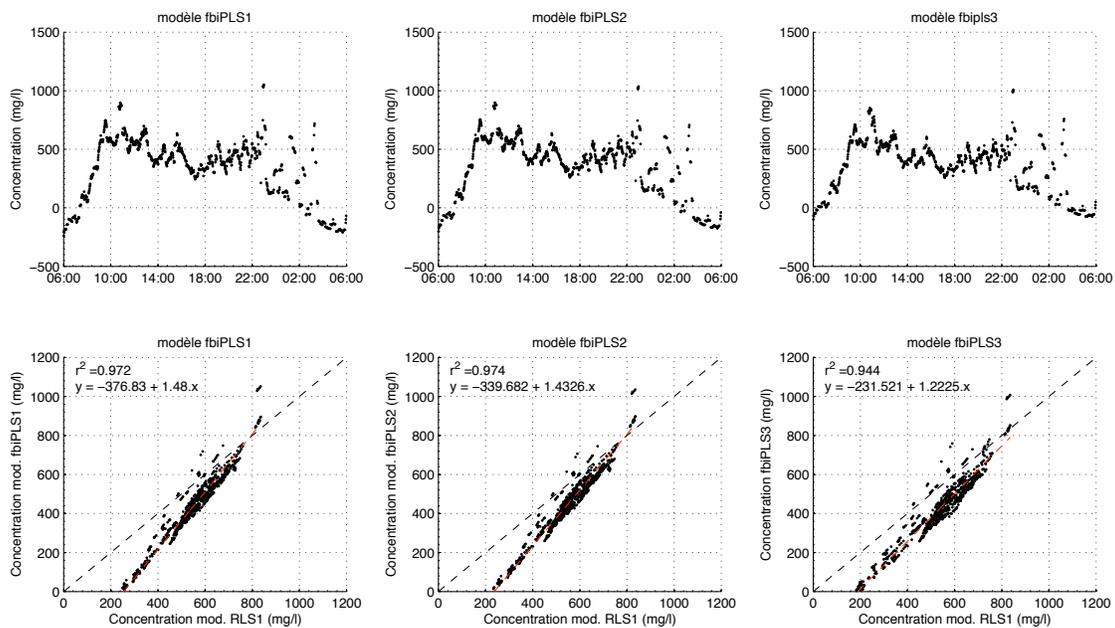


Fig. C.1 – Concentrations en DCO obtenues avec les trois modèles fbiPLS pendant la journée du 18/04/2008

## C.1.2 modèles mwPLS

Les concentrations calculées avec les modèles mwPLS ont été tracées en FIG. C.2.

Le signal de concentration calculé avec le modèle mwPLS1 est le moins bruité. Il utilise en effet les absorbances mesurées entre 262,5 et 360 nm.

En période nocturne, les concentrations calculées avec les trois modèles sont positives. Ces modèles sont donc moins sensibles au changement de classe de spectre.

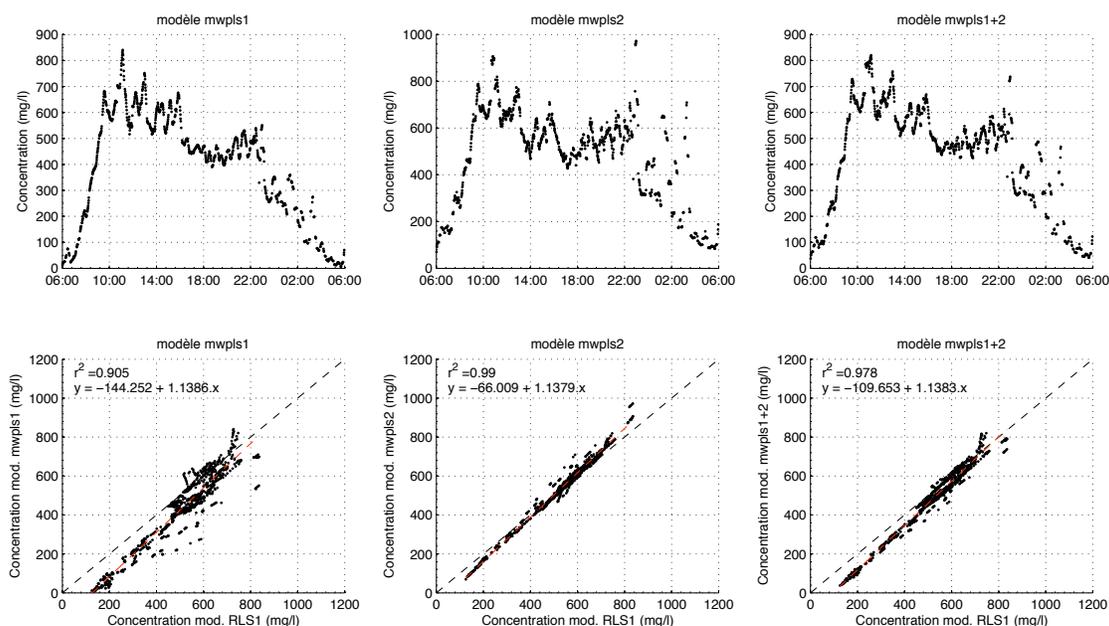


Fig. C.2 – Concentrations en DCO obtenues avec les 3 modèles mwPLS pendant la journée du 18/04/2008

## C.2 Mesure des concentrations en MES

Les concentrations calculées avec les trois modèles fbiPLS sont représentées en FIG. C.3.

On observe de grosses différences entre les trois modèles pour la prédiction des concentrations entre 10h00 et 14h00. Les concentrations calculées sur cette période avec le modèle fbiPLS2 sont les plus plausibles au vu des résultats donnés par les autres modèles. Par ailleurs le signal obtenu avec ce modèle est très peu bruité, ce qui constitue un avantage pour l'utilisation des mesures dans le cadre d'une gestion en temps-réel.

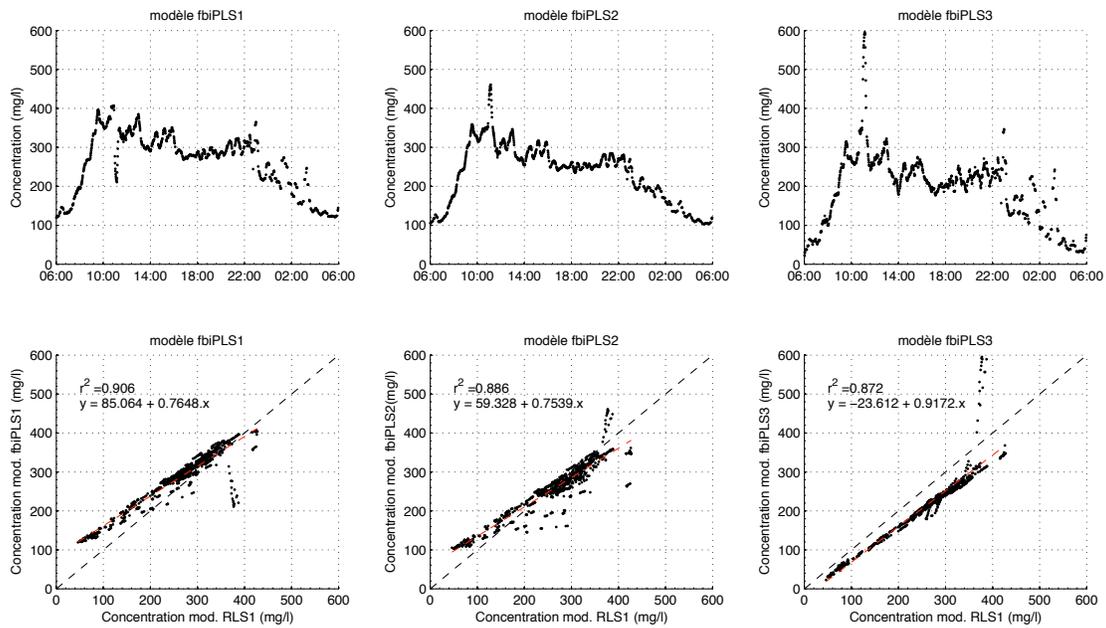


Fig. C.3 – Concentrations en MES obtenues avec les trois modèles fbiPLS pendant la journée du 18/04/2008

## Annexe D

# Régression linéaire effectuée à partir des échantillons mesurés en 2007

La régression a été effectuée en supprimant les trois valeurs les plus importantes d'absorbance qui avaient tendance à déséquilibrer la régression. La régression n'apporte aucun gain en prédiction pour les faibles valeurs d'absorbance ( $A < 2 \text{ m}^{-1}$ ) et donne par contre de meilleurs résultats pour les valeurs intermédiaires.

La différence entre les deux régressions porte essentiellement sur la prédiction des pics de concentrations de E. coli. Pour une valeur d'absorbance de  $30 \text{ m}^{-1}$  la première loi (basée sur 101 individus) donne  $9,3 \cdot 10^4$  et la seconde  $1,07 \cdot 10^5$  soit une différence de 30 000 E. coli/100 ml. La régression présentée dans cette annexe apparaît donc plus *sécuritaire* dans le sens où les pics de concentration de E. coli sont légèrement surévalués. Toutefois, aucun élément ne nous permet de choisir l'une ou l'autre de ces lois. Par ailleurs, la différence de prédiction obtenue sur les eaux du Stang-Alar à partir de ces deux lois aboutira probablement à une différence de concentration négligeable au niveau du site de baignade du fait de la dilution des eaux de la rivière dans la mer.

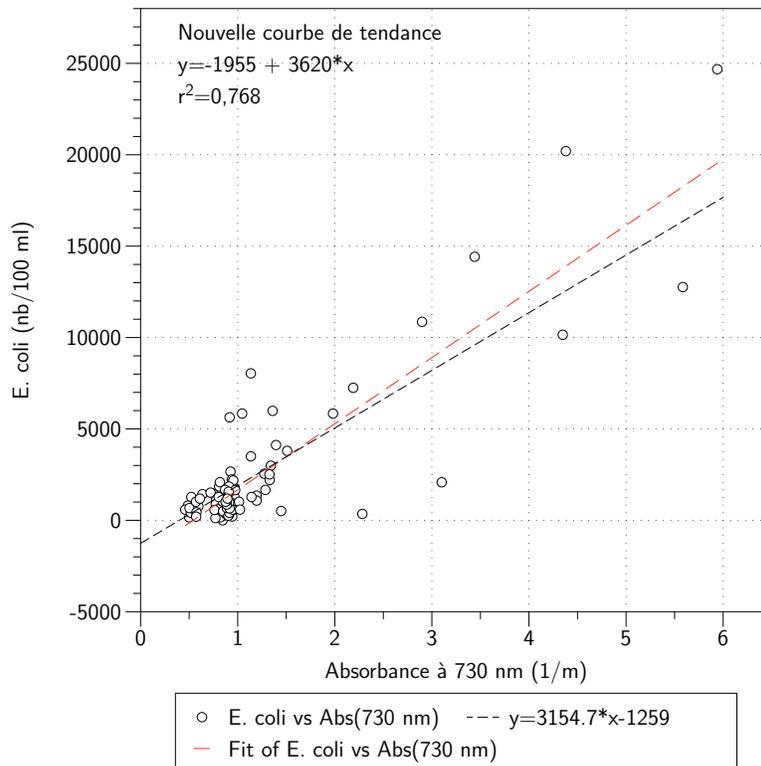


Fig. D.1 – Régression linéaire effectuée sur 98 individus

## Annexe E

# Comparaison des concentrations en *E. coli* mesurées et simulées en 2007

L'équation de régression utilisée est la suivante :

$$E. coli (nb/100ml) = 3154,7 \cdot A_{730} - 1259,1 \quad (E.1)$$

Nous avons limité la valeur maximum de concentration en *E. coli* à  $9,3 \cdot 10^4$ . Les résultats obtenus sur la saison de baignade sont présentés dans les graphiques suivants.

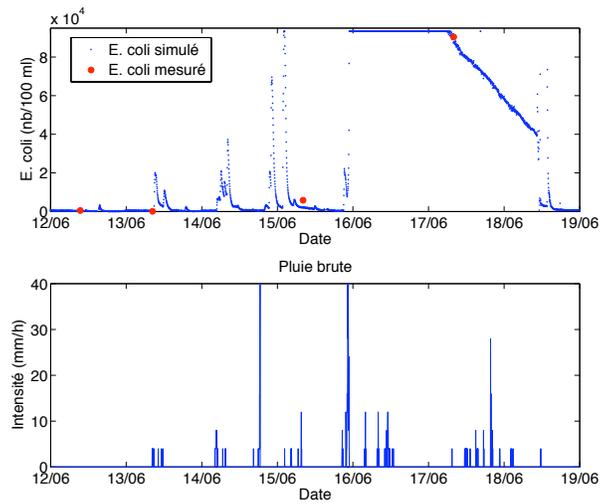


Fig. E.1 – Comparaison des concentrations en E. coli mesurées et simulées du 12 au 19/06/2007

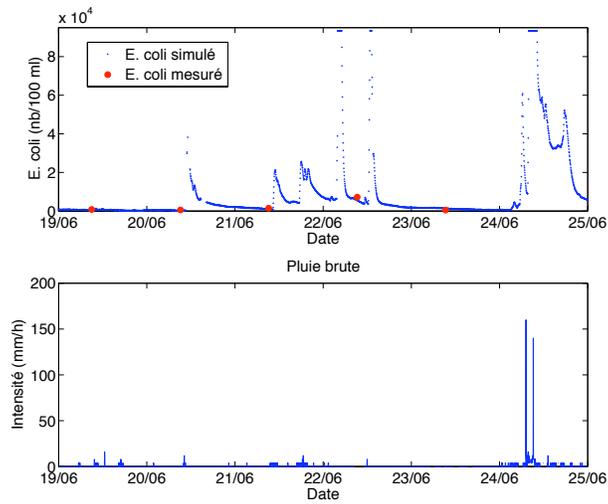


Fig. E.2 – Comparaison des concentrations en E. coli mesurées et simulées du 19 au 25/06/2007

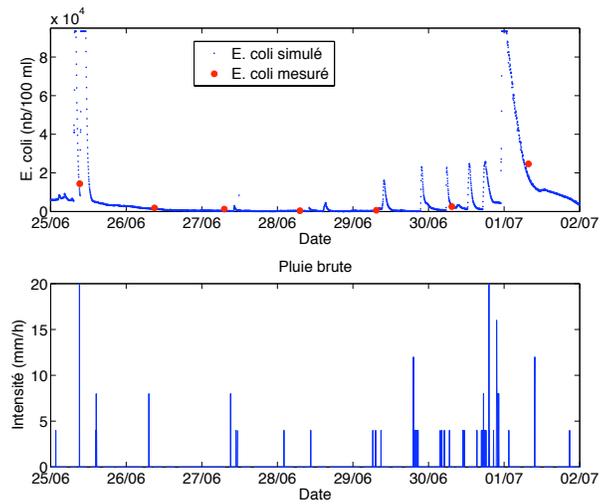


Fig. E.3 – Comparaison des concentrations en E. coli mesurées et simulées du 25/06 au 02/07/2007

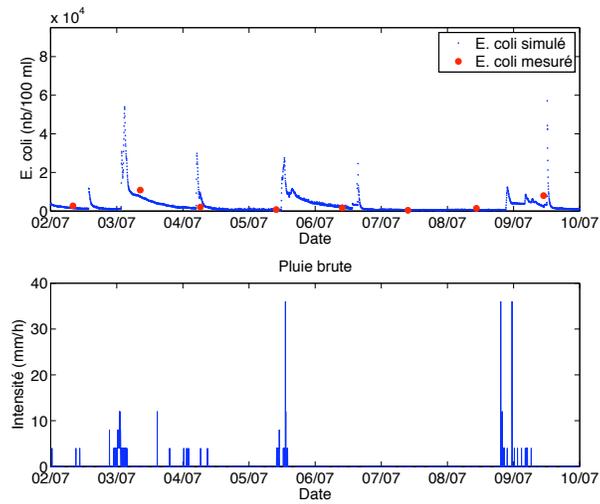


Fig. E.4 – Comparaison des concentrations en E. coli mesurées et simulées du 02 au 10/07/2007

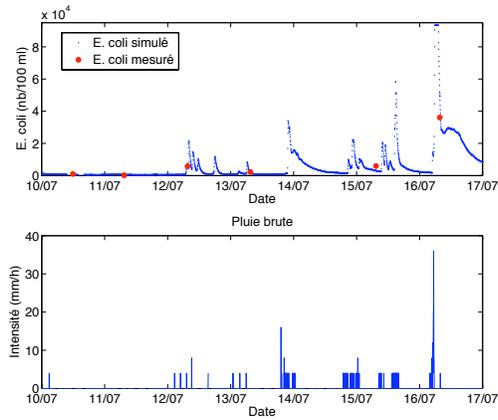


Fig. E.5 – Comparaison des concentrations en E. coli mesurées et simulées du 10 au 17/07/2007

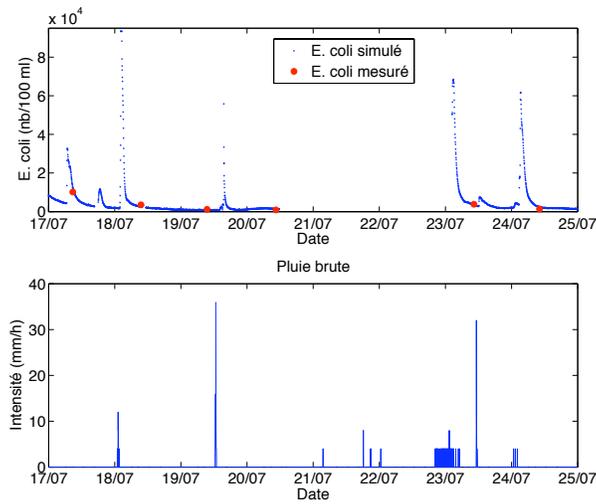


Fig. E.6 – Comparaison des concentrations en E. coli mesurées et simulées du 17 au 25/07/2007

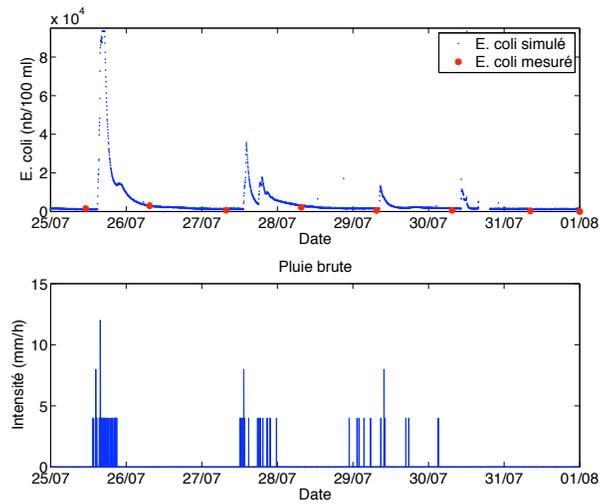


Fig. E.7 – Comparaison des concentrations en E. coli mesurées et simulées du 25/07 au 01/08/2007

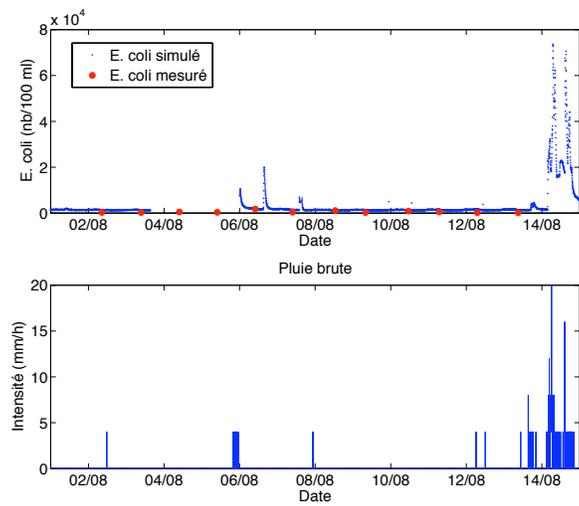


Fig. E.8 – Comparaison des concentrations en E. coli mesurées et simulées du 01 au 15/08/2007

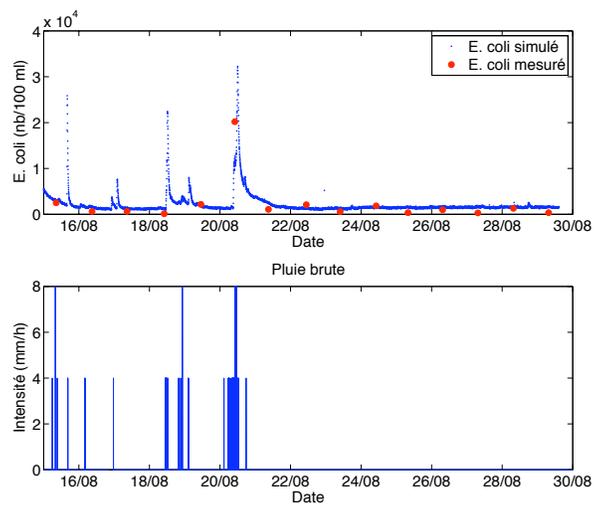


Fig. E.9 – Comparaison des concentrations en E. coli mesurées et simulées du 15 au 30/08/2007