

Titre: Analyse statistique de sensibilité du modèle Sancho
Title:

Auteur: John Cigana
Author:

Date: 1997

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Cigana, J. (1997). Analyse statistique de sensibilité du modèle Sancho [Master's thesis, École Polytechnique de Montréal]. PolyPublie.
Citation: <https://publications.polymtl.ca/6886/>

Document en libre accès dans PolyPublie

Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/6886/>
PolyPublie URL:

Directeurs de recherche: Michèle Prévost
Advisors:

Programme: Unspecified
Program:

UNIVERSITÉ DE MONTRÉAL

ANALYSE STATISTIQUE DE SENSIBILITÉ
DU MODÈLE **SANCHO**

JOHN CIGANA

DÉPARTEMENT DES GÉNIES CIVIL, GÉOLOGIQUE ET MINES
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE CIVIL)
NOVEMBRE 1997

© John Cigana, 1997.



National Library
of Canada

Acquisitions and
Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-38667-8

Canada

UNIVERSITÉ DE MONTRÉAL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

ANALYSE STATISTIQUE
DE SENSIBILITÉ
DU MODÈLE SANCHO

Présenté par : CIGANA John
en vue de l'obtention du diplôme de : Maîtrise ès Sciences Appliquées
a été dûment accepté par le jury d'examen constitué de :

M. DESJARDINS Raymond, M. Ing. ,Président
Mme PRÉVOST Michèle, Ph.D., membre et directrice de recherche
M. CLÉMENT Bernard, Ph.D., membre et codirecteur de recherche
M. MOORE Marc, Ph.D., membre

Dédicace

Alors un laboureur dit : Parle-nous du Travail.

Et il répondit disant :

« Vous travaillez pour vous maintenir au diapason de la terre et de son âme.

Car rester oisif, c'est laisser passer les saisons dans l'indifférence et sortir du cortège de la vie avançant majestueusement et en fière soumission vers l'infini.

Lorsque vous travaillez, vous devenez une flûte au cœur de laquelle le murmure lacinant des heures devient musique.

Lequel d'entre vous voudrait encore être un roseau indolent et silencieux, alors qu'autour de vous tout chante à l'unisson ?

Vous avez toujours entendu dire que le travail était une malédiction et le labeur une misère.

Mais je vous dis maintenant : lorsque vous travaillez, vous accomplissez en partie le plus vieux rêve de la terre dont vous êtes devenus les dépositaires aussitôt que ce rêve advint.

C'est en continuant à vous appliquer à votre travail que vous aimez véritablement la vie,

Et aimer la vie par le travail, c'est la rejoindre en sa nature la plus secrète.

Mais si, en un moment de désarroi, vous voyez dans votre naissance une affliction et dans votre nécessité d'assurer votre subsistance, une malédiction dont vous portez la marque sur le front, alors je vous dirais : seule la sueur de votre front saura en effacer la marque infamante.

Vous avez également entendu dire que la vie n'est que ténèbres et dans votre lassitude, vous ne pouvez que répéter ce que disent les éreintés.

Quant à moi, je vous assure que la vie en effet ne peut être que ténèbres, si on n'y rencontre pas l'élan d'une passion,

Et cette passion n'est qu'aveuglement si elle n'est pas guidée par la connaissance,

Et toute connaissance est vaine sauf dans un travail,

Et tout travail n'est que vacuité s'il se fait sans amour;

Lorsque vous travaillez avec amour, vous resserrerrez vos liens envers vous-même, envers les autres et envers Dieu. »

Khalil Gibran, *Le Prophète*

Remerciements

Les départs sont franchement ardu à vivre, parce qu'on réalise ce et surtout ceux qu'on quitte...

À défaut de vous le dire personnellement à chacun d'entre vous, sachez à quel point vous comptez tous pour moi. Merci pour les sourires. Merci pour les « Bonjour, comment ça va ? ». Merci pour l'encouragement. Merci pour les coups de pieds dans l'cul pour me secouer! Thanks for caring ! Ça fait toute la différence au monde.

Donc un gros merci à :

Anja, Anne, Anneli, Anouk, Annik, Annic, Benoit, Bonifacc, Carl, Catherine, Chantal, Charles, Christine, Denis A., Denis B., Émilie, Geneviève, Isabelle B., Isabelle M., Jacinthe, Jean-François, Julie, Josée, Kristina, Lina, Nathalie, Rouga, Robert, Shoukoufeh, Souleymane, Sylvianne, Stéphane B., Stéphane P., Patrick N.

Un chaleureux merci à mon mentor à travers les dédales de SANCHO, le Dr Patrick Laurent. Son aide de tous les instants, sa patience, ses encouragements et son amitité m'ont permis de passer à travers bien des obstacles.

Merci aussi au Dr Pierre Servais de nous avoir permis d'utiliser le modèle SANCHO. Sans son accord initial et ses judicieux conseils par la suite, ce mémoire n'aurait pas eu lieu.

J'aimerais aussi remercier ma directrice, le Dr Michèle Prévost. Il y de ça trois ans, j'ai eu la chance de débuter un PFE de trois crédits sous sa direction. J'ai adoré l'expérience. Dans la foulée, j'ai amorcé une maîtrise sous sa direction. Son dynamisme, ses

encouragements, sa rigueur et sa bonne humeur m'ont emballé et en font un exemple à suivre. Merci pour tout, Michèle.

Merci aussi au Dr Bernard Clément. Sans son aide, sa disponibilité, ses conseils, ses analyses et les méthodes qu'il m'a, tant bien que mal, inculquées ce mémoire n'aurait pas eu lieu.

Merci au professeur Raymond Desjardins d'avoir bien voulu présider mon jury.

Merci aussi au professeur Marc Moore d'avoir voulu siéger sur ce jury.

Finalement, je ne saurais passer sous silence l'appui inconditionnel de ma famille dans ce travail. Leur présence et affection me font toujours chaud au coeur. Grazie.

Résumé

La complexité grandissante des problèmes scientifiques actuels exige l'utilisation d'outils de calculs toujours plus puissants. L'avènement de l'informatique, et particulièrement de la micro-informatique, a mis l'ordinateur au même niveau que le crayon ou la calculatrice comme outil du scientifique. C'est ainsi que des modèles mathématiques de tous genres, auparavant impossibles à manipuler, sont devenus à la portée de tous et chacun.

Un de ces modèles est le modèle déterministe SANCHO. Il fait l'objet de cette étude de sensibilité. Développé par le Dr Gilles Billen et le Dr Pierre Servais de l'Université Libre de Bruxelles pour le compte de la Compagnie Générale des Eaux et du Syndicat des Eaux de l'Île de France, ce modèle calcule les densités bactériennes à partir de la consommation de CODB en réseau. Il est présentement utilisé dans l'industrie de la production de l'eau potable afin de prédire la dynamique bactérienne en réseau.

La valeur d'une réponse physique consiste à l'accompagner d'une indication de sa qualité. Cette indication de qualité s'appelle l'incertitude ou variabilité. L'objectif de cette étude consistait donc à déterminer les paramètres actuels du modèle SANCHO qui provoquent le plus de variabilité des variables de réponse du modèle. Si nous faisons l'hypothèse que le modèle SANCHO représente bien la dynamique bactérienne en réseau, cette étude permettra d'approfondir notre compréhension de la recroissance en identifiant sur ces paramètres critiques.

Deux méthodes complémentaires ont été mises à contribution afin d'étudier la variabilité de ce modèle : la planification d'expérience et la méthode de Taguchi.

La première méthode exploite un plan factoriel partiel (PFP) de 64 essais. Cette première analyse nous a permis de découvrir que certains paramètres influencent la variabilité des six variables de réponse. Ces paramètres critiques sont donc : la mortalité bactérienne

(k_d), le nombre maximal de sites disponibles à la fixation biologique (SPs), le taux maximum d'utilisation des substrats directs (B_{max}), la concentration initiale de substrat directement utilisable (S) et le coefficient de croissance bactérien (Y).

La seconde méthode, soit la méthode des plans croisés de Taguchi (PCT), requiert un total de 672 essais. En contrepartie, cette seconde méthode nous informe de l'importance des paramètres critiques sur la variabilité des variables de réponses. Ces paramètres critiques sont donc : le nombre maximal de sites disponibles à la fixation biologique (SPs), le taux maximum d'utilisation des substrats directs (B_{max}) et la concentration initiale de substrat directement utilisable (S_0).

À la lumière de cette étude, on peut conclure que le modèle SANCHO est sensible à certains paramètres expérimentaux. Ces paramètres critiques sont mentionnés ci-haut. Les travaux futurs sur le modèle SANCHO devraient porter sur la mesure plus précise de ces paramètres. Il n'en demeure pas moins que le modèle, dans son état actuel, simule correctement l'activité bactérienne en réseau de distribution. Des mesures in-situ de ces paramètres seraient souhaitables afin de mieux prédire la dynamique bactérienne en réseau.

Le modèle SANCHO a été développé par l'Université Libre de Bruxelles (ULB) avec le support de la Compagnie Générale des Eaux (CGE) et du SÉDIF (Syndicat des Eaux De l'Île de France). L'utilisation dans le cadre de la Chaire en Eau Potable du modèle SANCHO découle d'une entente d'utilisation intervenue entre l'École Polytechnique de Montréal, Ville de Laval, le SÉDIF, L'ULB et la CGE. De plus, les activités de recherche communes CGE-John Meunier Inc., Ville de Laval, Ville de

Montréal, Vallée Lefebvre et Associés (VLA) et l'ULB dans le groupe de la Chaire encadrent ces efforts de recherche.

Abstract

Over the last few years the use of personal computers has grown tremendously in the population. Likewise, they have become a very powerful tool to solve engineering problems that were, not to long ago, impossible to fathom.

The SANCHO model is one of these tools. Developed at the Université Libre de Bruxelles by Dr. Billen and Dr. Servais, this deterministic model predicts bacterial densities in drinking water distribution systems in relation to the consumption of Biodegradable Dissolved Organic Carbon (BDOC). This model is presently being used in the drinking water industry.

This study proposes to study the sources of variability of the SANCHO model. Through statistical analysis, the influence of each of the initial parameters on the variability of the response variables of the SANCHO model will be studied.

Two methods will be used to accomplish this task : designed experiments and a Taguchi approach.

The first method consists of a 64 run partial fractional design. The results of this analysis show that five initial parameters have an influence on the variability of the response variables. These are : bacterial mortality (k_d), the maximum capacity for bacterial fixation (SPs), the maximum rate of direct substrates uptake (B_{max}), the initial concentration of direct substrates (S_0) and the growth yield coefficient (Y).

The second method of analysis consists of a 672 essay Taguchi crossed array. Again, the variability of the response variables is being studied. Critical parameters for this method are found to be : the maximum capacity for bacterial fixation (SPs), the

maximum rate of direct substrates uptake (B_{max}) and the initial concentration of direct substrates (S_0).

This work has shed some light on the sensitivity of the SANCHO model to some of its initial parameters. These are indicated above. Future work on the SANCHO model should be undertaken to measure each of these parameters with a greater precision. The SANCHO model predictions are nonetheless in good agreement with measures taken on the distribution system. In-situ measures of these parameters are preferable in order to offer a better prediction of the bacterial dynamics in the distribution system.

The SANCHO model was developed at the Université Libre de Bruxelles (ULB) with the support of the Compagnie Générale des Eaux (CGE) and of the SÉDIF (Syndicat des Eaux De l'Île de France). The use of this model at the Chair on Drinking Water is the fruit of an agreement between the École Polytechnique de Montréal, the City of Laval, the SÉDIF, the ULB and the CGE. Furthermore, the joint research activities of CGE-John Meunier Inc. , the City of Laval, the City of Montréal, Vallée Lefebvre (VLA) and the ULB in the Chair on Drinking Water has embraced in these research efforts.

TABLE DES MATIÈRES

DÉDICACE	IV
REMERCIEMENTS	VI
RÉSUMÉ	VIII
ABSTRACT	XI
TABLE DES MATIÈRES.....	XIII
LISTE DES TABLEAUX.....	XVII
LISTE DES FIGURES	XVIII
LISTE DES ABRÉVIATIONS	XX
INTRODUCTION.....	1
CHAPITRE 1. LE MODÈLE SANCHO	3
1.1 HISTORIQUE.....	3
1.2 AUTRES MODÈLES	4
1.3 NÉCESSITÉ DE L'ANALYSE DE SENSIBILITÉ.....	7
1.4 ÉQUATIONS DE SANCHO	8
1.4.1 Hydrolyse bactérienne exoenzymatique de la matière organique.	9
1.4.2 Croissance des bactéries fixes et libres à partir des produits d'hydrolyses	10
1.4.3 Interaction entre les surfaces internes des canalisations et les bactéries lors d'un processus d'adsorption réversible	11
1.4.4 Consommation du chlore causée par une réaction avec la matière organique et l'effet du chlore libre sur les biomasses libres et fixées.	12
1.5 VARIABLES DE RÉPONSES	13

CHAPITRE 2. MATÉRIEL ET MÉTHODES	13
2.1 STRUCTURE DU LOGICIEL.....	13
2.2 CONSTRUCTION D'UNE SIMULATION AVEC SANCHO	15
2.3 MÉTHODE DE MESURES DE CERTAINS PARAMÈTRES DU MODÈLE SANCHO	16
2.3.1 Carbone Organique Dissous Biodégradable (CODB)	16
2.3.2 Capacité maximale de fixation biologique (SPs)	16
2.3.3 Rendement de croissance (Y)	16
2.3.4 Taux de mortalité bactérienne (k_d)	17
2.3.5 Taux maximal d'utilisation de substrats directs (B_{max})	17
2.4 MÉTHODOLOGIE EMPLOYÉE	17
CHAPITRE 3. ANALYSE DE LA VARIANCE DU MODÈLE À L'AIDE D'UN PLAN FACTORIEL PARTIEL.....	19
3.1 CONSTRUCTION DU PLAN FACTORIEL PARTIEL	20
3.1.1 Modèle statistique 20	
3.1.2 Analyse de la variabilité	22
INTRODUCTION.....	27
THE SANCHO MODEL	28
EXPERIMENTAL DESIGN AND METHODS.....	31
EXPERIMENTAL DESIGN OF THE SANCHO MODEL	32
METHOD.....	33
STATISTICAL ANALYSIS OF THE 64 RUN DESIGN	34
IMPORTANCE OF ORTHOGONALITY AND INTERPRETATION.....	35
RESULTS AND DISCUSSION	36
B1- BIOLOGICALLY ATTACHED BACTERIAL BIOMASS (FIXED BIOMASS)	37
B2 - REVERSIBLY ADSORBED BACTERIAL BIOMASS (ADSORBED BIOMASS)....	41
B3 - FREE BACTERIA IN THE LIQUID PHASE (FREE BACTERIA)	43
H1 - RAPIDLY DEGRADABLE MACROMOLECULAR ORGANIC MATTER.....	45
H2 - SLOWLY DEGRADABLE MACROMOLECULAR ORGANIC MATTER	46
S - DIRECTLY UTILIZABLE MONOMERIC SUBSTRATES	46
INFLUENCE OF THE TIME STEP USED FOR COMPUTATION.....	47
CONCLUSION	48

REFERENCES.....	50
CHAPITRE 4. ANALYSE DE LA VARIANCE DU MODÈLE À L'AIDE DE PLAN CROISÉS DE TAGUCHI.....	64
4.1 MATRICE D'EXPÉRIMENTATION.....	65
4.2 MATRICE DE BRUIT	65
4.3 MODÈLE STATISTIQUE	67
INTRODUCTION.....	70
THE SANCHO MODEL.....	71
EXPERIMENTAL DESIGN AND METHODS	74
CROSSED ARRAY DESIGN	75
ANALYSIS OF THE MODEL	76
RESULTS AND ANALYSIS	77
B1- BIOLOGICALLY ATTACHED BACTERIAL BIOMASS (FIXED BACTERIAL BIOMASS).....	78
B2 - REVERSIBLY ADSORBED BACTERIAL BIOMASS (ADSORBED BIOMASS).....	81
B3 - FREE BACTERIA IN THE LIQUID PHASE (SUSPENDED BACTERIAL BIOMASS).....	82
H1 - RAPIDLY DEGRADABLE MACROMOLECULAR ORGANIC MATTER.....	86
H2 - SLOWLY DEGRADABLE MACROMOLECULAR ORGANIC MATTER	86
S - DIRECTLY UTILIZABLE MONOMERIC SUBSTRATE	87
CHAPITRE 5. DISCUSSION ET CONCLUSIONS	103
5.1 BIOMASSE BACTÉRIENNE FIXÉE (B1)	104
5.2 BIOMASSE BACTÉRIENNE ADSORBÉE (B2).....	109
5.3 BACTÉRIES LIBRES (B3)	110
5.4 MATIÈRE POLYMÉRIQUE RAPIDEMENT HYDROLYSABLE (H₁).....	111

5.5 MATIÈRE POLYMÉRIQUE LENTEMENT HYDROLYSABLE (H₂)	112
5.6 SUBSTRATS DIRECTEMENT UTILISABLES (S)	113
CHAPITRE 6. RÉFÉRENCES.....	115

Liste des tableaux

<u>TABLEAU 1.1</u> VARIABLES DE RÉPONSE DU MODÈLE SANCHO	13
<u>TABLEAU 2.1</u> PARAMÈTRES INITIAUX DU MODÈLE SANCHO	14
<u>TABLEAU 3.2</u> CHARACTERISTICS AND CONFOUNDING RULES OF THE 64 RUN DESIGN.....	54
<u>TABLEAU 3.3</u> PERCENTAGE OF VARIABILITY EXPLAINED FOR EACH RESPONSE VARIABLE	55
<u>TABLEAU 4.1</u> LIST OF INITIAL PARAMETERS INVOLVED IN THE SANCHO MODEL.....	93
<u>TABLEAU 4.2</u> PERCENTAGE OF VARIABILITY EXPLAINED FOR EACH RESPONSE VARIABLE	94
<u>TABLEAU 5.1</u> POURCENTAGE D'EXPLICATION DE LA VARIABILITÉ POUR CHACUNE DES SIX VARIABLES DE RÉPONSES SELON LE FPF ET LE PCT.....	104

Liste des figures

<u>FIGURE 1.1</u> CLASSIFICATION DES MODÈLES ÉTUDIÉS.....	5
<u>FIGURE 1.2.</u> PROCESSUS DU MODÈLE SANCHO (ADAPTÉ DE SERVAIS <i>ET AL.</i> 1995)	8
<u>FIGURE 2.1</u> SCHÉMA SIMPLIFIÉ DU SYSTÈME SANCHO	17
<u>FIGURE 3.1</u> PROCESSES OF THE SANCHO MODEL (ADAPTED FROM SERVAIS <i>ET AL.</i> , 1995).....	56
<u>FIGURE 3.2</u> PARETO GRAPH FOR THE RESPONSE VARIABLE B1(FIXED BACTERIA) AT 3 AND 8 HOURS OF RESIDENCE TIME	57
<u>FIGURE 3.3.</u> INFLUENCE OF A VARIATION IN BACTERIAL MORTALITY (K_D) ON THE RESPONSE VARIABLE FIXED BIOMASS	58
<u>FIGURE 3.4</u> PARETO GRAPHS FOR THE RESPONSE VARIABLE B2(ADSORBED BIOMASS) AT 3 AND 8 HOURS OF RESIDENCE TIME	59
<u>FIGURE 3.5</u> PARETO GRAPHS FOR THE RESPONSE VARIABLE B3(FREE BACTERIA) AT 3 AND 8 HOURS OF RESIDENCE TIME	60
<u>FIGURE 3.6</u> PARETO GRAPHS FOR THE RESPONSE VARIABLE H1 AT 3 AND 8 HOURS OF RESIDENCE TIME.....	61
<u>FIGURE 3.7</u> PARETO GRAPHS FOR THE RESPONSE VARIABLE H2 AT 3 AND 8 HOURS OF RESIDENCE TIME	62
<u>FIGURE 3.8</u> PARETO GRAPHS FOR THE RESPONSE VARIABLE S AT 3 AND 8 HOURS OF RESIDENCE TIME.....	63
<u>FIGURE 4.1.</u> RÉPRÉSENTATION DU PLAN CROISÉS DE TAGUCHI	66
<u>FIGURE 4.2</u> PROCESSES OF THE SANCHO MODEL (ADAPTED FROM SERVAIS <i>ET AL.</i> , (1995)).....	95
<u>FIGURE 4.3</u> REPRESENTATION OF THE TAGUCHI CROSSED ARRAY	96
<u>FIGURE 4.4.</u> PARETO GRAPH FOR THE VARIANCE OF THE RESPONSE VARIABLE B1(FIXED BACTERIA) AT 1, 3 AND 8 HOURS OF RESIDENCE TIME.....	97

<u>FIGURE 4.5. PARETO GRAPH FOR THE VARIANCE OF THE RESPONSE VARIABLE B2</u>	
(ADSORBED BIOMASS) AT 1, 3 AND 8 HOURS OF RESIDENCE TIME.....	98
<u>FIGURE 4.6 PARETO GRAPH FOR THE VARIANCE OF THE RESPONSE VARIABLE B3</u>	
(FREE BACTERIA) AT 1, 3 AND 8 HOURS OF RESIDENCE TIME	99
<u>FIGURE 4.7 PARETO GRAPH FOR THE VARIANCE OF THE RESPONSE VARIABLE H1</u>	
(LABILITY CLASS 1) AT 1, 3 AND 8 HOURS OF RESIDENCE TIME.....	100
<u>FIGURE 4.8 PARETO GRAPH FOR THE VARIANCE OF THE RESPONSE VARIABLE H2</u>	
(LABILITY CLASS 2) AT 1, 3 AND 8 HOURS OF RESIDENCE TIME.....	101
<u>FIGURE 4.9 PARETO GRAPH FOR THE VARIANCE OF THE RESPONSE VARIABLE S</u>	
(DIRECT SUBSTRATES) AT 1, 3 AND 8 HOURS OF RESIDENCE TIME	102

Liste des abréviations

BAM : Biological Accumulation Model

COA : Carbone organique assimiliable

CODB : Carbone Organique Dissous Biodégradable

DOE : Design of Experiments

MSU : Montana State University

PCT : Plan croisé de Taguchi

PFP : Plan factoriel partiel

ULB : Université Libre de Bruxelles

Introduction

Cette étude porte sur l'analyse de sensibilité du modèle SANCHO. Ce modèle prédit la croissance de biomasse bactérienne, la consommation de carbone organique dissous biodégradable (CODB) et la consommation de chlore dans un réseau de distribution d'eau potable. Cette recroissance, ou reviviscence bactérienne, en conduite d'eau potable est un sujet d'actualité pour les producteurs d'eau potable, les municipalités et la population en général. Une enquête récente de 164 fournisseurs d'eau potable rapporte qu'une grande proportion des réseaux américains montrent une présence récurrente de coliformes (Smith and Hess 1990). L'eau potable, une fois traitée en usine, séjourne parfois plusieurs heures dans un réseau de distribution avant d'être utilisée par le consommateur. Dans un intérêt de santé publique, on veut éviter l'apparition d'une biomasse bactérienne qui pourrait affecter négativement la qualité de l'eau potable refoulée ou être le point de départ d'un chaîne trophique d'organismes supérieurs. L'utilité d'un modèle déterministe, comme le modèle SANCHO réside dans l'évaluation et la prédiction de cette cinétique bactérienne. L'objectif de cette étude est de déterminer les sources de variabilité de ce modèle afin d'en améliorer l'efficacité de prédiction.

Le premier chapitre de ce mémoire présente le modèle SANCHO accompagné des équations, relations et concepts qui le constituent. Les raisons qui rendent nécessaire l'analyse de sensibilité de ce modèle sont aussi présentées.

Le second chapitre présente le matériel et les méthodes utilisée pour la réalisation de ce travail. Les méthodes utilisées relèvent de la branche des statistiques connue sous le nom de la planification statistique d'expérience. Outre la procédure pour l'accomplissement d'une simulation avec le modèle SANCHO, le détail des méthodes mathématiques et statistiques retenues sont fournies. L'essentiel des résultats sont présentés sous forme d'articles.

Suit ensuite une conclusion des travaux entrepris ainsi que des recommandations.

Les annexes seront composées de différents articles connexes traitant du modèle SANCHO et d'un modèle cousin au modèle SANCHO, soit le modèle CHABROL.

Le modèle SANCHO a été développé par l'Université Libre de Bruxelles (ULB) avec le support de la Compagnie Générale des Eaux (CGE) et du SÉDIF (Syndicat des Eaux De l'Île de France). L'utilisation dans le cadre de la Chaire en Eau Potable du modèle SANCHO découle d'une entente d'utilisation intervenue entre l'École Polytechnique de Montréal, Ville de Laval, le SÉDIF, L'ULB et la CGE. De plus, les activités de recherche communes CGE-John Meunier Inc., Ville de Laval, Ville de Montréal, Vallée Lefebvre et Associés (VLA) et l'ULB dans le groupe de la Chaire encadrent ces efforts de recherche.

Chapitre 1. Le modèle SANCHO

Le présent chapitre vise à présenter le modèle SANCHO. On élabore d'abord un bref historique montrant les origines du modèle SANCHO. Il est à noter que le modèle SANCHO a été développé à l'Université Libre de Bruxelles (ULB) avec le support de la Compagnie Générale des Eaux (CGE) et du SÉDIF (Syndicat des Eaux De l'Île de France). L'utilisation dans le cadre de la Chaire en Eau Potable du modèle SANCHO découle d'une entente d'utilisation intervenue entre l'École Polytechnique de Montréal, Ville de Laval, le SÉDIF, L'ULB et la CGE. De plus, les activités de recherche communes CGE-John Meunier Inc., Ville de Laval, Ville de Montréal, Vallée Lefebvre et Associés (VLA) et l'ULB dans le groupe de la Chaire encadrent ces efforts de recherche. Divers modèles utilisés par d'autres chercheurs sont résumés et discutés. Finalement, les concepts ainsi que les relations mathématiques qui sous-tendent le modèle SANCHO sont exposés.

1.1 Historique

Le modèle SANCHO simule de manière déterministe les densités de biomasse bactérienne présentes dans les conduites d'eau potable à partir des concentrations de carbone organique dissous biodégradable (ou CODB) et des concentrations de chlore (Servais et al. 1995). Le CODB représente la quantité de substrat disponible pour les populations bactériennes. Le CODB est mesuré selon la méthode de Servais-Billen (Servais et al. 1989). L'énoncé du modèle SANCHO est inspiré de la similitude existant entre la croissance de la biomasse bactérienne sur les parois des grains de charbon actif dans les filtres biologiques et la croissance de la biomasse bactérienne sur les parois des conduites dans les réseaux de distribution.

En effet, de la même façon que les bactéries s'adsorbent et se désorbent, se fixent et croissent sur les grains de charbon actif biologique (CAB), la biomasse bactérienne est régie par les mêmes mécanismes sur les parois des conduites d'un réseau de distribution. Le modèle original décrivant l'utilisation du CODB dans les filtres biologiques se nomme le modèle CHABROL (Billen et al. 1992). Bien qu'utilisant les mêmes concepts et équations, les deux modèles se différencient par les valeurs numériques attribuées à certains paramètres et par l'incorporation de l'impact du chlore dans le modèle SANCHO.

Ces deux modèles ont comme ancêtre commun le modèle H3SB (Servais 1986). Ce modèle relie la dégradation de la matière organique dans la rivière Meuse (Belgique) à la présence de bactéries planctoniques en se basant sur les mêmes relations microbiologiques que le modèle SANCHO.

1.2 Autres modèles

L'observation d'une augmentation du nombre de bactéries dans l'eau traitée et durant sa distribution est communément appelée recroissance bactérienne. Cette recroissance est présumée associée à la prolifération de biomasse bactérienne présente dans le réseau de distribution.

De nombreuses études sur la dynamique bactérienne ont été réalisées récemment, soit à l'échelle pilote (Block et al. 1993; Mathieu 1992), soit à l'échelle réelle (Desjardins et al. 1997; Desjardins et al. 1991; Laurent et al. 1997; Prévost et al. 1997a; Servais et al. 1995). Ces travaux ont permis une meilleure compréhension des phénomènes et processus menant à la recroissance bactérienne (Laurent et al. 1997; Servais et al. 1995). Combinée à la puissance accrue des ordinateurs, les chercheurs ont élaboré divers modèles biologiques ou mathématiques pouvant simuler la recroissance en réseau.

Les modèles actuellement étudiés peuvent se diviser en deux grandes catégories : les modèles empiriques et les modèles déterministes. Un modèle empirique est uniquement basé sur une relation statistique tirée de mesures expérimentales entre la variable de réponse et les paramètres. Un modèle déterministe se base plutôt sur les relations ou lois scientifiques pour relier la variable de réponse aux paramètres.

Une fois la structure du modèle connue, on peut discuter de la fonction du modèle. Deux utilités se distinguent : les modèles de balancement hydraulique et les modèles microbiologiques de l'eau ou une combinaison des deux.

La figure 1.1 montre où se situent les modèles utilisés par d'autres chercheurs.

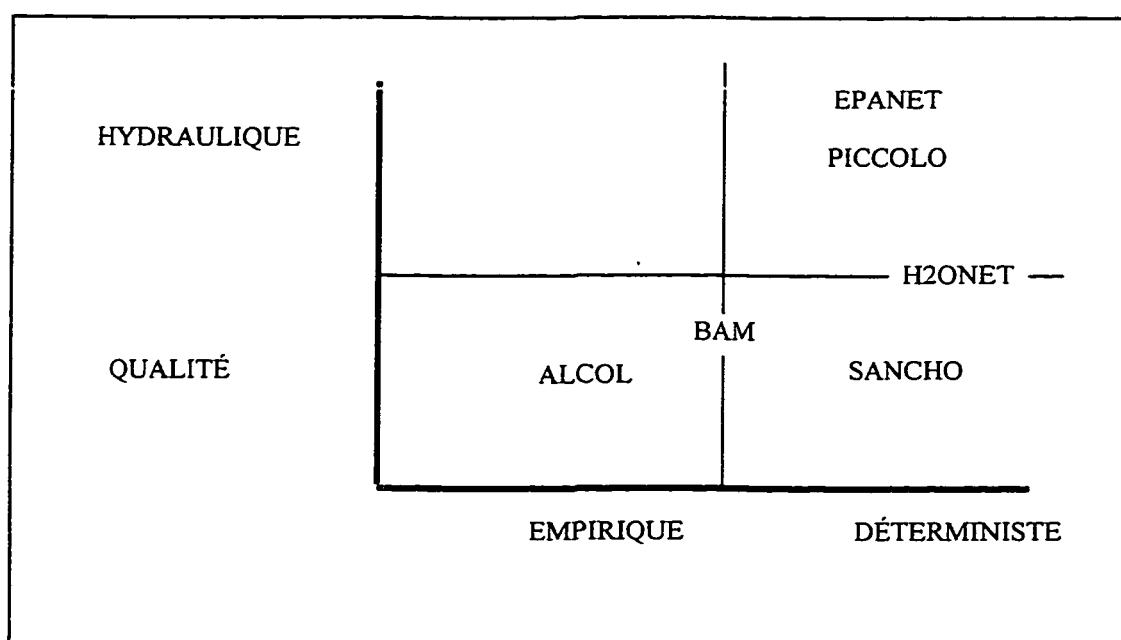


Figure 1.1 Classification des modèles étudiés.

Dans la catégorie des modèles empiriques traitant de la qualité de l'eau distribuée, on peut mentionner le modèle ALCOL. Développé par (Volk and Joret 1994), ce modèle évalue le risque d'apparition de coliformes totaux en se basant, sur la moyenne pour le

réseau, de la température, de la concentration résiduelle du désinfectant, du CODB et de l'abondance de bactéries mesurée par microscopie à l'épifluorescence. Ce modèle ne nécessite pas une connaissance approfondie du réseau ni même de son hydraulique.

Le modèle BAM (Biological Accumulation Model) chevauche l'empirique et le déterminisme. BAM simule la croissance d'un biofilm et l'utilisation d'un substrat mesuré par la méthode du carbone organique assimilable (COA) dans une conduite. L'écoulement piston désiré est obtenu en juxtaposant en série plusieurs réacteurs complètement mélangés. Ce modèle est utilisé à la Montana State University (MSU) par l'équipe du Dr. Anne Camper. BAM utilise un format de base de données, il devient donc facile de spécifier et modifier les paramètres. L'utilisation de ce modèle est limitée par le nombre élevé de paramètres difficilement mesurables empiriquement (Camper et al. 1993). Il n'existe présentement pas d'analyse de la sensibilité de ce modèle.

Parmi les modèles déterministes, les modèles de calibration hydrauliques sont très nombreux. Le modèle EPANET (Rossman 1993) distribué gratuitement par l'USEPA (United States Environmental Protection Agency), nécessite une connaissance détaillée de chacune des conduites du réseau. EPANET calcule alors le sens de l'écoulement, le débit dans chaque conduite ainsi que la pression aux noeuds. On peut aussi suivre l'évolution d'un traceur conservatif dans l'écoulement. Ce modèle possède aussi la capacité de simuler des phénomènes de diffusion dans le liquide et entre le liquide et la paroi. Dans la même veine, le modèle PICCOLO (Pirou et al. 1996) requiert une très bonne connaissance de la structure du réseau (diamètre, longueur, etc) et fournit, de la même façon, les vitesses d'écoulement en conduite, les pressions aux noeuds et un calcul des temps de séjour.

À l'intersection du modèle hydraulique et de qualité, se situe le modèle H2ONET de MWSoft Inc. (Boulos et al. 1995). Ce modèle utilise un algorithme de calcul perfectionné (« Hybrid Method ») qui permet de simuler le réseau complet en régime permanent ou en

dynamique. En plus des résultats complets sur l'hydraulique des réseaux, des paramètres de qualité de l'eau et le temps de résidence sont fournis. Ce modèle allie la principale fonctionnalité de EPANET, soit le balancement hydraulique, avec un module de qualité de l'eau. H2ONET utilise la version 12 de AutoCAD pour les présentations graphiques. Le coût élevé du produit ainsi que la complexité du système informatique requis est son principal désavantage.

SANCHO (Servais et al. 1995)

Le modèle SANCHO a été développé par les Dr. Billen et Servais de l'ULB (Université Libre de Bruxelles). Ce modèle déterministe prédit les densités bactériennes à partir de la consommation de CODB en réseau (Servais et al. 1995). Il a été testé avec succès d'abord en Europe et au Canada. Une assez bonne concordance a été établie entre les résultats observés et ceux prédits (Laurent et al. 1997.; Servais et al. 1995). Ceci a poussé à en analyser les sources de variabilité. Ce dernier modèle fait l'objet d'une analyse de sensibilité qui est le thème de ce mémoire.

1.3 Nécessité de l'analyse de sensibilité

Les résultats positifs obtenus par l'utilisation du modèle SANCHO dans divers réseaux nous poussent à croire que la formulation de ce modèle ainsi que ses concepts sous-jacents sont justes. Des améliorations futures à la formulation du modèle ne peuvent venir que d'une compréhension meilleure de ses forces et faiblesses. Les sources de variabilité du modèle SANCHO sont donc étudiées lors de cette étude. Entre autre, on cherche à déterminer quels paramètres initiaux sont les plus influents quant à la variabilité totale du modèle.

1.4 Équations de SANCHO

Les processus pris en compte dans le modèle SANCHO sont décrits en détail par (Servais et al. 1995). Les processus sont donc :

1. l'hydrolyse bactérienne exoenzymatique de la matière organique;
2. la croissance des bactéries fixes et libres à partir des produits d'hydrolyses;
3. l'interaction entre les surfaces internes des canalisations et les bactéries selon un processus d'adsorption réversible;
4. la consommation du chlore par réaction avec la matière organique et l'effet du chlore libre sur les biomasses libres et fixées.

La figure 1.2 présente schématiquement ces processus.

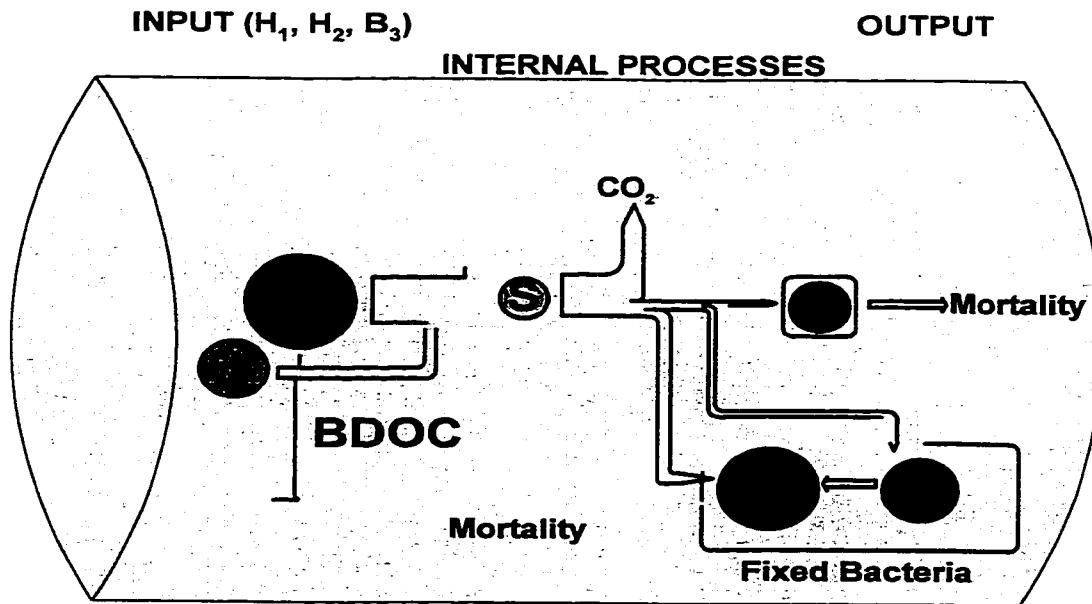


Figure 1.2. Processus du modèle SANCHO (adapté de Servais *et al.* 1995)

1.4.1 Hydrolyse bactérienne exoenzymatique de la matière organique.

Dans la formulation du modèle SANCHO, la source de carbone organique disponible pour les bactéries est le Carbone Organique Dissous Biodégradable ou CODB. La mesure du CODB présent dans l'eau distribuée d'un réseau peut-être déterminé selon la méthode de Servais-Billen. (Servais et al. 1989). Dans la nature, la matière organique est présente surtout sous la forme de macromolécules. Or la croissance bactérienne est liée à l'assimilation de petites molécules organiques (tels les acides aminés libres, mono- et oligosaccharides et acides organiques) qui peuvent être directement assimilées (Rogers 1961). Ces petites molécules sont appelées »substrats directs» dans la définition du modèle SANCHO (Servais et al. 1995). Les macromolécules de matière organique peuvent à leur tour être hydrolysées en molécules plus courtes, donc assimilables. La cinétique de dégradation des macromolécules en monomères est fonction de la nature plus ou moins complexe des macromolécules. Dans le modèle SANCHO, les macromolécules qui sont dégradées rapidement par hydrolyse exoenzymatique sont dénommées H_1 alors que les macromolécules demandant un temps d'hydrolyse plus long sont regroupées comme faisant partie du pool de H_2 dans ce modèle.

Dans le modèle SANCHO, la valeur totale du CODB est donc fractionnée en trois classes de substrats, soit:

S: Substrats monomériques directement assimilables (mg C/l)

H_1 : Matière organique polymérique pouvant être hydrolysée rapidement par les exoenzymes (mg C/l)

H_2 : Matière organique polymérique pouvant être hydrolysée lentement par les exoenzymes (mg C/l)

L'hydrolyse exoenzymatique des fractions H_1 et H_2 produit des substrats monomériques directement assimilables (S). Le taux d'utilisation de la fraction H_1 et H_2 est régi par une cinétique de Michaelis-Menten:

$$\frac{dH_i}{dt} = e_{i,MAX} \frac{H_i}{H_i + k_{H_i}} B \quad (1.1)$$

où

$e_{i,MAX}$ = Taux maximal d'hydrolyse exoenzymatique de la classe de substrats i (h^{-1})

k_{H_i} = Constante de demi-saturation du substrat H_i (mg C/l)

B = Biomasse bactérienne (mg C/l)

Parallèlement à l'hydrolyse des classes de substrats H_1 et H_2 , les substrats directs (S) sont métabolisés par les bactéries conformément à une cinétique de Michealis-Menten. Le taux d'utilisation du substrat S est donc défini comme :

$$\frac{dS}{dt} = B_{max} \frac{S}{S + k_S} B \quad (1.2)$$

B_{max} = Taux maximal d'utilisation du substrat S (h^{-1})

k_S = Constante de demi-saturation du substrat S (mg C/l)

B = Biomasse bactérienne (mg C/l)

1.4.2 Croissance des bactéries fixes et libres à partir des produits d'hydrolyses

Des réactions de croissance et de mortalité sont associées à la présence combinée de biomasse bactérienne et d'un substrat en quantité limitée, défini par le flux de substrat du CODB. La croissance dans le modèle SANCHO est régie par un terme de microbiologie classique connu sous le nom de rendement de croissance (Y). Ce terme représente la fraction constante de biomasse produite par unité de substrat consommé. De la même façon, la croissance de biomasse est accompagnée par sa mortalité. Un terme de mortalité

bactérienne (k_d) est introduit dans le modèle SANCHO. Il représente une cinétique de premier ordre quant à la biomasse et ses unités sont exprimées en h^{-1} . L'activité bactérienne (A) ainsi que la mortalité bactérienne (k_d) sont affectées par les variations de températures mesurées dans l'eau distribuée (Lehman et al. 1975) ont proposé la relation empirique suivante pour dépeindre la variation de l'activité biologique (A) en fonction de la température (T).

$$A = A(T_{opt}) e^{-\frac{(T_{opt}-T)^2}{(T_{opt}-T_i)^2}} \quad (1.3)$$

où T_{opt} représente la température optimale de croissance de la communauté et T_i représente l'étendue de la fonction sigmoidale. Les paramètres T_{opt} et T_i sont donc définis selon les relations suivantes:

$$T_{opt}(T) = 40 - \left(\frac{20 - T}{2} \right) \quad (1.4)$$

$$T_i(T) = 18 - \left(\frac{20 - T}{2} \right) \quad (1.5)$$

1.4.3 Interaction entre les surfaces internes des canalisations et les bactéries lors d'un processus d'adsorption réversible

Un isotherme de Langmuir est adopté pour décrire l'attachement des bactéries libres sur les bactéries formant le biofilm. Le taux d'adsorption des bactéries est exprimé en fonction de la capacité maximale de support biologique (SB), d'une constante d'adsorption (k_{ads}), du nombre de bactéries libres disponibles pour l'adsorption (B_3) et le nombre de site disponibles (SB-B₂).

$$Taux\ d' adsorption = k_{ads} B_3 (SB - B_2) \quad (1.6)$$

Le taux de désorption est relié au nombre de bactéries adsorbées (B_2) et à une constante de désorption (k_{des}).

$$Taux\ de\ désorption = k_{des} B_2 \quad (1.7)$$

Lorsque ces deux taux de réaction sont égaux, l'état d'équilibre est atteint. On obtient donc la formulation suivante de l'isotherme de Langmuir:

$$B_2 = SB \frac{B_3}{\frac{k_{des}}{k_{ads}} + B_3} \quad (1.8)$$

1.4.4 Consommation du chlore causée par une réaction avec la matière organique et l'effet du chlore libre sur les biomasses libres et fixées.

La présence de chlore libre dans les réseaux d'eau potable contribue au contrôle de la dynamique bactérienne en réseau. Le chlore a un double effet sur la biomasse bactérienne :

1. l'augmentation de la mortalité par lyse bactérienne;
2. la modification de l'activité des souches bactériennes.

Dans le cadre de ce projet, nous avons considéré comme nulle la concentration de chlore dans le réseau. L'objectif visé par cette simplification est de simuler la recroissance bactérienne dans une situation critique, soit en absence de biocide. La bibliographie montre aussi que la recroissance bactérienne se produit principalement lorsque la

température dépasse 14°C (Colbourne et al. 1992; LeChevallier 1990). La température a donc été maintenue constante à 20°C.

1.5 Variables de réponses

Les 6 variables de réponse du modèle SANCHO ainsi que leurs unités sont présentées au Tableau 1.1.

Tableau 1.1 Variables de réponse du modèle SANCHO

N°	Variable de Réponse	Nom	Unités
1	B ₁	Bactéries fixées biologiquement	µg C/ cm ²
2	B ₂	Bactéries adsorbées de façon réversible	µg C/ cm ²
3	B ₃	Bactéries libres en suspension	µg C/L
4	H ₁	Matière organique macromoléculaire rapidement hydrolysable	mg C/L
5	H ₂	Matière organique macromoléculaire lentement hydrolysable	mg C/L
6	S	Substrats monomériques directement utilisables	mg C/L

Chapitre 2. Matériel et Méthodes

2.1 Structure du logiciel

La version actuelle du modèle SANCHO nécessite l'introduction de 19 paramètres initiaux. Le tableau 2.1 résume ces 19 paramètres, leur symboles ainsi que leurs unités.

Tableau 2.1 Paramètres initiaux du modèle SANCHO

No	Symbol	Nom du paramètre	Unités
1	$e_{1\max}$	Taux maximum d'hydrolyse exoenzymatique	h^{-1}
2	k_{h1}	Constante de demi-saturation du CODB de	mg C/L
3	$e_{2\max}$	Taux maximum d'hydrolyse exoenzymatique	h^{-1}
4	k_{h2}	Constante de demi-saturation du CODB de	mg C/L
5	B_{\max}	Taux maximum d'utilisation des substrats	h^{-1}
6	k_s	Constante de demi-saturation des substrats	mg C/L
7	Y	Rendement de croissance bactérien	-
8	α	Fraction de la biomasse bactérienne recyclée	-
9	k_d	Taux de mortalité bactérienne	h^{-1}
10	k_{ads}	Taux d'adsorption des bactéries sur le support	$\mu\text{gC/mL } h^{-1}$
11	k_{des}	Taux de désorption des bactéries du support	h^{-1}
12	SPs	Capacité maximale de fixation biologique sur	$\mu\text{gC/cm}^2$
13	Fractio	Fraction de la capacité maximale de fixation	%
14	□	Temps d'analyse	h
15	H_{10}	Matière organique macromoléculaire	mg C/L
16	H_{20}	Matière organique macromoléculaire	mg C/L
17	S_0	Substrat monomérique directement utilisable	mg C/L
18	B_{30}	Bactéries libres dans la phase liquide	$\mu\text{g C/L}$
19	dtsej	Incrément de temps	h

Ces paramètres sont introduits de façon séquentielle dans le logiciel. Certains modifications nécessaires ont été effectuées au logiciel d'origine afin d'en améliorer l'exploitation:

- Une modification du code de SANCHO afin de pouvoir exploiter comme variables certains paramètres initialement considérés comme des constantes. En effet, les paramètres $e_{1\max}$, $e_{2\max}$, B_{\max} , k_{h1} , k_{h2} , k_s , k_{ads} , k_{des} , Y et k_d étaient considérés comme des constantes dans la version originale du logiciel. Après ces modifications, tous ces termes sont accessibles à l'utilisateur.

- Une modification du code de SANCHO en ajoutant des structures de fichiers de sortie.

Ceci permet de :

1. présenter les paramètres initiaux de la simulation effectuée dans un fichier de sortie;
2. présenter les résultats numériques des 6 variables de réponse (H_1 , H_2 , S , B_1 , B_2 , B_3) dans un fichier de sortie et ce pour chacun des incrémentations de temps de résidence (extension .SOR).

2.2 Construction d'une simulation avec SANCHO

Pour chacune des simulations effectuées avec SANCHO une feuille de simulation doit être remplie. Cette feuille est présentée en Annexe 1. On y indique les valeurs numériques des paramètres nécessaires à la simulation. En plus des 19 valeurs de paramètres, on note les informations relatives à la simulation : la date, le nom du fichier de sortie et le nom du modèle utilisé.

L'avantage d'utiliser un fichier de sortie plutôt d'une sortie graphique est double:

1. il permet de confirmer que les paramètres initiaux de la simulation sont bel et bien conformes à ceux désirés;
2. il permet d'exploiter les résultats numériques pour des analyses subséquentes plutôt que d'uniquement observer les tendances des courbes. Cet avantage est nécessaire pour l'analyse statistique qui est proposée.

2.3 Méthode de mesures de certains paramètres du modèle SANCHO

2.3.1 Carbone Organique Dissous Biodégradable (CODB)

La méthode utilisée pour la mesure du CODB consiste à mesurer la différence de Carbone Organique Dissous (COD) à l'aide d'un analyseur de carbone . La différence entre le COD initial et le COD final après une incubation de 30 jours à 20°C en présence de bactéries indigènes tel que la méthode de Servais-Billen (Servais et al. 1989) le décrit. L'erreur de mesure sur ce paramètre est de 0.1 mg C/l (MacLean 1995). Une fois le CODB total connu, il peut être subdivisé en ses trois composantes : H₁, H₂ et S.

2.3.2 Capacité maximale de fixation biologique (SPs)

Un examen par microscopie à balayage des parois d'une conduite indique que le biofilm présent est disparate, discontinu et non-homogène. Les mesures effectuées par (Servais et al. 1992a) sur le réseau de la banlieue parisienne montrent une fixation biologique variant entre 0.2 et 0.5 $\mu\text{gC/cm}^2$. Dans les réseaux de distribution étudiés, le biofilm bactérien ne dépasse jamais la valeur de 2 $\mu\text{g C/cm}^2$. Cette valeur a été retenue pour le terme SPs.

2.3.3 Rendement de croissance (Y)

Le rendement de croissance (Y) est le rapport entre la biomasse bactérienne produite et l'utilisation de la matière organique. La technique consiste à stériliser l'échantillon d'eau par une filtration sur membrane de 0.2 μm et ensuite la réensemencer par un inoculum d'eau filtrée. La pente du graphique de la production bactérienne en fonction de la mesure d'utilisation totale du CODB permet d'estimer le rendement de croissance. L'erreur expérimentale de cette mesure est reliée à la qualité de la droite de régression.

2.3.4 Taux de mortalité bactérienne (k_d)

La mortalité bactérienne (k_d) est la mesure de la décroissance de la biomasse bactérienne en fonction du temps. La méthode utilisée par (Servais et al. 1989), dans sa description du modèle SANCHO, consiste à suivre la disparition d'un marqueur radioactif dans l'ADN. Ce marqueur utilisé est la thymidine tritiée. La disparition de ce marqueur radioactif dans l'ADN est attribué à la mort des cellules qui le contenait.

2.3.5 Taux maximal d'utilisation de substrats directs (B_{max})

Le terme B_{max} représente le taux maximal d'utilisation des substrats directs. La détermination expérimentale de B_{max} est effectuée en évaluant la production de biomasse bactérienne à partir du taux de synthèse du DNA bactérien. Le taux de synthèse du DNA est estimé par la mesure du taux d'incorporation de thymidine dans le DNA. L'ajout de différentes doses de glucose marqué à la thymidine dans des échantillons permet de déterminer la valeur de B_{max} . La qualité de cette valeur dépend de la dispersion de la droite de régression.

2.4 Méthodologie employée

La planification statistique d'expérience ou le design d'expérience (en anglais DOE pour Design Of Experiments) représente un ensemble de méthodes, stratégies et outils appliqués au processus expérimental. Considérons un système général tel qu'illustre à la Figure 2.1. Les informations présentées entre parenthèses indiquent les résultats pour notre cas particulier.

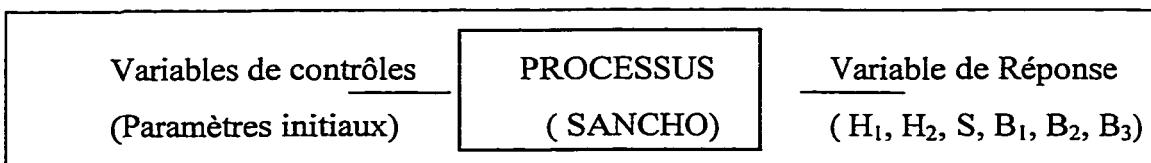


Figure 2.1 Schéma simplifié du système SANCHO

Les variables de contrôle de ce système sont les paramètres initiaux qu'il faut introduire dans le logiciel pour effectuer une simulation. Le processus représente le logiciel SANCHO avec ses équations mathématiques et ses fonctions microbiologiques. Finalement, les variables de réponse sont les réponses fournies par le modèle : les bactéries quelles soit Fixées (B_1), Adsorbées (B_2) et Libres (B_3) et les substrats rapidement biodégradables (H_1), lentement biodégradables (H_2) et directement assimilables (S). L'objectif de la planification d'expérience est de faire varier de façon planifiée les intrants (variables de contrôle) dans le but d'en évaluer les effets sur les extrants (variables de réponse) et d'en extraire un maximum connaissance quant au processus étudié.

La méthode traditionnelle voudrait qu'on fasse varier une seule variable à la fois. Dans cette approche, toutes les variables sont maintenues constantes et une seule varie selon un arrangement prédéterminé. Le succès de cette méthode dépend beaucoup de l'expertise et de l'intuition de l'expérimentateur. De plus, le nombre d'essais nécessaires pour obtenir les informations requises ne peut être déterminé à l'avance.

Au contraire, le design d'expérience préconise de faire varier de façon conjointe plusieurs variables de manière déterminée. Le design d'expérience permet, en effet, d'obtenir une estimation plus précise des effets principaux, d'explorer les impacts des interactions entre divers facteurs et d'étendre la zone de validité des conclusions retenues (Box and Draper 1987; Cox 1992).

Notre objectif particulier lors de ce projet est donc d'utiliser ces méthodes afin de dégager lequel ou lesquels des paramètres initiaux qui induisent le plus de variabilité dans les variables de réponse du modèle SANCHO.

Comme nous l'avons déjà dit, la planification d'expérience représente un amalgame de méthodes et stratégies qui mènent à une meilleure compréhension du système. Plusieurs méthodes sont présentées dans la littérature. On note entre autres: l'analyse de la variance (ANOVA), l'analyse de régression, la méthode des Moindres Carrées et divers plans factoriels.

Sans discuter en détail chacune de ces approches, il apparaît que deux design seront étudiés : un plan factoriel partiel (PFP) et une analyse selon les plan croisés de Taguchi (PCT).

Les plans factoriels permettent, selon (Box and Draper 1987) :

- de faciliter la comparaison et la critique du modèle;
- d'obtenir des estimés de paramètres avec une variance constante;
- d'être obtenus à partir de calculs simples.

Les plans croisés ont été introduits par le statisticien japonais Genichi Taguchi afin de comprendre la variabilité dans un processus industriel. Ce design combine les avantages d'un plan factoriel partiel tout en ajoutant une emphase sur la définition de la variabilité des variables de réponses.

Chapitre 3. Analyse de la variance du modèle à l'aide d'un plan factoriel partiel

Dans le cas qui nous intéresse, un plan factoriel complet requiert $2^{19} = 524\ 288$ simulations. Un tel design est à la fois long à exécuter et peu profitable quant à l'information réelle qu'il peut amener. En effet, beaucoup de redondances et de dédoublements sont introduits par ce plan exhaustif. Un plan factoriel partiel, de résolution adéquate, est beaucoup plus économique quant au nombre de simulations requises tout en fournissant les facteurs d'importance.

3.1 Construction du plan factoriel partiel

Des analyses antérieures ont montré qu'un plan de résolution III n'est pas suffisant pour départager correctement les chaînes d'interactions créées. Rappelons que la résolution d'un plan est le pouvoir discriminant des effets. Par exemple, un plan de résolution IV permet de départager la chaîne des effets confondus de 4 éléments A * B * C * D ainsi que toutes les chaînes contenant moins que 4 éléments. Évidemment, toutes les chaînes d'interactions contenant plus que 4 éléments demeurent partiellement confondues.

Nous avons donc utilisé la procédure FACTEX du logiciel SAS afin de produire divers plans de résolution IV. Ces plans contiennent soit 64, 128 ou 256 simulations avec divers degrés de complexité. Les raisons et contraintes qui nous on amenés à choisir un plan de 64 essais sont expliqués dans l'article présenté dans ce chapitre. Le design de 64 essais est présenté dans cet article.

3.1.1 Modèle statistique

Le modèle statistique sous-jacent peut être exprimé très succinctement de façon matricielle. On a donc pour une variable de réponse :

$$Y_{64 \times 1} = X_{64 \times 62} \beta_{62 \times 1} \quad (3.1)$$

où :

Y = Représente la matrice des variables de réponses. C'est une matrice de 64 lignes et 1 colonne. Elle est construite à partir des résultats numériques produits par SANCHO pour chacune des six variables de réponse.

X = Représente la matrice du plan expérimental. Cette matrice orthogonale est entièrement définie lorsque le choix du plan est fait. Elle est codée uniquement avec des +1 et des -1, selon la modalité de chaque paramètre. C'est une matrice de 64 lignes par 62 colonnes. Les 62 colonnes sont construites à partir des 19 colonnes des effets principaux et des 43 colonnes d'effets confondus.

\square = Représente la matrice des demi-effets à estimer. C'est une matrice de 62 lignes et 1 colonne. Encore une fois, elle est construite à partir des 19 effets principaux et 43 effets confondus.

En multipliant par l'inverse, on isole la matrice $\square \square$

$$X^{-1}Y = X^{-1}X\beta \quad (3.2)$$

$$\beta = (X^{-1}X)^{-1}X^{-1}Y \quad (3.3)$$

Or si X est orthogonal ($X^{-1}X$) devient une matrice diagonale $D = \text{diag}(d_1, d_2, \dots, d_j, \dots, d_{64})$.

Ainsi,

$$\beta = D^{-1}X^{-1}Y \quad (3.4)$$

De plus, on a que chacun des x_{ij} de la matrice X est codé en ± 1 . La matrice étant orthogonale, on a que $\sum_{i=1}^k x_{ij} = 0$ et donc que d_j est égal à une constante k .

Ainsi,

$$\beta = \frac{1}{k} \sum_{i=1}^k x_{ji} y_i = \frac{1}{k} \sum_{i=1}^k (\pm 1) y_i \quad (3.5)$$

L'estimation des demi-effets \square_j est donc uniquement une combinaison linéaire des réponses calculées y_i .

3.1.2 Analyse de la variabilité

La variabilité totale d'une réponse peut-être décomposée en deux termes distincts : la variabilité expliquée par le modèle et la variabilité expliquée par l'erreur résiduelle.

Ainsi

$$SS_{\text{Totale}} = \sum_{i=1}^k (y_i - \bar{y})^2 = SS_{\text{Modèle}} + SS_{\text{Résiduelle}} \quad (3.6)$$

$$SS_{\text{Modèle}} = \sum_{i=1}^k (\hat{y} - \bar{y})^2 \quad (3.7)$$

$$SS_{\text{Résiduelle}} = \sum_{i=1}^k (\hat{y} - y_i)^2 \quad (3.8)$$

où

y_i = valeur numérique de la réponse

\bar{y} = valeur moyenne des y_i

\hat{y} = valeur estimée du paramètre y_i

L'orthogonalité de la matrice X est très importante pour l'analyse de la variabilité du système, en l'occurrence le modèle SANCHO. En général, si la matrice X n'est pas orthogonale, la décomposition de la variabilité associée au modèle ($SS_{\text{Modèle}}$) est liée à l'ordre d'introduction des paramètres X_i . Ainsi ,

$$SS_{\text{Modèle}} = SS(X1) + SS(X2/X1) \neq SS(X2) + SS(X1/X2) \quad (3.9)$$

Dans notre cas, avec une matrice orthogonale, la décomposition de la somme des carrés est unique. De plus, chacun des termes est estimable de façon indépendante des autres.

$$SS_{\text{Modèle}} = SS(X_1) + SS(X_2) + SS(X_3) + \dots + SS(X_{19}) \quad (3.10)$$

Il peut être démontré que dans ce cas,

$$SS(X_i) = SS(\square_j) = nk \square_j^2 \quad (3.11)$$

et que $SS_{\text{Modèle}} = \sum_{i=1}^{62} SS(X_j) = n \cdot k \cdot \sum_{i=1}^p \beta_i^2 \quad (3.12)$

Ainsi, le pourcentage de la variabilité du modèle qui est expliquée par un seul paramètre est donc :

$$X_j = 100 \cdot \frac{\beta_j^2}{\sum_{i=1}^{62} \beta_i^2} \quad (3.13)$$

On obtient donc, à la suite de ce calcul, 62 estimés de demi-effets estimés. En ordonnant de façon décroissante ces effets, on arrive à construire un diagramme de Pareto départageant les paramètres importants de ceux qui le sont moins. L'article présenté ci-après, soumis à la revue Water Research, présente les résultats obtenus lors de l'analyse selon la méthode du design d'expérience. Cette méthode permet d'explorer l'espace expérimental le plus complètement possible en fixant les modalités minimales et maximales des paramètres et en les testant à ces points.

Sensitivity Analysis of the SANCHO model:**Identification of critical parameters using Design of Experiments (Part 1)**

John Cigana ¹, Michèle Prévost ¹, Bernard Clément ², and Pierre Servais ³

¹ Department of Civil Engineering, Ecole Polytechnique of Montréal, C.P. 6079 , Succ. Centre-Ville, Montréal, Québec, Canada, H3C 3A7,

² Department of Applied Mathematics and Industrial Engineering, Ecole Polytechnique of Montréal, C.P. 6079 , Succ. Centre-Ville, Montréal, Québec, Canada H3C 3A7

³ Researcher, Groupe de Microbiologie des Milieux Aquatiques, Université Libre de Bruxelles, Campus de la Plaine, C.P. 221 , Boulevard du Triomphe, B-1050, Bruxelles, Belgium

Nomenclature

$e_{1\max}$ Maximum rate of exoenzymatic hydrolysis of macromolecular BDOC of class 1,

h^{-1}

k_{h1} Half-saturation constant of exoenzymatic hydrolysis of class 1 compounds, mg

C/L

$e_{2\max}$ Maximum rate of exoenzymatic of hydrolysis of macromolecular BDOC of class

2, h^{-1}

k_{h2} Half-saturation constant of exoenzymatic hydrolysis of class 2 compounds, mg

C/L

B_{\max} Maximum rate of direct substrates S uptake, h^{-1}

K_S Half-saturation constant of direct substrates S, mg C/L

Y Bacterial growth yield, -

\square Fraction of bacterial biomass recycled as H_1 , -

k_d Mortality rate, h^{-1}

k_{ads} Adsorption rate constant of bacteria onto the support, $\square \text{gC/ml h}^{-1}$

k_{des} Desorption rate constant of bacteria, from the support, h^{-1}

SPs Maximum capacity for 'biological' bacterial fixation, $\square \text{gC/cm}^2$

SBs Maximum capacity for bacterial adsorption defined as a fraction of SPs, %

\square Residence time of analysis, h

$dtsej$ Incremental step of time, h

- H₁₀ Initial concentration of macromolecular organic matter of lability class 1 (Quickly hydrolysable), mg C/L
- H₂₀ Initial concentration of macromolecular organic matter of lability class 2 (Slowly hydrolysable), mg C/L
- S₀ Initial concentration of direct substrates, mg C/L
- B₃₀ Initial free bacteria in the liquid phase, $\mu\text{g C/L}$
- H₁ Concentration of macromolecular organic matter of lability class 1 (Quickly hydrolysable), mg C/L
- H₂ Concentration of macromolecular organic matter of lability class 2 (Slowly hydrolysable), mg C/L
- S Concentration of direct substrates, mg C/L
- B₁ Biologically attached bacterial biomass, $\mu\text{g C/cm}^2$
- B₂ Reversibly adsorbed bacterial biomass, $\mu\text{g C/cm}^2$
- B₃ Free bacteria in the liquid phase, $\mu\text{g C/L}$

Abstract

SANCHO is a time dependent deterministic model, developed at the ULB (Université Libre de Bruxelles) that simulates bacterial dynamics, free chlorine decay and biodegradable organic matter (BOM) consumption in water distribution systems (DS). The software version of the SANCHO model is used in the drinking water industry to predict bacterial regrowth in pipes and to assist utilities in complying with drinking water regulations. The objective of this work was to determine the key parameters of the SANCHO model and thus identify the critical parameters of bacterial regrowth with respect to bacterial biomass (fixed, adsorbed and free bacteria) and BDOC components (H_1 , H_2 and S). Our results indicate that five parameters have an undeniable influence on the model response. These parameters are : bacterial mortality (k_d), the maximum capacity for bacterial fixation (SP_s), the bacterial yield (Y), the initial concentration of direct substrates (S_0) and the maximum rate of uptake of direct substrates (B_{max}).

Introduction

The understanding of the causes of changes in water quality in distribution system (DS) is a complex task. Water quality can be estimated by its physical, chemical and biological properties. One of these is the presence of bacteria in the treated and distributed water. Relevant local authorities set up the different standards. Indicator organisms, such as coliform bacteria, are commonly used to monitor microbiological water quality of drinking water. A recent study as shown that a high proportion of the 164 American water utilities surveyed had experienced recurring coliform episodes (Smith et al. 1990). Although

not a new problem (AWWA 1930), bacterial regrowth in DS is of rising concern in light of recent studies and multiple non compliance events. Bacterial regrowth may be linked to an increase in gastroenteric illnesses (Payment et al. 1991) and to the development of nuisance organisms like *Asellus* and *Naïs* in the DS (AWWA, 1995). The classic response to a regrowth event is to increase the disinfectant dosage (Camper, 1995) or to change disinfectant type (LeChevallier et al. 1996). Not only this has sometimes proven to be ineffective in controlling coliform regrowth (Camper 1995) but may give rise to an undesirable increase in the concentration of disinfection by-products. An alternative approach resides in reducing the amount of substrates required for bacterial growth (the biodegradable organic matter or BOM) in the treated water. Several studies (Prévost et al. 1997b; Servais et al. 1995) have shown that this was an effective way to control bacterial regrowth in DS.

The SANCHO model

In order to predict the respective role of BOM and chlorine in controlling bacterial regrowth in DS, a time-dependent static model, the SANCHO model, was developed at the Université Libre de Bruxelles (Servais et al. 1995). The SANCHO model simulates bacterial dynamics of growth and mortality, free chlorine decay and biodegradable organic matter utilisation in a water distribution system. The reader may find the processes of the SANCHO model described in greater length in (Servais et al. 1995); and

(Billen et al. 1992). An overview of the model is presented in Figure 1 showing the six response variables, their symbols and their units.

A group of four processes are taken into account by the SANCHO model as shown in Figure 1 :

1. the interaction between organic matter and bacteria. This includes the exoenzymatic hydrolysis of biodegradable organic matter into direct substrates by bacteria ;
2. the growth of free and fixed bacteria on the hydrolysis products and bacterial mortality which releases organic matter ;
3. the reversible adsorption-desorption of bacteria onto the inner pipe surface and bacterial fixation onto pipe surfaces ;
4. the chemical consumption of free chlorine and the impact of free bacteria on free and fixed bacterial activity and mortality;

The nutrients used by bacteria in the SANCHO model are described as biodegradable dissolved organic carbon or BDOC. BDOC is divided in three classes : H1, the rapidly hydrolysable polymeric carbon, H2, the slowly hydrolysable polymeric carbon and S, the directly utilisable monomeric carbon. In this model, BDOC is thus represented as the sum of H1, H2 and S.

The SANCHO model has been tested in Europe and North America with some success. There is a general agreement between observed and predicted suspended biomass values and BDOC consumption using the SANCHO model (Laurent et al. 1997) ; Servais, 1995 #1353]. The experimental results used for this testing were obtained from different sources of water, treatment plants and over a wide range of temperature and on varying pipe diameters and pipe materials (Laurent et al. 1997; Servais et al. 1995). The model and its software version are rather complex and their validation is on-going. Further improvements in the predictive ability of the model can only come from a deeper understanding of its components, its strengths, its weaknesses and its limits of applicability. Indeed, if we hypothesise that SANCHO correctly depicts bacterial dynamics, then critical parameters of the model are also critical to regrowth in DS. To move in that direction, it appears necessary to complete a sensitivity analysis.

A simplified version of the model was used for the purpose of conducting the sensitivity analyses. Two major effects were not analysed : temperature and chlorine. Using a simplified version allows for in depth testing of the equations predicting bacterial regrowth and substrate utilisation. The temperature was held constant at 20°C and the theoretical diameter of the network pipe remained at 0.15 m for all the simulations. This maximises the Surface to Volume (S/V) ratio thus yielding optimum growth conditions. Finally, the impact of chlorine was not investigated. These three simplifications were made to offer a simple situation in which the model could be tested for the prediction of bacterial growth.

This simplified version of the model uses 19 initial parameters to accomplish the task. Table 1 lists the 19 necessary input parameters given to the SANCHO model. Given an initial set of parameters, the model will predict the evolution of bacterial biomass and BDOC concentrations for a given time of residence. The software version of the SANCHO model is used in the drinking water industry in order to predict bacterial regrowth in networks. Since key parameters may change with residence time, two significative residence time where chosen: 3 and 8 hours. This choice was based on the several biomass curves yielded by the SANCHO model. It is between these two residence time that most of the bacterial activity was observed in the absence of chlorine and in small diameter pipes.

The objective of this research was to gain a better understanding of the sensibility of the SANCHO model. In this outlook, we ask:

- What are the critical parameters of the SANCHO model?
- Is the model unduly sensitive to a few parameters ?

To answer these questions, a statistical analysis using a design of experiment approach was carried out.

Experimental Design and Methods

Mathematical literature is chock full of methods and strategies in order to analyse any set of data and obtain its critical parameters (Box and Draper 1987; Cox 1992). We have opted in our work for the fractional factorial plans of experiments with a design of experiments (DOE) approach. The advantages of DOE are many: it allows multitudes of comparisons to be made and so facilitates model criticism; it also gives rise to simple calculations and provides highly efficient estimates of parameters (Box and Draper 1987). The ultimate choice of a design will result from the trade-off between the total number of runs necessary to obtain a valid result and the additional intricacy provided by the additional runs.

Experimental Design of the SANCHO Model

One might be tempted, in analysing a particular model, to vary only one of the parameters single-handedly and study the fluctuation of the response variable. It would be a daunting task to study this model with a trial and error approach. Indeed, a complete two-level factorial plan for our 19 parameters would require 2^{19} simulations which is 524 288 simulations for a two-level design. DOE, on the other hand, permits the testing of many factors simultaneously. A special problem rises from the large number of initial parameters needed to properly run the model. By fractionating the factorial plan, we can limit the total number of experiments necessary. Initially, several factorial designs were explored. These preliminary analysis had shown that a level III resolution design of 32 runs was not sufficient to correctly reveal each main effect. A minimal resolution of level IV was therefore necessary to isolate the main effects from the interaction effects. An

orthogonal, two-level, partial factorial design of 64 experiments was chosen to obtain a resolution IV experimental design with a minimal number of simulations. The characteristics and confounding rules of the level IV orthogonal factorial design are presented in Table 2.

Method

Each of the 19 initial parameters in Table 1 is attributed a value of either +1 or -1. A value of +1 indicates that the parameter takes the maximum value presented in Table 1. Likewise, a value of -1 represents the minimal value of the range. The values used for the maximum and minimum modalities were chosen based on the published values used with the SANCHO model (Laurent and Servais 1995) ; (Laurent et al. 1997); (Servais et al. 1995). An arbitrary range around these published values was then fixed but within the range of variability which could be expected in drinking water systems. This choice was made in order to study a larger fraction of the experimental range. A complete simulation is obtained when all 19 initial parameters are introduced in the model with their modalities. For example, for the first run of the 64 run design, X9 has a value of +1. This means that a numerical value of 0.08 h^{-1} is introduced in the SANCHO model for the parameter X9. So on and so forth for the remaining 18 parameters. The results of the simulations are written in an output file. From these results, further numerical analysis can be done.

Statistical analysis of the 64 run design

Using the data produced from the 64 simulations runs, it is possible to perform a linear adjustment of the variability of the regression model according to the main and interaction effects. The underlying model describing the variability of each response variable is:

$$Y = \beta_0 + \sum_{j=1}^{19} \beta_j X_j + \sum_{k=1}^{43} \gamma_k Z_k + \epsilon \quad (\text{Eqn. 3.14})$$

where :

Y : Numerical value of the response variable as predicted by the SANCHO model

β_0 : General constant of the regression model

X_j : ± 1 coding variable

Z_k : Chain of 2-factor interactions of $X_j * X_{j'}$ ($j \neq j'$)

β_j : Half-effect estimate of main effect

γ_k : Half-effect estimate of interaction effect

ϵ : Error term

Closer analysis of Equation 1 shows that the variability of the model for a given response variable can be attributed to either a main effect or to an interaction effect. Indeed, we see that there are 19 main effects intervening, one for each of the initial parameters, and 43

confounded interaction effects representing all 171 interactions possible. The Experimental design module of the STATISTICA software (StatSoft 1996) was used to analyse the data at both 3 and 8 hours of residence time.

Importance of orthogonality and interpretation

The orthogonality of the factorial design is of prime importance. Indeed, if we lack orthogonality, the decomposition of the variability of each parameter is linked to the order of introduction of the given parameter, as shown in Equation 2.

$$SS_{\text{model}} = SS(X_1) + SS(X_2/X_1) + \dots + SS(X_n/X_1) \neq SS(X_2) + SS(X_2/X_1) + \dots + SS(X_n/X_1) \quad (\text{Eqn. 3.15})$$

where $SS(X_i)$ = the sum of square of the parameter X_i

and $SS(X_j/X_i)$ = the sum of square of the parameter X_j knowing X_i

With orthogonality, the total variability of the model is directly the sum of the individual variabilities.

$$SS_{\text{model}} = SS(X_1) + SS(X_2) + SS(X_3) + \dots + SS(X_n) = \sum_{i=1}^n SS(X_i) \quad (\text{Eqn. 3.16})$$

This way, the percentage of variability explained by a single parameter can be easily calculated:

$$\text{Percentage of the total variability explained by parameter } X = \frac{SS(X_i)}{\sum_{i=1}^n SS(X_i)} \quad (\text{Eqn. 3.17})$$

The resulting percentage of explanation ensuing for each parameter can be sorted and ranked in a Pareto graph. A Pareto graph is ... The importance of a parameter is measured by the relative explanatory power it has among all effects estimates. The STATISTICA (StatSoft 1996) analysis of data yields a list of 19 half-effect estimates of main effects (\square_j) and 43 half-effect estimates of interaction effects (\square_k). These 62 estimates are then sorted in descending order. Resulting Pareto graphs for the six response variables are shown in Figures 2 to 7. Using Table 1, each of the estimates are then related back to their biological signification. Usually, a maximum cumulative percentage of explanation is decided on arbitrarily, for example 70% of the variability. Main and interaction effects would then be added until that percentage is reached. We decided to pursue a less arbitrary method of explaining variability. The percentages of explanation were added until the interaction effects no longer contained preceding main effects. In other words, the first interaction effect not composed of main effects of greater importance would halt the summation of the explanatory power which would be further commented on.

Results and discussion

Table 3 summarises the percentage of variability explained by the critical initial parameters of the SANCHO model for the six response variables and the critical parameters for each response variable.

These six response variables can be divided in two categories : biomass and substrate terms. The first three response variables B_1 , B_2 and B_3 are all terms relating to the presence of bacterial biomass, be it fixed, adsorbed or free, in DS. The last three response variables H_1 , H_2 and S all track the presence of nutrient substrates available for bacteria to the water.

B1- Biologically attached bacterial biomass (fixed biomass)

Figure 2 represents the main and interaction effects sorted in decreasing order for the response variable B_1 (fixed bacteria) at residence times of 3 and 8 hours. Closer analysis of the Pareto graph at 3 hours shows that a small number of effects account for a large amount of the explained variability of the SANCHO model. Very quickly, the explanatory power of the effects drops off reaching values close to zero. From this Pareto graph, we see that bacterial mortality rate (or k_d) (parameter X9 from Table 1) explains 20.3% of the variability of the SANCHO model by itself. Next in line is the maximum capacity for biological fixation (or SPs) with 18.4% of explanatory power. The third term of influence is a confounded interaction term of the first two main effects (k_d and SPs). The fourth most important term is the initial value of directly utilisable substrates (S_0)

which explains 5.9% of the total variability. The next term is not the confounding effect of any of the preceding main effects. We therefore would have to halt the process of adding terms.

From Table 3, we see that 66.0% of the variability of the response variable fixed bacteria can be explained by 6 terms: the three initial parameters of the model k_d , SPs and S_0 and the three interactions between these three initial parameters (k_d^*SPs , SPs^*S_0 and $k_d^*S_0$). We therefore consider these three initial parameters : k_d (mortality rate), SPs (maximum capacity for biological fixation) and S_0 (initial concentration of directly utilisable substrates) as critical parameters of the SANCHO model for the response variable fixed bacteria at 3 hours of residence time.

From Figure 2, we can also see that parameters k_d and SPs still remain as very influential at 8 hours of residence time, explaining respectively 11 and 8% of the total variability. The third term, bacterial growth yield (Y) also explains 8% of the variability of the response variable. Using the same technique as described earlier to set the critical parameters according to their contribution, we see that 71% of the variability can be explained by the six initial parameters k_d , SPs, Y, S_0 , k_{ads} and k_{des} and their interactions at 8 hours. It can also be noticed that the slope is not as steep as with the Pareto graph at 3 hours of residence time.

Some explanations can be brought forward concerning this difference in the critical parameters between at the two residence time. Earlier in the DS, the mortality rate (k_d) is of prime importance, as are the number of accessible sites for colonisation (SPs) and the initial quantity of substrate present (S_0). At 8 hours, mortality and site accessibility remain important but in a smaller way and the growth yield becomes more important as do the adsorption-desorption rates (k_{ads} , k_{des}).

These results appear to be consistent with results obtained by other researchers. Pipe material and biodegradable organic matter were already surmised by (Camper, 1993 #1904, Camper, 1995 #1961 and Prévost, 1997 #2001]. Indeed, a biofilm is an accumulation of microorganisms on pipe surfaces, which is our definition of fixed bacteria. Pipe material in the SANCHO model is taken into account by the parameter SPs. Our results suggest that pipe material, in the form of the maximum capacity for biological fixation or SPs, greatly influence the variability of fixed bacteria. Likewise, organic matter is represented by the SANCHO model by BDOC. Our results point out that only one of the components of BDOC, the initial concentration of direct utilisable substrates (S_0), stands out as a major factor. Camper (1995) did not identify mortality as a critical parameter of the occurrence of fixed biomass although a systematic sensitivity analysis was not performed. Our results indicate that the impact of the mortality rate on fixed bacteria is paramount. Indeed, the mortality rate is actually first at both 3 and 8 hours with respectively 20.3% and 11% of the total variability. Fixed biomass is therefore mostly influenced by the value of three initial parameters: k_d (mortality rate), SPs (maximum

capacity for biological fixation) and S_0 (the initial concentration of directly utilisable substrates). Figure 3 illustrates the importance of bacterial mortality on the density of fixed biomass on the pipe walls. The y-axis of this figure represents the amount of fixed bacteria, in $\mu\text{gC}/\text{cm}^2$, as predicted by the SANCHO model whereas the x-axis represents the residence time, in hours, of the drinking water in a 0.15m diameter pipe. This figure was produced by attributing eight increasing values to the bacterial mortality term. The bacterial mortality term (k_d) ranged from 0.01 h^{-1} to 0.08 h^{-1} with 0.01 h^{-1} increments. All other eighteen parameters were kept constant to the average value of the range shown in Table 1. This figure indicates the direct influence of bacterial mortality on fixed biomass when average conditions are placed on the other parameters. A great difference in the value of fixed bacteria can be noticed from Figure 3. At 1 hour of residence time and with $k_d = 0.01 \text{ h}^{-1}$, fixed bacteria reaches $0.392 \mu\text{g C}/\text{cm}^2$ whereas with bacterial mortality set at 0.08 h^{-1} , fixed bacteria drops sharply to $0.195 \mu\text{g C}/\text{cm}^2$. A twofold difference in the fixed bacteria occurs at this residence time of 1 hour. We nonetheless notice that at greater residence times, all eighth simulations converge toward the same value of fixed biomass, around $0.015 \mu\text{g C}/\text{cm}^2$, except for the first two simulations where more time is required to reach this plateau. This figure pinpoints the direct influence of bacterial mortality on the amount of fixed bacteria found in the DS. The lowest mortality rate will permit the generation of twice the fixed bacteria, all other parameters being equal, than the highest mortality rate. A complete analysis of the influence of critical parameters on the three biomass terms can be found in (Cigana et al. 1997). We also notice from Figure 2 that parameters of influence vary according to the residence time. Indeed, we see that

effects of small importance at 3 hours of residence time may gain influence at 8 hours of residence time. Conversely, important effects at 3 hours of residence time may loose importance at 8 hours of residence time. This also substantiates our hypothesis that the influence and importance of effects vary with the time of residence.

B2 - Reversibly adsorbed bacterial biomass (adsorbed biomass)

Figure 4 shows the sorted main and interaction effects in a Pareto graph for the adsorbed biomass at a residence time of 3 and 8 hours. Adsorbed bacteria represent bacteria that are reversibly attached to the surface where several attachment-detachment cycles are possible. They represent, in mass, a very small amount of the total biomass present in the pipe after the initial colonisation, ranging to less than 5% of the total bacterial biomass. The most influential parameter is the maximum capacity for biological fixation or SPs with 5.6% of the explained variability. The number of adsorbed bacteria in the SANCHO model is determined from the Langmuir isotherm adjusted to take into consideration the amount of sites remaining. One would expect that the parameters describing the isotherm and influencing the number of remaining adsorption sites (SBs) would be the most influential. However, the total number of sites SPs came out as the most critical parameter. This may be explained by the fact that as the number of sites occupied by fixed bacteria increases, the number of sites available for adsorption will be reduced. Table 3 shows that about 40% of the variability of the response variable is explained by six of the 19 initial parameters at 3 hours of residence time: maximum capacity for

biological fixation (SPs), the adsorption constant, (k_{ads}), the fraction of SPs already filled (SBs), the desorption constant, (k_{des}), the concentration of initial directly utilisable substrates, (S_0) and the mortality rate, (k_d), and their interactions. We can note that each individual effect, be it main or interaction, has a smaller overall influence.

The sorted main and interaction effects for the response variable adsorbed bacteria at 8 hours of residence time are overlaid on the same scale. As can be seen on the graph, the same parameters appear critical at 8 hours although in a different order (SPs, k_{ads} , k_d , k_{des} , SBs and S_0). Table 3 indicates that these 6 parameters and their interactions account for close to 40% of the variability of this response variable. We notice in the case of adsorbed bacteria that influential effects follow the same trend and do not vary much at 3 and 8 hours of residence time.

Further in the distribution system, initial colonisation and subsequent free bacteria concentration densities are strongly influenced by adsorption/desorption. The adsorption and desorption processes are governed by four parameters: the adsorption constant, (k_{ads}), the desorption constant (k_{des}), the maximum capacity of biological fixation (SPs) and the utilised fraction of the maximum capacity (SBs). It would be expected to find that these four parameters would be influential to adsorbed bacteria in the DS. Our results show that all four of these parameters are critical in the determination of the variability of adsorbed bacteria. To these four parameters, growth related parameters also contribute to explain the variability of adsorbed bacteria: the maximum capacity for biological fixation

(SPs) and the initial concentration of directly utilisable substrates (S_0). Nevertheless, it is reassuring to find adsorption and desorption terms coupled with the maximum capacity for biological fixation (SPs) are the critical terms for the response variable adsorbed biomass.

B3 - Free bacteria in the liquid phase (free bacteria)

Figure 4 shows the Pareto graph for the response variable free bacteria at 1, 3 and 8 hours of residence time. The first impression we get from this figure is that the effects of influence are different at 3 and 8 hours. The surprising results obtained at 8 hours of residence time warranted some further investigation. To confirm the tendency between results at 3 and 8 hours of residence time, an additional analysis was completed at a residence time of 1 hour. The most influential parameters at those residence times are S_0 , the initial concentration of direct substrates, the bacterial growth yield, Y , and B_{max} , the maximum rate of uptake of direct substrates, as well as their interactions. Table 3 indicates the cumulative percentages of explanation by critical parameters. Results obtained at 8 hours of residence time indicate that all 62 of the effects account for about 1.5% of the total variability at 8 hours of residence time. The total variability explained by these three main effects and their interactions is 24% at 3 hours of residence time and 52% at 1 hour of residence time. The same analysis is performed at 8 hours of residence time. Although in a slightly different order, the same three parameters are critical to the response variable free bacteria at 8 hours of residence time. The percentage of explained variability by the initial parameters is thus very low for the response variable free

bacteria. The SANCHO model has been tested in various full-scale DS including a wide range of residence time, water quality, temperature and pipe material. A very good agreement for predicted and measured free bacteria has been observed in several networks (Laurent and Servais 1995), [Servais, 1995 #1353]. Experimental data from full scale DS suggests that free bacteria are the result of detachment from fixed bacteria (Servais et al. 1992a). The value of free bacteria is overall quite small as compared to the value of fixed biomass, about 5%. These results suggest that the prediction of suspended solids biomass is sensitive to a small number of parameters early in the DS. Later in the DS, as there is an increase in this biomass pool, a shift from growth of suspended bacteria to detachment is observed. No single parameter or group of parameters is critical to the prediction of suspended bacterial levels.

The sharp decrease in our ability to identify critical parameters to predict the biomass of suspended bacteria with increasing residence time may be caused by a number of reasons:

- Early in the DS (at 1 hour of residence time), the number of suspended bacteria is influenced by growth related parameters (Y), by the initial concentration of directly utilisable substrates (S_0) and the ability to use it (B_{max}). This can be understood since the changes in suspended bacterial biomass in the front end of the DS would be dependant upon the establishment of growth (fixed and free biomass) at short period of time.
- Later in DS, once the substrate has been exhausted, the amount of suspended biomass is driven by the detachment and death of fixed bacteria earlier in the DS.

The decrease of the relative importance of the parameters driving the response variable at shorter residence times can be explained by the much more complex network of parameters defining the levels of adsorbed and fixed bacterial biomass.

H1 - Rapidly degradable macromolecular organic matter

Figure 6 shows the Pareto graph for the main and interaction effects for the response variable H_1 (rapidly hydrolysable macromolecular organic matter) at 3 and 8 hours of residence time. We notice that graphs at both residence times follow the same trends. As can be seen on this graph, 67% of the variability of the response variable H_1 can be explained by the initial value of H_1 introduced in the model, H_{10} , at 3 hours of residence time. By adding the different interactions present, 82% of the total variability can be explained by three parameters H_{10} , SP_S and k_d , as seen in Table 3. Nonetheless, H_{10} is overwhelmingly the most critical of the three at 3 hours of residence time. As for the response variable H_1 at 8 hours of residence time, the percentage explained by the initial value of H_1 drops to 31%. Overall, 60% of the total variability is accounted for when adding its interactions as shown in Table 3.

The evolution of substrates is directly influenced by the amount of substrates entering the system. We can conclude that the response variable H_1 found in the DS, is therefore very sensitive to the initial measure of H_1 . The method of measurement of H_1 and its accuracy are therefore of great importance to this response variable.

H2 - Slowly degradable macromolecular organic matter

Figure 7 shows the Pareto graph for the main and interaction effects for the response variable H₂, the slowly degradable macromolecular organic matter at 3 hours and 8 hours of residence time. Trends are clear for the response variable H₂. A very high percentage (98%) of the variability of this response variable is explained by three of the initial parameter at 3 hours as shown in Table 3. The overwhelming parameter of the three is the initial value of slowly biodegradable macromolecular organic matter, H₂₀. We very clearly see on the same graph that 91% of the total variability of this response variable is explained by the initial value of H₂ with small influences by the mortality rate (k_d) and initial direct substrates (S₀). This was to be expected since slowly hydrolysable macromolecular organics are not likely to be degraded by bacteria within the relatively short residence times tested (<8 hours). Again, we see that the response variable H₂ found in the DS is therefore very sensitive to the initial measure of H₂ and that its accurate estimation is critical.

S - Directly utilizable monomeric substrates

Figure 8 shows the Pareto graph for main and interaction effects of the response variable S, the direct substrates concentration at 3 and 8 hours of residence time. Closer inspection of this graph shows that the variability of the response variable S is also very well defined by two critical parameters of the SANCHO model : S₀, the initial value of direct substrates and B_{max}, the maximum rate of uptake for direct substrates. These parameters

explain 78% of the total variability of the response variable at 3 hours as indicated in Table 3. The most influential of these terms remains above all S_0 , the initial value of the parameter direct substrates. Similar trends are observed at 8 hours of residence time. Table 3 shows that 60% of the total variability can be explained by three parameters : S_0 , the initial concentration of direct substrates, B_{max} , the maximum rate of uptake for direct substrates and \square the fraction of dead bacteria recycled as H_1 . The trend observed for S_0 follows that of the two other substrates response variables H_1 and H_2 . The initial value of the response variable is essential in determining the variability of the response variable S . We also note that B_{max} is of some importance in the variability of S , since the utilisation rate of S is determined by the values used for the Michaelis-Menten kinetic model of substrate uptake.

Influence of the time step used for computation

For each simulation, an incremental time step (dtsej) is defined prior to computation. This time step defines the residence time in the successive subsections modelled to pseudo-equilibrium. The influence of this incremental time step in the variability of each response variable was studied. As shown in Table 1, the two values tested were either 0.1 or 1 hour. It appears that the parameter dtsej is actually of very little importance in the variability of the response variables. In the best case, this time step accounts for 2.8% of the total variability of the response variable free bacteria at 3 hours of residence time and 1.8% of the total variability at 8 hours of residence time. For all other response variables, be it at 3 or 8 hours of residence time, the variability associated to the incremental time

step is always smaller than 2%. It appears that valuable simulation time can be gained by accomplishing the simulations at a time step of 1 hour without adversely affecting the variability and validity of the results. This information is important if dynamic modelling is implemented to reflect unsteady fluxes of nutrients.

Conclusion

Design of experiments (DOE) is a valid approach in order to analyse and extract the important parameters of the SANCHO model. DOE holds some potential in assisting engineers in their conceptual work by identifying the critical parameters affecting regrowth in water DS. Not only does it reduce the number of necessary simulations but it allows for the enhancement and ranking of the most important effects. It then becomes possible to:

1. to pursue a more detailed sensitivity analysis using a limited number of parameters;
2. to focus the efforts on the most important analytical procedures, to develop or to apply, in order to quantify these key parameters.

In a general manner, we can conclude from Table 3 that it is easier to identify the sources of variability of substrates in a DS model than those of bacterial biomass, growth and decay in DS being more intrinsically complex phenomena. The SANCHO model simulates the complex phenomenon development of fixed and suspended bacteria in DS. Accordingly, the critical parameters of the model vary with each of the response variables as shown in Table 3. Some of the parameters nonetheless appear to have some influence

over all of the six response variables. These parameters are : the mortality rate, (k_d), the maximum capacity for bacterial fixation, (SPs), the initial value of the monomeric substrate that is directly utilisable (S_0), the maximum rate of uptake of direct substrate (B_{max}) and the bacterial yield (Y). Two parameters are microbiological in nature (k_d , Y), one is a substrate term (S_0), one determines the substrate uptake rate (B_{max}) and the last one a pipe characteristic (SPs). In the SANCHO model, SPs is a global simplification of many complex effects which determine bacterial biomass in model DS. SPs is not measured experimentally but rather fixed at maximum level based on experimental values. SPs may be considered as a simplified concept including the parameters critical to the determination of the fixed biomass: attachment-detachment, shear stress, etc... Estimating the many parameters influencing detachment in pilot scale systems operated under steady state conditions may be feasible but remains a difficult task. Estimating these parameters in complex full-scale DS, operated under varying conditions of shear stress, reflecting drinking water consumption, appears a daunting task. A simplified approach may prove more useful to predict regrowth in full scale DS. In order to reduce and control the variability of the response variables, we must better insure the adequate measure of each of the critical parameters. Further work on the measurement of these parameters appears to be of great importance. This specially holds true for the response variables pertaining to substrates (H_1 , H_2 and S) which are very sensitive to the initial value of each of these respective parameters. This implies that proper characterisation of BDOC is paramount to valid modelling results since it defines the flux of nutrients and the growth of fixed and suspended bacteria. The partitioning of BDOC into H_1 , H_2 and S

is done using the Servais-Billen method (Servais et al. 1989). This procedure is tedious and would benefit from automation of the biovolume measurement. Significant variations in the seasonal composition of BDOC have been reported by Servais et al. (1994) and Laurent et al. (1997). This pinpoints to the importance of conducting several characterizations of a water source before relying on model predictions. An alternative solution may be to test a range of BDOC values when completing simulations.

Finally, the SANCHO model can become an important decision making tool for water distribution systems. By better understanding its sources of variability, the SANCHO model may yield more accurate predictions. Future work on the SANCHO model will focus on obtaining confirmation of these results with a second, unrelated method.

References

AWWA. (1930). Committee on Water Supply, Bacterial aftergrowth in distribution systems. American Journal of Publications Health, 20, 485.

Billen, G., Servais, P., Bouillot, P. et Ventresque, C. (1992). «Functioning of Biological Filters Used in Drinking Water Treatment-the Chabrol Model.» J Water SRT - Aqua, 231-241.

Box, G. et Draper, N. (1987). Empirical model-building and response surfaces, John Wiley & Sons, New York. 699 p.

Camper, A. K. (1995). Factors influencing biofilm growth in drinking water distribution systems, Ph.D., Montana State University, Bozeman, Montana, U.S.A.

Cox, D. R. (1992). Planning of experiments, , New York. 300 p.

Laurent, P. (1995). Colonisation bactérienne des réseaux de distribution: étude expérimentale et modélisation, Ph.D., Université Libre de Bruxelles, Bruxelles.

Laurent, P., Servais, P., Prévost, M., Gatel, D. et Clément, B. (1997). Testing the SANCHO model on distribution systems. Journal of American Water Works Association, 89(7), 92-103.

Payment, P., Franco, E., Richardson, L. et Siemiatycki, J. (1991). Gastrointestinal Health Effects Associated with the Consumption of Drinking Water Produced by Point-of-Use Domestic Reverse-Osmosis Filtration Units. Applied and Environmental Microbiology, 57(4). 945-948.

Servais, P., Anzil, A. et Ventresque, C. (1989). Simple Method for Determination of Biodegradable Dissolved Organic Carbon in Water. Applied and Environmental Microbiology, 55(10). 2732-2734.

Servais, P., Billen, G., Laurent, P., Lévi, Y. et Randon, G. (1992). Studies of BDOC and Bacterial Dynamics in the Drinking WAtter Distribution System of the Northern Parisian Suburbs. Revue des sciences de l'eau, 5(Spécial COA), 69-89.

Servais, P., Laurent, P., Billen, G. et Gatel, D. (1995). Development of a model of BDOC and bacterial biomass fluctuations in distribution systems. Revue des sciences de l'eau, 8, 427-462.

Smith, D. B., Hess, A. F. et Hubbs, S. A. (1990). Survey of distribution system coliform occurrence in the United States. Journal of American Water Works Association, San Diego.

StatSoft, I. (1996). Statistica for Windows, , Tulsa, OK. p.

Tableau 3.1 List of the parameters of the simplified version of the SANCHO model and range of values used for the two level analysis.

No	Desig	Symb	Name of parameter	Units	Range
1	X1	e_{1ma}	Maximum rate of exoenzymatic of		0.4 - 1.5
2	X2	k_{h1}	Half-saturation constant of class 1	mg C/l	0.1 - 0.5
3	X3	e_{2ma}	Maximum rate of exoenzymatic of		0.1 - 0.4
4	X4	k_{h2}	Half-saturation constant of class 2	mg C/l	1.0 - 4.0
5	X5	B_{max}	Maximum rate of substrate uptake of		0.3 - 2.0
6	X6	K_s	Half-saturation constant of direct	mg C/l	0.02 -
7	X7	Y	Bacterial yield	-	0.1 - 0.3
8	X8	\square	Fraction of bacterial biomass	-	0.75 -
9	X9	k_d	Mortality rate	h^{-1}	0.01 -
10	X10	k_{ads}	Adsorption rate constant of bacteria		0.6 - 3.0
11	X11	k_{des}	Desorption rate constant of bacteria		0.2 - 3.0
12	X12	SPs	Maximum capacity for 'biological'		0.2 - 3.0
13	X13	SBs	Fraction of X12 i.e. SBs of the	%	10 - 95
14	X14	\square	Time of analysis	h	3 - 8
15	X15	H_{10}	Macromolecular organic matter of	mg C/l	0.02 -
16	X16	H_{20}	Macromolecular organic matter of	mg C/l	0.02 -
17	X17	S_0	Monomeric substrate that is directly	mg C/l	0.02 -
18	X18	B_{30}	Free bacteria in the liquid phase	$\square g C/l$	0.1 - 10
19	X19	$dtsej$	Incremental step of time	h	0.1 - 1

Tableau 3.2 Characteristics and confounding rules of the 64 run design

DESIGN	Factor confounding rules
	$X_7 = X_1 X_2 X_3 X_4 X_5 X_6$
	$X_8 = X_3 X_4 X_5 X_6$
	$X_9 = X_2 X_4 X_5 X_6$
2^{19-13} IV	$X_{10} = X_1 X_4 X_5 X_6$
	$X_{11} = X_2 X_3 X_5 X_6$
64 runs	$X_{12} = X_1 X_3 X_5 X_6$
	$X_{13} = X_2 X_3 X_4 X_6$
level IV	$X_{14} = X_2 X_3 X_4 X_5$
resolution	$X_{15} = X_1 X_2 X_4$
	$X_{16} = X_1 X_3 X_4 X_6$
	$X_{17} = X_1 X_2 X_6$
	$X_{18} = X_1 X_3 X_4 X_5$
	$X_{19} = X_1 X_2 X_5$

Tableau 3.3 Percentage of variability explained for each response variable

Response	Residence	Cumulative %	Critical parameters for each response
Variable	time	of variabilities	variable in order of importance
B₁	3h	66.0	kd SPs S0
	8h	70.5	kd SPs Y S0
B₂	3h	38.9	SPs Kads SBs Kdes S0 kd
	8h	40.1	SPs Kads kd Kdes SBs S0
B₃	1h	52.5	S0 Bmax Y
	3h	23.9	S0 Bmax Y
	8h	8.9	S0 Bmax Y
H₁	3h	82.0	H10 SPs kd
	8h	60.1	H10 kd
H₂	3h	97.8	H20 S0
	8h	91.0	H20 kd S0
S	3h	77.7	S0 Bmax
	8h	59.7	S0 Bmax □

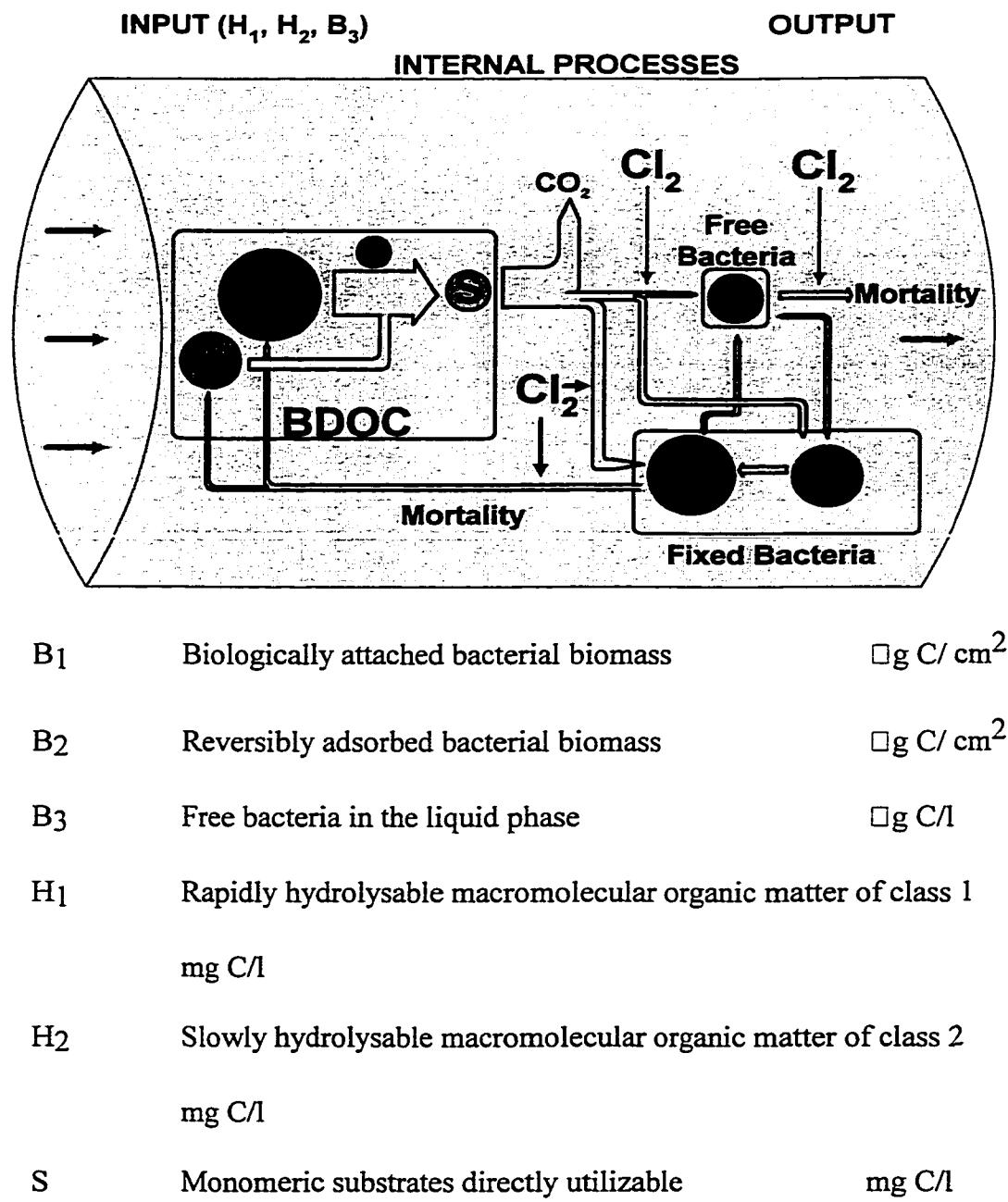


Figure 3.1 Processes of the SANCHO model (adapted from Servais *et al.*, 1995)

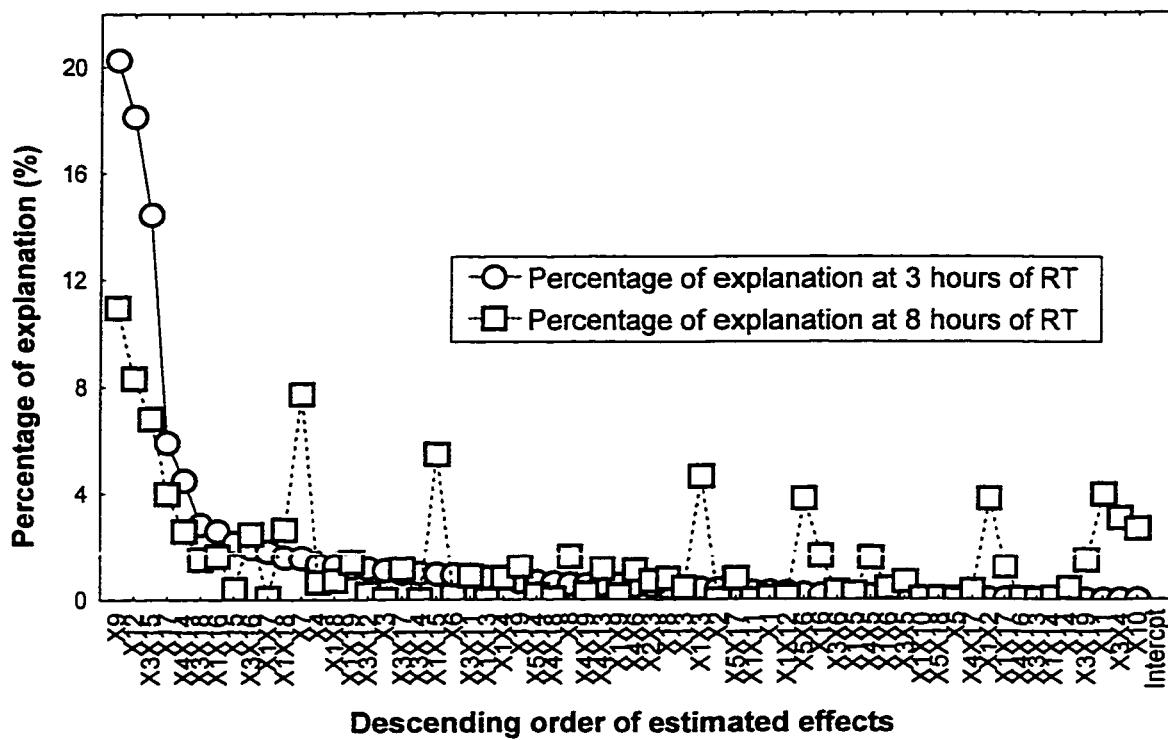


Figure 3.2 Pareto graph for the response variable B1(fixed bacteria) at 3 and 8 hours of residence time

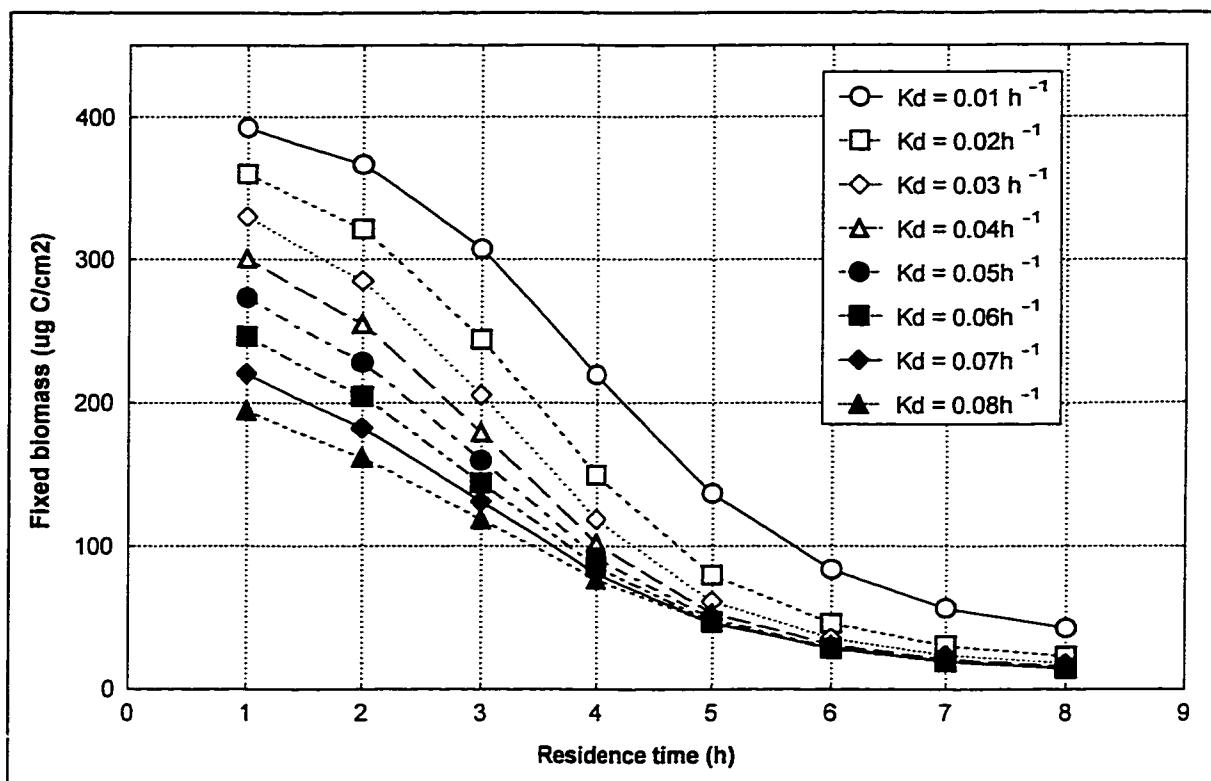


Figure 3.3. Influence of a variation in bacterial mortality (k_d) on the response variable fixed biomass

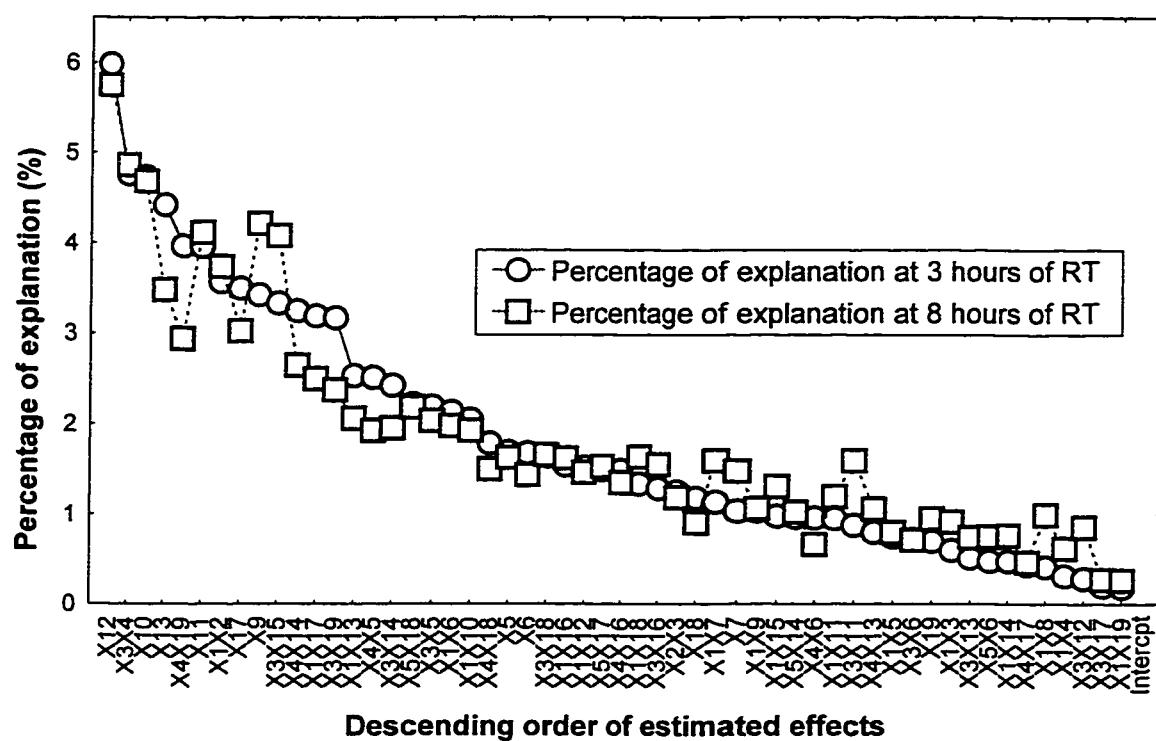


Figure 3.4 Pareto graphs for the response variable B2(adsorbed biomass) at 3 and 8 hours of residence time

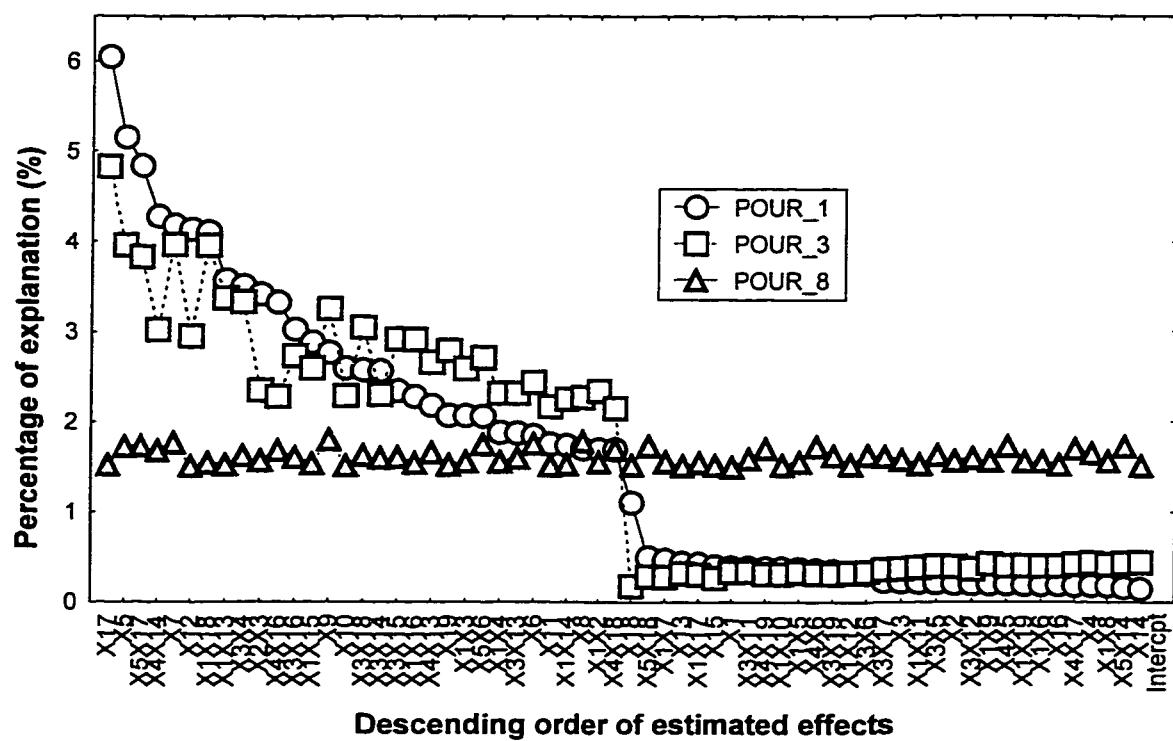


Figure 3.5 Pareto graphs for the response variable B3(free bacteria) at 3 and 8 hours of residence time

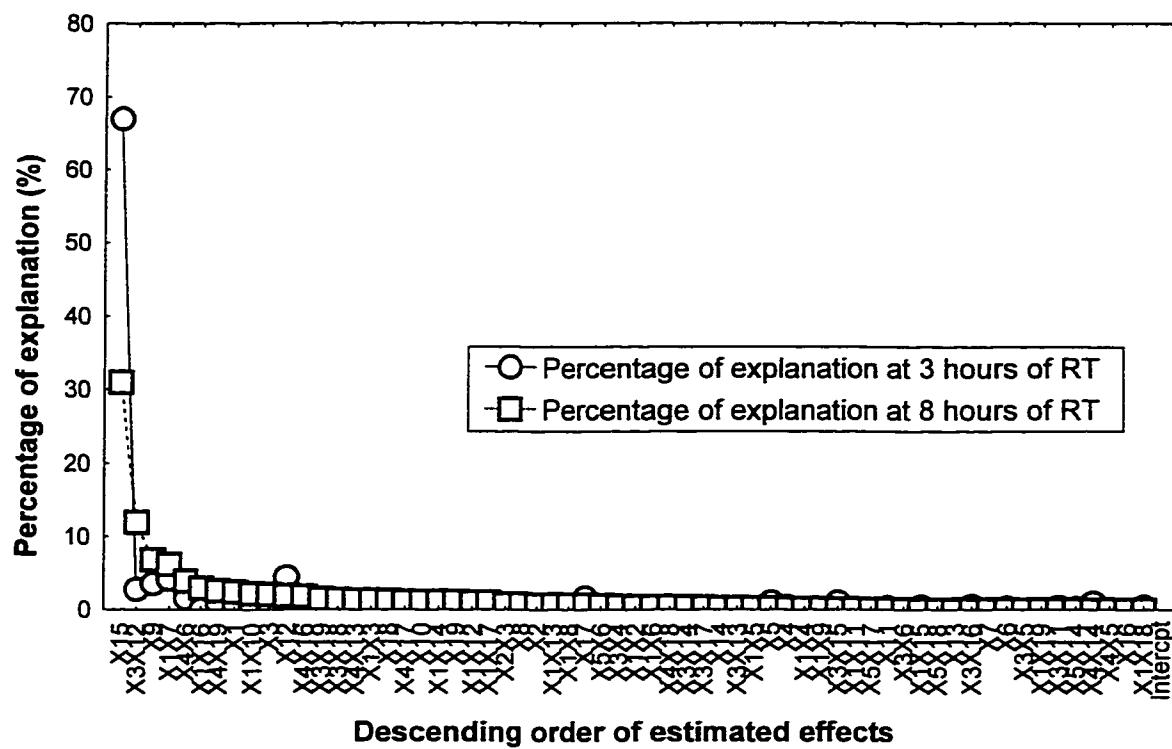


Figure 3.6 Pareto graphs for the response variable H1 at 3 and 8 hours of residence time

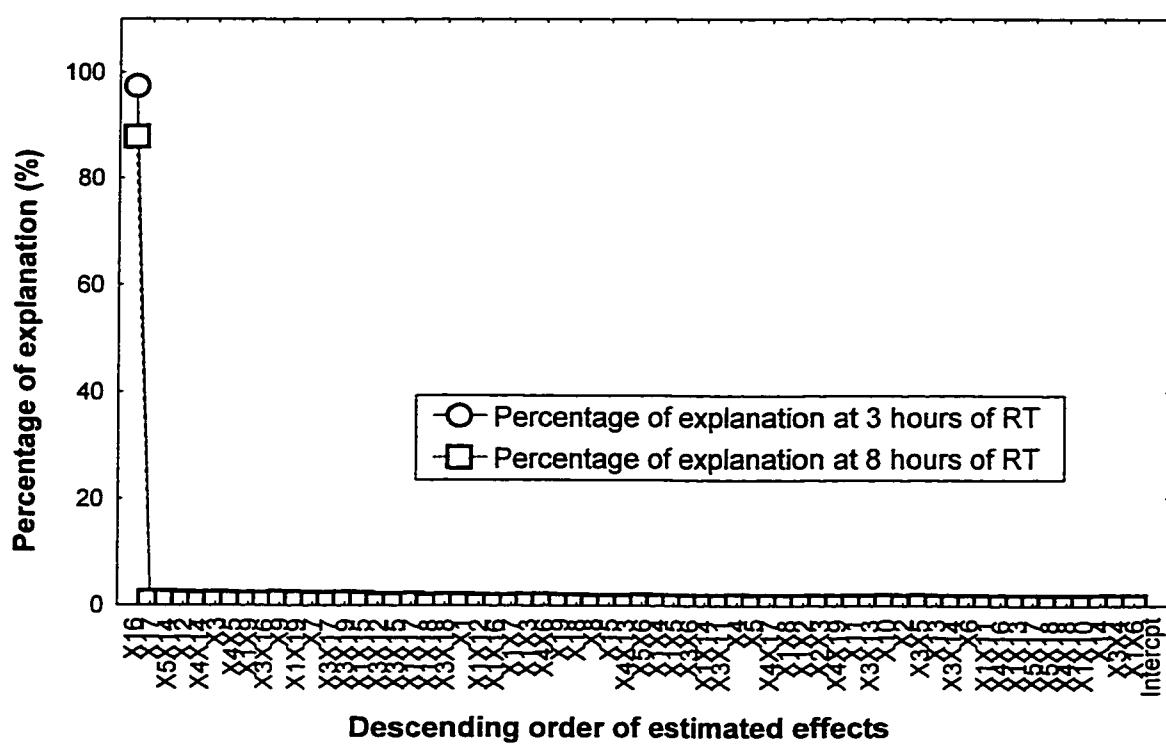


Figure 3.7 Pareto graphs for the response variable H2 at 3 and 8 hours of residence time

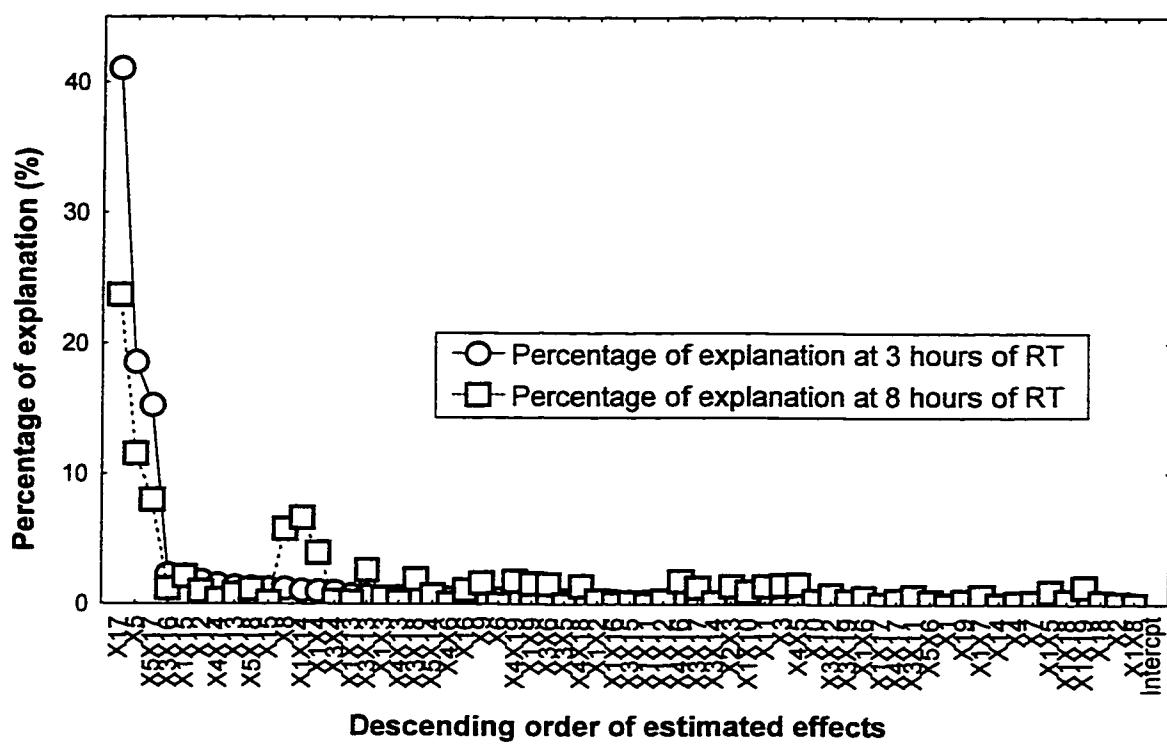


Figure 3.8 Pareto graphs for the response variable S at 3 and 8 hours of residence time

Chapitre 4. Analyse de la variance du modèle à l'aide de Plan croisés de Taguchi

Genichi Taguchi est un statisticien japonais qui a introduit plusieurs nouveaux concepts dans le domaine de la planification statistique d'expériences. Une de ses contributions a été de reconnaître l'importance de la planification d'expérience dans l'amélioration de la qualité des produits. En effet, la variabilité de qualité de fabrication d'un produit contribue à sa faible qualité. Taguchi a proposé le Design de Paramètres Robustes (« Robust Parameter Design ») ayant comme objectif de créer un produit insensible (ou robuste) aux facteurs principaux influençant la variabilité de ce produit (Taguchi 1987) .

Les mêmes stratégies utilisées par Taguchi dans l'industrie manufacturière peuvent être mises à contribution pour obtenir les sources de variabilité du modèle SANCHO. Taguchi reconnaît deux types de facteurs : les facteurs contrôlables et les facteurs non-contrôlables. Les paramètres contrôlables peuvent être modifiés par l'opérateur. La vitesse de rotation d'un tour, le débit d'eau dans un conduit ou le volume d'une solution sont autant d'exemples de paramètres qui peuvent être contrôlés par l'expérimentateur. Ces paramètres sont regroupés dans une matrice appelée la matrice d'expérimentation. Par opposition, l'humidité relative et la température sont difficilement contrôlables par l'expérimentateur et représentent donc un bruit de fond influençant la variabilité. Ces paramètres non-contrôlables sont aussi regroupés dans une matrice appelée matrice de bruit.

La seconde méthode d'analyse retenue pour connaître les sources de variabilité du modèle SANCHO est donc fondée sur l'analyse de la variabilité générée lorsqu'on perturbe légèrement (matrice de bruit) les paramètres initiaux du modèle (matrice d'expérimentation). La combinaison de ces deux matrices permet de créer un plan d'expérience où sera testée la variabilité des six variables de réponses du modèle SANCHO.

Cette deuxième méthode est donc appliquée après une première analyse selon le DOE, dont les résultats sont décrits au Chapitre 3. Avec cette deuxième méthode, onze des 19 paramètres clés ont été retenus. Ces paramètres sont présentés dans l'article accompagnant ce chapitre.

Le travail décrit dans ce chapitre permettra de tester la sensibilité du modèle SANCHO à une erreur d'estimation de ces 11 paramètres critiques, dans l'objectif de préciser les exigences de précision dans la mesure expérimentale de ces paramètres.

4.1 Matrice d'expérimentation

La matrice d'expérimentation est composée des paramètres qui sont contrôlés lors d'une simulation de SANCHO. Onze paramètres ont été retenus pour ces expériences. Le choix des paramètres a été effectué à la lumière des résultats obtenus lors de l'analyse par le PFP et de considérations pratiques. En effet, l'ajout de paramètres supplémentaires aurait rapidement gonflé le nombre total de simulations à effectuer. Un plan de 32 essais à deux modalités a été utilisé pour cette matrice. Les 11 colonnes de cette matrice représentant chacun des 11 paramètres utilisés dans ce modèle. Ce design est présenté à l'annexe 1. Les valeurs codées minimales (-1) et maximales (+1) ont été choisies en fonction d'études précédentes (Laurent 1995); Laurent, 1997 #1776].

4.2 Matrice de bruit

La matrice de bruit permet d'introduire des perturbations mineures aux éléments retenus dans la matrice d'expérimentation afin d'en étudier l'influence sur la variabilité. Selon la stratégie initiale de Taguchi, les éléments de cette matrice représentent le bruit de fond introduit dans le système. Toutefois, les mêmes onze facteurs de la matrice d'expérimentation ont été introduits dans cette matrice de bruit. La matrice utilisé est un design de Plackett-Burman à 20 essais. La variation induite autour des valeurs moyennes a elle été établie à 20%. Cette perturbation permet de s'assurer de couvrir un espace

expérimental suffisant et d'obtenir des résultats représentatifs. Une colonne supplémentaire a été introduite pour tenir compte de la moyenne des paramètres d'entrée. Seulement les onze premières colonnes de cette matrice ont été utilisées. Ce design est présenté à l'annexe 2.

La figure 3.1 illustre un plan croisé de Taguchi. La matrice de bruit est placée en haut à droite de la figure alors que la matrice d'expérimentation est plutôt placée à gauche en bas. Le croisement de ces deux matrices génère un total de 672 réponses à traiter.

Matrice d'expérimentation

Essai	Paramètre d'entrée			
	X1	X2	...	X11
1	+1	+1		-1
2	+1	+1		+1
...				
32	-1	-1	...	-1

Paramètre d'entrée	Essai				
	1	2	...	20	21
X1	+1	-1			0
X2	+1	+1			0
...					
X11	-1	-1			0

Variable de réponse

y_{ij}

Sommaire des réponses

y_1	s_1^2	$\log(s_1^2)$
....		
y_{32}	s_{32}^2	$\log(s_{32}^2)$

Figure 4.1. Réprésentation du plan croisé de Taguchi

où

X1, X2...X11 : Paramètre d'entrée (facteur) (valeur codée)

y_{ij} : Valeur de la variable de réponse à l'essai i de la matrice d'expérimentation et l'essai j de la matrice de bruit où $i = 1, 2, \dots, 32$ et $j = 1, 2, \dots, 21$

La moyenne des valeurs des variables de réponses est exprimée par :

$$\bar{y}_i = \frac{1}{21} \sum_{j=1}^{21} y_{ij} \quad (4.1)$$

Alors que la variance des variables de réponses est plutôt exprimée par :

$$s_i^2 = \frac{1}{20} \sum_{j=1}^{21} (y_{ij} - \bar{y}_i)^2 \quad (4.2)$$

4.3 Modèle statistique

L'objectif de cette analyse est l'étude de la variabilité des variables de réponses du modèle SANCHO. Afin de mieux départager l'importance des effets, le logarithme des variances est étudié. Ainsi, le modèle de régression utilisé pour étudier la variance est de :

$$\log(s^2) = \beta_0 + \sum_{j=1}^{11} \beta_j X_j + \sum_{k=1}^{15} \gamma_k Z_k \quad (4.3)$$

où

$\log(s^2)$: Logarithme de la variance

$X_1 \dots X_{11}$: Paramètres d'entrée (valeur codée)

$Z_1 \dots Z_{15}$: Chaîne d'interaction des paramètres (valeurs codées)

β_0 : Constante générale

$\beta_1 \dots \beta_{11}$: Effets principaux de $X_1 \dots X_{11}$

□_{1..11} : Effets d'interactions de X1..X11

Les résultats de cette analyse sont présentés sous la forme de graphique de Pareto. Dans ces graphiques, l'abscisse représente les 26 effets (11 effets principaux et 15 effets d'interactions) en ordre décroissant et l'ordonnée est exprimée comme le pourcentage d'explication de chacun des effets (principaux ou interaction). Le pourcentage d'explication de chaque effet est exprimée de la façon suivante :

$$\text{Pourcentage d'explication : } = 100 \cdot \frac{\beta_j^2 \text{ ou } \gamma_k^2}{\sum_{j=1}^{11} \beta_j^2 + \sum_{k=1}^{15} \gamma_k^2} \quad (4.4)$$

Les diagrammes de Pareto obtenus pour les 6 variables de réponses aux trois temps de résidence retenus sont présentés dans l'article suivant comme deuxième partie d'un article soumis à la revue Water Research. Dans cette seconde démarche, nous avons utilisé un Plan Croisé de Taguchi (PCT) dans l'objectif d'étudier la variabilité induite par une erreur d'estimation des paramètres sur les variables de réponses du modèle SANCHO.

Sensitivity Analysis of the SANCHO model:**Robustness Analysis Using a Taguchi Approach (Part 2)**

John Cigana ¹, Michèle Prévost ¹, Bernard Clément ², and Pierre Servais ³

¹ Department of Civil Engineering, Ecole Polytechnique of Montréal, C.P. 6079 , Succ. Centre-Ville, Montréal, Québec, Canada, H3C 3A7,

² Department of Applied Mathematics and Industrial Engineering, Ecole Polytechnique of Montréal, C.P. 6079 , Succ. Centre-Ville, Montréal, Québec, Canada H3C 3A7

³ Researcher, Groupe de Microbiologie des Milieux Aquatiques, Université Libre de Bruxelles, Campus de la Plaine, C.P. 221 , Boulevard du Triomphe, B-1050, Bruxelles, Belgium

Abstract

SANCHO is a time dependent deterministic model, developed at the ULB (Université Libre de Bruxelles) that simulates bacterial dynamics, free chlorine decay and biodegradable organic matter (BOM) consumption in water distribution systems (DS). The software version of the SANCHO model is used in the drinking water industry to predict bacterial regrowth in pipes and to assist utilities in complying with drinking water regulations. The objective of this work was to determine the sensitivity of the response

variable of the SANCHO model to errors of estimation of the critical parameters identified by a sensitivity analysis using a DOE approach. Our results indicate that four parameters have an undeniable influence on the variance of the bacterial biomass in the distribution system predicted by the model. These parameters are: bacterial mortality (k_d), the maximum capacity for bacterial fixation (S_P s), the initial concentration of direct substrates (S_0) and the maximum rate of uptake of direct substrates (B_{max}). With these results in hand, we can identify the key parameters which need very accurate estimations and focus the analytical efforts to quantify them.

Introduction

The understanding of the causes of changes in water quality in distribution system (DS) is a complex task. Water quality can be estimated by its physical, chemical and biological properties. One of these is the presence of bacteria in the treated and distributed water. Relevant local authorities set up the different standards. Indicator organisms, such as coliform bacteria, are commonly used to monitor microbiological water quality of drinking water. A recent study as shown that a high proportion of the 164 American water utilities surveyed had experienced recurring coliform episodes (Smith et al. 1990). Although not a new problem (AWWA 1930), bacterial regrowth in DS is of rising concern in light of recent studies and multiple non compliance events. Bacterial regrowth may be linked to an increase in gastroenteric illnesses (Payment et al. 1991) and to the development of nuisance organisms like *Asellus* and *Naïs* in the DS (AWWA, 1995). The classic response to a regrowth event is to increase the disinfectant dosage (Camper, 1995) or to change

disinfectant type (LeChevallier et al. 1996). Not only this has sometimes proven to be ineffective in controlling coliform regrowth (Camper 1995) but may give rise to an undesirable increase in the concentration of disinfection by-products. An alternative approach resides in reducing the amount of substrates required for bacterial growth (the biodegradable organic matter or BOM) in the treated water. Several studies [Servais, 1995 #1353; Prévost *et al.*, 1997] have shown that this was an effective way to control bacterial regrowth in DS.

The SANCHO model

In order to predict the respective role of BOM and chlorine in controlling bacterial regrowth in DS, a time-dependent static model, the SANCHO model, was developed at the Université Libre de Bruxelles (Servais et al. 1995). The SANCHO model simulates bacterial dynamics of growth and mortality, free chlorine decay and biodegradable organic matter utilization in a water distribution system. The reader may find the processes of the SANCHO model described in greater length in (Servais et al. 1995); and (Billen et al. 1992). An overview of the model is presented in Figure 1 showing the six response variables, their symbols and their units.

A group of four processes are taken into account by the SANCHO model :

1. the interaction between organic matter and bacteria. This includes the exoenzymatic hydrolysis of biodegradable organic matter into direct substrates by bacteria ;
2. the growth of free and fixed bacteria on the hydrolysis products and bacterial mortality which releases organic matter ;
3. the reversible adsorption-desorption of bacteria onto the inner pipe surface and bacterial fixation onto pipe surfaces ;
4. the chemical consumption of free chlorine and the impact of free bacteria on free and fixed bacterial activity and mortality;

The nutrients used by bacteria in the SANCHO model is described as biodegradable dissolved organic carbon or BDOC. BDOC is divided in three classes : H1, the rapidly hydrolysable polymeric carbon, H2, the slowly hydrolysable polymeric carbon and S, the directly utilisable monomeric carbon. In this model, BDOC is thus represented as the sum of H1, H2 and S.

The SANCHO model has been tested in Europe and North America with some success. There is a general agreement between observed and predicted suspended biomass values and BDOC consumption using the SANCHO model (Laurent et al. 1997); (Servais et al. 1995). The experimental results used for this testing were obtained from different sources of water, treatment plants and over a wide range of temperature and on varying pipe diameters and pipe materials (Laurent et al. 1997; Servais et al. 1995). The model and its software version are rather complex and their validation is on-going. Further

improvements in the predictive ability of the model can only come from a deeper understanding of its components, its strengths, its weaknesses and its limits of applicability. Indeed, if we hypothesise that SANCHO correctly depicts bacterial dynamics, then critical parameters of the model are also critical to regrowth in DS. To move in that direction, it appears necessary to complete a sensitivity analysis.

A simplified version of the model was used for the purpose of conducting the sensitivity analyses. Two major effects were not analysed: temperature and chlorine. Using a simplified version allows for in depth testing of the equations predicting bacterial regrowth and substrate utilisation. The temperature was held constant at 20°C and the theoretical diameter of the network pipe remained at 0.15 m for all the simulations. This maximises the Surface to Volume (S/V) ratio thus yielding optimum growth conditions. Finally, the impact of chlorine was not investigated. These three simplifications were made to offer a simple situation in which the model could be tested for the prediction of bacterial growth.

This simplified version of the model uses 19 initial parameters to accomplish the task. Table 1 lists the 19 necessary input parameters given to the SANCHO model. Given an initial set of parameters, the model will predict the evolution of bacterial biomass and BDOC concentrations for a given time of residence. The software version of the SANCHO model is used in the drinking water industry in order to predict bacterial regrowth in networks. Since key parameters may change with residence time, two

significative residence time where chosen: 3 and 8 hours. This choice was based on the several biomass curves yielded by the SANCHO model. It is between these two residence time that most of the bacterial activity was observed in the absence of chlorine and in small diameter pipes.

The objective of this research was to gain a better understanding of the sensibility of the SANCHO model. In this outlook, we ask:

- What are the critical parameters that need to be measured with better accuracy?

To answer these questions, a statistical analysis using a Taguchi approach was carried out.

Experimental Design and Methods

One of the contributions of Taguchi in the field of quality engineering was to recognise the need to improve product design with statistically designed experiments. In our case, we have been working on a model and its software version. Product variability is seen as a major contributor to poor quality. Taguchi introduced Robust Parameter Design with the goal to design a product which is robust(not sensitive) to factors influencing variability in the performance of the product (Taguchi 1987). Two types of factors are recognised : control (design) factors and noise factors. Control factors are parameters that can be controlled experimentally. For example, the flowrate and pipe diameter can be set

to given values. Examples of noise factors are temperature, amount of material, wear, etc... These noise factors can be considered as tolerances around the mean values.

Taguchi introduced a crossed array statistical design to study the impact of noise factors on the performance characteristics, the response variable. This idea can even be applied to design control factors which cannot be strictly controlled around their nominal values. It is this approach that we have used in this paper.

We have chosen a set of 11 input parameters of the SANCHO model as shown in Table 1. This choice was based on the result of a previous study (Cigana et al. 1997) and of practical considerations. Indeed, this method requires a large number of simulations. Any increase in the number of initial parameters would have greatly increased the number of runs necessary. In our case, the set of noise factors is identical to the set of control factors. Our goal is to study the impact on the response variable of deviations or errors of estimation from nominal values.

Crossed Array Design

We used the 11 input parameters in a 32 run inner array statistical design where each input is varied at two levels. The low level value (-1) and the high level value (+1) are set at the known minimum and maximum values of the input parameters (Cigana et al. 1997). The induced variation near the mean values (+1 or -1) was fixed arbitrarily at 20% of this latter value. Not only does this cover a large experimental space, but induces

sufficient variability in the response variable as to render representative results. The 32 run design is a standard orthogonal design. The inner array has a resolution IV permitting to separate 11 main effects and 15 chains of confounded interactions effects as shown in Table 2, whereas the outer array is a standard 20 run Plackett-Burman design. (ref Stats)

Figure 2 illustrates the crossed array design in a two-dimensional experimental space. Each input parameter takes on two coded values +1 or -1 giving four nominal combinations of (X1, X2): (+1,+1), (+1,-1), (-1,+1), (-1,-1) represented by the middle point in each square. Around each of these points, an outer array is built with four circles. These points represent deviations of 20% from the nominal values. The final design is obtained by crossing each row of the inner array with every run of the outer array. The complete design has (32x 31) or 672 runs representing experimental simulations that were submitted to the SANCHO model. Our main interest was to study the impact of the perturbations of each of the 11 input parameters on the variance of the response variables.

Analysis of the model

We have modelled the $\log(s^2)$ response has :

$$\log(s^2) = \beta_0 + \sum_{j=1}^{11} \beta_j X_j + \sum_{k=1}^{15} \gamma_k Z_k \quad (\text{Eqn. 4.5})$$

where

$\log(s^2)$ = Variance of the predicted response variable

X1,...,X11 : Coded input parameters

Z1, ..., Z15: Coded chain of interactions given in Table

$\square \square$ General constant of the regression model

$\square 1, \square \square$ Half main effect of X1...X11

$\square 1, \square \square$ Half interaction effect of X1...X11

The results have been presented under the form of Pareto graphs. The percentage of explanation of each effect (main or interaction) on the variability of the response variable is expressed as :

$$\text{Percentage of explanation} = 100 \cdot \frac{\beta_j^2 \text{ or } \gamma_k^2}{\sum_{j=1}^{11} \beta_j^2 + \sum_{k=1}^{15} \gamma_k^2} \quad (\text{Eqn. 4.6})$$

Results and Analysis

As shown on Figure 1, three of the six response variables are related to the bacterial biomass present in the distribution system. The three forms under which the bacterial biomass is represented in this model are: fixed, adsorbed or free. The three other response variables represent the lability classes of the biodegradable dissolved organic carbon: H₁, the concentration of macromolecular organic matter of lability class 1 (quickly hydrolysable), H₂, the concentration of macromolecular organic matter of

lability class 2 (Slowly hydrolysable) and S , the concentration of directly assimilable substrates.

B1- Biologically attached bacterial biomass (fixed bacterial biomass)

Figure 3 shows the parameters of influence of the variance of the response variable fixed bacteria in decreasing order as predicted by the SANCHO model at 1, 3 and 8 hours of residence time. The y-axis represents the percentage of explanation of the variance for each estimate whereas the x-axis represents the 26 estimated effects (11 main and 15 interaction effects) ranked in decreasing order. This permits to highlight the influence of residence time on the critical parameter.

It can be noticed that the parameters of influence are common at 1 and 3 h of residence time. The most influential parameter on the variance of the response variable fixed bacteria is SPs (maximum capacity for bacterial fixation) with respectively 55% and 47% of explanatory power at 1 and 3 hours of residence time. The two other parameters of importance at these residence time are S_0 and B_{max} . An entirely set of parameters and chain of interaction dominate the results at 8 hours of residence time. The variance of the fixed bacteria at this residence time is greatly influenced by the value of B_{max} with 55% of explanatory power and the initial value of H_1 with 17% of explanatory power.

Table 3 presents the first parameter of importance of the variance for all six response variables. From this table, it can be seen that SPs (maximum capacity for bacterial

fixation) is the most important parameter in explaining the variance of fixed bacterial biomass at 1 hour and 3 hours of residence time. The parameter SPs in the SANCHO model represents a simplification of the many parameters involved in controlling the maximum density of fixed biomass in a DS. The parameter SPs takes into account the impact of the hydraulic shear stress of detachment in a global value, which could be considered as resulting from unsteady water fluxes (nutrient and shear). It also includes the effect of pipe material, which has recently been shown to dominate over nutrient levels in various studies (Camper, 1995; Rompré et al. 1997).

At 8 hours of residence time, it appears that the maximum rate of substrate uptake becomes most critical to the mean value of this response variable. The shift of the key parameters with increasing residence time reflects the different conditions of growth. At the shorter residence times, the number of available sites for bacterial colonisation, expressed in the SANCHO model by the value of SPs, is the driving in the prediction of the variance of the fixed bacterial biomass. At longer residence times, substrate utilisation parameters such as B_{max} , the maximum rate of substrate uptake, becomes dominating. High values of B_{max} will strongly influence bacterial biomass at longer residence times by reducing the flux of nutrients later in the DS. These results therefore support the experimental evidence that pipe material is an important factor controlling regrowth. These results are in concordance with those obtained with the DOE approach, but they pinpoint to the importance of choosing appropriate SPs values to predict fixed bacterial biomass.

The growth of fixed bacteria can be linked back to the maximum uptake rate of direct substrates. In the SANCHO model, the uptake of direct substrates is controlled by a Michealis-Menten relationship where B_{max} represents the maximum rate of uptake of direct substrates per unit of bacterial biomass. The results from the DOE approach also identified the amount of direct substrates (S_0) as one of the parameters critical to the variability of fixed bacterial biomass. Two analyses converge on the importance of the concentration of direct substrates in the variability of the regrowth phenomena.

The numeric value of the parameter SPs is based on two estimations :

- laboratory measurements of the amount of sites available for adsorption of bacteria on several pipe material, which may lead to attachment. This experimental value is considered to be a minimum value of SPs, since the number of sites available for growth is most likely higher than the number of adsorption sites.
- On the other hand, the maximum value of the parameter SPs is defined by the maximum value of the biomass measured in pilot and full-scale DS.

Servais et al. (1992) have shown that the fixed biomasses on pipe surfaces never exceed a value of 2 mgC/cm^2 . Our results show the great influence of this parameter in the variability of fixed bacterial biomass. Knowing this, it appears that the estimation of SPs should be given more attention in order to minimise the variance of the predicted value of fixed bacterial biomass. Experimental measurement on biofilm reactors using different materials and under varying shear stress would be useful.

B2 - Reversibly adsorbed bacterial biomass (adsorbed biomass)

Adsorbed bacterial biomass includes the reversibly attached biomass on the pipe material where a dynamic state of attachment-detachment exists. Adsorbed biomass represents a very small proportion (from less than 0.01% to 1%) of the total bacterial biomass present in the pipe and is therefore very elusive to measure. Only initial adsorption kinetics can be measured experimentally on virgin material or on pipe conditioned material.

Figure 4 expresses the evolution, through time, of the critical parameters pertaining to the variance of the response variable of adsorbed bacterial biomass. At all three residence times the most influential parameter is SPs. At 1 hour of residence time, SPs accounts for 46% of the explanatory power, 38% at 3 hours of residence time and 28% at 8 hours, as shown in Table 3. These results lead to the conclusion that SPs (the maximum capacity for bacterial fixation) governs the variance of the adsorbed bacterial biomass.

Adsorption can only take place when the sites for adsorption, expressed by the parameter SBs, are available. In absence of chlorine and in most of the experimental conditions tested during this sensitivity analysis, the number of sites available for adsorption are very small since most of the sites are occupied by attached bacteria. The parameter SPs has been shown to be the most critical parameter for fixed bacterial biomass.

B3 - Free bacteria in the liquid phase (suspended bacterial biomass)

Free bacteria represents the bacteria found in the bulk liquid phase of the distributed water. Free bacteria either originate from cellular growth of free bacteria or from the detachment of fixed bacteria from the pipe wall. The detachment of fixed bacteria has been shown to be the main source of free bacteria and that growth of free bacteria is not sufficient to explain the amount of free bacteria found in DS (Mathieu 1992; Servais et al. 1995; van der Wende et al. 1989).

Figure 5 expresses the evolution, through time, of the critical parameters pertaining to the variance of the response variable. This again permits to highlight the influence of residence time on the critical parameter. It can be noticed that similar trends are observed at 3h and 8h of residence time, but different factors stand out as critical at 1 hour of residence time.

At 1 hour of residence time, it can be seen that B_{max} explains 35% of the variance of the response variable free bacteria followed by the initial concentration of direct substrates (S_0) and the bacterial yield (Y) with respectively 17% and 15% of explanatory power. At 3 hours of residence time, the initial concentration of direct substrates (S_0) is the most influential with 27% of explanatory power. In second place comes bacterial yield (Y) with 20% of the explained variability. At 8 hours of residence time, the term most

influential with regards to the variance is the initial concentration of direct substrates (S_0) followed by the bacterial yield (Y) with 22% of explanatory power. Again, it can be seen that the influence of some parameters vary greatly through time. The value of B_{max} drops from a value of 35%, at 1 hour, to almost zero at 8 hours of residence time. The influence of S_0 doubles from 15% to 30% when residence time goes from 1 hour to 8 hours. Likewise, the influence of the initial concentration of rapidly hydrolysable substrates (H_{10}) greatly increases during the 8 hour period going from almost zero to 13%. This again shows the impact of residence time on the relative importance of parameters.

It appears from our analysis that in the front end of the DS (1 hour of residence time) the maximum uptake capacity of direct substrates (B_{max}) rules the variance of this response variable. The growth of free and suspended bacteria in the bulk water is linked to the parameter B_{max} found in the Michealis-Menten relationship of the SANCHO model by the following equation (Servais et al. 1995) :

$$\text{Direct substrate uptake} = B_{max} \frac{S}{S+k_s} B \quad (4.7)$$

where

B_{max} = Maximum capacity of uptake of direct substrates

S = Concentration of direct substrates

k_s = Half saturation constant

B = Bacterial biomass present (fixed and suspended)

This probably reflects the fact that at the earlier residence time :

1. the substrate is probably not limiting and the $S/(S + k_s)$ fraction is close to the unit giving more weight to the value of B_{max} .
2. the maximum rate of direct substrate uptake will be critical in determining how early the fixed biomass will become established in the DS. Since most suspended bacteria comes from detachment, B_{max} will impact directly suspended bacterial biomass.

At 3 hours of residence time, it is the initial concentration of direct substrates (S_0) by the way of the same fraction $S/(S + k_s)$, that becomes more important in the variance of the response variable. Finally, at 8 hours of residence time there is a divergence in the terms of influence for the variance of the response variable. Later in the DS, the flux of direct substrates is ensured by the hydrolysis of H_1 into S and by the remaining fraction of S_0 in the bulk liquid. These parameters dominate as source of variability at 8 hours of residence time. It may be surprising not to find any parameters relating to detachment in the critical parameters for suspended bacterial biomass. Two types of parameters in the SANCHO model are related to attachment and detachment. First, the desorption constant, k_{des} , the adsorption constant, k_{ads} , the maximum capacity for adsorption ,SBs, and the maximum capacity for bacterial fixation, SPs. Although the parameter SPs is of prime importance for the fixed bacteria and adsorbed bacteria response variable, SPs did not stand out as a critical factor. The fact that the adsorption related parameters do not contribute significantly most probably reflects the very small size of the adsorbed

bacterial biomass as compared with the suspended and even more the fixed bacterial biomass.

Several studies have been undertaken to predict the detachment of fixed bacteria, the major source of suspended bacteria, while taking into account the major mechanisms which may come into play. (Stewart et al. 1993) proposed a detachment rate proportional to the derivative of the velocity in the pipe ; (Characklis and Marshal 1989) used a detachment rate proportional to the square of the pipe velocity. (Dukan et al. 1996) proposed a model which predicts detachment based on the sheer stress and growth.

In the SANCHO model, suspended bacteria are produced by three means : 1) growth of suspended bacteria, 2) desorption and 3) detachment proportional to growth, since the probability of attachment decreases as the number of available attachment sites decreases. The SANCHO model does not take into account the effect of shear stress.

Expressed in term of SANCHO's parameters, the following inequality has to be respected in order to validate this detachment model for non chlorinated biofilms:: $Y \cdot B_{\max} \geq K_{des} + k_d$. Using the range of values furnished in Table 1, it is possible to find a subset of values where this inequality is respected in the SANCHO model. This tends to show that the Langmuir isotherm used in the SANCHO model is at least valid over an albeit limited range of the parameters. Perhaps minor modifications to this isotherm would bring out the importance of the detachment of fixed bacteria in the presence of free bacteria

H1 - Rapidly degradable macromolecular organic matter

Figure 6 expresses the evolution, through time, of the critical parameters pertaining to the variance of the response variable H1. The y-axis of this figure represents the percentage of explanation of the variance for each estimate whereas the x-axis represents the 26 estimated effects (11 main and 15 interaction effects) ranked in decreasing order. This again permits to highlight the influence of residence time on the relative weight of critical parameters.

The variance of the response variable H1 is almost entirely explained by the initial value of H1 at 1 hour of residence time. At 3 hours of residence time this percentage drops to 75% for the value of H1. Finally, at 8 hours of residence time, the most influential parameter becomes the bacterial mortality with 35% of explanatory power as shown in Table 3.

At 1 and 3 hours of residence time, it is the initial concentration of rapidly hydrolysable organic matter that governs the variance of the response variable H1. This highlights the importance of obtaining an accurate measure of the concentration of BDOC with the Servais-Billen method (Servais et al. 1989), from which the initial concentration of H1 is taken. At 8 hours of residence time, it is rather the term of bacterial mortality (k_d) which governs the variance of H1. A previous study had already shown that the initial value of H1 was responsible for the greater part of the variance of the response variable. This situation is confirmed by the results obtained by the Taguchi method.

H2 - Slowly degradable macromolecular organic matter

Figure 7 expresses the evolution, through time, of the critical parameters pertaining to the variance of the response variable H1. The y-axis of this Figure represents the percentage of explanation of the variance for each estimate whereas the x-axis represents the 26

estimated effects ranked in decreasing order. This again permits to highlight the influence of residence time on the critical parameter.

By the same token, it can be noticed that almost 100% of the variability of the response variable as shown in table 3.

At all three residence times, the variance of the response variable slowly hydrolysable organic matter (H_2) is solely determined by the initial concentration of H_2 . Again, this shows the importance of correctly characterising the initial BDOC of the water into its three lability classes in order to obtain accurate predictions in the DS later on. Once again these results confirm earlier results obtained by (Cigana et al. 1997) using a different method.

S - Directly utilizable monomeric substrate

Figure 8 expresses the evolution, through time, of the critical parameters pertaining to the variance of the response variable S . The y-axis of this Figure represents the percentage of explanation of the variance for each estimate whereas the x-axis represents the 26 estimated effects (11 main and 15 interaction effects) ranked in decreasing order. This again permits to highlight the influence of residence time on the critical parameter.

At 1 and 3 hours of residence time, the variance of the response variable S is governed by the initial value of S_0 . This can be explained by the same Michealis-Menten relationship discussed for the free bacteria response variable. The parameter B_{max} appears to become more important in the variance of the response variable at 8 hours of residence time. The importance of correctly characterising the water into its three lability classes is again apparent. Only by inserting a correct initial value of S , can the variance of the response variable be controlled.

Conclusions

Design of experiments (DOE), in the form of the Taguchi method, is a valid approach to study the variance of the response variables of the SANCHO model. Although a substantive number of simulations are required ($n=672$), the results provided highlight the critical parameters. Table 3 indeed shows that the critical parameters, with respect to the variance, vary from response variable to response variable and also with residence time.

In a general manner, we can conclude from Table 3 that it is easier to follow the sources of variability of substrates in DS than those of biological biomass, growth and decay in DS being more intrinsically complex phenomena. The SANCHO model simulates the complex phenomenon of bacterial regrowth in networks. Accordingly, the critical parameters of the model vary with each of the response variables as shown in Table 3. Some of the parameters nonetheless appear to have some influence over all of the six response variables. These parameters are : the maximum capacity for bacterial fixation, (SPs), the initial value of the monomeric substrate that is directly utilizable (S_0) and the maximum rate of uptake of direct substrate (B_{max}). We see that one of the three parameters are microbiological in nature (B_{max}) , one is a substrate terms (S_0) and the last one a pipe characteristic (SPs). In order to reduce and control the variability of the response variables, we must better insure the adequate measure of each of the critical parameters. Further work on the measure of these parameters should be of great importance. Also, we see that the response variables pertaining to substrates(H_1 , H_2 and S) are very sensible to the initial value of each of these respective parameters. This justifies the Servais-Billen (Servais et al. 1989) method of measuring BDOC.

Acknowledgements

The study was carried within research program of the Industrial Drinking Water Chair of the École Polytechnique of Montréal. The authors therefore wish to thank the Chair partners for their financial support, NSERC, the City of Montréal, the City of Laval, École Polytechnique de Montréal, John Meunier Inc., Compagnie Générale des Eaux Group, the consulting firm Vallée Lefebvre et associés and the Université Libre de Bruxelles (Belgium). The model SANCHO was studied under an agreement with CGE and Université Libre de Bruxelles (Belgium) and we would like to express our gratitude to the SEDIF (Syndicat des Eaux d'Ile de France) for the extended use of the upgraded model version.

References

- Billen, G., Servais, P., Bouillot, P., et Ventresque, C. (1992). Functioning of Biological Filters Used in Drinking Water Treatment-the Chabrol Model. *Dans «J Water SRT - Aqua»*, Vol. 41, pp. 231-241.
- Block, J. C., Haudidier, K., Paquin, J. L., Miazga, J., et Lévi, Y. (1993). Biofilm accumulation in drinking water distribution systems. *Biofouling* 6, 333-343.
- Camper, A. K. (1995). Factors influencing biofilm growth in drinking water distribution systems. Ph.D., Montana Stage University, Bozeman, Montana, U.S.A.
- Camper, A. K., Hayes, J. T., Jones, W. L., et Zelver, N. (1993). Persistance of Coliforms in Mixed Populaiton Biofilms. *Dans «AWWA - WQTC»*, Miami, Floride.
- Characklis, W. G., et Marshal, K. C. (1989). «Biofilms,» John Wiley and Sons.

- Cigana, J., Clément, B., Prévost, M., et Servais, P. (1997a). Sensibility Analysis of the SANCHO model using Design of Experiments. *Dans «Quality and Productivity Research Conference»*. American Statistical Association, Orlando, Florida.
- Cigana, J., Laurent, P., Prévost, M., Clément, B., et Servais, P. (1997b). Key parameters affecting regrowth in distribution systems using the SANCHO model. *Dans «AWWA-WQTC»*, Denver, CO.
- Colbourne, J. S., Dennis, P. J., Rachwal, A. J., House, S., Keevil, W., et Mackerness, C. (1992). The Operational Impact of Growth of Coliforms in London's Distribution System. *Dans «AWWA - Water Quality Technology Conference»*, pp. 11 pages, Toronto, Ontario.
- Dukan, S., Levi, Y., Pirion, P., Guyon, F., et Villon, P. (1996). Dynamic Modelling of Bacterial Growth in Drinking Water Networks. *Wat. Res.* **30**, 1991-2002.
- Laurent, P. (1995). Colonisation bactérienne des réseaux de distribution: étude expérimentale et modélisation. Ph.D., Université Libre de Bruxelles, Bruxelles.
- Laurent, P., et Servais, P. (1995). Fixed Bacterial Biomass Estimated by Potential Exoproteolytic Activity. *Can. J. of Microbiol.* **41**, 749-752.
- Laurent, P., Servais, P., Prévost, M., Gatel, D., et Clément, B. (1997). Testing the SANCHO model on distribution systems. *J. Am. Water Works Assoc.* **89**, 92-103.
- LeChevallier, M. W. (1990). Coliform Regrowth in Drinking Water: A Review. *J.AWWA* **82**, 74-86.

- LeChevallier, M. W., Lee, R. G., et Young, J. S. (1991). «Application of Biological Processes to Current Water Treatment Practices.». AWWASCo.
- Mathieu, L. (1992). La matière organique et la stabilité biologique des eaux au cours de leur distribution. Thèse de doctorat, Université de Nancy I, France.
- Servais, P., Anzil, A., et Ventresque, C. (1989). Simple Method for Determination of Biodegradable Dissolved Organic Carbon in Water. *Appl. Environ. Microbiol.* **55**, 2732-2734.
- Servais, P., Billen, G., Laurent, P., Lévi, Y., et Randon, G. (1992a). Studies of BDOC and Bacterial Dynamics in the Drinking WAtter Distribution System of the Northern Parisian Suburbs. *Revue des sciences de l'eau* **5**, 69-89.
- Servais, P., Laurent, P., et Billen, G. (1992b). «Etude de la colonisation bactérienne des réseaux de distribution.». Groupe de Microbiologie des Milieux Aquatiques-Université Libre de Bruxelles.
- Servais, P., Laurent, P., Billen, G., et Gatel, D. (1995). Development of a model of BDOC and bacterial biomass fluctuations in distribution systems. *Revue des sciences de l'eau* **8**, 427-462.
- Servais, P., Laurent, P., Prévost, M., et Gatel, D. (1996). Contrôle de la croissance bactérienne dans les réseaux de distribution. *Dans «Hydrotop 96»*, Marseilles.
- Stewart, P. S., Peyton, B. M., Drury, W. J., et Murga, R. (1993). Quantitative observations of heterogeneities in *Pseudomonas aeruginosa* biofilms. *Applied Environmental Microbiologiy* **59**, 327-329.

Taguchi, G. (1987). «System of Experimental Design,» International Publications, White Plains, NY.

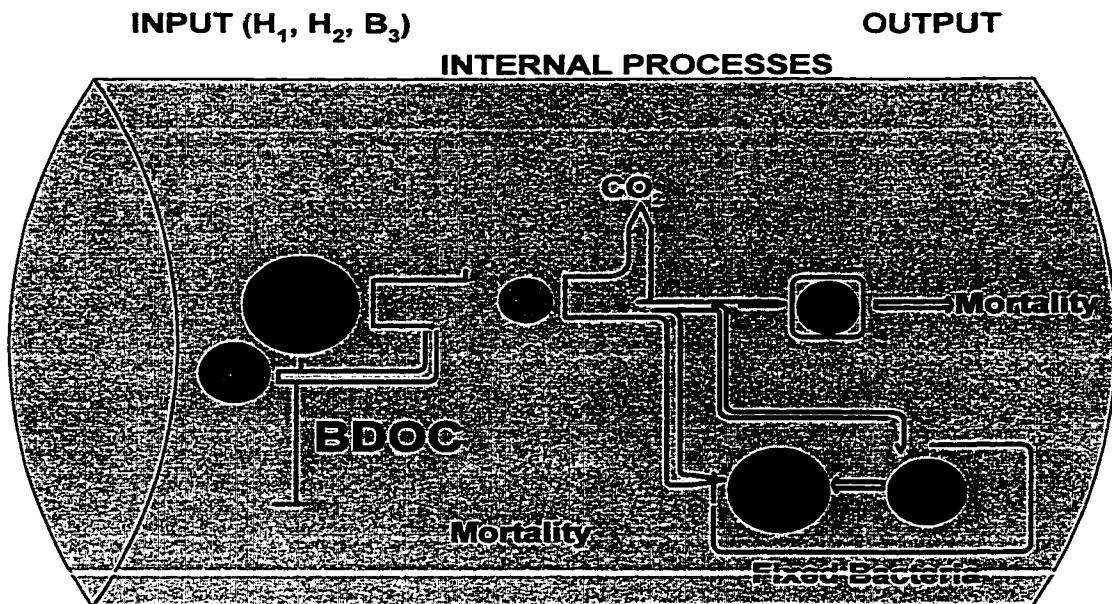
van der Wende, E., Characklis, W. G., et Smith, D. B. (1989). Biofilms and Bacterial drinking Water Quality. 23 .

Tableau 4.1 List of initial parameters involved in the SANCHO model

No	Design	Symbol	Name of parameter	Units	Mean	Low	High
		variable			modalit	modality	
1	A	H ₁₀	Macromolecular organic matter of lability class 1 (Quickly degradable)	mg C/l	0.26	0.02	0.5
2	B	H ₂₀	Macromolecular organic matter of lability class 2 (Slowly degradable)	mg C/l	0.26	0.02	0.5
3	C	S ₀	Monomeric substrate that is directly utilizable	mg C/l	0.26	0.02	0.5
4	D	SPs	Maximum capacity for biological bacterial fixation	μg C/cm ²	1.6	0.2	3.0
5	E	Fraction	Fraction of X12 i.e. SBs of the support = Fraction* SPs	%	52.5	10	95
6	F	k _{ads}	Adsorption rate constant of bacteria onto the support	(μg C/ml) ⁻¹	1.8	0.6	3.0
7	G	k _{des}	Desorption rate constant of bacteria from the support	h ⁻¹	1.6	0.2	3.0
8	H	k _d	Mortality rate	h ⁻¹	0.045	0.01	0.08
9	I	Y	Bacterial yield	-	0.2	0.1	0.3
10	J	K _s	Half-saturation constant of direct substrates S	mg C/l	0.11	0.02	0.2
11	K	B _{max}	Maximum rate of substrate uptake of direct substrates S	h ⁻¹	1.15	0.3	2.0

Tableau 4.2 Percentage of variability explained for each response variable

Response Variable	Time of residence	Explained variability (%)	First critical parameters for each response variable
B₁	1h	56.0	SPs
	3h	49.0	SPs
	8h	57.0	B _{max}
B₂	1h	46.0	SPs
	3h	38.0	SPs
	8h	28.0	SPs
B₃	1h	34.0	B _{max}
	3h	27.0	S ₀
	8h	31.0	S ₀
H₁	1h	96.0	H ₁₀
	3h	73.0	H ₁₀
	8h	34.0	k _d
H₂	1h	100.0	H ₂₀
	3h	100.0	H ₂₀
	8h	99.0	H ₂₀
S	1h	86.0	S ₀
	3h	38.0	S ₀
	8h	38.0	B _{max}



B ₁	Biologically attached bacterial biomass	$\mu\text{g C/cm}^2$
B ₂	Reversibly adsorbed bacterial biomass	$\mu\text{g C/cm}^2$
B ₃	Free bacteria in the liquid phase	$\mu\text{g C/L}$
H ₁	Rapidly degradable macromolecular organic matter of class 1	mg C/L
H ₂	Slowly degradable macromolecular organic matter of class 2	mg C/L
S	Monomeric substrate that is directly utilisable	mg C/L

Figure 4.2 Processes of the SANCHO model (adapted from Servais *et al.*, (1995))

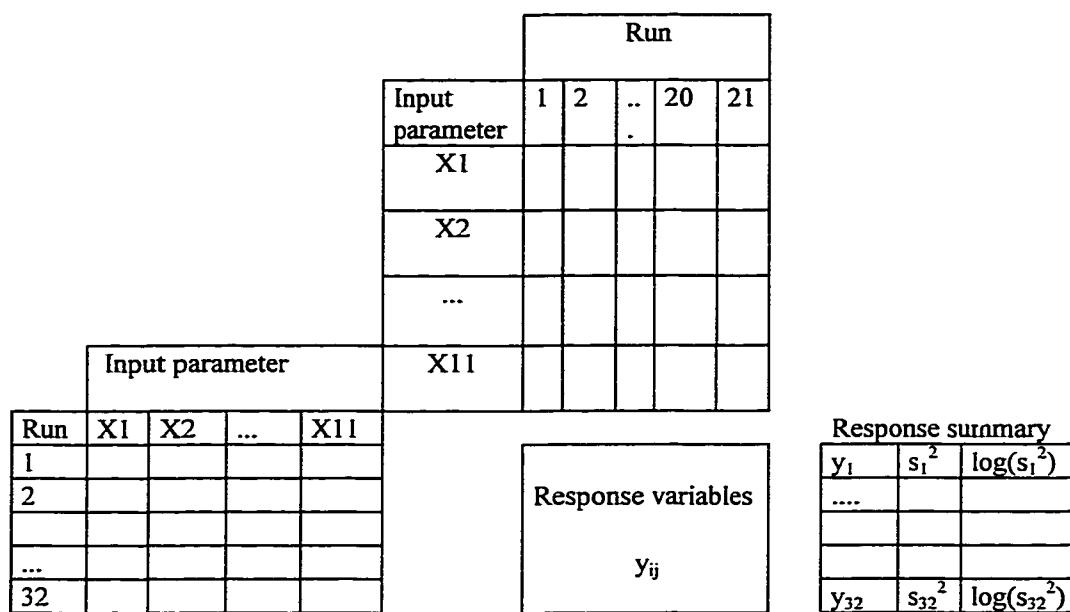


Figure 4.3 Representation of the Taguchi crossed array

X1, X2...X11 : Input (factor) parameter (coded values)

y_{ij} : Response variable at run i of inner array and run j of outer array

where $i = 1, 2, \dots, 32$ and $j = 1, 2, \dots, 21$

$$\bar{y}_i = \frac{1}{21} \sum_{j=1}^{21} y_{ij}$$

$$s_i^2 = \frac{1}{20} \sum_{j=1}^{21} (y_{ij} - \bar{y}_i)^2$$

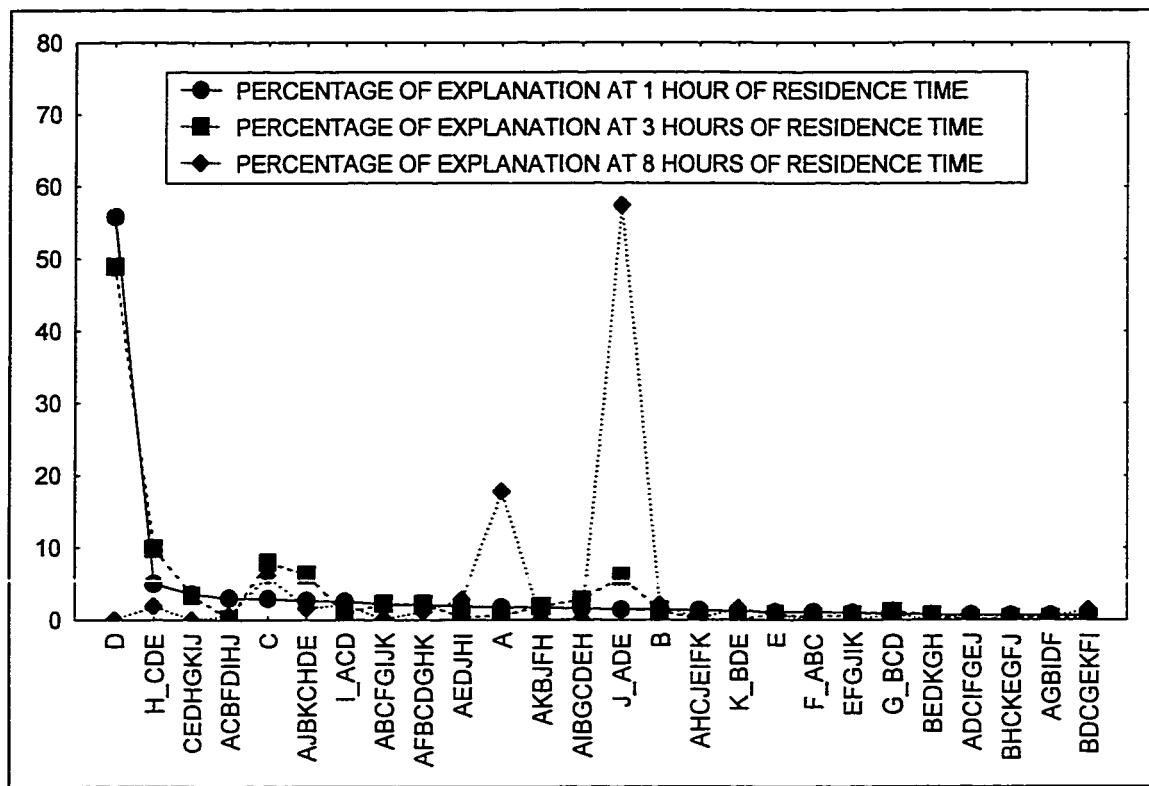


Figure 4.4. Pareto graph for the variance of the response variable B1(fixed bacteria) at 1, 3 and 8 hours of residence time

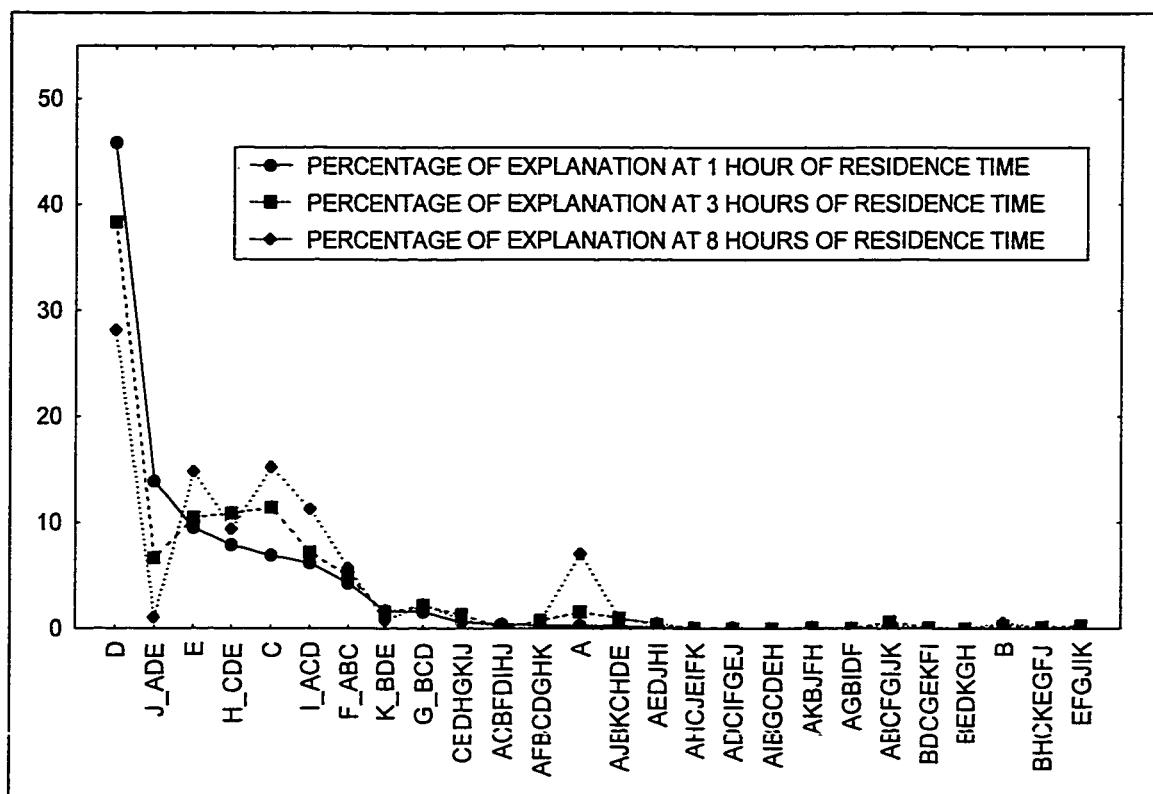


Figure 4.5. Pareto graph for the variance of the response variable B2(adsorbed biomass) at 1, 3 and 8 hours of residence time

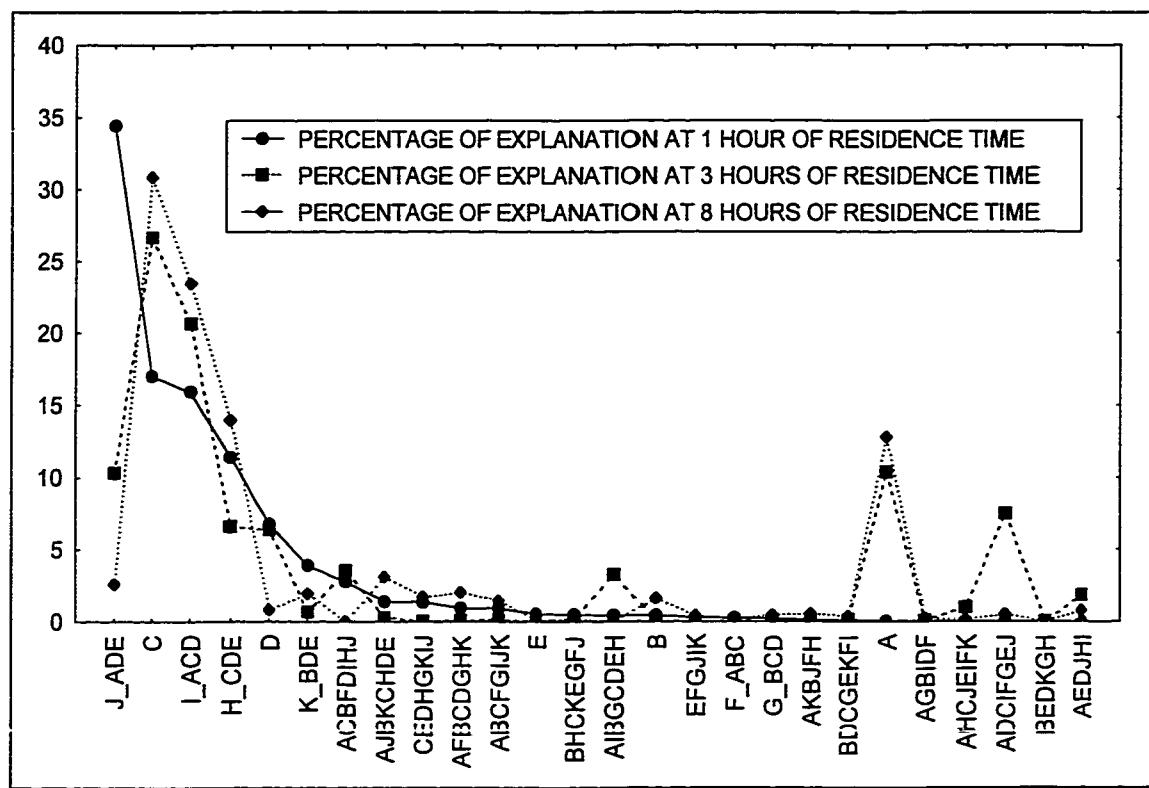


Figure 4.6 Pareto graph for the variance of the response variable B3(free bacteria) at 1, 3 and 8 hours of residence time

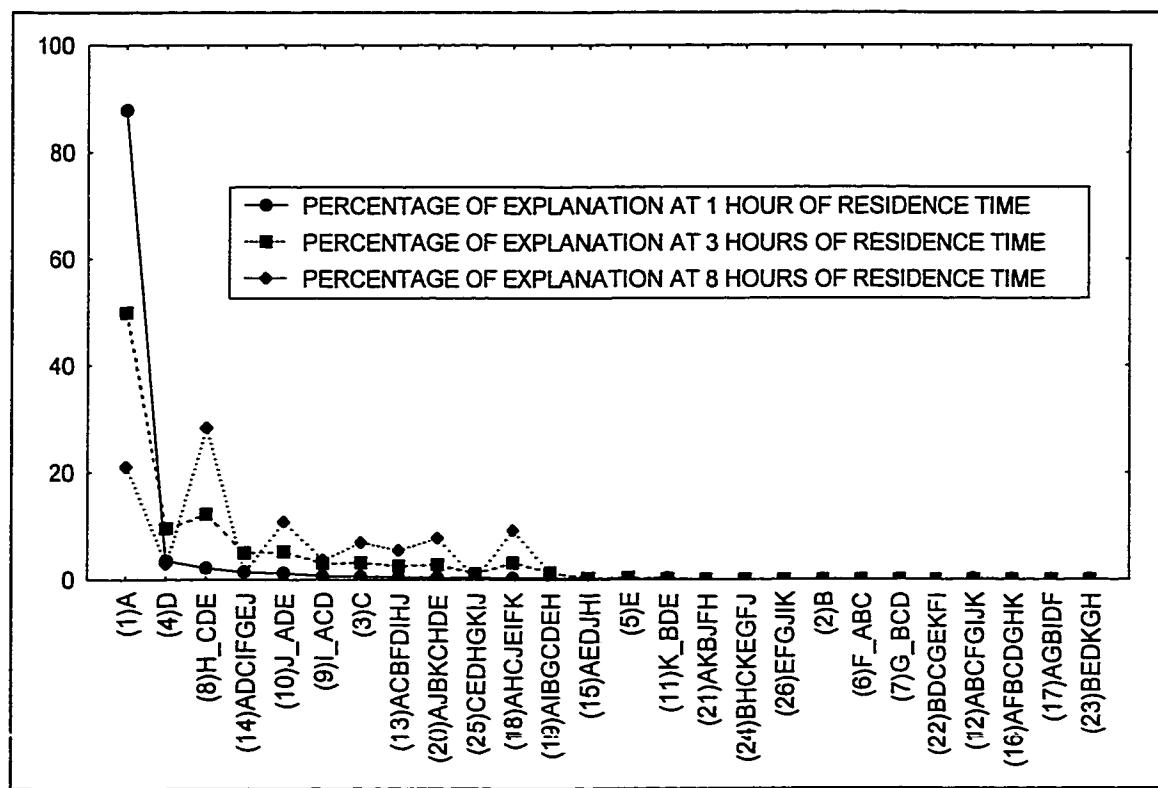


Figure 4.7 Pareto graph for the variance of the response variable H1(liability class 1) at 1, 3 and 8 hours of residence time

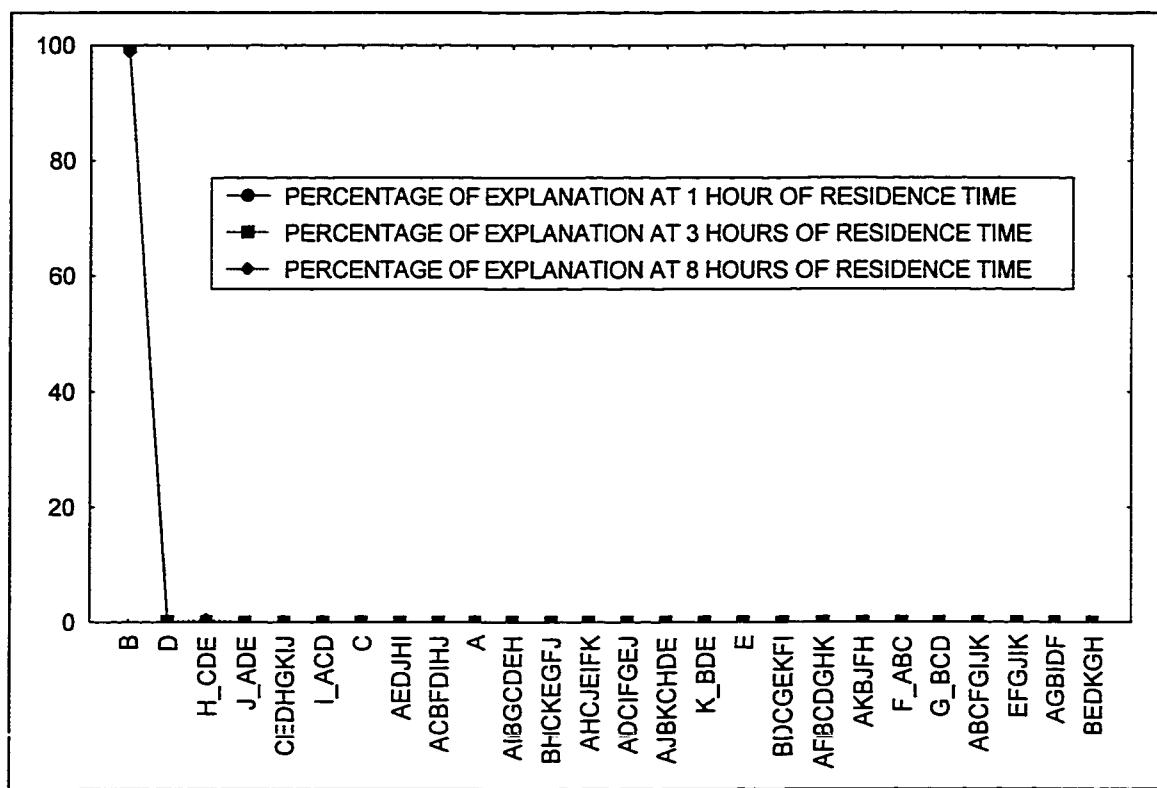


Figure 4.8 Pareto graph for the variance of the response variable H2(liability class 2) at 1, 3 and 8 hours of residence time

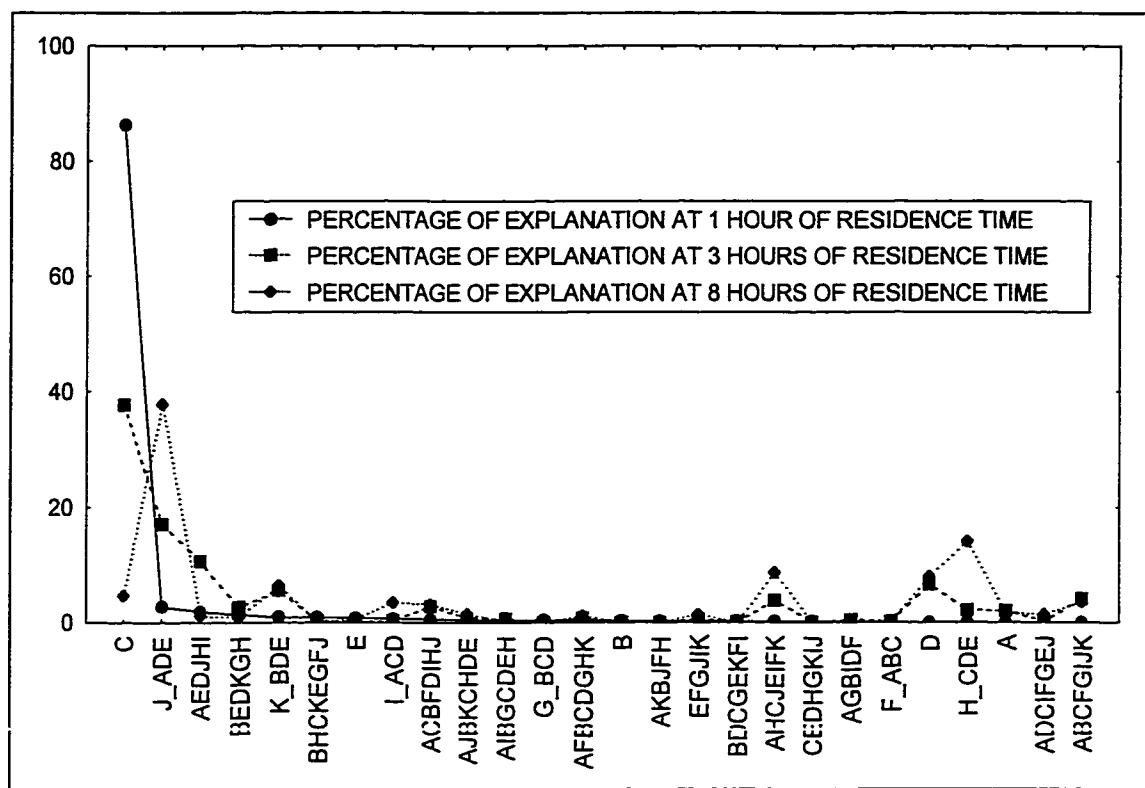


Figure 4.9 Pareto graph for the variance of the response variable S (direct substrates) at 1, 3 and 8 hours of residence time

Chapitre 5. Discussion et Conclusions

L'objectif de cette étude est d'étudier les paramètres du modèle ayant une influence indue sur la variabilité des variables de réponses. Pour atteindre cet objectif, nous avons combiné deux types d'analyses statistiques :

- la première soit le plan factoriel partiel (PFP), qui a été appliqué à deux modalités sur 19 paramètres du modèle. L'analyse a été faite dans un espace expérimental très large, chacun des paramètres prenant une valeur minimale et une valeur maximale, et deux temps de séjour étant étudiés (3 et 8 heures) ;
- la deuxième, le plan croisé de Taguchi (PCT), a été utilisé à trois modalités, (valeur moyenne, +20% et -20%) sur un sous-ensemble de 11 paramètres identifiés comme critiques par la première méthode. Cette méthode a été appliquée à un espace expérimental plus restreint à la fois par le nombre des paramètres à l'étude, mais aussi par l'écart des valeurs des paramètres par rapport à des valeurs moyennes. Par ailleurs, trois temps de séjour sont étudiés soit 1, 3 et 8 heures.

Ce chapitre est une synthèse des résultats expérimentaux obtenus avec les deux méthodes d'analyses, le PFP et le PCT. La comparaison des résultats sera effectuée pour les six variables de réponses. Le Tableau 5.1 résume les pourcentages d'explication de chacune des mêmes variables de réponses.

Tableau 5.1 Pourcentage d'explication de la variabilité pour chacune des six variables de réponses selon le FPF et le PCT.

	Plan Factoriel Partiel (PFP)			Plan Croisé de Taguchi (PCT)		
Variable de réponse	Temps de séjour	Variabilité expliquée (%)	Paramètres critiques pour la variable de réponse en ordre d'importance	Temps de séjour	Variabilité expliquée (%)	Premier paramètre critique pour chaque variable de réponse
B₁	-	-	-	1h	56,0	SPs
	3h	66,0	k_d , SPs, S_0	3h	49,0	SPs
	8h	70,5	k_d , SPs, Y,	8h	57,0	B _{max}
B₂	-	-	-	1h	46,0	SPs
	3h	38,9	SPs, k_{ads} SBs, k_{des} , S_0 k_d	3h	38,0	SPs
	8h	40,1	SPs, k_{ads} , k_d	8h	28,0	SPs
B₃	1h	52,5	S_0 , B _{max} , Y	1h	34,0	B _{max}
	3h	23,9	S_0 , B _{max} , Y	3h	27,0	S ₀
	8h	8,9	S_0 , B _{max} , Y	8h	31,0	S ₀
H₁	-	-	-	1h	96,0	H ₁₀
	3h	82,0	H ₁₀ , SPs, k_d	3h	73,0	H ₁₀
	8h	60,1	H ₁₀ , k_d	8h	34,0	k_d
H₂	-	-	-	1h	100,0	H ₂₀
	3h	97,8	H ₂₀ , S ₀	3h	100,0	H ₂₀
	8h	91,0	H ₂₀ , k_d , S ₀	8h	99,0	H ₂₀
S	-	-	-	1h	86,0	S ₀
	3h	77,7	S_0 , B _{max}	3h	38,0	S ₀
	8h	59,7	S_0 , B _{max} , □	8h	38,0	B _{max}

5.1 Biomasse bactérienne fixée (B1)

Les bactéries fixées sur les parois des conduites des réseaux de distribution sont le résultat d'une colonisation bactérienne. Dans des conditions d'eau potable, cette biomasse

forme un biofilm disparate et non-homogène de quelques dizaines de microns d'épaisseur à la surface de la conduite où des réactions de croissance et de mortalité se produisent.

La première des six variables de réponses du modèle SANCHO est la biomasse bactérienne fixée. Tel qu'indiqué au tableau 5.1, la méthode du PFP indique que la mortalité bactérienne (k_d), la capacité maximale de fixation biologique (SPs) et la concentration initiale de substrats directs (S_0) expliquent 66% de la variabilité totale à 3 heures de temps de résidence et 71% de la variabilité totale de cette variable de réponse à 8 heures de temps de résidence. Lorsque 3 modalités sont considérées, comme avec le plan croisé de Tagushi, le paramètre SPs qui domine en expliquant à lui seul la moitié de la variabilité totale à 1 heure et 3 heures de temps de résidence alors qu'à 8 heures de temps de résidence, c'est le paramètre B_{max} qui explique 57% de la variabilité.

Selon la méthode des PFP, la variabilité des bactéries fixées est influencée surtout par un paramètre microbiologique, la mortalité bactérienne et en second lieu par SPs. Dans un espace expérimental plus restreint à trois modalités, la méthode des PCT montre que la biomasse fixée varie surtout en fonction de SPs, paramètre regroupant l'effet du matériau des conduites ainsi que l'impact des conditions du réseau (forces de cisaillement, etc.). La prédominance de l'influence de SPs avec la méthode du PCT peut aussi être le résultat de :

- Le nombre d'essais et leurs effets estimés. La méthode du PFP est accomplie avec 64 essais générant 19 effets principaux et 43 effets combinés, soit 62 effets au total. Par contre, la méthode des PCT nécessite 672 essais tout en générant seulement 26 effets (11 principaux et 15 d'interaction). La méthode des PCT concentre l'importance des effets estimés car elle effectue un ajustement du modèle de régression avec moins d'effets et en présence de plus d'essais.
- Le type de perturbation du modèle. La méthode du PFP requiert de connaître uniquement la modalité minimale et maximale de chacun des paramètres. Avec cette méthode, les seules valeurs numériques possibles d'un paramètre se situent donc à la

valeur minimale et maximale. La méthode du PCT multiplie valeurs numériques possibles pour chaque paramètre. En effet, la modalité minimale et maximale d'un paramètre est définie de la même façon. Toutefois, une variation de 20% de part et d'autre de ces extrêmes est allouée, soit quatre nouvelles valeurs. Chaque paramètre peut alors prendre six valeurs numériques. La méthode du PCT permet donc de balayer un plus grand nombre de situations possibles que la méthode des PFP, mais dans un espace expérimental moins large.

Le paramètre qui ressort comme ayant une importance sur la variabilité des bactéries fixées est le terme SPs selon les deux méthodes employées. SPs n'est pas mesuré mais défini à partir des mesures du potentiel maximum d'adsorption (SBs) qui peut être mesuré expérimentalement sur différents matériaux. Le nombre de sites pouvant servir à l'attachement est vraisemblablement supérieur à celui pouvant servir pour l'adsorption. La valeur de SPs, la capacité maximale pour l'attachement bactérien est aussi ajustée pour tenir compte des conditions environnementales qui peuvent limiter la croissance d'un biofilm dans des conditions d'eau potable, comme le cisaillement, sur la base des valeurs expérimentales mesurées dans des réseaux pilotes et industriels. En effet, l'ensemble des mesures expérimentales tend à suggérer un maximum de densité de la biomasse bactérienne fixée à la surface des conduites d'eau potable.

L'écart entre SBs et SPs a varié durant la première analyse. SBs n'a pas été considéré pour la deuxième analyse.

Compte tenu de l'influence de ce paramètre sur la prédiction de la biomasse bactérienne fixée en réseau, les différents mécanismes pouvant influencer SBs et SPs sont discutés ci-après.

Adsorption-désorption

Le taux d'adsorption dans SANCHO se définit comme :

$$\text{Déposition} = k_{dep} \cdot B_3 (SB - B_2) = k_{dep} \cdot B_3 \cdot SB - k_{dep} \cdot B_3 \cdot B_2 = K B_3 - k_{dep} B_3 B_2$$

Dans le modèle SANCHO, le détachement des bactéries fixées est proportionnel au nombre de bactéries adsorbées.

$$\text{Désorption} = k_{des} B_2$$

(van der Wende et al. 1989), (Dukan et al. 1996) et (Piriou et al. 1996) proposent aussi un modèle de déposition proportionnel à la concentration des bactéries libres retrouvées dans l'eau distribuée, soit:

$$\text{Déposition} = k_{dep} \cdot B_3$$

Détachement par cisaillement

La vitesse d'écoulement de l'eau dans les conduites du réseau crée un cisaillement à la paroi du biofilm qui arrache une partie de ce biofilm. Selon (Dukan et al. 1996) et (Piriou et al. 1996), le calcul des efforts de cisaillement requiert une connaissance du réseau et de son hydraulique, soit le débit d'eau transporté, la longueur et rugosité des conduites, leur diamètre ainsi que leur disposition. Ces informations doivent être connues pour la totalité du réseau.

Si l'incorporation de l'effet du cisaillement apparaît comme une amélioration souhaitable, il apparaît toutefois difficile ou impossible à effectuer. En effet, du moins dans les réseaux de distribution industriels, les vitesses d'écoulement, et par le fait même les forces de cisaillement à la paroi, varient en fonction des cycles de consommation. Le biofilm en place est donc soumis à des cycles quotidiens de variations des forces de cisaillement et même à des vitesses très élevées lors des interventions spéciales (débit d'incendie ou vidanges orientées). De plus, les vitesses moyennes d'écoulement sont généralement très faibles dans les conduites de petit diamètre. Le détachement provoqué à ces vitesses pourrait alors être négligeable particulièrement à la surface où les vitesses

sont quasi-nulles. Enfin, dans une étude avec des réacteurs annulaires, (Peyton 1992) conclut que le cisaillement ne peut pas être relié au taux de détachement.

Influence du taux de croissance

Selon (Stewart et al. 1993) une modification de la nature du substrat cause une modification dans les paramètres k_s et B_{max} . L'augmentation du taux de croissance augmente la probabilité de détachement des bactéries.

L'effet d'une modification de la composition du substrat est aussi pris en compte par le modèle SANCHO, car le substrat, le CODB, est divisé en classes d'aptitude à la biodégradabilité. Une modification de la nature du substrat se traduit par une modification de la croissance bactérienne. De plus le modèle tient compte de l'impact de la croissance directement en diminuant la probabilité d'attachement après adsorption en fonction de la proportion des sites occupés sur le nombre de sites totaux.

Lors d'une étude de la cinétique de détachement des bactéries fixées dans un réacteur annulaire (Peyton 1992) conclut que le taux de détachement est le mieux modélisé grâce à une expression tenant compte de du taux d'utilisation du substrat (B_{max}), du coefficient de croissance réel (Y) et de l'épaisseur du biofilm (L_f), sous la forme :

$$Dé sorption = k_{des} \cdot Y \cdot B_{max} \cdot L_f$$

Notons qu'en eau potable, l'épaisseur du biofilm est faible. Cette façon permet d'exprimer le taux de détachement de façon indépendante du terme SPs.

Il est donc recommandé de :

Il est donc recommandé de :

- Valider les valeurs du paramètre SPs à l'aide d'essais en conditions contrôlées. Ces essais devraient être effectués sur différents matériaux soumis à diverses forces de cisaillement.
- Maintenir la définition de l'isotherme d'adsorption de Langmuir et le détachement proportionnel à la croissance et au nombre de sites disponibles pour le modèle SANCHO.
- Ne pas inclure l'effet du cisaillement sur le détachement.

5.2 Biomasse bactérienne adsorbée (B2)

Les bactéries adsorbées représentent, dans définition du modèle SANCHO, les bactéries déposées à la surface du matériau. À cet endroit, plusieurs cycles réversibles d'attachement et de détachement sont possibles. En terme d'abondance, les bactéries adsorbées représentent une quantité très faible de la biomasse totale présente dans la conduite. La biomasse bactérienne fixée n'est majoritaire qu'en début de colonisation. Aucune mesure expérimentale de la quantité de bactéries adsorbées n'est possible, étant donné leur faible nombre.

Les deux méthodes de calculs utilisées, le PFP et le PCT, convergent vers le paramètre SPs comme étant le plus influent au niveau de la variabilité des bactéries adsorbées. La méthode du PFP indique que le terme SPs accompagné des autres termes décrivant la cinétique d'adsorption-désorption expliquent 40% de la variabilité totale de la variable à 1 heure et 3 heures de temps de résidence. Quant à la deuxième méthode, celle des PCT, le terme de SPs ressort clairement comme le plus influent sur la variabilité de cette variable de réponse et ce aux trois temps de résidence étudiés. La méthode des PFP met en lumière l'influence conjuguée de tous les termes reliés à la cinétique d'adsorption. En effet, les paramètres k_{ads} , k_{des} , SPs ainsi que SB sont responsables de la variabilité totale étudiée. Toutefois, la méthode du PCT attribue plutôt la variabilité de la variable de réponse

principalement à la valeur de SPs. Notons que le paramètre SB n'a pas été retenu pour cette deuxième série d'analyses.

On peut cependant conclure que le terme SPs demeure le terme clé expliquant la variabilité de cette variable de réponse. Il est donc recommandé, comme pour améliorer la prédiction de la biomasse bactérienne fixée, de mieux définir la valeur de SPs.

5.3 Bactéries Libres (B3)

Le terme de la variable de réponse de la biomasse bactérienne libre représente l'ensemble de la flore bactérienne vivant et se reproduisant dans l'eau interstitielle distribuée dans le réseau de distribution. Bien que représentant moins de 5% de la masse totale de la biomasse, la présence de bactéries viables dans l'eau refoulée est la cause des dépassements de normes microbiologiques. Ces bactéries peuvent aussi être l'origine d'une chaîne trophique dans le réseau de distribution.

La variabilité de cette variable, selon la méthode des PFP, est fortement influencée par la concentration initiale en substrats directs (S_0), la capacité maximale d'absorption de substrats directs (B_{max}) et le coefficient de croissance (Y). La seconde méthode employée, celle des PCT, donne des résultats concordants en identifiant les paramètres B_{max} et S_0 comme critiques à cette variable de réponse. Les résultats obtenus semblent indiquer que la capacité maximale d'utilisation du substrat (B_{max}) et la concentration initiale en substrats directs (S_0) règlent la variabilité des bactéries libres.

L'abondance des bactéries libres dans les réseaux de distributions ne peut provenir uniquement de leur croissance (Servais et al. 1992b). En effet, les coefficients de croissance sont trop faibles pour expliquer l'augmentation de la biomasse en bactéries libres. Les bactéries libres proviennent du détachement des bactéries fixées (Servais et al. 1995). La dépendance des bactéries libres envers les paramètres S_0 et B_{max} est toutefois logique. En tête de réseau (très court temps de séjour), la croissance des bactéries libres est totalement dépendante de la quantité de substrat directement disponible et du taux

maximal d'utilisation de ce substrat, tel que défini dans la relation de Michaelis-Menten. Dès qu'un biofilm s'établit (>1 heure de temps de séjour), c'est à nouveau la quantité de substrats directement assimilables et le taux d'utilisation qui régit le développement du biofilm, et subséquemment la densité des bactéries libres qui s'en détachent. Plus le temps de séjour augmente plus les sources de variabilité sont complexes, reflétant la variabilité de la répartition de la biomasse fixée en fonction du temps de séjour dans les conditions expérimentales testées pour les analyses de sensibilité.

Notons que le paramètre B_{max} a été mesuré expérimentalement sur diverses molécules simples (sucres et acides aminés). Sa valeur est donc assez bien définie. Il est toutefois recommandé de mesurer le paramètre S_0 in-situ afin de bien caractériser la dynamique des bactéries libres en tête de réseau. La mesure de la concentration en substrat est effectuée à l'aide de la méthode de Servais-Billen (Servais et al. 1989).

5.4 Matière polymérique rapidement hydrolysable (H_1)

Le carbone organique dissous biodégradable (CODB) est la source de carbone utilisé par les bactéries présentes dans les réseaux de distribution. Ce CODB est fractionné en trois composantes d'aptitude à la biodégradation. La fraction du CODB qui est rapidement hydrolysée est le terme H_1 . La mesure du CODB est effectuée selon la méthode Servais-Billen (Servais et al. 1989). La première méthode, soit celle des PFP, montre que l'évolution de H_1 est fortement influencée par la valeur initiale de la concentration en matière organique rapidement hydrolysable, H_{10} . Cette tendance est observée aussi bien à 1 heure qu'à 3 heures de temps de résidence. De la même façon, la seconde méthode utilisée, les PCT, confirme l'importance de la concentration initiale en matière organique rapidement hydrolysable (H_{10}) à 1 heure et 3 heures de temps de résidence. Ce n'est qu'à 8 heures de temps de séjour que le terme de mortalité bactérienne (k_d) devient plus important. On constate à quel point la mesure initiale de H_1 se répercute sur la variabilité de la variable de réponse.

Il est donc recommandé de :

- Mesurer la concentration initiale en matière polymérique rapidement hydrolysable in-situ, en prélevant un échantillon de l'eau du réseau, afin de bien caractériser la charge en carbone organique introduite dans le modèle SANCHO. Cette mesure devrait être effectuée à différentes périodes de l'année afin d'obtenir une bonne idée de la variation saisonnière de la composition du CODB.

5.5 Matière polymérique lentement hydrolysable (H₂)

La fraction du CODB lentement hydrolysée est identifiée par le terme H₂. La mesure du CODB est effectuée selon la méthode Servais-Billen (Servais et al. 1989). Les résultats de l'analyse de sensibilité par la méthode des PFP, montre que cette variable est fortement influencée par la valeur initiale de la concentration en matière polymérique lentement hydrolysable, H₂₀. Cette tendance est observée aussi bien à 1 heure qu'à 3 heures de temps de résidence. La méthode d'analyse par les PCT fournit des résultats concordants. De la même façon, la concentration initiale en matière organique lentement hydrolysable explique entièrement la variabilité de cette variable de réponse et ce aux trois temps de séjour. On constate à quel point la mesure initiale de H₂ introduite dans le modèle se répercute sur la variabilité de cette variable de réponse. Le terme de H₂ représente le CODB lentement hydrolysable, il est donc normal de remarquer la forte influence de la concentration initiale de H₂ à un temps de résidence aussi court.

Il est donc recommandé de :

- Mesurer la concentration initiale en matière polymérique lentement hydrolysable in-situ, en prélevant un échantillon de l'eau du réseau, afin de bien caractériser la charge en carbone organique introduite dans le modèle SANCHO. Cette mesure devrait être effectuée à différentes périodes de l'année afin de cerner la variabilité de la composition du CODB.

5.6 Substrats directement utilisables (S)

La fraction du CODB est directement utilisée par les bactéries pour leur croissance est identifiée par le terme S.

Les résultats par la méthode d'analyse des PFP indiquent que cette variable de réponse est fortement influencée par sa valeur initiale de concentration en substrats directement utilisables, S_0 . Cette situation est vraie aussi bien à 1 heure qu'à 3 heures de temps de résidence. De la même façon, l'analyse de cette variable de réponse selon la méthode des PCT, montre que la concentration initiale en substrats directs (S_0) explique une grande partie de la variabilité à 1 heure et à 3 heures de résidence. Ce n'est qu'à 8 heures de temps de séjour que le paramètre B_{max} devient important.

Il est donc recommandé de :

- Mesurer la concentration initiale en substrats directs in-situ, en prélevant un échantillon de l'eau du réseau, afin de bien caractériser la charge en carbone organique introduite dans le modèle SANCHO. Cette mesure devrait être effectuée à différentes périodes de l'année afin de cerner l'ampleur des variations annuelles de cette fraction du CODB.

Il faut noter qu'un équilibre n'est vraiment atteint dans un réseau de distribution d'eau potable qu'après une période de fonctionnement très longue. En effet, particulièrement dans le cas de réseaux de conduites métalliques sujettes à la corrosion, l'atteinte de cet équilibre peut prendre plusieurs années. Les conditions de croissances défavorables en réseau de distribution, soit les faibles concentrations de substrat, la présence d'oxydants, la température faible et les variations des conditions d'écoulement et donc de décrochage, mènent à des taux de croissance faibles qui nécessitent des temps d'équilibrage élevés (plusieurs semaines à plusieurs mois) dans des conditions d'exploitations stables. Or les

conditions d'exploitation des réseaux sont instables reflétant les cycles de demande d'eau. Finalement, l'estimation des temps de séjour, définissant les flux de substrat, est particulièrement difficile en réseaux maillés, et sa définition fine pour des fins de modélisation apparaît impossible.

A la lumière de ces éléments, une prédiction en temps réel de la dynamique bactérienne apparaît actuellement impossible peu importe le logiciel ou modèle utilisé. L'utilisation de modèles statiques, tel le modèle SANCHO, permet de prédire la recroissance bactérienne dans différents secteurs du réseau.

Chapitre 6. Références

- AWWA. (1930). Committee on Water Supply, Bacterial aftergrowth in distribution systems. American Journal of Publications Health, (20), p. 485.
- BILLEN, G., SERVAIS, P., BOUILLOT, P. ET VENTRESQUE, C. (1992). "Funtioning of Biological Filters Used in Drinking Water Treatment-the Chabrol Model." J Water SRT - Aqua, p. 231-241.
- BLOCK, J. C., HAUDIDIER, K., PAQUIN, J. L., MIAZGA, J. ET LÉVI, Y. (1993). Biofilm accumulation in drinking water distribution systems. Biofouling, (6), p. 333-343.
- BOULOS, P. F., ALTMAN, T., JARRIGE, P.-A. ET COLLEVATI, F. (1995). Discrete simulation approach for network-water-quality models. Journal of Water Resources Planning and Management, 121(1), p. 49-60.
- BOX, G. ET DRAPER, N. (1987). Empirical model-building and response surfaces, John Wiley & Sons, New York. 699 p.
- CAMPER, A. K. (1995). Factors influencing biofilm growth in drinking water distribution systems, Ph.D., Montana Stage University, Bozeman, Montana, U.S.A.
- CAMPER, A. K., HAYES, J. T., JONES, W. L. et ZELVER, N. (1993). Persistance of Coliforms in Mixed Populaiton Biofilms. AWWA - WQTC, Miami, Floride.
- CHARACKLIS, W. G. ET MARSHAL, K. C. (1989). Biofilms, John Wiley and Sons.
- CIGANA, J., LAURENT, P., PRÉVOST, M. ET CLÉMENT, B. (1997). Key Parameters Affecting Regrowth in Distribution Systems using the Sancho model. AWWA-WQTC, Denver CO, 14 p.

- COLBOURNE, J. S., DENNIS, P. J., RACHWAL, A. J., HOUSE, S., KEEVIL, W. ET MACKERNESS, C. (1992). The Operational Impact of Growth of Coliforms in London's Distribution System. AWWA - Water Quality Technology Conference, Toronto, Ontario, 11 p.
- COX, D. R. (1992). Planning of experiments , New York. 300 p.
- DESJARDINS, R., JUTRAS, L. ET PRÉVOST, M. (1997). Évolution de la qualité de l'eau dans le réseau de distribution de la Ville de Montréal. Revue des sciences de l'eau, 10(2), p. 167-184.
- DESJARDINS, R., LAVOIE, J., LAFRANCE, P. ET PRÉVOST, M. (1991). Comparaison de l'évolution de la qualité de l'eau dans deux réseaux de distribution. Sciences et techniques de l'eau, 24(4), p. 321-331.
- DUKAN, S., LEVI, Y., PIRION, P., GUYON, F. ET VILLON, P. (1996). Dynamic Modelling of Bacterial Growth in Drinking Water Networks. Water Research, 30(9), p. 1991-2002.
- LAURENT, P. (1995). Colonisation bactérienne des réseaux de distribution: étude expérimentale et modélisation , Université Libre de Bruxelles, Bruxelles.
- LAURENT, P. ET SERVAIS, P. (1995). Fixed Bacterial Biomass Estimated by Potential Exoproteolytic Activity. Canadian Journal of Microbiology, 41(8), p. 749-752.
- LAURENT, P., SERVAIS, P., PRÉVOST, M., GATEL, D. ET CLÉMENT, B. (1997). Testing the SANCHO model on distribution systems. Journal of American Water Works Association, 89(7), p. 92-103.

LECHEVALLIER, M. W. (1990). Coliform Regrowth in Drinking Water: A Review.

J.AWWA, 82(11), p. 74-86.

LEHMAN, J., BOTKIND, D. ET LIKEN, C. (1975). The assumptions and rationals of computer model of phytoplankton dynamics. Limnology and oceanography, 20, p. 343-364.

MACLEAN, R. (1995). Enlèvement du COB (carbone organique biodégradable) par la filtration biologique: optimisation des méthodes de mesure et des paramètres d'exploitation, M.Sc.A., École Polytechnique, Montréal.

MATHIEU, L. (1992). La matière organique et la stabilité biologique des eaux au cours de leur distribution. Thèse de doctorat, Université de Nancy I, France.

PAYMENT, P., FRANCO, E., RICHARDSON, L. ET SIEMIATYCKI, J. (1991).

Gastrointestinal Health Effects Associated with the Consumption of Drinking Water Produced by Point-of-Use Domestic Reverse-Osmosis Filtration Units. Applied and Environmental Microbiology, 57(4), p. 945-948.

PEYTON, B. M. (1992). Kinetics of biofilm detachment, Ph.D., Montana State University, Bozeman, Montana.

PIRIOU, P., DUKAN, S., LÉVI, Y., GUYON, F. ET VILLON, P. (1996). Modélisation du comportement des biomasses bactériennes libres et fixées dans les réseaux de distribution d'eau potable. Revue des sciences de l'eau, 9(3), p. 381-406.

- PRÉVOST, M., COALLIER, J., ROMPRÉ, A., SERVAIS, P., LAURENT, P. ET LAFRANCE, P. (1997a). Suspended Bacterial Biomass and Activity in Drinking Water Distribution Systems: Impact of Water Treatment. Water Research, in press.
- PRÉVOST, M., ROMPRÉ, A., COALLIER, J. ET SERVAIS, P. (1997b). Impact of reducing biodegradable organic matter on fixed and suspended biomass in distribution systems. Microbial Impact of Biological Filtration, AWWA, ed., 427 p.
- ROGERS, H. J. (1961). The dissimulation of high molecular weight organic substances. The Bacteria, Academic Press, New-York.
- ROSSMAN, L. A. (1993). EPANET Users Manual, , Cincinnati.
- SERVAIS, P. (1986). Étude de la dégradation de la matière organique par les bactéries hétérotrophes en rivière. Développement d'une démarche méthodologique et application à la Mause belge, , Université Libre de Bruxelles, Bruxelle, Belgique.
- SERVAIS, P., ANZIL, A. ET VENTRESQUE, C. (1989). Simple Method for Determination of Biodegradable Dissolved Organic Carbon in Water. Applied and Environmental Microbiology, 55(10), p. 2732-2734.
- SERVAIS, P., BILLEN, G., LAURENT, P., LÉVI, Y. ET RANDON, G. (1992a). Studies of BDOC and Bacterial Dynamics in the Drinking Water Distribution System of the Northern Parisian Suburbs. Revue des sciences de l'eau, 5 (Spécial COA), p. 69-89.
- SERVAIS, P., LAURENT, P., BILLEN, G. ET GATEL, D. (1995). Development of a model of BDOC and bacterial biomass fluctuations in distribution systems. Revue des sciences de l'eau, 8, p. 427-462.

- SERVAIS, P., LAURENT, P., BILLEN, G. ET LEVI, Y. (1992b). Étude de la colonisation bactérienne des réseaux de distribution. Revue des sciences de l'eau, 87(6), p. 321-326.
- SMITH, D. B. ET HESS, A. F. (1990). Control of Distribution System Coliform Regrowth. AWWA - Water Quality Technology Control, San Diego, Californie, 23 p.
- SMITH, D. B., HESS, A. F. ET HUBBS, S. A. (1990). Survey of distribution system coliform occurrence in the United States. Journal of American Water Works Association, San Diego.
- STATSOFT, I. (1996). Statistica for Windows, , Tulsa, OK
- STEWART, P. S., PEYTON, B. M., DRURY, W. J. ET MURGA, R. (1993). Quantitative observations of heterogeneities in *Pseudomonas aeruginosa* biofilms. Applied and Environmental Microbiology, 59(1), p. 327-329.
- TAGUCHI, G. (1987). System of Experimental Design. International Publications, White Plains, New-York.
- VAN DER WENDE, E., CHARACKLIS, W. G. ET SMITH, D. B. (1989). Biofilms and Bacterial drinking Water Quality, 23 p.
- VOLK, C. ET JORET, J. C. (1994). Paramètres prédictifs de l'apparition des coliformes dans les réseaux de distribution d'eau d'alimentation. Revue des sciences de l'eau, 7, p. 131-152.