



**HAL**  
open science

# Contribution à la mise en place d'un système multicapteurs pour la surveillance des eaux de surface : fusion de données et aide à la décision

Ahcene Ameziane Azzam

► **To cite this version:**

Ahcene Ameziane Azzam. Contribution à la mise en place d'un système multicapteurs pour la surveillance des eaux de surface : fusion de données et aide à la décision. Autre. Université Henri Poincaré - Nancy 1, 2005. Français. NNT : 2005NAN10188 . tel-01747419

**HAL Id: tel-01747419**

**<https://hal.univ-lorraine.fr/tel-01747419>**

Submitted on 29 Mar 2018

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

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



## AVERTISSEMENT

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

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

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

Contact : [ddoc-theses-contact@univ-lorraine.fr](mailto:ddoc-theses-contact@univ-lorraine.fr)

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

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

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

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



UMR 7039

UFR Ecole Supérieure des Sciences et Technologies de l'Ingénieur de Nancy  
Ecole Doctorale IAE + M  
DFD Automatique et Production Automatisée

## Thèse

Présentée pour l'obtention du titre de

**Docteur de l'Université Henri Poincaré, Nancy 1**  
Spécialité Automatique

Par **Ameziane AZZAM**

**CONTRIBUTION A LA MISE EN PLACE D'UN SYSTEME MULTI-  
CAPTEURS POUR LA SURVEILLANCE DES EAUX DE SURFACE :  
FUSION DE DONNEES ET AIDE A LA DECISION**

Soutenance publique prévue le 06 décembre 2005 devant la Commission d'examen composée de :

Rapporteurs :	M. POLIT	Professeur, Université de Perpignan
	J.C. TRIGEASSOU	Professeur, Université de Poitiers
Examineurs :	H. BEGORRE	Professeur, Université Henri Poincaré Nancy 1
	J.C. BLOCK	Professeur, Université Henri Poincaré Nancy 1
	M. ROBERT	Professeur, Université Henri Poincaré Nancy 1

## Remerciements

Après ces laborieuses années de thèse, je tiens à remercier très sincèrement et tout particulièrement mon directeur de thèse Michel ROBERT, sans qui cette thèse n'aurait jamais vu le jour.

Je tiens à remercier ma femme Radia, pour avoir été à mes cotés et porté avec moi le poids de ces longues années.

Je tiens aussi à remercier mes parents, mes sœurs, Miminette et mes amis proches pour avoir cru en moi jusqu'à l'aboutissement de la thèse.

# Tables des matières

<b>Introduction</b>	<b>4</b>
<b>Généralités sur l'eau</b>	<b>4</b>
<b>Contexte du travail de recherche</b>	<b>5</b>
<b>Références bibliographiques</b>	<b>8</b>
<b>Références Internet</b>	<b>9</b>
<b>CHAPITRE I : Dispositif de mesure – Système multi-capteurs</b>	<b>10</b>
<b>I.1 Introduction</b>	<b>10</b>
<b>I.2 La mesure</b>	<b>14</b>
<b>I.3 Capteur classique et chaîne de mesure</b>	<b>14</b>
<b>I.4 Capteur intelligent et système multi-capteurs</b>	<b>16</b>
I.4.1 Le capteur intelligent	16
I.4.2 Systèmes multi-capteurs	17
I.4.3 Capteurs intelligents dans le cadre d'une architecture multi-capteurs	20
<b>I.5 Acquisition et transfert de données</b>	<b>31</b>
<b>I.6 Conclusion</b>	<b>34</b>
<b>Références bibliographiques</b>	<b>36</b>
<b>Références Internet</b>	<b>38</b>
<b>CHAPITRE II: Tests de toxicité et traitements statistiques</b>	<b>39</b>
<b>II.1 Introduction</b>	<b>39</b>
<b>II.2 Tests de toxicités et plans d'expériences</b>	<b>39</b>
II.2.1 Tests de toxicités multi-facteurs et plans d'expériences	41
II.2.2 Tests de toxicité mono-facteurs	48
<b>II.3 Espérance, Variance, variabilité, covariance et écart-type</b>	<b>49</b>
II.3.1 Variable centrée réduite	50
<b>II.4 Corrélation</b>	<b>51</b>
II.4.1 Coefficient de corrélation	51
II.4.2 Application et résultats de la méthode de base	52
II.4.3 Corrélation par arbre de classification	54
<b>II.5 Analyse en composantes principales (ACP)</b>	<b>56</b>
II.5.1 Définitions	56
II.5.2 Applications et résultats	56
<b>IV.6 Perspectives temporelles</b>	<b>57</b>
<b>II.7 Conclusion</b>	<b>61</b>
<b>Bibliographie :</b>	<b>63</b>
<b>Références Internet</b>	<b>65</b>
<b>CHAPITRE IIII : Fusion de données dans un environnement multi-capteurs</b>	<b>66</b>

<b>III.1</b>	<b>La fusion de données</b>	<b>66</b>
<b>III.2</b>	<b>Modélisation graphique et application à la fusion de données :</b>	<b>68</b>
<b>III.3</b>	<b>Méthodologie</b>	<b>69</b>
III.3.1	Introduction	69
III.3.2	Seuillage par Identification d'une variation dans les séries temporelles	71
III.3.3	Etablissement de masques de pollutions sur la base de pré-alertes	86
<b>III.4</b>	<b>Conclusion</b>	<b>91</b>
	<b>Bibliographie</b>	<b>93</b>
	<b>Références Internet</b>	<b>96</b>
	<b>CHAPITRE IV : Aide à la décision</b>	<b>97</b>
<b>IV.1</b>	<b>Introduction</b>	<b>97</b>
<b>IV.2</b>	<b>Aide à la décision</b>	<b>97</b>
<b>IV.3</b>	<b>Adaptation méthodologique à notre contexte</b>	<b>101</b>
<b>IV.4</b>	<b>Evaluation des Biocapteurs</b>	<b>107</b>
<b>IV.5</b>	<b>Système d'alerte mis en place</b>	<b>110</b>
<b>IV.6</b>	<b>Réflexion sur le positionnement de la station</b>	<b>111</b>
<b>IV.7</b>	<b>Conclusion</b>	<b>112</b>
	<b>Références bibliographiques</b>	<b>113</b>
	<b>Références Internet</b>	<b>115</b>
	<b>Conclusions et perspectives</b>	<b>116</b>
	<b>Annexe 0 : Présentation du pôle de l'eau</b>	<b>119</b>
	<b>Annexe I : Détail du dispositif de la station</b>	<b>120</b>
	<b>Annexe II : Récapitulatif des grandeurs</b>	<b>124</b>
	<b>Annexe III : Extraits du décret n° 89-3 du 3 janvier 1989 (amendé) relatif aux eaux destinées à la consommation humaine à l'exclusion des eaux minérales naturelles</b>	<b>126</b>
	<b>Annexe IV : Tests de pollutions et architecture du dispositif</b>	<b>132</b>
	<b>Annexe V : Exemple de calcul de concentrations de polluants sur le test « agricole 1 »</b>	<b>134</b>
	<b>Annexe VI : Graphiques des tests de pollutions de 9 capteurs</b>	<b>137</b>
	<b>Annexe VII : Observations des tests de pollutions</b>	<b>139</b>
	<b>Annexe VIII : Compléments théoriques</b>	<b>141</b>
	<b>Annexe IX : Maintenance des capteurs</b>	<b>148</b>
	<b>Annexe X : Autres réactions du 22 juin 2002</b>	<b>152</b>
	<b>Glossaire</b>	<b>155</b>

# Introduction

## **Généralités sur l'eau**

L'eau est une ressource indéniablement précieuse, et au niveau mondial une personne sur quatre a à sa disposition moins de 20 litres par jour {1}. La tendance ne cesse de se dégrader pour qu'en 2020 la proportion s'élève à 3 personnes sur 8. A titre comparatif, la consommation par habitant en Europe oscille entre 150 et 200 litres/jour, en Suisse à 240 litres/jour et dans certaines villes des Etats-Unis entre 500 et 600 litres/jours {2}.

En France (2002) {3}, 70% de l'eau potable distribuée est puisée dans les nappes souterraines (forages/sources), le complément provient des eaux de surface : lacs et rivières. Tous ces gisements sont renouvelables : la pluie et la neige contribuent à les remplir, à raison de 200 milliards de m<sup>3</sup> par an, sur lesquels la consommation totale n'excède pas un huitième.

La consommation d'eau potable de l'agglomération nancéenne, avoisine les 100.000 m<sup>3</sup>/jour. Cette eau provient exclusivement de la Moselle. Sur l'ensemble de l'eau distribuée, seuls 2% sont destinés à un usage purement alimentaire, ce qui n'exclut pas d'exiger un niveau de qualité très stricte sur les 100% produits. Si les traitements appliqués sont aujourd'hui relativement bien maîtrisés, leurs performances sont assujetties en partie à la qualité des eaux brutes. Par définition, celle-ci est très variable, ce qui complique nécessairement la gestion des installations de production d'eau potable. Traiter une eau de surface reste complexe en toute saison. En effet, les précipitations hivernales rendent le cours d'eau torrentiel, d'où une turbidité accrue de l'eau, nécessitant un traitement soutenu. D'autre part, les débits faibles en été, posent des problèmes liés à l'apparition d'algues et en une concentration des matières organiques, nécessitant un autre type des traitements complexes. De plus la vulnérabilité des ressources naturelles s'accroît avec l'augmentation de l'activité humaine. Aussi la pollution de l'eau peut avoir plusieurs:

- Les rejets des industries et stations d'épuration urbaines, en amont des stations de pompage ;
- L'agriculture industrielle par le lessivage de pesticides et d'engrais;
- Le lessivage des voies routières et autoroutières à proximité des cours d'eau ;
- Les retombées de micro-polluants émis dans l'atmosphère par les industries et la circulation automobile ;
- Les actes de malveillances et vidanges/déstockages sauvages.

La dégradation de la qualité des eaux brutes induite par ces pollutions, néfastes pour la santé publique, implique des effets et prises de mesures qui peuvent se traduire principalement par :

- Un arrêt momentané du pompage et de la production, voire de la distribution d'eau potable ;
- L'utilisation d'une eau de moindre qualité et l'augmentation d'additifs chimiques ;

Cette situation appelle, en conséquence, à une surveillance étroite de la qualité des eaux brutes.

Afin de prévenir l'usage d'une ressource trop contaminée, la directive européenne du 3 novembre 1998 [DIRECT-NOV1998] relative à la qualité des eaux destinées à la consommation humaine définit les paramètres et valeurs paramétriques à surveiller. Ces paramètres ont été subdivisés en trois catégories :

- les paramètres micro-biologiques : *Escherichia coli* et entérocoques (Partie A de l'Annexe 1 de la directive)
- les paramètres chimiques : l'arsenic, les pesticides et le chrome (Exemples de paramètres de la partie B de l'Annexe 1 de la directive)
- et les paramètres indicateurs tel que : la saveur, l'odeur, la turbidité et la couleur (Exemples de paramètres de la partie C de l'Annexe 1 de la directive)

En complément à la directive précédente, celle du 23 octobre 2000 [DIRECT-OCT2000], transposée dans le droit français par le décret du 20 décembre 2001 [DECRET-DEC2001], fixe des échéances sur la surveillance de la qualité de l'eau :

- 2004 : établir l'état des lieux des ressources en eau. C'est le premier état des lieux à établir en application de la directive sus-mentionnée. Celle-ci n'impose pas la satisfaction de toutes les exigences, de part la disparité de situation des états membres, mais construit les bases d'un état de lieux avec un minimum de données sur la qualité des milieux,
- 2006 : mettre en œuvre de réseaux de surveillances en ayant construit des réseaux de données sur les milieux
- 2009 : planifier la gestion qui arrêtera les objectifs de qualité et quantité des eaux retenues pour 2015 et le programme d'action qui définira pour chaque district les dispositions de nature réglementaire ou mesures de base dans la perspective d'être opérationnelle fin 2012.
- 2015 : atteindre des objectifs de bon état de la ressource: Il s'agira de définir les masses d'eau ou les groupes de masses d'eau risquant de ne pas répondre aux objectifs fixés par la directive.

## **Contexte du travail de recherche**

La CUGN (Communauté Urbaine du Grand Nancy), exploitant du réseau d'eau potable de l'agglomération nancéenne, compte parmi ses prérogatives la surveillance de la qualité de l'eau prélevée, conformément à l'échéance de 2006. Celle-ci mentionne la mise en œuvre de réseaux de surveillances après avoir préalablement construit des réseaux de données sur les milieux. C'est au cœur de cette problématique que s'est inscrit en 1999 le projet européen LIFE 99 ENV/F/000492 : « **PROTECTION ET SURVEILLANCE MULTI-PARAMETRES DES RESSOURCES AQUATIQUES** », pour lequel NANCIE (Centre International de l'Eau de NANCY), présenté en Annexe 0, a été le coordinateur de part son rôle de coordination de projets dans le domaine de l'eau et de l'environnement.

Dans ce cadre, la mise en place d'une station d'alerte à la pollution en amont de la station de pompage en rivière, s'explique par la nécessité de la surveillance qualitative de la ressource en eau destinée à la production d'eau. La station d'alerte composée de différents capteurs, surveille des paramètres physico-chimiques : pH, conductivité, oxygène dissous, température, turbidité, chlorophylle, potentiel redox, auxquels s'ajoutent des paramètres dits « biologiques » : mouvements valvaires de moules, signaux électriques de poissons, luminescence de bactéries. L'intérêt des collectivités locales à investir dans les dispositifs de surveillance de l'eau ressource s'est fait conformément aux exigences de la directive



européenne du 3 novembre 1998 [DIRECT-Nov1998], qui impose le respect de seuils sur les niveaux de produits toxiques pour la production d'eau potable, et implique de pouvoir anticiper le franchissement de ces seuils. Dans ce contexte, une des difficultés tenait de la nécessité d'être informé « à temps », permettant d'avoir une marge de sécurité raisonnable pour la gestion d'une situation de crise. L'autre difficulté était de disposer dans de pareils cas, d'une information simple à comprendre, malgré la profusion d'informations brutes issues de la station d'alerte.

Mon travail de recherche a eu pour objectif de mettre en forme les résultats de l'analyse et de mettre à disposition sous forme d'une information simple dans des délais qui permettent une prise de décision rapide. Cette information doit définir le degré de pollution, dans la perspective d'alimenter le système de décision garant d'une action, selon une échelle d'alerte, qui décidera de l'arrêt du pompage par exemple. Cette prise de décision passe par une concertation entre autorités locales et spécialistes du domaine qui d'une part ont la responsabilité de toute une population et d'autre part le déclenchement d'un plan de secours qui ne doit pas se faire intempestivement.

La méthodologie développée dans ce travail a visé à transformer les informations issues de la collecte des données brutes, en information fusionnée dans la perspective d'une aide à la décision. La fusion de données permet de qualifier les pollutions après avoir établi un état des réponses de chaque capteur, par rapport à différentes pollutions testées en circuit confiné. Sur la base de ces états appelés « masques de pollutions », l'aide multicritère à la décision permet quantifier le degré de gravité de ces pollutions.

La surveillance de la ressource, au travers de cette méthodologie, repose sur l'exploitation commune des deux types de données recueillies via la station multi-capteurs. Celle-ci intègre des paramètres physico-chimiques : pH, conductivité, oxygène dissous, température, turbidité, chlorophylle, potentiel redox ; et des paramètres biologiques : mouvements valvaires de moules, signaux électriques de poissons, luminescence de bactéries. Ces données sont fusionnées en vue d'une contribution à l'établissement d'alertes en fonction du degrés de gravité indiqué par l'aide à la décision, qui implique des actions pouvant être :

- soit préventives pour des pollutions « mineures », qui correspondent à l'anticipation sur les dosages d'additifs au niveau de la station de production et distribution d'eau potable,
- soit de sécurité dans l'éventualité de pollutions « majeures », la prise de décision consiste par exemple à arrêter le pompage sur le cours d'eau et son déroutement vers un bassin de sécurité.

Les différentes étapes du travail de recherche entrepris, sont définis en quatre chapitres:

Dans le premier chapitre, est défini le système multi-capteurs structure centrale du dispositif de recueil de données. Ce système créé par l'étude, est décrit de manière à expliquer les choix des capteurs physico-chimiques et biologiques. Il est utile de préciser que ces choix coïncident avec le démarrage du projet européen, et de ce fait, sont antérieurs au commencement des travaux de recherches.

Le second chapitre, présente les tests de toxicité et plans d'expériences effectués sur l'ensemble des paramètres permettant l'évaluation des réponse des capteurs dans des conditions de pollution en milieu confiné. Le choix des pollutions à tester s'est fait sur la base des risques les plus potentiellement appréhendés sur le cours d'eau étudié. Il s'agit, en l'occurrence, de pollutions : aux métaux lourds, aux hydrocarbures, industrielles, agricoles et écologiques. Les réponses obtenues permettent de connaître les seuils de sensibilité des

capteurs par pollutions pour être ultérieurement exploités en groupes. Puis sont présentés les résultats obtenus lors de traitements statistiques effectués sur les données des tests de toxicité.

Le troisième chapitre, traite de fusion de données en se basant sur des méthodes de détection d'alertes sur des signaux brut. Après des simulations comparatives, la méthode de comparaison de variance est choisie. Celle-ci appliquée à la fusion des données de la station multi-capteurs repose sur le recensement des seuils de sensibilité d'un ensemble de capteurs ayant réagi à un test de pollution particulier. Le groupe de capteurs ainsi constitué représente un masque de pollution, permettant d'alimenter une base de connaissance.

Au quatrième chapitre, est développée l'approche d'aide à la décision, après le rappel des principes fondamentaux. Sur la base des masques qui qualifient chaque pollution, est appliquée l'aide multicritère à la décision permettant de quantifier la gravité des pollutions. Puis est présenté la synthèse du système fonctionnel et son implantation sur le terrain.

Enfin, est présenté la conclusion qui établit un bilan de l'étude, en faisant part des recommandations et perspectives inhérentes. Cette synthèse peut être considérée comme étant un guide méthodologique pour l'exploitation d'un tel type de station multi-paramètres. Celle-ci est destinée à la surveillance de la qualité de l'eau, avec un certain nombre de pièges et problèmes à contourner lors de la mise en place d'une station de ce type, qui doit répondre à des exigences de délais d'actions précédant la contamination du réseau d'eau potable.

*Les références bibliographiques et Internet sont regroupées à la fin de chaque chapitre et respectivement indiquées dans le texte entre [crochets] et entre {accolades}. Un glossaire des principaux termes employés dans le texte est disponible à la fin du document.*

## **Références bibliographiques**

- [DIRECT-Juil1980] DIRECTIVE 80/778/CEE DU CONSEIL du 15 juillet 1980 relative à la qualité des eaux destinées à la consommation humaine
- [DIRECT-Oct1981] DIRECTIVE 81/858/CE DU CONSEIL du 19 octobre 1981 portant adaptation en raison de l'adhésion de la Grèce de la directive 80/778/CEE.
- [DIRECT-Dec1990] DIRECTIVE 90/656/CE DU CONSEIL du 04 décembre 1990 relative aux mesures transitoires applicables en Allemagne concernant certaines dispositions communautaires en matière de protection de l'environnement.
- [DIRECT-Dec1991] DIRECTIVE 91/692/CE DU CONSEIL du 23 décembre 1991 visant à la standardisation et à la rationalisation des rapports relatifs à la mise en oeuvre de certaines directives concernant l'environnement.
- [DIRECT-Nov1998] DIRECTIVE 98/83/CE DU CONSEIL du 3 novembre 1998 relative à la qualité des eaux destinées à la consommation humaine.
- [DIRECT-Oct2000] DIRECTIVE 2000/60/CE DU PARLEMENT EUROPÉEN ET DU CONSEIL du 23 octobre 2000 établissant un cadre pour une politique communautaire dans le domaine de l'eau. Extrait du journal officiel des communautés européennes, pages L 371/1-L371/72.
- [DECRET-Dec2001] DECRET n°2001-1220 du 20 décembre 2001 relatifs aux eaux destinées à la consommation humaine.

## ***Références Internet***

{1} <http://www.eaufrance.com/>

{2} <http://www.attac.org/indexfr/index.html>

{3} [http://www.cc-bievre-liers.fr/html/vie/eau\\_potable.html](http://www.cc-bievre-liers.fr/html/vie/eau_potable.html)

# CHAPITRE I : Dispositif de mesure – Système multi-capteurs

## 1.1 Introduction

Cette ressource si précieuse qu'est l'eau, ne pose relativement moins de soucis d'abondance dans un pays tel que la France par rapport à une grande partie du monde. Se pose alors un autre problème, qui est la préservation de la qualité de l'eau. Cette notion de qualité peut se quantifier par les normes imposées par les différents organismes nationaux et européens, concernant les eaux destinées à la consommation humaine. Nous citerons le *décret n°89-3 du 3 janvier 1989 relatif aux eaux destinées à la consommation humaine à l'exception des eaux minérales*.

En effet, les rejets de l'industrialisation et l'agriculture intensive durant ce dernier demi-siècle ont généralisé la pollution des eaux de surface et souterraines à l'origine de notre consommation d'eau potable. Prenons le cas des pesticides {1}, on les retrouve dans l'eau des rivières, mais aussi dans l'eau de pluie, dans l'air et sous forme de résidus dans les aliments.

Bien entendu, cette contamination généralisée de l'environnement constitue une menace pour la santé. La faune et la flore sont également touchées et menacées par cette pollution. Cet enjeu crucial, qu'est la préservation de la qualité de l'eau est induit en amont par une vigilance des rejets industriels et agricoles et en aval par une surveillance des prélèvements d'eau. La Figure II.1 présente les résultats d'une enquête conduite par l'Institut Français de l'Environnement (IFEN) en juillet 2000 {1}. Cette étude est relative à la contamination des eaux de surface par les pesticides durant la période 1997-1998.

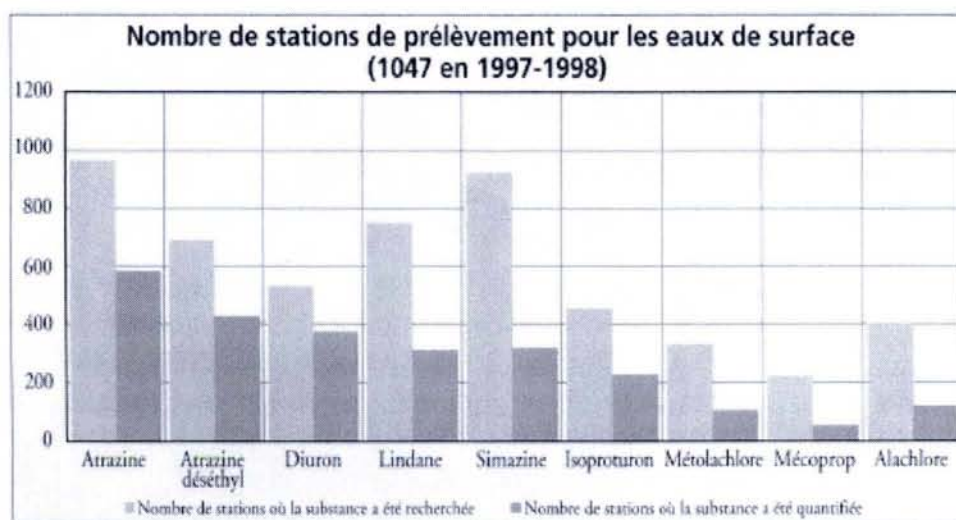
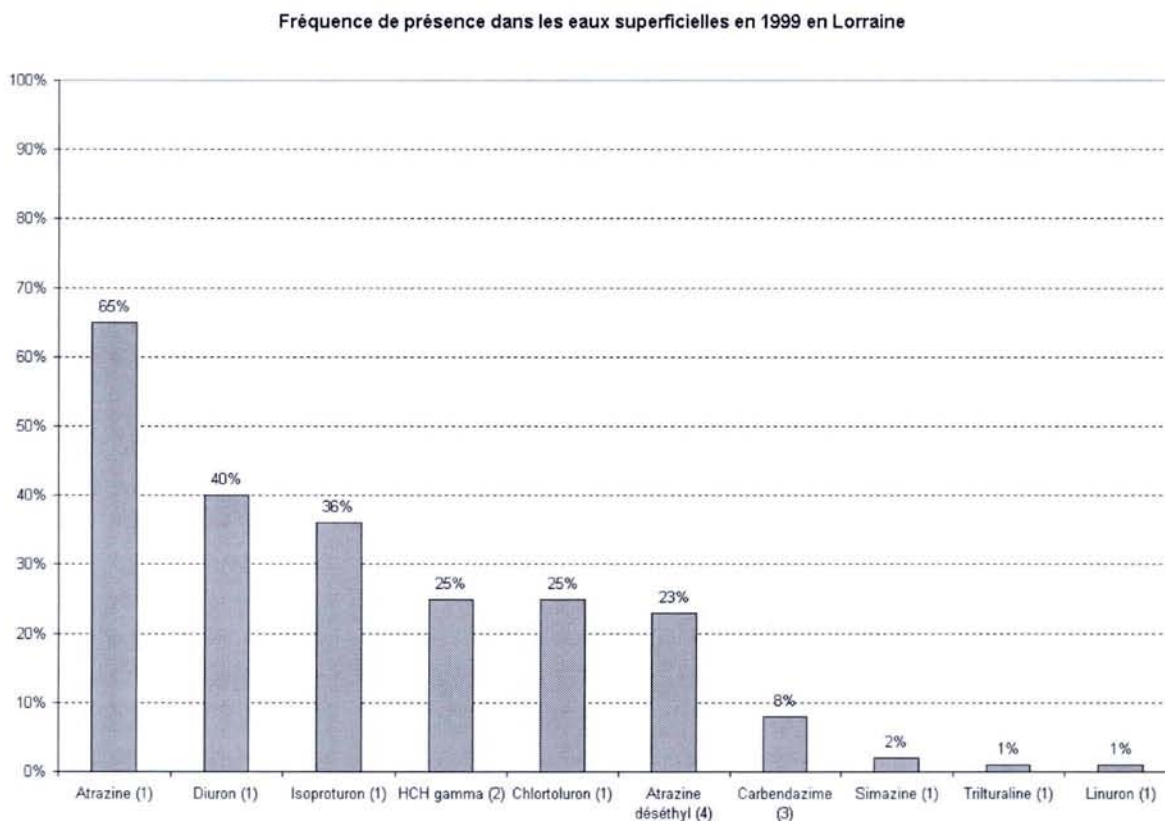


Figure II.1 : « Etude sur la contamination des eaux de surface »

Sur cet échantillon de 1047 stations on peut constater, que la contamination des eaux par des substances toxiques est chose courante. Ce qui implique de définir des normes et des seuils, permettant de garantir une eau la plus saine possible. Cela sous-entend un certain niveau de tolérance, et qu'une eau totalement dépourvue de polluants est une notion relative voire utopique.

Plus localement une *étude de vulnérabilité* a été réalisée sur le bassin versant de la Moselle. Elle révèle les risques majeurs de pollution sur cette ressource.

En l'occurrence l'étude [MAZU-2000] menée par la DIREN (Direction Régionale de l'Environnement) de Lorraine, en collaboration avec l'AERM (Agence de l'Eau Rhin-Meuse) et le CSP (Conseil Supérieur de la Pêche) révèle que l'Atrazine est la substance toxique la plus souvent présente dans les eaux de Lorraine en 1999 (dans 65% des analyses) comme l'indique la figure II.2 :



(1) : herbicide, (2) : insecticide, (3) : fongicide, (4) : produit de dégradation

Figure II.2 : « La qualité des cours d'eau lorrains en 1999 »

De ce type d'observations, se justifie l'utilité de la mise en place d'une station d'alerte, prévenant assez tôt de l'arrivée d'une pollution au niveau même du processus de pompage. En effet, la moindre suspicion sur la qualité de l'eau entraîne le pompage sur un bassin de sécurité, dont l'autonomie est de l'ordre de la semaine. Ce dispositif est déjà en place, à travers une surveillance classique des paramètres physico-chimiques pris individuellement par rapport à des seuils prédéterminés. L'opérateur désigné pour la surveillance de l'évolution des paramètres a la lourde responsabilité de donner l'alerte en cas de pollution avérée. En effet, la prise de décision doit se faire en un laps de temps très court.

Dans notre optique, l'opérateur doit se trouver conforté dans sa décision par notre dispositif d'alerte, dont la finalité est l'aide à la décision sur l'observation d'un ensemble de données sous une forme simplifiée. Le principe consiste à fusionner un ensemble de paramètres physico-chimiques associés à des capteurs biologiques, pour produire des indicateurs de pollutions. Ces indicateurs, selon leur importance déclencheront ou pas une alerte. Dans les cas majeurs, cela signifie le basculement du pompage sur le bassin de sécurité ; dans les cas mineurs, cela conduira à une anticipation sur le pré-traitement de l'eau à effectuer dans l'usine de production d'eau potable.

On peut, alors, poser la question du tout sécuritaire, qui consiste à préconiser le « basculement à tout prix sur le bassin de sécurité quelle que soit l'importance de la pollution ». L'inconvénient d'une telle démarche est l'épuisement des réserves pour des pollutions mineures que l'on peut traiter en usine de traitement, impliquant ainsi un fonctionnement intempestif des équipements, mais aussi le recours à un traitement spécifique de l'eau dû à la nature stagnante du bassin.

De ce fait, le recours à cette solution exige une réflexion se voulant la plus précise possible, enjeu de notre système destiné à se rapprocher au mieux de la réalité d'une pollution. Le but étant de quantifier son degré d'importance selon des indicateurs de pré-alertes, et de la qualifier selon son type.

La notion d'aide à la décision trouve ainsi toute son importance pour « ne pas passer à côté » d'une pollution en anticipant son l'évolution afin d'éviter la contamination du réseau d'eau potable domestique. La performance temporelle de ce processus passe par une concession sur certaines méthodes de calculs laborieuses lourdes, pour la production d'actions instantanées, trouvant ainsi un compromis entre le temps de calcul et le temps de réponse. Ce point est détaillé en chapitre IV.

Les aspects à prendre en compte, sont les distances entre le lieu de pompage, la station d'alerte et l'usine de traitement. Ces éléments permettent de déterminer la marge de sécurité lors d'une pollution en cours de transfert vers l'usine de traitement.

A titre indicatif, le temps de transfert de l'eau, destinée à l'agglomération nancéenne, prend moins d'une demi-heure du lieu de pompage vers l'usine de traitement. C'est durant ce laps de temps que doit s'effectuer le traitement des données et la prise de décision. En effet, en cas de faible pollution, la procédure consistera à élever les dosages d'additifs chimiques aseptisant, alors que dans des cas de pollutions considérables, il sera question de passer à un bassin de sécurité sur l'usine d'eau, le temps de purger et vidanger la canalisation de transfert et le passage sur un bassin de sécurité au lieu du pompage, ce qui évite ainsi la contamination avérée du réseau d'eau potable.

L'objectif du système d'alerte multi-capteurs implanté en amont du pompage, est d'avoir une appréciation qualitative de la situation. Cela est possible de part leur sensibilité sur un ensemble de polluants dont les niveaux ne sont pas nécessairement élevés, ce que n'apporte pas une gestion classique fondée sur une appréciation individuelle des capteurs.

Sur ce principe, la démarche qui est décrite au chapitre suivant consiste à simuler un ensemble de pollutions, pour pouvoir les qualifier en situation réelle en fonction de la réaction de l'ensemble des capteurs.

Un autre apport du système multi-capteurs, est de pouvoir éviter au mieux, le déclenchement de plans d'alertes de manière intempestive. En effet, du fait que nous disposons d'au plus une demi-heure pour décider d'un déroutage du pompage, le temps ne nous permet pas de décider à la légère du passage à des bassins de sécurité se répartissant :

- au niveau de l'usine d'eau, où de précieuses réserves sont épuisées,
- au niveau du bassin de sécurité sur le lieu de pompage où la réserve est certes d'une semaine, mais dont les caractéristiques statiques de l'eau nécessitent un traitement spécifique additionnel.

Concernant l'analyse de données, elle s'effectue dès le prélèvement de l'échantillon de pompage, après quoi chaque capteur produit sa mesure de manière synchronisée, toutes les cinq minutes dans les conditions expérimentales, et toutes les minutes dans les conditions réelles d'exploitations. Malgré une distance de 10 Km entre la station de pompage/multi-

capteurs et la centrale de supervision, les mesures ainsi produites sont instantanément transférées vers cette centrale appelée GTC (Gestion technique centralisée), via une ligne spécialisée pour le traitement ultérieur de fusion de données utile à l'aide à la décision.

La condition de validité du système repose sur une évaluation de la qualité de l'eau avant l'arrivée de cette dernière ainsi mesurée à l'usine de traitement mitoyenne de la GTC et distante de 10 Km du lieu de pompage. Ce qui correspond ainsi à la demi-heure d'acheminement de l'eau, temps à l'issue duquel le résultat doit être connu.

En figure II.3 est représenté le schéma du processus de rapatriement de l'eau en parallèle de celui des données :

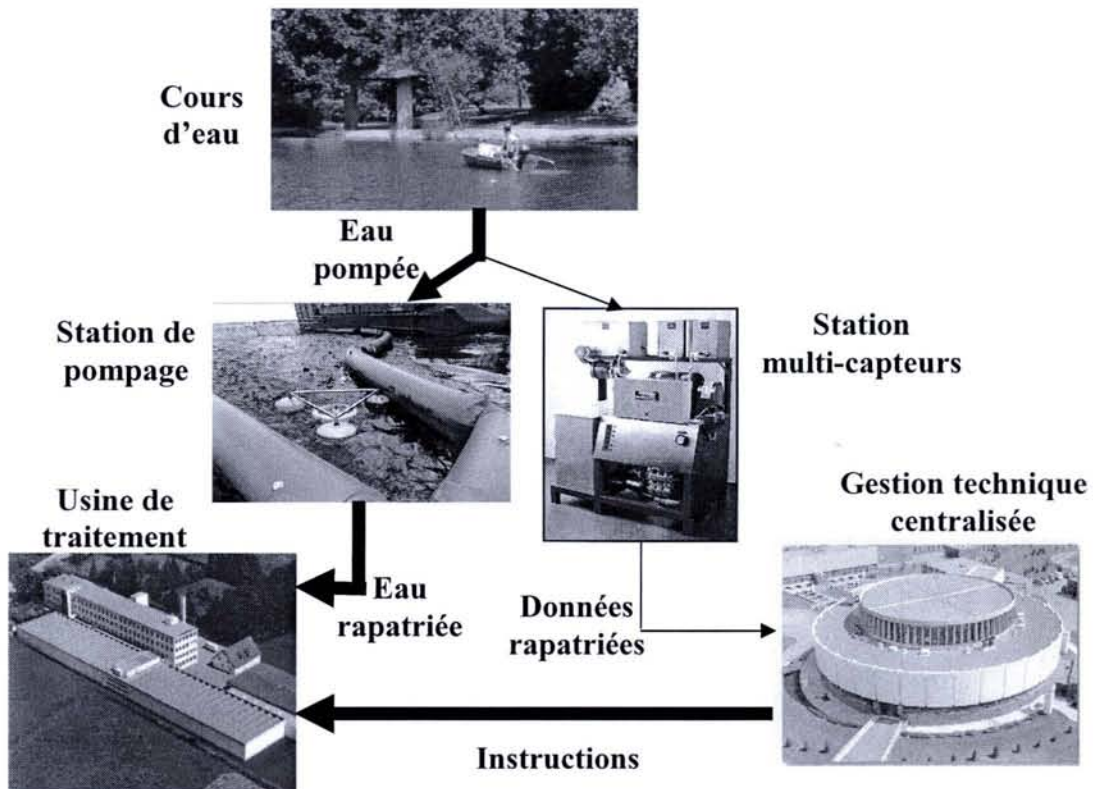


Figure II.3 : « Schéma du processus d'acheminement de l'eau »

Ce schéma dont les éléments sont détaillés dans les points qui suivent, montre explicitement, qu'un échantillon de l'eau pompée va directement à la station multi-capteurs, implantée sur le même site que la station de pompage, alors que la centrale de supervision GTC (Gestion technique centralisée) mitoyenne de l'usine de traitement est délocalisée sur un autre site.

La présentation de notre système d'alerte dans ce chapitre passe par la définition de certains principes de base, en l'occurrence celui de la mesure puis celui du concept de système multi-capteurs. Le dispositif mis en place par nos soins, a en effet nécessité une étude préliminaire respectant certaines contraintes liées à l'environnement de la station d'alerte. L'originalité de notre système, consiste à associer des capteurs physico-chimiques classiques à des capteurs biologiques représentant le milieu naturel sous forme de chaîne trophique. L'intérêt réside dans la confrontation du comportement des capteurs biologiques aux paramètres physico-chimiques, quantifiant ainsi de nouveaux indicateurs de pollutions notamment dans le cadre des tests de toxicité détaillés ultérieurement au chapitre III.

Le choix des capteurs a été adapté d'une part en fonction du cours d'eau à étudier qu'est la Moselle, d'autre part en prenant en compte leur compatibilité à communiquer en



terme de modes de transfert de données, pour un recueil de données centralisé, harmonisé et homogène. Ce point, laborieux pour cause d'hétérogénéité des équipements entre fabricants est présenté en fin de chapitre dans le cadre du dispositif de communication et transfert de données.

## 1.2 La mesure

Le terme « mesure » présente de nombreuses acceptions dans la langue française ; il est donc préférable d'utiliser en métrologie le terme « mesurage » qui est l'action de quantifier ce qui n'était que qualifié [BOIS-1989]. Cela sous-entend la capacité de pouvoir faire correspondre grâce à une échelle absolue ou relative, un ou plusieurs nombres à une situation ou un phénomène donné. Cette correspondance doit être facilement reproductible, transmissible et, pour avoir un sens, associée à une estimation du degré de certitude que l'on peut attacher au résultat [SELI-1992].

Selon l'*International Standardization Organisation* (I.S.O), « La mesure est un ensemble d'opérations ayant pour objet la détermination de la valeur d'une grandeur ». Cette définition est imprécise dans le sens où elle ne fait aucune référence, en particulier, aux moyens mis en œuvre pour son obtention et nécessite donc un développement pour mieux saisir ce concept et son évolution [PRIE-1995].

N'a-t-on pas écrit que : "*l'homme est la mesure de toutes choses*" Protagoras  
(sophiste grec 485-411 avJC)

Jusqu'au XVIIIème siècle il n'existait aucun système de mesure unifié. Nombre de ces mesures étaient empruntées à la morphologie humaine. Leur nom en conservait fréquemment le souvenir : le doigt, la palme, le pied, la coudée, le pas ou encore la brasse. Ces unités de mesures n'étant pas fixes, variaient d'une ville à l'autre, d'une corporation à l'autre, mais aussi selon la nature de l'objet mesuré.

Les mesures de volume et celles de longueur n'avaient aucun lien entre elles. Pour chaque unité de mesure les multiples et sous multiples s'échelonnaient de façon aléatoire, ce qui rendait tout calcul extrêmement laborieux. Pour comprendre les difficultés qu'entraînaient de tels systèmes, il convient de considérer le mode actuel de la mesure du temps, survivance de l'ancien système de subdivisions. A mesure que l'industrie et le commerce prenaient de l'ampleur, la nécessité d'une harmonisation s'imposait, pour cela les scientifiques réformèrent cet état de fait, l'idée étant d'assurer l'invariabilité des mesures en les rapportant à un étalon universel issu d'un phénomène naturel. C'est ainsi qu'ont été fixées les bases de l'unité des mesures.

Le **Système international d'unité** (SI), successeur du système métrique, est officiellement né en 1960. Ce système permet de rapporter toutes les unités de mesure à un petit nombre d'étalons fondamentaux, et de consacrer tous les soins nécessaires à améliorer sans cesse leur définition.

## 1.3 Capteur classique et chaîne de mesure

La grandeur physique objet de la mesure, est désignée comme étant le "*mesurande*". Les opérations concourant à la détermination de sa valeur numérique constituent son mesurage [ASCH-1991]. Celui-ci est possible en utilisant un élément appelé "*corps d'épreuve*" ou "*transducteur*". Le principe consistant à faire correspondre une grandeur d'entrée à une grandeur de sortie selon un conditionnement déterminé.

La figure II.4 représente une schématisation du capteur traditionnel :

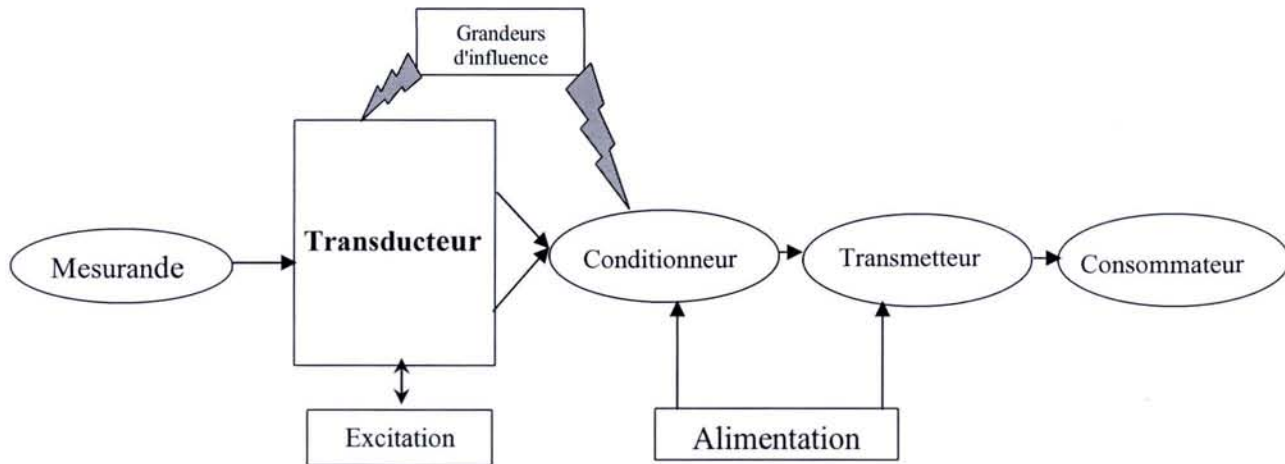


Figure II.4 : « Capteur traditionnel »

La schématisation du capteur traditionnel, correspondant à une synthèse de la chaîne de mesure, nous montre que l'origine du mesurage doit être définie, correspondant à la notion d'étalonnage [ASCH-1991 ; KERO-2002 ; NF-1984 ; NUSS-1987 ; PARA-1986 ; ROBERT-1998].

En effet, l'étalonnage comprend l'ensemble des opérations établissant, la relation entre les valeurs indiquées par un appareil de mesure, et les valeurs connues correspondantes d'une grandeur mesurée.

La phase de mise en place d'un étalonnage précis dans le cadre de notre installation, a été un impératif indéniable de par l'enjeu que peut représenter la réalité d'une mesure, en particulier celle des capteurs biologiques dont la maîtrise et le recul sont moindres par rapport à celle des capteurs physico-chimiques.

La garantie d'un bon étalonnage passe par la "répétabilité" des mesures. Celle-ci témoigne alors de la capacité du capteur à délivrer des valeurs de sortie identiques, dans les limites spécifiées chaque fois que ce capteur est utilisé dans des conditions précisées. Cette étape a été initialement validée dans le contexte de notre environnement de travail qu'est la rivière et dont les paramètres entrants diffèrent d'une prise à une autre. En effet, cela s'explique du fait de l'environnement non contrôlé du milieu naturel. Il fallait, dans ce contexte, se baser sur une entrée d'eau à paramètres les moins fluctuants possibles, en fermant le circuit d'eau et re-bouclant la sortie à l'entrée.

En effet, l'eau rentrant dans l'installation est sous forme de circuit ouvert, comme représenté en figure II.5 :

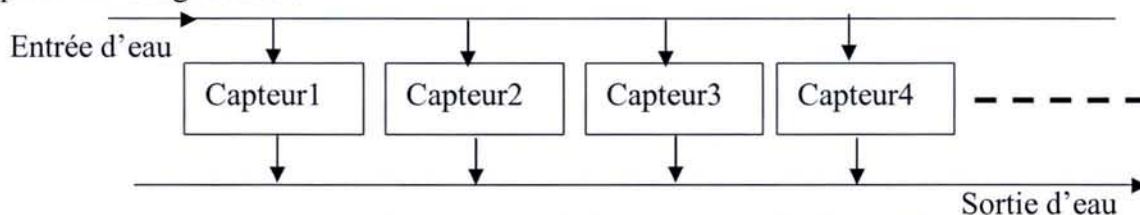


Figure II.5 : « Description du circuit d'alimentation »

Il est utile de préciser la raison de la disposition des capteurs sous forme d'architecture parallèle. Celle-ci réside dans l'intérêt de confronter les réactions de chacun d'entre eux en même temps, garantissant ainsi une simultanéité synchronisée de la mesure. Cette notion d'architecture parallèle est en l'occurrence développée au point suivant.

## I.4 Capteur intelligent et système multi-capteurs

### I.4.1 Le capteur intelligent

Le capteur intelligent correspond principalement à l'intégration dans le corps du capteur d'un organe de calcul interne (microprocesseur, micro-contrôleur), d'un système de conditionnement du signal (programmable ou contrôlé) et d'une interface de communication. Plus largement, le concept de capteur intelligent se décompose ainsi [ROBE-1993]:

- un ou plusieurs transducteurs,
- des conditionneurs spécifiques,
- d'une mémoire,
- d'une alimentation,
- d'un organe intelligent interne permettant un traitement local et l'élaboration d'un signal numérique,
- d'une interface de communication.

Ces composants sont reliés selon l'architecture de la figure II.6 :

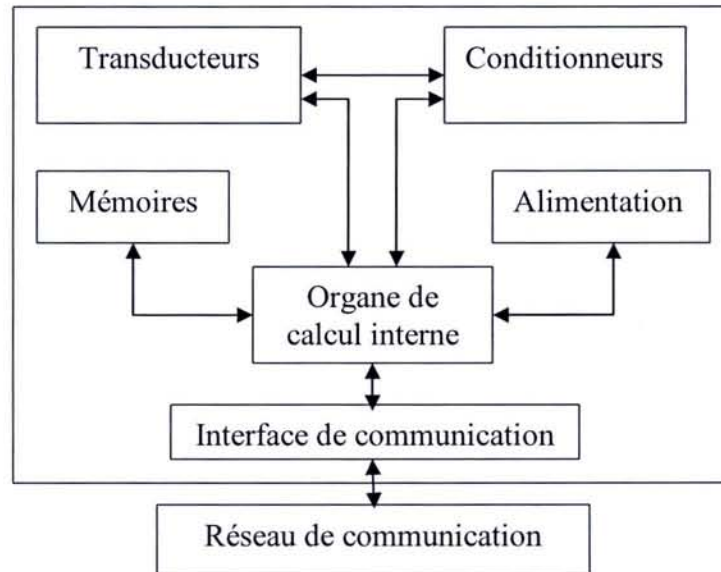


Figure II.6 : « Architecture générique de capteur intelligent [ROBE-1993] »

Ces fonctionnalités permettent de déterminer les caractéristiques d'un système multi-capteurs.

Dans notre optique d'acquisition de l'information et de décision en vue d'une action, notre système est assimilable à un S.A. (système d'automatisation) composé à la fois d'éléments matériels et logiciels comme décrit dans [GNAE-1991], à savoir :

- les dispositifs de traitement : capteurs, actionneurs, régulateurs, automates, ...
- les dispositifs de mémorisation : mémoires, disques, ...
- les dispositifs de communication : lignes, réseaux, ...

Le système peut être scindé, de manière fonctionnelle en quatre sous-ensembles : système d'information, système de décision, système de communication, système d'application [CIAM-1991], représenté en figure II.7 :

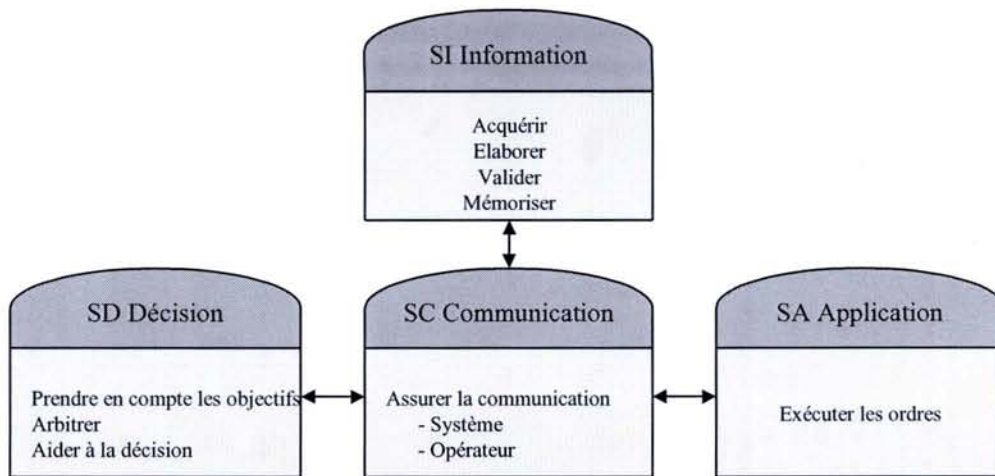


Figure II.7 : « Décomposition d'un S.A. en sous-systèmes »

Le système d'information acquiert, valide et mémorise l'ensemble des informations nécessaires au fonctionnement du S.A. et gère la base de données validée, utilisée par le système de décision.

Le capteur est un des constituants du système d'information ; à minima, tout capteur est un producteur qui alimente une base de données gérée par le système d'information, à partir d'informations caractérisant son environnement physique, émanant du processus et d'une source de temps.

Le système de décision décide du mode de fonctionnement du S.A. et/ou seconde l'opérateur dans ce choix, en fonction de la stratégie et des objectifs émanant des niveaux supérieurs à savoir conduite, maintenance, gestion technique et gestion de production.

Le système de communication garantit la circulation du flux d'informations à l'intérieur du système ainsi que l'interfaçage avec l'extérieur ; le réseau de terrain et l'interface homme-machine sont les constituants du système de communication.

Le système d'application agit sur le processus en fonction des consignes du système de décision.

Seul le capteur permet d'obtenir une représentation de l'état du processus. Cette représentation doit être la plus précise possible afin qu'elle n'engendre pas d'action erronée sur le processus ; Or les mesures peuvent être entachées d'erreurs dont les sources peuvent être nombreuses [NOIZ-1991]. Aussi, il est nécessaire de leur associer un degré de confiance : la **crédibilité**.

Cette notion de crédibilité [LEBO-1989] caractérise l'aptitude du dispositif à délivrer dans ces conditions de fonctionnement, des informations de mesure justifiant un degré de confiance requis. On procède à leur validation, d'une part par confrontation des mesures identiques par redondance matérielle et d'autre part par confrontation des mesures différentes par redondance analytique. Il est cependant nécessaire de choisir le capteur en fonction de ses caractéristiques métrologiques et de ses critères de sûreté de fonctionnement.

#### 1.4.2 Systèmes multi-capteurs

Dans le cadre de notre démarche, un système multi-capteurs peut être considéré comme étant un système d'acquisition de données brutes de part l'aspect individuel de leurs acquisitions, en vue de produire une information fusionnée, alimentant un centre décisionnel, qui à son tour actionne une ou plusieurs commandes adéquates. C'est généralement le cas

lorsqu'on ne dispose pas d'un capteur générant une information directement proportionnelle au paramètre que l'on veut mesurer.

Les données acquises par des systèmes multi-capteurs fournissent les moyens de définir une base de connaissance et le scénario d'une situation donnée. Les données y sont validées et intégrées en tant que connaissance fusionnée. Cela est possible à partir de la mise en forme du niveau du signal, qui le rend exploitable ou compréhensible par l'utilisateur.

On trouvera sur la figure II.8 la représentation symbolique d'un système multi-capteurs évoluant dans un environnement dont la stabilité est incertaine [GEE-2000], en l'occurrence celui d'un site nucléaire à démanteler pour lequel nous en faisons l'analogie avec l'eau étudiée dans notre station multi-capteurs.

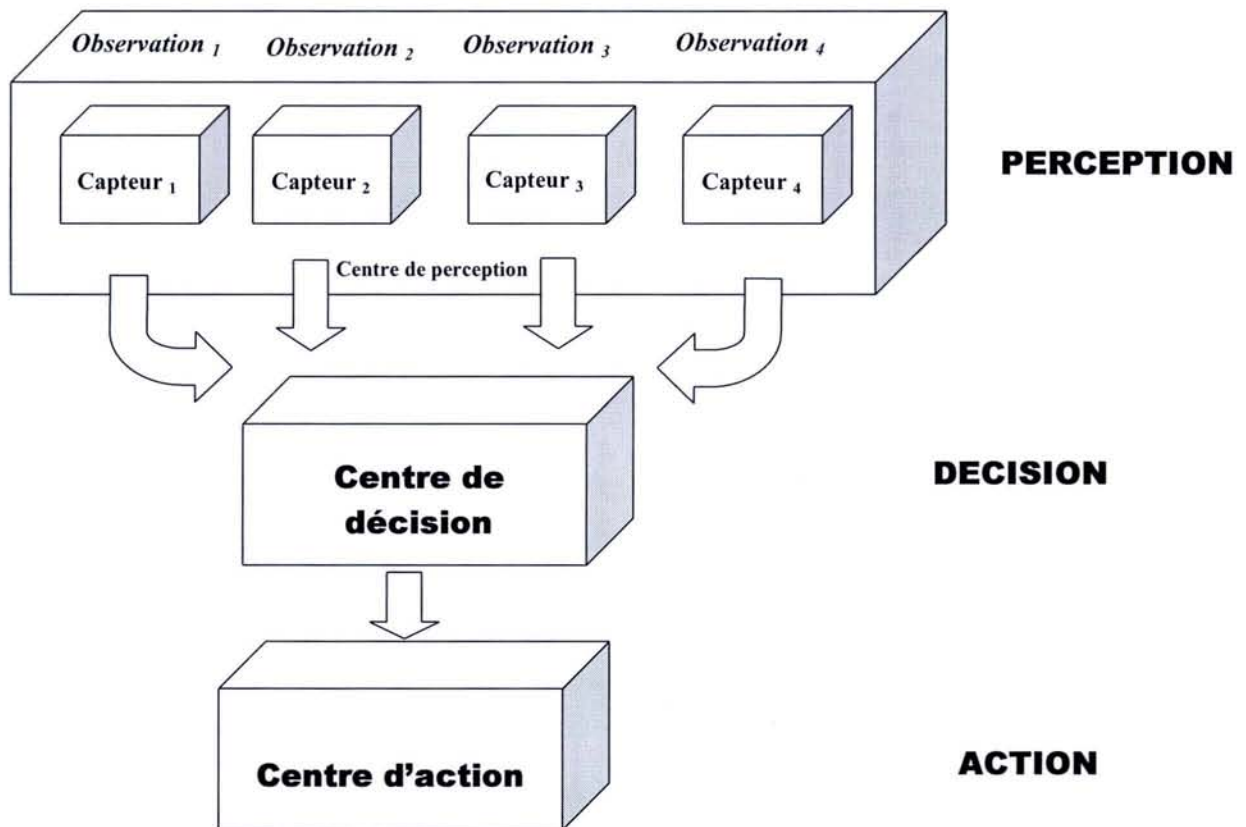


Figure II.8 : « Configuration du système multi-capteurs »

L'approche multi-capteurs en question, permet de faire en sorte que l'information de rétroaction soit transmise du centre de décision afin que les actions inhérentes dépendent en temps réel du centre de perception. Dans le contexte en question, il s'agit de télécommander un robot en fonction des observations fournies par les capteurs de perceptions. Le but est la conduite d'identification de scène et d'exécuter l'activité de démantèlement basée sur l'instruction reçue du centre de décision qui à son tour dépend de la perception acquise.

La disposition des capteurs en parallèle est indispensable dans ce contexte dans le sens où le centre de décision exige un recueil synchronisé des informations, contrairement à une disposition en série où la problématique de la décision est distribuée, ce qui implique un sondage périodique des capteurs [VISW-1998].

L'analogie entre notre système et celui du système de démantèlement du site nucléaire se justifie, de part le fait que nous considérons l'ensemble des capteurs comme étant des éléments de perception de l'environnement. D'autre part, leur disposition en parallèle répond aux mêmes exigences d'une prise décision basée sur un processus de collecte

d'informations en parallèle et synchronisé. Le centre de décision reçoit, fusionne et synthétise les informations fournies et en fonction des alertes, les actions sont soit préventives soit de sécurité. Comme mentionné en introduction du chapitre, le délai de traitement de l'information n'excède pas le temps d'arrivée de l'eau à l'usine de traitement, correspondant à la demi-heure.

Rappelons qu'en terme de bassin de sécurité, il faut en distinguer deux types, l'un correspond au grand bassin sous forme de lac artificiel mitoyen de la station de pompage et dont l'autonomie est d'une semaine pour notre agglomération et d'autre part le bassin intermédiaire de réserve situé à l'usine d'eau et dont l'autonomie est moindre.

La figure II.9 montre l'analogie faite dans notre contexte par rapport à l'architecture du site de démantèlement nucléaire :

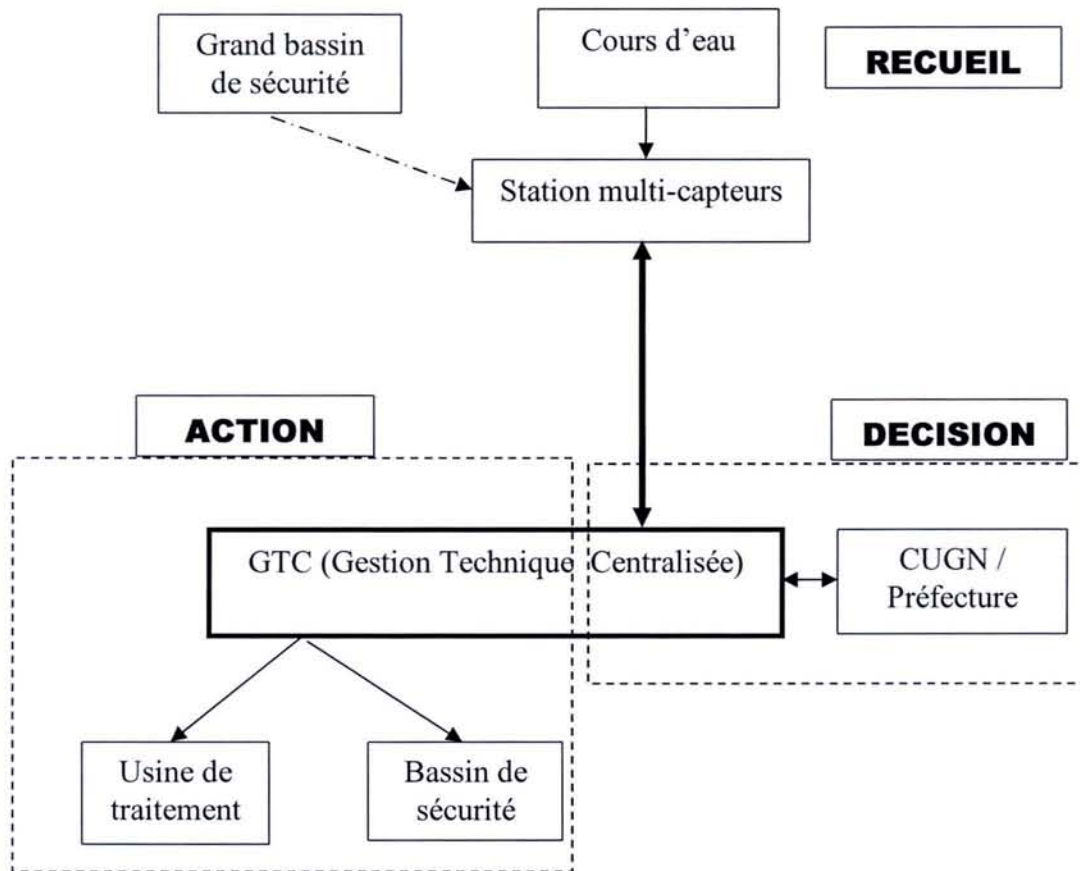


Figure II.9 : « Configuration analogue dans notre contexte »

De manière évidente, la station multi-capteurs équivaut au centre de perception. La particularité de notre contexte, est le rôle central que joue la GTC qui perçoit d'une part les données de la station de multi-capteurs, d'autre part fait partie du centre de décision en collaboration avec la CUGN/ Préfecture, et enfin coordonne les actions à entreprendre auprès de l'usine de traitement, avant de décider au cas extrême le basculement vers les bassins de sécurité. Il faut observer que l'action de sécurité, est peu fréquente, mais il suffit d'imaginer les effets dévastateurs autant d'un point de vue santé publique qu'économique que peut engendrer une pollution conséquente non dérivée à temps, pour justifier la mise en place d'un tel système.

### I.4.3 Capteurs intelligents dans le cadre d'une architecture multi-capteurs

L'idée de départ de la mise en place de notre chaîne de capteurs, était basée sur la reconstitution d'une suite d'êtres vivants se nourrissant les uns les autres dans la chaîne alimentaire [AZZ-2001], ce qui correspond à la chaîne trophique depuis la bactérie jusqu'aux poissons. L'intérêt est d'évaluer l'effet de la pollution aux différents niveaux de cette chaîne alimentaire. L'étape qui s'en est suivie, a consisté à rechercher une chaîne la plus représentative possible de la Moselle, cours d'eau à évaluer.

Le choix de cette suite d'êtres vivants en tant que bio-capteurs plus sensibles au milieu naturel que les capteurs classiques, a pour but de contribuer à prévenir l'apparition d'une pollution plus précocement et de manière globale. Les avantages à utiliser des organismes vivants sont nombreux, en l'occurrence :

- rapidité de détection car le seuil de détection des capteurs est faible,
- élargissement de la gamme des substances pouvant être détectées,
- possibilité d'identifier des pollutions propres à chaque compartiment de l'écosystème aquatique, car chaque espèce utilisée a une sensibilité plus prononcée pour une catégorie de toxique.

En effet, le moindre changement du comportement « normal » du bio-capteur peut être considéré comme un indicateur de pollution. Ainsi, la finalité de notre étude conduit à quantifier et qualifier cette pollution en la confortant par les mesures issues des capteurs physico-chimiques. De ce fait, la prévention s'enrichit par la complémentarité de la surveillance classique physico-chimique associée à la surveillance biologique. C'est dans ce contexte, que l'on conçoit l'approche multi-trophique en tant que démarche novatrice globale, qui permet d'appréhender les changements qualitatifs des écosystèmes aquatiques en général.

Si l'on considère effectivement que chaque organisme biologique est doté d'un profil de sensibilité toxicologique qui lui est propre, on conçoit aisément que le recours à différents taxons (**voir glossaire**) biologiques engage la surveillance dans un champ de détection élargi des dégradations potentielles des milieux hydriques. Pour ce faire, cinq réactifs biologiques, représentant cinq maillons différents de la chaîne alimentaire ont été sélectionnés [METH-2003] :

- 1) Les bactéries, parmi les organismes du niveau trophique des **décomposeurs**. Elles minéralisent la matière organique, alors disponible pour les organismes hétérotrophes (**voir le glossaire**).
- 2) Les algues, représentant le niveau trophique des **producteurs primaires**. Ces organismes photosynthétiques exploitent les substances minérales pour la transformer en matière organique.
- 3) Les daphnies, parmi les premiers maillons de la chaîne des **consommateurs**. Ce sont des herbivores.
- 4) Les moules, parmi les seconds maillons de la chaîne alimentaire. Elles se nourrissent de phytoplancton et de bactéries.
- 5) Les poissons, parmi les derniers maillons de la chaîne trophique.

En terme de sensibilité connue, les Daphnies (puces d'eau) sont plus sensibles au phénol et la turbidité, les algues aux herbicides, les poissons au cyanure, les moules à la conductivité et l'oxygène dissous, et les bactéries aux hydrocarbures. D'autres variations de

paramètres et substances sont aussi détectées par ces organismes, c'est ce que nous avons démontré tout au long des tests de toxicité détaillés au prochain chapitre.

La démarche a ensuite consisté à retenir 5 bio-capteurs industriels impliquant ces différents taxons (**voir le glossaire**). Ces êtres vivants, étant chacun représenté par un capteur biologique, produisent un mesurande en fonction de réactions physico-chimiques s'opérant sur la qualité de l'eau, d'où le couplage de cette chaîne trophique aux capteurs physico-chimiques, classiquement utilisés de manière individuelle pour la surveillance de la qualité de l'eau. L'intérêt de la démarche est d'intercepter des phénomènes de pollutions bénins en exploitant la forte sensibilité des organismes vivants face aux variations du milieu naturel. C'est de là que naît la notion de pré-alerte, qui peut être autant quantifiée par les grandeurs biologiques que physico-chimiques associées, d'où l'appellation station d'alerte multi-paramètres.

Les exigences attendues d'un bio-détecteur dans notre contexte, font qu'elles doivent répondre à une détection précoce quelle que soit la nature du milieu surveillé, en l'occurrence :

- La capacité des organismes à rester sous les conditions des tests de toxicité, sans qu'ils en pâtissent dans leur comportement.
- Une **"bonne" sensibilité** des animaux-tests. Ces derniers doivent être capables de réagir rapidement à une large gamme de polluants et à de faibles seuils de concentration, c'est ce qui est observé tout au long de tests de toxicités explicités dans les prochains chapitres.
- L'exploitation d'une réaction biologique qui autorise une **mesure rapide, fiable, automatique**, avec un transfert à distance des données et la possibilité de déclencher une alarme.
- **L'absence ou la maîtrise d'influences extérieures**, hormis les variations de la qualité de l'eau, sur la réponse biologique développée par l'organisme.
- **La facilité d'obtention des organismes tests** (contrôle de leur élevage), ainsi que la possibilité d'une standardisation de ces derniers.

L'ensemble des capteurs est interconnecté de manière à ce que le rapatriement et le traitement de données puissent se faire de manière synchrone et centralisée, comme détaillées en « Annexe II » et symbolisées comme suit :

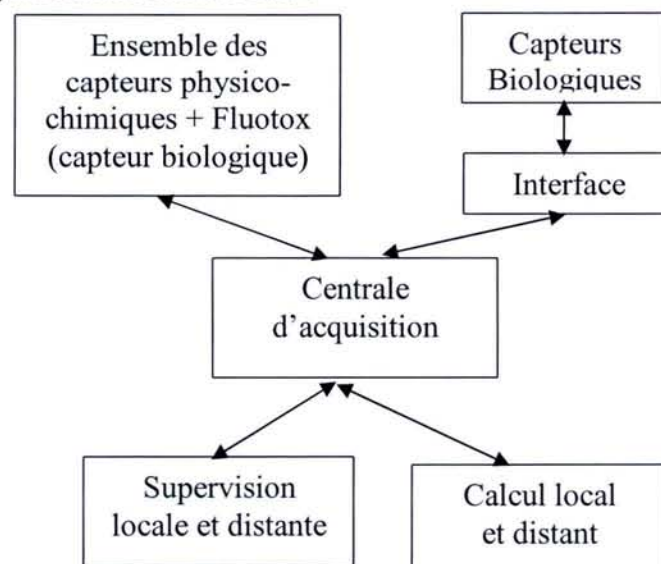


Figure II.10 : « Description générale du dispositif de la station »



Comme indiqué en figure II.10, l'installation a été conçue en fonction des contraintes et architectures de transfert de données des différents capteurs. Cela s'explique d'une part par l'existence de capteurs possédant leurs propres systèmes de conditionnement et de transfert de données, et d'autre part de capteurs dépendant d'un PC interface intermédiaire effectuant le conditionnement et l'envoi de données. Les aspects modes de transmissions et centrales d'acquisition sont mentionnés en fin de chapitre et détaillés en « Annexe II».

Il en ressort en outre, qu'une installation de ce type est assimilée à ensemble de capteurs intelligents, dont les caractéristiques principales apparaissent, en l'occurrence :

- la communication et le transfert de données,
- l'exploitation partagée de données pour un résultat unique,
- plusieurs transducteurs et conditionneurs spécifiques,
- un organe intelligent interne permettant des traitements locaux et l'élaboration d'un signal numérique.

Dans cette optique, nous mettons en évidence ci-après, de manière individuelle, les interfaces et mode de fonctionnement de chacun des capteurs de l'installation.

#### **I.4.3.1 Capteurs biologiques**

Le choix de ces capteurs s'est fait sur la base de la chaîne trophique afin d'évaluer l'impact d'une pollution au niveau de chaque maillon de la chaîne alimentaire, et d'autre part de pouvoir couvrir une large champs de sensibilité en cas de pollution. Chacun de ces capteurs est équipé d'une interface PC qui intègre leur propre logiciel de supervision. L'ensemble de ces capteurs inclut :

- Le Bactéria-test : étude de la luminescence bactérienne.
- Le Fluotox : analyse de l'activité photosynthétique d'algues.
- L'Aquatox : exploite l'activité de nage des daphnies.
- Le Dreisseina : exploitation de l'activité valvaire des moules.
- Le Gymnotox : détection des signaux émis par des poissons produisant des ondes électromagnétiques

##### **I.4.3.1.1 Le Bactéria-test :**

Cet analyseur de toxicité mesure le pourcentage d'inhibition de la luminescence qui varie en fonction de la pollution. Le principe de ce test se base sur l'émission de luminescence de bactéries aquatiques luminescentes, la toxicité entraîne ainsi une chute de l'émission de luminescence.

Les principales spécifications du *Bacteria-test* se caractérisent par :

- la fréquence de mesure : 1 par heure,
- la durée de la mesure : 30 minutes,
- le seuil d'alerte programmable de 0 à 100%,
- le diagnostic d'erreur automatique,
- la télétransmission des données.

Les mesures de toxicité sont établies par rapport à un témoin (eau de référence non toxique). Le protocole du test d'inhibition de la luminescence bactérienne est basé sur le principe de comparaison de la figure II.11 :

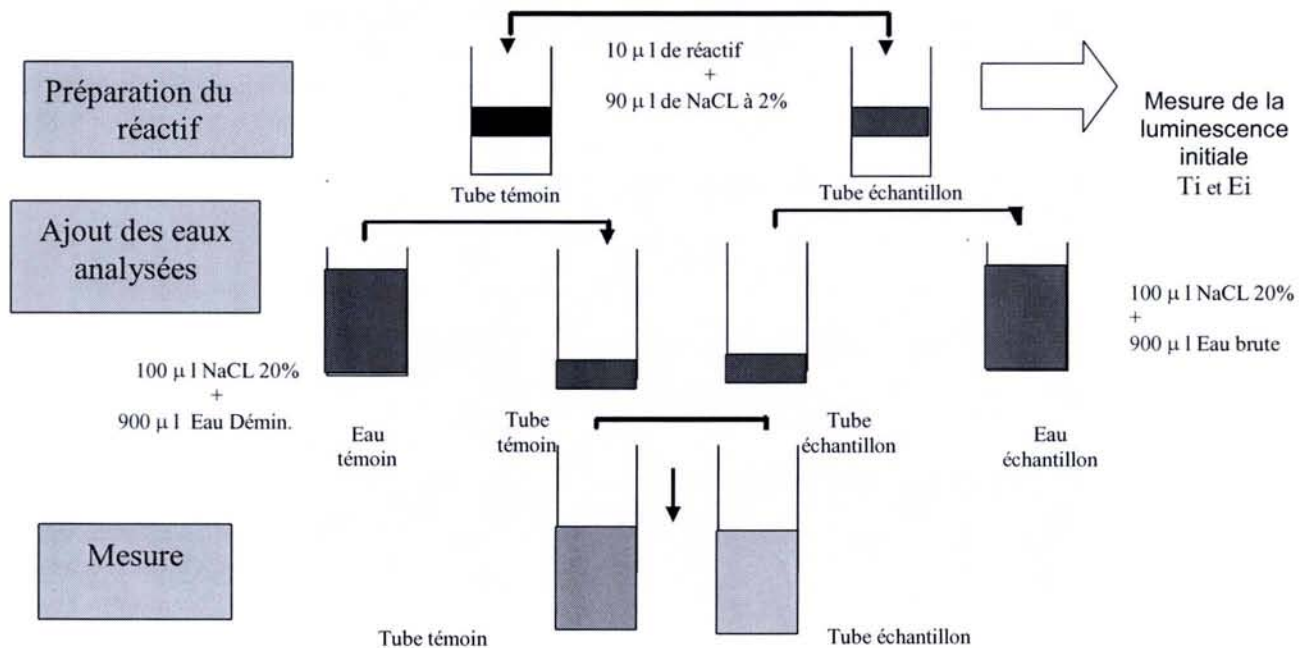
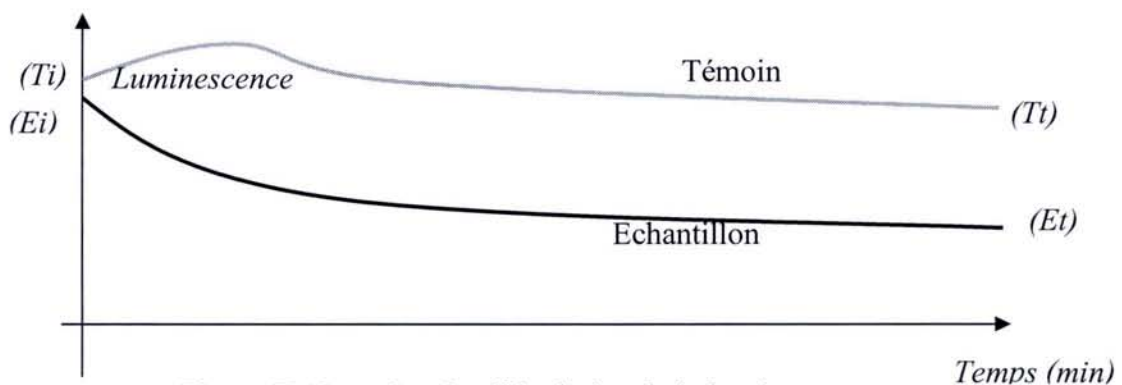


Figure II.11 : « Mesure de la luminescence de chacune des solutions »

Ce descriptif correspond au protocole de mesure classique pratiqué en laboratoire.

Les tests précédents se concrétisent comme suit :



Le calcul du pourcentage d'inhibition de luminescence a pour formule :

$$\partial\% = \frac{E_i \times \frac{T_t - E_t}{T_i}}{E_i \times \frac{T_t}{T_i}} \times 100 \quad (2.1)$$

- $E_i$  et  $T_i$  représentent les luminescences initiales respectives des solutions échantillon et témoin,
- $E_t$  et  $T_t$  représentent les luminescences à l'instant « t » respectives des solutions échantillon et témoin.

Ce calcul se fait grâce à l'interface PC directement relié au plateau de travail qui recueille les données.

La description matérielle de cet analyseur est présentée en « Annexe I ».

#### I.4.3.1.2 Le Fluotox :

Le Fluotox exploite l'activité photosynthétique de cellules algales *Scenedesmus Subsicatus*. Les algues sont immobilisées sur une membrane de cellulose, placée dans le courant d'eau à tester. Chaque cellule algale renferme un chloroplaste contenant de la chlorophylle et un corpuscule plus foncé.

Dans le bio-capteur, les algues sont immobilisées sur une membrane filtrante, à travers laquelle passe le courant d'eau dont l'on évalue la qualité, il est en l'occurrence régulé en température pour limiter la présence de parasites. Le bio-capteur réalise une excitation des algues grâce à une source lumineuse (lumière bleue modulée à 400 Hz) permettant de faire émettre aux algues une lumière qui est captée par une phot cellule. La photosynthèse s'effectue alors normalement, en dehors de la présence d'une pollution susceptible d'augmenter la fluorescence au niveau de la membrane.

En effet, lors d'une pollution, la photosynthèse de l'algue est perturbée dans son processus chimique. Il en résulte que l'énergie lumineuse captée par les antennes chlorophylliennes ne participe plus à la chimie de la plante; de ce fait l'énergie solaire captée se transforme en chaleur et en rayonnement de fluorescence. L'amplitude de ce rayonnement indique le taux d'inhibition de l'algue.

Ce type de phénomène est du coup, irréversible pour la survie des algues, nécessitant leur renouvellement immédiat. A titre indicatif, en périodes dites « normales », le renouvellement des algues s'effectue périodiquement à hauteur d'un renouvellement par mois.

Le bio-capteur réalise l'excitation des algues microscopiques chlorophylliennes et le relevé de fluorescence selon le schéma de fonctionnement de la figure II.13 :

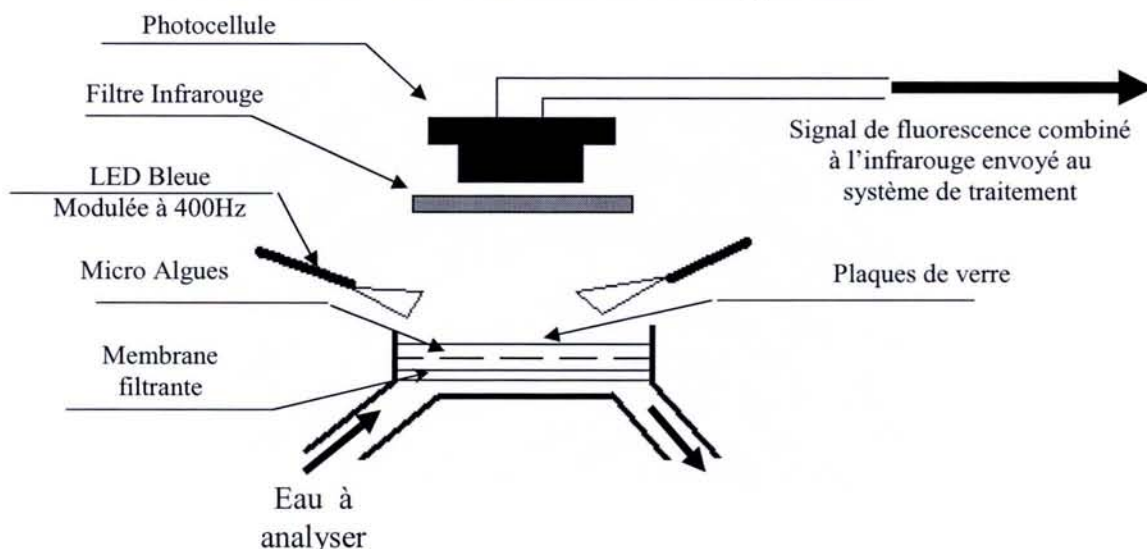


Figure II.13 : « Schéma de principe du Fluotox »

La capture de la luminescence est faite au moyen d'une phot cellule associée à un filtre infrarouge. Le signal de fluorescence est alors envoyé au système électronique où il est immédiatement traité.

### I.4.3.1.3 L'Aquatox :

Cet analyseur intègre des pompes permettant d'alimenter en eau un circuit contenant deux cuves. Celles-ci appelées chambres d'essais contiennent des daphnies préalablement élevées. Les daphnies sont des colonies d'insectes aquatiques qui ont une activité de nage qui varie selon la qualité de l'eau. Le principe de mesure se base sur l'analyse de l'activité de nage des daphnies. Plus exactement, la particularité de la daphnie est sa capacité à réagir à la lumière : elle est photosensible. En présence d'une source lumineuse, elle se dirige vers celle-ci. L'éclairage de ces chambres provoque donc le mouvement des daphnies qui est saisi par des sondes infrarouges. En cas de pollution, leur activité de nage est diminuée. La figure A1.2 de l'«Annexe I» explicite le fonctionnement des chambres d'essais constituées de deux sources lumineuses activées de manière intermittente dans les chambres d'essais.

Le nombre de daphnies introduites dans les chambres d'essais est préalablement fixé. La source lumineuse située au-dessus des sondes infrarouges est activée sur une période de trois minutes durant laquelle les daphnies sont attirées par la source lumineuse. Dans leur mouvement les daphnies coupent les rayonnements infrarouges engendrant un comptage.

Plus l'eau est trouble moins les daphnies sont attirées; c'est ainsi que l'on conçoit une notion relative de pollution pour ce capteur due à un contexte turbide. Après trois minutes, l'éclairage initial est interrompu, au même moment un autre éclairage est activé à l'autre extrémité de la chambre d'essai pendant une période de deux minutes. Cette période s'appelle la phase de repos qui permet de rassembler toutes les daphnies et ainsi de suite le cycle recommence.

Dans le cas de notre installation, la chambre d'essai de référence est utilisée dans le même cadre d'analyse que la chambre de mesure, ce qui rend la mesure redondante permettant de conforter une fiabilité de mesure. Il est à noter que les daphnies sont renouvelées chaque semaine, pour éviter des erreurs de comptage dues à la reproduction. En effet, elles ont tendance à se multiplier après 8 à 9 jours, ce qui fausse les calculs.

### I.4.3.1.4 Le Dreisseina :

Le bio-détecteur *Dreisseina*-Monitor exploite l'activité valvaire des moules [BORCH-1992 ; BORCH-1994]. En l'occurrence, il exploite deux types de réponses que sont le pourcentage de moules ouvertes d'une part et le nombre de mouvements valvaires par moule et par heure d'autre part. Le principe de mesure est décrit selon le schéma de principe de la figure II.14 :

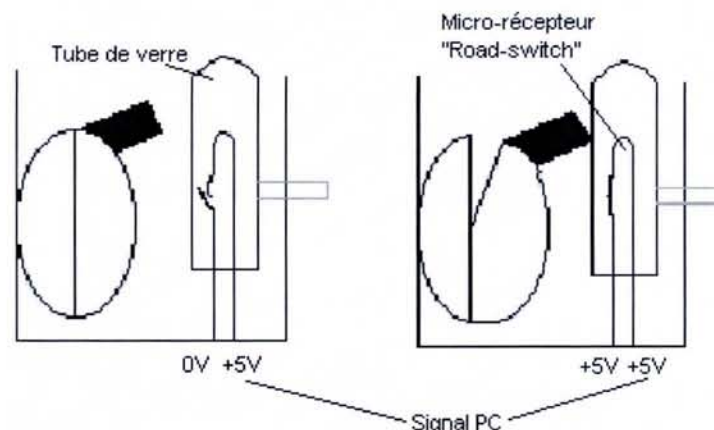


Figure II.14 : « Principe de fonctionnement du Dreisseina »

Chaque moule est reliée à un micro-récepteur actionné par un aimant collé à une des coquilles de la moule fixée, de manière à créer un champ magnétique en fonction de l'ouverture de la moule. Chaque changement d'état du micro-récepteur, est enregistré. L'analyse des informations porte sur 84 moules (deux canaux de 42 moules chacun).

L'interface PC envoie un signal 5V à chaque connecteur, le signal de retour est :

- soit 0 V (connecteur ouvert – moule fermée),
- soit 5 V (connecteur fermé – moule ouverte).

L'analyse des données se fait comme suit :

- l'enregistrement des données sur l'ouverture ou la fermeture des valves,
- calcul du pourcentage moyen de moules ouvertes et nombre moyen de mouvements valvaires (par moule par heure) sur un intervalle de temps constant,
- comparaison de l'analyse enregistrée au comportement normal préalablement évalué, sachant qu'une moule a au moins un mouvement valvaire / 48H. Dans le cas où ce mouvement n'aurait pas lieu, on détecte soit un défaut de connecteur ou la mort de la moule.

De ce constat, il ressort que les valves des moules sont normalement ouvertes pour assurer les échanges alimentaires et respiratoires avec le milieu ambiant. Si une pollution survient, l'animal ferme naturellement ses valves et augmente la fréquence d'ouverture pour tester si le milieu est redevenu propice. Ce réflexe lui permet de s'isoler de conditions défavorables pour assurer correctement ses fonctions physiologiques.

#### **1.4.3.1.5 Le Gymnotox :**

Le Gymnotox est un détecteur biologique de pollution des eaux. Il exploite les signaux électriques naturellement émis par des poissons originaires d'Amérique du sud en l'occurrence l'*Apteronotus Albifrons* [GYMNO-1999]. Ce poisson émet des décharges bio-électriques selon la variation des caractéristiques physico-chimiques des eaux. Ce procédé permet de quantifier toutes variations brutales physico-chimiques induites par une pollution dans l'eau, à partir de l'analyse de la fréquence des décharges électriques du poisson [THOM-1996]. Il possède une activité physiologique particulière capable d'émettre un courant électrique faible, de l'ordre de quelques millivolts. Cela s'explique par le fait qu'il possède des électro-récepteurs et d'un organe électrique à l'origine d'un signal ondulatoire faiblement électrique mais suffisant pour y discerner une variation de comportement. D'autre part sa fréquence est élevée (entre 750 et 1250 Hz). Les signaux électriques émis par le Gymnotox sont transmis via les électrodes à une carte d'acquisition d'un PC. La forme des décharges électriques est quasi-sinusoïdale, et ne présente aucune discontinuité entre deux périodes consécutives.

Le Gymnotox est une sorte d'aquarium fonctionnant en circuit ouvert. Il est alimenté régulièrement par une eau réchauffée, à raison de 100l/h. Elle est thermorégulée avec une précision de +/- 0.1°C, sachant que ces poissons tropicaux ne supportent pas l'eau froide et de ce fait, un régulateur de température la maintient aux alentours de 25°C. L'eau ainsi régulée, alimente alors 3 bacs de contention (avec tubes reliés à des électrodes), renfermant chacun un poisson. En effet, un tube en PVC transparent est placé dans chaque bac de contention. Les poissons électriques s'y logent très spontanément comme c'est souvent le cas en milieu naturel. Ce tube joue plusieurs rôles : il sécurise le poisson d'une part, et facilite le recueil des signaux bioélectriques d'autre part. En effet, il favorise le positionnement du poisson dans le plan tracé par les électrodes de captage des signaux.

En terme de principe de fonctionnement la figure II.13 montre le circuit que parcourt l'eau dans le GYMNOTOX selon les étapes suivantes :

- l'eau pénètre dans le circuit du **GYMNOTOX** par l'intermédiaire d'une vanne manuelle. Elle s'achemine ensuite vers une vanne automatique qui régule le débit et la pression de manière constante,
- le surplus de pression et de débit est assumé par la vanne de décharge, située entre la vanne manuelle et la vanne automatique,
- à la sortie de cette dernière l'eau est thermorégulée grâce à un réchauffeur constitué de résistances chauffantes,
- l'eau alimente alors le premier bac des trois bacs de contention, contenant l'espace de bio-détection, d'un volume unitaire de 15 litres. Ces bacs sont adaptés à des poissons de 8 à 15 cm. Sur le couvercle de chaque bac est greffé un tube PVC transparent, ce qui permet une surveillance visuelle du comportement des poissons. Ces tubes permettent de sécuriser le poisson qui se place spontanément à l'intérieur. Cette configuration facilite en outre la récupération des signaux électriques.

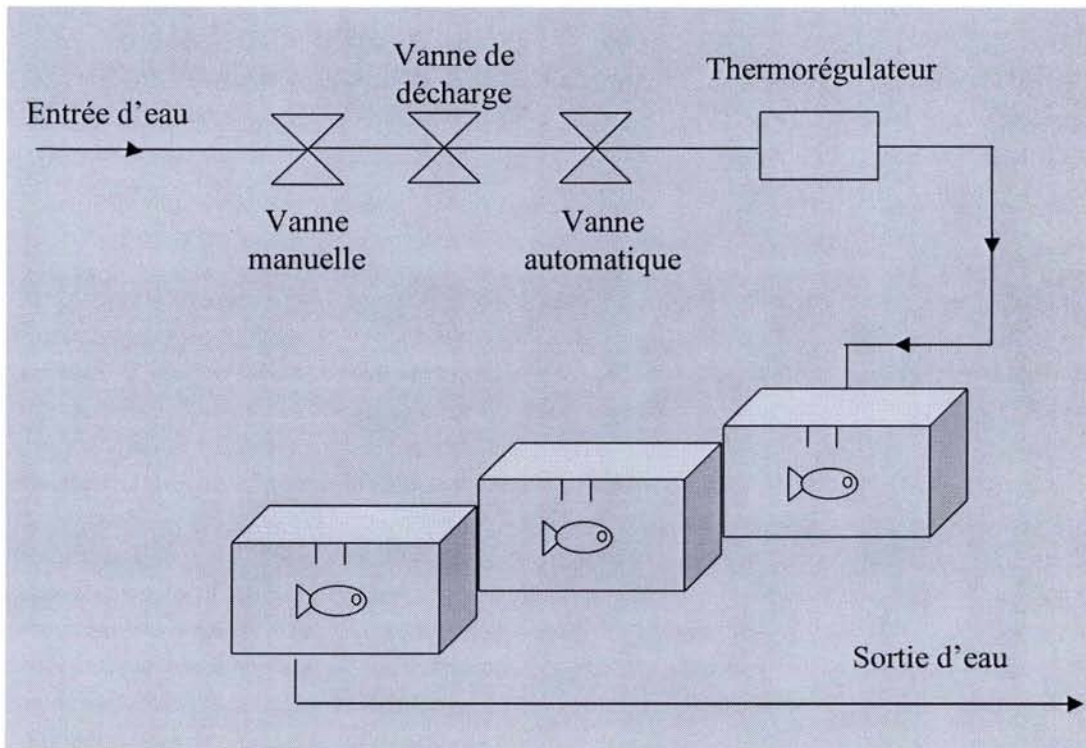


Figure II.15 : « Schéma simplifié de fonctionnement du GYMNOTOX »

- Deux tubes plongeants sur chaque couvercle servent à maintenir :
  - les entonnoirs pour assurer le déversement de la nourriture,
  - les sondes de température pour l'auto-calibrage entre la température et la fréquence des poissons.
- Les trois bacs sont implantés de manière à ce que l'écoulement d'un bac à l'autre se fasse en cascade. Le liquide pénètre par le haut et traverse diagonalement le bac jusqu'en surface pour alimenter le bac suivant. L'eau est brassée par les bulles diffusées dans chaque bac.

En cas de pollution, la variation de fréquence est un indicateur de la variation de la qualité de l'eau. La particularité de ces poissons est leur forte sensibilité multi-risques qui couvre ainsi un large panel de polluants.

#### **I.4.3.2 Capteurs physico-chimiques**

Ce sont des capteurs de type analogique disposant d'afficheurs numériques et reliés au système de supervision/recueil de données. L'ensemble de ces capteurs inclut :

- une sonde multi-paramètres (t°, pH, conductivité, O<sub>2</sub>, potentiel Redox),
- un turbidimètre autonettoyant (information générale sur la qualité de l'eau),
- deux fluorimètres (détection des hydrocarbures, solvants, alcools),
- un analyseur Nitrate/Ammonium.

##### **I.4.3.2.1 La sonde multi-paramètres (t°, pH, conductivité, O<sub>2</sub>, potentiel Redox)**

Ce transmetteur multi-paramètres dispose de réactifs, d'électrodes et d'un thermomètre permettant d'enregistrer les mesures correspondantes aux trois grandeurs recueillies par ce multi-capteur. Les gammes de mesures sont représentées sur le tableau suivant :

<b>Grandeur (résolution et précision)</b>	<b>Gamme</b>	<b>Calibration</b>
pH (0,1)	2,00 à 12,00	4 – 20 mA
°C (1°C)	0°C à 50°C	4 – 20 mA
Conductivité (salinité)	0-1999 µS/cm	4 – 20 mA
Oxygène dissous	0-60 mg/L	4 – 20 mA
Redox (100mV)	0 à 400mV	4 – 20 mA

Tableau II.1 : «Gamme et calibration de la sonde»

Les gammes peuvent être réglées à 20% au-delà ou en dessous des valeurs seuils. Les trois grandeurs sont continuellement affichées sur leur afficheur respectif. Une alarme peut être réglée en fonction d'une valeur pré-programmable. Les données 4-20 mA sont transmises vers le superviseur à travers la centrale d'acquisition.

##### **I.4.3.2.2 Le turbidimètre :**

###### **Définition de la turbidité :**

La turbidité correspond à la caractéristique d'un liquide contenant des matières en suspension et/ou des composés chimiques colorés qui en atténuent la transparence. Cela correspond à la réduction de la transparence d'un liquide due à la présence de particules en suspension. Elle se mesure en faisant passer un faisceau lumineux à travers l'échantillon à tester, elle est ainsi déterminée en mesurant la lumière qui est diffusée par les particules en suspension.

## Le capteur :

La turbidimétrie est régie selon deux normes :

- la norme américaine qui préconise l'utilisation d'une source lumineuse visible, dans ce cas l'unité est le NTU (Nephelometric Turbidity Unit),
- la norme européenne qui préconise l'utilisation d'une diode infrarouge, qui présente l'avantage d'être moins sensible à la couleur des échantillons, dans ce cas l'unité de mesure est le FTU (Formazine Turbidity Unit), qui correspond à la norme de notre station. La gamme de mesure s'étale de 0 à 500 FTU correspondant à un courant de sortie de 4 – 20 mA.

L'unité de commande à microprocesseur affiche en continu la turbidité de l'échantillon d'eau. Les mesures fournies par le turbidimètre sont transmises à la centrale d'acquisition qui les transmet au PC superviseur.

Le schéma de principe représenté en figure II.16 illustre la méthode de mesure de la turbidité :

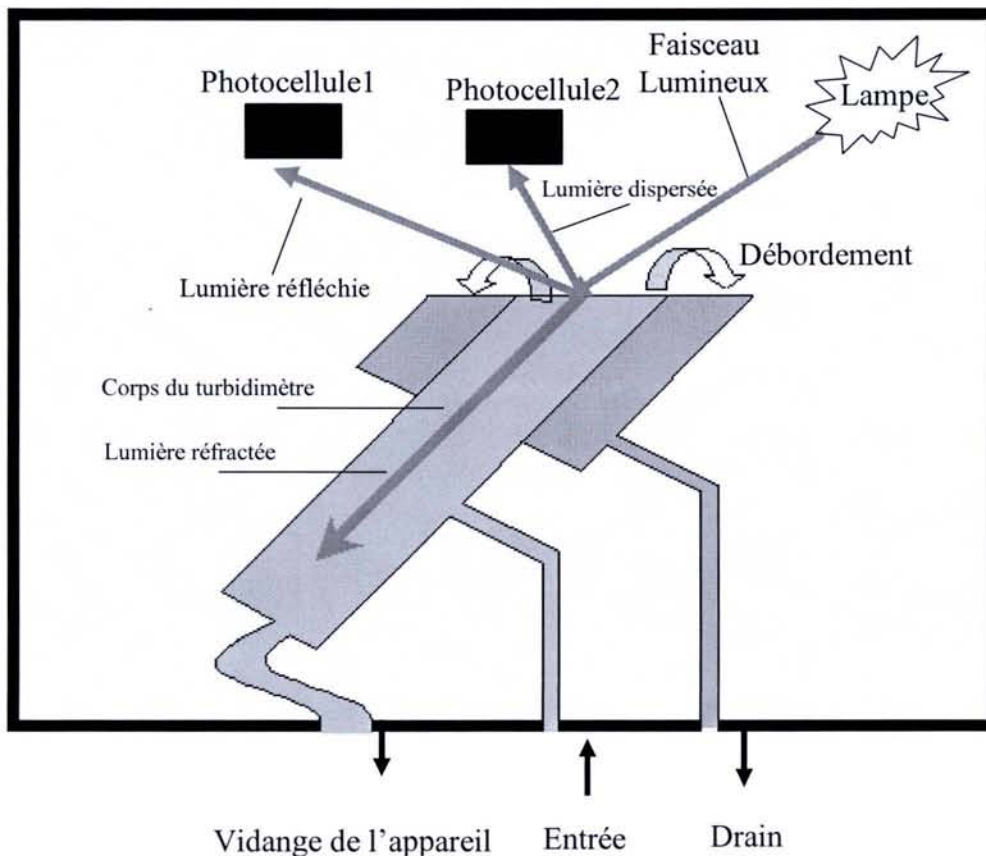


Figure II.16 : « Schéma de principe d'un turbidimètre »

Il est utile de préciser que ce schéma, diffère sensiblement d'un modèle à un autre, sans pour autant modifier le principe.

L'échantillon d'eau à évaluer transite au centre d'un tube incliné, qui alimente à son tour un récipient à débordement, spécialement conçu pour produire l'écoulement d'un jet régulier. Ce récipient est disposé autour du tube central, formant ainsi une surface optiquement plane. Cela permet d'une part d'éliminer d'éventuelles bulles d'air, d'autre part



de produire par le trop plein une charge hydraulique constante, assurant ainsi la vidange régulière de l'eau.

Ainsi afin de calculer la turbidité, l'intensité de la lumière dispersée fait un angle de 90 degrés par rapport au trajet de la lumière incidente. A cet effet, un faisceau modulé d'une source lumineuse infrarouge est projeté sur la surface plane pénétrant l'eau du récipient dans l'axe du jet d'eau sous un angle aigu. Ce faisceau ainsi réfléchi, produit la diffusion de lumière, mesurée par deux récepteurs sous 90°, disposés à deux hauteurs différentes. La lumière ainsi recueillie par ces photo-récepteurs correspond à la mesure de la turbidité. L'évaluation du rapport des deux mesures permet de corriger les effets d'une éventuelle coloration de l'eau.

La turbidité caractérise la réduction de la transparence d'un liquide due à la présence de matières non dissoutes mesurées à un angle de 90° par rapport à la lumière incidente (néphélométrie). La mesure de cette propriété traduit essentiellement la présence de matières en suspension minérales (à l'occasion de crues) ou organiques (pollution ou développement de phytoplancton). Une turbidité inférieure à 30 NTU caractérise une eau de très bonne qualité pour ce paramètre.

Lors d'une crue, la turbidité peut atteindre momentanément des niveaux très élevés (>1000 NTU) sans pour autant que cela se traduise par une détérioration systématique de la qualité compte tenu du caractère bref de cette détérioration {3}, {4}.

En terme d'étalonnage, le turbidimètre utilisé dans notre installation, le WTM 500, est étalonné en usine à la Formazine. Afin d'éviter des manipulations ultérieures de Formazine, un bloc de référence en verre est incorporé dans l'appareil, aux caractéristiques de diffusion définies et stables. Sa valeur est mémorisée et indiquée dans le certificat d'étalonnage de l'appareil.

Il est par la suite question d'un étalonnage interne automatique à intervalles programmables, réglé par l'utilisateur. Cet étalonnage s'effectue par mélange des réactifs qui constituent alors les étalons de l'appareil. L'étalonnage du turbidimètre se fait automatiquement en interne chaque jour à heure programmée et après chaque maintenance du bol de réception.

#### **I.4.3.2.3     *Le Fluorimètre:***

Cet appareil permet de mesurer le dosage de la concentration d'une substance par la mesure de l'intensité de son rayonnement de fluorescence. L'appareil est composé de deux sondes, l'une pour les hydrocarbures et l'autre pour la chlorophylle. Le principe de mesure est le même dans les deux cas, il s'agit de la détection par fluorescence. Celle-ci est produite par l'action d'un faisceau de lumière monochromatique, qui en pénétrant dans un échantillon provoque une émission lumineuse à une longueur d'onde plus grande.

Dans le cas de la sonde hydrocarbure, le faisceau lumineux est de type U.V (ultra-violets), sa longueur d'onde permet de détecter les composants des huiles minérales tel que les hydrocarbures appelés HPA (Hydrocarbures Polycycliques Aromatiques). En effet, ils produisent de la lumière visible lorsqu'ils sont éclairés par les U.V. La photodiode réceptrice recueille la longueur d'onde résultante, permettant ainsi de calculer la quantité d'hydrocarbure selon l'échelle préétablie. L'unité de mesure est le PPM<sup>1</sup> (parties par million) qui s'exprime sous forme de quantité en µg/l de HPA : 1 µg/l de HPA = 1 PPB (parties par milliard) = 1/1000 PPM.

---

<sup>1</sup> A titre indicatif les PPM représente les unités de poids de toute substance donnée par million d'unités équivalentes de poids d'un sol séché à l'étuve; ou, pour une solution de sol ou d'autres substances, unités de poids du soluté par million d'unités de poids de solution, il en est de même pour le PPB {5}.

Dans le cas de la chlorophylle, le principe est le même que celui exposé précédemment. Les différences sont la source lumineuse excitatrice qui est une LED bleue de longueur d'onde de 450  $\eta m$  et la photodiode réceptrice équipée d'un filtre de détection optimisé pour la fluorescence de la chlorophylle A dont la longueur d'onde seuil maximal est de 683  $\eta m$ . L'unité de mesure correspond à la concentration de chlorophylle qui est calée sur l'échelle des longueurs d'onde.

#### I.4.3.2.4 *L'analyseur Nitrate/Ammonium:*

Cet analyseur mesure la concentration de Nitrate et d'Ammonium contenue dans l'eau en mg/l, ces mesures sont possibles à travers des réactifs renouvelables tous les six mois.

Les principales spécifications de cet analyseur se caractérisent par :

- Une précision de +/- 5% ou 0,1mg/l.
- Une échelle de 0,1 à 100mg/L pour l'Ammonium et 0,5 à 100mg/l pour le Nitrate.
- Fréquence de mesure périodique toutes les 2 minutes 30 secondes.
- Quatorze alarmes programmables.
- Une répétabilité de lecture de +/- 3% ou 0.06mg/l.

### I.5 *Acquisition et transfert de données*

Notre installation est composée de plusieurs postes locaux qui assurent l'interconnexion entre les instruments de mesure et les postes de supervision local et distant, tout ceci par l'intermédiaire de bus de terrain et lignes de télécommunication pour les postes distants. En effet la station d'alerte est au centre d'une télésurveillance à partir d'un poste opérateur qui centralise les données de cette station parmi d'autres. Nous pouvons schématiser cette structure par la figure II.17 :

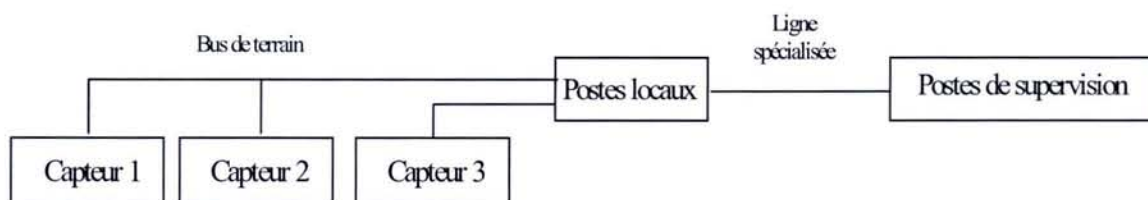


Figure II.17 : « Architecture de communication pour une supervision distante »

Les capteurs de l'installation sont reliés par le bus de terrain représenté sur la figure II.17. Il correspond au MODBUS qui est un protocole de communication qui définit une structure hiérarchique (un maître et plusieurs esclaves). L'accès physique est basé sur une transmission série de type half-duplex. L'interface électrique autorise la connexion en mode point-à-point (RS232) ou mode multipoint (RS485/RS422 et boucle de courant 20mA).

Le système de supervision local est essentiellement destiné à la configuration du S50 (centrale d'acquisition décrite en détail en « Annexe I ») et à l'observation de l'état des différents capteurs, il est également délocalisé vers la GTC (Gestion Technique Centralisée) afin de minimiser les opérations de maintenance sur le site.

La description faite en figure II.18 concerne le « Détail du dispositif de la station d'alerte », mettant en évidence les différents éléments rentrants dans la chaîne de communication ainsi que les modes de transmission. Le détail et la définition des différents éléments composant la station est décrit en Annexe I sur la Figure A1.1 .

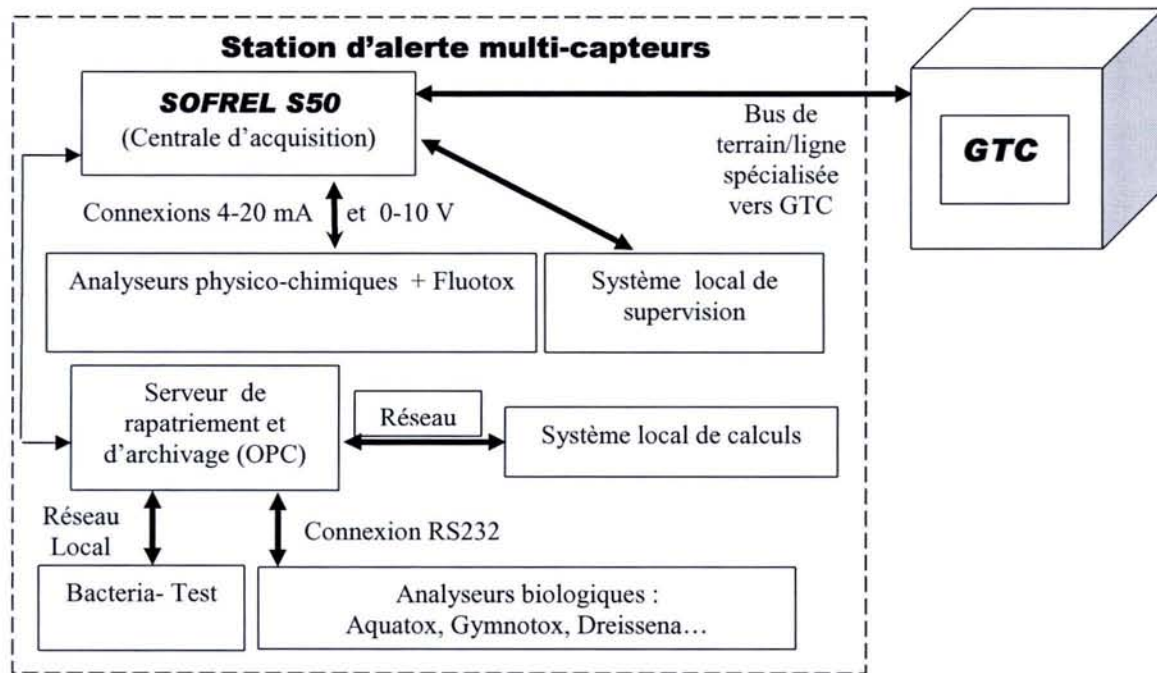


Figure II.18 : « Schéma simplifié de la station d'alerte »

A titre indicatif, la GTC centralise les informations issues d'horizons divers, au-delà de celles concernant notre station multi-capteurs dédiée à la surveillance des eaux de surface. En effet d'autres prérogatives lui sont assignées, en l'occurrence, la gestion des crues ainsi que la surveillance des rejets urbains issus des réseaux d'assainissement. A titre illustratif un schéma de supervision multi-site de la GTC, est décrit sur la figure II.19 :

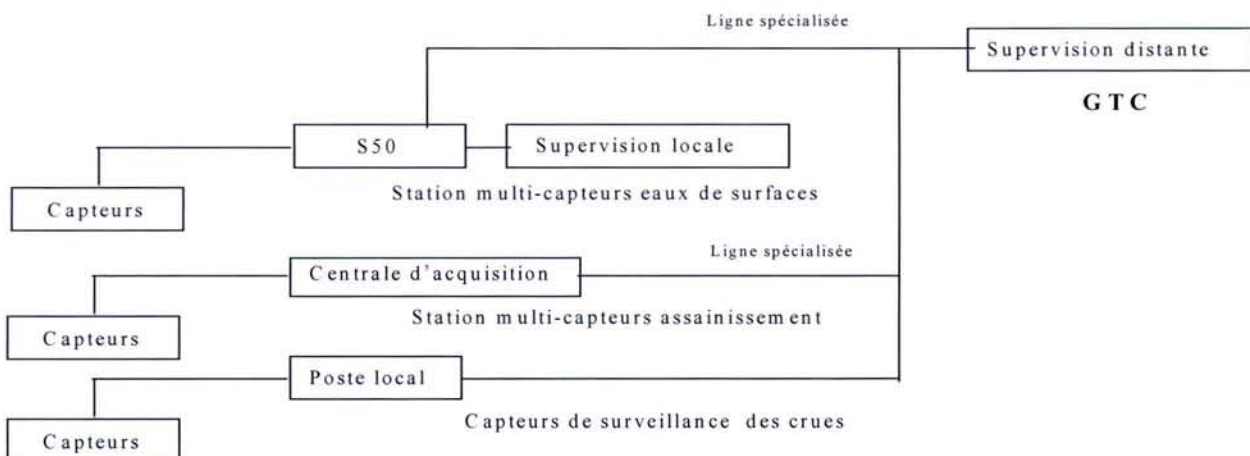


Figure II.19 : « Supervision multi-site »

De ce constat, on se rend bien compte de la complexité de la gestion des données et de l'enjeu en terme de gain de temps que représente la simplification d'une telle gestion. En effet, l'opérateur qui veille à la surveillance du flux d'informations en vue d'activer une éventuelle alerte, a fort à faire face aux nombreux écrans de supervision. Ainsi, si chacune des stations supervisées possède un ensemble de paramètres à surveiller, la tâche s'avère ardue pour la surveillance de plusieurs stations en parallèle. Cela nous amène à mettre en avant l'intérêt de notre approche visant à **synthétiser l'information** accompagnée d'un conseil fourni par **l'aide à la décision**. Ces deux points développés plus en détail dans les chapitres IV et V, cœur de notre méthodologie, sont gérés directement au niveau de la GTC. Le traitement de données inhérent, est synchronisé à partir de la centrale d'acquisition S50, qui centralise, en temps réel, les données issues des capteurs.

SCD UHP NANCY 1  
Bibliothèque des Sciences  
Rue du Général de Gaulle - CS 20148  
54601 VILLERS LES NANCY CEDEX

## 1.6 Conclusion

Notre démarche de mise en adéquation des différents capteurs, n'a pas été si simple dans le sens où il fallait d'abord respecter un choix de capteurs biologiques retraçant la chaîne trophique, d'autre part y associer les capteurs physico-chimiques représentant au mieux les paramètres principaux à relever sur le cours d'eau. Ces choix devaient par ailleurs s'harmoniser avec le réseau de transmission de données, et lorsque ce n'était pas le cas, il fallait trouver des solutions de substitutions comme dans le cas du Gymnotox. Celui-ci avait un mode de transmission propriétaire inadapté pour une installation mutualisée, d'où l'adaptation d'un mode de rapatriement spécifique.

Cette démarche pratique s'inscrivant dans le cadre de la mise en place d'une architecture multi-capteurs, a le mérite de pouvoir fournir un cadre rigoureux pour notre contexte de recherche qu'est la fusion de donnée associée à l'aide à la décision, dont le processus dépend fortement de l'architecture de communication que représente l'installation. En effet, on constate que tous les maillons de rapatriement et de supervision ont un retour vers l'élément transmetteur de l'information, ce qui rentre dans le cadre d'un processus d'amélioration continue plus explicitement détaillé en chapitre V.

L'autre intérêt mis en exergue dans ce chapitre, est la démarche explicative nécessaire pour la mise en place d'une station multi-capteurs de ce type. Elle peut être considérée comme une démarche de référence, de part son aspect novateur sur l'utilisation de l'approche trophique et aussi en termes d'architecture de communication et de rapatriement de données.

En terme de synchronisation, au-delà du fait que l'architecture parallèle garantit une distribution de la même à l'entrée de chaque capteur, la transmission des données par ces derniers se fait à hauteur d'une résolution de rapatriement de cinq minutes, à l'exception du « Bacteria-test » où le test prend une demi-heure à se faire. Il faut aussi préciser que cette eau testée est un échantillon de l'eau pompée et rapatriée vers l'usine de traitement ce qui permet d'avoir une traçabilité précise de l'eau qui sera traitée.

Un autre apport important lié à l'utilisation des capteurs biologiques, en leur qualité de prototypes, est l'exploitation d'innovations permettant un cadre pratique pour leur validation. En effet, les fournisseurs de bio-capteurs ont pu, grâce aux remarques émises tout au long de notre étude, améliorer leur produit, et aboutir à des bio-capteurs commercialisables, d'où l'intérêt de notre étude à promouvoir de telles innovations, notamment, la mise en application de l'un des brevets de NANCIE par rapport au transducteur de base du Gymnotox, qu'est le *poisson électrique*.

Un autre produit de l'innovation locale, est le Fluotox développé en collaboration avec le CRAN-ESSTIN (Centre de Recherche en Automatique de Nancy - Ecole Supérieure des Sciences et Technologies de l'Ingénieur) et la société ARNATRONIC aboutissant à développer un kit d'enseignement pour le traitement des signaux en temps réel [ORY]. Ce bio-capteur a été en l'occurrence exclusivement exploité lors de notre étude, pour sa dernière étape de validation avant commercialisation. Ce capteur a été particulièrement performant de par sa mise au point n'ayant posé aucun problème de fonctionnement, et surtout l'exploitation de ses données sur un maximum de tests de toxicité. Pour le reste des capteurs biologiques, il s'agit de prototypes allemands utilisés pour la surveillance du Rhin. L'innovation apportée ayant été de rassembler tous ces capteurs d'horizons divers, développés dans des cadres individuels et d'en faire, dans notre contexte, une exploitation commune.

Concernant les délais de sécurité, en cas de pollution, nous disposons du temps de latence lié au transfert de l'eau, ainsi que des bassins intermédiaires isolés de l'eau instantanément rapatriée. En cas d'atteinte d'un seuil maximal, le degré d'anticipation est si moindre que les dispositifs de sécurité sont systématiquement utilisés, sans nécessairement avoir la connaissance pertinente. Notre approche du traitement et fusion des données, consiste à trouver un degré de sensibilité intermédiaire, permettant de prévenir l'arrivée d'une pollution en tablant sur l'association des capteurs. Cela permet de disposer de l'information avant la pollution, sous forme de pré-alertes en cours d'évolution avant l'atteinte des seuils d'alertes classiques.

## Références bibliographiques

- [ASCH-1991] Asch G., *Les capteurs en instrumentation industrielle*, Editions DUNOD, 4<sup>ème</sup> édition, 1991.
- [AZZ-2001] Azzam A.A., "Traitement de données d'une station de protection et surveillance multi-paramètres des ressources aquatiques", in: Actes Journées Scientifiques 2001 du GMCEA, p. 196--211, Nancy, octobre 2001.
- [BOIS-1989] Boisseau J.F., *Méthodologie de la mesure*, Techniques de l'ingénieur, Mesures et contrôle, 1989, R140, page 120.
- [BORCH-1992] Borchering J. Another early warning system for the detection of toxic discharges in the aquatic environment based on valve movements of the freshwater mussel, *Dreissena polymorpha*. *Limnologie Aktuell*, vol. 4, pp. 127-146, 1992.
- [BORCH-1994] Borchering J., Volpers M. The 'Dreissena-Monitor' - First results on the application of this biological early warning system in the continuous monitoring of water quality. *Wat. Sci. Tech.*, vol. 29(3), pp. 199-201, 1994.
- [CIAM-1991] CIAME G.T. système, *Fonctionnalités systèmes des constituants intelligents*, Mesucora 1991, Paris Nord Villepinte.
- [GEE-2000] Gee L. A. and Abidi M.A., *Multisensor Fusion for Decision-based Control Cues*, In *Signal Processing, Sensor Fusion, and Target Recognition IX*, Proceedings of SPIE Vol. 4052(2000), Pages 249-257.
- [GNAE-1991] Gnaedinger E., *Analyse fonctionnelle et modèle générique des capteurs intelligents, Application à la surveillance de l'anesthésie*, Thèse, Université Henri Poincaré, Nancy 1, Décembre 1991.
- [KERO-2002] Kéroack F., Guay P., INTRODUCTION à L'ÉTALONNAGE Rencontre de l'ACMV 18 juin 2003
- [LEBO-1989] Lebourgeois F., *Crédibilité de la mesure*, 2<sup>ème</sup> Forum Production Automatisée, Arles 1989.
- [MAZU-2000] Mazuer P. et Matte J.L. (DIREN Lorraine) C. Riou (Agence de l'Eau Rhin-Meuse) et S. Manné (Conseil Supérieur de la Pêche), La qualité des cours d'eau lorrains en 1999, *Exploitation des données du réseau national de bassin*, DIREN Lorraine, Décembre 2000.
- [METH-2003] Azzam A.A., Andreoli A., Thomas M., Cecile J.L., Jonca G., Guide méthodologique du projet Life ENV/F/000492, Protection et surveillance multi-paramètres des ressources aquatiques, NANCIE avril 2003.
- [NOIZ-1991] Noizette J.L., *Méthodologie de conception des capteurs intelligents, Application à un capteur granulométrique*, Thèse, Université de Nancy I, Novembre 1991.
- [NUSS-1987] Nussbaumer H., Informatique Industrielle III, *Automates programmables commande et réglages, capteurs*, Presses polytechniques rommandes, 1987.
- [PARA-1986] Paratte P.A., Robert P., *Système de mesure*, Editions DUNOD, 1986, pages 370.
- [PRIE-1995] Prieur G., Nadi M., La mesure et l'instrumentation, Etat de l'art et perspectives, *Editions MASSON*, 1995, pages 726.

- [ROBE-1993] Robert M., Marchandiaux M., Porte M., *Les capteurs intelligents et méthodologie d'évaluation*, Editions HERMES, 1993.
- [ROBERT-1998] Robert M., Contribution au développement d'une méthodologie d'étalonnage de capteurs : Application à un transmetteur intelligent de température. Thèse, Université de Nancy I, 1998.
- [SELI-1992] Selianov M.N., *Definitions of Metrologie, Quantity and Measurement*, Traduit de Izmeritel'naya Technica, N°2, Février 1992, pages 11 à 15.
- [THOM-1996] Thomas M. Biodétection en temps réel. Contribution à la réalisation d'un nouveau système de détection biologique précoce de la pollution des eaux de surface, fondé sur l'exploitation des décharges auto-stimulatrices d'un poisson électrique tropical : *Apteronotus albifrons* (Apteronotidae - Gymnotiformes). Thèse de doctorat, Université Henri Poincaré, Nancy 1, pp. 1-240, 1996.
- [VISW-1998] R.Viswanathan, Thomopoulos S.C.A., Tumuluti R., "Optimal Serial Distributed Decision Fusion", IEEE Transactions on Aerospace and Electronic Systems, Volume AES-24, numéro 4, juillet 1998.

**Bibliographie technique :**

- [GYMNO-1999] Guide technique du GYMNOTOX, E.R.I (Equipement Réalisation Ingénierie).
- [NF-1984] Norme Française, NF X 07-001, *Vocabulaire international des termes fondamentaux et généraux de métrologie*, Décembre 1984.
- [ORY] Mu.Psi-Fibula, un kit DSP simple et efficace pour l'enseignement de l'industrie. Jean-Marie Ory, CRAN (Centre de recherche en Automatique de Nancy), ESSTIN (Ecole Supérieure des Sciences et Technologies de l'Ingénieur).

BIBLIOTHÈQUE NANCY 1  
 Université des Sciences  
 Rue du Prince Léopold - CS 20148  
 54601 VILLERS LES NANCY CEDEX



## **Références Internet**

{1} Les pollutions par les pesticides : Le rapport de l'Institut Français de l'Environnement IFEN, juillet 2000 : un bilan 1997-1998 sur les pesticides dans les eaux :

<http://perso.wanadoo.fr/francois.veillerette/pesticides1.htm>

<http://www.ifen.fr/>

{2} Bureau nationale de métrologie :

<http://www.bnm.fr>

{3} Notions d'Ecotoxicologie (**voir le glossaire**) :

<http://www.isims.be/database/coursOL/coursOL10.doc>

{4} Principes de calcul de turbidité :

[http://www.shom.fr/fr\\_page/fr\\_act\\_oceano/turbidite/quiditurbid.htm](http://www.shom.fr/fr_page/fr_act_oceano/turbidite/quiditurbid.htm)

{5} Glossaires d'hydrologie et biologie :

[http://sis.agr.gc.ca/siscan/glossary/parts\\_per\\_million.html](http://sis.agr.gc.ca/siscan/glossary/parts_per_million.html)

[http://195.101.101.220/eau/serveurs\\_donnees/veille\\_hydro/glossaire.html](http://195.101.101.220/eau/serveurs_donnees/veille_hydro/glossaire.html)

<http://www.futura-sciences.com/>

## CHAPITRE II: Tests de toxicité et traitements statistiques

### II.1 Introduction

Le milieu naturel varie en continu en fonction des grandeurs physiques qui y interagissent. Ce type d'interactions nous intéresse particulièrement en situation de pollution. A cet effet, le décryptage de ces interactions passe par une campagne de tests, dont on simule les conditions en fonction des types de pollutions à étudier. L'intérêt de cette démarche consiste à améliorer notre connaissance du milieu naturel (*milieu des eaux de surface*) à travers les mesures de chaque capteur. De ces mesures seront extrapolés les indicateurs statistiques utiles à la base de connaissance du système de fusion de données (système détaillé au chapitre IV).

La complexité de cette analyse réside principalement, dans le fait que ces grandeurs physiques varient de manière imprévisible, indépendamment des phénomènes accidentels. L'analyse préventive se basera sur une observation à posteriori des grandeurs et de leur prise en compte dans leur ensemble. Le nombre de paramètres régissant un milieu naturel tel que celui des eaux de surface étant abondant, nous nous sommes limités à un choix de grandeurs couvrant un panel des risques de pollutions les plus usuellement appréhendées par les collectivités exploitant le cours d'eau. Cela justifie le choix des grandeurs physico-chimiques mesurées associées à la chaîne trophique que constitue la chaîne de capteurs biologiques.

Dans cette optique, nous avons procédé à une campagne de tests de toxicités permettant d'évaluer la réactivité des capteurs de la station, en vue d'exploiter ces résultats pour notre système d'alerte, celui-ci étant basé sur les seuils à préétablir selon un profil de pollution. Les analyses statistiques qui s'en suivent, sont destinées à avoir une meilleure connaissance des synergies entre paramètres et être à l'affût des signes avant coureurs d'une alerte à la pollution.

### II.2 Tests de toxicités et plans d'expériences

Nous abordons les tests de toxicités et plans d'expériences en même temps, du fait qu'une partie des tests ont été effectués en s'appuyant sur la méthode des plans d'expériences, destinés à réduire le nombre d'expérimentations. En effet, le plan d'expérience en question, a concerné particulièrement la simulation des pollutions agricoles dans lesquelles plusieurs toxiques interviennent, d'où l'intérêt de faire varier les teneurs de chacun et tenant compte du maximum de scénarios possibles. Cela nous amène, dans ce qui suit, à aborder les tests de toxicités dont le déroulement est détaillé en annexe V, puis nous abordons les plans d'expériences en tant qu'outil mathématique appliqué au cas particulier des simulations de pollutions agricoles.

La procédure de tests pour la simulation de pollution a été mise en place pour enrichir la base de connaissance destinée à affiner le système de synthèse d'alerte. Le principe consiste à évaluer, en même temps, le comportement d'un ensemble de capteurs ayant réagi à une pollution connue, que l'on simule. Lorsque ce comportement réapparaît en situation réelle, on peut donc identifier le type de pollution le plus rapidement possible. Cela permet de mieux cerner la pollution, à travers les signes avant coureurs que présenteront les capteurs ayant réagi, en faisant le parallèle avec la base de connaissances préétablie grâce aux simulations.

La détermination des pré-alertes passe par l'étude de la dynamique du signal lors de la réaction de chaque capteur face à une pollution donnée. Le signal produit par les capteurs,

est par conséquent analysé afin d'y associer une pollution donnée. En effet, l'information pertinente ne se limite pas à savoir quels capteurs ont réagi à une pollution, mais aussi à quantifier la variation de chaque signal, d'où la notion de correspondance d'un ensemble de pré-alertes à un profil de pollution.

Comme mentionné précédemment, les tests choisis ont été faits sur une base de risques usuellement appréhendés par les collectivités exploitant le cours d'eau. Celui-ci traverse une région à vocation agricole, industrielle et est traversé par des routes à grande circulation, ce qui justifie les tests réalisés, subdivisés en cinq types :

- pollutions agricoles se produisant le plus couramment après lessivages des sols, présence en l'occurrence d'une haute turbidité, d'herbicide, de nitrates, de phosphates et de pesticides lors de pulvérisations massives ou vidanges volontaires ou accidentelles, en l'occurrence l'Atrazine (Herbicide interdit depuis juin 2003) et le Diuron (insecticide),
- pollution au métal lourd pouvant provenir autant de l'industrie que des lessivages ; le choix s'est porté sur le Chrome VI, considéré par les autorités locales et experts comme ayant une forte probabilité de présence dans le cours d'eau dont il faut surveiller la teneur, notamment de part le fait qu'il rentre dans la composition des revêtements anti-corrosion automobile,
- pollution industrielle due aux déversements industriels, dans notre cas nous avons fait un mélange de soude, d'acide sulfurique et de chrome,
- pollution aux hydrocarbures, en cas de vidanges massives ou lessivage des routes par temps de pluie ou accident, en l'occurrence du fuel, gasoil et acétone,
- pollution écologique représentant un système anoxique : chute des oxygènes dissous avec présence de sulfure de sodium en milieu acide.

Cette subdivision, reflétant la plage de risque la plus usuellement appréhendée par les collectivités, a nécessité un choix de polluants les plus représentatifs par famille sur le critère de forte probabilité d'apparition. Ce choix ne couvrant pas la totalité des polluants, s'explique par le fait que couvrir la totalité des polluants, s'avère irréalisable financièrement et matériellement de part le grand nombre de polluants référencés, et que par conséquent une classification par famille s'imposait.

En effet, les tests consistent à injecter une gamme de toxiques recouvrant un large panel de polluants inspirés de la liste d'une soixantaine de polluants mentionnés dans le *décret n°89-3 du 3 janvier 1989 relatif aux eaux destinées à la consommation humaine à l'exception des eaux minérales* [Décret n°89-3]. Un détail des substances impliquées est présenté en Annexe III.

Tous les tests de pollutions ont été effectués sur la base d'un facteur à faire varier, à l'exception de la pollution agricole, pour laquelle plusieurs facteurs interagissent. En effet ce type d'expérimentation nécessite de faire varier, à la fois, plusieurs facteurs selon deux modalités au moins. Cela engendre une prise en compte de toutes les combinaisons possibles, pouvant atteindre un nombre élevé de tests. Pour cette raison, nous avons opté pour la méthode des plans d'expériences, qui tiennent compte de la multiplicité de facteurs sans pour autant multiplier le nombre d'expériences. Dans ce contexte, la notion de **facteur** symbolise la grandeur physico-chimique ou toxique que l'on fait varier tout au long d'un test, alors que la **modalité** correspond à la valeur prise par le facteur.

Il en découle que les tests de toxicités effectués se déclinent selon les deux variantes : multi-facteurs et mono-facteurs, que l'on décrit ci-après.

## II.2.1 Tests de toxicités multi-facteurs et plans d'expériences

### II.2.1.1 Définitions

Dans le dictionnaire, le mot «expérience» peut avoir les définitions suivantes :

- « Epreuve, visant à étudier un phénomène »
- « Scientifiquement, fait provoqué ou attendu pour vérifier une hypothèse, une loi, et arriver à une connaissance théorique de la façon dont se passent les choses ».

La notion d'expérience s'oppose ainsi à celle d'observation ou d'enquête. Dans le cas des études par enquête, on observe en effet les phénomènes auxquels on s'intéresse, tels qu'ils se présentent « naturellement », sans les influencer en aucune façon, tandis que dans le cas des études par expérience, on provoque les phénomènes auxquels on s'intéresse, en maîtrisant certains facteurs, et on observe les conséquences qui en résultent [ DAG - 1997]. On ne doit cependant pas perdre de vue le fait qu'au-delà des quelques facteurs qui sont alors maîtrisés ou contrôlés, subsistent en réalité un nombre important de facteurs non contrôlés.

En effet, toutes les grandeurs agissant sur le phénomène à étudier ne peuvent être considérées. Seules les variables importantes, que l'on pourrait qualifier de principales, sont examinées [BENO-1992].

Le problème qui se pose est de connaître la « loi » reliant le phénomène à étudier, en l'occurrence la pollution, aux variables considérées. La réponse à ce problème consiste à réaliser un nombre d'expériences nécessairement fini de manière à prendre en considération l'ensemble des combinaisons.

De ces définitions, il découle que toute expérience nécessite une planification, d'où la notion de plans d'expériences. Les principes de l'approche classique, en l'occurrence l'approche COST (*Changing One Separate factor at Time*), consistaient à modifier un paramètre à la fois jusqu'à ce que plus aucune modification ne soit obtenue sur l'ensemble des résultats recensés [JOHA-2002]. Cette approche s'est avérée très inefficace comme l'a montré Fisher en 1925, de part le fait de modifier un facteur à la fois :

- ne donne pas d'information sur la position de l'optimum,
- ne fournit aucune information, lorsque plusieurs facteurs sont changés en même temps, ignorant ainsi les interactions entre facteurs,
- n'est pas efficace pour un nombre élevé d'expériences,
- fournit moins d'informations au sujet de la **variabilité**, notion abordée ultérieurement.

A partir de ces constats, ces méthodes ont été améliorées principalement par Yule, Box, Student, Hunter, Scheffe, Cox, Taguchi, pour aboutir à la définition des **Plans d'expériences**, qui fournissent un cadre rigoureux pour la résolution des problèmes d'optimisation des tests.

Cette constante recherche de l'optimum peut se résumer par cette citation de George Box (1992) : "...We have a large reservoir of engineers (and scientists) with a vast background of engineering know-how. They need to learn statistical methods that can tap into knowledge. Statistics used as catalyst to engineering creation will, I believe, always result in the fastest and most economical progress..."

Les plans d'expériences peuvent être définis comme étant un ensemble d'expériences faisant l'objet d'une planification précise qui se concrétise sous la forme d'un protocole expérimental.

L'apport le plus important des plans d'expériences réside dans le fait qu'ils fournissent un cadre mathématique rigoureux permettant de modifier tous les facteurs simultanément, et ce, avec le plus petit nombre d'expériences.

### II.2.1.2 Facteurs et modélisation

Afin de mener à bien ces expériences, il est utile définir les tenants et les aboutissants de celles-ci, qui consistent à connaître les effets d'un ensemble de toxiques sur les grandeurs de l'installation multi-capteurs.

De ce fait, les paramètres constituant les toxiques sont assimilés en tant qu'entrées du système multi-capteurs et les réactions des capteurs comme étant les réponses de ce système, ce qui correspond à la schématisation suivante :

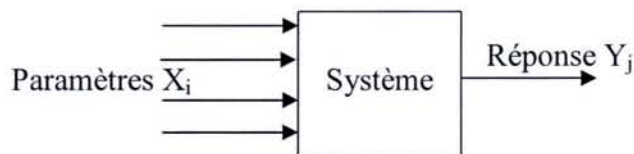


Figure III.1 : « Représentation entrées/sortie »

Ainsi, le niveau d'une réponse  $Y_j$  peut être exprimé en fonction d'une sélection de paramètres ( $X_1, X_2, X_i$ ) :  $Y_j = f(X_1, X_2, X_i)$ .

Il est à noter que les facteurs peuvent être :

- des facteurs contrôlés ou facteurs maîtrisés,
- des facteurs non contrôlés,
- des facteurs quantitatifs ou facteurs mesurables,
- des facteurs qualitatifs ou facteurs non mesurables mais identifiables.

L'objet des plans d'expériences est de quantifier l'influence des paramètres sur la réponse à partir du résultat des expérimentations. Pour cela, il existe deux stratégies d'expérimentation principales [BAI - 1996], [DAG - 1997], [FOW - 2000] et [VIG - 1991] :

#### II.2.1.2.1 Les plans factoriels complets :

Une expérimentation factorielle complète, permet l'exploration de l'ensemble des combinaisons des niveaux de facteurs. Ce qui implique que dans le cas d'une expérimentation à  $x$  facteurs selon  $y$  modalités, le nombre total d'expériences est donné selon la formule suivante :

$$\text{Nombre de combinaisons} = y^x \quad (\text{III.1})$$

Afin de schématiser les entrées sorties, considérons l'exemple [VIG - 1991] de 2 facteurs notés A, B. On s'intéressera à leurs effets sur une sortie Y selon 2 modalités, obtenant un plan  $2^2$  qui correspond au nombre de variantes des réponses notées:  $Y_1, Y_2, Y_3, Y_4$ .

En ayant considérées toutes les expériences, il s'agit d'un plan complet ou plan factoriel, comme indiqué sur le tableau suivant :

Facteur A	Facteur B	Réponse Y
1	1	$Y_1$
2	1	$Y_2$
1	2	$Y_3$
2	2	$Y_4$

Tableau III.1 : « Tableau d'un plan  $2^2$  »

Le modèle de la réponse s'écrit sous sa forme symbolique [VIG – 1991] :

$$\hat{Y} = I + A + B \quad (\text{III.2})$$

$\hat{Y}$  signifie qu'il s'agit des Y estimés par le modèle, et I représente la moyenne générale des réponses qui s'écrit comme suit :

$$I = \frac{1}{4} \sum_{i=1}^4 Y_i \quad (\text{III.3})$$

La représentation détaillée du modèle se présente sous la forme suivante :

$$\hat{Y} = I + [a_1 \ a_2]A + [b_1 \ b_2]B \quad (\text{III.4})$$

$a_1$  est l'effet de A au niveau 1

$a_2$  est l'effet de A au niveau 2

$b_1$  est l'effet de A au niveau 1

$b_2$  est l'effet de A au niveau 2

Or si nous considérons l'interaction des facteurs entre eux, nous devons tenir compte d'un effet supplémentaire, qui donne la réponse sous sa forme symbolique qui s'exprimera comme suit :

$$\hat{Y} = I + A + B + AB \quad (\text{III.5})$$

Cette prise en compte des interactions entre facteurs augmente considérablement le nombre d'expériences, dans ce cas notamment avec le doublement du nombre d'expériences. Il est à préciser, que dans le cadre des plans factoriels, des tableaux ont été édifiés pour la détermination des différents effets de chaque interaction. Il en ressort que les effets principaux ainsi que les effets des interactions sont des combinaisons linéaires des réponses pour lesquelles la somme des coefficients est nulle.

La modélisation d'une réponse, comprend un terme pour chacun des facteurs et chacune des interactions. Les coefficients sont représentés sous forme de moyenne générale, d'effets de facteurs ainsi que d'interactions entre facteurs. La comparaison entre les termes permet de voir l'importance des interactions entre les facteurs, le degré d'influence entre eux, ainsi que leur importance pour la détermination de la réponse.

Cette modélisation dite de représentation qualitative permet de déterminer ainsi, le nombre de degrés de liberté qui correspond au nombre de coefficients indépendants nécessaires à sa détermination.

La notion de degré de liberté est utile pour décrire la taille nécessaire d'une expérimentation et la quantité d'information qui peut en être extraite. Le nombre de degrés de liberté d'une matrice d'expériences correspond à :

$$ddl_{exp} = \text{nombre de expériences} - 1 \quad (\text{III.6})$$

De même, le nombre de degrés de liberté nécessaire à la description des effets d'un facteur correspond à :

$$ddl_F = \text{nombre de niveaux} - 1 \quad (\text{III.7})$$

Le nombre de degrés de liberté total nécessaire pour la description des effets de l'ensemble des facteurs est égal à la somme des nombres de degrés de liberté associés à chacun d'eux :

$$ddl_T = \sum ddl_F \quad (\text{III.8})$$

Pour un facteur possédant  $p$  variantes, le modèle comporte  $p$  effets dont  $(p-1)$  indépendants en prenant les effets entre eux deux à deux. La détermination des effets d'un facteur à  $p$  variante nécessite donc  $(p-1)$  degrés de liberté.

Il est à noter que l'augmentation de la précision des estimations des effets est possible grâce à la répétition du plan que l'on appelle réplique. Les réponses correspondant à chaque combinaison de facteurs seront calculées sous forme de moyenne des réponses des différentes répliques.

Les interactions non nécessairement désirées complexifient l'établissement d'expériences. La démarche optimale dans notre contexte, consiste à choisir des facteurs n'interagissant pas entre eux.

En considérant l'exemple simple de 7 facteurs à faire varier, selon deux modalités **1** et **2**, 128 combinaisons sont représentés dans le tableau III.2. Ce qui indique, que le nombre d'expériences de ce plan factoriel devient considérable lorsque le nombre de facteurs augmente, d'où l'intérêt d'optimiser le nombre d'expériences.

<i>Essai</i>	<i>Facteur A</i>	<i>Facteur B</i>	<i>Facteur C</i>	<i>Facteur D</i>	<i>Facteur E</i>	<i>Facteur F</i>	<i>Facteur G</i>
<i>1</i>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
<i>2</i>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>2</b>
<i>3</i>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>2</b>	<b>1</b>
<i>.</i>	<i>.</i>	<i>.</i>	<i>.</i>	<i>.</i>	<i>.</i>	<i>.</i>	<i>.</i>
<i>.</i>	<i>.</i>	<i>.</i>	<i>.</i>	<i>.</i>	<i>.</i>	<i>.</i>	<i>.</i>
<i>127</i>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>1</b>
<i>128</i>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>

Tableau III.2 : « Tableau d'un plan  $2^7$  »

La grande faiblesse de la méthode des plans factoriels est qu'elle comporte un nombre d'expériences excessives par rapport à la quantité d'information nécessaire à la compréhension des effets des facteurs. L'utilité pratique de ce type de méthode, n'a de sens que dans le cas d'un petit nombre de facteurs et de modalités.

### II.2.1.2.2 *Les plans réduits*

Appelée aussi plans fractionnaires, cette méthodologie consiste à tester une partie du plan factoriel. Les combinaisons sont choisies afin de fournir une information suffisante pour la compréhension des effets des facteurs. La fraction de plan choisi doit être en ce sens orthogonale avec les autres fractions du plan. Cette notion d'orthogonalité implique que l'effet de chaque facteur puisse être évalué de façon indépendante des effets des autres facteurs.

A titre d'exemple, il existe  $(2^n - 1)$  manières de couper un plan  $2^n$  en deux parties et plusieurs manières de couper à nouveau le plan  $2^{n-1}$  obtenu en deux parties ce qui conduit à un quart de plan. On voit que le nombre de solutions possibles dès qu'on aura atteint des fractions  $\frac{1}{8}$  ou  $\frac{1}{16}$ ... est très grand. Parmi toutes les solutions possibles, des tables standards ont pu être établies sous forme de tableau respectant les règles d'orthogonalités [VIG – 1991].

Concrètement ces plans d'expériences, se substituent à l'ensemble des combinaisons irréalisable financièrement et matériellement. En effet, la multiplication d'expériences, augmente considérablement le coût d'un projet.

Dans notre cas, le plan d'expérience défini, permet de préserver l'intégrité des résultats escomptés, en garantissant une couverture des combinaisons de facteurs sans pour autant tester l'ensemble des combinaisons. Il est ainsi possible d'étudier l'ensemble du comportement du système multi-capteurs en ayant effectué une partie des tests de toxicités possibles.

Le compromis trouvé dans notre contexte, est de choisir des facteurs avec un nombre de combinaisons d'expériences minimal en obtenant une précision équivalente à celle que nous aurions si nous testions toutes les combinaisons. Aucun des effets de chacun des paramètres n'est lié à celui d'un autre paramètre, n'apportant mathématiquement aucun biais à son estimation, c'est ce qu'on appelle la condition d'orthogonalité. **Afin de garantir cette condition d'orthogonalité, le mélange de polluants a été minutieusement choisi, tout en restant dans le cadre représentatif des pollutions agricoles.**

La notion d'orthogonalité, se traduit par le fait que si deux actions sont disjointes (n'ayant pas de facteurs communs), cela implique qu'elles sont orthogonales, et de même si, à chaque modalité de l'une, sont associées toutes les modalités de l'autre le même nombre de fois. Ainsi un plan est orthogonal si toutes les actions deux à deux sont orthogonales.

Notre liberté de choix de facteurs connus pour ne pas interagir entre eux, motive l'utilisation de la méthode *Taguchi*<sup>2</sup> [VIG – 1991], [MAT - 1997], [FOW-2000] et [AZAI - 2003] qui est présentée ci-après en proposant des plans orthogonaux pré-définis.

---

<sup>2</sup> En 1949, le laboratoire des communications électriques de la *Nippon Telephone and Telegraph Compagny* entreprit un projet de modernisation du système national de télécommunication. Connu pour ses compétences en ingénierie et pour sa manière d'appliquer les plan d'expériences dans un but purement « productif », G. Taguchi fut de promouvoir la productivité à l'intérieur de la section recherche et développement du laboratoire. C'est là que prirent forme les premières fondations de ce que G. Taguchi appelle l'ingénierie de la qualité [FOW- 2000].



### II.2.1.3 Méthodologie de Taguchi mise en oeuvre

Cette méthodologie propose un nombre prédéfini de tables orthogonales standard se présentant sous forme de tableau où chaque ligne correspond à une combinaison différente des facteurs. Chaque colonne correspond à un contraste (facteur) différent, et toutes les colonnes ont la propriété d'être orthogonales 2 à 2. Les tables de *Taguchi* sont identifiées par la notation suivante :

$$L_N (p^n) \quad (\text{III.9})$$

où N correspond au nombre de lignes,

n au nombre de facteurs,

p au nombre de modalités de chacun des facteurs.

De ce fait, la détermination du nombre d'expériences dépend de :

- la sélection des facteurs,
- du choix de modalités par facteurs,

Ces choix sont faits à partir d'une connaissance a priori du domaine d'expérimentation, en prenant en considération deux conditions minimales :

- en premier lieu que l'on ait autant de résultats que de degrés de liberté ; cela veut dire plus explicitement, avoir au moins autant d'expériences que de facteurs,
- en second lieu, l'indépendance des facteurs entre eux et en considérant cette hypothèse dès leur choix.

### II.2.1.4 Application des plans de Taguchi aux tests de pollutions agricoles:

L'idée de base est de définir un petit ensemble d'expériences, dans lequel tous les facteurs pertinents varient systématiquement. Cet ensemble n'inclut habituellement pas plus de 10 ou 20 expériences. Les plans d'expériences mis en place, concernent uniquement les seuls tests multi-paramètres de la campagne de tests de toxicité, en l'occurrence celle de la simulation des pollutions agricoles.

Dans le contexte des tests mis en oeuvre, les produits toxiques choisis sont représentatifs des risques que représente l'agriculture massive sur un cours d'eau. Nous avons en l'occurrence retenu:

- le **phosphate**,
- les **nitrate**s,
- l'**Atrazine** (pesticide massivement utilisé par les producteurs de maïs qui traitent 80% de leurs surfaces avec ce produit, interdit depuis 30 juin 2003, le risque demeure de part certains stocks inépuisés malgré la distribution prohibée),
- l'**Isoproturon** utilisé pour traiter les céréales comme l'orge.

Ces facteurs dans le cadre des plans d'expériences, n'ont aucune interaction chimique connue, d'où leur interdépendance.

S'ajoute à cela la simulation d'un facteur physique qui apparaît lors de ces pollutions, en l'occurrence la simulation d'une forte turbidité de l'eau grâce à l'ajout de Kaolin. Cette poudre blanche qui résulte de l'altération des roches granitiques, a la propriété de provoquer facilement de fortes turbidités, sans pour autant interagir sur la nature chimique

des polluants. C'est d'ailleurs cette absence d'interaction entre polluants qui garantit l'orthogonalité du plan, de par l'inexistence de combinaisons chimiques pouvant provoquer un effet autre que ceux des toxiques en question.

En choisissant de faire varier les toxiques selon deux modalités tout en maintenant le niveau de turbidité constant grâce au mélangeur de Kaolin mis en place (voir Annexe IV), nous aboutissons dans le cadre d'un plan complet à  $2^4 = 16$  expériences, ce qui s'avère très contraignant en terme de :

- coût de personnel mobilisé en considérant que chaque expérience de part sa préparation et déroulement prend une journée,
- maintenance et nettoyage des capteurs devant se faire plus fréquemment,
- coût de recyclage et impact sur l'environnement.

Ces contraintes nous conduisent au choix d'une table de *Taguchi* orthogonale, simplifiant le nombre d'expériences, en conservant la richesse de l'information obtenue au même titre que si nous avions effectué toutes les combinaisons

Dans notre cas, la table orthogonale de *Taguchi* nous a permis de réaliser 5 expériences, au minimum, avec rajout de 2 expériences supplémentaires selon le tableau III.2. Ce passage de 16 à 7 expériences est garanti par l'hypothèse que les facteurs choisis, n'interagissent pas entre eux, et de ce fait permet de choisir une table orthogonale de *Taguchi*. Ainsi, l'information fournie, est suffisante pour représenter la simulation des pollutions agricoles. Il est important de préciser, que dans la pratique des méthodes *Taguchi*, il est recommandé de respecter l'ordre donné dans la table, dans le sens où les facteurs difficiles à faire varier sont placés dans les colonnes qui varient le moins, afin de limiter la durée de l'expérimentation [FOW - 2000].

N° essai	Facteurs essayés				Résultats de l'essai
	A	B	C	D	
1	1	1	1	1	R1
2	2	1	1	1	R2
3	1	2	1	1	R3
4	1	1	2	1	R4
5	1	1	1	2	R5
6	2	1	2	1	R6
7	1	2	1	2	R7

Tableau III.2 : « Table orthogonale de Taguchi  $L_7 (2^4)$  »

Les essais facteurs ont été désignés comme suit :

- A représente les Nitrates,
- B représente les Phosphates,
- C représente l'Atrazine,
- D représente l'Isoproturon,

Ce qui permet de construire le tableau III.3, présentant les modalités sous la forme des concentrations mises en œuvre :

N° de pollution agricole	Facteurs essayés				Facteur fixe
	Nitrate (mg/L)	Phosphate (mg/L)	Atrazine (mg/L)	Isoproturon (mg/L)	
1	25	2,5	5	5	450
2	50	2,5	5	5	450
3	25	5	5	5	450
4	25	2,5	10	5	450
5	25	2,5	5	10	450
6	50	2,5	10	5	450
7	25	5	5	10	450

Tableau III.3 : « Tableau de concentrations en polluant désirés lors des tests de pollution »

La première ligne représente les sept expériences de pollutions agricoles désirées. En « Annexe V » est détaillé le calcul de la concentration des différents facteurs, pour les modalités de la pollution « agricole 1 ». Nous tenons à préciser que l'invariabilité du facteur de turbidité, permet de limiter notre étude aux effets chimiques d'une pollution agricole. En effet, la variation de ce facteur ne ferait qu'augmenter le nombre d'expériences sans pour autant apporter d'informations pertinentes sur l'essentiel de notre objectif.

A titre indicatif, en « Annexe VI » sont présentés quelques graphiques de réactions de capteurs, en l'occurrence, les 8 capteurs physico-chimiques (pH, eH, °C, Conductivité, Turbidité, CH<sub>2</sub>, O<sub>2</sub>, Chlorophylle) ainsi que le capteur biologique qu'est le Fluotox.

Nous tenons à préciser que le maintien de la turbidité à un niveau élevé et figé, est voulu, sur le principe qu'une pollution agricole draine une forte turbidité et que ce paramètre n'a pas lieu d'être assujéti à une étude supplémentaire sur la variation des constituants d'une pollution agricole, qui pour notre part nous intéresse que sur la partie toxique.

## II.2.2 Tests de toxicité mono-facteurs

Les autres tests de toxicités, voir « Annexe VI », concernent la variation de la modalité d'un unique facteur, ce qui n'exige pas l'établissement de plans d'expériences particuliers. Ces tests permettent de dresser les niveaux de réactions de chaque capteur de manière groupées. En « Annexe VII », sont mentionnés les graphiques de réactions de certains capteurs, nous rappelons parmi ces tests :

- le test de pollution industrielle, essentiellement composée d'acide sulfurique,
- le test de pollution au Carbofuran, qui est un insecticide particulièrement toxique largement utilisé pour traiter les cultures légumières,
- le test de pollution à l'hydrocarbure sur la base d'un mélange d'hydrocarbures tel qu'il pourrait se trouver dans la nature dans le cadre d'un lessivage de route ou de pollutions accidentelles. Notons que ce test a été effectué en dernier, pour cause de caractère persistant du produit, notamment dans les canalisations pour lesquelles le nettoyage était laborieux,

- le test de pollution au métal lourd, le Chrome VI, difficile à déterminer, car ceux-ci sont naturellement présents dans les cours d'eau. Ce n'est que lorsque les concentrations dépassent un certain seuil que l'on peut conclure à une contribution significative des activités humaines.
- Le test de pollution écologique dans le cas d'un système anoxique, représentant la chute des oxygènes dissous, simulée par une présence de sulfure de sodium en milieu acide.

A titre de synthèse en « Annexe VI », sont décrites les observations de base sur valeurs brutes de l'ensemble des tests de pollutions y compris les tests agricoles, ainsi que le fonctionnement des capteurs au cours de l'ensemble des simulations effectuées.

Les résultats présentés, donnent une information sur les données disponibles relatives aux tests de pollutions et une visualisation des réactions dites apparentes, sachant que l'étude de comparaison de variance détaillée au prochain chapitre, permet de mieux expliciter les variations significatives.

**Le processus de détermination de la réaction des capteurs s'appuie sur deux points** : une analyse des valeurs brutes du signal, et une analyse des pentes de variations inhérentes. Ces observations ont été faites sur la base de la variabilité des paramètres par rapport à leur variation maximale au cours de la journée.

Chaque test a été étudié individuellement. Ainsi une réaction, qualifiée d'intermédiaire, l'est au regard des réactions des autres capteurs au cours du même test. Elles ne sont pas à négliger : cela dénote dans certains tests que certains capteurs sont plus sensibles que d'autres et que pour faire réagir le capteur de manière plus significative, la concentration de toxique doit être augmentée pour atteindre le degré de sensibilité du capteur concerné.

En complémentarité de ces tests, l'analyse des données statistiques issues des capteurs, permet de fournir une connaissance a priori au système d'information. A cet effet, nous définissons et commentons brièvement dans ce qui suit, les principales notions utiles à l'étude statistique. Un complément théorique de ces définitions est détaillé en Annexe VIII.

### **II.3 Espérance, Variance, variabilité, covariance et écart-type**

Les différentes notions abordées ci-après, sont les bases des principaux calculs statistiques effectués sur nos jeux de données, en l'occurrence l'étude de variance. Avant d'aborder les notions de variances, il est utile définir la notion d'espérance mathématique qui correspond à la moyenne pondérée, par la probabilité, de toutes les valeurs des événements possibles d'une variable aléatoire.

Cela correspond à calculer, le produit de la valeur de chaque résultat possible par sa probabilité d'apparition et suite à quoi on fait la somme de tous les produits ainsi obtenus :

$$E(X) = \sum xP(x) \quad (\text{III.11})$$

où  $E(X)$  = l'espérance mathématique de la variable  $X$ ,  
 $x$  = toute valeur que peut prendre la variable  $X$ ,  
 $P(x)$  = la probabilité  $d$  de la valeur  $x$ .

Cela nous amène à pouvoir définir la variance correspondant à :

$$\mu_2 = E\left((X - E(x))^2\right) = \sigma^2 = \text{variance de } X = V(X) = E\left(X^2\right) - E(X)^2 \quad (\text{III.12})$$

La variance traduit la dispersion de la distribution de la variable aléatoire (*v.a.*) autour de sa valeur moyenne. Etant un carré, la dimension de la variance n'est pas celle de la moyenne. C'est pourquoi on utilise plus souvent l'écart type, noté  $\sigma$ , qui est la racine de la variance.

A ce titre, il est utile de définir la notion de variabilité qui est mesurée à partir du coefficient de variation (*CV*). C'est l'écart type exprimé comme un pourcentage de la moyenne.

$$CV = 100 \frac{\sigma}{\bar{x}} \quad (\text{III.13})$$

$x$  correspondant à la variable aléatoire et  $\sigma$  l'écart type

L'idée est de rendre comparable la variabilité de plusieurs jeux de données quand les unités de mesure sont différentes. Il est quand même important de noter que le *CV* dépend de la moyenne ; il n'est donc utile que lorsque les jeux de données sont de même type et avec des moyennes positives

Concernant la variance, elle traduit aussi la notion d'incertitude. Plus la variance est faible, moins le résultat de l'expérience aléatoire est incertain. A la limite, une *v.a.* de variance nulle conduit à des expériences strictement identiques. Dans ce cas, le phénomène est complètement déterministe, il n'y a donc plus aucune raison de garder la notion de variable aléatoire.

La variance a également des propriétés intéressantes vis à vis de la combinaison linéaire de *v.a.* :

Soient  $X_1$  et  $X_2$  deux *v.a.* :

$$V(X_1 + X_2) = V(X_1) + V(X_2) + 2\text{cov}(X_1, X_2) \quad (\text{III.14})$$

où  $\text{cov}(X, Y)$  est la **covariance** des *v.a.*  $X$  et  $Y$ , définie par :

$$\text{cov}(X, Y) = \mu_{1,1} = E(XY) - E(X)E(Y) = E[(X - E(X))(Y - E(Y))] \quad (\text{III.15})$$

La covariance peut être vue comme le moment centré conjoint d'ordre 1 de deux *v.a.* Si les deux *v.a.* sont indépendantes, alors leur covariance est nulle (mais la réciproque n'est pas vraie en général).

Par ailleurs, soit  $X$  une *v.a.* et  $a$  et  $b$  deux constantes, on a :

$$V(aX + b) = a^2 V(X) \quad (\text{III.16})$$

### II.3.1 Variable centrée réduite

On appelle variable aléatoire centrée réduite, une *v.a.*  $Y$  construite par :

$$Y = \frac{X - E[X]}{\sqrt{V[X]}} \quad (\text{III.17})$$

C'est le moyen le plus classique pour normaliser une *v.a.*, par construction, on obtient :

$$E(Y) = 0 \text{ et } V(Y) = 1 \quad (\text{III.18})$$

## II.4 Corrélation

A partir des définitions précédentes, il est possible de définir mathématiquement le coefficient de corrélation, utile à notre étude de paramètres qui suit. A titre récapitulatif, il existe plusieurs types de tests de corrélations. Parmi les principales corrélations :

- La corrélation paramétrique de Pearson,
- la corrélation non-paramétrique de Spearman,
- et corrélation non-paramétrique de Kendall.

Dans notre contexte, il s'agit d'utiliser la corrélation pour la mesure de la relation linéaire entre deux ou plusieurs variables quantitatives deux à deux, ce qui correspond à notre choix pour la corrélation paramétrique de Pearson.

Statistiquement, la corrélation linéaire de *Pearson* est la covariance de deux variables centrées réduites. Cette mesure de liaison a été développée par *Francis Galton* et *Karl Pearson* [SCH – 1984].

Le coefficient de corrélation ainsi calculé est borné entre « -1 et 1 ». Il permet d'avoir un aperçu sur l'interdépendance entre paramètres. La valeur -1,00 représente une parfaite corrélation négative, alors que la valeur +1,00 représente au contraire une parfaite corrélation positive. La valeur 0,00 représente une absence de corrélation (indépendance), la réciproque n'est pas systématique.

### II.4.1 Coefficient de corrélation

La relation entre deux *v.a.* peut être quantifiée par la covariance comme vue précédemment. Cependant, à l'image de la moyenne et de la variance, la covariance est un moment donc possède une dimension ce qui la rend plus difficile à interpréter. C'est pourquoi on utilise plus généralement le **coefficient de corrélation**, indicateur sans dimension, défini par:

$$\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\mu_{1,1}}{\sqrt{\mu_2(X) \mu_2(Y)}} \quad (\text{III.19})$$

Le coefficient de corrélation mesure la qualité de la relation linéaire entre deux variables aléatoires  $X$  et  $Y$  de la forme  $Y=aX+b$ , avec les propriétés suivantes :

- $\forall X, Y: \rho(X,Y) \in [-1,1]$
- Si  $X$  et  $Y$  sont indépendantes, alors  $\rho(X,Y)=0$ .
- $\forall X, Y : \forall a_1, a_2, b_1, b_2 \in \mathfrak{R} (a_1 a_2 \neq 0) : \rho(a_1 X + b_1, a_2 Y + b_2) = \text{sign}(a_1 a_2) \rho(X, Y)$   
le signe «  $\text{sign}(a_1 a_2)$  » étant déterminé par le signe du produit des coefficients  $a_1$  et  $a_2$ , indiquant ainsi le «sens» de la relation.
- S'il existe une relation linéaire entre  $X$  et  $Y$  alors  $\rho(X,Y) = \pm 1$ .

On peut réécrire la relation sur la variance d'une somme de *v.a.* en utilisant le coefficient de corrélation:

$$\sigma^2(X_1+X_2) = \sigma^2(X_1) + \sigma^2(X_2) + 2\rho(X_1, X_2)\sigma(X_1)\sigma(X_2) \quad (\text{III.20})$$

Et en généralisant, on obtient :

$$\sigma^2\left(\sum_{i=1}^{i=n} X_i\right) = \sum_{i=1}^{i=n} \sigma^2(X_i) + 2 \sum_{i=1}^{i=n-1} \sum_{j>i}^{j=n} \rho(X_i, X_j) \sigma(X_i) \sigma(X_j) \quad (\text{III.21})$$

## II.4.2 Application et résultats de la méthode de base

Le test « Pearson » est la méthode de corrélation de base avec laquelle nous avons effectué nos comparaisons, puis que nous avons appliqué de manière plus élaborée à travers les arbres de classification.

Rappelons que le coefficient de corrélation de *Pearson* suppose que la relation est linéaire entre les deux *v.a.*  $X$  et  $Y$  à comparer, en l'occurrence :

- Relation linéaire positive parfaite  $\Rightarrow \rho(X, Y) = 1$
- Relation linéaire négative parfaite  $\Rightarrow \rho(X, Y) = -1$
- Absence de relation  $\Rightarrow \rho(X, Y) = 0$

Entre ces valeurs extrêmes

- Le signe du coefficient indique le «sens» de la relation
- La valeur (absolue) du coefficient indique la force de la relation

Exemples:

- Relation positive «faible»  $\Rightarrow 0,38$
- Relation négative forte (non parfaite)  $\Rightarrow -0,96$

Nous avons étudié statistiquement un jeu de données contenant environ 8000 analyses, effectuées entre le 1er janvier 1999 et le 30 novembre 1999, pour lequel nous avons récapitulé les coefficients de corrélation dans le tableau III.4 :

	pH	Turb.	Cond.	O <sub>2</sub>	Chlorures	Ammo.	Temp.
pH	1	-0,14	0,52	0,34	0,36	0,04	0,46
Turb.		1	0,15	0,0029	0,13	0,08	0,15
Cond.			1	0,14	0,61	0	0,62
O <sub>2</sub>				1	0,29	0,02	0,41
Chlorures					1	0,07	0,47
Ammo.						1	0,01
Temp.							1

Tableau III.4 : « Tableau de corrélation de janvier à novembre 1999 »

Ce tableau de corrélation renseigne sur plusieurs paramètres. On peut en extraire les couples suivants et leur coefficient de corrélation :

- Conductivité – Chlorures (0,61) : Cette corrélation peut s'expliquer chimiquement par une proportionnalité entre la présence d'ions chlorures [Cl] et la conductivité.
- Conductivité – Température (0,62) : Cette deuxième corrélation, peut s'expliquer par une influence de la température sur les processus chimiques qui favorisent l'augmentation de la conductivité.

Concernant les autres tendances, nous ne pouvons apporter de conclusions, du fait de la faible corrélation entre paramètres.

D'autre part, nous relevons l'absence de corrélation significative entre l'ammonium (NH<sup>4+</sup>) et tous les autres paramètres. L'ammonium (NH<sup>4+</sup>) résulte le plus souvent de la décomposition anaérobie des matières organiques azotées. De part les valeurs du coefficient de corrélation, il en ressort que ce processus, n'a aucune dépendance avec les autres variables.

Il en est de même pour la turbidité, qui est un indicateur de la quantité de colloïdes en suspension. Leur présence, dépend à priori de paramètres autres que ceux étudiés.

Il est utile de préciser, que ces observations sont spécifiques au cours d'eau étudié, et que certaines ne peuvent s'appliquer qu'à notre contexte. Il faut aussi savoir qu'une étude de ce type sur une période de onze mois, tient compte des effets de toutes les saisons et que certains paramètres n'ont pas les mêmes types de variations selon les saisons. Notons également que dans les cas de valeurs erronées ou manquantes, nous avons fait appel à des techniques de remplacement de données manquantes, en l'occurrence l'utilisation de moyennes.

Complémentairement à cette étude faite sur un jeu de données de onze mois, nous nous sommes intéressés à l'étude des données, suivant le cours d'une journée, selon les huit paramètres physico-chimiques que sont : le pH, le potentiel Redox, la température, la conductivité, la turbidité, les hydrocarbures, l'oxygène dissous et la chlorophylle.

Le tableau de corrélations III.5 [AZZ(a) - 2002], n'a pas pour objectif de généraliser les corrélations qui pourraient en sortir, mais de pouvoir d'une part conforter certaines tendances et d'autre part montrer la disparité des relations de corrélations d'une journée à une autre, dans notre cas le choix s'est porté sur trois journées consécutives.

	Redox			Temp.			Condu.			Turbi.			Hydro.			O <sub>2</sub>			Chloro.		
pH	-0,52	<b>-0,83</b>	-0,61	0,02	<b>0,69</b>	<b>0,86</b>	<b>-0,76</b>	0,45	<b>-0,84</b>	-0,35	0,4	<b>-0,78</b>	-0,56	0,24	0,07	<b>0,84</b>	<b>0,78</b>	<b>0,64</b>	-0,35	-0,6	0,36
Redox				0,24	0,4	-0,5	0,12	<b>0,77</b>	<b>0,85</b>	-0,23	0,2	0,8	0,1	0,33	-0,1	-0,48	0,66	-0,13	-0,37	<b>0,7</b>	-0,4
Temp.							-0,32	0,1	<b>0,74</b>	-0,48	-0,5	<b>0,82</b>	-0,02	0,08	0,12	-0,3	<b>0,85</b>	<b>0,84</b>	-0,42	-0,18	0,41
Condu.										0,37	0,22	<b>0,8</b>	<b>0,72</b>	0,42	0	<b>-0,72</b>	0,07	-0,32	0,41	<b>-0,7</b>	-0,25
Turbi.													0,19	0,09	-0,13	-0,04	-0,05	-0,6	<b>0,91</b>	0,04	-0,4
Hydro.																<b>-0,69</b>	0,09	0,14	0,24	0,56	0,22
O <sub>2</sub>																			-0,1	0,24	0,46

**En gras** : les coefficients de corrélations absolues supérieures à 0,69, que l'on considère significative.

Tableau III.5 : « Récapitulatif des corrélations pour les trois journées du 28 au 30 juin 2001 »

En synthèse de ces trois journées, on peut établir certaines tendances, en l'occurrence, l'existence d'une forte corrélation positive entre le pH et l'oxygène dissous. D'autre part, certaines corrélations sont le fruit de circonstances particulières, notamment concernant la température avec les autres paramètres. Par exemple, la température et l'oxygène dissous qui sont fortement corrélés sur les deux dernières journées atteignant des valeurs de facteur de corrélation supérieures à 0,8 contrairement à la première journée où l'on est proche de zéro. Même type de constat entre la turbidité et la chlorophylle, où l'on atteint 0,91 s'expliquant du fait qu'une forte présence d'algues induit une forte turbidité.

Il est important de souligner que ces observations ne concernent que des journées particulières et ne peuvent donner lieu à une généralisation. Ceci s'explique que d'une part, ces périodes ne reflètent pas les variations d'une saison et que d'autre part, une dérive ou une panne des capteurs sur de courtes périodes ne peuvent être facilement décelées et peuvent



fausser certaines interprétations. Néanmoins, nous avons pu conforter certaines tendances et pu montrer le caractère aléatoire de certaines corrélations.

### II.4.3 Corrélation par arbre de classification

L'objectif est de construire un ensemble de classes sur la base de mesure de «dissemblances» entre capteurs. Par cette notion, nous entendons la réutilisation de la corrélation précédente sous sa forme inverse. Le calcul de cette «dissemblances» est défini par la mesure de la distance que l'on explicite tout au long de ce point.

La nécessité de pouvoir grouper les capteurs par classes permet d'identifier le niveau de relations pouvant exister entre les capteurs sous formes d'ensembles. Il s'agit en l'occurrence, d'une étude de corrélation permettant une analyse de similarités, ne se restreignant pas qu'à deux capteurs. Il en ressort que l'on peut construire, un arbre englobant toutes les classes de corrélations qui à leur tour sont évaluées entre elles. Pour cela, nous avons utilisé l'outil des arbres de classification [SY09-2004]. La classification est l'organisation d'un ensemble en classes naturelles ou classes homogènes via le calcul d'un critère comme indiqué sur la figure III.2.

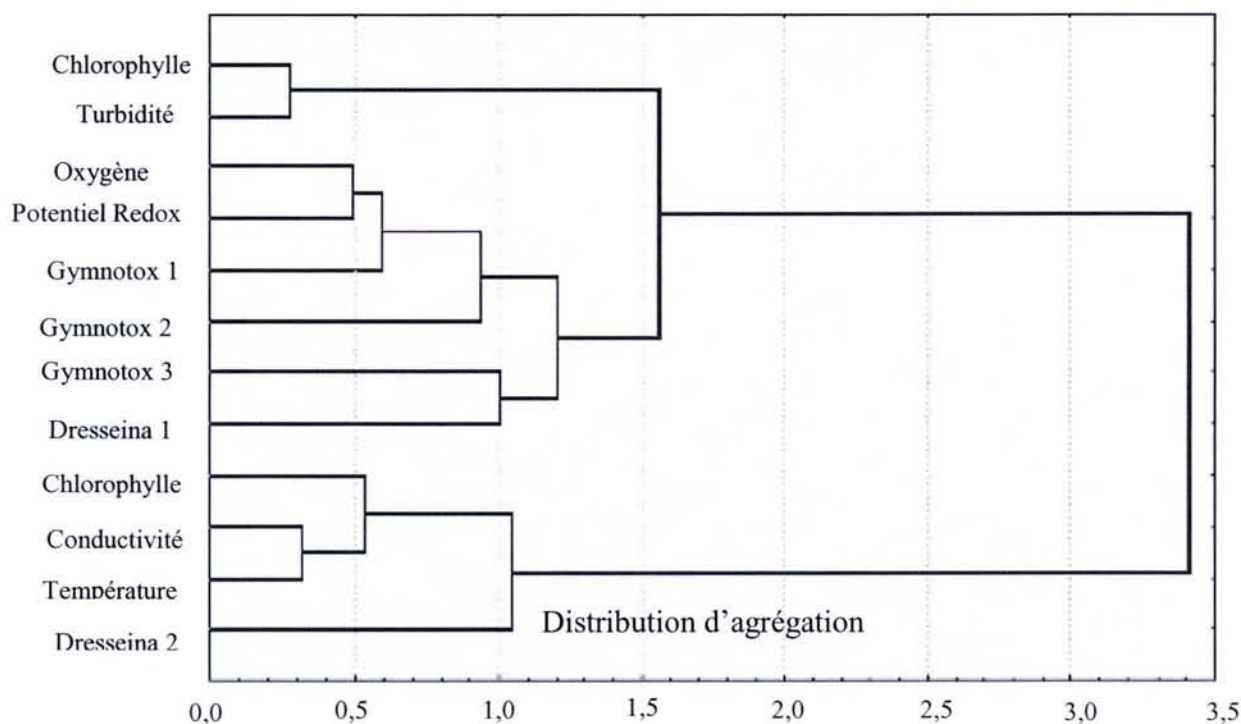


Figure III.2 : « Arbre de Classification »

L'arbre hiérarchique horizontal présenté sur cette figure commence à gauche par le tracé de chaque objet individuel dans une classe unique. Puis, imaginons que petit à petit nous assouplissions notre critère, de sorte que chaque objet ne soit plus seul dans sa classe. Ainsi, de plus en plus d'objets sont regroupés dans des classes de plus en plus larges, conduisant à des classes composées d'éléments de plus en plus dissemblables. Finalement, à la dernière étape, tous les objets forment une seule classe.

Dans ce tracé, l'axe horizontal représente la distance d'agrégation. C'est pourquoi, pour chaque nœud du graphique (le lieu d'une formation d'une nouvelle classe), nous pouvons lire rapidement la distance du critère auquel les éléments ont été agrégés en une seule classe.

Il est à noter que la constitution de cet arbre de classification, s'est faite sur la base de données de la période du 2 au 10 novembre 2001. Nous avons donc constitué une base de données sous forme d'un mélange entre des paramètres physico-chimiques et biologiques, ce qui a constitué une matrice de 12 colonnes et 2600 lignes. Les trois notations de Gymnotox représentent les fréquences des trois poissons, alors que les deux notations de Dresseina représentent le pourcentage de moules ouvertes sur les deux canaux de moules. Cette séparation des canaux sur deux classes différentes dénote une fragilité du capteur lors du renouvellement des êtres vivants qui prennent du temps à se réadapter, faussant ainsi les mesures, sans que le capteur soit défaillant.

Avant l'analyse détaillée de l'arbre, il est utile de rappeler deux notions élémentaires :

- **Mesure de distance** : La classification hiérarchique [SY09-2004] utilise les dissemblances ou distances entre les objets lors de la formation des classes. Par exemple, pour classer des plats culinaires, on pourrait prendre en compte le nombre de calories qu'ils contiennent, leur prix, etc. La distance choisie dans notre cas est la distance (1 - r Pearson) qui correspond à (1 - le coefficient de corrélation). Ainsi, plus le coefficient entre 2 objets est grand, plus la distance entre ces deux objets sera petite.
- **Règle d'agrégation** : A la première étape, lorsque chaque objet représente sa propre classe, les distances entre ces objets sont définies par la mesure de la distance choisie. En revanche, après l'agrégation de plusieurs objets, comment déterminer la distance entre deux nouvelles classes ? En d'autres termes, il faut une règle d'agrégation pour déterminer quand deux classes seront suffisamment similaires pour n'en former plus qu'une. Pour réaliser notre arbre, nous avons utilisé la méthode de Ward<sup>3</sup> (1963) [SY09-2004]. C'est une méthode qui utilise une analyse de la variance approchée afin d'évaluer la distance entre classes. En résumé, cette méthode tente de minimiser la somme des carrés des **distances** de tous les couples (hypothétiques) de classes pouvant être formées à chaque étape.

Il s'ensuit de ces notions l'**Analyse de l'arbre obtenu** :

La classification hiérarchique de la figure III.1 nous montre que plusieurs groupes de données semblent se détacher. En l'occurrence, si l'on coupe l'arbre de la figure à une distance de 1,5 (axe des abscisses), on peut voir que 3 groupes se forment :

- Chlorophylle, Turbidité
- Oxygène, Potentiel redox, Gymnotox 1, Gymnotox 2, Gymnotox 3 et pourcentage de moules ouvertes dans le canal 1 (%C1)
- Hydrocarbures, Conductivité, Température et pourcentage de moules ouvertes dans le canal 2 (%C2)

On peut donc supposer que ces 3 premiers groupes représentent une partition de l'ensemble des capteurs. On constate que les 3 circuits du Gymnotox sont dans le même groupe. Par contre, et c'est plus surprenant, le pourcentage de moules ouvertes dans le canal 1

---

<sup>3</sup> On utilise les barycentres des classes obtenues par réallocation itérative en leur donnant un poids égal à la somme des poids des individus de la classe. A partir de ces barycentres et de leur poids, on réalise une classification hiérarchique selon le critère du saut de Ward [SY09-2004]. En fonction de l'arbre de classification obtenu, on détermine en combien de classes il convient de regrouper l'ensemble des individus.

n'appartient pas au même groupe que le pourcentage de moules ouvertes dans le canal 2. On peut supposer ici qu'un des deux canaux était en dysfonctionnement.

Nous tenons à noter, le fort niveau de corrélation existant entre la Chlorophylle et la turbidité qui reflète bien l'accroissement de l'activité Chlorophyllienne liée à une forte turbidité. Il ressort de ces niveaux de classifications, la possibilité de dégager une tendance sur les groupes de réactions engendrées par des pollutions. L'intérêt d'une telle appréciation est la possibilité d'avoir un niveau de redondance appréciable pour une meilleure interprétation de la pollution.

## II.5 Analyse en composantes principales (ACP)

### II.5.1 Définitions

Il faut tout d'abord rappeler que l'objectif d'une analyse en composantes principales est de représenter graphiquement les relations entre variables quantitatives et de visualiser en même temps les individus qui sont dans la relation avec ces variables. Cette représentation graphique permet de saisir d'un seul coup d'œil ce que le calcul numérique ne peut fournir. A partir de cette double visualisation, l'analyste pourra interpréter les résultats en fonction de sa connaissance du sujet étudié.

Interpréter signifie donner un sens physique aux axes factoriels, à des regroupements, à des singularités, à des continuums, à des proximités entre variables ou entre individus [JAMBU - 2003]. Cette synthèse fondée sur une exploitation des résultats d'observations permet à l'analyste de conclure, de construire d'autres hypothèses et de mettre en jeu d'autres relations entre variables. Dans notre cas, il s'agira de conforter les modèles de fusions de données pour les profils d'alertes associés.

### II.5.2 Applications et résultats

Dans notre contexte, la décomposition en ACP<sup>4</sup> a été utile dans le sens où cela nous a permis de discerner le comportement de certains capteurs vis à vis d'autres, ainsi que la corrélation relative d'une famille de capteurs. Nous avons en l'occurrence comparé des capteurs biologiques dont les comportements statistiques ne sont pas parfaitement stables. En effet, à titre indicatif comme représenté sur la figure III.3, le cercle de corrélation montre bien la dissimilitude du signal de l'Aquatox, par rapport à celui des trois autres signaux choisis, en l'occurrence : le Fluotox, le turbidimètre ainsi que le pH-mètre.

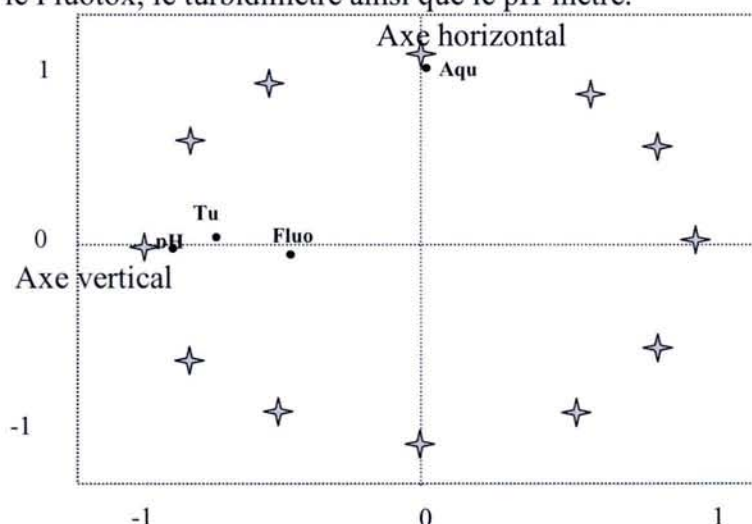


Figure III.3 : Exemple d'ACP pour le test de pollution agricole du 16 mai 2002.

<sup>4</sup> Pour plus de détail sur la théorie de l'analyse en composantes principales se référer au chapitre 5 de [SY09-2004] ou document électronique en référence Internet {4}.

La répartition des points sur les deux axes horizontal et vertical prédéfinis, représente la projection factorielle, de la contribution de la grandeur d'un capteur par rapport à un axe. Cette contribution par rapport à un axe est échelonnée de 0 à 1. Si celle-ci est proche de 1, la grandeur concernée est représentée par cet axe, si au contraire elle est proche de 0, l'individu est très mal représenté par cet axe.

En terme d'interprétation graphique, plus les paramètres sont proches de l'axe et assez loin de l'origine, plus la contribution est importante. Nous relevons à cet effet, deux groupes de corrélations en l'occurrence :

- la bonne corrélation du pH et de la turbidité par rapport à l'axe vertical et à moindre mesure celle du Fluotox,
- d'autre part, la bonne corrélation de l'Aquatox par rapport à l'axe horizontal, qui s'explique du fait que c'est le seul individu qui participe à la contribution de cet axe.

Il en ressort, que sur le principe de l'orthogonalité de l'axe vertical par rapport à l'axe horizontal, cela signifie que les paramètres contribuant à ces axes respectifs ne sont pas linéairement corrélés entre eux.

Cette méthode d'analyse ayant été appliquée avec les mêmes paramètres sur d'autres journées, ne produit pas les mêmes distances mais conforte bien la distinction de l'Aquatox par rapport aux autres capteurs. Cela nous permet de bien discerner certaines interprétations sur la variation des signaux de l'Aquatox. En effet, en fonction de la période de renouvellement des êtres vivants de l'appareil, celui-ci n'a pas les mêmes comportements, ce qui nous conduit à diminuer le degré de confiance sur cet appareil malgré une richesse d'informations sur les périodes dites normales.

On peut en conclure que cette analyse permet d'observer graphiquement la relative corrélation d'un groupe de capteurs mais aussi les éventuelles décorrelations. Ce gain d'appréciation permet de mieux imaginer l'interprétation de certains événements.

#### **IV.6 Perspectives temporelles**

Concernant l'étude statistique préliminaire, elle a été faite sur une base statique en vue d'une interprétation des données, néanmoins une étude dynamique sous forme d'estimations peut être envisagée afin d'enrichir le système déjà mis en place.

En effet, cette démarche prédictive doit s'inscrire dans un cadre complémentaire sans qu'elle ne se substitue à la démarche statique que constitue le paramétrage de base de notre système. Un complément sur la théorie des estimateurs et régressions est détaillé en « Annexe VIII ».

Cette piste envisagée conjointement par le CIRSEE (Centre International de Recherche sur l'Eau et l'Environnement) et le NANCIE [ETU-2001], a été suivie pour tenter d'estimer, en temps réel, le signal d'un capteur en fonction de celui d'autres capteurs avec lesquels il est corrélé. Cette régression temps réel, tient compte de l'état précédent de la grandeur à estimer. En l'occurrence, dans notre cas, nous avons estimé le comportement du pourcentage de moules ouvertes à l'instant (t) en fonction de l'état des capteurs hydrocarbure, température et conductivité à l'instant (t) et l'état des moules à l'instant (t-1). Ce groupe de capteurs, n'est pas anodin, puisqu'il s'agit du troisième groupe de corrélation, déduit par la méthode des arbres de classifications.

La figure III.4, illustre la comparaison entre le signal réel et le signal estimé :

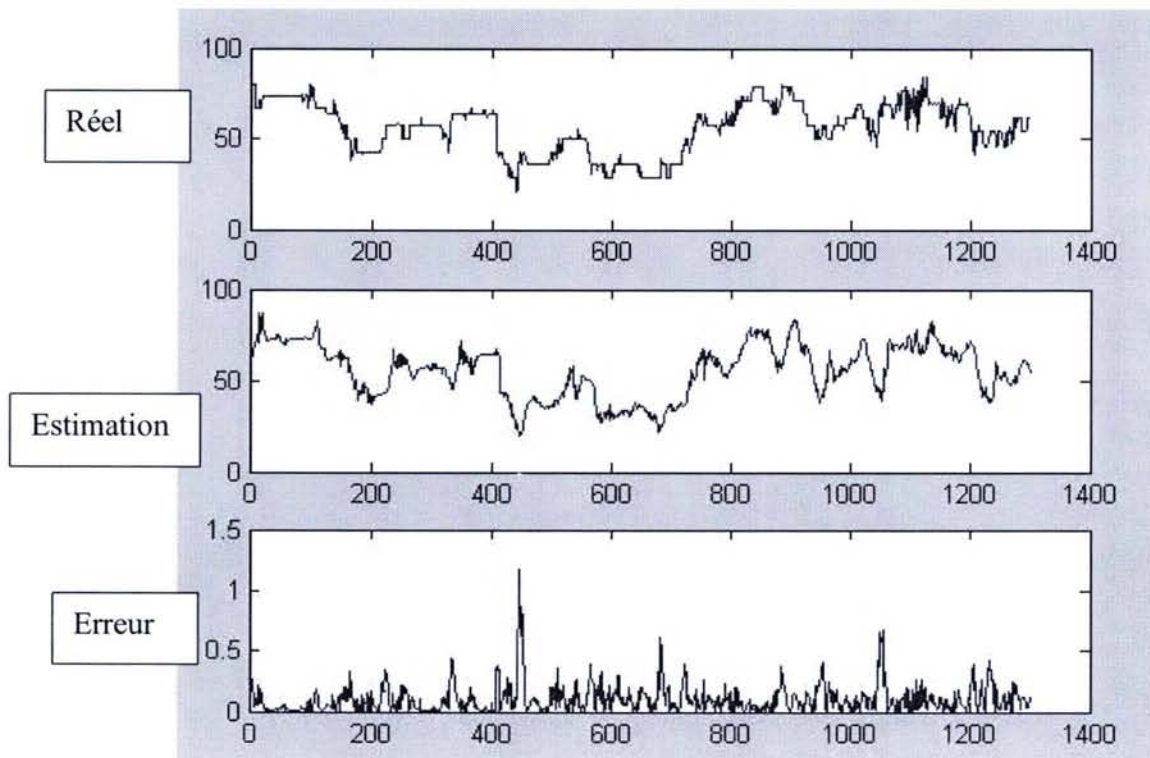


Figure III.4 : « Comparaison entre l'estimation et le signal réel du % de moules ouvertes »

Sur ces graphes, la première courbe représente la vraie valeur du pourcentage de moules ouvertes, la deuxième représente la valeur estimée, la troisième représente l'erreur correspondante. L'axe des abscisses indique le nombre d'échantillons de mesure. A raison d'un échantillon toutes les 5 minutes, cela représente 1300 échantillons équivalents à quatre jours et demi consécutifs.

Le calcul de l'erreur moyenne, nous donne une valeur de 7%, résultat tout à fait correct pour une estimation dans le contexte de notre environnement de variables.

Le but de cette régression est d'estimer, le pourcentage de moules ouvertes en fonction de sa progression précédente. Cet exemple est certes particulier mais il démontre bien qu'on a la possibilité d'améliorer une appréciation que l'on a d'une situation donnée.

Une utilisation potentielle de cette approche, peut consister à prédire le comportement sur un horizon à définir, la comparaison entre mesure et prédiction pouvant conduire à isoler de nouvelles situations de pollution.

Une autre approche intéressante pour quantifier et détecter un changement brutal dans un signal, est l'étude des séries temporelles dans le but d'identifier une saisonnalité qui peut éventuellement être observée sur certains signaux sous forme de périodicité.

En effet, il n'est pas évident de modéliser les signaux étudiés sous cette forme, du fait que ces grandeurs dépendent du comportement aléatoire de l'eau régi par le jour et la nuit, les saisons, l'activité humaine ainsi que par un mélange de paramètres « boîte noire » complexes à modéliser.

Néanmoins, nous avons pu constater sur certaines périodes que ces signaux, présentent des caractéristiques intéressantes de séries temporelles périodiques. Il s'agit dans ce type de cas, d'avoir recours à une modélisation de type SARIMA (*Seasonnal Autoregressive Integrated Moving Average*), pour des séries temporelles dont on a identifié la périodicité. Nous pouvons citer dans ce contexte l'étude [BONN- 1999] à travers laquelle des modèles de prévision ont été développés pour estimer les niveaux de nappes selon les saisons d'une année. Dans notre étude nous avons pu observer cette périodicité sur une journée régie par le lever du jour et la tombée de la nuit. En effet, ce changement étant régulier, il est tout indiqué pour étudier une périodicité qui influe directement la chimie de l'eau au cours du temps. Afin d'illustrer cette périodicité de manière apparente, nous présentons en figure III.5, trois exemples de signaux physico-chimiques, observés tout au long d'une semaine :

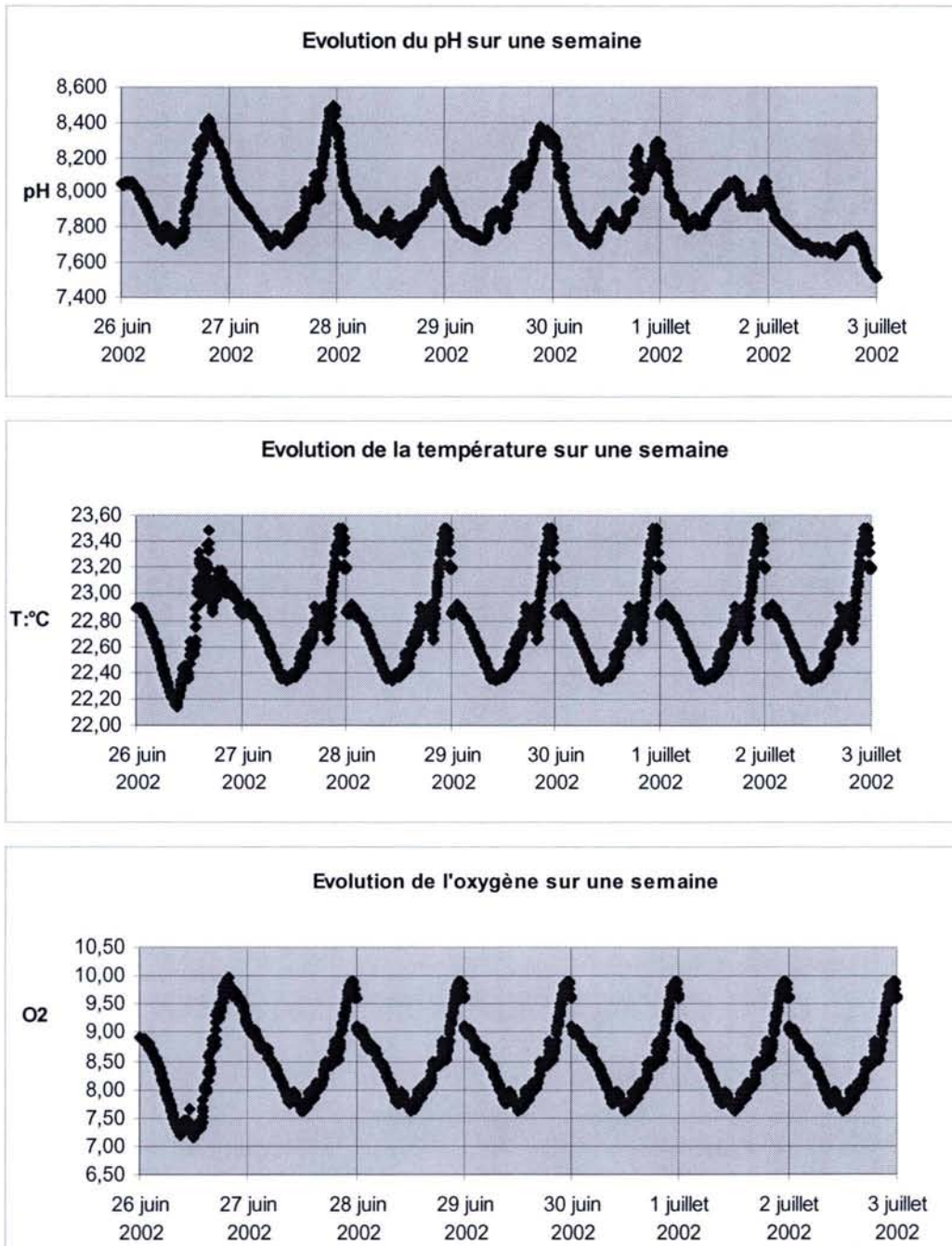


Figure III.5 : « Observation du pH, de la température et de l'oxygène sur la semaine du 26 juin au 2 juillet 2002 »

Ces mesures ont toutes débuté à minuit et se sont effectuées toutes les 5 minutes, aboutissant à 288 points de mesures par journée. L'observation qui en ressort pour le pH est que l'on discerne bien une forme de sinusoïde asymétrique qui se répète selon la même tendance chaque jour, sans qu'aucune ne soit l'identique de l'autre. Il demeure, néanmoins, qu'un cycle de hausse et de baisse du pH, se distingue clairement pour pouvoir envisager une prévision.

Plus nettement, dans les cas de la température et de l'oxygène, en dehors de la première journée, la périodicité est flagrante, pouvant être confortée dans la durée au-delà d'une semaine même pour un système ouvert tel que le nôtre. Nous pouvons même expliquer la différence de signal sur la première journée du 26 juin par rapport aux autres jours, par le fait qu'un phénomène inhabituel a pu survenir.

## II.7 Conclusion

La réaction des capteurs biologiques face à la plupart des tests de toxicité, permet d'établir une contribution qualitative, à une couverture la plus exhaustive possible d'un ensemble de risques et de dégradation d'un cours d'eau.

En effet, les capteurs biologiques réagissent à beaucoup de substances. Leur niveau de performance diffère selon l'organisme test, la réponse exploitée, les techniques de mesure, le polluant (nature, concentration, mode d'action) ou encore les caractéristiques physico-chimiques de l'eau.

Il en résulte qu'il demeure parfois difficile de connaître la cause exacte de leur réaction, sans avoir à combiner les réactions des capteurs physico-chimiques. C'est dans cet esprit que nous avons entrepris une démarche de fusion de données et d'aide à la décision, respectivement décrites dans les deux prochains chapitres, exploitant l'atout que constitue l'association d'informations en parallèle, que sont d'une part celles des capteurs biologiques réagissant précocement à une pollution et d'autre part celles des capteurs physico-chimiques dont le début de réaction donne une indication sur le type de pollution à subir.

Il faut aussi savoir que si les tests de toxicité effectués ne couvrent pas l'ensemble des familles de toxiques référencés, il n'en demeure pas moins que c'est une base de travail pour la construction de la base de connaissance. Celle-ci est soumise à un apprentissage au fil du temps en enregistrant tout événement inédit pouvant survenir. En effet, chaque nouvelle réaction, pendant l'exploitation réelle du dispositif d'alerte, doit aboutir à une analyse laboratoire permettant de déterminer le type de pollution. Cette information validée, enrichit ainsi la base connaissance, ce qui contribue à affiner la détermination des types de pollutions.

Un autre apport de ces tests est l'appréciation de la « robustesse » des capteurs face à des situations de pollutions réelles ainsi qu'un rythme effréné de maintenance et de nettoyage (*voir le détail des coûts et rythme de maintenance en annexe IX*). Cela nous a permis, d'observer le niveau de fiabilité en situation d'utilisation critique, notamment les bio-capteurs, dont les êtres vivants ont subi des concentrations de polluants selon un cycle inhabituel. A cet effet, il a été convenu de renouveler les êtres vivants plus fréquemment que prévus, afin de ne pas les accommoder à un milieu constamment pollué et ne pas perdre ainsi les caractéristiques originelles de ces capteurs en terme de niveau de sensibilité.

Il est à noter d'ailleurs, que les choix de concentrations de polluants s'étant faits sur des bases de seuils tolérés, ils n'ont pas induit de mortalité anormale des bio-capteurs, ce qui prouve leur capacité à enregistrer une pollution. Il n'en demeure pas moins que la mortalité a plus été remarquée lors de manipulations de substitution d'êtres vivants. Cela a notamment été le cas des bactéries qui ne s'adaptaient pas à leur nouveau milieu naturel, de même dans le cas de défauts de fonctionnements du Gymnotox, où la production de micro-bulles due au réchauffement de l'eau et l'absence de dégazage, provoquait une forte mortalité des poissons. Ce type d'incidents malencontreux, a freiné considérablement le rythme des simulations, et a entraîné l'invalidation de certaines mesures.

Tout cela ne remet pas en cause la pertinence de réaction des bio-capteurs, mais dénote une certaine fragilité à manipuler les êtres vivants, selon un protocole complexe à mettre au point. Dans ce cas précis, nous pouvons citer le cas du Dresseina où la méthode de classification sépare les deux canaux sur deux classes différentes. Cela s'explique par le fait que le Dresseina, est un capteur souvent en mode maintenance, du fait de la complexité de la



procédure de changement des moules qui implique un temps d'adaptation faussant ainsi les résultats.

Ces constats, permettent de préjuger sur la mise à niveau et le maintien de certains capteurs au sein de l'installation, c'est ce qui est synthétisé en détail au chapitre V.

Concernant les plans d'expériences, ils n'ont certes pas concerné l'ensemble des tests de toxicité, mais nous ont permis de faire l'économie d'un bon nombre de tests, ainsi qu'une masse moins importante de données à interpréter, sans affecter la qualité de l'information produite.

Au-delà des tests de fiabilités sur les capteurs biologiques, les tests de toxicités ont permis d'évaluer globalement l'ensemble des capteurs pour chaque pollution testée. La richesse livrée par ces mesures brutes passe par l'étude statistique des données. C'est ce que nous avons tenté d'explicitier à travers les tests de corrélations, qui ont permis d'établir certaines tendances, mais ont été aussi employés en tant que méthode d'évaluation des distances dans le cadre des méthodes de classification. En effet, ces méthodes permettent de classer les capteurs par groupes partageant certaines caractéristiques statistiques. Le but consiste à avoir une connaissance à priori des tendances d'évolution des mesures par groupes de capteurs.

C'est aussi ce qu'apportent les décompositions en ACP en terme de classification par ensemble de capteurs corrélés, tout ceci en réduisant l'espace des paramètres permettant d'obtenir une représentation lisible sous forme de carte factorielle. L'approche ACP, a aussi le mérite, dans notre contexte, de pouvoir évaluer la cohérence des capteurs. En effet, la propriété de l'ACP pour la détection de défauts et d'anomalies a été mise à l'épreuve dans l'exemple traité précédemment, cela nous a permis, en l'occurrence, d'évaluer les performances de l'Aquatox qui est un capteur prototype biologique, sensé réagir lors du test de pollution agricole et être communément corrélé avec les autres capteurs ayant réagi à cette pollution. Cela n'a pas été le cas dans la plus part de nos analyses, ce qui a été l'une des raisons d'avoir éliminé ce capteur de l'installation.

L'ensemble de ces approches introduit aussi la notion de masques de pollutions abordée au prochain chapitre. Celle-ci permet d'établir un profil de comportement des capteurs pour chaque pollution testée.

Ces différents résultats implémentés dans la base de connaissance du système de traitement de données, contribuent à enrichir la connaissance à priori que doit avoir l'opérateur chargé de la surveillance du milieu. Il en ressort qu'une exploitation optimale de ce flux de données passe par une synthèse de celles-ci. C'est ce que notre démarche de fusion de données, abordée au prochain chapitre, permet de quantifier qualitativement, fournissant ainsi l'information précise nécessaire à notre système d'aide à la décision.

## **Bibliographie :**

- [AZAI - 2003] Azaïs J.M., Composantes de la variabilité : Plans d'expériences, Publications du laboratoire de statistique et probabilité, Université Paul Sabatier, janvier 2003.
- [AZZ(a) - 2002] Azzam A.A., "Etude statistique et traitement de données pour la synthèse d'alertes à partir d'une station multi-capteurs", in : Actes Journées Scientifiques 2002 du GMCEA, p. 157--170, Nantes, Octobre 2002.
- [BAI-1996] Baillargeon G., Méthodes statistiques de l'ingénieur, volume2, Université du Québec à Trois-Rivières, Les Editions SMG, 1996.
- [BENO-1992] Benoist D., Notions sur les plans d'expériences, Publication de l'Institut Français du Pétrole, Editions TECHNIP, juillet 1992.
- [BONN- 1999] Liliane BONNAL, Pascal FAVARD : Prévisions à court terme du niveau des aquifères: le cas de la nappe de Beauce Cahiers d'économie et sociologie rurales, n° 53, 1999.
- [Décret n°89-3] Décret du 3 janvier 1989 relatif aux eaux destinées à la consommation humaine à l'exclusion des eaux minérales naturelles.
- [DAG - 1997] Dagnelie P., "Plans d'expériences : Application à l'entreprise", Chapitre2 : *La planification des expériences et l'analyse de la variance : Une introduction*, Jean-Jacques DROESBEKE, Jeanne FINE, Gilbert SAPOTRA éditeurs, Editions TECHNIP, juillet 1997.
- [ETU-2001] CIRSEE-NANCIE : Etude statistique sur les données de la station multi-capteurs LIFE 99, 2001.
- [FOW-2000] Fowlkes W.Y. & Creveling C. M., L'Ingénierie Robuste : Méthodes Taguchi en conception, Editions DUNOD 2000.
- [JAMBU - 2003] Jambu M., Méthodes de base de l'analyse des données 3<sup>ème</sup> Edition, EYROLLES 1999.
- [JOHA-2002] Johan T. & Svante W. *Introduction to Statistical Experimental Design - What is it? Why and Where is it Useful?*. Editorial in homepage of Chemometrics, University of Queensland, Australia & Umeå University, Sweden, Août 2002.
- [MAT - 1997] Mathieu D., Phan-Tan-Luu R., "Plans d'expériences : Application à l'entreprise", Chapitre6 : *L'approche Taguchi*, Jean-Jacques DROESBEKE, Jeanne FINE, Gilbert SAPOTRA éditeurs, Editions TECHNIP, juillet 1997.
- [METH-2003] Azzam A.A., Andreoli A., Thomas M., Cecile J.L., Jonca G., Guide méthodologique du projet Life ENV/F/000492, Protection et surveillance multi-paramètres des ressources aquatiques, NANCIE avril 2003.

- [SY09-2004] Ambroise C. et Govaert G., Analyse de données et data mining, Support de cours du module SY09 de l'Université de Technologie de Compiègne, printemps 2004 {4}.
- [SCH – 1984] Scherrer, B., Biostatistique, Édition Chicoutimi: Gaëtan Morin, section 4.2.2, chapitre 17, 1984.
- [VIG – 1991] Vigier M., Pratique des plans d'expériences : Méthodologie de Taguchi et compléments, Les Editions d'Organisation, 1991.

## ***Références Internet***

{1} Notions de statistiques :

[http://www.math-info.univ-paris5.fr/smel/lexique/t\\_student/t\\_student.html](http://www.math-info.univ-paris5.fr/smel/lexique/t_student/t_student.html) et

<http://www.math-info.univ-paris5.fr/smel/cours/ts/node14.html>

{2} Notions de statistiques :

[http://www.imep-cnrs.com/avner/Dea\\_u3/node7.html](http://www.imep-cnrs.com/avner/Dea_u3/node7.html)

{3} Notions de statistiques :

<http://rfv.insa-lyon.fr/~jolion/STAT/>

{4} Notions de statistiques :

[http://www.hds.utc.fr/sy09/documents/poly\\_1.pdf](http://www.hds.utc.fr/sy09/documents/poly_1.pdf)

## CHAPITRE III : Fusion de données dans un environnement multi-capteurs

### III.1 La fusion de données

La fusion de données, étape qui précède l'aide à la décision, peut être établie selon plusieurs approches. Notre souci premier a été de développer un système le plus réactif possible face à une pollution pouvant contaminer une agglomération en l'espace d'une demi-heure dans le pire des cas. Cette exigence passe par l'utilisation de méthodes de calculs temps réel, faisant abstraction de calculs rétroactifs. Cette contrainte implique la combinaison de méthodes simples. Celles-ci, doivent s'intégrer dans une architecture de décision compensant au mieux le côté élaboré que peuvent apporter les méthodes de fusions classiques mais inadaptées à notre contexte temps réel.

La fusion de données est une discipline transversale à tous les domaines ayant recours à une analyse de données de sources hétérogènes. On l'utilise essentiellement dans les domaines de l'automatique, de la défense, de l'aérospatiale, de l'imagerie et de l'environnement. De ce fait, plusieurs définitions sont proposées dans la littérature :

- En tant que fusion d'informations, c'est l'ensemble des théories et des outils pour exploiter la synergie de différents jeux de données de sorte que le résultat basé sur les données fusionnées soit meilleur que le résultat basé sur chaque source individuelle {1}.
- Selon [WALD - 1999]: *"Data fusion is a formal framework in which are expressed the means and tools for the alliance of data originating from different sources. Data fusion aims at obtaining information of greater quality; the exact definition of 'greater quality' will depend upon the application"*.
- Selon le NOAA (National Oceanic and Atmospheric Administration) data Centers : *"Data fusion is the seamless integration of data from disparate sources" {2}*.

Les techniques de fusion de données combinent des données de capteurs multiples, et l'information relative aux bases de données associées, afin que l'on puisse approcher l'exactitude que l'on pourrait obtenir lors de l'utilisation d'un seul et unique capteur [LLIN-1998]. En effet, les données multi-capteurs fournissent des données de base simples ayant l'avantage d'être significatives statistiquement. Leur fusion permet, d'autre part, d'augmenter l'exactitude avec laquelle une mesure peut être observée et caractérisée.

Un modèle de fusion de données a été proposé par le JDL (Joint Directors of Laboratories) groupe de travail sur la fusion de données [LEXI-1991], [KESS-1992], dans lequel sont identifiés différents niveaux d'interactions, associés à des sources données. Celles-ci sont par la suite fusionnées pour être destinées à l'opérateur humain, comme indiqué à la figure IV.1 :

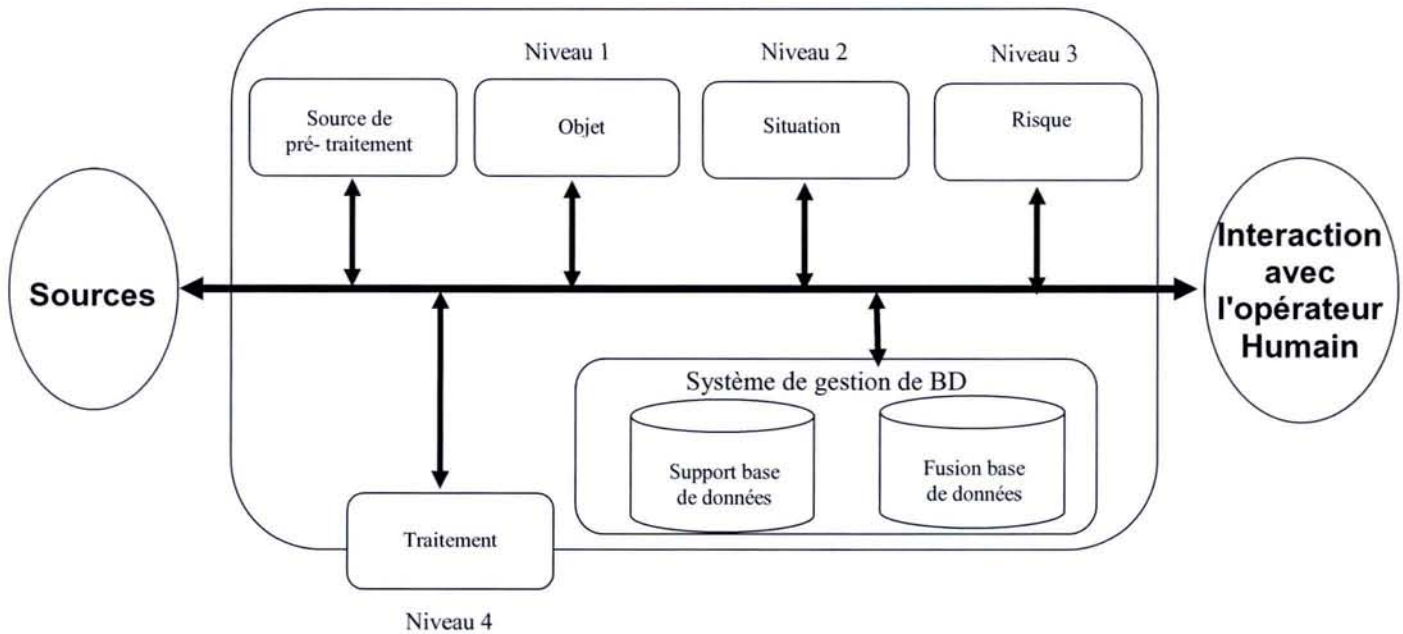


Figure III.1 : «Modèle de processus de fusion de données»

Le modèle proposé en figure III.1 illustre parfaitement le contexte de la fusion de données auquel est confronté un opérateur humain. Les différents éléments du modèle peuvent être présentés comme suit :

Les sources : fournissent des informations destinées aux différents niveaux allant des données brutes issues des capteurs à l'information fusionnée destinée à l'opérateur humain.

Source de pré-traitement : permet au processus de fusion de données de se concentrer sur les données les plus pertinentes concernant la situation actuelle réduisant ainsi la charge du processus de fusion de données. Ceci est accompli à travers une observation intermédiaire permettant d'assigner les données aux processus appropriés.

Objet (Niveau 1) : niveau de traitement 1 combinant l'information localisée, paramétrique, et d'identité afin d'assurer la représentativité des différents objets. Les fonctions principales sont l'estimation, la prévision, et l'assignation des classifications.

Situation (Niveau 2) : niveau de traitement 2 essayant de développer une description contextuelle des événements observés, ce qui détermine la signification d'un ensemble d'entités et incorpore des conditions ambiantes, des connaissances à priori et des observations.

Risque (Niveau 3) : niveau de traitement 3 projetant la situation actuelle dans le futur dans le cadre de situations de risques. Dans notre cas cela correspond à l'étude de vulnérabilité des milieux aquatiques.

Traitement (Niveau 4) : un méta-processus, concerné par d'autres processus. Les fonctions principales du niveau 4 sont : suivre en temps réel et à long terme l'exécution de la fusion de données, identifier l'information requise pour améliorer les niveaux de fusion de données, et assigner les actions aux capteurs pour réaliser les objectifs.

Système de base de données : c'est la fonction auxiliaire de la fusion de données, qui permet de gérer l'extraction de données, le stockage, l'archivage, la compression, et la protection de données.

Interaction opérateur humain : En plus de fournir une interface entre l'entrée humaine et la communication de la fusion de données résultante, elle inclut des méthodes d'aide à la décision destinées à l'opérateur.

Sur ces principes il est possible de fonder l'approche globale de fusion de sachant que l'aide à la décision est abordé au prochain chapitre.

Nous sommes ainsi amenés à lier le concept de fusion de données à celui d'un système multi-capteurs d'où la multiplicité de méthodes possibles pour cette problématique.

Selon les domaines d'utilisation, plusieurs méthodes analytiques peuvent être envisagées. Nous citerons les méthodes mathématiques pouvant être adaptées à des systèmes multi-capteurs liés à l'environnement :

- théorie des réseaux de neurones [DAI-1999], [SHANG-1999]
- théorie des réseaux Bayésiens [DEG-1992], [JEN-1996], [PEARL-1988],
- théorie de l'évidence (Dempster-Shafer) [BLOCH-1998], [PARI-2001], [SAND-2002], [SMET – 2000], [TCHA - 1997], [VANO – 2000],
- la logique floue [LI-2000], [KUO-2000].

Bien que ces méthodes ne soient pas directement exploitées dans l'étude, elles contribuent au développement de la méthode préconisée que l'on pourrait qualifier d'hybride. Le choix pragmatique est lié aux contraintes temps réel du dispositif. En effet, les méthodes classiques permettent la résolution de problèmes de perception et de décision par apprentissage en modélisant la connaissance sur le système observé, mais ne permettent pas un traitement temps réel d'un flux de données, pour produire une alerte dont la fiabilité dépend du temps nécessaire à sa production.

**Il faut convenir que dans ce contexte, on entend par temps réel une notion d'instantanéité du résultat différé du temps de rapatriement de données et de calcul qui n'excède pas l'ordre de la minute. Cette notion est valable tout au long de notre contexte, sur ce même principe qu'est l'instantanéité, après exécution d'un ensemble d'opérations.**

### **III.2 Modélisation graphique et application à la fusion de données :**

Un système d'aide à la décision peut être fondé sur un modèle de réseau Bayésien. Ce réseau est un langage de représentation de connaissances permettant de préciser graphiquement les dépendances probabilistes parmi les propositions et les événements [PEARL-1988].

Les réseaux Bayésiens ou diagrammes d'influence (*belief network*) constituent un des modèles de représentation des connaissances utilisables en analyse de décision. Un réseau Bayésien, également appelé réseau causal est un graphe dirigé, acyclique dans lequel les nœuds représentent les variables et les arcs précisent les dépendances probabilistes entre les variables. Il permet d'afficher graphiquement les variables d'un problème de décision et les relations ou influences entre ces variables, qui peuvent mener à des décisions complexes.

Les réseaux Bayésiens centrent l'attention du décideur exclusivement sur les composants du problème en relation avec la tâche de décision. Exclure l'information non pertinente du diagramme d'influence facilite le travail du décideur. Cela lui permet de gagner du temps puisqu'il existe moins de variables à interpréter [DEG-1992].

Dans notre contexte, une étude préliminaire de modélisation par graphes Bayésiens a été conjointement menée avec le CIRSEE (Centre International de Recherche sur l'Eau et l'Environnement) [RAPP-2001] sur une base de données d'environ 7900 données pour 7 variables physico-chimiques, assez fournie pour envisager la modélisation des relations entre variables par le réseau Bayésien dont la structure est représentée en figure IV.2 :

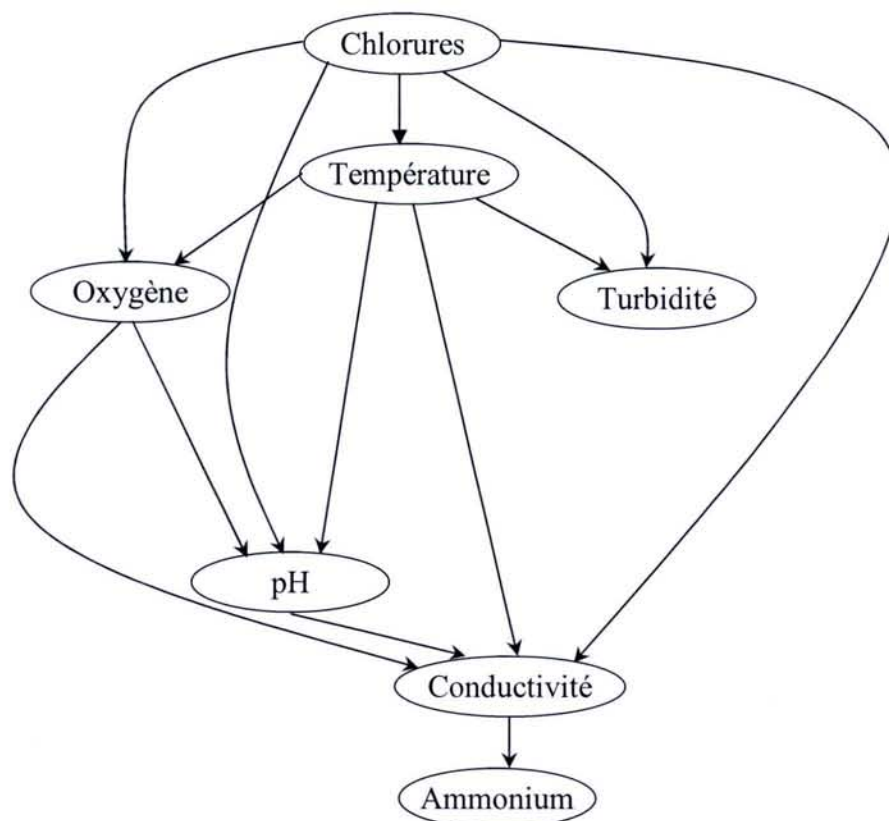


Figure IV.2: Modélisation graphique des paramètres physico-chimique

Ce type de représentation, ne peut être appliqué de manière précise dans un environnement aussi fluctuant que celui d'un cours d'eau, notamment concernant les probabilités de franchissement entre paramètres qui varient considérablement d'un jeu de données à un autre. Néanmoins, l'intérêt d'une telle représentation graphique, permet de faire apparaître l'existence de corrélations selon les probabilités de franchissement préétablies.

### III.3 Méthodologie

#### III.3.1 Introduction

Notre approche basée sur la combinaison de méthodes de calculs ciblées et immédiates, permet de fournir à l'opérateur des données synthétiques et exploitables dans des délais raisonnables, et prêtes à être intégrées dans un processus de décision.

Dans ce cadre, le principe consiste à établir pour chaque toxique testé, des seuils de réactions propres à chaque capteur. La détermination de ces seuils se fait sur plusieurs états de la mesure brute, à laquelle sont appliquées des méthodes de **comparaisons par fenêtres glissantes**. Ces méthodes sont explicitées tout au long du point suivant, au travers de méthodes d'identification de rupture dans les séries temporelles. Elles permettent d'établir des pré-alertes avant l'atteinte du niveau de tolérance normal fixé par les extraits du *décret n° 89-3 du 3 janvier 1989 (amendé) relatif aux eaux destinées à la consommation humaine à l'exclusion des eaux minérales naturelles* détaillés en « Annexe III ».



Une fois ces seuils recensés, un profil de variation de grandeur de chaque capteur est établi pour chacune des pollutions, permettant de prévoir l'apparition probable de celles-ci.

Plus concrètement, l'évaluation des paramètres se base sur des critères de variations. L'affectation de niveau de seuils aux paramètres ayant réagi à une pollution, détermine le profil de celle-ci. Ces évaluations initiales des seuils ont été possibles grâce aux tests de toxicité. Le recours à une base de connaissance pour des événements similaires contribue à l'affinement de ces profils tout au long de l'exploitation du système. L'architecture de fusion, représentée en figure IV.3 reflète notre approche, comme préconisé dans [Gee-2000]. Cette architecture met en scène le scénario d'une fusion sur la base d'un système multi-capteurs, d'une situation donnée et d'une base de connaissance.

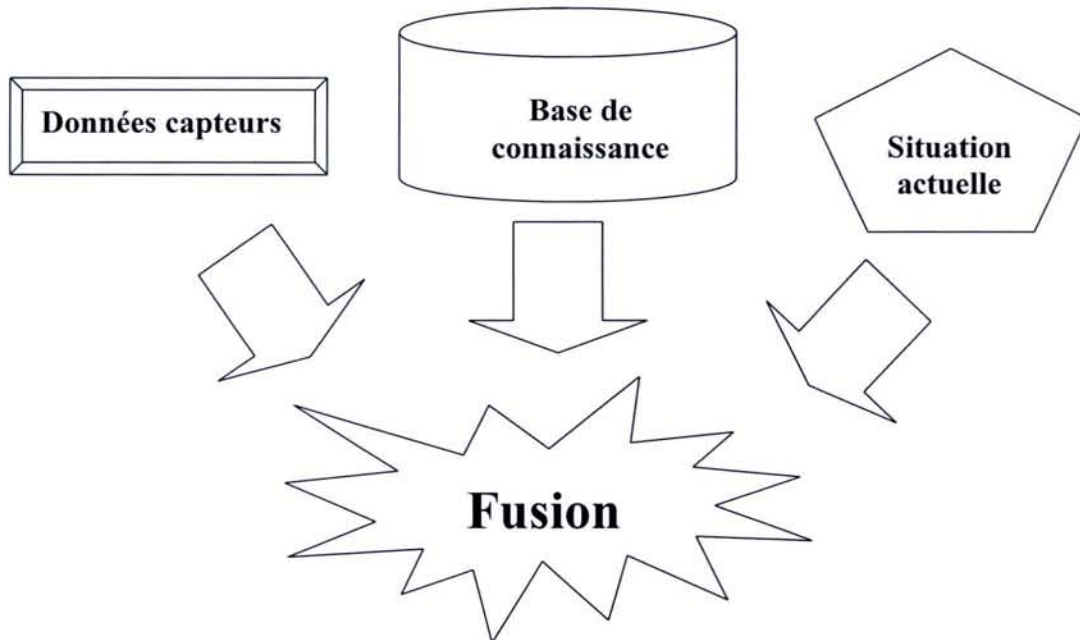


Figure IV.3 : Réseau d'un système de fusion [Gee-2000]

Cette base de connaissance intègre le seuillage sur valeurs brutes, qui permet d'intégrer l'approche élémentaire classique consistant à évaluer la réaction des capteurs face à une pollution. En effet, ce seuillage sur les grandeurs brutes s'avère nécessaire, dans le sens où il contribue à évaluer la performance de notre approche basée sur des données dérivées des valeurs brutes. Notre système passant par des systèmes de calculs dérivés, doit nécessairement réagir avant les seuils bruts, or si une pollution est détectée par un seuillage simple avant que notre système de pré-alertes n'aie réagi, cela signifie qu'il y a dysfonctionnement.

Ce type d'incidents, n'est en effet pas à exclure. Il conduit dans le cas de capteurs uniques ou n'ayant pas d'équivalent en terme de sensibilité sur un type de pollution, à un passage sur les bassins de sécurité, le temps du rétablissement de la panne.

Le seuillage de base s'effectue sur les seuils de tolérance des données brutes, mentionnés dans le décret de l'« Annexe III ». Nous avons, en l'occurrence, appliqué ces recommandations sur le seuillage des valeurs brutes des paramètres physico-chimiques de la station. Se pose alors le problème de l'évaluation du seuillage des capteurs biologiques correspondant à des indicateurs d'anormalités.

Il faut savoir, par exemple, que concernant le Gymnotox fonctionnant sur une base de trois poissons, la plage de fréquence balayée par l'espèce varie de 750 Hz à 1250 Hz et que le calage du signal se fait selon le poisson. Cela correspond à un étalonnage relatif, ne

correspondant pas à fixer un seuil absolu. Il en ressort que la notion de seuillage sur valeur brute dans ce cas n'a aucun sens.

Cela nous amène à avoir un autre type d'approche consistant à évaluer une variation au cours du temps, en l'occurrence, pouvoir identifier une variation dans une série temporelle. Cette approche, consiste à se baser sur les tests de pollutions pour établir les états de pré-alertes conditionnés selon un ou plusieurs critères, caractérisés par des seuils particuliers sur fenêtres glissantes, moyennes ou dérivées. Notre choix pour ces méthodes de détection sur séries temporelles, s'avère suffisant dans notre contexte de génération de pré-alertes, permettant de détecter le début d'une variation brusque sur le signal à évaluer.

### III.3.2 Seuillage par Identification d'une variation dans les séries temporelles

Avant d'explicitier notre méthodologie de détection de pollution, nous présentons brièvement dans ce point les méthodes principales de rupture de modèle dans les séries temporelles, puis nous définissons et comparons trois types de tests statistiques permettant la détection de changement brusque dans une série temporelle.

Dans ce contexte, deux approches principales permettent l'évaluation d'une série temporelle :

- D'une part l'approche utilisant un modèle, basée sur le test de l'erreur de prédiction à un pas qui est défini comme étant la différence entre la valeur du signal et la prédiction calculée à l'aide du modèle utilisant les valeurs du signal aux instants précédents. Cette approche est plutôt utilisée sur un signal susceptible d'être représenté par le même modèle en dehors de l'apparition d'événements particuliers [SCHE – 1991], ce qui n'est pas le cas adéquat dans notre contexte qu'est le milieu naturel.
- D'autre part l'approche permettant de comparer deux modèles, sur la base de fenêtres glissantes, permettant une comparaison relative du signal par rapport à son passé immédiat, c'est dans ce contexte que notre méthodologie est développée.

A cet effet, la première approche de ce type a été introduite par Bodenstein et Praetorius [BODE – 1977] pour la segmentation de signaux EEG (Electro Encéphalo Graphique). Elle se schématise, en l'occurrence, par deux fenêtres glissantes juxtaposées.

A titre de rappel, la fenêtre glissante est une section du signal de durée fixe  $T$  se déplaçant sur l'axe temporel d'un pas  $\tau$ .

$$\begin{cases} T = N \Delta \\ \tau = p \Delta \end{cases} \quad (4.1)$$

où  $\Delta$  est la période d'échantillonnage du signal,  $N$  et  $p$  sont deux entiers représentant respectivement la taille et le nombre de mesures de déplacement de la fenêtre, ce qui correspond à  $N$  échantillons du signal pris consécutivement à des intervalles de temps  $\Delta$ . Un déplacement équivaut à retirer de cette fenêtre les  $p$  mesures les plus anciennes et à rajouter à la place  $p$  nouvelles mesures.

Par rapport à l'approche proposée initialement, où il était question de fixer une fenêtre de référence au début du signal, dans notre cas précis, il est question de comparer une portion récente du signal à une portion de son passé immédiat, afin de limiter les fausses alarmes liées au contexte de milieu naturel. En effet, la portion d'un signal en début de journée différera par

rapport à d'autres portions de signal tout au long d'une journée sans pour autant déclencher une alerte de pollution. Il en ressort que la variante de comparaison de fenêtres glissantes juxtaposées, notamment utilisée dans [VAEZ –1990], se présente en figure IV.4 comme suit :

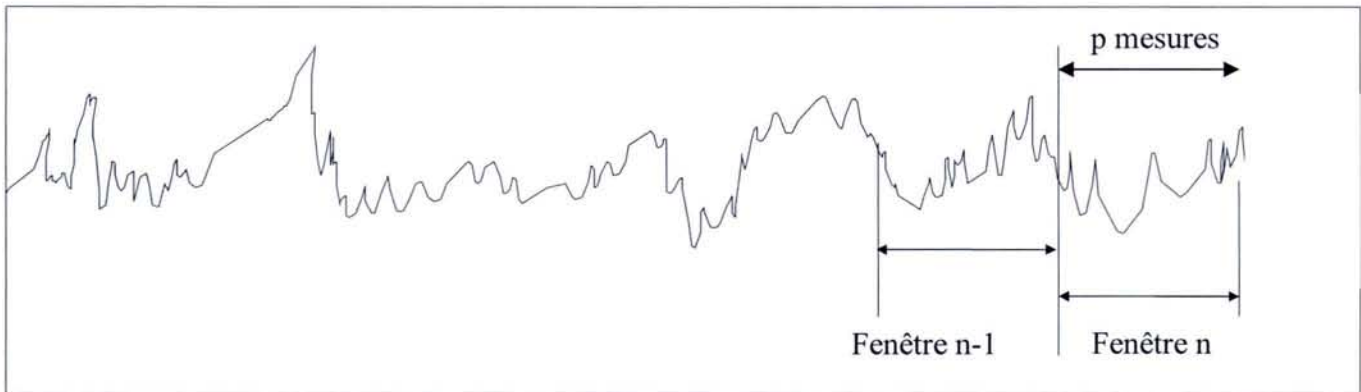


Figure IV.4 : Fenêtres glissantes juxtaposées

La première fenêtre, appelée « fenêtre de référence », est comparée à la fenêtre qui la suit, sur la base d'une fonction discriminante ou indicateur de rupture. Les fenêtres sont par la suite, successivement déplacées du pas de déplacement  $\tau$ . Ainsi, la disposition des deux fenêtres glissantes permet la comparaison de deux sections différentes du signal contenant les mesures les plus récentes.

Une autre approche, utilisant deux modèles a été introduite par Basseville et Benveniste [BASS – 1983] ainsi que Appel et Brandt [APPE – 1983], qui consiste à avoir deux fenêtres glissantes sur deux horizons temporels, l'un ne tient compte que du court terme (fenêtre glissante) et l'autre conserve tout le passé du signal à partir d'une date de référence (fenêtre croissante).

Cette méthode a l'avantage d'être peu affecté par un changement brusque du signal, et de ce fait réduit les fausses alarmes. Dans le cas où il est tenu compte de tout le passé du signal, on peut aussi parler de fenêtre croissante comme schématisé en figure IV.5, il est donc question de comparer une portion récente du signal par rapport la totalité de son passé.

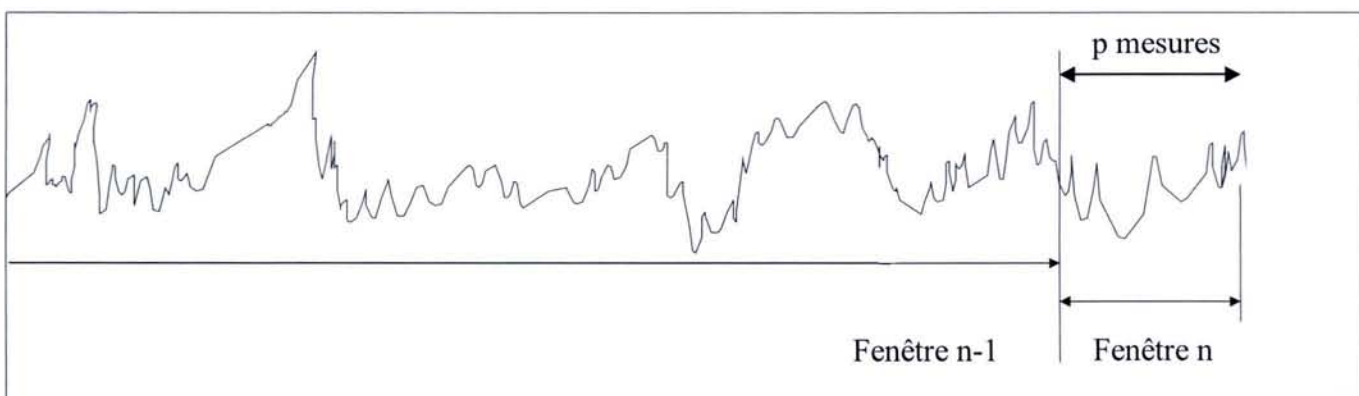


Figure IV.5 : Cas de la fenêtre croissante associé à une fenêtre glissante

Cette fenêtre « croissante » comme son nom l'indique, croît au cours du temps et tient certes compte des variations du signal, mais n'est pas intéressante dans notre cas où l'exploitation du signal est dynamique. Cela engendrerait une fenêtre de plus en plus grande sur le long terme, devenant lourde à gérer en terme de calcul temps réel. D'autre part, le calcul de la moyenne ou variance associée à la « fenêtre n-1 » long terme, reflétera de moins

en moins l'évolution du signal au fil du temps et deviendra inefficace en terme de précision sur le présent immédiat du signal. A titre d'illustration dans notre contexte d'environnement naturel, les changements de niveaux des signaux d'une saison à une autre, affectent considérablement la différence de moyenne entre le long terme et le court terme.

Pour toutes ces raisons, nous avons opté pour l'autre type de fenêtre qui ne tient compte que du court terme, en l'occurrence les fenêtres glissantes juxtaposées comme schématisées en figure IV.4. Cette approche favorise la détection d'un changement brutal sur le court terme du signal et de ce fait tiens compte de l'évolution relative de celui-ci au cours du temps, notamment dans le cadre des changements saisons.

La méthode de comparaison de modèles étant choisie, plusieurs fonctions discriminantes (ou fonction de rupture) et calcul de distances sont possibles, pour permettre d'identifier un changement brutal dans le signal.

Dans ce qui suit, nous citons brièvement deux types de calculs de distances basés sur des modèles AR (Auto-Régressif), mais se présentant comme étant des méthodes lourdes pour une implantation temps réel. Nous détaillons par la suite trois fonctions discriminantes que nous avons choisies de comparer, qui permettent pour leur part une implantation temps réel.

En terme de distance, l'une des premières connues, est la divergence de Kullback, issue des développements effectués par Kullback [KULL – 1959] ainsi que des travaux de Shore [SHOR – 1981] sur la cross-entropie conditionnelle calculée entre deux lois de densités. Cette distance a été notamment utilisée dans [NOWA – 1989] pour la détection de défauts dans les séries temporelles.

Une autre distance, issue du calcul du rapport de vraisemblance généralisé, est la distance de Chernoff principalement développée dans [APPE – 1983] dans le cadre de l'identification récursive. Il faut rappeler que le test rapport de vraisemblance généralisé, a été introduit par Willsky et Jones [WILL – 1974], permettant le calcul récursif du rapport de vraisemblance.

Concernant, les trois tests statistiques que nous avons choisis de détailler et de comparer, en l'occurrence, le test des sommes cumulées de Hinkley, test de comparaison de moyennes et le test de comparaison de variances, ils permettent l'implantation temps réel recherchée et surtout de détecter un changement de niveau du signal brut ou d'identifier une évolution de pente brusque.

La problématique qui suit, consiste à régler la taille des deux fenêtres glissantes juxtaposées puis quantifier les changements brusques à travers des seuillages.

### III.3.2.1 Test des sommes cumulées de Hinkley par détection de sauts de moyenne

Introduit par Page [PAGE - 1954], le test des sommes cumulées a été étudié par Hinkley [HINK 1971] en vue de son application à la détection de saut de moyennes dans les séquences aléatoires gaussiennes. Ce test a été notamment appliqué dans [MERCK – 1988] pour la localisation des défauts sur des produits sidérurgiques. Dans le cas échéant, nous définissons le test, en considérant les moyennes connues avant et après le saut.

Soit  $\{ y_k \}$  une séquence gaussienne de variance  $\sigma^2$  et de moyenne telle que sa moyenne :

$$\mu_n = \begin{cases} \mu_0 & \text{si } n \leq r-1 \\ \mu_1 & \text{si } n \geq r \end{cases} \quad (4.2)$$

La problématique consiste à détecter un changement dans la moyenne  $\mu_n$ , d'estimer l'instant de changement  $r$  et éventuellement les moyennes  $\mu_0$  et  $\mu_1$  avant et après le saut.

La détection consiste à choisir entre l'hypothèse de non-changement :

$H_0 : r > n$  à laquelle nous associons, la loi de probabilités des observations  $y_k, p_0(y_k)$

et l'hypothèse de changement :

$H_1 : r \leq n$  à laquelle nous associons, la loi de probabilités des observations  $y_k, p_1(y_k)$

Le rapport de vraisemblance calculé entre ces deux hypothèses est :

$$\prod_{k=r}^n \frac{P_1(y_k)}{P_0(y_k)} \quad (4.3)$$

où  $p_i$  est la densité de probabilités gaussienne des observations ( $y_n$ ) de moyenne  $\mu_i$  ( $i = 0, 1$ ).

Son logarithme est donné par :

$$\begin{aligned} \Lambda_n(r) &= \frac{(\mu_1 - \mu_0)}{\sigma^2} \sum_{k=r}^n \left( y_k - \frac{(\mu_1 + \mu_0)}{2} \right) \\ &= \frac{1}{\sigma} S_r^n(\mu_0, \nu) \end{aligned}$$

$$\text{où } S_i^j(\mu_0, \nu) = \nu \sum_{k=i}^j \left( y_k - \mu_0 - \frac{\nu}{2} \right) \quad (4.4)$$

La notation  $S_i^j$  s'appelle somme cumulée et  $\nu = \mu_1 - \mu_0$  représente l'amplitude du saut.

L'estimation, au sens du critère du maximum de vraisemblance sous l'hypothèse  $H_1$ , de l'instant du saut  $r$ , est donnée par :

$$\begin{aligned} \hat{r}_n &= \arg \max_{1 \leq r \leq n} \left( \prod_{k=0}^{r-1} p_0(y_k) \prod_{k=r}^n p_1(y_k) \right) \\ &= \arg \max_{1 \leq r \leq n} S_r^n(\mu_0, \nu) \end{aligned} \quad (4.5)$$

Sous l'hypothèse  $H_1$ , la quantité  $\Lambda_n(r)$  doit être significativement différente de zéro.

Par conséquent, on peut utiliser le détecteur suivant :

$$G_n = \max_r S_r^n(\mu_0, \nu) \begin{matrix} > \\ < \end{matrix} \begin{matrix} H_1 \\ S \\ H_0 \end{matrix} \quad (4.6)$$

où  $S$  est un seuil.

En définitive, l'hypothèse  $H_1$  est retenue lorsque, à l'instant  $k$ , on a pour la première fois :

$$G_n = S_1^n(\mu_0, \nu) - \min_{1 \leq k \leq n} S_1^k(\mu_0, \nu) > S \quad (4.7)$$

L'estimation de l'instant de saut est donc le premier instant pour lequel la valeur minimale  $S_1^k$  a été atteinte.

Le détecteur peut être calculé récursivement dans le cas d'un saut positif ainsi :

$$G_n = G_{n-1} + y_n - \mu_0 - \frac{\nu}{2} \quad (4.8)$$

dans le cas d'un saut négatif, le détecteur est calculé ainsi :

$$G_n = G_{n-1} + y_n - \mu_0 + \frac{\nu}{2} \quad (4.9)$$

### III.3.2.2 Test de comparaison de moyennes

Ce test de comparaison correspond à celui utilisé en statistique pour comparer deux échantillons issus de deux populations de moyennes inconnues, de variances inconnues, mais supposées égales [C.E.A - 1978].

Si les valeurs, utilisées appartiennent à la même population et sont normalement distribuées, alors la fonction  $f(n)$  suit une distribution de *Student* à  $(N_1 + N_2 - 2)$  degrés de liberté. Pour deux fenêtres glissantes et juxtaposées, la fonction de comparaison  $f(n)$  se présente ainsi :

$$f(n) = \frac{m_1(n) - m_2(n)}{\sqrt{\frac{\sigma_1^2(n)}{N_1} + \frac{\sigma_2^2(n)}{N_2}}} \quad (4.10)$$

où  $m_1(n)$  et  $m_2(n)$  sont les  $n^{\text{ièmes}}$  moyennes arithmétiques du signal calculées sur les fenêtres glissantes, et où  $\sigma_1^2(n)$  et  $\sigma_2^2(n)$  sont les variances de ce signal, estimées à partir des fenêtres glissantes respectives.  $N_1$  et  $N_2$  représentent les tailles respectives des fenêtres glissantes «  $n$  » et «  $n - 1$  ».

Les deux moyennes arithmétiques sont calculées par les sommations qui suivent, où  $k$  représente l'instant courant, défini dans la relation (4.10),  $n$  l'indice de la fenêtre glissante et  $p$  le pas de déplacement des fenêtres.

$$K = np + N_1 + N_2 \quad (4.11)$$

Sur la fenêtre 1 de taille  $N_1$ , la  $n^{\text{ième}}$  moyenne à l'instant  $k$  est donnée par l'expression :

$$m_1(n) = \frac{1}{N_1} \sum_{i=0}^{N_1-1} \Delta y_{k-i} \quad (4.12)$$

Cette expression se simplifie en remplaçant  $\Delta y_{k-i} = y_{k-i} - y_{k-i-1}$  ce qui donne :

$$m_1(n) = \frac{1}{N_1} \left[ y_k - y_{k-N_1} \right] \quad (4.13)$$

Sur la fenêtre 2 de taille  $N_2$ , l'expression de la moyenne est la suivante :

$$m_2(n) = \frac{1}{N_2} \sum_{i=N_1}^{N_1+N_2-1} \Delta y_{k-i} \quad (4.14)$$

Ce qui donne après simplification :

$$m_1(n) = \frac{1}{N_2} \left[ y_{k-N_1} - y_{k-N_1-N_2} \right] \quad (4.15)$$

Concernant le calcul des variances  $\sigma_1^2(n)$  et  $\sigma_2^2(n)$ , il est estimé à partir des fenêtres glissantes selon les formules suivantes :

$$\sigma_1^2(n) = \frac{1}{N_1-1} \left\{ \sum_{i=0}^{N_1-1} \Delta y_{k-i}^2 - N_1 [m_1(n)]^2 \right\} \quad (4.16)$$

$$\sigma_2^2(n) = \frac{1}{N_2-1} \left\{ \sum_{i=N_1}^{N_1+N_2-1} \Delta y_{k-i}^2 - N_2 [m_2(n)]^2 \right\} \quad (4.17)$$

Ces expressions de moyennes et de variances peuvent être écrites de manières récursives, pour une fenêtre glissante de taille N, selon les expressions générales suivantes :

$$m(n) = m(n-1) + \frac{1}{N} \left\{ \sum_{i=0}^{p-1} \Delta y_{k-i} - \Delta y_{k-N_1-i} \right\} \quad (4.18)$$

$$\sigma^2(n) = \sigma^2(n-1) + \frac{1}{N-1} \left\{ \sum_{i=0}^{p-1} (\Delta y_{k-i}^2 - \Delta y_{k-N-i}^2) - N [m^2(n) - m^2(n-1)] \right\} \quad (4.19)$$

Il en ressort de ces deux dernières expressions, que l'algorithme de calcul récursif sur fenêtre glissante se base respectivement sur les deux sommateurs suivants :

$$a_k = \sum_{i=0}^{p-1} (\Delta y_{k-i} - \Delta y_{k-N_1-i}) \quad (4.20)$$

$$b_k = \sum_{i=0}^{p-1} (\Delta y_{k-i}^2 - \Delta y_{k-N-i}^2) \quad (4.21)$$

Après simplification, les sommateurs s'écrivent :

$$a_k = [y_k - y_{k-p} - y_{k-N} + y_{k-N-p}] \quad (4.22)$$

$$b_k = b_{k-1} + \Delta y_{k-N-p}^2 + \Delta y_k^2 - \Delta y_{k-N}^2 - \Delta y_{k-p}^2 \quad (4.23)$$

Le nombre d'opérations lors du calcul de la fonction de comparaison de moyennes peut encore être réduit en prenant un pas de déplacement égal à celui de la taille des deux fenêtres de tailles  $N_1$  et  $N_2$  égales, donc  $p = N_1 = N_2$ , ce qui revient à écrire :

$$\left\{ \begin{array}{l} m_1(n) = m_2(n-1) \\ \text{et} \\ \sigma_2^2(n) = \sigma_1^2(n-1) \end{array} \right. \quad (4.24)$$

### III.3.2.3 Test de comparaison de variances

Ce test de comparaison, au même titre que le précédent, correspond à celui utilisé en statistique pour comparer deux échantillons issus de deux populations de moyennes inconnues, de variances inconnues, mais supposées égales [C.E.A - 1978].

Ce test représenté par la fonction de comparaison de variances  $f(n)$  définie en (4.24), permet de décider, compte tenu de la loi de distribution supposée de deux populations, si les variances des échantillons sont égales ou non. Dans le cadre de valeurs appartenant à la même population qui est normalement distribuées, la fonction  $f(n)$  suit une distribution de *Fischer* qui s'écrit :

$$f(n) = \frac{\sigma_1^2(n)}{\sigma_2^2(n)} \quad (4.25)$$

$\sigma_1^2(n)$  et  $\sigma_2^2(n)$  sont les  $n^{\text{ièmes}}$  variances du signal, estimées à partir des fenêtres glissantes respectives.  $N_1$  et  $N_2$  représentent les tailles respectives des fenêtres glissantes «  $n$  » et «  $n - 1$  ».

Le calcul des variances est équivalent à celui des expressions (4.16) et (4.17), en remplaçant les valeurs des moyennes après simplification, ce qui donne :

$$\sigma_1^2(n) = \frac{1}{N_1(N_1-1)} \left\{ (N_1) \sum_{i=0}^{N_1-1} \Delta y_{k-i}^2 - \left[ \sum_{i=0}^{N_1-1} \Delta y_{k-i} \right]^2 \right\} \quad (4.26)$$

$$\sigma_2^2(n) = \frac{1}{N_2(N_2-1)} \left\{ (N_2) \sum_{i=N_1}^{N_1+N_2-1} \Delta y_{k-i}^2 - \left[ \sum_{i=N_1}^{N_1+N_2-1} \Delta y_{k-i} \right]^2 \right\} \quad (4.27)$$

L'expression récursive de la variance, pour une fenêtre glissante de taille  $N$  et un pas de déplacement  $p$  est :

$$\sigma^2(n) = \sigma^2(n-1) + \frac{1}{N(N-1)} \left\{ N \sum_{i=0}^{p-1} (\Delta y_{k-i}^2 - \Delta y_{k-N-i}^2) - s(n)^2 + s(n-1)^2 \right\} \quad (4.28)$$

où  $s(n)$  s'exprime sous forme récursive par :

$$s(n) = s(n-1) + \left\{ \sum_{i=0}^{p-1} (\Delta y_{k-i} - \Delta y_{k-N-i}) \right\} \quad (4.29)$$

après simplification des sommes  $s(n)$  s'écrit :

$$s(n) = s(n-1) + [y_k - y_{k-p} - y_{k-N} + y_{k-N-p}] \quad (4.30)$$

Au même titre que précédemment nous obtenons les mêmes sommateurs  $a_k$  et  $b_k$  obtenus en (4.21) et (4.22).

De même, Afin de réduire le nombre d'opérations, lors du calcul de la fonction de comparaison de variances, nous prenons un pas de déplacement  $p$  des deux fenêtres égal à leurs tailles  $N_1$  et  $N_2$ , ce qui donne :

$$\sigma_2^2(n) = \sigma_1^2(n-1) \quad (4.31)$$



### III.3.2.4 Simulations et résultats

Les simulations qui suivent sur la base des trois tests détaillés précédemment, nous permettent de les évaluer et les comparer. A cet effet, nous avons choisi de simuler des séquences aléatoires gaussiennes, sur lesquelles sont simulés différents types de fronts de pollutions. Les signaux issus des tests statistiques associés au signal brut testé, sont disposés de manière superposée sur un même graphique, selon l'ordre suivant :

- le signal brut à tester,
- la fonction de comparaison de moyennes appliquée au signal brut,
- la fonction de comparaison de variances appliquée au signal brut,
- la fonction de comparaison de variances inverse appliquée au signal brut,
- et le test de Hinkley, utilisant parallèlement les deux détecteurs positif et négatif.

Tout au long des exemples qui suivent, la réactivité des tests statistiques dépend des décalages induits du pas des deux fenêtres glissantes.

#### 1 Cas d'un changement brutal, sur la base d'un front descendant :

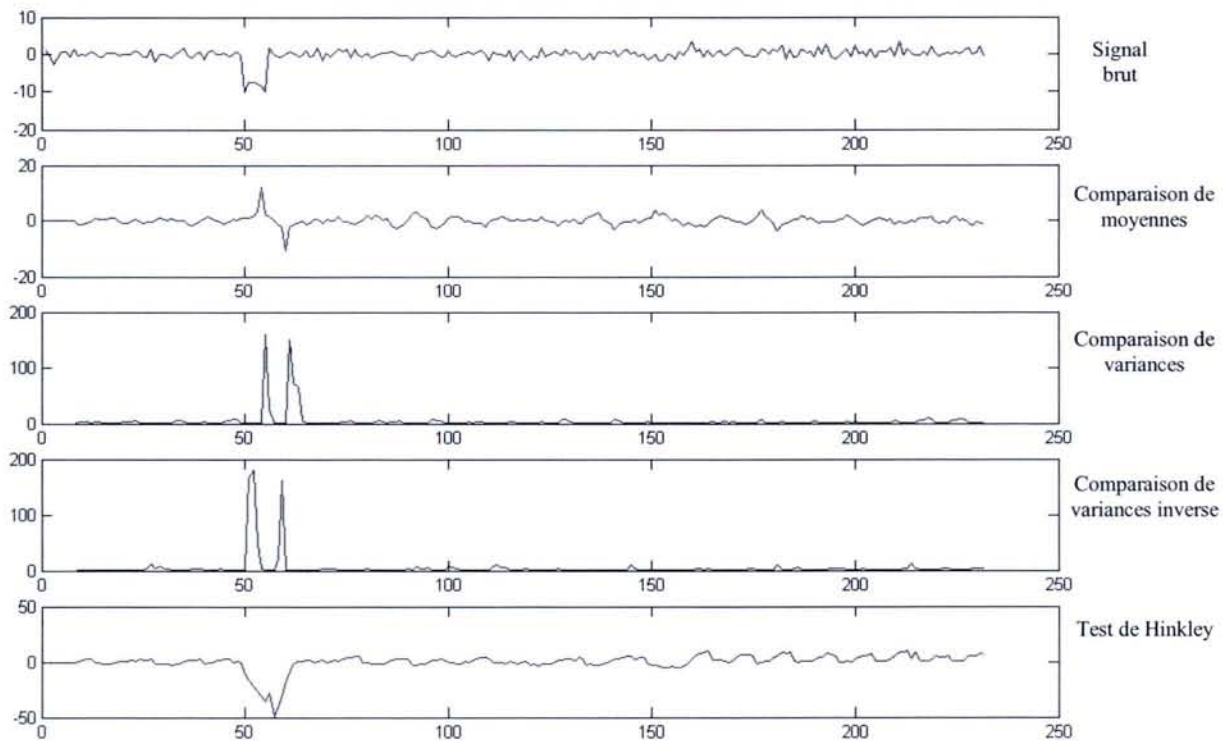


Figure IV.6 : Simulation d'un front descendant

Il ressort de ces graphiques, que le front descendant, est détecté par l'ensemble des tests statistiques, que ce soit à la chute du niveau du signal qu'à sa remontée, de différentes manières, en l'occurrence :

- sur le test de comparaison de moyennes, la chute du niveau du signal se manifeste par un pic positif, alors que la hausse par un pic négatif,
- les deux tests de comparaisons de variances, réagissent de manière équivalente, et détectent autant le saut négatif que le saut positif par des pics positifs,
- concernant le test des sommes cumulées d'Hinkley, les pics sont plus larges et accompagnent la tendance du signal.

## 2 Cas d'un changement brutal, sur la base d'un front montant :

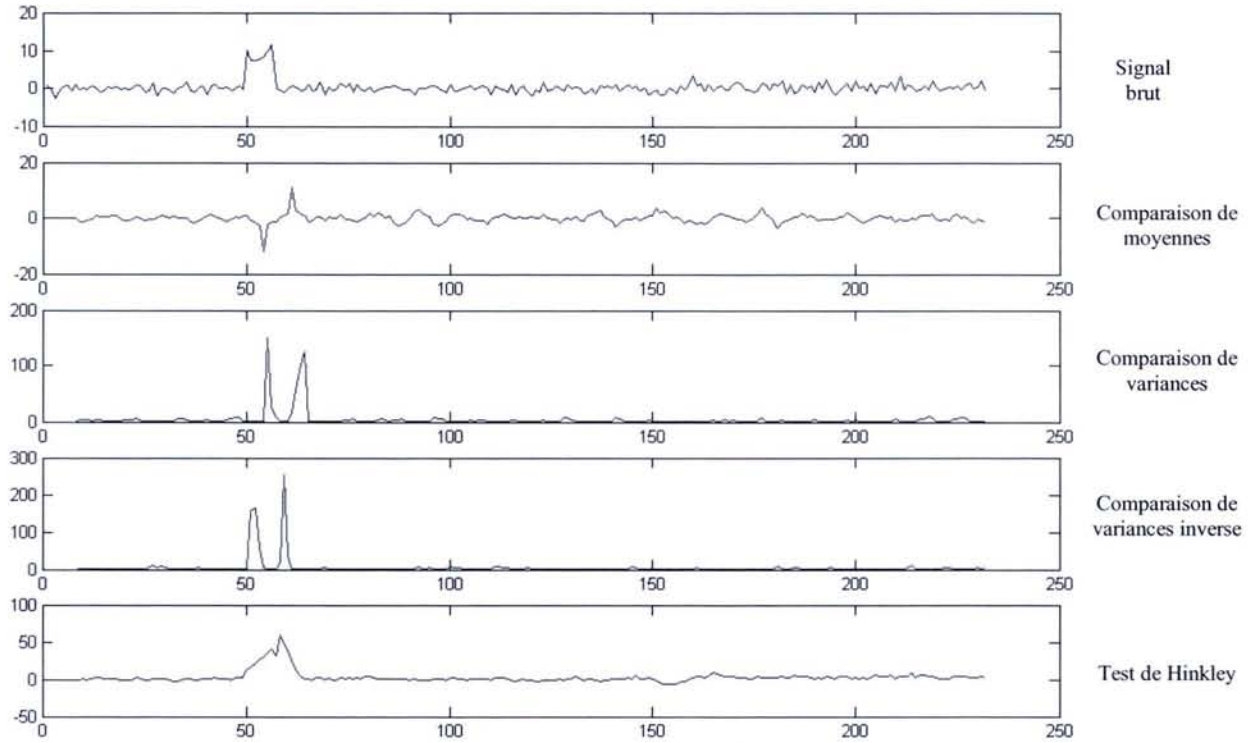


Figure IV.7 : Simulation d'un front montant

Les résultats de cette simulation confortent ceux de la précédente, en retrouvant le même profil de réaction respectivement décrit précédemment pour chaque type de test. Il est tout de même intéressant de vérifier la véracité de ces résultats dans la continuité. D'autre part, l'autre intérêt de ce front inverse, est de connaître le sens des variations.

## 3 Cas d'un changement brutal, sur la base d'une succession de deux fronts opposés :

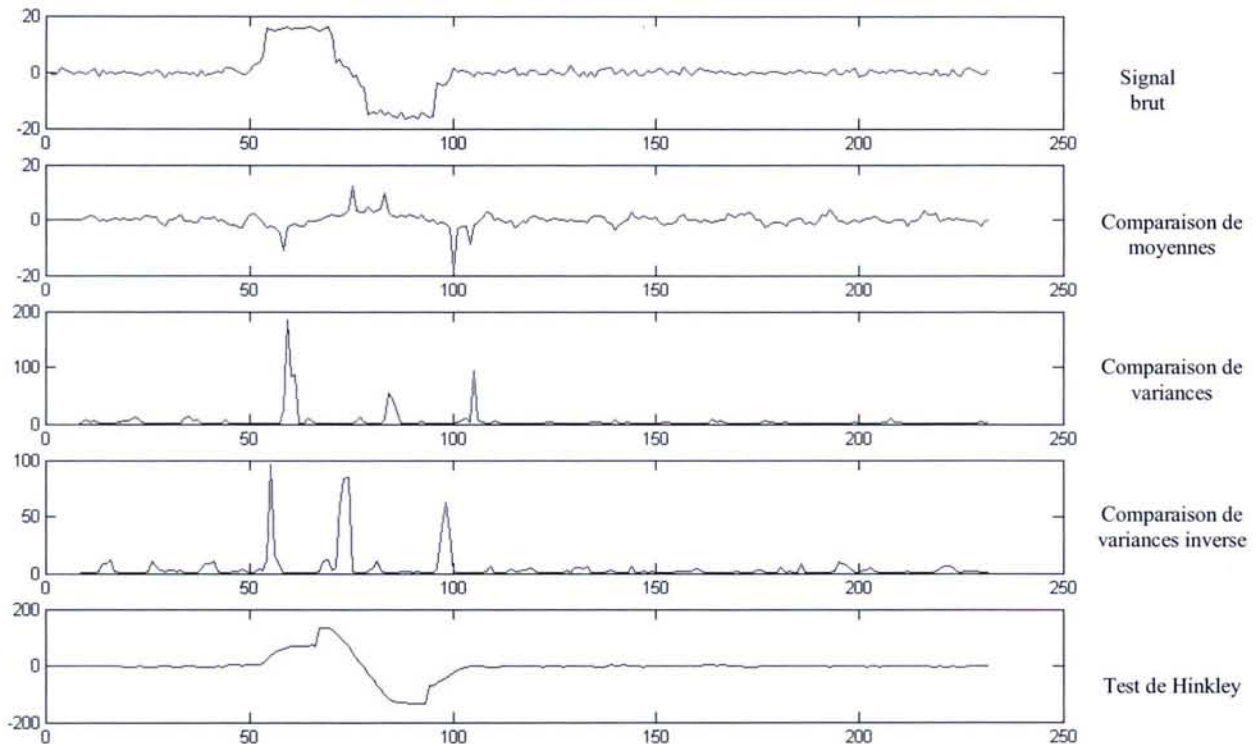


Figure IV.8 : Simulation de deux fronts opposés

Ce signal testé, se caractérise par deux fronts opposés de tailles équivalentes, et des sauts intermédiaires progressifs. La simulation de cette dernière spécificité, s'avère intéressante, du fait que certaines pollutions dans la nature se manifestent de manière graduelle.

Il en ressort sur le signal de la comparaison de moyennes, la réactivité du test sur le front montant, puis les deux chutes intermédiaires successives, pour enfin détecter le deuxième front montant. Une anomalie s'avère tout de même de taille, concernant une variation secondaire qui affecte l'importance de la détection des variations principales. Cela dénote, une sensibilité du détecteur à des variations mineures et donc sujet à des fausses alarmes. Concernant les autres tests, les tendances sont respectées.

Les tendances de réactions étudiées, il reste à établir les limites de la taille des fenêtres à exploiter. A cet effet, nous proposons dans ce qui suit, une étude comparative de trois longueurs de fenêtres, permettant de justifier le choix de la taille des fenêtres utilisées précédemment, qui est de quatre points de mesure.

4 Cas de petites fenêtres glissantes de deux points de mesure appliquée aux deux fronts opposés :

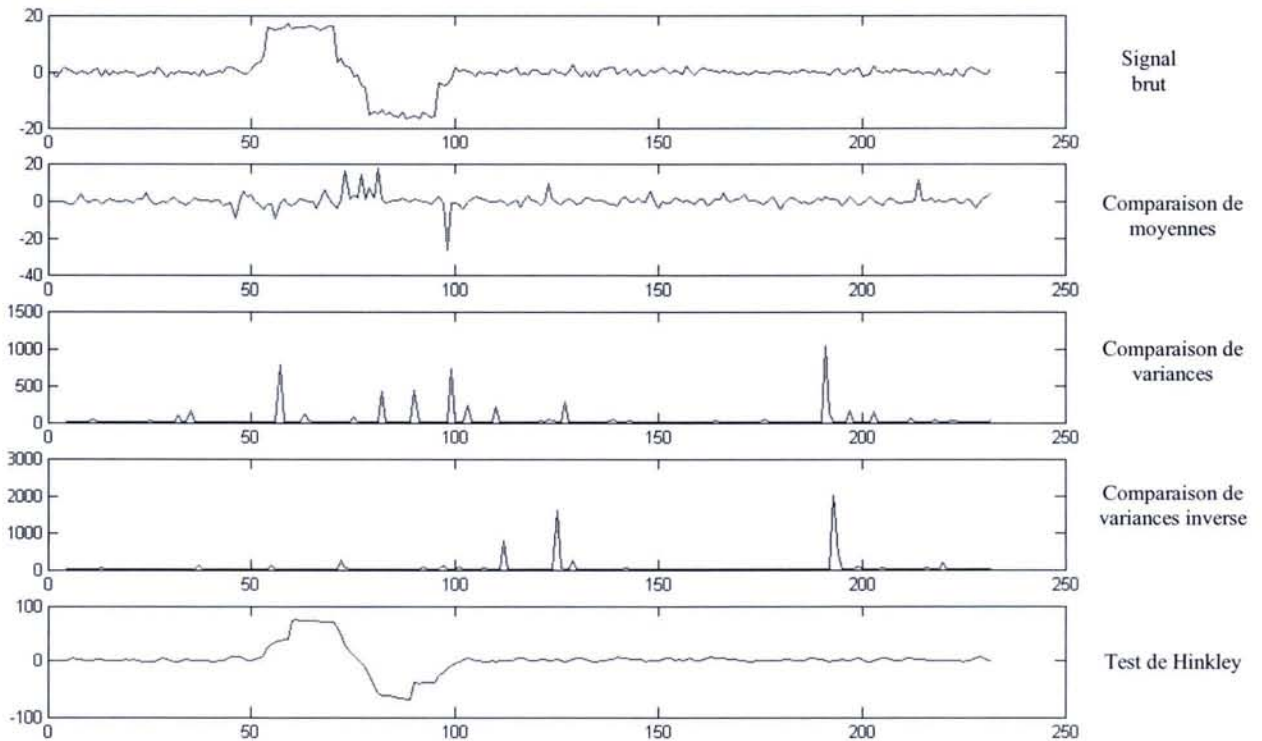


Figure IV.9 : Application de fenêtres à deux points de mesures

Les résultats de cette simulation, étaient prévisibles concernant le test de comparaison de moyenne et les deux tests de comparaison de variances. En effet, un raccourcissement des fenêtres à comparer, sur deux points de mesures, favorise une forte variabilité, conduisant à de fausses alarmes. Concernant, le signal des sommes cumulées de Hinkley, celui-ci épouse quasiment la forme du signal à tester, mais n'est pas assez fin, pour favoriser l'apparition d'une information, différente du signal à tester.

5 Cas de fenêtres glissantes de six points de mesure appliquée aux deux fronts opposés :

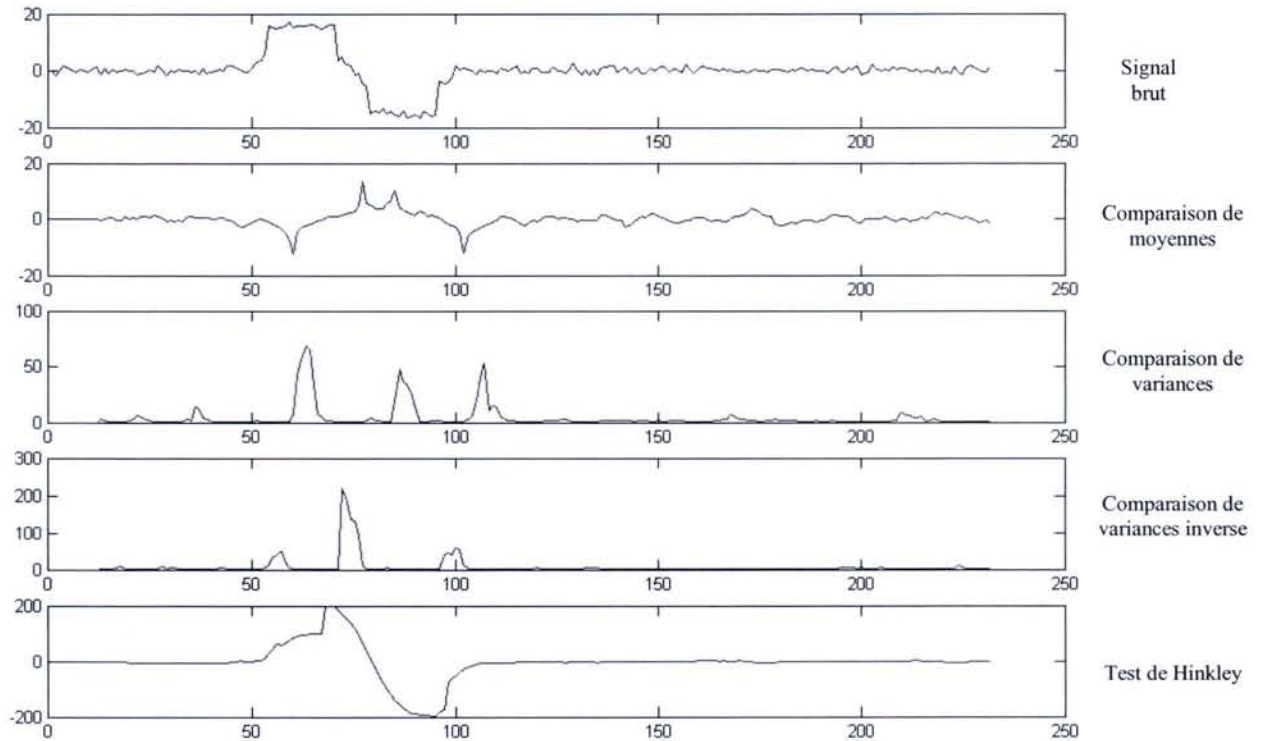


Figure IV.10 : Application de fenêtres à six points de mesures

Cette simulation, fait ressortir fidèlement l'ensemble des variations du signal brut. La particularité de cette simulation, se caractérise par un plus grand balayage des portions du signal, ce qui explique une accentuation du pic intermédiaire de la comparaison de variances inverse, qui englobe les deux chutes successives du niveau du signal brut.

6 Cas de fenêtres glissantes de dix points de mesure appliquée aux deux fronts opposés :

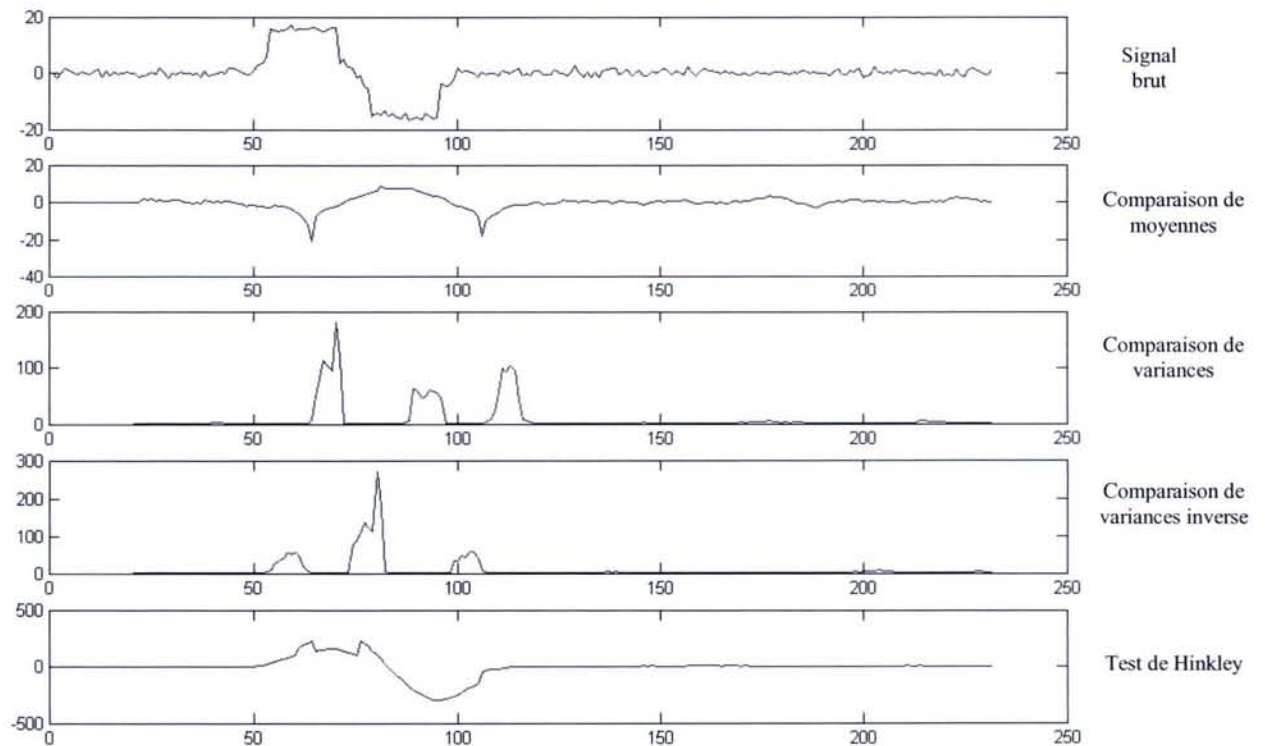


Figure IV.11 : Application de fenêtres à dix points de mesures

Sur cette dernière simulation, on constate une perte d'information sur le signal de la comparaison de moyennes, où les deux chutes successives ne sont plus apparentes. La largeur des pics s'accroît sur les signaux de comparaison de variance sans pour autant faire disparaître l'information pertinente, de même concernant le test de sommes cumulées de Hinkley.

Au terme de ce comparatif, il en convient que notre choix de fenêtre se porte sur une taille comprise entre quatre et six points de mesures, plage entre laquelle les pertes d'informations et les fausses alarmes n'ont pas été constatées.

Concernant les méthodes comparées lors de ces simulations, nous avons constaté que la première apparition de changement brutal, a toujours été détectée. Il en ressort tout de même quelques faiblesses sur la méthode de comparaison de moyennes qui a tendance à générer des fausses alarmes conséquentes. Concernant les autres tests statistiques, que sont les comparaisons de variances et les sommes cumulées de Hinkley, ils conviennent autant l'un que l'autre à notre problématique, à condition de déterminer les critères de seuillage. Dans notre cas, le choix s'est porté sur la méthode de comparaison de variances qui permet d'obtenir des pics distincts quantifiables pour notre problématique de détection de pré-alerte.

Cette quantification des pré-alertes, passe par une évaluation de la sensibilité de la méthode de détection choisie. En effet, nous avons simulé plusieurs niveaux de variations permettant d'évaluer le seuil minimal de sensibilité.

Les graphiques qui suivent, représentent le résultat de la fonction de comparaison de variance, appliquée sur un signal sur lequel ont été respectivement simulé, plusieurs pourcentages de variations (5%, 10%, 15%, 20%, 50%) entre le 54<sup>ème</sup> et le 60<sup>ème</sup> échantillon du signal, en ayant sur le premier graphique comme référence, le résultat de la fonction de comparaison de variances appliquée au signal initial.

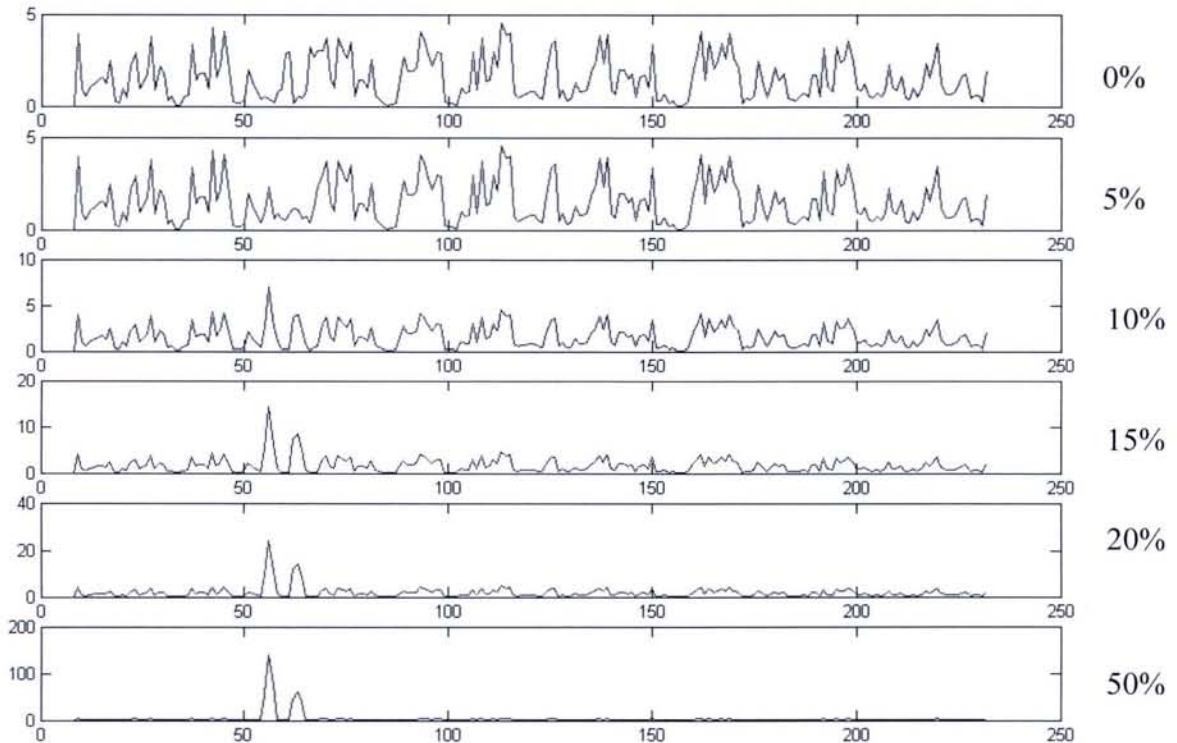


Figure IV.12 : Application de fenêtres à dix points de mesures

Au vu de ces graphiques, il apparaît clairement, que la variation de 5% n'a aucune incidence sur le signal de la fonction de comparaison de variances. Cela s'explique par le fait

que la variabilité du signal oscille au moins de 5%, ce qui rend toute éventuelle détection impossible à ce niveau.

A 10%, nous constatons la détection du début de la variation brutale, et à 15 % de manière plus distincte, la détection du début de la variation ainsi que celui de la fin de variation. Enfin la tendance se confirme pour être plus accentuée à 20% et de manière plus évidente à 50%.

Il en ressort que la fonction de comparaison de variances, réagit de manière appréciable, avec une sensibilité approximative de 10 %, ce qui est satisfaisant dans le cadre de notre étude, du fait que les seuils de tolérance sont au-delà des 10%, ce qui nous conforte considérablement dans notre choix de méthode de détection de variations brutales. **En effet, statistiquement, le rapport des deux variances permet d'éliminer relativement les bruits du signal brut, et d'avoir un calcul basé sur la pure variation du signal**, ce qui permet d'obtenir une finesse de détection appréciable.

Nous avons d'autre part conforté ces résultats, sur les signaux obtenus lors des tests de toxicité, en comparant la méthode de comparaison variances à celle de comparaison de moyennes. Il est bien établi, que les variations brutales ont été bien identifiées sur notre fonction de comparaison de variances, ce qui n'était pas systématiquement le cas sur la fonction de comparaison de moyennes.

Au vu de l'ensemble de ces résultats, cette fonction de comparaison de variances met en évidence, suffisamment, la variation anormale d'un signal brut. C'est à ce titre que nous l'adoptons comme méthode d'application à notre dispositif de détection.

D'autre part, à l'issue des ces simulations, nous constatons que les variations sont relatives à l'échelle sur laquelle doivent être constatées celles-ci. Cela signifie que pour chaque signal étudié, il faut adapter une quantification du niveau seuillage. L'avantage de notre démarche fusion de données, associée aux tests de toxicité, permet d'établir les échelles de variations de chaque pollution donnée et pour chacune des grandeurs ayant réagi. D'autre part, la prise en compte d'un ensemble de variations, limite le risque de fausses alertes.

Afin d'être plus explicite et d'étayer les résultats de la fonction de comparaison de variances sur un cas réel de pollution, nous commentons ci-après les résultats d'une pollution réelle.

### **III.3.2.5 Confrontation à une situation de pollution réelle**

Nous avons choisi d'illustrer la méthode mise en place à partir de l'incident qui a eu lieu le 22 juin 2002 [AZZ(b) – 2002]. En effet, une pollution réelle a eu lieu au cours de cette journée qualifiée d'historique par les services de la CUGN. Cet événement a été principalement détecté par l'analyseur d'ammoniaque de la CUGN qui a enregistré des valeurs allant jusqu'à 2g/l de  $\text{NH}_4^+$ , ce qui représente une forte pollution de la Moselle.

En parallèle, cette pollution a été aussi détectée par certains pilotes de la chaîne biologique et physico-chimique de notre station, notamment le Fluotox comme illustré sur les graphiques de la figure IV.13 et IV.14.

Il faut savoir que les mesures débutent à minuit et s'effectuent toutes les 5 minutes, aboutissant à 288 points par jour. On voit nettement sur le graphique du relevé du signal brut, une première variation à 6h30 négligeable, puis la distinction de deux fronts prononcés de pollutions à 11h30 et 21h30 du fait que le niveau de signal du Fluotox en temps normal n'est pas aussi élevé.

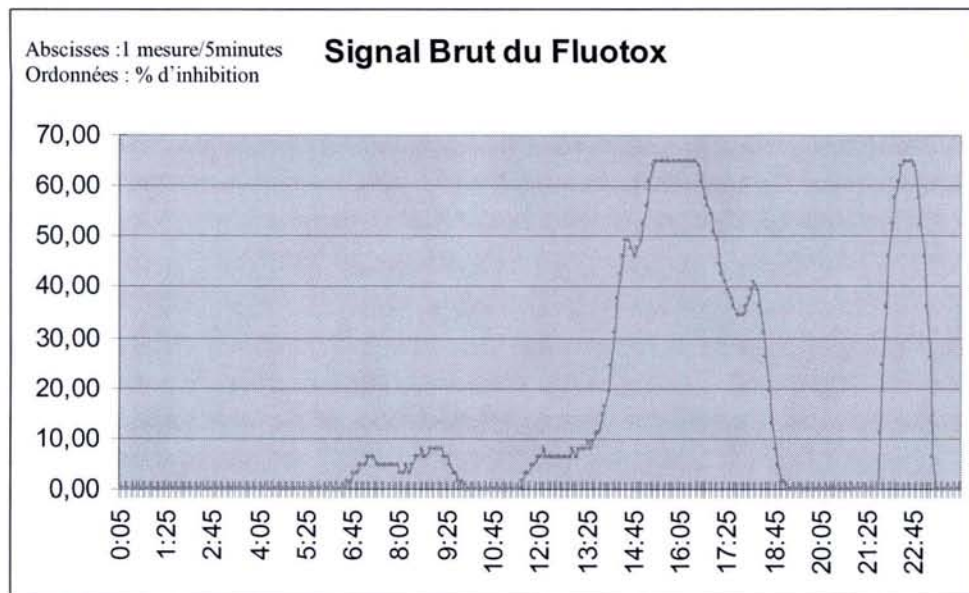


Figure IV.13 : Signal brut du Fluotox au cours de la journée du 22 juin

Concernant le premier front de pollution, les variations brutales s'opèrent entre 14h15 et 15h00, ce qui est pris en compte par la fonction de comparaison de variance avec le décalage attribué à la fenêtre glissante. Il en est de même pour le deuxième front de pollution aux alentours de 21h25. Cette constatation conforte le but recherché, consistant à la production de pré-alerte avant atteinte de seuil maximal, sur la base de variation brutale.

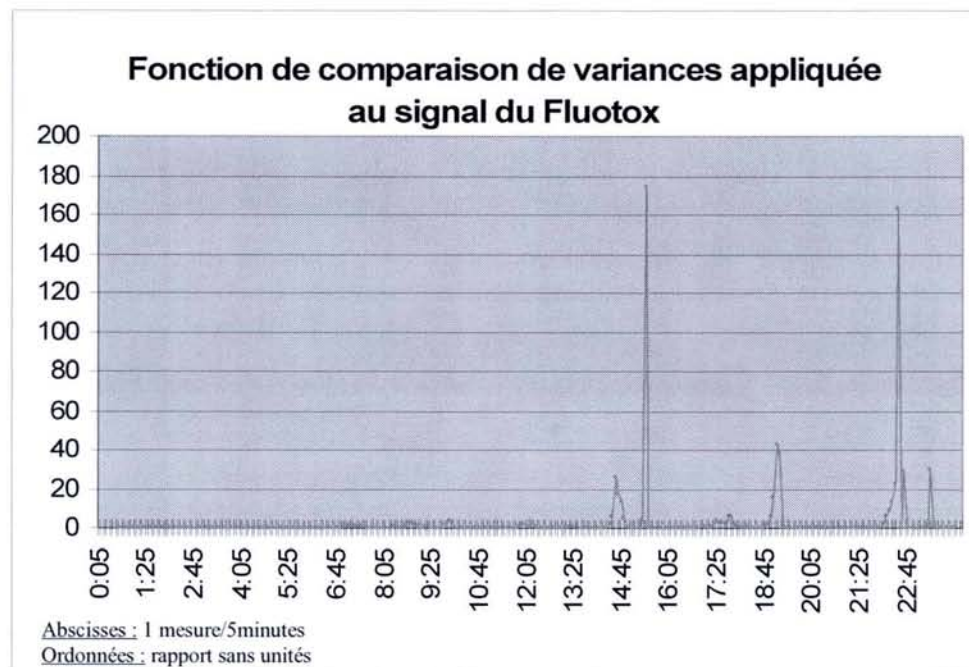


Figure IV.14 : Fonction de comparaison de variance appliquée sur le signal du Fluotox

Or, au vu du long décalage temporel de la fenêtre glissante, l'alerte est produite après le passage d'une bonne partie du front de pollution, ce qui remet en cause l'efficacité du système de détection. En effet, la résolution de mesure a été de 5 minutes pour la plus part des capteurs, ce qui correspond à des fenêtres glissantes à pas de vingt minutes pour quatre points de mesure. Il en ressort que pour une exploitation réelle du dispositif, cela demeure insuffisant en terme de précision, et considérablement long pour le signalement d'une alerte. Cela

conduit à optimiser la résolution de mesure, rendant notre dispositif de détection réaliste pour une résolution de mesure d'au moins 1 minute. Cela permet ainsi d'avoir en même temps des fenêtres plus réduites (moins de retard sur le résultat) et plus de points de calculs (plus de précision). Cela diminue considérablement la taille temporelle des fenêtres, et offre la possibilité de détecter des événements beaucoup plus courts dans des délais plus brefs.

Il faut savoir d'autre part, qu'en confrontant un événement aussi toxique que celui-ci, à l'échelle normalisée des tests de toxicités correspondants, les niveaux sur les fonctions de comparaison de variance, permettent de qualifier la teneur du niveau de toxicité. A titre indicatif, sur la campagne des tests de toxicité, l'ammoniaque a été testée sur des concentrations de l'ordre du « milligramme » (10mg/l) correspondant aux normes tolérées décrites en « Annexe III » alors que la quantité détectée était de l'ordre du « gramme » (2g/l). En effet, lors de ce test de toxicité, le pic de détection de la fonction de comparaison de variance appliquée au Fluotox est atteint dès les premiers instants de la pollution réelle, ce qui montre que le Fluotox n'a pas autant réagi que lors de la pollution réelle. Cela dénote bien de la haute toxicité de la pollution réelle.

Sur un aspect biologique, il faut associer à cet événement la diminution de la concentration en chlorophylle débutant aux alentours de 12h00. Elle s'explique par la forte dose de  $\text{NH}_4^+$  présent dans le milieu, amenant à déduire que l'ensemble de la flore et la faune a été intoxiquée par cette pollution.

Le relevé de mesure recoupe bien le fait que la photosynthèse de l'algue (Bio-détecteur du Fluotox) ait été perturbée dans son processus chimique, ce qui provient du fait que l'énergie lumineuse captée par les antennes chlorophylliennes ne participe plus à la chimie de la plante, conduisant l'énergie solaire captée à se transformer en chaleur et donc en rayonnement de fluorescence.

En distinguant bien les fronts de pollutions tout au long de cette journée particulière du 22 juin 2002, cela laisserait supposer qu'il y a eu des déversements volontaires. La nature de cette pollution en fonction des résultats de notre station, fait penser à une pollution de type agricole (excès d'engrais et /ou herbicides).

Le travail d'investigation mené par la CUGN, a permis d'identifier la distance de la pollution par rapport au lieu de pompage à hauteur de 5 km, ceci ayant été calculé sur la base du signal relevé associé au débit de la Moselle.

Les autres capteurs biologiques, en l'occurrence, l'Aquatox et le Dresseina ont respectivement réagi, mais de manière moins pertinente ne correspondant pas nécessairement aux périodes de pollutions. Ceci dénote les propriétés atypiques de ces capteurs, dont les signaux bruts sont mentionnés en « Annexe X ». Nous tenons à déplorer l'absence des données du Gymnotox qui était en état de maintenance. En effet, de part les caractéristiques de détection multirisques de ce capteur, il aurait sûrement réagi pertinemment.

Concernant le Bacteria-Test, dont les mesures s'effectuent toutes les 30 minutes, ne nous permet pas de constater une variation brutale, qui laisserait augurer d'une variation significative. La comparaison de variances, n'as pas d'intérêt dans ce contexte du fait que la durée de la fenêtre est inférieure à la résolution de mesure, ce qui obligerai d'allonger la durée de la fenêtre, ne correspondant pas aux objectifs temporels. Nous en déduisons que le Bactéria-Test, ne peut s'inscrire dans les objectifs temps réel du dispositif de pré-alerte.

Concernant les capteurs physico-chimiques (voir l'Annexe X), il n'y a pas eu de variations notables pouvant justifier l'association de la pollution à la variation du signal du capteur. Nous tenons à préciser, que concernant le pH la variabilité est mitigée et n'a rien à



voir avec la pollution au  $\text{NH}_4^+$ , mais peut s'expliquer par une dérive ponctuelle du capteur ou d'une modification naturelle du pH sans que cela corresponde nécessairement à une pollution.

Ce type de constatations, nous amène à montrer l'importance de la notion de croisement de données détaillée ci-après sous forme de fusion de données. En effet, dans le cas précis de cette pollution, nous avons pu établir les horaires de la pollution, sa teneur et son lieu, à travers les capteurs que sont le Fluotox et l'analyseur ammonium. Cela nous a permis d'éviter des interprétations erronées qui auraient pu être produites par des capteurs défectueux. C'est dans ce contexte, que nous définissons la notion de masques de pollutions qui fait intervenir plusieurs capteurs à la fois, basés sur les observations des tests de pollutions.

### **III.3.3 Etablissement de masques de pollutions sur la base de pré-alertes**

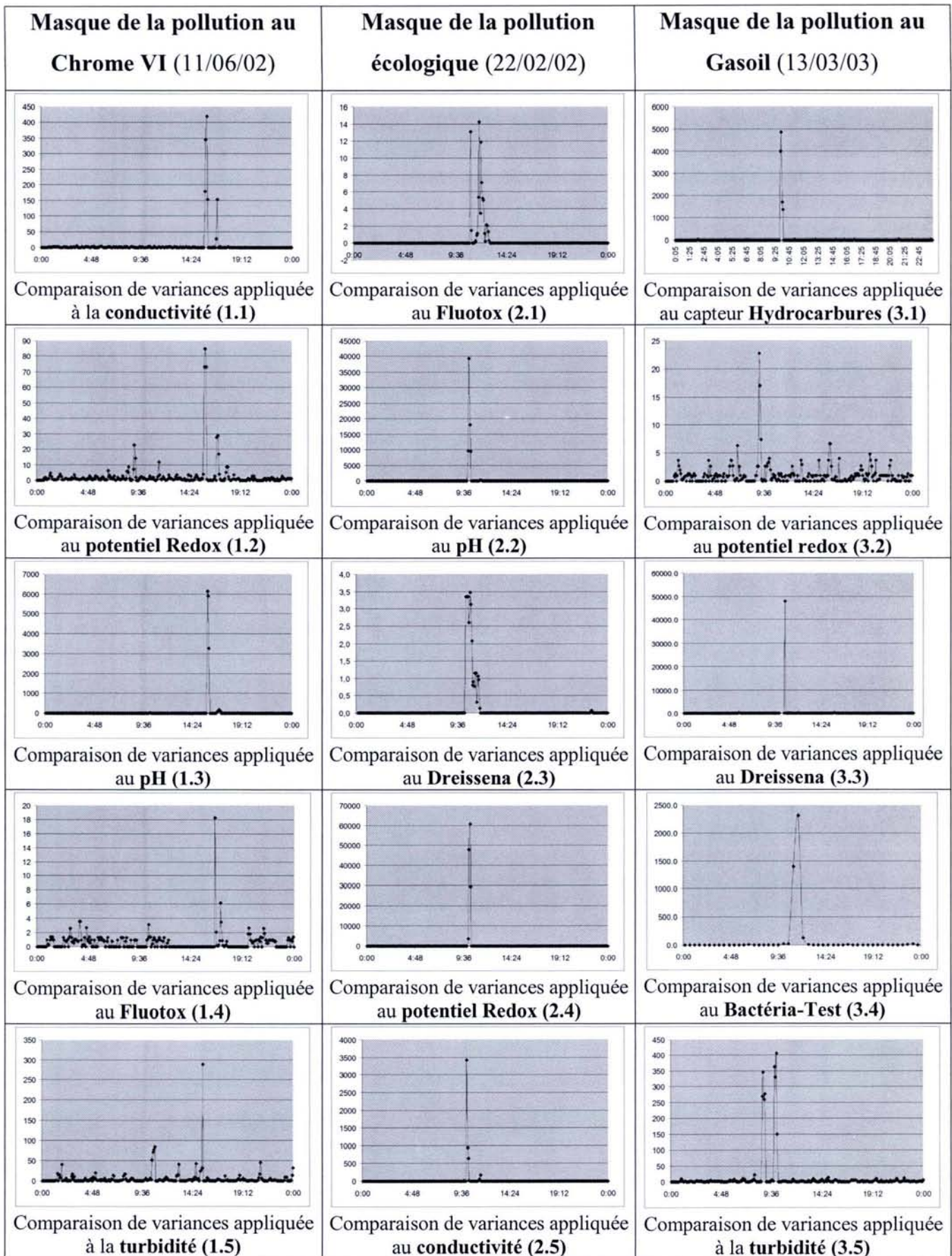
Cette phase correspond à l'exploitation des tests de pollution. Elle repose sur une classification des différents comportements des capteurs vis à vis de chaque pollution. Le but est d'identifier des pollutions, pour lesquelles il n'y a pas de capteurs spécifiques, d'où l'intérêt d'exploiter une information mutualisée issue de l'apport de chacun des capteurs. Cette approche nous permet de créer des classes de réactions pour une pollution déterminée. L'information complémentaire pertinente à extraire consiste à établir un seuillage intermédiaire caractéristique des pré-alertes, ce qui permet d'anticiper sur l'apparition de pollutions avant l'atteinte des seuils maximaux.

Afin d'avoir une idée préliminaire de notre approche, le tableau de réaction des capteurs face aux tests de toxicité en « Annexe VII », donne une idée générale sur la base du signal brut, de la réactivité des capteurs face aux injections de polluants, mais s'avère insuffisante pour affirmer la variation anormale de mesures sur des valeurs brutes.

Notre principe consiste à considérer que l'échelle de chaque capteur est régie par un seuil minimum et un seuil maximum, nous pouvons exprimer les valeurs en pourcentage après normalisation. Ces conditions permettent d'identifier des états de pré-alertes, contribuant à constituer des profils de pollutions que l'on dénomme « masques de pollutions ».

Les seuillages établis à ce niveau de l'étude, affinent les seuils de sensibilités des valeurs brutes en y détectant les prémices de déclenchement d'une pollution avant l'atteinte du seuil maximal sur la valeur brute, d'où l'introduction de la notion de pré-alerte. Le but recherché étant d'appréhender le début d'une variation brutale par rapport à une évolution dite « Normale ». L'application de ces seuils se base sur le principe de calcul d'indices de comparaisons statistiques de variances. Ces indices permettent de déterminer l'instant où apparaît le changement dans la série temporelle et surtout de quantifier l'intensité de la variation.

Il est en effet plus précis de modéliser la réaction des capteurs à partir de la méthode de comparaisons de variances. Le principe consiste à relever pour chaque test de pollution, le pic de la fonction de comparaison de variance de chaque capteur ayant réagi. Ce pic correspond au niveau maximum en cas d'atteinte des niveaux de tolérance fixés par la législation. L'établissement des échelles de pré-alertes est spécifique à chaque capteur. Elles sont bornées et normalisées sur la base de ce pic maximal qui relègue en second lieu l'importance d'autres pics secondaires, sans qu'ils soient pour autant négligés. Chaque groupe de capteurs ayant réagi à la pollution, forme alors un masque de pollution. Une fois ces niveaux relevés, ils servent de base pour l'initialisation du profil de pollution concerné par la réaction des capteurs. Afin d'illustrer notre méthode, ci-après un tableau représentant trois masques établis sur la base des tests de toxicités.



« Représentation graphique de la comparaison de variances sur trois profils de pollution »

Sur chaque colonne de ce tableau, sont représentées, les comparaisons de variances des signaux ayant mis en évidence une variation brutale du signal par rapport à des variations dites « normales ». Sur la base des graphiques précédents quelques commentaires s'imposent :

- le pic maximal de chaque graphique correspond à l'apparition de la pollution,
- les pics secondaires apparaissant dans les graphes (1.1, 1.2 et 1,4) sont des indicateurs de baisse de la pollution,
- les doubles pics dans les graphes ( 2.1 et 3.5) correspondent à deux variations brutales successives de la pollution.

L'échelle de pré-alerte à partir de laquelle est évalué chaque paramètre, est basé sur une comparaison par rapport au seuil maximal ainsi déterminé. Afin d'illustrer plus explicitement l'établissement des masques de pollutions, par la méthode des comparaisons de variances, sur la figure IV.15 est présenté le masque de pré-alerte à la pollution au **Chrome VI**, où est fixé le seuil maximal de réaction. Il faut rappeler que l'obtention de ces valeurs maximales est faite sur la base des seuils de tolérance à la toxicité, et peuvent être malheureusement largement dépassées lors de pollutions hors normes.

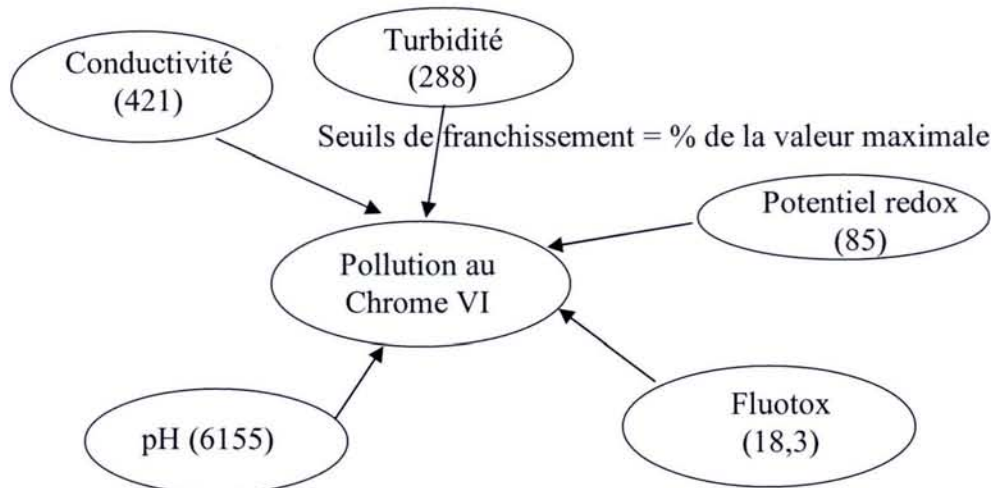


Figure IV.15 : Exemple de masque de pollution

Les seuils de franchissement correspondent à un niveau sur l'échelle de la fonction de comparaison de variances de chaque capteur. Ce niveau est relatif à l'atteinte du niveau maximal déterminé lors des tests de toxicité (valeur représentée entre parenthèses). Le niveau intermédiaire étant fixé à 50%, il peut être abaissé ou rehaussé en fonction de la sensibilité recherchée pour déclencher la pré-alerte d'un type de pollution. Cette appréciation passe par une exploitation du dispositif dans des conditions réelles, permettant d'obtenir la finesse de détection voulue.

Complémentairement à l'exploitation qui permet d'avoir l'expertise nécessaire pour le paramétrage, une autre condition d'efficacité du dispositif de pré-alerte passe par un enrichissement de la base de connaissance, en effectuant d'autres tests de toxicités soit pour conforter certains tests ou réaliser de nouveaux autres types de tests non pris en compte.

En effet, au-delà des pollutions identifiées, nous associons à la panoplie de masques, la pollution *boîte noire*, qui a de fortes chances de se manifester de part le fait que l'on n'a pas nécessairement exploré toutes les sortes de pollutions. Cette pollution *boîte noire* que nous

notons "X", comme mentionné en figure IV.16, sert de base d'entrée pour le rajout d'autres profils de pollution dans notre base d'événements. Sa détermination se base sur les indicateurs de seuils préétablis sur un groupe de capteurs ayant réagi et ne correspondant à aucun des profils de pollutions connues. En d'autres termes cela représente l'ensemble des capteurs ayant eu une réaction significative à un instant donné et dont le profil n'a pas été recensé dans notre base d'événements issue des tests de pollution.

La procédure d'enregistrement de cette pollution X, passe par l'analyse du prélèvement d'eau effectué au moment de la détection de cet évènement, permettant ainsi de qualifier la pollution pour l'associer au profil de capteurs ayant permis de la détecter.

En effet, en dehors des cas déjà expérimentés, la mise en place de détection de pollution « boîte noire », se déclenche sur le critère de la variation d'au moins deux paramètres à la fois. Cela permet d'enrichir la base de données en terme d'incidents ou de pollutions. Suite aux analyses (selon la pertinence de l'évènement), **basées sur un prélèvement automatique d'échantillon d'eau**, il devra être possible d'affecter à cette boîte noire un type d'évènement qui peut être soit une pollution soit un phénomène naturel dû au climat.

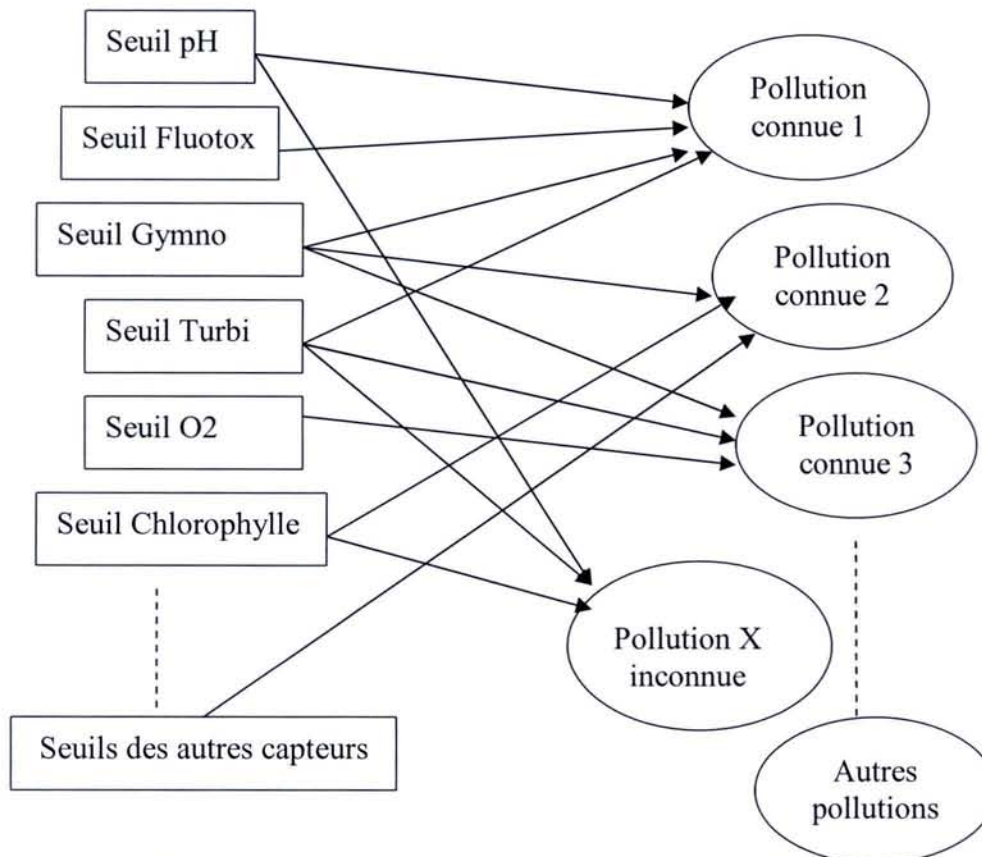


Figure IV.16 : Identification d'un masque de pollution inconnue

Dans le cas où l'analyse serait infructueuse, cela dénotera d'une erreur de mesure ou de transmission. Au-delà de ces considérations, il est utile de préciser qu'en cas de défauts avérés des capteurs ou du système de communication, des alarmes permettent d'annuler la procédure d'enregistrement.

Par ailleurs, il est important de noter que notre étude a aussi pris en compte certains événements naturels pouvant être assimilés à des pollutions. En effet la hausse de la chlorophylle, lors de fortes périodes d'ensoleillement, cause une eutrophisation de l'eau qui

est en l'occurrence causée par le développement d'algues. Cela nécessite un traitement spécifique de l'eau, qui doit être associé au déclenchement d'une pré-alerte, correspondant à un masque à mettre en place. Nous pouvons aussi citer le cas des pluies entraînant une hausse de la turbidité des eaux, ce qui est également le cas de la fonte des neiges associée à la période d'épandage.

Tous ces types de phénomènes naturels doivent être pris en considérations avant toute interprétation hâtive d'une pollution accidentelle, ce type d'informations devant être un pré-requis pour les opérateurs décisionnels, en intégrant ce type d'événements à notre base de connaissance de masques de pollutions.

Un autre type d'information contribuant à rendre qualitative notre base de connaissances, est la détection d'un front descendant de pollution, comme constaté précédemment. En effet, nous avons pu nous en rendre compte lors des simulations de pollutions pour certains paramètres, en l'occurrence, la détection du début de pollution coïncidant avec le front montant de l'injection du polluant engendre par la suite la détection de la fin de la pollution lors du front descendant de fin d'injection du polluant. Cette seconde détection doit donc être considérée comme un retour à la normale, devant être pris en compte par l'opérateur de surveillance.

En figure IV.17, est donné un exemple illustratif du phénomène sur le signal de l'Aquatox, lors de la simulation de pollution agricole du 15 mai 2002, et pour lequel nous avons appliqué la fonction de comparaison de variances :

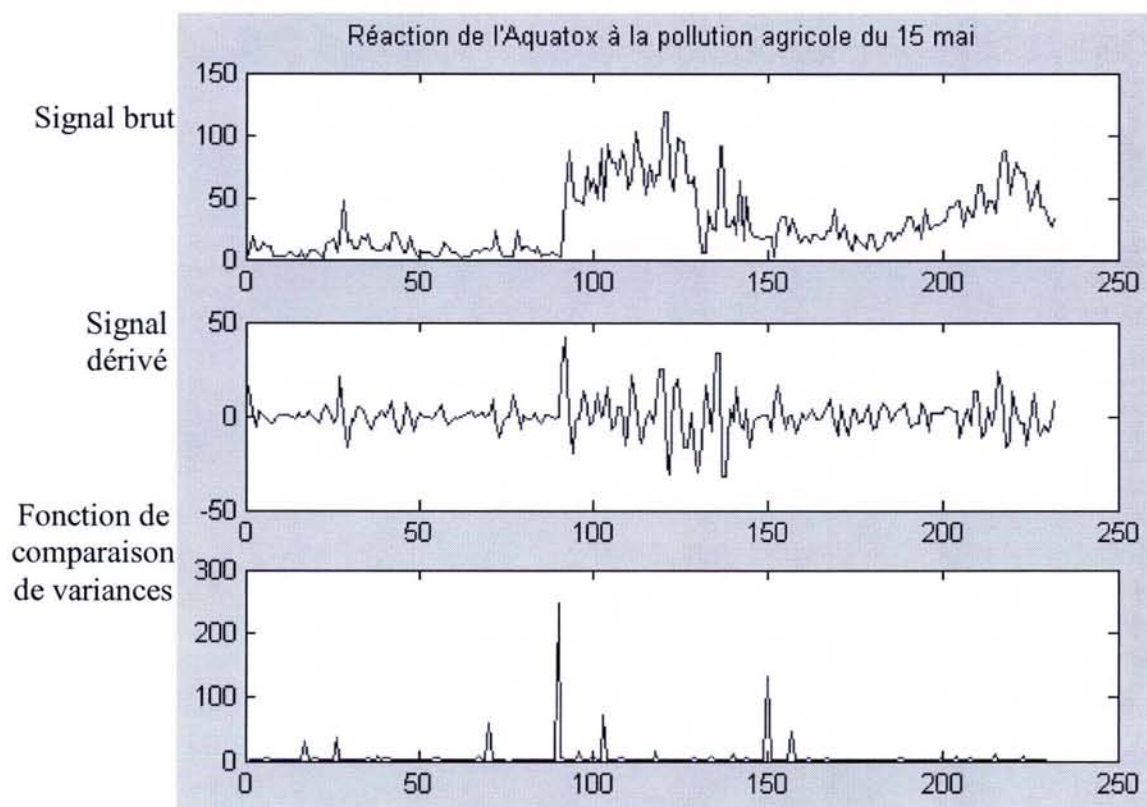


Figure IV.17 : Mise en évidence du front montant et descendant d'une pollution

En effet, on discerne bien la mise en évidence du front montant ainsi que le front descendant. Le pic de détection sur le front descendant n'est pas négligeable et constitue un événement autant inhabituel que celui du début d'une pollution. Il doit, à cet effet, être considéré comme un événement utile à prendre en compte, lors d'interprétations postérieures.

### III.4 Conclusion

Au-delà de la fonction préventive, les résultats du système d'alerte permettent d'enquêter sur la nature et la source de la pollution après identification. En effet, la chasse aux pollueurs ou aux pollutions intentionnelles, et dont il faut tarir la source, passe par une procédure de recherche et d'investigation. En effet lors de la pollution du 22 juin 2002, à partir de la courbe d'évolution de la pollution associée au débit de la rivière, il a été possible d'identifier le lieu de la pollution. Sa nature a été qualifiée « d'agricole volontaire » du fait qu'il n'y avait pas de pluies pouvant provoquer l'érosion des terres, ce qui peut être par contre le cas en hiver lors de pollutions aux métaux. Dans le cas des pollutions agricoles, nous avons un effet indirect dû à l'épandage d'engrais ou insecticides, d'où une vigilance accrue de discernement dans le cadre d'une analyse *à posteriori* de l'événement. Celle-ci doit tenir compte de l'éventualité de la cause naturelle.

De ces constatations, il est intéressant de s'interroger sur la pertinence d'analyseur coûteux tel que l'analyseur ammoniacal et de le remplacer par un bio-capteur moins complexe et coûteux qu'est le Fluotox, dont le spectre de détection ne se limite pas à un unique produit.

Il est à cet effet, intéressant à ce niveau de redéfinir ce que nous entendons par capteurs et analyseurs physico-chimiques : un **capteur** est de fonctionnement plus simple puisqu'il est de type sonde et effectue des mesures en ligne (par exemple capteur à pH, température, etc...), alors qu'un **analyseur** nécessite un prélèvement, parfois un pré-traitement de l'eau brute (analyseur de métaux lourds, d'hydrocarbures...) et met en jeu une réaction physico-chimique plus ou moins complexe et ciblée sur un produit spécifique. Il serait donc hasardeux de s'aventurer à définir quel bio-capteur remplacerait quel(s) analyseurs. Ces bio-capteurs réagissent face à des pollutions que les analyseurs détectent, mais également face à d'autres pollutions. **L'information donnée par un bio-capteur est plus riche que celle fournie par un analyseur.** Bien sûr nous ne connaissons plus la teneur exacte de certains composants, par exemple en ammonium à tout moment, est-ce cela l'important ? Non. L'important est que notre système qualifie la pollution pour une prise de décision adéquate.

S'affranchir totalement de l'utilité d'un analyseur en ligne, n'est donc pas impossible. En effet, l'objectif de l'exploitation de la station à moyen terme, est d'exploiter la précocité des informations fournies par les capteurs biologiques associée à la précision des capteurs physico-chimiques classiques, pour pallier aux analyseurs en ligne. Cela nous dispenserait d'associer à chaque risque un analyseur spécifique, alors que notre système est exploitable dans un cadre beaucoup plus global. L'atteinte de cet objectif, passe par un enrichissement de la base de connaissance lors d'événements inédits à analyser pour une qualification (procédure détaillée au point suivant). Il s'agira aussi à l'exploitant d'établir d'autres tests de toxicités pour conforter les hypothèses sur les nouveaux événements recensés, cette veille devant s'inscrire dans la durée.

En terme de recueil de données, l'établissement des critères d'alertes n'aurait pas été possible sans avoir réuni une base de données quantitative et qualitative. Cette base de données issue de sources multiples, dans un environnement multi-capteurs doit tenir compte :

- de la synchronisation de la période de rapatriement des données au niveau de chaque capteur,

- de la réduction de la résolution de rapatriement, permettant de gagner en précision de calcul et l'obtention rapide de résultats issus des fonctions de comparaisons de variances,
- des dérives et pannes de capteurs,
- de la maintenance périodique des capteurs nécessitant un arrêt de leur fonctionnement,

Cette étape de validation a été une condition *sine qua none* pour la suite du traitement de données, exigeant une continuité sur la série de données à traiter avec le maximum de capteurs en fonctionnement. En effet, certains dysfonctionnements ponctuels ou maintenances périodiques, posaient des difficultés à trouver des séries de données assez longues pour pouvoir être exploitées. En effet, afin de minimiser ces périodes inexploitable, lorsque la maintenance ou le dysfonctionnement était court, nous avons procédé à des méthodes de remplacement des valeurs manquantes par « moyennages » des échantillons précédents, ce qui n'était pas acceptable sur de longues coupures pour remplacer de longues séries. Ces désagréments ont été des éléments déterminants, dans le choix final des capteurs à maintenir au sein de la station, comme précisé au prochain chapitre.

Un autre aspect important à préciser, est l'exploitation des masques de pollutions qui sont fondés sur une base de réactivité des capteurs durant les tests de pollutions. Ces tests ont certes été suffisants pour la mise en place de masques de réactivité, mais pas assez pour l'affinement des seuils. Ces masques constituant une base de référence pour l'exploitation réelle du système, doivent être absolument évolutifs autant en terme de seuillage que de rajouts d'autres paramètres dans la configuration.

S'ajoute à cela, le rôle du masque « X » servant à lancer une pré-alerte à une pollution non identifiée et pour laquelle après analyse, on puisse donner un nom au masque nouvellement identifié. Cette pollution « X » met en évidence, l'intérêt d'avoir une base de connaissance des pollutions à partir de laquelle il est possible de construire l'identification d'autres types de pollutions. Cela est possible grâce aux seuils déjà préétablis des capteurs mais sur une base de combinaison différente produisant ainsi, le masque d'un nouveau type de pollution.

## **Bibliographie**

- [APPE – 1983] Appel U., Brandt A.V., *Adaptative sequential segmentation of piecewise time series*, in Inf. Sci., 29, pages 27-56, 1983.
- [AZZ(b) - 2002] Azzam A.A., “Traitement de données d’une station de protection et surveillance multi-paramètres des ressources aquatiques”, in : POLLUTEC , Lyon, Novembre 2002
- [BASS – 1983] Basseville M., Benviste A., *Sequential detectyion of abrupt changes in spectral characteristics of digital signals*, in IEEE Trans. On Inform. Theory; 29, pages 709-724, 1983.
- [C.E.A - 1978] Statistique appliquée à l’exploitation des mesures. Commissariat à l’énergie atomique. Editions Masson 1978.
- [BLOCH-1998] Bloch I., Vidal-Madjar D.: *Introduction of neighborhood information in evidence theory and application to data fusion of radar and optical images with partial cloud cover*, In Pattern Recognition Volume 31, Number 11, November 1998, pages 1811-1823.
- [BODE – 1977] Bodenstein G., Praetorius H.M., *Feature extraction from electroencephalogram by adaptive segmentation*, Proceedings of the IEEE, Pages 642-652, 1977.
- [DAI-1999] Dai, X. Long and Siamak Khorram, "Remotely Sensed Change Detection based on Artificial Neural Networks," Photogrammetric Engineering & Remote Sensing, vol.65, no.9, 8p., September 1999.
- [DEG-1992] Nouvelles Méthodes de Traitement de l’Information en Médecine Rédacteurs : P. Degoulet, A. Hénaut et P. Corvol Volume 5 – 1992.
- [GEE-2000] Gee L. A. and Abidi M.A., *Multisensor Fusion for Decision-based Control Cues* , In Signal Processing, Sensor Fusion, and Target Recognition IX, Proceedings of SPIE Vol. 4052(2000), Pages 249-257, 2000.
- [HINK 1971] Hinkley D.V., Inference about the change point from cumulative sum tests. Biometrika, 58, 509-523, 1971.
- [JEN-1996] Jensen F., An Introduction to Bayesian Networks. Springer Verlag. New York, 1996.
- [KESS-1992] Kessler et al., *Functional Description of the Data Fusion Process, report prepared for the Office of Naval Thechnology, published by the Naval Air Development Center, Warminster, PA, January 1992.*
- [KULL – 1959] Kullback S., *Information Theory and Statistics*, New York Wiley 1959.
- [KUO-2000] Kuo R.J., *Multi-sensor integration for on-line tool wear estimation through artificial neural networks and fuzzy neural network*, In Engineering Applications of Artificial Intelligence 13 (2000) 249-261, January 2000.



- [LEXI-1991] Data Fusion Lexicon, published by the Data Fusion Sunpanel of the Joint Directors of Laboratories Technical Panel for C3 ( F. E. White, Code 4202, NOSC, San Diego, CA), 1991.
- [LI-2000] Li Rong, Wang Ping, Hu Wenlei, *A novel method for wine analysis based on sensor fusion technique*, In Sensors and Actuators B 66 2000 246–250, January 2000.
- [LLIN-1998] Llinas J., Hall D.L., An introduction to multi-sensor data fusion, 0-803-4455 mars 1998 IEEE.
- [NOWA – 1989] Nowakowski S., Détection de défauts dans les séries temporelles ; Doctorat de l’Université Henry Poincaré spécialité Automatique, soutenu le 13/06/1989.
- [MERCK – 1988] Merckle J., Stratégies de détection de rupture de modèle appliquées à la recherche et à la localisation des défauts sur des produits sidérurgiques; Doctorat de l’Université Henry Poincaré spécialité Automatique, soutenu le 21/04/1988.
- [PAGE – 1954] Page E. S., Continuous inspection schemes. *Biometrika*, 41, 100-115, 1954.
- [PARI – 2001] Parikh C.R., Pont M.J. and N.B. Jones, “Application of Dempster-Shafer theory in condition monitoring systems: A case study”, *Pattern Recognition Letters*, 22 (6-7): 777-785, 2001.
- [PEARL-1988] Pearl J. In: Probabilistic reasoning in intelligent systems: Networks of plausible inference. Brachman (ed). Morgan Kaufmann, 1988.
- [RAPP-2001] Rapport intermédiaire du projet LIFE 1999, NANCIE, LYONNAISE DES EAUX – CIRSEE, mars 2001.
- [SAND – 2002] Ferson S., Kreinovich V., Ginzburg L., Myers D. S., and Sentz K.: Constructing Probability Boxes and Dempster-Shafer Structures, SAND REPORTS 2002 - Sandia National Laboratories Albuquerque, New Mexico 87185 and Livermore, California 94550, Unlimited Release Printed January 2003.
- [SCHE – 1991] Sheer-Dorr C., Méthodes paramétriques d’analyse du signal EEG : Application à détection, localisation et analyse spectrale de fuseaux de sommeil; Doctorat de l’Université Henry Poincaré spécialité Automatique, soutenu le 27/09/1991.
- [SMET – 2000] Smets Ph., Data Fusion in the Transferable Belief Model. Proc. 3rd International Conference Information Fusion, Paris, France (2000) PS21-33.
- [SHANG-1999] Shang-Liang Chen, *Data fusion neural network for tool condition monitoring in milling machining*, In *International Journal of Machine Tools & Manufacture* 40 (2000) 381–400, June 1999.
- [SHOR – 1981] Shore J., *Minimum cross entropy spectral analysis*. *IEEE Trans. On ASSP*, Vol. 45, N° 2, pages 230-237, 1981.
- [TCHA - 1997] Tchamova A., *Evidence reasoning theory with application to the identity estimation and data association systems*, In *Mathematics and Computers in Simulation* 43 (1997) 139 – 142.

- [VANO – 2000] Vannoorenberghe P., Lefevre E., Colot O. : "Application de la théorie des fonctions de croyance à la surveillance de l'environnement"; Rencontres Francophone sur la Logique floue et ses applications (LFA'2000); La Rochelle France; Octobre 2000; pp 229-236.
- [VAEZ –1990] Vaezi-Nejad H., Détection de défauts d'instruments de mesure ; Doctorat de l'Université Henry Poincaré spécialité Automatique, soutenu le 23/01/1990.
- [WALD-1999] Wald L., A European proposal for terms of reference in data fusion. *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXII, Part 7, 651-654, 1998, or Wald L., Some terms of reference in data fusion. *IEEE Transactions on Geosciences and Remote Sensing*, 37, 3, 1190-1193, 1999.
- [WILL –1974] Willsky A.S., Jones, *A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems*, in IEEE Trans. On A.C., février 1976.

## **Références Internet**

{1} Groupe de travail sur la fusion de données de l'université canadienne de Laval :

<http://fusion.scg.ulaval.ca/reunion1.htm>

{2} *Le NOAA (National Oceanic and Atmospheric Administration) data Centers*

<http://www.noaa.gov/>

# CHAPITRE IV : Aide à la décision

## IV.1 Introduction

L'analyse de données désigne l'ensemble des méthodes à partir desquelles on collecte, organise, résume, présente et étudie des données pour permettre d'en tirer des conclusions et de prendre des décisions [JAMBU - 2003].

Ce concept de décision peut être défini comme un choix entre actions, solutions ou alternatives [LAAR-2000]. La complexité de ce choix provient du fait qu'il couronne l'aboutissement d'un processus dynamique dépendant de plusieurs facteurs pouvant être plus ou moins long selon le problème considéré. En d'autres termes, la décision est rarement une action circonscrite dans le temps, mécanique ou stéréotypée.

D'autre part la notion de décideur désigne en dernier ressort l'entité qui apprécie le 'possible' et les finalités, qui exprime les références et qui est sensé les faire prévaloir dans l'évolution du processus en question [ROY-1985].

Il est notamment dit que, la théorie de la décision consiste en l'élaboration de modèles statistiques, à partir de connaissances a priori, en vue d'un choix optimal d'une décision [FLE - 2001]. La figure V.1 met en évidence, les différents éléments concourant à la prise de décision.

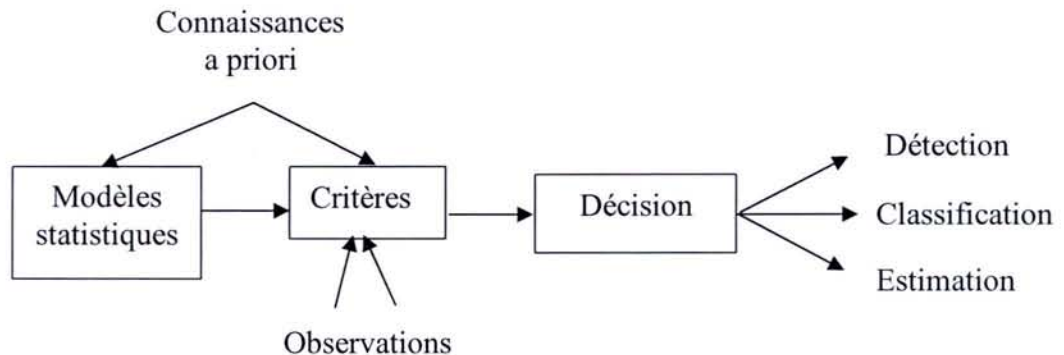


Figure V.1 : « Eléments concourant à la prise d'une décision »

Dans un système complexe comme le nôtre, le décideur étant un opérateur généraliste, n'est pas l'homme de l'étude et doit en outre être le plus autonome possible, d'où l'intérêt d'intégrer une méthodologie d'aide à la décision la plus intelligible possible. Celle-ci doit impliquer le moins d'acteurs possibles, dans le but de préserver l'enjeu temps réel de notre problématique, en l'occurrence éviter la contamination du réseau d'eau potable.

## IV.2 Aide à la décision

Selon Bernard Roy [ROY-1993] : « l'aide à la décision est l'activité de celui qui, prenant appui sur des modèles clairement explicités mais non nécessairement complètement formalisés, aide à obtenir des éléments de réponse aux questions que se pose un intervenant dans un processus de décision, éléments concourant à éclairer la décision et normalement à recommander ou simplement à favoriser, un comportement de nature à accroître la cohérence entre l'évolution du processus d'une part, les objectifs et le système de valeurs desquels cet intervenant se trouve placé d'autre part ».

Cette définition sous-entend, la modélisation formalisée selon les préférences du décideur et d'autre part l'analyse de solutions et l'évaluation de leurs conséquences.

Afin d'avoir une idée concrète de l'aide à la décision, la figure V.2 illustre la mise en situation d'un opérateur face à différentes sources de données :

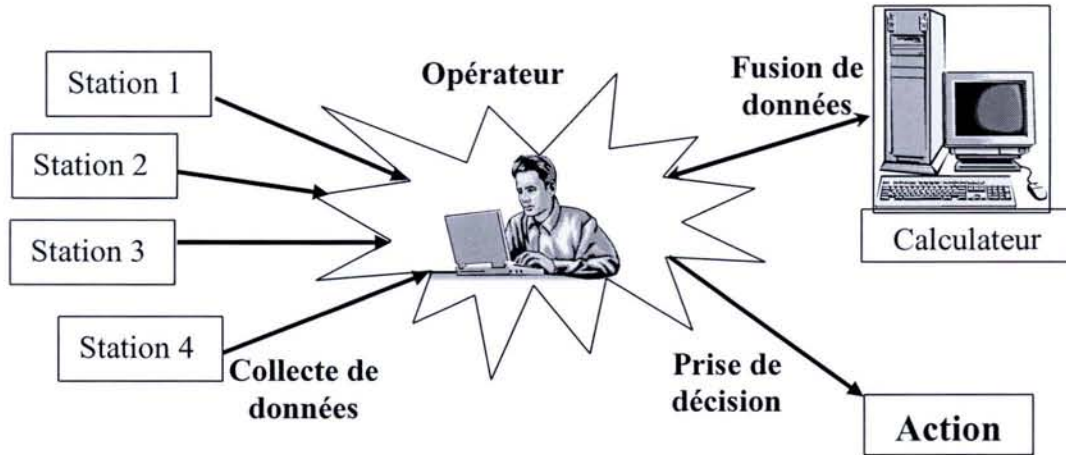


Figure V.2 : « Mise en situation de l'opérateur »

Les données en question sont issues de différentes stations. Pour chacune d'entre elles, il y a plusieurs capteurs à observer et lors d'une situation de crise, l'opérateur doit faire face à ce flux de données en prenant une décision dans les plus brefs délais, ce qui relève d'une grande responsabilité.

Dans ce contexte, notre station n'est pas la seule à être supervisée, s'ajoute à cela le mode de transmission de données sans fusion ni synthèse, à travers lequel les informations arrivent classiquement sous la forme suivante représentée en figure V.3 :

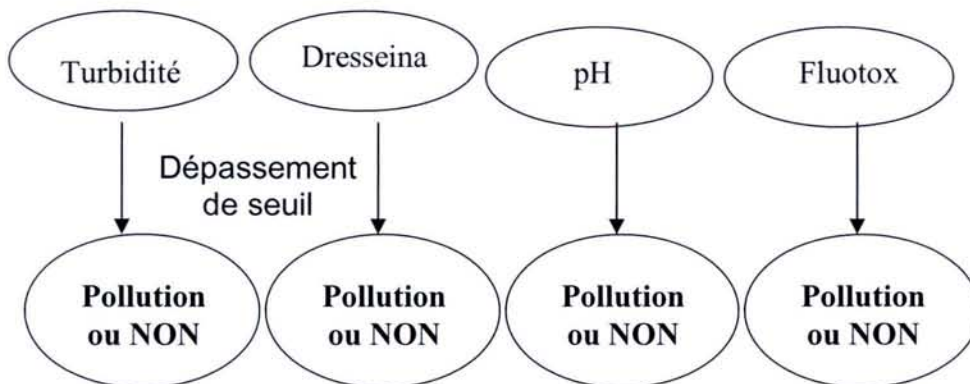


Figure V.3 : « Analyse mono-signal »

La décision se faisant sur la base de l'atteinte d'un seuil élevé d'une analyse mono-signal fait que l'on a :

- une information dispersée,
- des alertes binaires,
- des interprétations différées.

A partir de cet état de faits, on voit déjà un premier apport de la fusion de données schématisé en figure V.4 :

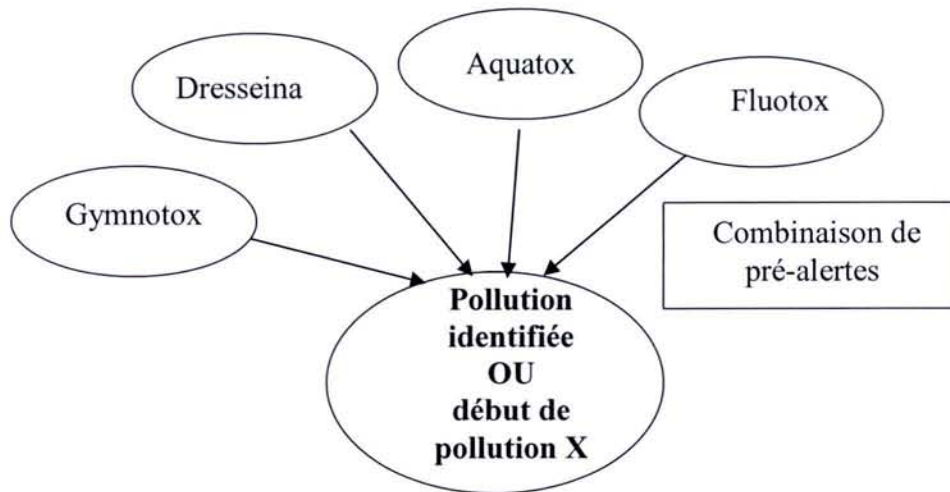


Figure V.4 : « Analyse multi-signal »

Cela permet ainsi d'avoir :

- une information unique à partir de la station concernée,
- une prévention d'une alerte sur la base d'un ensemble de pré-alertes,
- une identification instantanée du type de pollution sur la base de connaissance des tests de toxicité effectués.

Les problèmes de gestion de l'environnement sont par essence complexes ; leur appréhension nécessite une simplification de la réalité par la modélisation, comme par exemple, l'approche systémique. Le gestionnaire de l'environnement devra effectuer des comparaisons et des choix en se basant sur des modèles descriptifs, décrivant les principales composantes de l'objet d'étude, des modèles relationnels, qui déterminent les relations existant entre les composantes du système ou des modèles de simulation.

Le modèle d'aide à la décision proposé par [PICT-2000], représenté en figure V.5, en tant qu'outil d'aide à la décision, basé lui aussi sur les comparaisons de variances, met en exergue les facettes de la communication dans le cadre d'une prise de décision :

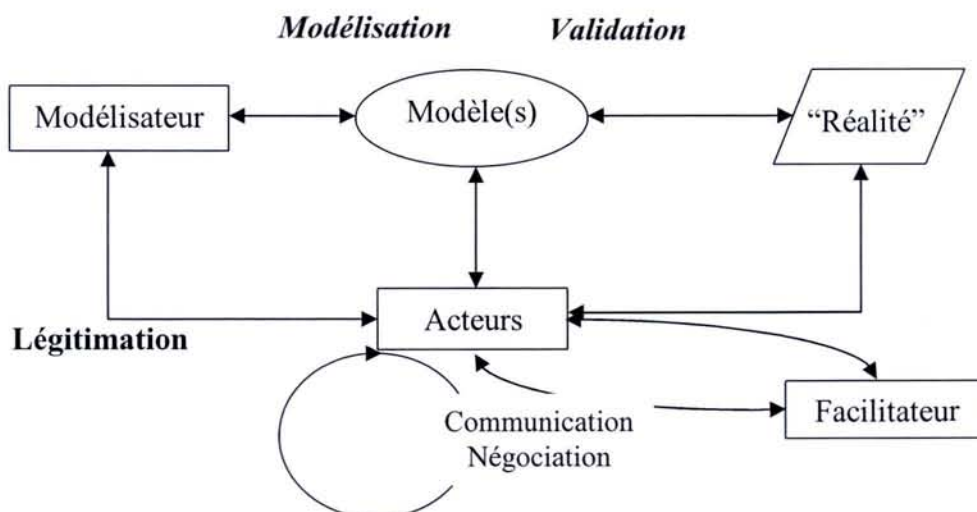


Figure V.5 : « Modèle d'aide à la décision de l'Outil : " l'expert " »

Ce type d'architecture montre bien les interactions entre différents éléments d'un processus de décision, notamment celles auxquelles sont confrontés les acteurs d'une prise de décision. L'intérêt en environnement concerne des acteurs différents, impliqués de manière diverse, et avec des perceptions souvent divergentes du problème et de ses solutions.

De manière plus explicite, l'opérateur est un maillon du cycle de l'information modélisé par [JAMBU - 2003] selon la méthodologie représentée en figure V.6. Les différentes étapes sont positionnées selon le cycle de Deming<sup>5</sup>. Ce cycle comporte quatre étapes dont les noms, Plan, Do, Check, Act, sont traduits respectivement par:

- **Préparer** : Cette étape, la plus importante, consiste à cerner les objectifs et préparer le conditionnement de données en vue de son étude. Le cycle complet repose sur cette étape. Dans notre cas cette étape correspond à la stratégie de conception de la station d'alerte en vue de confronter les données produites aux tests de toxicités, base à partir de laquelle les profils de pré-alertes seront déduits
- **Développer** : Cette étape consiste à réaliser l'essai comparatif ou l'expérience, de référence à petite échelle, en respectant bien les dispositions définies à la première étape, dans notre cas cette étape correspond aux tests de toxicité.
- **Comprendre** : Il s'agit d'étudier les résultats, dans notre cas il est question de la fusion de données, plus explicitement l'établissement de profils d'alertes.
- **Agir** : Il s'agit de l'étape de décision et passage à l'acte, c'est au bout de cette étape que l'opérateur est confronté à sa responsabilité.

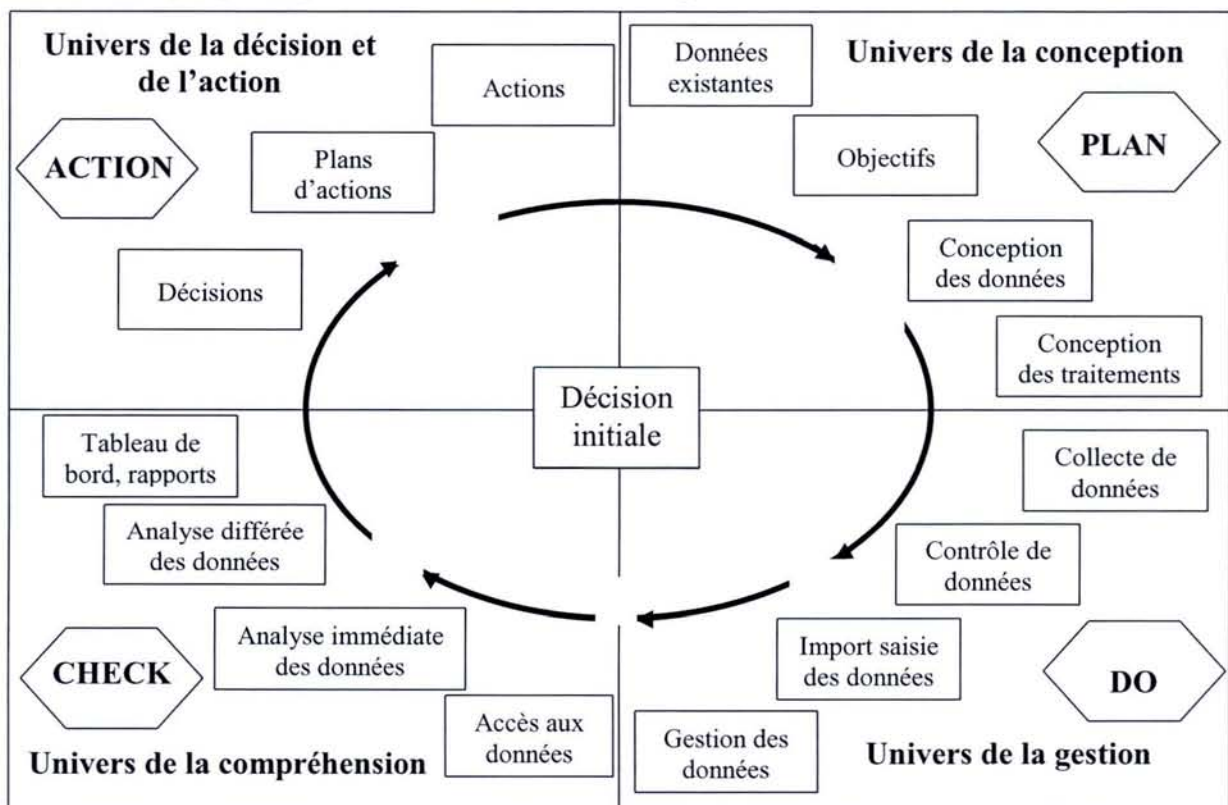


Figure V.6 : Méthodologie de l'information (Plan, do, check, action)

<sup>5</sup> A titre indicatif, cette méthode d'aide à l'apprentissage, ainsi que la conduite et amélioration d'un produit ou d'un processus ont été initialement proposées en juillet 1950 par *Edwards Deming (1900-1993)* au comité directeur du patronat japonais, et a été adopté en tant que méthode de management, qui a permis en l'occurrence de révolutionner l'économie japonaise.

Le fait d'avoir formalisé notre modèle d'information sous forme de cycle a pour objectif de le rendre évolutif de part sa forme de modèle d'amélioration continue. En effet, cette approche méthodologique est une aide à l'apprentissage, ainsi qu'une conduite d'amélioration de processus. Dans notre cas les modèles explicités sont représentés par les profils de pollutions issus de la fusion de données.

Ce cycle montre bien que malgré l'aspect synthétique des informations fusionnées, l'opérateur demeure soumis à une responsabilité de décision ultime, difficile à prendre. De ce constat, il est nécessaire de définir les critères selon lesquels l'opérateur doit se prononcer et selon quelles priorités il doit les faire passer, tout cela en vue d'une décision pour une action prédéfinie. Ceci nous amène à définir ci-après l'aide multicritère à la décision et son adaptation à notre contexte.

### IV.3 Adaptation méthodologique à notre contexte

L'approche classique des problèmes de décision, correspondant à l'optimisation d'une unique fonction d'évaluation, montre certaines faiblesses de traitement à une époque où le flux de données et le nombre de calculs qui incombe à l'exploitant d'un système d'information n'est que grandissant. C'est dans cet esprit, que les méthodes d'aide multicritères à la décision ont été développées, permettant ainsi de gérer un ensemble de contraintes à la fois tout en préservant la dimension temporelle.

En effet, avant l'apparition de l'analyse multicritère, les problèmes de décision se ramenaient le plus souvent à l'optimisation d'une fonction économique. Cette approche avait le mérite de déboucher sur des problèmes mathématiques bien posés mais qui n'étaient pas toujours représentatifs de la réalité car :

- la comparaison de plusieurs actions possibles se fait rarement suivant un seul critère ;
- les préférences sur un critère sont, dans bien des cas, difficilement modélisables par une fonction ;
- lorsqu'il y a plusieurs objectifs, il est impossible de les atteindre tous à la fois [MAYE-1994].

Dans le contexte de notre problématique de décision, il s'agit de gérer des données à la fois qualitatives et quantitatives en vue de produire une décision dans des délais raisonnables, ce qui par analogie nous amène à opter pour les méthodes multicritères qui conviennent pour la synthèse d'informations d'horizons divers, destinées à l'aide à la décision.

Afin d'être plus explicite et d'étayer la présentation de l'aide multicritère à la décision, plusieurs définitions peuvent en découler pour le même objectif. En l'occurrence, Vincke [VINC-1989] définit l'aide multicritère comme suit : « *L'aide multicritère à la décision vise, comme son nom l'indique, à fournir à un décideur des outils lui permettant de progresser dans la résolution d'un problème de décision où plusieurs points de vue, souvent contradictoires, doivent être pris en compte* ».

Selon Ben Mena [BEN-2000], à propos des méthodes d'aide multicritère à la décision : « *Par leur manière d'intégrer tout type de critères, ces procédures semblent mieux permettre de se diriger vers un judicieux compromis plutôt qu'un optimum souvent désuet. La diversité de ces méthodes réside dans la façon d'effectuer la synthèse de l'information contenue dans chaque critère* ».

Il apparaît clairement que cette méthode d'analyse aide à synthétiser un grand nombre d'informations, souvent très diverses et peu comparables, en fonction des objectifs du



décideur. Il est notamment dit, selon Martel [MART-1988] que l'aide multicritère à la décision œuvre à apporter un éclairage et des explications à une catégorie de problèmes selon quoi :

- Plusieurs critères quantitatifs et qualitatifs sont pris en considération ;
- ces critères sont souvent hétérogènes dans le sens où les mesures s'effectuent sur différentes échelles ;
- ces critères peuvent être conflictuels ;
- ces critères sont considérés d'inégale importance.

Cette méthode développée initialement dans le cadre de la recherche opérationnelle dont l'objet est la préparation des décisions incombant à une autorité responsable, est couramment utilisée dans des problèmes de gestions économiques. Son adaptation à notre contexte environnemental, se justifie par l'enjeu socio-économique qu'engendre la surveillance d'une ressource aussi stratégique que l'eau.

L'aide multicritère à la décision suppose l'existence d'un décideur (une personne confrontée à un choix). Cependant si la responsabilité d'une décision incombe à un individu clairement identifié, celle-ci est souvent résultante d'interactions entre de multiples acteurs.

Dans notre cas il s'agira de l'opérateur de surveillance qui aura comme interlocuteurs privilégiés un ou plusieurs experts pour conforter son analyse ainsi que les autorités locales pour déclencher l'action à entreprendre.

Avant de décrire l'adaptation de cette méthode à notre approche, il est important de définir certains concepts de base ainsi que les fondements des principales méthodes multicritères.

En l'occurrence, selon Vincke [VINC-1989], un critère est une fonction "  $g$  ", définie sur l'ensemble des actions "  $A$  ", qui prend ses valeurs dans un ensemble totalement ordonné, et qui représente les préférences de l'utilisateur selon un point de vue.

L'exemple de la moyenne d'un élève est un critère déterminant la valeur de l'élève. Selon le contexte, la définition de la "valuation" (manière de donner une valeur à un critère) peut changer. En revanche, la définition d'un critère demeure identique.

Parmi les différentes méthodes d'aide à la décision, se trouvent les méthodes ELECTRE (**EL**imination **Et** **Ch**oix **TR**aduisant la **RÉ**alité) initialement développées par Bernard Roy en 1968 [VINC-1989] et PROMETHEE (**P**reference **R**anking **O**rganisation **METH**od for **E**nrichment **E**valuations) développée par Brans, Vincke et Mareschal qui est une variante de ELECTRE [BRAN-1986].

Il découle de ces méthodes, l'établissement de matrices de décisions, liant les actions :  $a_i, i=1, \dots, k$  aux critères  $g_j, j=1, \dots, n$ , dans lesquelles l'évaluation de l'action  $a_i$  suivant le critère  $g_j$  noté  $g_j(a_i)$  implique l'établissement de combinaisons prédéfinies liant une action à plusieurs critères. Se rajoute à ces critères, l'affectation de pondérations, selon leur importance dans la prédétermination de l'action. Le rôle de ces poids peut être qualitatif ou quantitatif selon l'approche employée. Il en ressort que la fiabilité de l'approche dépend essentiellement des poids devant refléter le plus fidèlement possible les préférences du décideur.

Dans notre cas, l'adaptation de l'aide multicritère à notre contexte se base sur le fait qu'une action est déclenchée en fonction de plusieurs critères auxquels ont été affectés des poids prédéfinis à partir d'une méthode d'évaluation directe basée à son tour sur d'autres

critères d'évaluations. Il en découle une évaluation de la situation sous forme de valeur qui n'est rien d'autre que la fonction d'utilité. Cette fonction ainsi produite reflète au mieux la réalité des préférences du décideur.

Dans notre démarche méthodologique, l'étape initiale avant toutes évaluations commence par le recensement des différents paramètres que sont les actions à entreprendre et les critères inhérents associés à leurs poids respectifs.

Le recensement des actions selon le type de décisions à prendre s'établit selon l'ordre suivant :

- arrêt des pompages sur le cours d'eau et passage sur le bassin de sécurité, en cas de pré-alertes majeures,
- instruction d'augmentation de dosage préventif anti-polluant pour l'usine de traitement en cas de pré-alertes intermédiaires non négligeables,
- maintien du pompage habituel et dosage préventif en deçà des seuils de pré-alertes intermédiaires.

Concernant le recensement des critères motivant la prise de décision, ils dépendent des capteurs concernés par le masque de pollution établi lors de la fusion de données ou de l'ensemble des capteurs ayant réagi de manière prononcée. Ces capteurs pris un par un sont évalués sur la base du seuil de leur rapport de variance étalé sur leurs plages de pré-alertes, comme précédemment défini en fusion de données.

Concernant l'affectation des poids à attribuer pour chaque critère que représentent les capteurs, il s'agit d'effectuer un paramétrage délicat à déterminer. En effet, on se base sur une connaissance experte, pas nécessairement acquise pour certains capteurs, pour lesquels nous n'avons pas le recul suffisant, en l'occurrence, la quasi-totalité des capteurs biologiques utilisés à titre de prototypes dans notre station. Il en ressort un intérêt à établir des critères rendant ces affectations de pondérations, les plus crédibles possibles, en l'occurrence :

- la fiabilité des capteurs en terme de maintenance,
- leur niveau d'entretien en terme de disponibilité et de périodicité de mobilisation.

Ces pondérations demeurent fixes pour chaque capteur, mais sont normalisées selon l'ensemble de capteurs considérés par le masque de pollution en question.

Pour illustrer ces notions, dans le tableau V.1, est représenté l'exemple de cinq critères concernant une pollution prédéterminée, en l'occurrence une pollution agricole, à laquelle sont associées les trois actions précédemment citées.

	Critères				
	<i>g1</i>	<i>g2</i>	<i>g3</i>	<i>g4</i>	<i>g5</i>
<b>Actions</b>					
<i>a1</i>	<i>X1</i>	<i>X2</i>	<i>X3</i>	<i>X4</i>	<i>X5</i>
<i>a2</i>					
<i>a3</i>					
<b>Pondérations</b>	<i>P1</i>	<i>P2</i>	<i>P3</i>	<i>P4</i>	<i>P5</i>

Tableau V.1 : Illustration d'action selon des critères pondérés

A partir des données de ce tableau il est possible de définir une **fonction d'utilité F** intégrant les paramètres suivants :

- $g_1, \dots, g_5$  représentent les capteurs concernés par la pollution,
- $P_1, \dots, P_5$  représentent les pondérations respectives affectées aux différents capteurs et déterminées à partir d'une évaluation de la **crédibilité** de ces capteurs que l'on détaille ultérieurement (voir le tableau V.4 ), et dont les valeurs sont normalisées selon une échelle de 0 à 1 et dont la somme est l'unité "1" .
- $X_1, \dots, X_5$  sont les valeurs affectées aux fonctions de comparaisons de variances et que l'on représente sous forme de pourcentage basé sur une échelle de valeurs spécifique à chacun des capteurs, que l'on discrétise sous la forme suivante (Tableau V.2) :

	Capteur1	Capteur2	Capteur3	Capteur4	Capteur5
$X$ le rapport de variance	100% si rapport < $w_1$	100% si rapport < $w_2$	100% si rapport < $w_3$	100% si rapport < $w_4$	100% si rapport < $w_5$
	50 % si $w_1 < \text{rapport} < x_1$	50 % si $w_2 < \text{rapport} < x_2$	50 % si $w_3 < \text{rapport} < x_3$	50 % si $w_4 < \text{rapport} < x_4$	50 % si $w_5 < \text{rapport} < x_5$
	0% si rapport > $x_1$	0% si rapport > $x_2$	0% si rapport > $x_3$	0% si rapport > $x_4$	0% si rapport > $x_5$

Tableau V.2 : Discrétisation des niveaux de rapports de variances

Le paramétrage de ces paliers est basé sur les tests de toxicité et doit être conforté, dans le cadre de l'affinement du système, au fil de son exploitation.

- $a_1, a_2, a_3$  représentent les trois actions possibles dans notre contexte de surveillance ; ils sont en l'occurrence régis par une fonction d'utilité normalisée prédéfinie selon le calcul suivant :

$$F = P_1X_1 + P_2X_2 + P_3X_3 + P_4X_4 + P_5X_5 \quad (V.1)$$

La fonction d'utilité telle que définie précédemment, permet une évaluation globale de la situation à apprécier. Cette fonction correspondant à un pourcentage variant de 0% à 100% correspond à la normalisation héritée des grandeurs qui la régissent. Le fractionnement de cette échelle, permet d'établir des actions selon des plages prédéfinies que l'on classe selon le tableau V.3 :

Actions	Plage de la fonction d'utilité	Qualification du type de situation
$a_1$ : Arrêt pompage/ Passage au bassin de sécurité.	$70\% < F \leq 100\%$	Cas de pré-alertes majeures.
$a_2$ : instruction d'augmentation de dosage préventif anti-polluant pour l'usine de traitement.	$30\% < F \leq 70\%$	Cas de pré-alertes intermédiaires non négligeables,
$a_3$ : maintien du pompage habituel et dosage préventif en deçà des seuils de pré-alertes intermédiaires.	$0\% < F \leq 30\%$	Cas de situation dite normale, sans pollution considérable

Tableau V.3 : Correspondance fonction d'utilité/actions

D'un point de vue pratique, il s'agit en l'occurrence d'établir ce type de tableau pour chaque masque de pollution ainsi que pour celles dites inconnues notées "X", et pour lesquelles l'enregistrement du nouvel événement y est systématiquement effectué dans le système d'information. Il est conforté par une analyse laboratoire qui identifie et classe le type de pollution.

En effet, dans le cas d'une pollution inconnue tel que définie au point **IV.3.5** du chapitre précédent, l'enregistrement de l'événement, dans la base de connaissance, passe par une affectation des capteurs ayant réagi au profil de pollution "X". Cela permet en l'occurrence, de définir la fonction d'utilité associée et de la normaliser en conséquence.

La prédétermination initiale de ces grandeurs n'implique pas une configuration figée. En effet un affinement de notre configuration issue de la simulation, peut s'avérer différent du contexte réel d'une pollution grandeur nature. Il en ressort la nécessité d'un calage des valeurs tout au long de l'exploitation du système. Concrètement, cela passe, soit par une modification sur les intervalles de déclenchement d'actions, soit par une redistribution des niveaux de pondérations ou en dernier lieu sur les niveaux de pré-alertes.

De ce fait, l'affectation d'un bon niveau de crédibilité à l'information fournie par le système de fusion de données, trouve tout son intérêt pour que l'on puisse l'utiliser en tant qu'information principale.

En effet, l'information de pré-alerte issue de l'association des niveaux de seuils de chaque capteur, est confortée par l'affectation de poids par rapport à leurs **crédibilités** respectives comme préconisées dans le **tableau V.1**. Ces pondérations sont classiquement affectées à partir d'une connaissance experte du dispositif. Celle-ci est fondée sur le jugement de l'historique de fonctionnement des capteurs sans avoir nécessairement recours à une évaluation statistique. Du coup, ce type démarche peut avoir une évaluation subjective de la réalité, notamment pour les capteurs biologiques pour lesquels on a un trop faible recul pour leur affecter la pondération adéquate.

Pour pallier à ce type d'affectations dénuées de fondements mathématiques, nous proposons d'associer la **crédibilité** à la notion de **disponibilité** du capteur. En effet, les méthodes multicritères d'aide à la décision permettent d'attribuer un sens à ces pondérations dès lors qu'il est possible de quantifier la crédibilité. De ce fait, nous sortons du contexte classique, où tout à priori est levé sur des capteurs ayant été en panne et pour lesquels des améliorations ont été apportées, donc une disponibilité plus élevée et une pondération mise à jour.

Il en ressort que pour prédéterminer ces poids sur la base de la disponibilité de fonctionnement des capteurs, il est nécessaire d'identifier les niveaux d'indisponibilités à qui sont en l'occurrence : le **niveau de panne** et le **niveau d'entretien**. Ces deux cas rassemblent le niveau d'immobilisation du capteur, que nous avons choisi d'évaluer annuellement sur une période de 365 jours. Le principe consiste à considérer que s'il y'a intervention, au cours d'une journée, celle-ci est considérée comme n'étant pas comptabilisée dans les disponibilités. Cela nous permet d'en déduire le nombre de jours de disponibilité du capteur que l'on formule comme suit :

$$Disponibilité = 365 - (Panne + Entretien) \quad (V.2)$$

l'unité de la formule (V.2) étant le nombre de jours.

Sur la base de la formule V.2, il est possible de calculer des pourcentages sur une année, en l'occurrence le pourcentage de maintenance ( $M_i$ ) et le pourcentage d'entretien ( $E_i$ ), qui par la suite sont soustrait au pourcentage de nombre de jours d'une année (100%), pour enfin déduire le pourcentage de jours de disponibilité comme détaillé sur le tableau V.4:

	Capteur1,.....	.....	Capteur K
1) Maintenance/Panne	M <sub>1</sub> %.....	.....	M <sub>k</sub> %
2) Entretien	E <sub>1</sub> %.....	.....	E <sub>k</sub> %
<b>Disponibilité</b>	(100-(M <sub>1</sub> +E <sub>1</sub> ))%....	.....	((100- (M <sub>k</sub> +E <sub>k</sub> ))%)

Tableau V.4 : Détermination du pourcentage de disponibilité

Le pourcentage de jours de disponibilité étant défini, nous pouvons le normaliser **pour chaque capteur** en jeu durant **la pollution en question**. En effet, la normalisation personnalisée des poids à un type de pollution spécifique, s'explique par notre souci de préserver l'unicité de l'échelle de la fonction d'utilité. En effet le poids affecté à un capteur au cours d'une pollution, n'est pas nécessairement identique pour le même capteur au cours d'un autre type de pollution, du fait que ce n'est pas nécessairement le même groupe de capteurs qui a réagi à ce second type de pollution.

La détermination des poids se fait en fonction des capteurs concernés par la pollution et en normalisant les pourcentages de disponibilité, de manière à respecter la condition énoncée à partir du tableau V.1, relatif à la somme des poids devant être égale à 1, respectant ainsi l'unicité de la fonction d'utilité.

Pour illustrer le calcul, continuons sur l'exemple précédent d'une pollution agricole pour laquelle cinq capteurs C1, C2, C3, C4, C5 ont réagi et dont les disponibilités respectives sont : D1 = 90%, D2=60%, D3=75%, D4=25%, D5= 50 %

La normalisation des disponibilités indépendantes du type de pollution, fera qu'elle personnalisera les poids respectifs en fonction des capteurs ayant réagi, selon les règles de trois suivantes, leur affectant ainsi un niveau de crédibilité:

$$P1 = \frac{D1 * 100\%}{(D1 + D2 + D3 + D4 + D5)} = 30\% \quad (V.3)$$

$$P2 = \frac{D2 * 100\%}{(D1 + D2 + D3 + D4 + D5)} = 20\% \quad (V.4)$$

$$P3 = \frac{D3 * 100\%}{(D1 + D2 + D3 + D4 + D5)} = 25\% \quad (V.5)$$

$$P4 = \frac{D4 * 100\%}{(D1 + D2 + D3 + D4 + D5)} = 8,33\% \quad (V.6)$$

$$P5 = \frac{D5 * 100\%}{(D1 + D2 + D3 + D4 + D5)} = 16,66\% \quad (V.7)$$

En ayant déterminé les poids de crédibilité (P<sub>i</sub>) et la sensibilité des capteurs représentés par le rapport de variances (X<sub>i</sub>) la fonction d'utilité est complétée. Nous pouvons ainsi quantifier au sein d'une même fonction la sensibilité des capteurs face à une pollution en tenant compte d'un critère de fiabilité. Selon le niveau de la fonction d'utilité, nous adoptons les règles du tableau V.3, correspondant aux actions à entreprendre.

Afin de vérifier la préservation de l'échelle, il suffit de mettre les sensibilités à leur maximum (100%) :

$$F = P1(100\%) + P2(100\%) + P3(100\%) + P4(100\%) + P5(100\%) \quad (V.8)$$

Etant donné que la somme des poids vaut l'unité, notre fonction d'utilité respecte l'échelle du **tableau V.3**.

En terme de résultat, si l'on prend le cas du Gymnotox qui est un capteur biologique performant de part le large panel de pollution qu'il couvre, mais dont le niveau de panne était élevé en étant crédible qu'à hauteur de 25%, cela montre bien notre souci de conjuguer au sein d'une même fonction, crédibilité et sensibilité, de manière à minimiser le crédit à affecter, pour un capteur qui peut fournir des données erronées.

Nous pouvons aussi citer l'exemple d'un autre capteur biologique qu'est l'Aquatox où l'on doit nécessairement changer les daphnies, hebdomadairement ce qui immobilise le capteur au moins un jour par semaine. Nous constatons d'ailleurs que sur l'ensemble des capteurs biologiques, excepté le Fluotox, le niveau de disponibilité demeure en deçà de celui des capteurs physico-chimiques. L'intérêt de la crédibilité inhérente, est d'introduire les capteurs biologiques dans le cycle de décision, avec une crédibilité certes minime, mais qui permet d'avoir l'information conduisant à la prise de la bonne décision.

Il n'en demeure pas moins que l'amélioration des nouvelles versions de capteurs biologiques tend à réduire l'écart de disponibilité par rapport aux capteurs physico-chimiques, d'où l'intérêt de mettre à jour annuellement les disponibilités, rejoignant ainsi notre **principe de modèle d'amélioration continue**.

#### **IV.4 Evaluation des Biocapteurs**

Après avoir exploité les bio-capteurs tout au long de cette étude, une phase sélective a été entreprise pour passer à l'étape d'exploitation réelle du dispositif. En effet, cette exploitation passe par une autonomie de la station et un entretien minimal, réduisant les immobilisations et la mobilisation de personnel. Or, il s'avère que certains bio-capteurs, comme cité précédemment, ont nécessité une attention particulière. D'autre part, la sensibilité mitigée produite par certains bio-capteurs ainsi que la redondance des uns avec les autres, ont permis d'opter pour certains choix. Toutes ces considérations ont été motivées par les arguments que l'on détaille comme suit :

##### Le Bacteria-test

*Inconvénients :*

- Au cours du projet, certains problèmes sont survenus sur l'appareil, notamment une panne des photo-multiplieurs, éléments essentiels pour la mesure. La culture de bouillon bien que standardisée a aussi posé quelques problèmes durant le projet. Il n'a pas été possible de déterminer la cause précise, de la mauvaise qualité persistante des bouillons obtenus (malgré de nombreux essais : changement de la souche, temps d'incubation plus ou moins long...).
- Fréquence des mesures non adapté à l'objectif à savoir rapide et continu (30 min = trop long).
- Pas de protection vis-à-vis de l'extérieur, l'appareil n'est pas confiné dans un boîtier hermétique (comme Aquatox ou Fluotox, par exemple).
- Laboratoire obligatoire.
- Ventilateur et alimentation des photo-multiplieurs souvent en panne.

#### *Avantages :*

- Pas de détérioration du réactif biologique lors d'une pollution.
- Test classique, connu depuis longtemps.
- Bonne simplicité générale du système : Visualisation claire de la réaction.

#### Fluotox

##### *Inconvénients :*

- Renouvellement du buvard algal à chaque pollution importante (intervention sur site).

##### *Avantages :*

- Très peu de maintenance et système de sécurité performant (notamment en cas d'arrêt d'eau, fuite électrique ou hydraulique).
- Nécessité de régulation de la température (incluse dans appareil) bien maîtrisée.
- Périodicité de la collecte des informations (possible toutes les secondes).
- Facilité de récupération et d'interprétation des résultats via le synoptique et via le rapatriement.

#### Aquatox

##### *Inconvénients :*

- Encrassement des tubes (entretien difficile de l'échangeur thermique/ serpentin).
- Maintenance des tuyaux d'alimentation difficile, ce qui implique un technicien spécialisé, formé à cet effet.
- Filtration et changement des daphnies impliquant un temps maximal de cinq jours d'immobilisation en chambre, dû au nombre fixe de daphnies devant être présentes dans chaque chambre.
- Fragilité des éléments : têtes de pompes et serpentins de chauffe en verre.
- Difficulté d'interprétation des résultats bruts.

##### *Avantages :*

- Bien isolé de l'extérieur sauf la station d'élevage qui se trouve à l'extérieur du module
- Facilité de se procurer et d'élever les organismes tests.
- Maîtrise des influences extérieures et mesure fiable, rapide et automatique.

#### Dreissena

##### *Inconvénients :*

- Temps de maintenance élevé dû à la formation de filaments entraînant un dérèglement de la position des moules. (Note: si l'entretien est négligé, il n'y a pas assez de moules opérationnelles d'où des "fausses alertes" ou pas d'alertes car elles sont mises hors mesures par l'ordinateur).
- Pas de sécurité en cas de changement des débits d'alimentation.

#### *Avantages :*

- Subsistance des moules après pollution (mortalité corrélée avec la température).
- Facilité de récupération des moules (présentes en nombre dans le milieu naturel).
- Fonctionne à  $T^{\circ} > 5^{\circ}\text{C}$  (pas de système de chauffe nécessaire ni de filtration).
- Maîtrise des influences extérieures : cependant le claquement d'une porte, une agitation de personnes autour semblent perturber le signal, du fait que l'on observe une fermeture des moules.

#### Gymnotox

##### *Inconvénients :*

- Validité des résultats basée sur l'étalonnage, s'il y a mauvais étalonnage, le résultat est faussé.
- Problème lié au système de chauffe: création de micro-bulles due à la différence entre les températures d'entrée et de sortie. En effet la régulation de température se fait sur la température de sortie du thermo-plongeur et non pas sur la température dans les bassins des poissons, ce qui implique de longues pannes.
- Fragilité des poissons : notamment en hiver (micro-bulles pouvant induire une forte mortalité des poissons) et face à la maladie.
- Nécessité de vidanger et de nettoyer complètement l'installation après alarme de pollution.

##### *Avantages :*

- Longue autonomie si le poisson est bien acclimaté lors de son introduction.
- Mesure en continue et lecture / interprétation aisées.
- Réactivité importante, couvrant un large panel de pollution.

Il ressort de ces bilans ainsi que des résultats obtenus tout au long de l'étude un choix pragmatique :

- lié à la maintenance et la disponibilité,
- une pertinence de la réactivité des capteurs face aux pollutions,
- les efforts que doivent consentir les fournisseurs pour améliorer leurs bio-capteurs.

De ces critères, deux capteurs ont été retenus, en l'occurrence, le Fluotox dont les niveaux de disponibilité et de fiabilité sont équivalents à ceux des capteurs physico-chimiques, et le Gymnotox dont les propriétés de détection multirisques ont été prouvées ainsi que son évolutivité depuis les premiers soucis de fonctionnements, rendant ainsi sa disponibilité plus élevée.

Il en ressort que retenir un capteur au détriment d'un autre, n'est pas si préjudiciable de part la redondance de sensibilité entre capteurs biologiques. D'autre part, le maintien d'une installation de qualité, importe plus, que d'avoir l'ensemble de la chaîne trophique avec de fortes contraintes de maintenances.



## IV.5 Système d'alerte mis en place

Le choix des capteurs de la chaîne trophique a certes l'avantage de pouvoir observer une pollution à différents stades de la chaîne alimentaire aquatique du cours d'eau, mais leur exploitation dans le cadre d'un dispositif d'alerte, relève de considérations pragmatiques ne nécessitant pas leur maintien total comme constaté lors de l'évaluation précédente.

En terme de communication, le système dont la vocation est d'aider à la décision l'opérateur, doit être le plus souple possible en terme de circulation de l'information. Cela implique un respect du cycle d'information, en terme de synchronisation temporelle mais aussi de manière à s'implémenter dans un processus de modèle d'amélioration continue. La figure V.7 détaille le dispositif en place :

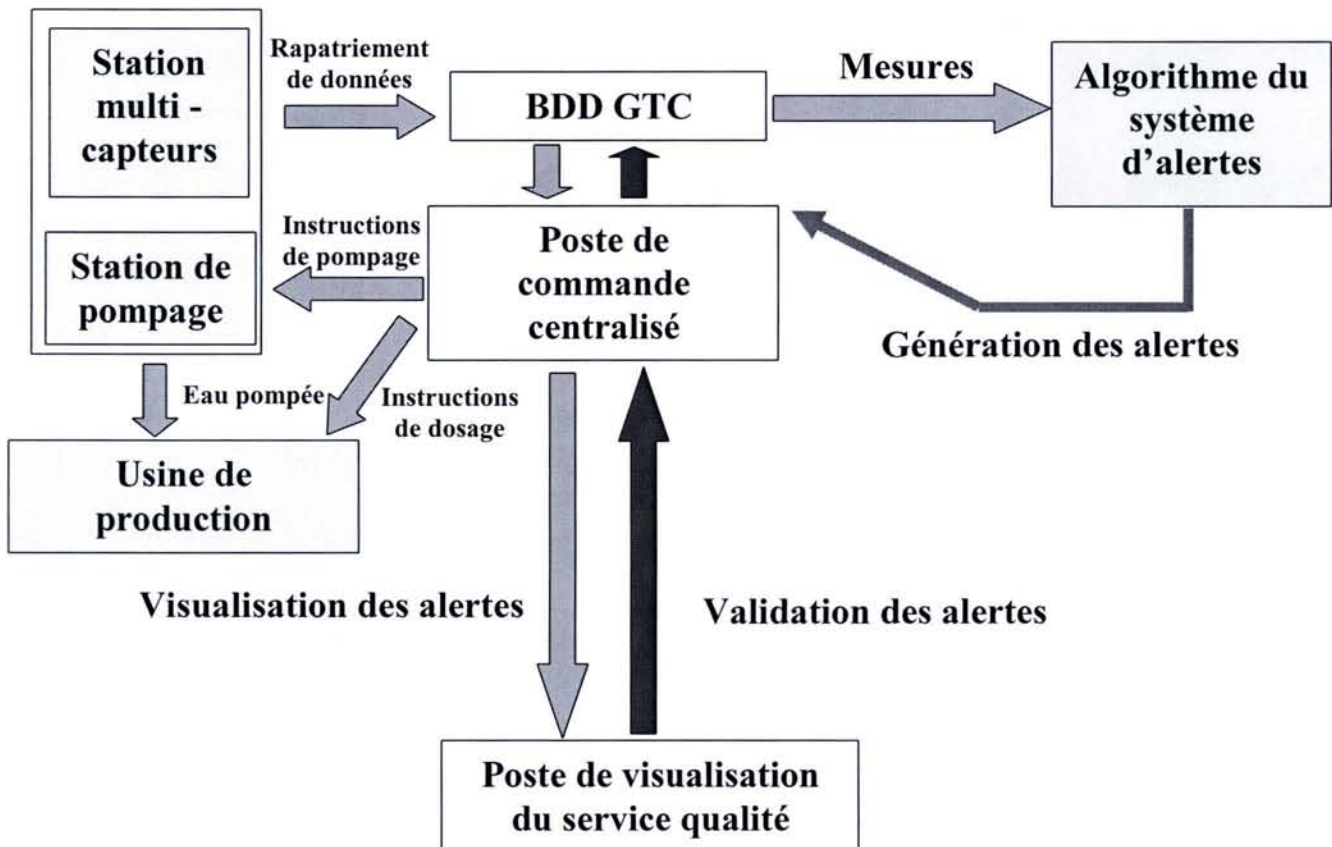


Figure VI.7 : « Dispositif mis en place »

Le système d'information, tel que décrit sur la figure IV.7, englobe les différents maillons rentrants dans la chaîne de décision, en l'occurrence :

- La station de pompage : Délocalisée à une dizaine de kilomètres de la GTC, les ordres de pompage peuvent être des commandes sur le débit de pompage, d'arrêt et passage au bassin de sécurité.
- La station multi-capteurs : Implantée au lieu même du pompage, de manière à synchroniser le prélèvement de l'échantillon d'eau à celui envoyé à l'usine de traitement, elle conserve les mêmes caractéristiques décrites au **chapitre II**, à l'exception d'un ensemble de simplifications, en l'occurrence :

1. Le maintien de deux capteurs biologiques que sont le Fluotox et Gymnotox, sur les cinq initiaux, comme justifié précédemment.
  2. Cette simplification du nombre de capteurs biologiques, permet de supprimer le serveur interface OPC (voir **annexe I**), et du fait de sa suppression, modifie le mode de rapatriement du Gymnotox qui se fait directement sans passer par l'interface.
  3. Se rajoute à ces simplifications, le superviseur local qui est transféré au sein du poste de visualisation du service qualité ainsi que serveur de calcul local implémenté au niveau de l'algorithme de fusion de données. Ces systèmes rapatriés au sein même de la GTC, permettent une gestion centralisée de la station distante et évite une maintenance délocalisée des systèmes informatiques.
- L'usine de production d'eau : Chargée de la distribution de l'eau, elle a pour rôle de pré-traiter l'eau en fonction de certaines instructions de dosages, dépendant de la qualité de l'eau constatée.
  - Le poste de visualisation du service qualité : C'est le poste opérateur décisionnel, il est directement relié aux hautes instances décisionnelles de l'agglomération. Les informations synthétisées y sont directement acheminées afin d'être validées pour produire une décision.
  - Le poste de commande centralisé : C'est le poste opérationnel, exécutif d'une décision aboutissant à ordonner l'action à entreprendre.
  - La base de données GTC : C'est à ce niveau que sont stockées la base d'événements et la configuration du système d'alerte. Elle a pour rôle d'alimenter l'algorithme du système d'alerte.
  - L'algorithme du système d'alerte : C'est le cœur de notre système dans lequel est implémenté l'algorithme de fusion de données et aide à la décision, les alertes produites peuvent être configurées en fonction du type d'événements souhaités, sachant que l'information principale à produire est concentrée au sein de la fonction d'utilité.

Nous pouvons rajouter à cette description synthétique du système d'information, que l'opérateur décisionnel a un regard sur l'ensemble des informations rapatriées autant synthétisées que brutes. Cela montre bien la complexité de sa tâche à pouvoir concilier des informations d'horizons divers, notre contribution est de lui apporter les informations qualitatives utiles à sa prise de décision.

#### **IV.6 Réflexion sur le positionnement de la station**

La station d'alerte positionnée sur le lieu de pompage, limite la marge de temps nécessaire à la production d'une alerte, au temps de rapatriement de l'eau à l'usine de traitement. Il en ressort, un intérêt à positionner la station en amont du processus de pompage, se posera alors le problème du tronçon incontrôlé entre la station d'alerte et la station de pompage. Dans ce cas précis nous pouvons appliquer le principe de « ne pas mettre tous ses œufs dans le même panier », où on maintiendrait les capteurs physico-chimiques sur le pompage et les capteurs biologiques en amont. La fusion serait alors complémentaire mais avec moins de paramètres sur les profils de pollutions. Le fait de mettre les capteurs physico-chimiques sur le pompage, s'expliquera que s'il y a pollution entre les deux stations relativement proches. Cette pollution n'aura pas le temps de se diluer et aura un effet tellement important qu'elle sera détectable par les capteurs physico-chimiques.

## IV.7 Conclusion

Un tel système de décision trouve toute son importance dans le cadre de situations exceptionnelles. En effet, la pré-alerte à une pollution imminente, permet l'arrêt du pompage sur le cours d'eau pollué et le basculement vers le bassin de sécurité. Cela permet d'éviter un grand nombre de désagréments pour l'agglomération nancéenne, on peut citer :

- L'éventuelle intoxication de la population, ayant des effets sanitaires à grande échelle,
- Le surdosage d'additifs dans les usines de traitement/distribution d'eau, rendant l'eau de moindre qualité,
- La purge des conduites d'eau contaminées, laborieuse et très coûteuse,
- L'arrêt de la distribution d'eau qui cause des désagréments autant auprès de la population qu'auprès des industriels et services sanitaires tel que les hôpitaux.

A titre d'exemple, nous pouvons faire le parallèle en observant à petite échelle ce que peuvent endurer certains villages, lors de pollutions de leur nappe phréatique par des engrais nitrés ou insecticides.

En mesurant les risques d'une pollution, et en constatant l'impact désastreux que cela implique, on se rend bien compte de la prise de conscience à surveiller la qualité de l'eau. A ce titre, il suffit d'une fois pour que l'amortissement d'un tel système soit justifié, par rapport aux énormes coûts et désagréments engendrés par une pollution détectée après contamination du réseau d'eau potable.

Il faut aussi retenir que la modélisation de la décision, sous forme de modèle d'amélioration continu nécessitant l'intégration d'une connaissance experte, atténue les considérations qui la qualifient de vague, imprécise, incertaine et subjective.

L'autre aspect à prendre en compte pour un temps de réponse correct, est la situation géographique du lieu de pompage par rapport à la station d'alerte, ainsi que l'éloignement par rapport à l'usine de traitement. Cela permet d'évaluer la marge de sécurité à avoir lors d'une pollution en cours de transfert vers l'usine de traitement. L'accroissement de cette marge passe aussi par une diminution de la résolution de rapatriement de données qui peut descendre à l'échelle de la seconde sur l'ensemble des capteurs retenus. En effet, au cours l'étude, cette résolution se situait à l'ordre de la minute, du fait que certains capteurs ne pouvaient descendre au-delà.

## Références bibliographiques

- [ADP-2001] Aéroports de Paris, Analyse Multicritères, rapport interne 2001.
- [BEN-2000] Ben Mena S., Introduction aux méthodes multicritères d'aide à la décision, *Biotechnologie, Agronomie, Société et Environnement* 2000 4 (2), pages 83–93, février 2000.
- [BRAN-1986] Brans J.P, Vincke Ph., Marschal B., How to select and how to rank projects : The PROMETHEE method, *European journal of Op. Research*, volume 24, pages 228-238, 1986.
- [BOUY-1993] Bouyssou D. et ROY B. : "Decision-aid: an elementary introduction with emphasis on multiple criteria", *Investigación Operativa*, 2, 95–110, 1991 (with B. Roy) also in *Information Science and Technology*, 2, 109-123, 1993.
- [FLE – 2001] Fleury G., *Analyse spectrale: Méthodes non-paramétriques et paramétriques*, Edition Ellipses 2001.
- [JAMBU - 2003] Jambu M., *Méthodes de base de l'analyse des données 3<sup>ème</sup> Edition*, Eyrolles 1999.
- [LABB-2001] Labbé R., *Méthodes d'aide à la décision : Approche théorique et étude de cas développées*, Edition Ellipses 2001.
- [LAAR-2000] Laaribi A., *SIG et analyse multicritère*, Editions Hermes Science, 2000.
- [MART-1988] Martel J.M., *Aide Multicritère à la Décision*, Document Interne, Département d'opérations et Systèmes de Décision, Université Laval, septembre 1988.
- [MART-1999] *L'aide Multicritère à la décision : Méthodes et Applications*, Communication dans : *The 41st National Conference of the Canadian Operational Research Society (CORS) JUNE 7-9, 1999 – WINDSOR, ONTARIO*.
- [MAYE-1994] Maystre LY., Pictet J., Simos J. *Méthodes multicritères Electre. Description, conseils pratiques et cas d'application à la gestion environnementale*. Lausanne, Suisse : Presses polytechniques 1994.
- [MILL-1995] Million O., *Analyse multicritère; Rapport bibliographique de DEA en Production Automatisée*, soutenu le 12/07/1995 à l'Université Henry Poincaré.
- [MUCK-2003] Muckenhirm P., *Le système d'information décisionnel : Construction et exploitation*, Lavoisier 2003.
- [PICT-2000] Pictet J. du Bureau d'aide à la décision, Lausanne 2000, <http://www.aide-decision.ch/>

- [ROY-1985] Roy B., Méthodologie multicritère d'aide à la décision, Economica 1985.
- [ROY-1993] Roy B., Bouyssou D., Aide multicritère à la décision : méthode et cas, Economica 1993.
- [VINC-1989] Vinck Ph., L'aide multicritère à la décision, Editions de l'Université de Bruxelles, 1989.

## ***Références Internet***

{1} Site de l'Association Française Edwards Deming :

<http://www.fr-deming.org/afed-F12.pdf>

{2} Site du Laboratoire d'Analyse et Modélisation de Systèmes pour l'Aide à la Décision :

<http://l1.lamsade.dauphine.fr>

## Conclusions et perspectives

La surveillance d'une ressource d'eau à potabiliser, sujette à d'innombrables pollutions, se fait en fonction de la qualité et la disponibilité des informations représentant le cours d'eau. Celles-ci doivent être générées de manières fiables et rapides permettant une prise de décision objective, sûre et quasi-immédiate de l'opérateur chargé de la surveillance en cas de pollution avérée.

Ce travail de recherche ayant consisté à produire ce type d'informations, s'est déroulé dans un cadre où certains choix matériels étaient déjà en place sans nécessairement correspondre à l'utilisation scientifique souhaitée. En effet, sur la phase préliminaire dans un cadre dépourvu de contraintes matérielles et temporelles, le choix des capteurs devait se faire après une étude de sensibilité. D'autre part, le manque de fiabilité de certains capteurs prototypes ou commerciaux, lors des tests de toxicité, a produit des résultats clairsemés, d'où le souhait de reproduire d'autres campagnes de tests dans de bonnes conditions de fonctionnement. Ces contraintes nous ont, d'autre part, mené à revoir le choix initial des capteurs.

En effet, après avoir analysé les différents capteurs présélectionnés en amont de ce travail, nous avons éliminé pour des raisons pratiques certains capteurs dont nous avons constaté des limites, d'ordre temporel et de maintenance contraignante. Nous avons, à cet effet, retenu les paramètres dont la mesure pouvait se révéler fiable, automatisable, instructive et rapide. Cela permet d'assurer une autonomie relative pour une station nécessitant un minimum de maintenance pour un maximum de fonctionnement, et cela dans le but de minimiser l'intervention humaine tout en assurant la continuité de la mesure.

Les capteurs retenus sont :

- les capteurs des paramètres physico-chimiques : pH, conductivité, oxygène dissous, température, turbidité, chlorophylle, potentiel redox. La fiabilité de ces capteurs a permis de fournir des mesures en continu, respectant ainsi le critère d'autonomie souhaité.
- les capteurs des paramètres dits « biologiques » : le Fluotox basé sur la mesure de l'activité algale, et le Gymnotox basé la mesure des signaux électriques de poissons. Ces capteurs prototypes au début de l'étude, ont évolué tout au long de celle-ci, en tenant compte des contraintes du terrain pour aboutir à des produits commerciaux. Ces capteurs compensent relativement bien le rejet des autres capteurs par leur large couverture des risques toxiques ainsi qu'une réactivité assez précoce. Il n'en demeure pas moins que l'intégration des versions améliorées des capteurs exclus que sont le Dresseina basé sur l'activité valvaire des moules et le Bacteria-test basé sur la luminescence de bactéries contribueront fortement à l'élargissement du champ de détection. Ces capteurs n'ont certes pas évolué tout au long de l'étude, mais les remarques apportées ont permis ultérieurement de contribuer à l'amélioration de ces prototypes. Cette étude fournit une base de validation pour les capteurs prototypes retenus et une contribution à l'amélioration des autres capteurs qui valaient la peine d'être améliorés.

Le choix pragmatique de retenir ces capteurs permet de s'assurer d'un niveau de réponse suffisant pour mettre à l'épreuve le système de surveillance dans des conditions d'exploitation réelle. La station d'alerte multi-capteurs mise en place sur la base de mesures expérimentales doit se réajuster sur la base des mesures réelles. Sa mise à l'épreuve dans les conditions d'exploitation réelle recalera les données expérimentales et confortera son évolutivité. Cette dernière passera, d'une part, par la prise en compte d'événements accidentels permettant de dépasser le rôle consultatif et complémentaire qu'a actuellement la station, et d'autre part, par le rajout de capteurs respectant les contraintes de temps de réponse rapide et autonomie de fonctionnement.

En terme d'analyse de l'information, nous avons mis en place deux outils prenant en compte le milieu non déterministe que représente le milieu naturel des eaux de surface. La fusion de données, a consisté à mettre en place un processus d'interprétation de l'information. Celui-ci basé sur la réaction d'un groupe de capteurs, permet de **qualifier** une pollution. A titre d'exemple, dans le cas d'une pollution agricole, le Fluotox, le Gymnotox, le turbidimètre, le capteur Chlorophylle et le capteur conductivité réagissent significativement pour former le profil de pollution agricole. Le profil de capteurs ayant réagi est appelé masque de pollutions. Celui-ci est enregistré dans la base de connaissance du dispositif, permettant ainsi d'être confronté en cas d'événement similaire.

D'autre part l'aide à la décision, nous a permis de **quantifier** l'information issue de la fusion de données selon une échelle d'alerte. Cette échelle est subdivisée selon trois niveaux d'alerte:

- le niveau maximal est indicateur de l'action d'arrêt du pompage sur le cours d'eau et basculement sur un bassin de sécurité,
- le niveau intermédiaire correspond à l'action d'augmentation du dosage des additifs de prévention anti-pollution,
- et le niveau bas correspondant au maintien des dosages préventifs.

L'information issue, permet de faire correspondre une action adéquate à la pré-alerte quantifiée. Le type d'action à entreprendre dépend du calage de l'échelle établie à cet effet. Une amélioration pourrait être apportée à ce dispositif, consistant à établir une proportionnalité entre le degré de la pollution et la quantité d'additifs à apporter au pré-traitement de l'eau.

La complémentarité de ces deux outils permet de préserver une cohérence dans l'aide à l'interprétation des données et tient compte des exigences temporelles d'une alerte imprévisible et rare, mais dont les effets pourraient être dévastateurs.

Il en ressort que l'heureuse rareté des pollutions nécessitera sûrement de nouvelles campagnes de tests de toxicité, permettant ainsi de conforter et d'affiner les choix déjà entrepris, ainsi qu'un enrichissement de la base de connaissance par la simulation d'autres types de risques. L'avantage de telles campagnes en amont de l'étude, est d'avoir un système déjà au point, pouvant être mis à l'épreuve dans des conditions réelles d'exploitation et de décision.

En se basant sur les échéances fixées par la directive européenne du 23 octobre 2000, nous constatons que l'échéance de 2006 est bien respectée par la construction d'un réseau de surveillance. Ceci étant, notre dispositif doit s'affiner pour se conformer aux objectifs qualitatifs et quantitatifs qui seront fixés pour 2009, tout ceci en perspective d'être opérationnel en 2012 afin d'atteindre les objectifs de bon état de la ressource en 2015.



Enfin, pour que notre système soit réutilisable à moindre coût, sur d'autres sites demandeurs, il faudra tout d'abord établir une étude de vulnérabilité du milieu à surveiller. Cela identifiera le profil de pollution à surveiller, afin d'effectuer la campagne de tests de toxicité adéquate aux risques appréhendés. Celle-ci permettra de dégager l'information nécessaire pour l'initialisation des paramètres de seuillages qui diffèreraient en aval du même cours d'eau de part les éventuels affluents, rejets d'usines et autres paramètres.

## Annexe 0 : Présentation du pôle de l'eau

Nancy étant située en fond de vallée, a été confrontée durant son histoire à de fortes inondations, notamment durant les années 70, où les réflexions et études ont abouti à la construction d'importants ouvrages de rétentions, améliorations des réseaux hydrauliques et l'amélioration des connaissances dans le domaine de la mesure. Cela a donné un champ d'expérimentations dans lequel les chercheurs pouvaient s'impliquer sur un pôle unique à l'échelle de l'agglomération nancéenne. Ce pôle rassemble aujourd'hui plus de 300 chercheurs, ingénieurs et techniciens répartis sur des structures d'études, d'analyse, de recherche, d'expertise et d'assistance technique privée ou publique. Tout cela constitue ainsi le pôle de l'eau créé en 1984 dont l'animateur est le NANCIE (Centre International de l'Eau de Nancy). Cette association a été fondée en même temps que le pôle de l'eau, dans ce contexte, par la CUGN (Communauté Urbaine du Grand Nancy). L'objet de cette association est de promouvoir les activités de recherche dans les domaines de l'eau et de l'assainissement, le développement de l'innovation, la coopération internationale, le transfert de technologies et la formation.

Le NANCIE est également le lien privilégié entre les industriels et les chercheurs. Il joue ce rôle de médiateur en tant que centre d'expertise et de coordination, aux problématiques liées au cycle de l'eau et créer les liens entre le monde scientifique, les pouvoirs publics et l'industrie.

L'eau, constituant un domaine de recherche, fait qu'interviennent microbiologistes, physiciens, chimistes, hydrauliciens, automaticiens et informaticiens. Le NANCIE dans le cadre de l'animation du pôle de l'eau, contribue à fédérer ces différentes disciplines dans le cadre de projets nationaux et internationaux en mettant en œuvre un ensemble d'équipes.

Dans le contexte de l'étude menée, le NANCIE avait un besoin d'expertise scientifique spécialisée dans le domaine des capteurs et du traitement du signal. Ainsi est née la collaboration avec mon laboratoire de recherche qu'est le CRAN (Centre de Recherche Automatique de Nancy) spécialisé dans le domaine. Cela s'est concrétisé par une Convention Industrielle de Formation par la REcherche (CIFRE), financée par l'Association Nationale de la Recherche Technique (ANRT) et le NANCIE.

L'objectif de cette collaboration entre le NANCIE et le CRAN a consisté à mettre en place matériellement et méthodologiquement un système de surveillance de la qualité de l'eau, matérialisé par une station d'alerte à la pollution des eaux de surface.

## Annexe I : Détail du dispositif de la station

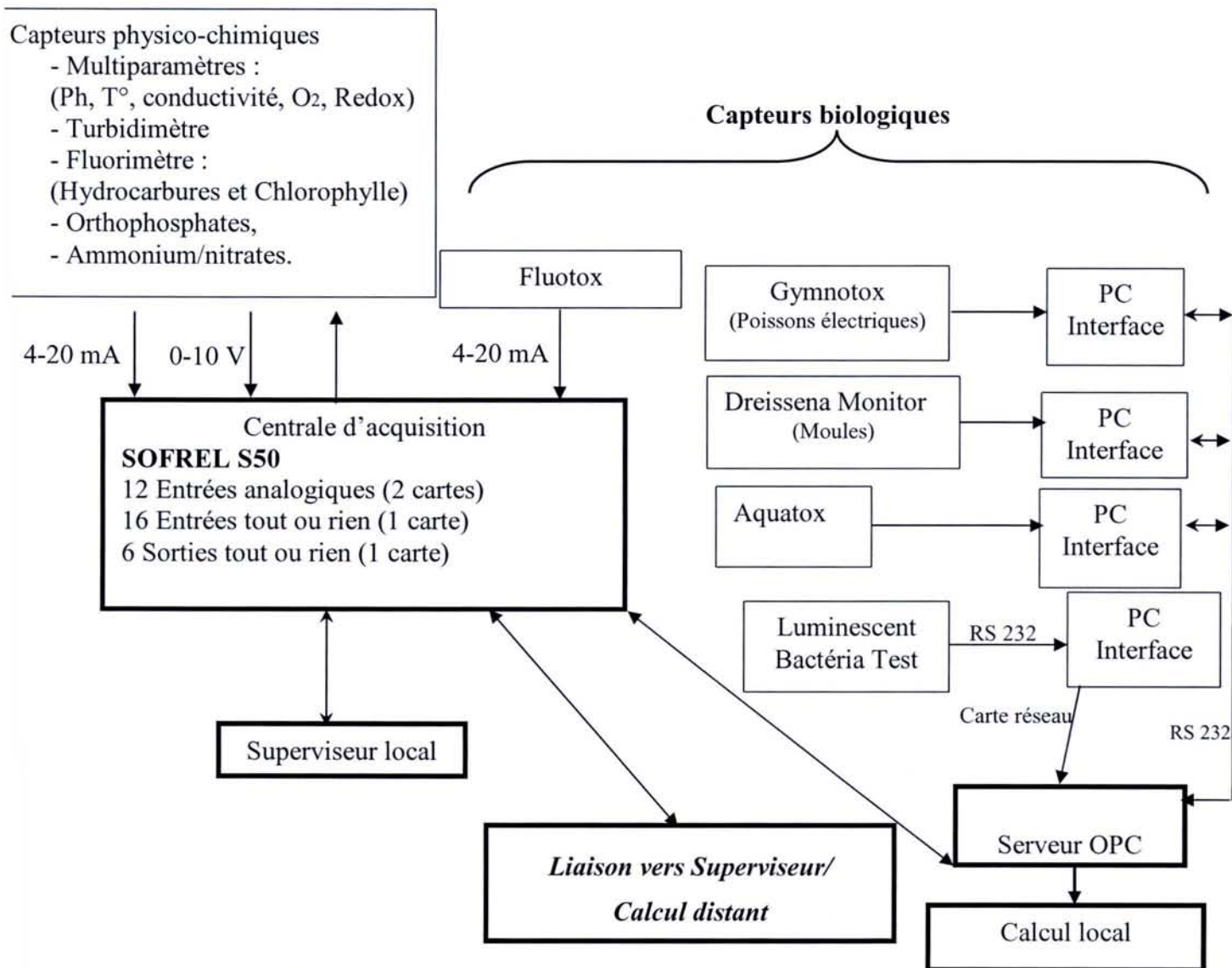


Figure A1.1 : « Détail fonctionnel de la station multi-capteurs »

### Éléments du système de transmission

#### 1 La centrale d'acquisition "S50"

Comme indiqué sur la figure ci-dessus la centrale d'acquisition S50, produit de la société SOFREL, est l'élément clef du dispositif, toutes les données y transitent soit en étant rapatriées soit pour y être redistribuées. Sa constitution matérielle se compose de :

- 12 Entrées analogiques (2 cartes)
- 16 Entrées tout ou rien (1 carte)
- 6 Sorties tout ou rien (1 carte).

La configuration du conditionnement des données ainsi que la détermination de résolution de rapatriement (période de scrutation) se fait grâce à une interface Minitel standardisé distante PCWIN (outils de télégestion pour poste central). Les données sont d'une part rapatriées des capteurs pour y être redistribuées après conditionnement vers les postes locaux ou distants de supervision, de calculs et d'archivages.

Il est à noter, qu'il est possible de programmer des alertes en fonction de critères de seuillage, pour les signaler sur GSM.

## **2 Système de rapatriement d'archivage de données ( Le serveur OPC)**

Ce système représente un PC interface permettant de rendre compatible les données issues des capteurs biologiques vers le S50, à l'exception du Fluotox qui possède une liaison classique 4-20 mA directement compatible avec le S50. Il est équipé du standard de communication OPC (OLE for Process Control) qui permet d'effectuer les tâches suivantes :

- Scrutation permanente des ports d'E/S (carte PCISER4 permettant de gérer 4 ports séries). Chaque analyseur local est connecté à un port série indépendant. Cas particulier pour le BACTERIA test, connecté au PC interface à l'aide d'une carte réseau, permettant au PC interface de scruter les données sur un répertoire partagé.
- Récupération, construction et décodage des chaînes de caractères en provenance des appareils.
- Sauvegarde des données à conserver (paramétrage utilisateur) dans des fichiers ASCII.

L'archivage de toutes les données se faisant sur le serveur OPC, il est relié d'une part en analogique au S50 et d'autre part par un câble réseau au système de calcul local.

## **3 Système de calcul local**

Ce système sous forme de PC de traitement de données est directement relié au serveur OPC, et reçoit les données de tous les capteurs. Un dispositif de calcul local y est installé, il joue un rôle local pour un opérateur désirant vérifier les paramétrages aux différents niveaux du système d'alerte, comme ça été le cas lors de la phase de mise en place. Il est délocalisé dans la configuration actuelle au niveau de la GTC (Gestion Technique Centralisée) dont le descriptif est détaillé au point suivant. Cette délocalisation est due au maintien d'un minimum de maintenance sur la station devant être la plus autonome possible.

## **Nomenclature des signaux de mesure sur la centrale d'acquisition**

Il s'agit de faire le bilan des divers signaux en entrées et sorties de chaque instrument de mesure. En effet, la multiplicité de types capteurs a fait que l'on a dû adapter dans certains cas des interfaces pour faire circuler les signaux de mesure vers la centrale d'acquisition qui est régie par les signaux suivants :

- **Boucle de courant 4-20mA** : son mode de fonctionnement est illustré par l'exemple suivant :
  - Un capteur de niveau piézométrique à sortie 4 à 20mA (intervalle de 16mA) a une gamme de mesure de 0 à 10 mètres de colonne d'eau, soit une pente de 62.5cm/mA (1000cm/16mA).
  - La centrale d'acquisition possède un convertisseur analogique/numérique sur 12 bits signés c.à.d qu'il donne 0 pour -20mA et 4096 ( $2^{12}$ ) pour +20mA soit une résolution de 0.009766 mA (40/4096).

- La résolution du mesurage enregistrée est donc de :  $(40/4096)*62.5 = 0.61\text{cm}$ , valeur dont il faut tenir compte dans les incertitudes de mesure.
- **Différence de potentiel** illustré par l'exemple suivant :
  - Un capteur spectrophotométrique à sortie 0 à 1 volt pour une gamme de mesure d'absorption de 0 à 100%, soit une pente de 100 % / Volt.
  - La centrale d'acquisition possède un convertisseur analogique/numérique -10 +10 volts sur 12 bits signés, soit une résolution de 20/4096 volts
  - La résolution du mesurage enregistrée est donc de :  $(20/4096)*100\% = 0.5 \%$ , valeur à vérifier par rapport aux besoins.
- **Impulsions tout ou rien** : Dans le cas de débitmètres ou de pluviomètres, ces instruments peuvent transmettre des impulsions de courtes durées à chaque incrément de volume ou basculement d'auget, à ce titre la centrale d'acquisition est équipée d'entrées de comptage.

### Aquatox :

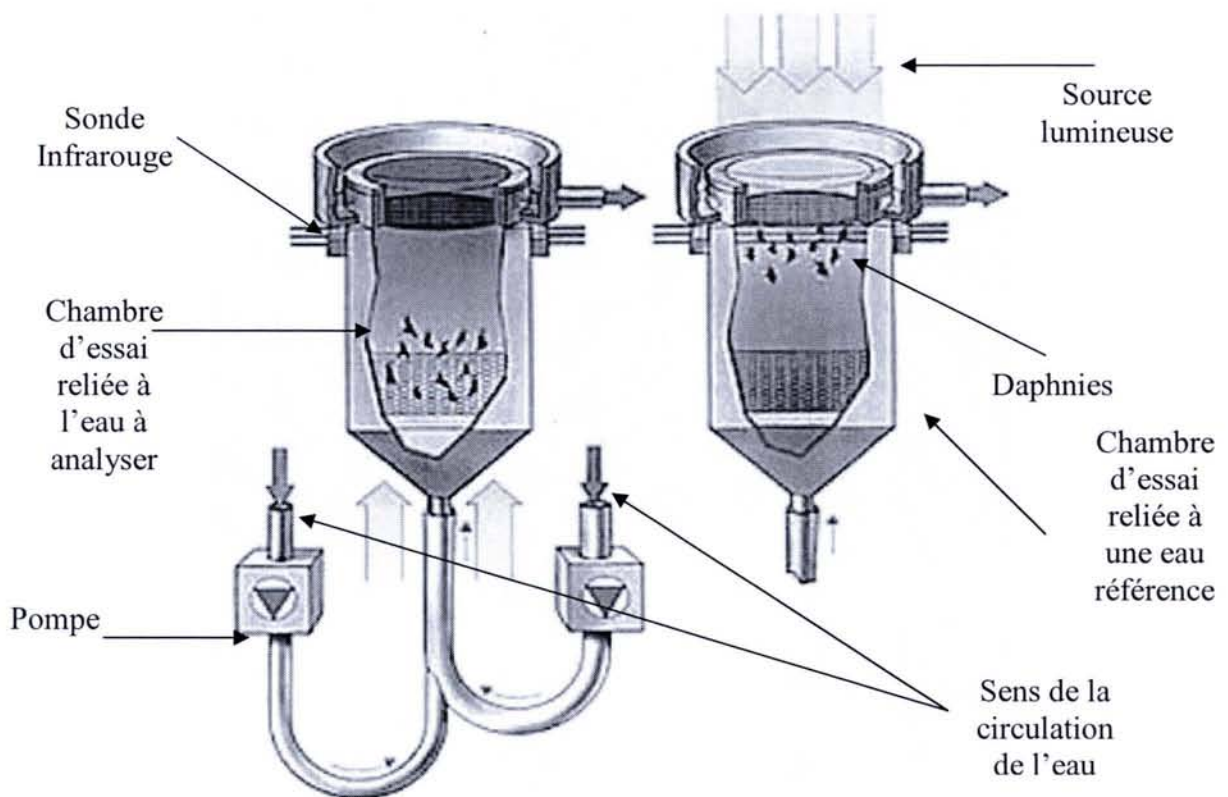


Figure A1.2 : « Détail du fonctionnement de l'Aquatox »

## Bactéria test:

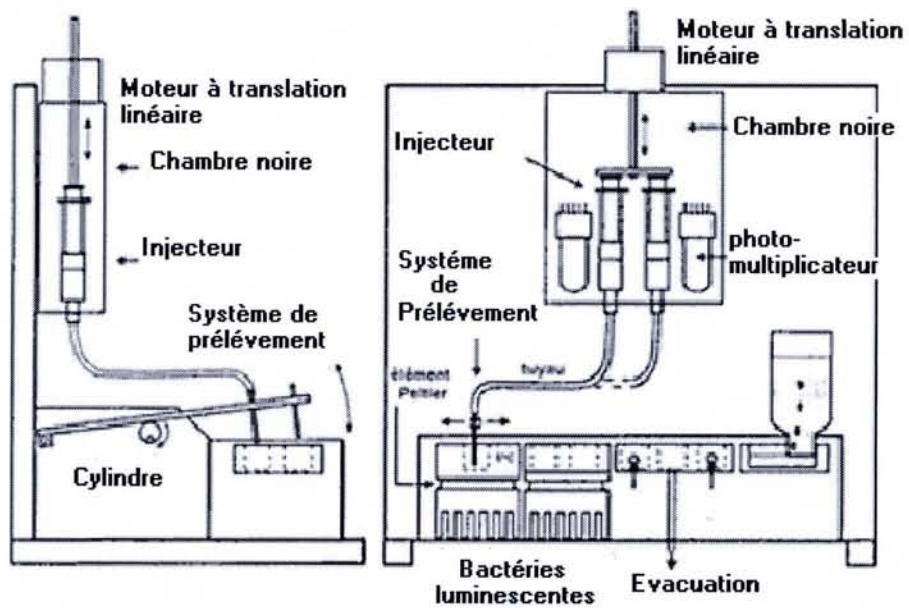


Figure A1.3 : « Schéma de principe du Bacteria-Test »

## Annexe II : Récapitulatif des grandeurs

La période de relevé des données sur le superviseur est de 5 minutes. Ces périodes diffèrent de la période d'échantillonnage du capteur. Le tableau suivant récapitule les différentes grandeurs :

Grandeur ( résolution et précision)	Gammes	Période. d'éch.	Sorties
<b>Capteurs Physico-Chimiques</b>			
<b>1- Sonde multiparamètres</b>			
PH (0,1)	2,00 à 12,00		4 – 20 mA
°C ( 1°C)	0°C à 50°C		4 – 20 mA
Conductivité (( $\mu$ S/cm)salinité)	0-1999 $\mu$ S/cm		4 – 20 mA
Oxygène dissous (0,1 mg O <sub>2</sub> /l)	0-60 mg/L		4 – 20 mA
<b>Potentiel Redox ( 100mV)A vérifier</b>	<b>0mV à 400mV</b>		<b>4 – 20 mA</b>
<b>2- Turbidimètre</b>			
FTU (0,1 Unité de turbidité)	0-500		4 – 20 mA
<b>3- Fluorimètre</b>			
Hydrocarbures (0,01ppm)	0-2500 (à vérifier)		0-10V
Chlorophile a (0,01 $\mu$ g/l)	0-100 $\mu$ g/l		0-10 V
<b>4- Analyseur Nitrate/Ammonium</b>			
Ammonium (0,01mg NH <sub>4</sub> /l)	0.1-100mg/l		4 – 20 mA
Nitrate(0,01mg NO <sub>3</sub> /l)	0.5-100mg/l		4 – 20 mA
<b>Capteurs Biologiques</b>			
1- Le Bactéria-test			

Inhibition des bactéries (%)	1-100		
<b>2- Le Fluotox</b>			
Chlorophylle (Inhibition au % $\mu$ g/l)	1-100	1s	4-20 mA
<b>3- L'Aquatox</b>			
% de passages de daphnies (0,1%)	1-100	10 enregistrements /heure	RS-232
<b>4- Le Dreissena</b>			
			RS-232
% Moules ouvertes (0,1%)	0 à 100	5mn (à vérifier)	
Mouvements valvaires (moyenne glissante - 0,1)	0-1	Moyenne sur 1h	
Moules exclues	0 à 42	Moyenne sur 1h	
<b>5- Le Gymnotox</b>			
Fréquence électriques de 03 poissons (1 Hertz)	750 Hz à 1250 Hz	10 secondes	RS-232

Il est noter que les grandeurs biologiques, sont directement interprétées par les interfaces logicielles auxquelles les capteurs y sont associés.



# **Annexe III : Extraits du décret n° 89-3 du 3 janvier 1989 (amendé) relatif aux eaux destinées à la consommation humaine à l'exclusion des eaux minérales naturelles**

## **Annexe I-1 : Limites de qualité des eaux destinées à la consommation humaine**

(Décret n° 90-330 du 10 avril 1990, article 10-I et II et Décret n° 95-363 du 5 avril 1995, article 12)

### A. Paramètres organoleptiques

L'eau ne doit pas présenter :

1. Une coloration dépassant 15 mg/l de platine en référence à l'échelle platine/cobalt.
2. Une turbidité supérieure à une valeur équivalente à 2 unités Jackson.
3. D'odeur, de saveur, pour un taux de dilution de 2, à 12 °C et de 3, à 25 °C.

### B. Paramètres physico-chimiques en relation avec la structure naturelle des eaux

1. A l'exception des eaux ayant subi un traitement thermique pour la production d'eau chaude, la température ne doit pas dépasser 25 °C.
2. Le pH doit être supérieur ou égal à 6,5 unités pH et inférieur ou égal à 9 unités pH; cette obligation ne s'applique pas aux eaux conditionnées non minérales.
3. Pour les substances suivantes, les valeurs des concentrations doivent être inférieures ou égales aux valeurs indiquées ci-après :

Chlorures : "200 mg/l (Cl)"

Sulfates : 250 mg/l (SO<sub>4</sub>)

Magnésium : 50 mg/l (Mg)

Sodium : 150 mg/l (Na) avec un percentile de 80

Potassium : 12 mg/l (K)

Aluminium total : 0,2 mg/l (Al)

(A l'exception des eaux ayant subi un traitement thermique pour la production d'eau chaude, pour lesquelles la valeur de 0,5 mg/l [Al] ne doit pas être dépassée.)

4. La quantité de résidus secs, après dessiccation à 180 °C, doit être inférieure ou égale à 1 500 mg/l.

### C. Paramètres concernant des substances indésirables

1. Pour les substances suivantes, les valeurs des concentrations doivent être inférieures ou égales aux valeurs indiquées ci-après :

Nitrates : 50 mg/l (NO<sub>3</sub>).

Nitrites : 0,1 mg/l (NO<sub>2</sub>).

Ammonium : 0,5 mg/l (NH<sub>4</sub>).

Azote Kjeldahl : "1 mg/l (en N)", N de NO<sub>3</sub> et NO<sub>2</sub> exclus.

2. L'oxydabilité au permanganate de potassium (KMnO<sub>4</sub>), mesurée après 10 minutes en milieu acide, à chaud, doit être inférieure ou égale à 5 mg/l en oxygène.

3. La teneur en hydrogène sulfuré doit être telle que ce composé ne soit pas détectable organoleptiquement.

4. La valeur de la concentration en hydrocarbures dissous ou émulsionnés, après extraction au CCl<sub>4</sub>, doit être inférieure à 10 microgrammes par litre.

5. La teneur en phénols doit être telle que les composés ne soient pas détectables organoleptiquement après ajout de chlore. En cas de détection, la concentration en phénols, exprimés en indice phénol C<sub>6</sub>H<sub>5</sub>OH, doit être inférieure ou égale à 0,5 µg/l, les phénols naturels ne réagissant pas au chlore étant exclus.

6. Pour les substances suivantes, les valeurs de concentrations doivent être inférieures ou égales aux valeurs indiquées ci-après :

Agents de surface réagissant au bleu de méthylène : 200 µg/l (exprimés en lauryl-sulfate).

Fer : 200 µg/l (Fe).

Manganèse : 50 µg/l (Mn).

Cuivre : 1 mg/l (Cu).

Zinc : 5 mg/l (Zn).

Phosphore : 5 mg/l (P<sub>2</sub>O<sub>5</sub>).

Argent : 10 µg/l (Ag).

"7. La teneur en fluor doit être inférieure à 1 500 microgrammes par litre (F) pour une température moyenne de l'aire géographique considérée comprise entre 8 °C et 12 °C et à 700 microgrammes par litre (F) pour une température moyenne de l'aire géographique considérée comprise entre 25 °C et 30 °C. Pour les températures moyennes comprises entre 12 °C et 25 °C, la teneur limite en fluor est calculée par interpolation linéaire."

#### D. Paramètres concernant des substances toxiques

Pour les substances suivantes, les valeurs des concentrations doivent être inférieures ou égales aux valeurs indiquées :

Arsenic : 50 µg/l (As).

Cadmium : 5 µg/l (Cd).

Cyanures : 50 µg/l (CN).

Chrome total : 50 µg/l (Cr).

Mercure : 1 µg/l (Hg).

Nickel : 50 µg/l (Ni).

Plomb : 50 µg/l (Pb).

Antimoine : 10 µg/l (Sb).

Sélénium : 10 µg/l (Se).

Hydrocarbures polycycliques aromatiques (H.P.A.) :

Pour le total des six substances suivantes : 0,2 µg/l.

Fluoranthène,

Benzo (3,4) fluoranthène,

Benzo (11,12) fluoranthène,

Benzo (3,4) pyrène,  
Benzo (1,12) pérylène,  
Indéno (1,2,3-cd) pyrène,  
Benzo (3,4) pyrène 0,01 µg/l.

#### E. Paramètres microbiologiques

1. L'eau ne doit pas contenir d'organismes pathogènes, en particulier de salmonelles dans 5 litres d'eau prélevée, de staphylocoques pathogènes dans 100 millilitres d'eau prélevée, de bactériophages fécaux dans 50 millilitres d'eau prélevée et d'entérovirus dans un volume ramené à 10 litres d'eau prélevée.
2. 95 p. 100 au moins des échantillons prélevés ne doivent pas contenir de coliformes dans 100 millilitres d'eau.
3. L'eau ne doit pas contenir de coliformes thermotolérants et de streptocoques fécaux, dans 100 millilitres d'eau prélevée.
4. L'eau ne doit pas contenir plus d'une spore de bactéries anaérobies sulfite-réductrices par 20 millilitres d'eau prélevée.
5. Lorsque les eaux sont livrées sous forme conditionnée, le dénombrement des bactéries aérobies revivifiables, à 37 °C et après vingt-quatre heures, doit être inférieur ou égal à 20 par millilitre d'eau prélevée ; à 22 °C et après soixante-douze heures, il doit être inférieur ou égal à 100 par millilitre d'eau prélevée. L'analyse est commencée dans les douze heures suivant le conditionnement.
- "6. Lorsque les eaux sont livrées sous forme conditionnée, l'eau ne doit pas contenir de *Pseudomonas aeruginosa* dans 100 ml."

#### F. Pesticides et produits apparentés

"Pour les insecticides organochlorés persistants, organophosphorés et carbamates, les herbicides, les fongicides, les P.C.B. et P.C.T., les valeurs des concentrations doivent être inférieures ou égales aux valeurs indiquées ci-après :

a) Par substance individualisée : 0,1 µg/l, à l'exception des substances suivantes :

Aldrine et dieldrine : 0,03 µg/l ;

Heptachlore et epoxyde d'heptachlore : 0,03 µg/l".

b) Pour le total des substances mesurées : 0,5 µg/l".

#### G. Paramètres concernant les eaux adoucies livrées à la consommation humaine

Les eaux adoucies livrées à la consommation humaine doivent satisfaire, en outre, aux exigences suivantes :

1. La dureté totale ne doit pas être inférieure à 15 degrés français.
2. L'alcalinité ne doit pas être inférieure à 2,5 degrés français.

Ces dispositions ne sont pas applicables aux eaux adoucies ayant subi un traitement thermique pour la production d'eau chaude.

---

**Annexe I-2 : Autres références de qualité des eaux destinées à la consommation humaine**

**(Décret n° 90-330 du 10 avril 1990, article 11)**

Les dispositions du troisième alinéa de l'article 2 s'appliquent lorsque est constatée l'une des situations suivantes :

1. La composition de l'eau, notamment sa teneur en anhydride carbonique libre et sa concentration en ions hydrogène, la rend agressive au carbonate de calcium;
2. La conductivité s'écarte de 400  $\mu$ S par centimètre à 20 °C;
3. Lors de traitement de l'eau avec des produits contenant des silicates, l'augmentation de la teneur en silicates dépasse 10 mg/l (SiO<sub>2</sub>);
4. La teneur en calcium est supérieure à 100 mg/l;
5. La valeur de saturation de l'oxygène dissous est inférieure ou égale à 75 p. 100, excepté pour les eaux souterraines;
6. La concentration en carbone organique total augmente par rapport à la concentration habituelle;
7. Le résidu sec des substances extractibles au chloroforme à pH neutre est supérieure à 0,1 mg/l;
8. La teneur en bore est supérieure à 1 000  $\mu$ g/l;
9. La teneur en composés organochlorés, autres que les pesticides et les produits apparentés, est supérieure à 1  $\mu$ g/l;
10. La teneur en chlore résiduel est supérieure à 0,1 mg/l;
11. La teneur en baryum est supérieure à 100  $\mu$ g/l;
12. Le dénombrement des germes totaux pour les eaux autres que conditionnées fait apparaître plus de 10 germes par millilitre à 37 °C ou plus de 100 germes par millilitre à 22 °C. Ces valeurs sont ramenées respectivement à 2 germes et 20 germes par millilitre pour les eaux désinfectées, à la sortie des stations de traitement;
13. L'eau contient des organismes parasites, des algues, d'autres éléments figurés (animalcules) ou des matières en suspension.

**Annexe I-3 : Exigences de qualité des eaux douces superficielles utilisées ou destinées à être utilisées pour la production d'eau destinée à la consommation humaine**

(Décret n° 91-257 du 7 mars 1991, article 2-III) **G** : valeur guide et **I** : valeur limite impérative

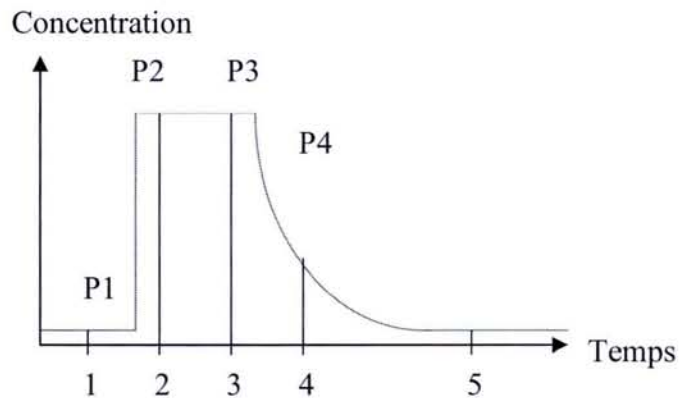
Groupes de paramètres	Paramètres	A1	A2	A3			
		G	I	G	I	G	I
Paramètres organoleptiques.	Coloration (après filtration simple) mg/l (échelle Pt)	10	20	50	100	50	200
	Odeur (facteur de dilution à 25 °C)	3		10		20	
Paramètres physico chimiques liés à la structure naturelle des eaux.	Conductivité µS/cm à 20 °C	1 000		1 000		1000	
	Température (°C)	22	25	22	25	22	25
	pH (unités pH)	6,5-8,5		5,5-9		5,5-9	
	Chlorures (mg/l Cl)	200		200		200	
	Sulfates (mg/l SO4)	150	250	150	250	150	250
	Matières en suspension (mg/l)	25					
	Demande biochimique en oxygène (DBO6) à 20 °C sans nitrification (mg/lO2)	< 3		< 5		< 7	
	Demande chimique en oxygène (DCO) (mg/l O2)					30	
Taux de saturation en oxygène dissous(% O2)	> 70		> 50		> 30		
Paramètres concernant les substances indésirables.	Nitrates (mg/l NO3)	25	50		50		50
	Ammoniaque(mg/l NH4)	0,05		1	1,5	2	4
	Azote Kjeldhal (NO3) excepté (mg/l N)	1		2		3	
	Hydrocarbures dissous ou émulsionnés après extraction par éther de pétrole (mg/l)		0,05		0,2	0,5	1
	Phénols (indice phénol) para-nitraniline 4 aminoantipyrine (mg/l)		0,001	0,001	0,005	0,01	0,1

	C6H5OH) Agents de surface réagissant au bleu de méthylène (mg/l lauryl- sulfate)	0,2		0,2		0,5	
	Substances extractibles au chloroforme (mg/l)	0,1		0,2		0,5	
	Fer dissous (mg/l Fe)	0,1		1	2	1	
	Manganèse (mg/l Mn)	0,05	0,3	0,1		1	5
	Cuivre (mg/l Cu)	0,02		0,05		1	
	Zinc (mg/l Zn)	0,5	0,05	1	5	1	
	Phosphore (mg/l P2 O5)	0,4	3	0,7		0,7	
	Fluor (mg/l F)	0,7-1		0,7-1,7		0,7-1,7	
	Bore (mg/l B)	1	1,5	1		1	
	Baryum (mg/l Ba)		0,1		1		1
Paramètres concernant les substances toxiques.	Arsenic (µg/l As)	10	50		50	50	100
	Cadmium (µg/l Cd)	1	5	1	5	1	5
	Cyanures (µg/l CN)		50		50		50
	Chrome total (µg/l Cr)		50		50		50
	Plomb (µg/l Pb)		50		50		50
	Mercure (µg/l Hg)	0,5	1	0,5	1	0,5	1
	Sélénium (µg/l Se)		10		10		10
	Hydrocarbures polycycliques aromatiques total 6 substances visées par décret du 03/01/89 (µg/l)		0,2		0,2		1
Pesticides.	Total : parathion, HCH, dieldrine (µg/l).		1		2,5		5
Paramètres microbiologiques.	Coliformes totaux 37 °C (100 ml)	50		5 000		50000	
	Coliformes thermotolérants (100 ml)	20		2 000		20000	
	Streptocoques fécaux (100 ml)	20		1 000		10000	
	Salmonelles	Absen- ce dans 5 000ml		Absence dans 1 000 ml			

## Annexe IV : Tests de pollutions et architecture du dispositif

L'ensemble des tests, est réalisé pour simuler un front de pollution tel qu'il pourrait se produire dans la nature accidentellement, la figure suivante nous donne l'allure dynamique du front de pollution injecté :

- 1 → début du test ( $t = 0$ )
- 2 → prise d'échantillon 15 minutes après le premier prélèvement
- 3 → arrêt de l'injection du polluant
- 4 → 15 minutes après l'arrêt de l'injection
- 5 → fin du test (95% du polluant a été éliminé)

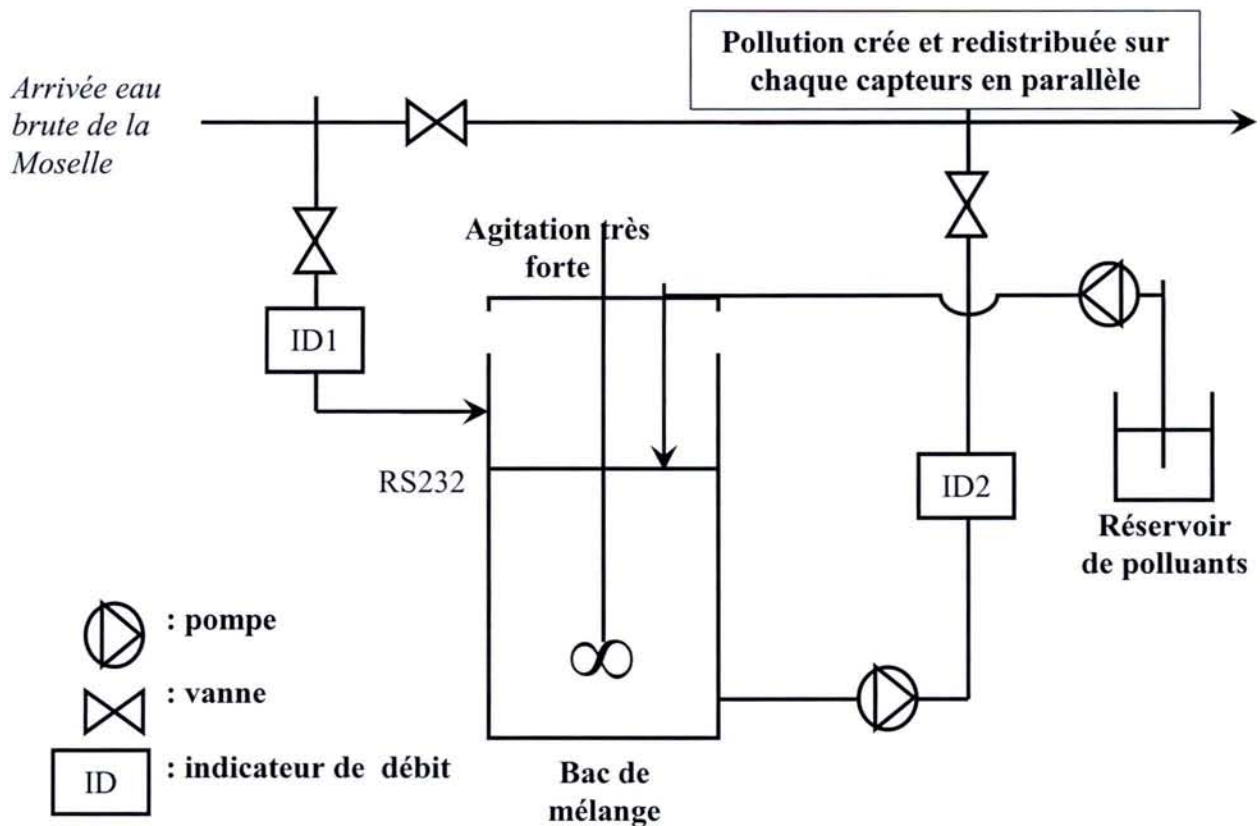


L'injection par un front de pollution a l'avantage de limiter les problèmes d'accoutumance liés aux réactifs biologiques. Afin de mettre en place ces tests, il a fallu concevoir un système de génération de pollution.

Le dispositif suivant a été monté de manière à pouvoir diffuser la pollution produite de manière uniforme et synchronisée sur chaque capteur en parallèle, la simulation de ces pollutions a nécessité l'isolation totale des rejets en utilisant le matériel suivant:

- piquage « by pass »
- bac de pollution
- système de mélangeur (palmes)
- pompe doseuse
- préleveur à mettre sur la boucle
- prévoir de la rhodamine pour test amont / aval

La figure suivante détaille la disposition des différents éléments:



« Descriptif du dispositif destiné à la réalisation des tests de pollution »

Il est à noter que lors d'un test, la totalité de la pollution qui traverse les différents pilotes est récupérée dans deux bâches disposées à l'extérieur du local afin d'être traitée

Ces déchets sont ensuite pompés pour rejoindre un filtre à sable, l'eau arrive par le haut et une hauteur d'eau de 1m40 est maintenue constante à l'aide d'un système de régulation et d'une électrovanne (lorsque la consigne fixée à 1m40 est dépassée l'eau peut s'écouler, l'électrovanne est ouverte et inversement).

L'eau filtrée est ensuite évacuée par le bas du filtre par gravité où elle rejoint un tonneau qui est équipé d'une poire de niveau branchée à une pompe à queue de cochon (lorsque la poire est en position haute la pompe se met en marche, la filière de traitement fonctionne et inversement), la pompe envoie l'eau filtrée sous pression vers les cartouches de filtration. Cette filière de traitement est composée de trois cartouches, la première est à 1 $\mu$  les deux suivantes sont à 5 $\mu$  et contiennent du charbon actif. L'intérêt d'utiliser la pompe à queue de cochon est de mettre sous pression l'eau filtrée afin que se soit le charbon actif des cartouches qui se colmate et non la membrane à 5 $\mu$ . L'eau filtrée est ensuite évacuée vers la Moselle.



## Annexe V : Exemple de calcul de concentrations de polluants sur le test « agricole 1 »

### -Pour le phosphate

Le produit utilisé pour obtenir la concentration à 2.5 mg/L de  $\text{PO}_4^{2-}$  est le Trisodium Phosphate ( $\text{Na}_3\text{PO}_4, 12\text{H}_2\text{O}$ )  $\text{MM} = 380.12 \text{ g/mol}$

### -Concentration molaire

$$C_M = C / \text{MM} (\text{PO}_4)$$

$$C_M = 2.5 \cdot 10^{-3} / (31 + 16 \cdot 4)$$

$$C_M = 2.63 \cdot 10^{-5} \text{ mol/L}$$

### -Masse de Trisodium Phosphate pour la bache V = 478 L

$$m = C_M \cdot \text{MM} \cdot V$$

$$m = 2.63 \cdot 10^{-5} \cdot 380.12 \cdot 478$$

$$m = 4.78 \text{ g}$$

Pour obtenir une concentration de 2.5 mg/L de  $\text{PO}_4^{2-}$  dans la bache il faut dissoudre 4.78 g de Trisodium Phosphate.

### -Masse de Trisodium Phosphate pour le fût V = 60L

La pollution du fût est transférée vers la bache par une pompe doseuse qui débite 50L/h, en parallèle, la bache est alimentée en eau de Moselle à compter de 3000L/h, on a donc un rapport de dilution (rd) de 60.

$$m = C_M \cdot \text{MM} \cdot V \cdot \text{rd}$$

$$m = 2.63 \cdot 10^{-5} \cdot 380.12 \cdot 60 \cdot 60$$

$$m = 35.9 \text{ g}$$

Pour obtenir la concentration voulue il faut introduire 35.9g de Trisodium Phosphate.

### -Pour le nitrate

Pour obtenir la concentration en  $\text{NO}_3^-$  de 25 mg/L, on utilise du Sodium Nitrate ( $\text{NaNO}_3$ )

$\text{MM} = 84.99 \text{ g/mol}$ .

### -Concentration molaire

$$C_M = C / \text{MM} (\text{NO}_3)$$

$$C_M = 25 \cdot 10^{-3} / (14 + 16 \cdot 3)$$

$$C_M = 4.0310^{-4} \text{ mol/L}$$

### -Masse de Sodium Nitrate pour la bache V = 478 L

$$m = C_M * MM * V$$

$$m = 4.0310^{-4} * 84.99 * 478$$

$$m = 16.4 \text{ g}$$

Pour la bache, il faut dissoudre 16.4g de Sodium Nitrate.

-Masse de Sodium Nitrate pour le fût V = 60L

$$m = C_M * MM * V * rd$$

$$m = 4.0310^{-4} * 84.99 * 60 * 60$$

$$m = 123.3\text{g}$$

Pour obtenir la concentration voulue il faut introduire 123.3 g de Sodium Nitrate.

-Pour l'atrazine

L'atrazine est une solution visqueuse, blanche, nous avons considéré une densité de 1. La concentration en atrazine de la solution est de 500 g/L, pour obtenir une concentration de 5 mg/L dans la bache il faut donc :

$$m = 5 \cdot 10^{-3} * 478$$

$$m = 2.39 \text{ g}$$

Il faut donc une quantité de 4.78g en considérant que dans 1 L de solution il y a 500g d'atrazine.

Pour le fût une quantité de 30 g est nécessaire si les conditions sont respectées (60 L dans le fût et 50L/h débit de la pompe doseuse et 3000L/h débit de l'eau de la Moselle arrivant dans la bache).

$$m = 5 \cdot 10^{-3} * 3000$$

$$m = 15 \text{ g}$$

La masse nécessaire est de 30g car la concentration en atrazine est de 500g/L

-Pour l'isoproturon

Ce sont les mêmes quantités nécessaires que dans le cas de l'atrazine, car la densité est de 1 également et la concentration en isoproturon dans la solution est de 500 g/L.

-Pour le Kaolin

Des tests ont été effectués afin de déterminer la quantité nécessaire pour obtenir la turbidité voulue lors des tests de pollution.

-1.8g de Kaolin dans 0.8L d'eau de la Moselle → turbidité > à 1000 NTU

-Dilution  $\frac{1}{2}$   $\rightarrow$  turbidité  $>$  à 1000 NTU

- Dilution  $\frac{1}{2}$   $\rightarrow$  turbidité à 600 NTU

Il y a eu dilution 4 fois ce qui correspond à 1.8g de Kaolin dans 3.2L d'eau  $\rightarrow$  turbidité à 600 NTU.

Pour la réalisation des tests on désire obtenir une turbidité de 450 à 500 NTU, pour la bâche de 478 L il faut donc :

$$m = [(478 * 1.8)/3.2] * (470/600)$$

$$m = 200g$$

Dans la bâche on prépare 2 béciers de 1.5L contenant chacun 110g de Kaolin, sans oublier d'ajuster le pH entre 9.5 et 10.5 car pendant l'ajout de Kaolin le pH diminue très fortement.

Pour le fût la masse à ajouter est plus importante (débit de l'eau de la Moselle arrivant dans la bâche 3000L/h) :

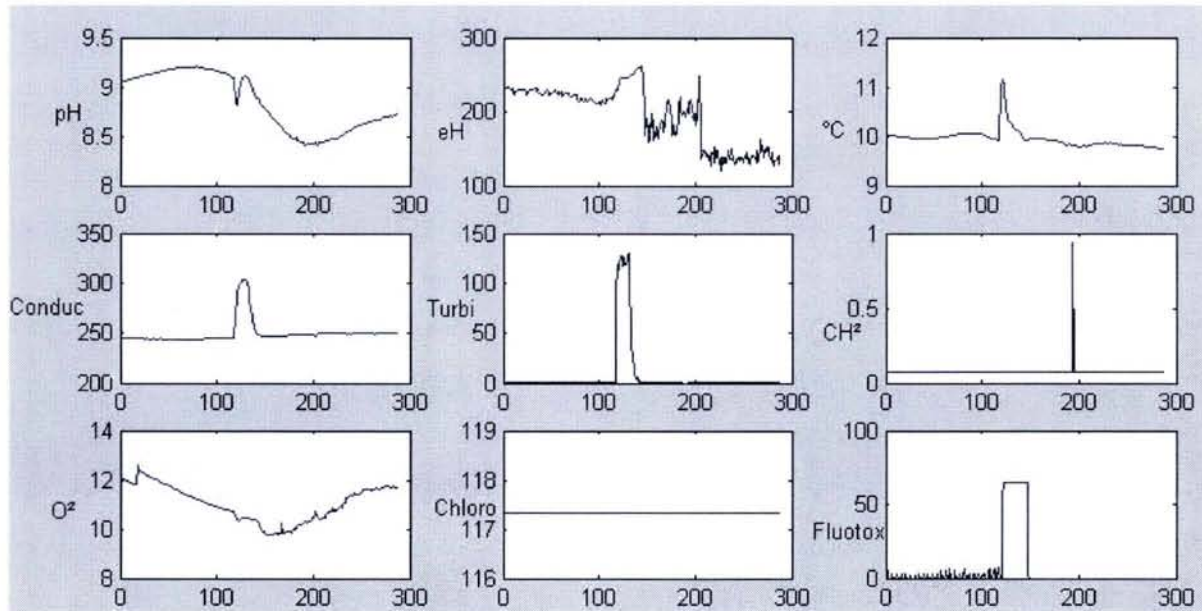
$$m = [(3000*1.8)/3.2] * (470/600)$$

$$m = 1300g$$

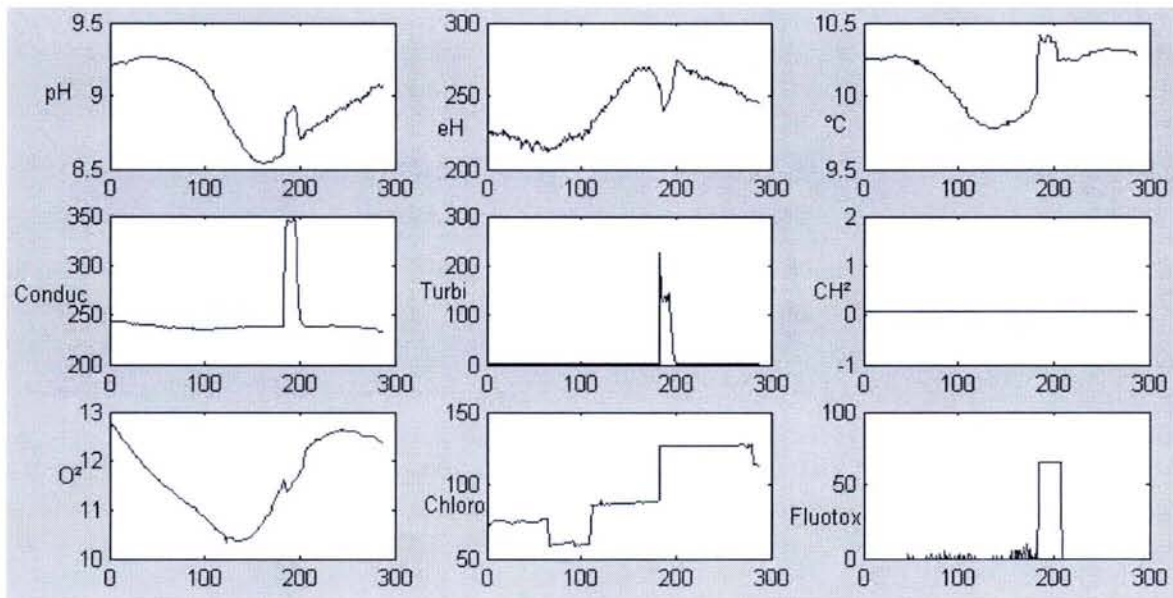
On introduit 1500g de Kaolin sous forme de 10 béciers contenant chacun 150g de Kaolin tout en ajustant le pH.

Les masses de Kaolin nécessaires pour obtenir une turbidité d'environ 470 NTU sont un peu dépassées lors de la préparation pour être sûr d'atteindre la turbidité désirée.

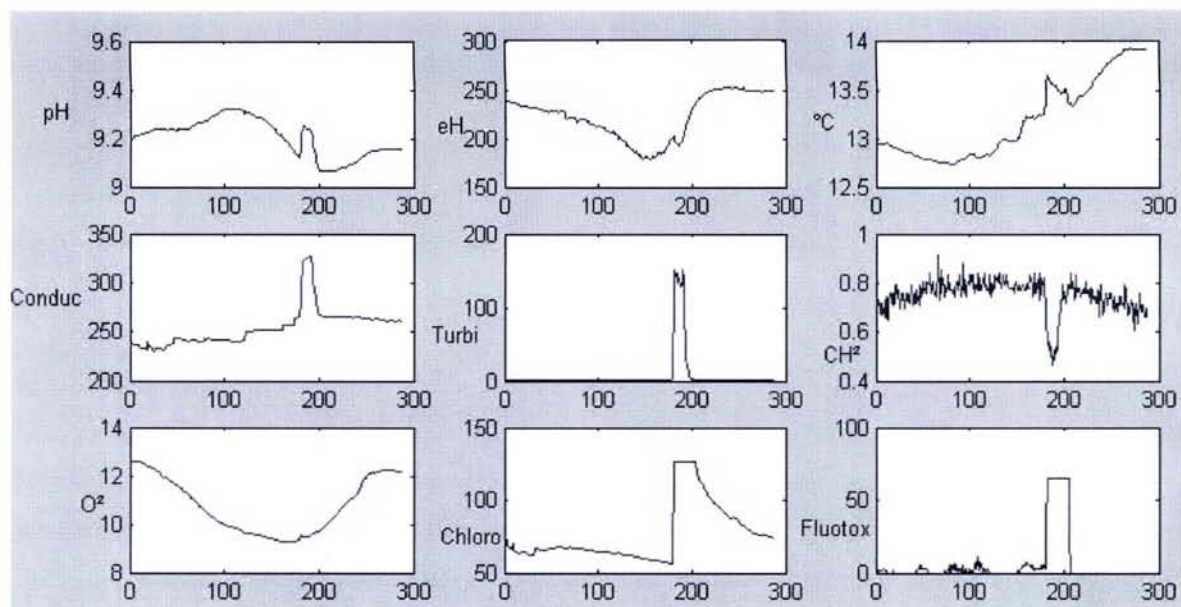
## Annexe VI : Graphiques des tests de pollutions de 9 capteurs



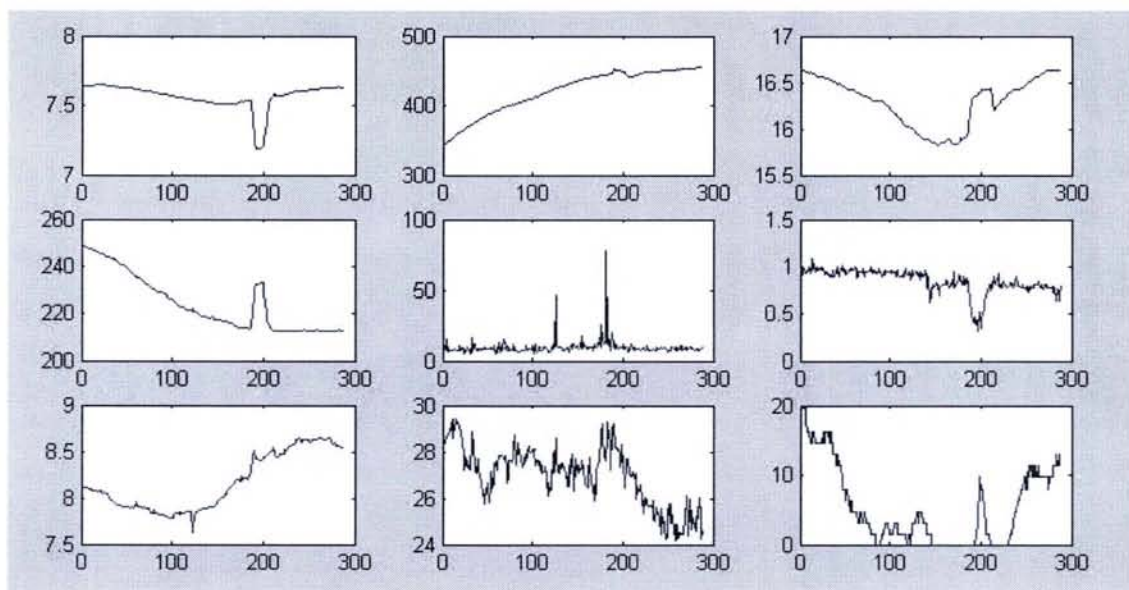
« Résultats du test Agricole1 du 16 avril 2002 »



« Résultats du test Agricole2 du 18 avril 2002 »



« Résultats du test Agricole3 du 23 avril 2002 »



« Résultats du test au Chrome 5 mg/l du 11 juin 2002 »

## Annexe VII : Observations des tests de pollutions

CATÉGORIE	Date	Type et concentration théorique	Réactions BIO					Réactions PHYSICO							
			G	D	F	B	A	ortho	ph	O2	eH	cond	turb	chloro	hydro
1	30/01/02	pH acide			X	X			X			X			
	31/01/02	pH basique	X		X	X	O		X			X			
	1/02/02	Conductivité 600µS/cm	X	!	O		O			F		X		F	
	4/02/02	Conductivité 800µS/cm		X			X			F		X		F	
	6/02/02	Turbidité 100 NTU					X	F					X		
	7/02/02	Turbidité 400 NTU	O				X	F					X		
2	11/02/02	O2 : anoxie		F		X	X		X	X	X	X			
	12/02/02	O2 : suroxy	F	F	F		X		X	X	O	O		O	
	13/02/02	PO4 5mg/L	X						X	F	F		F		
	14/02/02	PO4 10mg/L							X	X	F	F	O		F
	15/02/02	Nitrates 50mg/L											X		
	18/02/02	Nitrates 100mg/L											X		
	19/02/02	NH4+ 5 mg/L		O		O							O		
	20/02/02	NH4+ 10 mg/L		X		O							O		O
3	22/02/02	Pollution écologique	!	F	X		X		X	X	X	X		O	
	14/03/02	Atrazine 1mg/L		O	X						F	O		X	
	16/04/02	Pol° agri 1	O	X	X	X			X			X	X		
	18/04/02	Pol° agri 2	O	X	X	!	!		X	O		X	X	X	
	23/04/02	Pol° agri 3	X	X	X		F		F		F	X	X	X	F
	25/04/02	Pol° agri 4	X	X	X							X	X	X	
	30/04/02	Pol° agri 5	X	X	X	X	X		F		X	X	X	X	X
	16/05/02	Pol° agri 6		!	X				F			X	X		
	23/05/02	Pol° agri 7		X	X	!						O	X		
	28/05/02	carbofuran 1mg/L				O	X				O			O	
	6/06/02	CrVI 1mg/L			F					F	F	F	F		
	11/06/02	CrVI 5mg/L		F					X		O	O			F
	13/06/02	Pol° Industrielle		X			X		?		X	X		X	
	7/08/02	Pyralène 50mg/L					X				O		F	O	
4	29/10/02	Pol° agri comp 1	F	X	X	X	X		X	O	O	X	X	X	
	6/11/02	Pol° agri comp 2		O	X	O			O	F	O	X	X	X	
	8/11/02	Pol° agri comp 3	O	X	X				O	F	O	X	X	X	
	15/11/02	Pol° agri comp 4		O	X		F		O		O	X	X	X	
	5/12/02	Pol° agri comp 1 sans kaolin	!	!											
	12/02/03	pyralene 12,5mg/l		X		X	F			F	F		F	F	
	19/02/03	pyralene 25mg/l	!	X	O	X	X			F	F		X		
3	6/03/03	Gasoil 2,5mg/L		X		F	F		F		F	F		X	
	13/03/03	Gasoil 5mg/L		X		X	O		F		F	F		X	
	20/03/03	Cocktail gasoil/fuel		X		X	F					O		X	
	27/03/03	Acétone 0,1mg/l				O	F		F	F		F	F		

Tableau de Réactions capteurs face à des injections de polluants.

Le tableau doit être compris ainsi :

- X : signifie que le capteur a réagi de façon significative (signal au moins double)
- F : signifie qu'une réaction a été visible mais que celle-ci est faible.
- O : signifie qu'une réaction intermédiaire a été visible.
- ! : signifie que les données ne sont pas disponibles ou erronées mais que le technicien a observé une réaction ; l'importance de la réaction n'a pu être quantifiée.
- ■■■ : signifie qu'il n'a pas été possible de récupérer les données
- ■ : signifie que le capteur ne fonctionnait pas.
- : signifie que le capteur n'a pas réagi bien qu'il soit opérationnel, son comportement ne témoigne pas de présence de polluant.

Les capteurs biologiques sont référencés par :

G = Gymnotox

D = Dreissena Monitor

F = Fluotox

B = Bacteria Test

A = Aquatox

## Annexe VIII : Compléments théoriques

### Analyse de variance

L'analyse de variance (Analysis of Variance ou ANOVA), {1}, {2}, {3}, [FOW – 2000], permet de tester l'effet d'un facteur prenant un nombre fini de modalités. L'analyse de variance peut être vue comme une généralisation du *test de Student*.

Celui-ci consiste à tester l'égalité des espérances de deux échantillons en comparant la différence centrée réduite de leurs moyennes empiriques.

On souhaite tester les effets de  $k$  traitements qui ont été administrés respectivement à  $n_1, \dots, n_k$  individus. En analyse de variance, le paramètre susceptible d'influer sur les données étudiées s'appelle un facteur, et ses valeurs sont les modalités.

Dans le modèle probabiliste, chaque modalité correspond à un échantillon. Pour  $h=1, \dots, k$ , on note :  $X_1^{(h)}, \dots, X_{n_h}^{(h)}$  les variables aléatoires modélisant les données du  $h^{\text{ième}}$  groupe, supposées indépendantes et de même loi  $N(\mu_h, \sigma^2)$ . On suppose donc en particulier que la variance  $\sigma^2$  est constante, hypothèse qu'un test pourra valider.

On cherche à savoir si la variabilité observée dans les données est uniquement due au hasard, ou s'il existe effectivement des différences significatives entre les classes, imputables au facteur. Pour cela, on va comparer les variances empiriques (la variance empirique d'un échantillon est la moyenne des carrés des données centrées, sa racine carrée correspond à l'écart-type empirique) de chaque échantillon, à la variance de l'échantillon global, de taille  $n_1 + \dots + n_k = n$ . La moyenne des variances (pondérée par les effectifs) résume la variabilité à l'intérieur des classes, d'où le nom de variance intra-classes, ou variance résiduelle (la variance résiduelle, ou variance intra-classes est la moyenne pondérée des variances de chaque classe). La variance des moyennes décrit les différences entre classes qui peuvent être dues au traitement, d'où le nom de variance inter-classes, ou variance expliquée correspondant à la variance pondérée des moyennes de chaque classe.

Les tests de Fisher et *Student* servent à comparer les moyennes et les variances de deux échantillons gaussiens

Pour le modèle probabiliste, on considère deux échantillons indépendants.

- $X_1, \dots, X_{n_x}$  est un échantillon de la loi  $N(\mu_x, \sigma_x^2)$ ,
- $Y_1, \dots, Y_{n_y}$  est un échantillon de la loi  $N(\mu_y, \sigma_y^2)$ .

On note :

- $\bar{X} = \frac{1}{n_x} \sum_{i=1}^{n_x} X_i$  et  $\bar{Y} = \frac{1}{n_y} \sum_{j=1}^{n_y} Y_j$  les moyennes empiriques,



$$- S_x^2 = \frac{1}{n_x} \sum_{i=1}^{n_x} (X_i - \bar{X})^2 \text{ et } S_y^2 = \frac{1}{n_y} \sum_{j=1}^{n_y} (Y_j - \bar{Y})^2 \text{ les variances empiriques.}$$

Le résultat théorique permettant de comparer les moyennes empiriques suppose que l'on fasse l'hypothèse que les variances théoriques  $\sigma_x^2$  et  $\sigma_y^2$  sont égales. Le but du test de Fisher est de tester cette hypothèse :

$$H_0 : \sigma_x^2 = \sigma_y^2$$

## Régression linéaire simple et multiple

La régression linéaire est un cas particulier d'estimation très usité car très bien formalisé et correspondant à des modèles simples (car linéaires). C'est l'outil de base de la modélisation de données. Cela nous amène à définir brièvement les notions de bases d'estimations puis la régression simple et multiple appliquées à l'étude des paramètres de notre station.

### Régression simple

La notion de régression consiste à analyser une relation entre deux variables quantitatives et à l'exploiter pour estimer la valeur inconnue de l'une à l'aide de la valeur connue de l'autre. Plus explicitement la régression permet de résumer la relation existante entre une variable aléatoire dépendante Y et une ou plusieurs variables explicatives.

**La régression linéaire simple** correspond au cas de la régression de Y en X.

- La variable Y est appelée variable expliquée.
- La variable X est appelée variable explicative.

Il s'agit de calculer une fonction du premier degré liant les variables x et y.

La variable dépendante est toujours aléatoire la variable explicative contrôlée. La relation n'est pas univoque.

Elle est de la forme :  $\hat{y} = ax + b$

Les hypothèses :

- Homogénéité de la variance : les distributions conditionnelles de Y ont la même dispersion.
- La linéarité : la moyenne de chaque distribution de Y,  $E(Y_i/X_i)$  se situe sur une droite.
- L'indépendance des valeurs de  $Y_i$ .

Ce qui se traduit par : les erreurs sont normalement distribuées, elles sont indépendantes entre elles, et elles sont de variance constante.

Le calcul des coefficients de cette droite passe par la minimisation des carrés entre y et  $\hat{y}$ ,

$$\sum_{i=1}^n e_i^2 = \sum (y_i - \hat{y}_i)^2 \text{ qui peut aussi s'exprimer par } \sum_{i=1}^n (y_i - (ax_i + b))^2$$

le minimum d'une fonction se trouve en annulant sa dérivée première

$$S = \sum e_i^2$$

$$\frac{\partial S}{\partial b} = -2 \sum (y_i - b - ax_i) = 0$$

$$\frac{\partial S}{\partial a} = -2 \sum x_i (y_i - b - ax_i) = 0$$

ce qui conduit aux deux formulations suivantes :

$$a = \frac{S_{x,y}}{S_x^2}$$

ou la covariance de x et y,  $S_{x,y}$ , est :

$$S_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1} \text{ ou bien } S_{x,y} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n(n-1)}$$

et la variance de x :

$$S_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{(n \sum_{i=1}^n x_i^2) - (\sum_{i=1}^n x_i)^2}{n(n-1)}$$

et  $b = \bar{y} - a\bar{x}$

## Régression multi-linéaire :

### Définition mathématique

La régression multilinéaire a pour but l'étude de la relation entre :

- une variable à expliquer Y et
- p variables explicatives  $x_1, \dots, x_p$

Les variables explicatives (obtenues après étude des coefficients de corrélation) sont supposées connues. La variable Y est supposée dépendre des  $x_i$  selon le modèle suivant :

$$Y = b_0 + b_1 \cdot x_1 + \dots + b_p \cdot x_p$$

(calcul de b : estimateur des moindres carrés :  $b = (X'X)^{-1}X'Y$ )

### Estimation

A partir de données d'échantillons représentatifs, on va induire des résultats sur la population-mère (i.e. population dans laquelle les échantillons ont été prélevés).

Plus exactement, soit  $\theta$  un paramètre inconnu intervenant dans la loi de probabilité d'une variable aléatoire  $X$ . La loi de probabilité de cette variable aléatoire doit être connue

analytiquement (on choisit parmi les modèles existants la loi la plus appropriée au phénomène observé). Seule la valeur numérique du paramètre  $\theta$  intervenant dans cette loi de probabilité est inconnue.

Soient  $x_1, \dots, x_i, \dots, x_n$  les  $n$  valeurs prises par la v.a.  $X$  dans un échantillon de taille  $n$  prélevé dans la population-mère.

On appelle estimateur de  $\theta$ , que l'on note  $T_n$ , la fonction qui aux valeurs  $x_i$  de l'échantillon fait correspondre la valeur du paramètre  $\theta$ . On note la valeur numérique de cette estimation par :

$$\hat{\theta} = T_n(x_1, \dots, x_n)$$

Par définition,  $T_n$  est une fonction des réalisations d'une v.a.,  $T_n$  est donc une v.a. dont on peut chercher à déterminer les caractéristiques.

Exemple : On observe un phénomène de production de pièces manufacturées. Chaque pièce est associée à une mesure (un indicateur de qualité par exemple). Comme on ne peut pas vérifier chaque mesure, on procède à un échantillonnage qui nous fournit donc un échantillon. Supposons que la connaissance de la nature de cet indicateur nous permet de faire l'hypothèse qu'il obéit à une loi de probabilité normale. Le problème est maintenant, au vue de l'échantillon  $\{x_i\}$ , de proposer une valeur pour la moyenne de cette loi normale. Il faut procéder à une estimation du paramètre vrai  $\mu$  qui se traduit par la valeur  $\hat{\mu}$ . Il y a une infinité de manière possible parmi lesquelles on peut citer :

- $\hat{\mu} = \frac{1}{n} \sum_i x_i$
- $\hat{\mu} = \text{médiane} \{x_i\}$
- $\hat{\mu} = \text{mode} \{x_i\}$

Quel est le meilleur estimateur de la moyenne ? Existe-t-il ?

Sur ce simple exemple, est résumé le problème fondamental de l'estimation: quelle est la définition mathématique de **meilleur**?

La réponse est qu'il n'en existe pas. Alors comment comparer les estimateurs. Pour cela, on se sert de plusieurs critères, le plus souvent liés au bon sens:

- **le biais**: On souhaite que l'estimation ne soit pas systématiquement décalée par rapport à la valeur vraie.
- **la précision**: Si l'on répète l'estimation sur un autre échantillon, on souhaite obtenir une estimation cohérente, donc peu de variation d'un échantillon à l'autre. On parlera aussi d'efficacité.
- **la convergence**: Si l'on peut estimer la valeur du paramètre sur toute la population-mère, la valeur de l'estimation obtenue doit être la valeur vraie du paramètre.
- **la complexité**: Toute estimation nécessite un calcul donc un temps. On s'attachera donc à évaluer la complexité du calcul en fonction de la taille des données ( $n$ ).

- **la robustesse:** Dans tout cas concret, il existe des sources de perturbations. On souhaite que l'estimation ne soit pas sensible à la présence de valeurs aberrantes (*outliers* en anglais).

Ces différents critères ne sont pas forcément compatibles entre eux, et l'on retrouve des dilemmes classiques, précision/ robustesse, convergence /complexité.

### Estimateur convergent

Un estimateur  $T_n$  est convergent si la valeur estimée tend en probabilité vers la valeur vraie du paramètre, soit:

$$\exists n_0 \forall n > n_0 \forall \xi, \nu \text{ (arbitrairement petits)} P\left(|T_n(x_1, \dots, x_n) - \theta| < \xi\right) > 1 - \nu$$

Si l'estimation est exhaustive (l'échantillon est égal à la population-mère), alors la valeur vraie du paramètre est connue.

### Estimateur sans biais

Un estimateur  $T_n$  est dit sans biais lorsque son espérance mathématique est égale à la valeur vraie du paramètre.

$$E(T_n) = \theta$$

Un estimateur  $T_n$  est dit asymptotiquement sans biais si le biais diminue si la taille de l'échantillon augmente:

$$\lim_{n \rightarrow \infty} E(T_n) = \theta$$

### Estimateur robuste

Un estimateur est dit robuste si il est insensible à des petits écarts sur les hypothèses pour lesquelles il a été optimisé. Il y a deux sens au terme "petit": de petites variations sur toutes les données, ou des écarts importants sur un petit nombre de données. C'est le deuxième aspect qui est le plus mal pris en compte par les estimateurs classiques.

Ainsi, la robustesse traduit le plus souvent la résistance de l'estimation aux données aberrantes. On la définit mathématiquement par le plus petit nombre de données extrêmes qui modifie la valeur de l'estimation ramené à la taille de l'échantillon.

Considérons un échantillon constitué de  $n$  valeurs identiques  $a$ , auquel on ajoutera une perturbation sous la forme de valeurs extrêmes  $b \gg a$ . Pour estimer l'espérance mathématique, on peut utiliser la moyenne arithmétique qui donne bien sûr  $a$  sur l'échantillon. Cependant, cette estimation est modifiée dès l'introduction d'une nouvelle valeur,  $b$ , sa robustesse est donc de  $\frac{1}{n}$ . Par contre, la médiane de cet échantillon n'est pas modifiée si l'on ajoute une valeur extrême. En fait, la médiane ne sera modifiée que si le nombre de valeurs extrêmes est supérieur au nombre de valeurs initiales. On en déduit que la

robustesse de l'estimateur médiane est égale à  $\frac{n-1}{n}$  dont la valeur asymptotique est  $\frac{1}{2}$ .

## Estimateur efficace

La variance d'un estimateur représente sa précision. Pour tous les estimateurs (ayant même moyenne), il est possible de trouver celui dont la précision sera la meilleure, *i.e.* dont la variance sera la plus faible. On parle alors d'**estimateur à variance minimum**.

Lorsque l'on compare deux estimateurs, on dira également que  $T_n$  est plus **efficace** que  $T_n^*$  si  $V(T_n) < V(T_n^*)$ .

Une estimation est liée à un échantillon de taille finie. Si la population-mère est de taille infinie, il n'est pas possible d'avoir accès à la valeur vraie  $\theta$ . La précision que l'on pourra obtenir sur  $T_n$  ne pourra donc pas descendre en deçà d'une certaine limite (borne inférieure de la variance de l'estimateur ou *Minimum Variance Bound (MVB)*) qui est déterminée par **l'inégalité de Cramer-Rao**:

$$V(T_n) \geq \frac{\tau'(\theta)^2}{I(\theta)}$$

où  $I(\theta)$ , appelée quantité d'information de l'échantillon, est définie par:

$$I(\theta) = E \left[ \left( \frac{\partial \ln L(X, \theta)}{\partial \theta} \right)^2 \right]$$

$L(X, \theta)$  est appelée **fonction de vraisemblance** et se calcule par:

$$L(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta)$$

$f$  désignant la ddp de la v.a.  $X$  et  $\tau(\theta) = E[T_n]$ ,  $\tau'(\theta) = \frac{\partial \tau(\theta)}{\partial \theta}$

Si un estimateur atteint la limite inférieure, on parle alors de MVB estimateur. Cet estimateur est obligatoirement convergent et sans biais.

## Méthode du maximum de vraisemblance

Le critère d'efficacité permet de comparer des estimateurs. On peut aussi s'en servir pour construire un estimateur. Soit  $X$  une variable aléatoire de densité de probabilité  $f(x, \theta)$  connue analytiquement mais dont l'un des paramètres  $\theta$  est inconnu (numériquement). Le problème consiste donc à construire une expression analytique fonction des réalisations de cette variable dans un échantillon de taille  $n$ , permettant de trouver la valeur numérique la plus vraisemblable pour le paramètre  $\theta$ .

Si  $\{x_1, \dots, x_n\}$  sont des réalisations indépendantes de la v.a., on peut dire que

$\bar{x} = \begin{Bmatrix} x_1 \\ \vdots \\ x_n \end{Bmatrix}$  est une réalisation d'un vecteur aléatoire  $\vec{X} = \begin{Bmatrix} X_1 \\ \vdots \\ X_n \end{Bmatrix}$  dont les composantes  $X_i$  sont

indépendantes deux à deux.

L'approche consiste à chercher la valeur de  $\theta$  qui rend le plus probable les réalisations que l'on vient d'obtenir. La probabilité d'apparition a priori de l'échantillon en question peut alors être caractérisée par le produit des probabilités d'apparition de chacune des réalisations (puisque celles-ci sont supposées indépendantes deux à deux).

$$P(\bar{X} = \bar{x}) = \prod_{i=1}^n f(x_i, \theta)$$

La méthode du maximum de vraisemblance consiste à rechercher la valeur de  $\theta$  qui rend cette probabilité maximale. Comme nous l'avons vu plus haut, le produit des valeurs  $f(x_i, \theta)$  est aussi noté  $L(x_1, \dots, x_n, \theta)$  et appelé **fonction de vraisemblance**. La valeur  $\tilde{\theta}$  qui rend maximum la fonction de vraisemblance  $L$  est donc la solution de:

$$\frac{\partial \ln L}{\partial \theta} = 0 \Rightarrow \hat{\theta} : \frac{\partial^2 \ln L}{\partial \theta^2} < 0$$

L'emploi du logarithme sur la fonction  $L$  permet de passer de la maximisation d'un produit à celle d'une somme, le résultat restant le même car la fonction logarithme est monotone strictement croissante.

### Propriétés de la fonction de vraisemblance:

$$\begin{aligned} & - \int_{R^n} L(\bar{x}, \theta) d\bar{x} = 1 \\ & - \int_{R^n} \frac{\partial L(\bar{x}, \theta)}{\partial \theta} = 0 \\ & - E \left[ \frac{\partial \ln(\bar{x}, \theta)}{\partial \theta} \right] = 0 \\ & - E \left[ \left( \frac{1}{L(\bar{x}, \theta)} \frac{\partial L(\bar{x}, \theta)}{\partial \theta} \right)^2 \right] = -E \left[ \frac{\partial^2 \ln(\bar{x}, \theta)}{\partial \theta^2} \right] \end{aligned}$$

### Estimation par intervalle de confiance

Cette nouvelle approche est souvent préférée dans la pratique car elle introduit la notion d'incertitude. On cherche à déterminer l'intervalle  $[a, b]$  centré sur la valeur numérique estimée du paramètre inconnu  $\theta$  contenant la valeur vraie avec une probabilité  $\alpha$  fixée *a priori*. Cette probabilité permet de s'adapter aux exigences de l'application.

$$P[a < \theta < b] = \alpha$$

L'intervalle  $[a, b]$  est appelé intervalle de confiance et  $\alpha$  est le coefficient de confiance. Une estimation par intervalle de confiance sera d'autant meilleure que l'intervalle sera petit pour un coefficient de confiance grand. La donnée de départ, outre l'échantillon, sera la connaissance de la loi de probabilité du paramètre à estimer.

## Annexe IX : Maintenance des capteurs

Voici les consommables les plus souvent renouvelés et leur prix. :

Produits	Quantité	Coût (en euros HT)	Fréquence
<b>Orthophosphate ACA:</b>			
√ Location platine de filtration	<b>1 mois</b>	<b>707</b>	<b>Nettoyage 1fois/mois</b>
√ Solution standard de phosphate 1000mg/l	<b>1 litre</b>	<b>40</b>	<b>1 an</b>
√ Acide sulfurique → réactif	<b>1 L</b>	<b>12</b>	<b>3 mois</b>
√ Acide nitrique (300ml/décolmatage)	<b>1 L</b>	<b>15</b>	<b>3 mois</b>
√ (Acide ascorbique) Essai de décolmatage	<b>1 L</b>	<b>97</b>	<b>3 mois</b>
<b>Sonde pH:</b>			
• Tampon pH 4	<b>1</b>	<b>15</b>	<b>2/3 mois</b>
• Tampon pH 7	<b>1</b>	<b>15</b>	<b>2/3 mois</b>
• Sonde	<b>1</b>	<b>160</b>	<b>1 an</b>
<b>Redox:</b>			
• Tampon Redox	<b>1</b>	<b>15</b>	<b>2/3 mois</b>
• Sonde	<b>1</b>	<b>190</b>	<b>1 an</b>
<b>Oxygène : Kit de maintenance</b>	<b>1</b>	<b>130</b>	<b>6/8 mois</b>
<b>Conductivité: solution tampon (2 à 3 litres par étalonnage)</b>	<b>3</b>	<b>210</b>	<b>1 par étalonnage</b>

*Principaux coûts consommable pour capteurs physico-chimiques.*

Le planning des maintenances hebdomadaires des capteurs biologiques est le suivant :

	LUNDI	MARDI	MERCREDI	JEUDI	VENDREDI
<b>DREISSENA</b>	<ul style="list-style-type: none"> <li>Nettoyage des deux canaux contenant les moules → 45 min</li> <li>Réglage des moules → 1h15</li> </ul>	<ul style="list-style-type: none"> <li>Vérification du réglage des moules → 15 min</li> </ul>		Réglage des moules → 1h30	<ul style="list-style-type: none"> <li>Vérification du réglage des moules → 15 min</li> </ul>
<b>AQUA-TOX</b>	<ul style="list-style-type: none"> <li>Filtration des daphnies → 15 min</li> </ul>	Nettoyage des tuyaux + filtres + cuves 45 min Remise en route du pilote → 15 min	Filtration et changement d'eau des daphnies → 15 min		<ul style="list-style-type: none"> <li>Filtration des daphnies → 15 min</li> </ul>
<b>BAKTERIA</b>	Repiquage de la souche sur boîte de pétri → 30 min		Repiquage de la souche dans l'erlen et mise en culture à Messein → 30 min		<ul style="list-style-type: none"> <li>Nettoyage de l'installation et remise en route → 15 min</li> </ul>

Les maintenances du Fluotox et du Gymnotox ne sont pas précisées sur ce planning car elles sont mensuelles.

### **Voici un récapitulatif des caractéristiques vis-à-vis de la maintenance :**

#### **Bacteria :**

*Nb interventions :* 1-2 fois/semaine

*Durée totale :* 1h/semaine (plus si culture bouillon non délégué à un laboratoire)

*Observations :* Récupération des bouillons bactériens et matières en contact pour autoclavage. Redeviens opérationnel à la détection 1h à 1h30 après la fin de la maintenance.

**Le consommable essentiel** du Bacteria Test est son réactif biologique, à savoir le bouillon de bactéries obtenus à partir de la souche mère.

**La fréquence de renouvellement du bouillon** est de une à deux fois par semaine selon sa propension à se maintenir dans le temps (sa qualité). En effet, un bouillon correct doit une fois inséré dans l'appareil fournir des valeurs de l'ordre des 100 000 au minimum. Et l'on estime que des valeurs inférieures à 20 000 sont insuffisantes.

Remarque: Une pollution n'entraîne pas de destruction du réactif biologique, car seulement quelques millilitres sont prélevés à chaque mesure.

**Le bouillon** s'obtient à partir d'une souche mère *Vibrio fisheri* qu'il convient de renouveler tous les 3 mois. Un protocole de culture de cette souche permet de la maintenir ce laps de temps afin de créer les bouillons nécessaires. Le travail que cela implique peut être confié à un laboratoire. Il faut alors se référencer à la norme ISO11348-4 : 1998 "Qualité de l'eau - Détermination de l'effet inhibiteur d'échantillons d'eau sur la luminescence de *Vibrio fisheri*. (Essai de bactéries luminescentes) ". Le détail des opérations nécessaires se trouve dans la fiche de maintenance en annexe.

Des solutions déjà prêtes à l'emploi existent; ceci élimine l'utilisation d'un laboratoire, mais est beaucoup plus coûteux.

**( prix: environ 350 euros pour 10 flacons permettant de faire fonctionner pendant 10 semaines l'appareil).**

Il convient de préparer aussi une **solution saline à 20 %** (nécessité d'eau distillée et de sel).

#### **Fluotox :**

*Nb interventions :* 1 fois/ mois

*Durée totale :* 1h30min par mois

*Observations :* plus si pollution importante (renouvellement des algues); redevient opérationnel à la détection 1 heure après la fin de la maintenance

Cet appareil ne nécessite que très peu de matériel consommable.

On a besoin du filtre algal (sorte de buvard où sont stockées les algues servant au test) et du tuyau de la pompe péristaltique (élément se détériorant rapidement).

**Fréquence de renouvellement :** 1 fois par mois. **(prix: environ 200 euros/mois)**

Attention: une pollution peut détruire les algues présentes et cela nécessite alors le renouvellement du filtre algal.

Un plus de cet appareil réside dans la facilité de remplacement du filtre.



### Aquatox :

*Nb interventions :* 1-2 fois/semaine

*Durée totale :* 2 h/semaine

*Observations :* change selon saison (varie selon turbidité : + de MES → + d'encrassement)

Le consommable essentiel est le réactif biologique : *daphnia magna* (les daphnies) référencé sous le nom de Daphtoxkit test. Il renferme des œufs de daphnies. Il est également proposé des tubes de spiruline pour nourrir les daphnies.

L'appareil est fourni avec une **station d'élevage des daphnies** (utilisation du système de chauffe principale). Il est possible, pour démarrer un élevage, de se procurer des organismes auprès des différents laboratoires d'analyses des eaux, les daphnies étant des organismes utilisés pour la réalisation de tests de toxicité normalisés.

Il est évident que l'élevage de daphnies présente des avantages notables :

- ✦ Peut être placé sous l'appareil.
- ✦ Utilise le système de chauffage de l'appareil.
- ✦ Ne nécessite que de l'eau de *Volvic* et des algues pour nourriture des daphnies élevées.
- ✦ Permet la récupération des daphnies qui se reproduisent dans l'appareil (avec tamis et eau en pissette).
- ✦ Nécessité d'un système de refroidissement en été si T° > à 22°C.
- ✦ Grande mortalité dans l'élevage par temps d'orage (conditions chaudes, humides et changement brusque lorsque des précipitations surviennent)

**Il en ressort que le prix de consommable est très réduit. Cependant, une présence est nécessaire régulièrement** (cf. planning de maintenance).

### Dreissena :

*Nb interventions :* 2 fois/semaine

*Durée totale :* 2 à 4 h/semaine (2 h si 20 moules ; 4 h si 42 moules)

*Observations :*

- La durée de maintenance peut être réduite en ne réglant que 15 à 20 moules par canal au lieu des 42 prévues. Ceci permet d'avoir des résultats cohérents et de réduire l'entretien d'environ une heure. Après chaque maintenance, l'appareil redevient opérationnel au bout de 2 h (car les moules sont "traumatisés" par les changements et se ferme encore pendant un moment du fait de la réaction de protection)
- Le temps de maintenance est réduit en hiver du fait de la vitesse plus faible de formation des filaments *byssaux* de septembre à mars. En effet, avec les filaments *byssaux* produits, les moules réussissent à se décoller.
- **Attention :** il est important de couper avec un couteau les filaments *byssaux* avant de placer la moule sur un support et de la régler.

Le dreissena ne requiert pas d'autre matériel consommable que les moules.

Bien évidemment, il est également possible de collecter des organismes directement dans le milieu naturel car il s'agit là d'une espèce colonisatrice bien implantée en Europe et notamment dans les bassins de la région Lorraine. **Le prix est alors nul ou très réduit.**

Elles doivent être changées en cas de mort. En fonction de la qualité de l'eau, elles peuvent rester dans le système entre 6 à 12 mois. En pratique, elles sont renouvelées plus souvent pour mettre des moules plus petites donc plus sensibles.

#### **Gymnotox :**

*Nb interventions :* 1 fois /mois (plus si poissons défaillants (maladie..))

*Durée totale :* 5 h/mois (2 heures nettoyage bacs ; 3 h nettoyage tuyauterie)

*Observations :* L'étalonnage nécessaire après tout arrêt complet de l'appareil a une durée de 3 heures. L'appareil ne redevient fonctionnel qu'après cette période.

Le premier réactif consommable est le poisson *Apteronotus Albifron*.

Ces poissons étant populaires en aquariophilie, il est possible de se procurer ces organismes dans différentes animaleries ou directement auprès d'importateurs de poissons tropicaux.

**Le prix de l'*Apteronotus A.* varie selon sa taille.** Il convient de se fournir en poissons de taille voisine des 8-10 cm (ceci afin qu'ils puissent pénétrer dans les tubes aisément sans toutefois se faufiler dans d'autres organes de l'appareil rendant ainsi la récupération du signal impossible). La taille maximale acceptable du poisson pour pouvoir pénétrer dans les tubes collecteurs de signal est de 15 cm.

*Aptéronotus A.* (8-10cm) : **10 à 20 euros le poisson selon le fournisseur.**

Il faut prévoir également un **aquarium équipé pour conserver quelques poissons de réserve** afin de pouvoir remplacer dans les meilleurs délais les poissons utilisés dans le Gymnotox. En effet, ces poissons étant importés de Guyane ou pays au climat similaire, un délai de 3 semaines environ est souvent appliqué avant réception. Il convient de prévoir suffisamment de poissons de réserve pour maintenir un fonctionnement constant.

#### **Traitement :**

Les poissons sont sujets à une maladie provoquant leur mort rapide dès apparitions des premiers symptômes (gélatine, excroissance blanche cotonneuse au niveau des branchies). Il s'agit d'une mycose due au frottement du poisson dans le milieu (grains de sable trop gros, ...) et favorisé par le stress. Un traitement préventif à base de SERA Bactopur est nécessaire dans le (ou les) aquarium (s). Ce produit se trouve dans toute animalerie. (Voir notice)

Il convient également de maintenir un traitement préventif, car la propagation et l'aggravation de la maladie est assez fulgurante.

#### **Alimentation :**

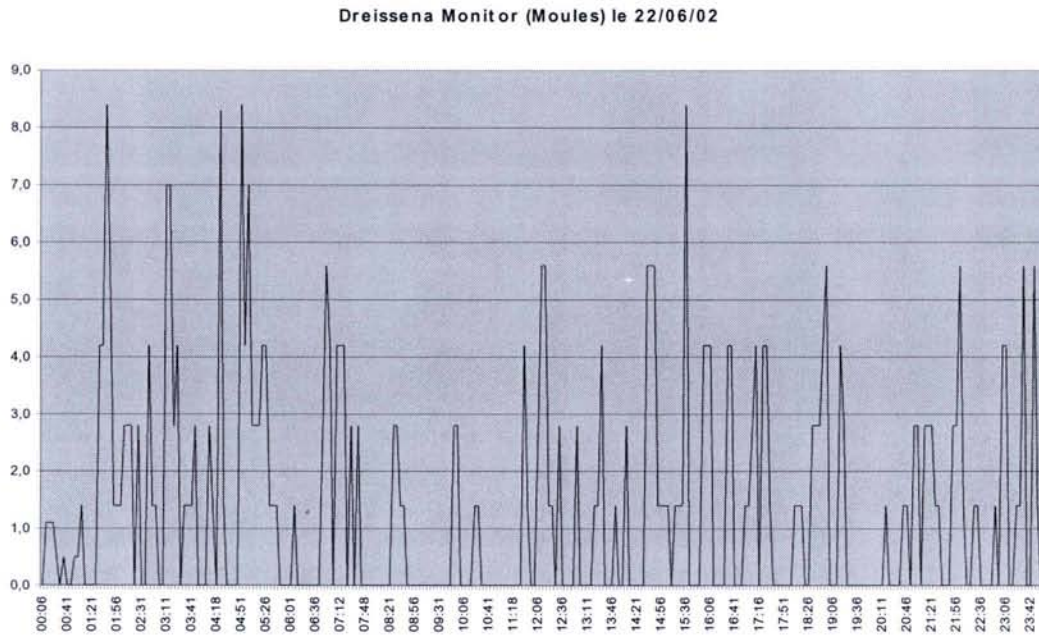
Les poissons ont besoin d'être nourris soit :

- A base de vers de vase (en aquarium ou si quelqu'un vient régulièrement au moins tous les deux jours sur site),
- A base de granulés : alimentation à partir de l'alimentateur automatique de l'appareil (sous réserve de bon fonctionnement: autonomie de 30 jours).

**A titre indicatif, le coût de 10 plaquettes de 100g est de 18 euros.**

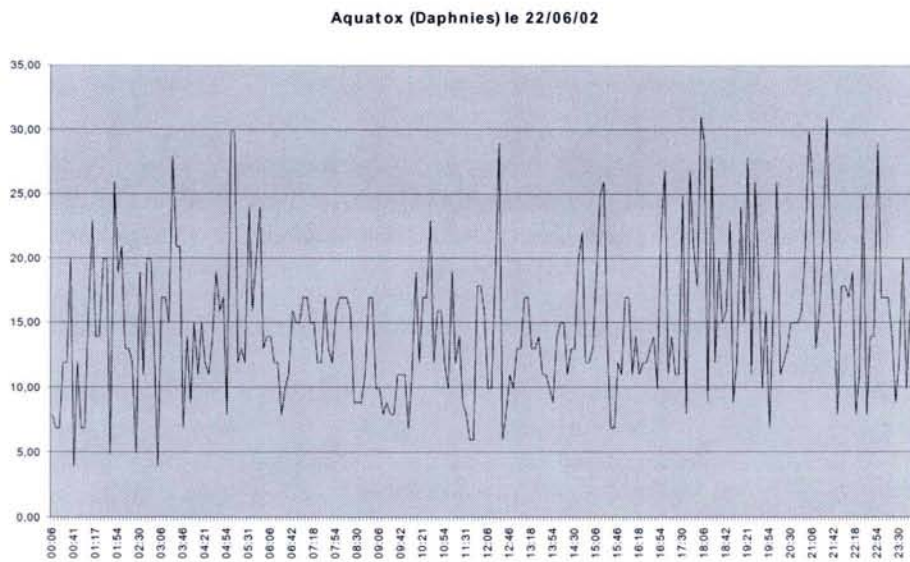
## Annexe X : Autres réactions du 22 juin 2002

### Le Dresseina



Graphique 1 : Nombre de mouvements valvaires en fonction du temps

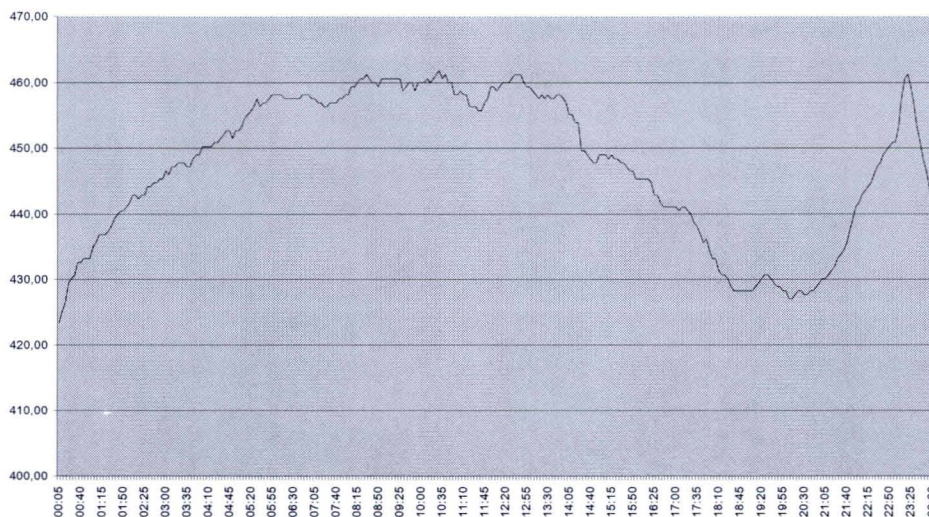
### L'Aquatox



Graphique 2 : Nombre d'impulsion à travers la cuve 1 en fonction du temps

## Le potentiel d'oxydo-réduction (eH)

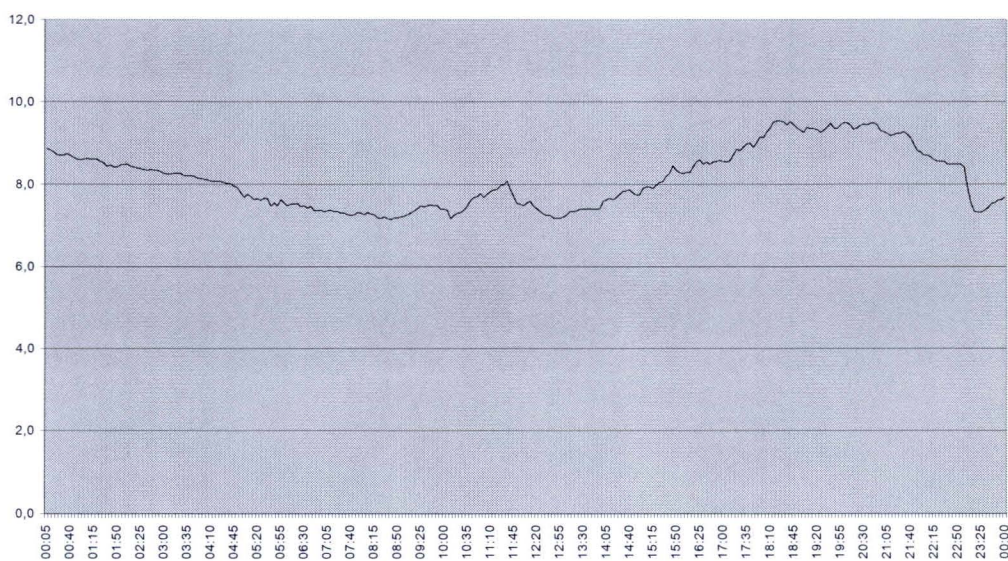
Evolution du potentiel Redox le 22/06/02



Graphique 3 : Variation du eH en fonction du temps

## L'oxygène dissous

Evolution des oxygènes dissous le 22/06/02

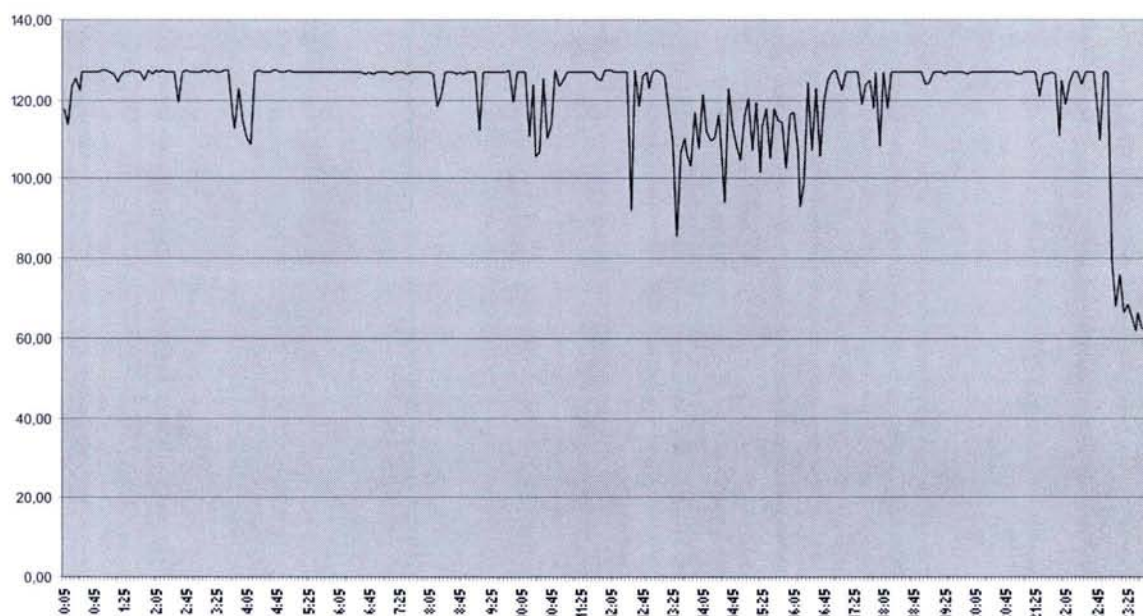


Graphique 4 : Variation de l'oxygène dissous en fonction du temps

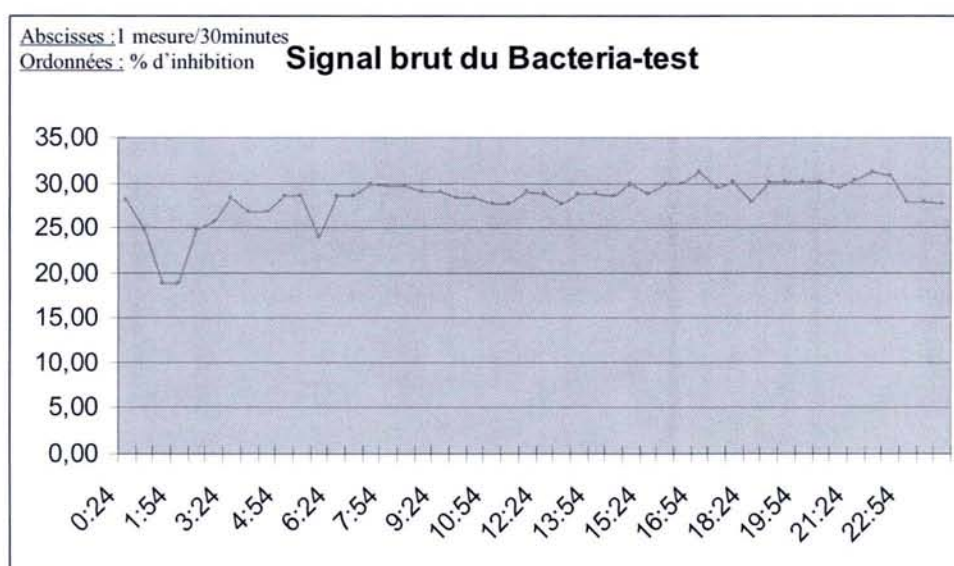
## La chlorophylle

SCD UHP NANCY 1  
Bibliothèque des Sciences  
Rue du Jardin Botanique - CS 20148  
54506 VILLERS LES NANCY CEDEX

Evolution de la Chlorophylle le 22/06/02



Graphique 5 : Variation de la chlorophylle en fonction du temps



Graphique 6 : Variation du Bacteria-test en fonction du temps au cours de la journée du 22 juin 2002

## Glossaire

- Anaérobie : Qualifie un microorganisme qui ne peut vivre qu'en l'absence d'oxygène ou dont le développement est possible en l'absence d'oxygène. Se dit de tout processus ou de toute réaction chimique se déroulant en l'absence d'oxygène.
- Anoxique : Caractérisant la chute des oxygènes dissous dans un milieu aquatique.
- Classification ascendante hiérarchique :  
méthode de classification hiérarchique partant des individus isolés assimilés à des classes et procédant, à chaque étape, par agrégation des deux classes les plus proches au sens de la norme choisie. Un arbre planaire hiérarchique permet de décrire de façon explicite la structure finale de la classification obtenue.
- Colloïdes : Substance finement dispersée, formée de particules de diamètres compris entre  $10^{-4}$  et  $10^{-7}$  cm.
- Ecotoxicologie : Est l'étude des conséquences écologiques de l'injection de substances qualifiées de polluants toxiques dans l'environnement.
- Hétérotrophes : Etres vivants tirant l'énergie nécessaire à leurs synthèses de matières organiques de la dégradation des matières organiques déjà formées.
- Taxon : Groupe d'organismes vivants qui descendent d'un même ancêtre et qui ont certains caractères communs. Les embranchements, classes, ordres, familles, espèces...sont des taxons.

Monsieur Ahcene Ameziane AZZAM

DOCTORAT DE L'UNIVERSITE HENRI POINCARÉ, NANCY-I  
en Automatique et Traitement du Signal

VU, APPROUVÉ ET PERMIS D'IMPRIMER N°1157

Nancy, le 29 mai 2006

Le Président de l'Université



## **TITRE**

CONTRIBUTION A LA MISE EN PLACE D'UN SYSTEME MULTI-CAPTEURS POUR LA SURVEILLANCE DES EAUX DE SURFACE :  
FUSION DE DONNEES ET AIDE A LA DECISION

## **RESUME**

La présente thèse porte sur la surveillance et l'évaluation de la qualité d'un cours d'eau qui alimente un réseau d'eau potable. Les données brutes recueillies à partir d'une installation multi-capteurs, sont transformées en information fusionnée. Cette fusion permet de qualifier les pollutions après avoir relevé les réponses de chaque capteur. Ceci se fait après avoir testé différentes pollutions en circuit confiné, en utilisant en partie, la méthode des plans d'expériences. Sur la base de ces états appelés « masques de pollutions », l'aide multicritère à la décision permet de quantifier le degré de gravité de ces pollutions. L'information ainsi produite, définit le degré de pollution, permet une aide à la décision pour éviter la contamination du réseau d'eau potable, garantissant une action, selon une échelle d'alerte, qui décidera de l'arrêt du pompage par exemple.

## **MOT-CLES**

Station multi-capteur, traitement statistique de données, fusion de donnée, aide à la décision, aide multicritère, plans d'expériences, masque de pollution

## **TITLE**

CONTRIBUTION TO INSTALLATION OF MULTI-SENSORS  
SYSTEM FOR MONITORING SURFACE WATER :  
DATA FUSION AND DÉCISION AID

## **ABSTRACT**

The present thesis investigates the monitoring and the evaluation of the quality of a river which supplies a drinking water network. The raw data collected from a multi-sensors installation, are transformed into data fusion information. This fusion makes it possible to qualify pollutions after having taken down the answers of each sensor. This is done after having tested various pollutions in confined circuit, while using partly, the method of the experimental plans. On the basis of these states called "masks of pollution", the multi-criterion assistance for the decision makes it possible to quantify the degree of seriousness of these pollutions. Information thus produced, defines the degree of pollution, allows a decision aid to avoid the contamination of the drinking water network, guaranteeing an action, according to a level of alarm, which will decide to stop the water pumping for example.

## **KEY WORDS**

Multi-sensors station, statistic data treatment, data fusion, decision aid, aide multi-criterion aid, experiences plan, masks of pollution