



# **Probabilistic design of wastewater treatment plants**

**Thèse**

**Mansour Talebizadehsardari**

**Doctorat en génie des eaux**

Philosophiae doctor (Ph.D.)

Québec, Canada

© Mansour Talebizadehsardari, 2015



## **RÉSUMÉ**

Dans cette étude, une méthode de conception probabiliste pour la conception d'usines de traitement des eaux usées (STEP) qui permet la quantification du degré de conformité aux normes de rejet en termes de probabilité a été proposée. La méthode de conception développée est une approche basée sur un modèle dans lequel les sources pertinentes d'incertitude sont exprimées en termes de fonctions de distribution de probabilité et leur effet combiné sur la distribution de la qualité de l'effluent sont quantifiés par simulation Monte Carlo.

Pour ce faire, une série de conceptions en régime permanent avec différents niveaux de sécurité ont été générés en utilisant des règles de conception et la probabilité correspondante de non-conformité aux normes de rejet ont été calculés en utilisant des simulations dynamiques avec différentes réalisations de séries temporelles d'affluent et différentes valeurs pour les paramètres du modèle de station d'épuration.

Pour générer différentes réalisations de séries temporelles d'affluent, un logiciel a été développé pour la génération de séries temporelles d'affluent tenant compte des conditions climatiques locales ainsi que les caractéristiques de base des réseaux d'égout connectés. En outre, différentes réalisations des paramètres du modèle de STEP ont été générés par l'échantillonnage des fonctions de distribution de probabilité et l'effet combiné de la variabilité de l'affluent et l'incertitude du modèle sur la distribution de la qualité des effluents a été calculé en exécutant un certain nombre de simulations Monte Carlo jusqu'à ce que la convergence de la distribution de la qualité des effluents a été atteinte. Une fois la convergence atteinte, la probabilité de non-respect d'une alternative de conception peut être calculée pour un certain niveau de qualité de l'effluent.

La méthode de conception probabiliste peut aider les concepteurs à éviter l'application de facteurs de sécurité conservateurs qui pourraient entraîner un dimensionnement trop petit ou

trop grand de stations d'épuration. En outre, le calcul de la probabilité de non-conformité comme un critère quantitatif peut aider les concepteurs et les décideurs à prendre une décision informée des risques en vue de la meilleure configuration de traitement, son dimensionnement, ainsi que le fonctionnement de l'usine pour la conception ou la mise à niveau des stations d'épuration.

**Mots-clés:** usine de traitement des eaux usées, de conception dans l'incertitude, de la conception en fonction du risque, analyse de l'incertitude, probabilité de non-conformité.

## **ABSTRACT**

In this study, a probabilistic design method for the design of wastewater treatment plants (WWTP) that enables the quantification of the degree of compliance to the effluent standards in terms of probability has been proposed. The developed design method is a model-based approach in which relevant sources of uncertainty are expressed in terms of probability distribution functions and their combined effect on the distribution of the effluent quality is quantified by Monte Carlo simulation.

To do so, a set of steady-state designs with different levels of safety is first generated using a design guideline and then the corresponding probability of non-compliance (PONC) to the effluent standards is calculated using dynamic simulations under different realizations of influent time series and different values for the WWTP model parameters.

To generate different realizations of the influent time series, a software tool was developed for synthetic generation of influent time series data considering the local climate conditions as well as basic characteristics of the connected sewershed. Moreover, different realizations of WWTP model parameters are generated by sampling from the probability distribution functions that are assigned to uncertain model parameters. The combined effect of influent variability and model uncertainty on the effluent distributions is calculated by running a certain number of Monte Carlo simulation runs until convergence of the effluent distribution is achieved. Once convergence is reached for the effluent distribution, the PONC for a design alternative can be calculated for a certain effluent standard.

The probabilistic design method can help designers avoid the application of conservative safety factors that could result in over-or under-sizing of WWTPs. Moreover, calculating the probability of non-compliance as a quantitative criterion can help designers and decision makers make risk-informed decisions on the best treatment configuration, sizing, and plant operation during the design or upgrading of WWTPs.

Keywords: Wastewater treatment plant, design under uncertainty, risk-informed design, uncertainty analysis, probability of non-compliance.

# TABLE OF CONTENTS

<b>RÉSUMÉ</b> .....	<b>III</b>
<b>ABSTRACT</b> .....	<b>V</b>
<b>TABLE OF CONTENTS</b> .....	<b>VII</b>
<b>LIST OF FIGURES</b> .....	<b>XI</b>
<b>LIST OF TABLES</b> .....	<b>XIX</b>
<b>NOMENCLATURE</b> .....	<b>XXI</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>XXIII</b>
<b>CHAPTER 1: INTRODUCTION</b> .....	<b>1</b>
<b>CHAPTER 2: LITERATURE REVIEW</b> .....	<b>5</b>
2.1 TRADITIONAL DESIGN USING DESIGN GUIDELINES .....	5
2.1.1 <i>ATV-DVWK-A 131E design guideline</i> .....	7
2.1.2 <i>Metcalf &amp; Eddy design guideline</i> .....	10
2.1.3 <i>WRC design guideline</i> .....	12
2.1.4 <i>Ten States Standards-recommended standards for wastewater facilities</i> .....	14
2.1.5 <i>Summary of design guidelines</i> .....	17
2.2 MODEL-BASED DESIGN AND OPTIMIZATION .....	18
2.3 ROBUST OPTIMAL DESIGN.....	23

2.4	PROBABILITY-BASED DESIGN.....	25
2.4.1	<i>Definition of the PONC based on the load-resistance concept.....</i>	26
2.4.2	<i>Types and characterization of uncertainty.....</i>	27
2.4.3	<i>Analytical methods for estimating the PONC.....</i>	32
2.4.4	<i>Monte Carlo Method for estimating the PONC.....</i>	38
2.5	COST ANALYSIS.....	44
2.6	PROBLEM STATEMENT.....	48
2.7	OBJECTIVES.....	50
<b>CHAPTER 3: PROPOSED PROBABILISTIC DESIGN METHODOLOGY.....</b>		<b>53</b>
3.1	GENERATING A SET OF STEADY STATE PRE-DESIGNS WITH DIFFERENT LEVELS OF SAFETY.....	56
3.1.1	<i>Design inputs and outputs using a design guideline and/or steady state model</i> 56	
3.1.2	<i>Monte Carlo simulations for generating a set of pre-designs.....</i>	57
3.2	PRELIMINARY EVALUATION AND SCREENING OF PRE-DESIGNS.....	59
3.3	QUANTIFICATION OF PONC USING MONTE CARLO SIMULATION.....	63
3.3.1	<i>Synthetic generation of influent time series.....</i>	66
3.3.2	<i>Uncertainty characterization and random generation of WWTP model parameters.....</i>	70
3.3.3	<i>Propagation of uncertainty and variability using Monte Carlo simulation....</i>	74
3.4	OUTPUT ANALYSIS.....	82
3.4.1	<i>Convergence test for one-dimensional Monte Carlo simulation.....</i>	84
3.4.2	<i>Convergence test for two-dimensional Monte Carlo simulation.....</i>	85
3.5	CALCULATION OF PONC AND TOTAL COST OF DESIGN ALTERNATIVES.....	86
3.5.1	<i>Calculation of PONC.....</i>	86
3.5.2	<i>Calculation of total cost.....</i>	88
<b>CHAPTER 4: INFLUENT GENERATOR FOR PROBABILISTIC MODELING OF NUTRIENT REMOVAL WASTEWATER TREATMENT PLANTS.....</b>		<b>91</b>

4.1	INTRODUCTION.....	92
4.2	METHODOLOGY.....	95
4.2.1	<i>Weather and air temperature generators.....</i>	98
4.2.2	<i>Influent generation in DWF conditions .....</i>	106
4.2.3	<i>Influent generation in WWF conditions.....</i>	107
4.2.4	<i>Bayesian model calibration of the CITYDRAIN sewer model.....</i>	111
4.2.5	<i>Synthetic influent generation .....</i>	112
4.3	DATA AND CASE STUDY .....	113
4.4	RESULTS AND DISCUSSION.....	115
4.4.1	<i>Synthetic generation of rainfall.....</i>	115
4.4.2	<i>Synthetic generation of air and bioreactor temperature .....</i>	120
4.4.3	<i>Multivariate auto-regressive model for DWF generation .....</i>	123
4.4.4	<i>CITYDRAIN model calibration and synthetic influent generation .....</i>	125
4.4.5	<i>Synthetic generation of influent time series .....</i>	132
4.5	CONCLUSION .....	133

**CHAPTER 5: APPLICATION OF A NEW PROBABILISTIC DESIGN METHOD FOR WASTEWATER TREATMENT PLANTS-CASE STUDY: EINDHOVEN**

**WWTP 135**

5.1	INTRODUCTION.....	136
5.2	PROPOSED PROBABILISTIC DESIGN METHOD .....	139
5.2.1	<i>Steady state pre-designs with different levels of safety.....</i>	139
5.2.2	<i>Preliminary Evaluation of and screening of design alternatives.....</i>	142
5.2.3	<i>Quantification of PONC and Total Cost.....</i>	143
5.2.4	<i>Data and case study.....</i>	146
5.3	RESULTS.....	150
5.3.1	<i>Steady state pre-designs with different levels of safety.....</i>	150
5.3.2	<i>Preliminary evaluation and screening of pre-designs .....</i>	154
5.3.3	<i>Quantification of PONC and the total cost for the selected designs .....</i>	157
5.4	CONCLUSION AND RECOMMENDATIONS .....	175

**CHAPTER 6: GENERAL CONCLUSIONS AND PERSPECTIVES.....181**

6.1 CONCLUSIONS.....181

6.1.1 *A practical framework for the probabilistic design of WWTPs .....181*

6.1.2 *Development of an influent generator using the basic characteristics of sewershed and climate data .....183*

6.1.3 *Rigorous calculation and communication of PONC concept .....184*

6.1.4 *Application of the proposed methodology to a real case study.....184*

6.2 PERSPECTIVES.....185

6.2.1 *Calculation of conditional PONC in view of future anthropological and climate change .....185*

6.2.2 *Including the uncertainty in the performance of technical components of WWTPs 185*

6.2.3 *Devising improved control strategies to reduce the risk of non-compliance 186*

**REFERENCES .....187**

**APPENDIX .....207**

## LIST OF FIGURES

Figure 2-1 Design inputs and outputs for design guidelines (top) and process models for design evaluation using process models (bottom)(adapted from Corominas et al. (2010)) ...	6
<b>Figure 2-2</b> Inputs, basis, and outputs of steady state (left) and dynamic models (right) (Ekama, 2009).....	21
<b>Figure 2-3</b> Uncertainty: a three-dimensional concept (Walker et al., 2003).....	28
<b>Figure 2-4</b> Sources of uncertainty (Tung and Yen, 2005) .....	30
<b>Figure 2-5</b> Different uncertainty levels (Walker et al., 2003).....	30
<b>Figure 2-6</b> Failure surface for a performance function with two uncertain parameters.....	36
<b>Figure 2-7</b> Monte Carlo simulation framework for the probabilistic design of WWTPs (Bixio et al., 2002) .....	40
<b>Figure 2-8</b> The relationship between the probability of compliance (1-PONC) and annual cost for two design alternatives (Adapted from Doby, 2004) ○ Alternative 1 △ Alternative 2 .....	43
Figure 3-1 Steps for the proposed probabilistic design of WWTPs .....	55
<b>Figure 3-2</b> Making a set of design alternatives using different input values. ....	58

Figure 3-3 Feasible pre-designs with total bioreactor smaller than  $10^5 \text{ m}^3$  and depth of the secondary clarifier lower than 2.5m. ....60

Figure 3-4 Generated pre-designs and cluster locations for a k-means clustering with 7 clusters.....61

**Figure 3-5** Calculation of effluent CDFs for the initial design alternatives .....63

**Figure 3-6** Calculating PONC for a design alternative .....65

Figure 3-7 Schematic of the proposed influent generator (detailed in Chapter CHAPTER 4:.).....67

**Figure 3-8** (a) Schematic of a RS and (b) LHS procedures.....72

Figure 3-9 Schematic of a two-dimensional Monte Carlo simulations.....76

Figure 3-10 A cloud of effluent CDFs resulting from a two-dimensional Monte Carlo simulation .....78

**Figure 3-11** Identifying a typical year-long influent time series .....81

Figure 3-12 CDF cloud generated for the effluent concentrations using a single year-long influent realization and different parameter values sample using a one-dimensional Monte Carlo simulation .....81

Figure 3-13 Aggregation of the simulated concentration time series ( $\Delta t=15$ ) into daily values and calculation of PONC using the empirical CDF (bottom) .....83

<b>Figure 3-14</b> Percentage change in the values of the average, 5 <sup>th</sup> , 50 <sup>th</sup> and 95 <sup>th</sup> percentiles as a function of Monte Carlo iterations .....	84
<b>Figure 3-15</b> Convergence of a two-dimensional Monte Carlo simulation.....	85
Figure 3-16 Calculation of PONC corresponding to an effluent standard.....	87
Figure 3-17 CDF for the number of non-compliance events in a year based on <i>N</i> years of simulation.....	88
<b>Figure 4-1</b> Schematic of the proposed influent generator .....	97
<b>Figure 4-2</b> Schematic of a two-state Markov chain with the two states being wet (W) or dry (D) and four transitions between them.....	98
Figure 4-3 Schematic of Bartlett-Lewis rainfall model .....	103
Figure 4-4 Four types of rainfall patterns for a three-hour rainfall sequence (Ormsbee (1989)).....	104
Figure 4-5 Schematic of virtual basin. a) filling of the basin in wet periods where .....	108
<b>Figure 4-6</b> Seasonal variation in the Markov chain-gamma model parameters.....	116
<b>Figure 4-7</b> A year-long realization of daily rainfall (right) versus an observed one (left).	116
<b>Figure 4-8</b> Cumulative distribution function of daily rainfall in the studied Eindhoven catchment .....	118

**Figure 4-9** Random generation of air and bioreactor temperature for one year for the Eindhoven WWTP: Seasonal variation in mean maximum and mean minimum air temperatures in wet days (a), in dry days (b), Seasonal variation in standard deviations of maximum and minimum air temperatures in wet days (c), in dry days (d), Randomly generated daily air temperature (e), Linear regression model between daily air and bioreactor temperatures (f), Bioreactor temperature time series with 15-minute temporal resolution (g), Average diurnal variation of bioreactor temperature (h), Observed historical bioreactor temperatures used in the analysis (i) .....122

**Figure 4-10** Variation of SBC with order of multivariate time series model .....123

**Figure 4-11** Observed and most likely simulated influent time series under DWF conditions .....124

**Figure 4-12** Posterior distribution of parameters for flow calibration where, runoff coeff, init loss, and perm loss are respectively the runoff coefficient, initial loss (mm), permanent loss (mm/day) parameters in the virtual basins model that is used in the CITYDRAIN model, K (sec) and X are the routing parameters used in the Muskingum method, and Sigma is the standard deviation of the residual error. The blue histograms represent the marginal posterior distribution of the individual parameters and the red scatter plots represent relationships corresponding to various combinations of parameters (Equation 4-14) .....128

**Figure 4-13** Posterior distribution of parameters used for TSS calibration where Ka is the accumulation coefficient (1/day), m\_lim is the maximum accumulated mass (kg/ha), We, and w are the calibration parameters (Equation 4-16) .....129

**Figure 4-14** Uncertainty bands for flow (top) and TSS concentration (bottom) in a 4-day wet weather period (Rain series in blue and maximum likelihood simulation in black) ....130

<b>Figure 4-15</b> CDFs of daily-averaged influent flow and concentration of influent pollutants .....	131
<b>Figure 4-16</b> CDFs of hourly-averaged influent flow and concentration of influent pollutants .....	131
<b>Figure 4-17</b> A 7-day realization of rainfall and influent time series (flow and composition) .....	132
Figure 5-1 Steps for the proposed probabilistic design of WWTPs .....	141
<b>Figure 5-2</b> Calculating PONC for a design alternative .....	145
Figure 5-3 Arial view and the schematics of of the Eindhoven WWPT (Schilperoort, 2011) .....	148
<b>Figure 5-4</b> Distribution of the generated 5000 pre-designs and the centroids locations corresponding to the k-means clustering with 7 centroids (i.e. the red dots) .....	152
<b>Figure 5-5</b> CDF graphs for the effluent of design alternatives .....	156
Figure 5-6 Location of final design alternatives in the space of design outputs.....	157
<b>Figure 5-7</b> Fluctuations in percentage change values of the average, 5 <sup>th</sup> , 50 <sup>th</sup> , and 95 <sup>th</sup> percentiles for the effluent NH <sub>4</sub> and TN distributions .....	160
<b>Figure 5-8</b> Daily mean effluent BOD distributions for 5 different design alternatives for one-dimensional Monte Carlo (i.e. influent variability and parameter uncertainty mixed)	

and the pragmatic Monte Carlo simulation (i.e. influent variability with model parameters set at the nominal and “worst case” parameter sets) .....161

**Figure 5-9** Daily mean effluent COD distributions for 5 different design alternatives (Alt1-Alt5) for one-dimensional Monte Carlo (i.e. influent variability and parameter uncertainty mixed) and the pragmatic Monte Carlo simulation (i.e. influent variability with model parameters set at the nominal and “worst case” parameter sets).....162

**Figure 5-10** Yearly mean effluent TSS distributions for 5 different design alternatives (Alt1-Alt5) for one-dimensional Monte Carlo (i.e. influent variability and parameter uncertainty mixed) and the pragmatic Monte Carlo simulation (i.e. influent variability with model parameters set at the nominal and “worst case” parameter sets).....163

Figure 5-11 Daily mean effluent NH<sub>4</sub> distributions for 5 different design alternatives (Alt1-Alt5) for one-dimensional Monte Carlo (i.e. influent variability and parameter uncertainty mixed) and the pragmatic Monte Carlo simulation (i.e. influent variability with model parameters set at the nominal and “worst case” parameter sets).....166

Figure 5-12 Yearly mean effluent TN distributions for 5 different design (Alt1-Alt5) alternatives for one-dimensional Monte Carlo (i.e. influent variability and parameter uncertainty mixed) and the pragmatic Monte Carlo simulation (i.e. influent variability with model parameters set at the nominal and “worst case” parameter sets).....167

**Figure 5-13** Relationship between PONC values (calculated using one-dimensional Monte Carlo simulation, i.e. Mixed and the pragmatic Monte Carlo simulation, i.e. Nominal and “Worst Case”) and the total bioreactor volumes corresponding to the final design alternatives in Table 5-4. ....170

**Figure 5-14** CDFs for the number of non-compliance days with NH<sub>4</sub> standard (i.e. daily mean NH<sub>4</sub> concentrations more than 2 mg/l) in a year calculated using one-dimensional

Monte Carlo simulation (i.e. Mixed) and the pragmatic Monte Carlo simulation (i.e. Nominal and “Worst case”). ..... 171

Figure 5-15 Total cost breakdown for the final design alternatives in Table 5-4 calculated using CAPDET ..... 174

**Figure 5-16** Relationship between PONC values (calculated using one-dimensional Monte Carlo simulation, i.e. Mixed and the pragmatic Monte Carlo simulation, i.e. Nominal and “Worst Case”) and the total cost corresponding to the final design alternatives in Table 5-4. .... 175



## LIST OF TABLES

<b>Table 2-1</b> Permissible aeration tank capacities and loadings in the 10 State Standards (Board, 2004) .....	15
<b>Table 2-2</b> Permissible surface overflow rates and peak solids loading rates in 10 State Standards (Board, 2004) .....	16
<b>Table 2-3:</b> Examples of total cost function (Doby, 2004).....	47
<b>Table 2-4</b> Comparison of different design approaches .....	49
<b>Table 4-1</b> Summary of the data type and their applications.....	114
<b>Table 4-2</b> Average rainfall amount and number of wet days for the Eindhoven catchment .....	119
<b>Table 4-3</b> Basic statistics of hourly rainfall data for the Eindhoven catchment.....	120
<b>Table 4-4</b> Correlation matrix for the generated and observed influent time series in DWF .....	125
<b>Table 4-5</b> Prior distribution of parameters and the values for the maximum likelihood function .....	127
<b>Table 5-1</b> Effluent standards .....	149

**Table 5-2** Range of values assigned to the ATV design guideline (uniform distribution).151

**Table 5-3** Dimensions of the preliminary selected design alternatives (cluster centroids) 153

**Table 5-4** Dimensions of the final design alternatives .....156

**Table 5-5** PONC values for different design alternatives calculated using one-dimensional Monte Carlo simulation (i.e. Mixed) and the pragmatic Monte Carlo simulation (i.e. Nominal and “Worst Case”).....168

**Table 5-6** Probability of having a certain number of non-compliance days with the NH<sub>4</sub> effluent standard in a year calculated using one-dimensional Monte Carlo (i.e. Mixed) and the pragmatic Monte Carlo simulations (i.e. Nominal and “Worst Case”).....172

## NOMENCLATURE

AFOSM	Advanced first order second moment
ASM	Activated sludge model
BOD	Biological oxygen demand
BSM	Benchmark simulation model
CDF	Cumulative distribution function
COD	Chemical oxygen demand
CSS	Combined sewage system
DREAM	Differential evolution adaptive Metropolis
DWF	Dry weather flow
F/M	Food to microorganisms ratio
GA	Genetic algorithm
HRT	Hydraulic retention time
MCMC	Markov chain Monte Carlo
MFOSM	Mean-Value first order second moment
MLSS	Mixed liquor suspended solids
MLVSS	Mixed liquor volatile suspended solids
NH <sub>4</sub>	Ammonia
P	Phosphorus
PDF	Probability distribution function
PONC	Probability of non-compliance
q <sub>A</sub>	Surface overflow rate
q <sub>sv</sub>	Sludge volume loading rate
SC	Secondary clarifier
SDNR	Specific denitrification rate
SS <sub>IST</sub>	Suspended solids concentration in the influent to the secondary clarifier
SVI	Sludge volume index
TKN	Total Kjeldahl nitrogen

TN	Total nitrogen
TSS	Total solids concentration
$T_{th}$	Sludge thickening time
WWF	Wet weather flow
WWTP	Wastewater treatment plant
ZSV	Zone settling velocity

## **ACKNOWLEDGEMENTS**

I would like to thank my supervisor Professor Peter A Vanrolleghem and my co-supervisor Dr Evangelia Belia as well as Dr Marc Neumann, Dr Cristina Martin and IWA's Design and Operation Uncertainty Task group (DOU) who had an important role in defining the project. I would also like to thank the other model *EAU* members that I had the opportunity to work with during the past four years. I would like to express my gratitude to our research partners namely Stefan Weijers, Youri Amerlinck, and Filip Claeys who provided us with materials and insights that were indispensable for the completion of this research work. I would like to thank the Natural Science and Engineering Research Council (NSERC), Primodal Inc, and the Hampton Roads Sanitation District (HRSD) for the financial support of this PhD research.

I would like to express my gratitude to the reviewing committee members: Professor François Anctil, Professor Paul Lessard, and Oliver Schraa for their constructive comments which substantially improved the contents of the thesis.

Last but certainly not least, I would like to express my deepest gratitude to my parents and my family who supported me throughout the different challenges in my life.



## **CHAPTER 1: INTRODUCTION**

Wastewater treatment plants (WWTPs) constitute one of the important water infrastructures and their performance in removing pollutants according to a set of effluent standards plays a pivotal role in protecting the environment and public health. The wastewater entering a WWTP system is usually composed of many different substances with different removal rates which depend on several environmental factors as well as the characteristics of the WWTP (i.e. WWTP size and configuration, process operation, etc). Although tracking all the different biological and physical processes that are taking place inside a WWTP system might not be possible at the micro level, years of experience in the design and operation of WWTPs, studies on pilot plants, and better understanding of the treatment processes have resulted in the development of guidelines for the design of WWTP systems. According to the design guidelines, representative values are determined for design inputs (e.g. influent flow and load, safety factors, effluent permits) and the dimensions of different treatment units are calculated using a set of experience-based rules, empirical equations, and process-based equations, most often under steady state conditions (ATV, 2000; Tchobanoglous et al., 2003). The performance of a WWTP that is designed according to a specific design guideline is contingent on proper selection of design inputs and efficient operational strategies which in turn depends to a great extent on the experience of design engineers.

Recent advances in the computational power of computers and better understanding of various treatment processes have resulted in the development of mathematical process-based models. Daigger and Grady (1995) identified three categories of models including stoichiometric, refined stoichiometric and kinetic/stoichiometric models and elaborated their underlying assumptions and applications. In models with full kinetics and stoichiometry (in this report they are simply referred to as dynamic models), the model constituents are structured into fundamental components and the underlying reactions are expressed in terms of mathematical equations and are capable of simulating the dynamics of

the treatment process (e.g. the family of Activated Sludge Models (ASM) (Henze et al., 2000)).

Dynamic simulation models have been used for checking the performance of treatment options for WWTPs (Gernaey et al., 2000; Bixio et al., 2002; Huo et al., 2006) and their application to evaluate the performance of various process configurations, and operation strategies is even advocated in the ATV design guideline (ATV, 2000). The application of process-based models has enabled users to check the performance of innovative configurations (that might not be accommodated in design guidelines) and operational strategies under different dynamic flow and loading scenarios. In other words, the incorporation of mathematical models (as a result of added knowledge gained on treatment processes) in conjunction with some of the rules and design methods in design guidelines could result in the improved design and operation of WWTPs. However, mathematical models only represent an approximation of reality and the issue of uncertainty should be considered in the application of models, if the simulation results are to be used for making decisions on the optimum plant sizing and operation (Belia et al., 2009).

The issue of uncertainty in design guidelines is currently dealt with by making conservative assumptions that are reflected in the selection of the design inputs (e.g. applying safety factors to influent concentrations, designing to more stringent effluent standard (Ekama et al., 1984; ATV, 2000; Tchobanoglous et al., 2003). Even when models are to be used for checking the performance of a design alternative, users need to define several loading scenarios that are representative of the site-specific conditions and use proper values for model parameters. However, in reality a WWTP is usually subjected to various dynamic influent flow and loading conditions that are unique to each plant. Hence, it might be difficult to define a generic set of critical and yet likely dynamic loading scenarios and proper values for model parameters that would result in the optimal sizing of a plant.

Given the importance of considering uncertainty during WWTP design, a relatively novel approach that is based on the quantification of uncertainty in design and operation of WWTPs has emerged recently (Rousseau et al., 2001; McCormick et al., 2007; Benedetti et al., 2013). The gist of these new probabilistic design methods is the characterization of relevant sources of uncertainties in terms of probability distribution functions and the calculation of the probability of non-compliance (PONC) with the effluent standards. One of the main advantages of these methods over the traditional design methods is the calculation of PONC (a quantitative factor) which could be used as an index when deciding on the best design alternative. However, probabilistic design methods require the application of advanced statistical analysis on top of the usual dynamic simulations used for modeling the behavior of WWTPs. In previous studies on probabilistic design, the main emphasis had been placed on considering the uncertainty of model parameters and their effect on the PONC. The effect of influent uncertainty (an important source of variability) which heavily depends on the site-specific conditions (e.g. characteristics of the connected sewershed, climate conditions) has been considered but not to a great extent (Rousseau et al., 2001; Bixio et al., 2002; Benedetti, 2006; Belia et al., 2012). In addition, little effort has been dedicated to the communication of the concept of PONC which may sound abstract to many designers and decision makers.

It should be noted that the performance of a plant could be degraded or interrupted for several reasons. For example, washout of biomass during a wet weather event, equipment failure (e.g. failure in blowers or pumping systems) and, inadequate process control, are among some of the causes that could result in non-compliance with effluent standards. In addition, the decisions regarding operation of WWTPs (e.g. bypassing of the influent flow in wet weather conditions to maintain stable plant operation) could also affect the treatment performance and hence the PONC. The main sources of uncertainty, considered in this study are the variability in flow and basic compositions of influent wastewater as well as model parameter uncertainty, in particular for simulating the biological processes (i.e. removal of wastewater constituents by biomass) and the settling of sludge inside the secondary clarifier.

The current research work aims at filling some of the gaps that exist in the probabilistic design of WWTPs. These include the development of a set of statistical methods for proper uncertainty characterization and random generation of uncertain model parameters and variable inputs, the development of a model-based quantitative assessment of the PONC and the incorporation of the proposed uncertainty analysis and modeling tools into the design of WWTPs and their application to a real case study constitute the main objectives of this research work.

This introductory chapter is followed by five more chapters that cover the different aspects of the study.

- ✓ **Chapter 2** provides a critical review on the different design methods and highlights the advantages and disadvantages of each of them. The main shortcomings of the previous studies are identified and the objectives of this research to rectify them are explained.
- ✓ **Chapter 3** provides an overview of the proposed probabilistic design method. In this chapter the different sources of information, the required modeling tools and analysis, as well as the different steps of probabilistic design are outlined.
- ✓ **Chapter 4** is dedicated to the methods that have been developed for the stochastic generation of influent time series considering the basic characteristics of the sewershed as well as the climate conditions of the region.
- ✓ **Chapter 5** shows how the proposed methodology can be used for an actual case study (the Eindhoven WWTP in the Netherlands). The different steps of the probabilistic design are implemented using the information and data of the Eindhoven case study and at the end some details are provided regarding the communication of the calculated PONC.
- ✓ **Chapter 6** summarizes the main highlights of this research and also identifies some of the unresolved issues in the field of probabilistic design of WWTPs that could be tackled in future.

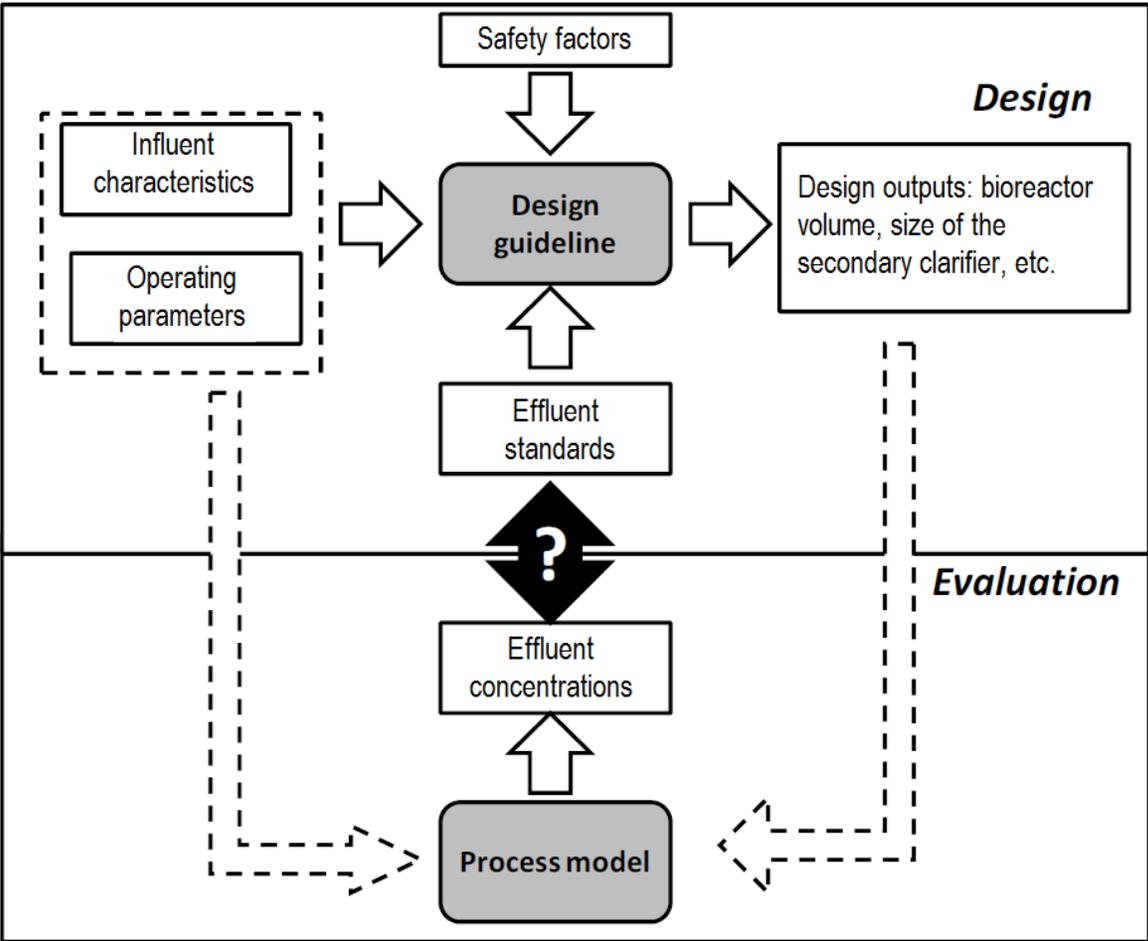
## **CHAPTER 2: LITERATURE REVIEW**

Wastewater treatment plants (WWTPs) are complex engineering systems whose failure to meet performance requirements can have detrimental effects on public health and the environment. The inputs and information required for the design of these water infrastructures are uncertain. Therefore engineers have to make their decisions on the design of a WWTP often under uncertainty. The literature review on the design of WWTPs shows that there are four main design methodologies, including: traditional design, model-based design with optimization, optimal robust design, and probability-based design.

### **2.1 Traditional design using design guidelines**

Years of experience in design and operation of WWTPs have resulted in the development of design guidelines that are being used nowadays by many design engineers worldwide. The recommended design parameters, mathematical equations, and also the way that design guidelines consider the effect of uncertain inputs in the design of a WWTP are among the main topics that are discussed briefly in this section. ATV-DVWK-A 131E (ATV, 2000), Metcalf and Eddy (Tchobanoglous et al., 2003), WRC (Ekama et al., 1984), and the Ten States Standards-recommended standards for wastewater facilities are the four important design guidelines that will be covered in detail. It should be noted that the application of some of the design guidelines that are discussed in this section may not be legally-binding (e.g. the Metcalf and Eddy guideline is rather a textbook with recommended design procedures that may be used for design. Conversely, the Ten States Standards is a design guideline whose application is obligatory in certain states/provinces in the North America). Therefore, the term “design guideline” used in this section refers rather to the design procedures that are typically used during a design project.

To avoid confusion regarding the inputs/outputs for design guidelines, the key terms that are used frequently throughout this section are illustrated in Figure 2-1. As indicated the inputs for design guidelines (i.e. top rectangular region in Figure 2-1) are classified into four categories (i.e. influent characteristics, operating parameters, safety factors, and effluent standards) and the outputs are the dimension of the different treatment units. Conversely, the inputs of a process-based model in addition to the influent characteristics and operating parameters include the dimensions of the plant (i.e. the outputs of a design guideline) and the outputs are effluent concentrations.



**Figure 2-1** Design inputs and outputs for design guidelines (top) and process models for

design evaluation using process models (bottom)(adapted from Corominas et al. (2010))

### **2.1.1 ATV-DVWK-A 131E design guideline**

One of the most commonly used design guidelines in Europe is the German “Standard ATV-DVWK-A 131E, Dimensioning of Single-Stage Activated Sludge Plants” (ATV, 2000). This guideline can be used for designing an activated sludge plant for the biological removal of COD, N, and P. German effluent standards require that the effluent limits be respected on 2-hour composite samples for at least four out of five samples with the fifth sample not surpassing the limit by more than 100%. To meet these stringent effluent standards, ATV-DVWK-A 131E uses conservative assumptions and safety factors (either to the influent design values or to the effluent standard concentrations to meet stringent effluent standards, e.g. setting  $\text{NH}_4$  concentration in the effluent from the bioreactors to zero) which could lead to an oversized plant (Bixio et al., 2002).

In ATV-DVWK-A 131E, after the determination of the required plant capacity and deciding on the type of treatment and configuration of the plant, the *design value*<sup>1</sup> of the solids retention time (SRT) must be set. The mathematical equation for the calculation of the design SRT comprises of three terms:

1) The safety factor SF which takes into account: a) variation of the maximum ammonium oxidizer growth rate caused by certain substances in the wastewater, short-term variations and/or pH shifts, b) the mean effluent concentration of the ammonium, c) the effect of

---

<sup>1</sup> In this report *design values* refer to the selected values for design inputs (influent, flow, effluent limits, stoichiometric and kinetic coefficients) to make a design under steady state conditions.

variations of the influent nitrogen load on the variations of the effluent ammonia concentration.

2) The value of 3.4 which is the product of the inverse maximum growth rate of ammonium oxidizers at 15°C (2.3d), and a factor of 1.6 to ensure sufficient oxygen transfer.

3) A temperature correction term for designing a system for design temperatures other than 15°C.

Depending on the size of the connected population contributing to the daily load, the SF should be in the range of 1.4 to 1.8 (depending on the size of the sewershed: lower safety factors for larger population equivalents). Although there are some recommendations for the selection of the safety factor value, selecting a proper value is a complicated task because the effect of several uncertain inputs has to be considered into only one safety factor.

For determining the size of the anoxic tanks for denitrification, the ratio of the anoxic volume to the total volume of the bioreactor is determined depending on the ratio of “nitrate to be denitrified” to the BOD concentration in the biological reactor. In determining the amount of nitrate to be denitrified a simple mass balance equation for nitrogen is written around the entire system. To be on the side of caution and estimate a conservative value for the amount of nitrate to be denitrified, the permissible concentration of ammonium in the effluent is set to zero and the selected value for the permissible concentration of nitrate in the effluent should be 60%-80% of the actually required value.

The recommended values for the ratio of the anoxic to the total volume of the bioreactor are tabulated in the ATV-DVWK-A 131E design guideline. Ratios of less than 0.2 or greater than 0.5 are not recommended.

The concentration of the *mixed liquor suspended solids* (MLSS) in the bioreactor is determined as a function of the anticipated settling characteristics of the sludge in the secondary clarifier and the concentration of sludge in the return sludge flow. The concentration of sludge in the return flow is assumed to be a certain percentage of the concentration of sludge at the bottom of the secondary settling tank which is estimated as a function of the anticipated *sludge volume index* (SVI) and *sludge thickening time* (Tth) through another empirical equation.

The total volume of the bioreactor is calculated by dividing the total mass (i.e. daily solids production  $\times$  SRT) by the concentration of MLSS in the bioreactor.

In ATV-DVWK-A 131E, the sizing of the secondary settling tank precedes the bioreactor sizing. However, in the end the concentration of sludge in the bioreactor which affects the sizing of both the bioreactor and secondary settling tank may change to find the optimum matching between the sizes of these two treatment units.

The area of the secondary settling tank can be estimated from the permissible *surface overflow rate* or the *sludge volume loading rate* and the design peak flow rate (i.e. the hourly peak flow rate, including the contribution of flow in wet weather conditions). Depending on the type of the secondary settling tank a permissible *sludge volume loading rate* ( $q_{SV}$ ) is selected (permissible values are provided in ATV-DVWK-A 131E); then the corresponding *surface overflow rate* ( $q_A$ ) is calculated by dividing the *sludge volume loading rate* by the product of the suspended solid concentration of the influent to the secondary settling ( $SS_{IST}$ ) tank and the SVI (i.e.  $q_A = \frac{q_{SV}}{SS_{IST} \times SVI}$  ).

If the calculated *surface overflow rate* is less than the permissible one (permissible values are provided in ATV-DVWK-A 131E) then the area of the secondary settling tank can be calculated by dividing the design peak flow rate by the *surface overflow rate*. The required *design values* should be selected in a way to cope with all possible flow and sludge settling characteristics (conservative values must be selected for the required *design values*).

### **2.1.2 Metcalf & Eddy design guideline**

The North-American Metcalf & Eddy guideline is another important guideline for the design of WWTPs. Design procedures for a wide range of treatment methods and configurations along with worked-out examples are provided in Metcalf & Eddy (Tchobanoglous et al., 2003). It uses detailed and usually process-based equations for the calculation of important design outputs.

The design SRT is estimated by multiplying the inverse specific growth rate of the nitrifying organisms with a nitrification safety factor. The specific growth rate is estimated using the process rate equations for the growth and decay of autotrophs in the ASM1 model. The effect of temperature on the maximum growth rate of the nitrifying organisms is considered explicitly (Henze et al., 1987).

The ratio of the expected peak to the average TKN loading estimated using the recorded influent data is selected for the nitrification safety factor. If no historical data are available, a value in the range of 1.3 to 2 will be selected for the nitrification safety factor. A typical range of values is also recommended for stoichiometric and kinetic coefficients that are necessary for the calculation of the specific growth rate. The typical values are the default values recommended in the ASM1 model (Henze et al., 1987). Because the reported nitrification kinetics covers a wide range of values, bench-scale or in-plant testing should be

performed to have a better estimate of site-specific nitrification kinetic values (Tchobanoglous et al., 2003).

For sizing the anoxic tank, the amount of nitrate that needs to be denitrified is determined by writing the mass balance equation for total nitrogen assuming that all of the influent TKN is nitrifiable and the effluent soluble organic nitrogen concentration is negligible (a conservative assumption). Moreover, it is assumed that the rate of denitrification depends on the concentration of *mixed liquor volatile suspended solids* (MLVSS) in the anoxic tank and another design parameter known as the *specific denitrification rate* (SDNR), which is a function of the *food to microorganisms* ratio (F/M), temperature, and the rate of internal recycle. The volume of the anoxic tank is estimated iteratively until an acceptable concentration of nitrate in the effluent is attained.

The equation for the estimation of daily solids production in a plant with nitrification accounts for heterotrophic biomass growth, cell debris from endogenous decay, nitrifying biomass, nonbiodegradable volatile suspended solids, and the inorganic solids in the influent. The equation for the estimation of daily solids production is derived by writing mass balance equations for the biomass of heterotrophic and nitrifying organisms around the entire system under steady state conditions. The process rate equations are similar to those in the ASM1 model. Typical values for the kinetics and stoichiometric coefficients are provided in the Metcalf & Eddy guideline.

As in the ATV-DVWK-A 131E, the total volume of the bioreactors is determined by dividing the total mass of solids (estimated by multiplying the daily solids production by SRT) by a selected concentration of MLSS in the bioreactor (recommended values are provided depending on the type treatment process. For example, for complete mixed activated sludge systems the recommend range for the MLSS in the bioreactors is 1500-4000 mg/l).

In the Metcalf & Eddy design guideline, the area of the secondary settling tank is calculated based on the *surface overflow rate* and the *solids loading rate*. When the *surface overflow rate* is used as the basis for dimensioning the area of the secondary settling tank, attention to the peak flow events and the use of a proper safety factor to ensure a *surface overflow rate* smaller than the *zone settling velocity* (ZSV) are important design considerations. Depending on the type of treatment, permissible design values for the *surface overflow rate* as well as *solids loading rate* can be found for both average and peak flow and solids loading conditions. The area of the secondary settling tank can be estimated by dividing the critical flow by the design *surface overflow rate* or the critical solids loading by the design *solids loading rate* (the larger area will be selected).

### **2.1.3 WRC design guideline**

The equations that are used in the South African WRC design guideline (Ekama et al., 1984) for dimensioning a WWTP are derived by using process rate equations and writing mass balance equations under steady state.

The design SRT is calculated by multiplying the inverse of the specific growth rate of the nitrifying microorganisms by a safety factor which accounts for that portion of sludge that might not be aerated. To calculate the specific growth rate of the nitrifiers, proper values for kinetic coefficients must be selected and adjustments should be made not only for the temperature but also for the effect of limiting alkalinity on the maximum specific growth rate. It is noted that if a value equal or greater than 1.25 is selected as the safety factor (for sufficient aeration of microorganisms) and the maximum specific growth rate of nitrifiers is adjusted for the temperature and alkalinity, then an effluent quality of 2-4 mgN/L during the lowest temperature and 1-3 mgN/L at 20 °C is achievable (Ekama et al., 1984).

The subdivision of the total reactor volumes into anoxic (base on the calculation of denitrification capacity) and aerated volumes can be calculated as a function of SRT and maximum specific growth rate of the nitrifiers. The recommended values for the unaerated to the total bioreactor volume are presented graphically and they should not be larger than 60%.

The total solids production is estimated by adding different types of solids entering or being produced in the system. These include the mass of ordinary heterotrophic organisms, endogenous residue, biodegradable organics, and also the inorganic suspended solids in the influent as well as the contribution of the ordinary heterotrophic organism to the inorganic suspended solids (about 15% of their volatile suspended solids or VSS (not originally in the WRC design guideline but developed later by Ekama and Wentzel (2004))). The related equations and their corresponding stoichiometric and kinetic coefficients required for the estimation of daily solids are the same as those provided by Marais and Ekama (1976).

For estimating the total volume of the bioreactors, the total mass of solids in the system should be divided by the design concentration of MLSS. The total mass of solids is estimated by multiplying the daily solids production with the selected SRT. The design concentration of TSS in the bioreactor can be set empirically from past experience or according to some cost analysis to minimize the cost of bioreactor and secondary settling tanks. In general the recommended range of values for the design concentration of TSS in the bioreactor with low strength ( $BOD_5$ , COD) wastewater and short SRT is from 2000 to 3000  $mgTSS/L$ . The TSS concentration in bioreactors with high strength and long SRT should be between 4000 and 6000  $mgTSS/L$ .

The design of the secondary settling tank is based on the flux theory according to which the surface overflow rate cannot exceed the critical recycle ratio (Henze et al., 2008). First, an empirical equation is used for calculating the settling velocity as a function of the sludge

concentration and the settling characteristics of the sludge (the adapted empirical equation is  $V_s = V_0 e^{-n.X}$  in which  $V_s$  is the settling velocity (m/d),  $X$  is the concentration of sludge ( $\text{kg/m}^3$ ) and  $V_0$  and  $n$  are constants describing settling characteristics of the sludge). Once, the settling velocity is calculated, the surface overflow rate (i.e. the flow divided by the area of the secondary clarifier) has to be equal or less than the settling velocity of the sludge at the entrance point to the settling tank (i.e.  $q_A \leq V_0 e^{-n.X}$ ). Considering the idealized settling tank assumption which will not occur in practice, designers are advised to apply some safety factors either to  $V_0$  or  $n$  or directly to the final area of the secondary clarifier that is calculated based on the flux theory.

#### ***2.1.4 Ten States Standards-recommended standards for wastewater facilities***

The 10 States Standards (Board, 2004) provides a set of design standards for rather conventional municipal wastewater collection and treatment systems for 10 states in the US (i.e. Illinois, Indiana, Iowa, Michigan, Minnesota, Missouri, New York, Ohio, Pennsylvania, Wisconsin) and Canada's Province of Ontario. The main objective of this guideline is to provide limiting values for different items corresponding to collection and treatment facilities so as to make the evaluation of plans and specifications possible for the reviewing authority. In addition, it establishes uniformity of practice among several states and provinces (Board, 2004).

According to this guideline the sizing of the bioreactors should be determined based on full scale experience, pilot plant studies, or process design calculations mainly based on SRT, food to microorganism ratio, volumetric organic loading rate and permissible MLSS inside the bioreactors. In general the sizing of the bioreactors should be in compliance with the permissible design parameters in Table 2-1. If using values significantly different than those

reported in Table 2-1, designers should provide reference to an actual plant and adequate data to support proper performance of their proposed alternatives.

**Table 2-1** Permissible aeration tank capacities and loadings in the 10 State Standards (Board, 2004)

Process	Organic loading (kg BOD <sub>5</sub> /m <sup>3</sup> .day)	F/M ratio (kg BOD <sub>5</sub> /kg MLVSS.day)	MLSS (mg/l)
Conventional Step aeration	0.64	0.2-0.5	1000-3000
Complete mix			
Contact stabilization	0.8	0.2-0.6	1000-3000
Extended aeration	0.24	0.05-0.1	3000-5000
Single stage nitrification			

The design of the secondary clarifier should be done in a way that both the thickening and solids separation requirements are met. To reach these objectives the sizing of the secondary clarifier must be based on the larger surface area calculated according to the surface overflow rate and solids loading rate. Table 2-2 indicates the permissible range of values for surface overflow rates and solids loading rates corresponding to different treatment processes that should not be exceeded for sizing of the secondary clarifier.

**Table 2-2** Permissible surface overflow rates and peak solids loading rates in 10 State Standards (Board, 2004)

Treatment process	Surface overflow rate at peak hourly flow rate (m <sup>3</sup> /m <sup>2</sup> .day)	Peak solids loading rate (kg/m <sup>2</sup> .day)
Conventional step aeration	49	244
Complete mix		
Contact Stabilization		
Carbonaceous stage of separate stage nitrification		
Extended aeration	41	171
Single stage nitrification	33	171
Two stage nitrification		
Activated sludge with chemical addition to mixed liquor for phosphorus removal	37	171

As mentioned earlier, the design recommendations in the 10 States Standards are mainly based on the years of experience of design and operation of full scale and pilot plans (there are no design equations) which enable the reviewing authorities to do their evaluation of plants. Therefore, the 10 states standards could be thought of as a design verification document, meaning that a proposed design alternative which is made by a design tool (weather it is based on a specific design guideline or a model) should be in compliance with design recommendations in the 10 States design standards.

### 2.1.5 *Summary of design guidelines*

Mathematical equations that are used in design guidelines are usually derived by assuming steady state conditions and writing mass balance equations around the system with simplified process equations. The estimation of the parameters that will affect the final sizing of the plant such as the SRT or daily solids production are not necessarily the same in all design guidelines. For example, in ATV-DVWK-A 131E an empirical equation is used for the estimation of the daily solids production whereas Metcalf & Eddy uses a rather different equation that takes into account the growth and decay of heterotopic and nitrifying organisms, nonbiodegradable organic matter in the influent, cell debris, as well as the inorganic matter in the influent. Therefore, the calculation of important design variables (e.g. total volume of bioreactor) may not be the same in different design guidelines.

In these guidelines uncertainty is taken into account by applying safety factors in by multiplying some of the design variables (e.g. volume of bioreactors) by a coefficient or making conservative assumptions regarding the *design values* for flow, influent, kinetic/stoichiometric coefficients and/or effluent limits (e.g. overestimation of the amount of nitrate which needs to be denitrified in ATV-DVWK-A 131E and Metcalf & Eddy). One of the problems with applying safety factors to different elements of design is the double accounting of a specific source of uncertainty. For example, in the ATV-DVWK-A 131E design guideline the SRT is multiplied by a safety factor to ensure that it is large enough for proper nitrification so that the effluent ammonia concentration meets the required effluent standard. However, on top of the safety factor mentioned above, another safety factor is applied to the required ammonia concentration in the effluent (i.e. a safety factor less than 1) to guard against negative factors that could cause non-compliance with the effluent standards. This could lead to a conservative and possibly expensive design without necessarily providing a worthwhile benefit (Doby, 2004).

Although there are some recommended ranges of values for the selection of safety factors, kinetic/stoichiometric coefficients, and settling parameters, the design engineers make the final selection of these values. Thus, the final design according to a specific guideline might be different since the engineers must use their subjective judgement in selecting proper values for safety factors and other design parameters.

Design guidelines assume steady state conditions and use a combination of empirical and simplified process equations. To cope with the different sources of uncertainty and variability various safety factors are applied and conservative assumptions are made during the design process which would lead to a WWTP that usually meets effluent standards but, usually, with a high cost (as is the case with ATV-DVWK-A 131E design (Benedetti et al., 2006b)). However, with such an approach the dynamics of the effluent cannot be predicted. Moreover, it is not possible to estimate the PONC for a specific design with respect to a certain effluent standard.

## **2.2 Model-based design and optimization**

Recent advances in computational methods and increased understanding of various treatment processes have led to the development of mathematical process-based models. These models have become attractive tools for engineers during the design process and to check different operation and control strategies (Hao et al., 2001; Salem et al., 2002; Copp, 2002; Flores-Alsina et al., 2014b). Daigger and Grady (1995) classified the biological wastewater treatment models into three categories: 1) stoichiometric, 2) refined stoichiometric, and 3) kinetic/stoichiometric.

Stoichiometric models use experience-based stoichiometric factors for overall biochemical conversions (e.g. kg oxygen required/kg COD removed). These models are very simple but

require extensive experience for determining proper stoichiometric coefficients. Also, they cannot simulate process dynamics.

In refined stoichiometric models (in this thesis they are simply referred as steady state models), the model constituents are structured into more fundamental components (e.g. distinction between particulate and filterable constituents, or distinction between biodegradable and nonbiodegradable organic matters) and both the stoichiometric and kinetics (under steady state) of the processes are considered. In these models, it is assumed that many bioprocesses have reached their completion (e.g. utilization of organic substrate). Hence, the dynamic bioprocesses are reduced to a set of stoichiometric equations only. Moreover, other processes that do not reach their completion are expressed in a simplified way (e.g. death regeneration of heterotrophic micro-organisms due to endogenous respiration) and the analysis is focused on the slowest processes that govern the sizing of the plant (e.g. nitrification for the activated sludge systems). In fact, some of the process-based equations that are used in design guidelines (explained in Section 2.1) have been derived from the same assumptions that are made for the development of steady state models. However, design guidelines provide users with more information regarding the determination of design inputs that could be the same as the inputs to a steady state models (e.g. influent characteristics).

The fact that they require much less information compared to kinetic/stoichiometric models (in this thesis they are simply referred as dynamic models) as well as the fact that they express relationships between design parameters (e.g. sludge age) and plant performance (e.g. predicted effluent concentration) in terms of a set of algebraic equations, has made steady state models an attractive tool for initial design of WWTPs (Ekama, 2009).

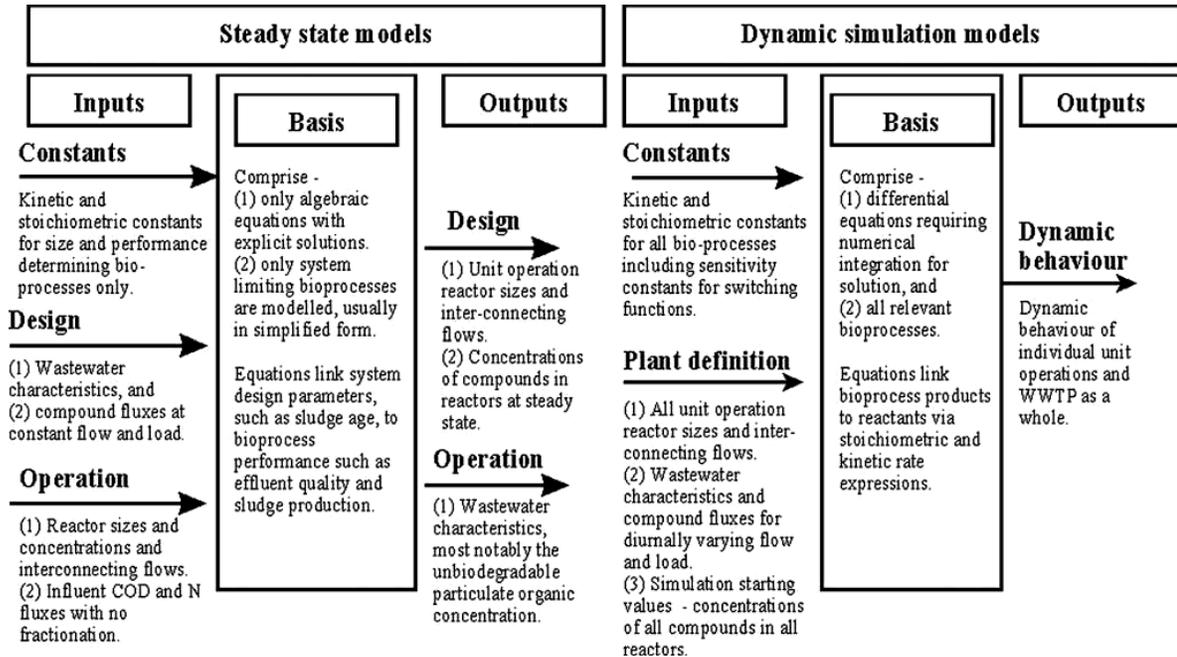
Development and successful application of a steady state model to several design and upgrade projects can be found in the work of Daigger et al. (1998). They developed a

steady state model called PRO2D for simulating the entire WWTP including the liquid/solid separation processes, granular media filters, solid thickening and dewatering units, biological processes such as activated sludge and anaerobic digesters. The ability of the PRO2D model to be coupled with dynamic models for making further adjustment to the initial design of the plant was noted as the main advantage of this steady state model. Wentzel et al. (1990) developed a steady model capable of modeling the release, uptake and removal of phosphorus in modified Bardenpho and UCT (University of Cape Town) system configurations under steady flow and loading conditions.

The application of steady state models for the design of WWTPs is advocated by Ekama (2009). In his study, a plant-wide mass balance-based steady state model was developed using bioprocess stoichiometry, capable of tracking the different products exiting the WWTP in the form of solid, liquid, and gas. He argued that the application of mass balance-based steady state models should precede the dynamic models as the former could conveniently provide the required inputs for the latter (Figure 2-2).

In dynamic models, the model constituents are also structured into fundamental components (similar to steady state models but with more details) but in contrast to the steady state models they are fully dynamic and thus capable of predicting the dynamics of the process. The well-known activated sludge models: ASMs (Henze et al., 2000) fall into the category of kinetic/stoichiometric models. These models provide a more comprehensive description of reality and tend to produce more accurate results. They predict process dynamics but are much more complex compared to the other previously mentioned models (Daigger et al., 1998). Dynamic simulation has been used to check the dynamic behavior of statically dimensioned WWTPs (Rousseau et al., 2001; Bixio et al., 2002; Huo et al., 2006; Corominas et al., 2010). The application of dynamic models to evaluate the performance of various process configurations, measurement and control technologies is even advocated in ATV (2000). However, it should be noted that the application of dynamic models in design,

should be preceded by a steady state model (Ekama, 2009) or a design guideline (Corominas et al., 2010).



**Figure 2-2** Inputs, basis, and outputs of steady state (left) and dynamic models (right) (Ekama, 2009)

In general, steady state models can be used in a trial-and-error fashion in which the dimensions of different treatment units are calculated given a specific set of influent characteristics and operating parameters, and effluent concentrations or vice versa. In order to search for an optimum combination of operating conditions and dimensions of a WWTP, some researchers have used steady state models in conjunction with formal optimization techniques to find optimum operating conditions and sizing of WWTPs (Doby et al., 2002).

Doby et al. (2002) used a steady state model and a genetic algorithm (GA) to identify low cost activated sludge designs that meet effluent standards. Model parameters, influent, and effluent *design values* were selected as inputs and a cost function was developed using

engineering judgment. Total cost was a function of primary clarifier, effluent penalties, and recycle costs and to ensure the treatment of wastewater with a high reliability, a large penalty relative to other costs was assigned to any sorts of noncompliance with the required effluent standards (see Table 2-3 for examples of cost functions). Rather than assigning a large cost to any noncompliance with the effluent standards, constraints can be imposed directly on the effluent values (Rivas et al. 2008).

The application of non-linear optimization techniques like GRG2 to the design of WWTPs using the ASM1 model (Henze, 2000) under steady state conditions has also been studied by Rivas et al. (2001). Compared to the non-linear optimization techniques, the GA optimization methods have the advantage of convergence to the global minimum and flexibility in defining the objective function. However, intensive computation, and the need for considerable knowledge on the part of the user about the efficient parameterization of the GA search (e.g. population size, initial generation of population, mutation rate, and etc), have been stated as the main drawbacks of GA approaches (Doby, 2004).

The number of design variables (i.e. those variables for which the optimum values are searched for) and also the cost function which is to be minimized are not the same in different studies. For example, Doby (2004) estimated the cost as a function of bioreactor volume, primary clarifier volume, effluent penalties, and recycle costs, whereas in the study of Rivas et al. (2008) the optimization was limited to the bioreactor volume alone. Limiting the optimization only to the performance of the bioreactor (e.g. the size of anoxic and aerated tanks, flow fractionations entering different tanks, SRT, and HRT) may not result in the optimum dimensioning of the entire WWTP. Because in a WWTP different units interact with each other, optimization should be performed for the entire plant and not be limited to a specific unit. For example, ATV-DVWK-A 131E (2000) states that optimization between the sludge volume loading rate to the secondary clarifier and the depth of the secondary clarifier could be an option, but in the optimization-based

methodology proposed by Rivas et al. (2008) the concentration of sludge in the influent to the secondary clarifier was fixed to a certain value.

Overall, in the model-based optimization design procedure, there is no clear-cut rule regarding the selection of design loads and model parameters. This is also tied to the choice of safety factors. The safety factors may be applied to the input, output, or some intermediate calculation within the model. However, such an approach may lead to a too conservative and possibly expensive design without necessarily providing a worthwhile benefit (Uber et al., 1991b). Furthermore, one should be very cautious in the application of models to those processes that are not represented by model equations. For example, Uber et al. (1991b) compared the results of optimized values of HRT presented by different researchers and showed that the type of model used in simulation and the imposed constraints on design variables may lead to different optimized sizings for the WWTP. Finally, the issue of uncertainty in plant performance stemming from different sources of uncertainty is not considered explicitly in this type of WWTP design (Uber et al., 1991b).

### **2.3 Robust optimal design**

One of the most important problems with the use of models for the design of WWTPs (whether they are used in a trial-and-error fashion or in conjunction with an optimization framework) is the uncertainty in model inputs and parameters and influent values. To tackle this problem Uber et al. (1991a) proposed a general robust optimal design procedure which is an extension of the traditional total cost optimization approach. They define robustness as the ability of the system to maintain a level of performance even if the actual parameter values are different from the assumed ones. Thus, a design is more robust if its performance is less sensitive to the variation of its inputs and model parameters.

In their approach, first-order sensitivity indices (the partial derivatives of different effluent concentration variables with respect to uncertain model inputs and parameters) are defined as robustness measures and these indices are incorporated either as an objective function or as constraints into an optimization framework (i.e. cost is considered as an objective function and robustness measure as a constraint, or the robustness measure is considered as an objective function and the cost as a constraint). The final optimized design variables are dependent on the formulation and the type of model by which the different processes are simulated. In other words, the issue of model-structure uncertainty (see Section 2.4 for details on different types of uncertainty) is not considered and it is assumed that the selected model equations are a good-enough representation of reality.

The robustness measures of Uber et al. (1991a) are estimated by taking partial derivatives of a specific effluent concentration with respect to all uncertain model inputs and parameters at their nominal value irrespective of the possible correlation that may exist among them. Hence, their possible interdependence is not considered. Although the model inputs and parameters are assumed to be uncertain, the robustness measures themselves are not subject to uncertainty as the robustness is calculated at the nominal values of the uncertain inputs and parameters (Uber et al., 1991a). Meanwhile, it should be emphasized that the notion of system robustness presented by Uber et al. (1991a) does not include the probabilistic concept of performance failure or the PONC (see 2.4 for details on PONC) with effluent standards as described below:

Overall, the main shortcomings of their proposed robust optimal methodology can be summarized as follows:

- 1) Their robustness measure is calculated at the nominal values of model inputs and parameters only.

- 2) It is assumed that all parameters are independent and any type of interdependence between them is ignored.
- 3) It does not provide the designers with the PONC to the effluent standards.

Uber et al. (1991b) applied their proposed methodology to the design of an activated sludge system. For their analysis, they used a set of modeling tools proposed by Tang et al. (1987) under steady state conditions with 55 uncertain inputs and parameters (including influent constituents). To generate a trade-off curve between total cost and the level of robustness, several upper constraints were imposed on total cost while the robustness index calculated as the weighted sum of the first order sensitivity indices of BOD and TSS concentration in the effluent to the uncertain inputs and parameters was minimized. The optimized design variables were compared to those obtained by a cost only optimization scheme and also to the recommended values by design guidelines (i.e. Metcalf and Eddy and Ten State Standards (Board, 2004)). The comparison indicated that the optimal robust methodology produces values closer to those provided by the usual design practice. The authors argued that since the recommended design practices and robust optimal design are not solely based on the minimization of cost, the final sizing of a WWTP using the robust optimal framework tends to be more consistent with the recommended design practices.

## **2.4 Probability-based design**

Since a WWTP is subjected to stochastic loads, the quality of treated wastewater would not be the same during different periods of the year. Therefore, estimating the PONC with the effluent standards for a set of design alternatives is essential in helping decision makers to make the right decision on the best configuration and sizing of a WWTP. Considering the PONC during the design of complex engineering systems leads us to a different design paradigm known as *probability-based* design. Potential failure of an engineering system to meet its required objectives is the result of the combined effects of inherent randomness of external loads and various uncertainties involved in the analysis, design, construction, and

operational procedures (Tung et al., 2006). Hence, to evaluate the probability of system failure to meet its expected performance requires uncertainty characterization and uncertainty analyses.

#### **2.4.1 Definition of the PONC based on the load-resistance concept**

Estimating the PONC requires the definition of noncompliance in terms of a mathematical expression and the probabilistic treatment of the problem. Tung et al. (2006) introduced the *load-resistance* interface for the description of failure to meet certain requirements (in this study failure to meet effluent standards) in terms of mathematical functions. In this approach, failure to meet certain requirements of an engineering system occurs when the *load* (external forces or demands) on the system exceeds its *resistance* (strength, capacity, or supply) of the system. For example, in the design of the urban drainage systems the *load* could be the inflow to the sewer system, whereas the *resistance* is the sewer conveyance capacity. In the economic analysis of a water system, the load could be the total cost, whereas the resistance is the total benefit.

The difference between the *resistance* and *load* yields the *performance function*. Equation 2-1 shows the *performance function* in mathematical terms:

$$W(X) = R(X_R) - L(X_L) \quad \text{Equation 2-1}$$

where  $W(X)$  is the performance function,  $R(X_R)$  the resistance of the system,  $L(X_L)$  the load,  $X_R$  and  $X_L$  are vectors of uncertain variables for estimating the *resistance* and the *load* respectively.

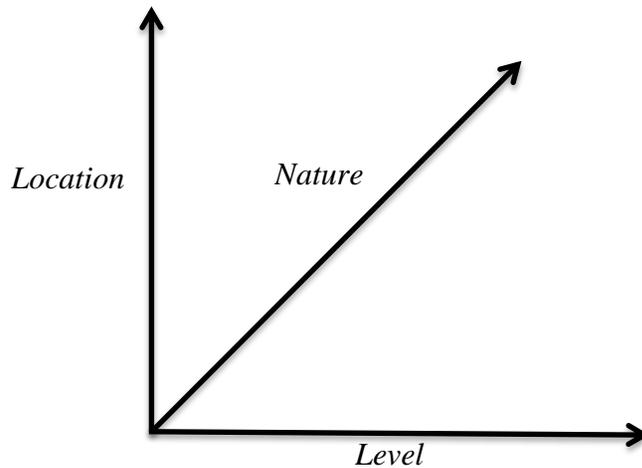
In the above equation *resistance* and *load* could be both functions of uncertain variables characterized in terms of probability distributions. Once the probability distribution of  $R(X_R)$  and  $L(X_L)$  are established as functions of uncertain vectors of variables (i.e.  $X_R$  and  $X_L$ ), the probability distribution of  $W(X)$  can also be estimated, and the probability of the corresponding vectors of variables for which the *performance function* is negative equals the system's probability of failure to meet certain requirements. For example, in water quality assessment the influent mass of pollutants to a WWTP system and the permissible pollutant concentration set by water quality regulations constitute the load and the resistance respectively.

#### **2.4.2 Types and characterization of uncertainty**

As mentioned earlier, calculation of the PONC requires the characterization of uncertain model inputs in terms of probability distributions. Uncertainty is attributed to the lack of perfect information concerning the phenomena, processes, and data involved in problem definition and resolution (Tung et al., 2006). Uncertainty could simply be defined as the degree of lack of knowledge about a system or degree of inability to perfectly describe its behavior or its existing state (Belia et al., 2009). In practice most of the decisions regarding the design of an engineering system are made under uncertainty. Many topologies of uncertainty have been developed in the literature but perhaps the most comprehensive one in the context of model-based decision can be found in the work of Walker et al. (2003). They defined three dimensions for the concept of uncertainty including *nature*, *location*, and *level* (Figure 2-3). The *nature* of uncertainty is classified into two main categories of *epistemic uncertainty* and *variability uncertainty* which are due to knowledge deficiency and uncertainty due to the inherent variability of phenomena respectively. These types of

uncertainties are referred to as *subjective* and *objective*<sup>1</sup> uncertainties respectively by Yen and Ang (1971).

The *location* of uncertainty refers to the different places where uncertainty manifests itself within the model complex. It includes: *context uncertainty* (i.e. uncertainty in the identification of the system boundaries), *model uncertainty* (i.e. both model structure uncertainty and model technical uncertainty arising from computer implementation of the model), *input uncertainty* (i.e. both inputs describing the reference system and external forces driving changes in the current system), *parameter uncertainty* (i.e. the uncertainties associated to the data and the different techniques used for model calibration), and *model output uncertainty* (i.e. the accumulated uncertainty caused by other types of model uncertainties) (Walker et al., 2003).



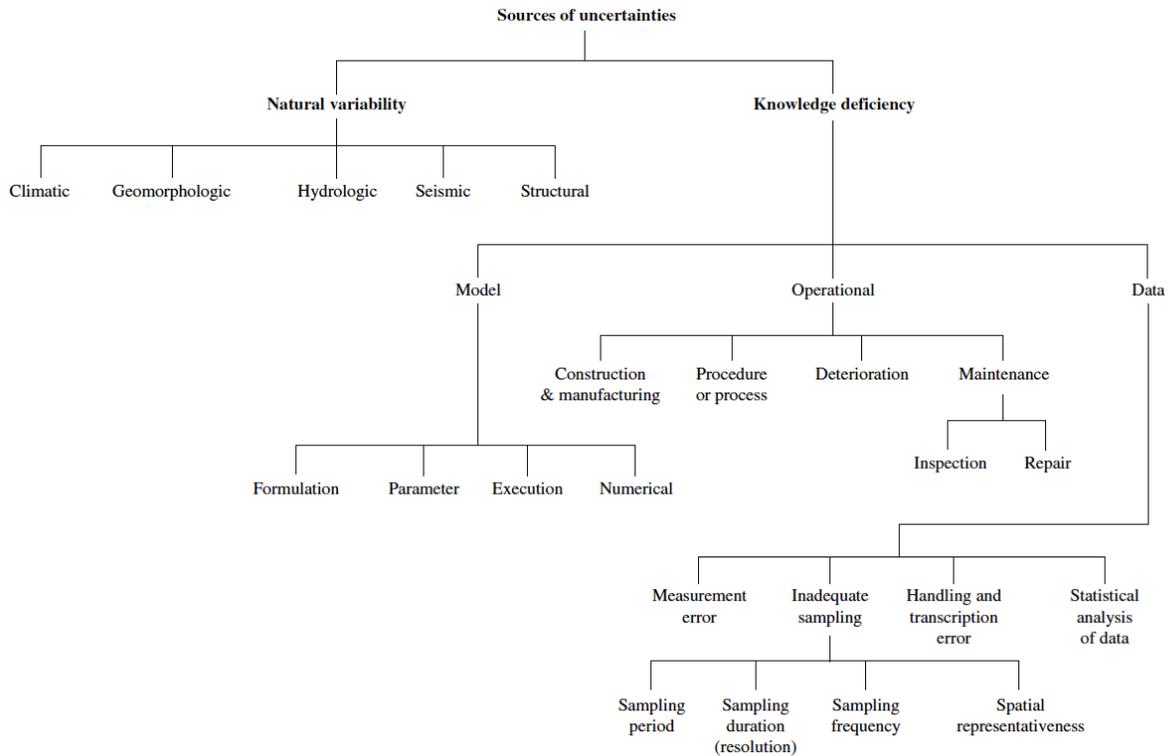
**Figure 2-3** Uncertainty: a three-dimensional concept (Walker et al., 2003)

---

<sup>1</sup> To avoid confusion, throughout this document we use *variability* for those types of inherent uncertainty and *uncertainty* for those that are related to knowledge deficiency.

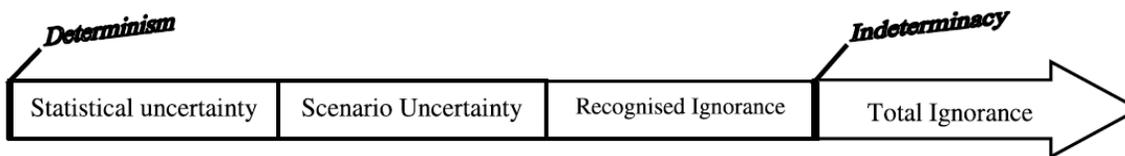
Tung and Yen (2005) also presented a more detailed and somewhat different classification of uncertainty than the one developed by Walker et al. (2003). According to their classification uncertainty is divided into two main categories of uncertainty due to natural variability and knowledge deficiency which is comparable to the nature dimension of uncertainty proposed by Walker et al. (2003). However, the identification of the location of uncertainty is different in the two classifications of uncertainty and the seemingly similar concept in these two classifications should not be confused with each other. For example the concept of model uncertainty by Walker et al. (2003) includes only the model structure uncertainty (corresponds to formulation in Figure 2-4), which stems from the form of the model itself, and model technical uncertainty (corresponds to execution and numerical in Figure 2-4), which arises from the computer implementation of the model. In the framework of Tung and Yen (2005) depicted in Figure 2-4, however, model uncertainty also includes parameter uncertainty. Meanwhile, in the framework of Tung and Yen (2005) nothing is mentioned regarding the dimensions of uncertainty as described by Walker et al. (2003).

According to Walker et al. (2003), the level of uncertainty manifests itself along a spectrum which begins with deterministic knowledge and ends with total ignorance. As indicated in Figure 2-5, the first level of uncertainty is the statistical uncertainty which refers to those uncertainties that can be described in statistical terms. Implicit in the characterization of a specific uncertainty in statistical terms is that the functional relationships of the model are good-enough descriptions of the phenomena being simulated, and the data used for the calibration of the model are representative of conditions to which the model is applied.



**Figure 2-4** Sources of uncertainty (Tung and Yen, 2005)

In contrast to the *statistical uncertainty* where the model relationships can reasonably represent reality, in *scenario uncertainty*, it is acknowledged that there is a range of possible outcomes but the mechanisms leading to these outcomes are not well-understood, hence they cannot be explained probabilistically (Walker et al., 2003). *Recognized uncertainty* and *total ignorance* are other levels of uncertainty in which the required basis for the development of scenarios is very weak.



**Figure 2-5** Different uncertainty levels (Walker et al., 2003)

To better highlight the role of uncertainty analysis identified at different *locations* in the various stages of the modeling process, Refsgaard et al. (2007) used almost the same classification of uncertainty developed by Walker et al. (2003) but further developed a framework for the modeling process and its interaction with the broader management process and the role of uncertainty at different stages in the modeling processes. Belia et al. (2009) developed a framework for identifying different sources of uncertainty introduced at the different stages of model-based design or optimization of WWTP projects. Uncertainty analysis of a system requires the use of probability and statistics. Depending on the type of uncertainty/variability and the availability of historical data, the various sources of uncertainty/variability can be expressed in an explicit manner.

Some researchers have characterized the input variability and parameter uncertainty in conjunction with a steady state model for estimating the PONC. Sharma et al. (1993) characterized the variability of flow, BOD and suspended solids in the influent in terms of lognormal probability distribution functions (PDFs). They used the recorded historical data for finding the parameters of the PDFs, and assumed independence among these three random variables. No uncertainties were assumed for model parameters. To capture the correlation between the values of flow and temperature in time, Doby (2004) characterized the variability of flow and temperature by repeating the available historical data (i.e. 5 years) over the lifetime of the project (i.e. 20 years) and then used regression equations to characterize the variability of other influent constituents like COD, ammonia, TKN, and total phosphorous (TP) as functions of flow and through regression equations. Model parameter uncertainties were characterized by truncated normal distributions whose parameters were determined using the values reported in the literature. Overall the uncertainty in model parameters is characterized using either a distribution with a finite support (e.g. uniform or triangular distributions) or a truncated distribution with infinite support (e.g. normal distribution) (Rousseau et al., 2001; Bixio et al., 2002; Benedetti et al., 2008). However, in cases where little information is available on parameters, it is

recommended to use uniform distributions to insure that the model prediction uncertainties are not underestimated (Benedetti et al., 2008).

Characterization of input and parameter uncertainty/variability for estimating the effluent variation using dynamic simulations has been reported in the literature. Rousseau et al. (2001) characterized the variability of flow by randomly generating a 6-month flow time series from a specified PDF at each Monte Carlo trial. To generate the time series of the influent constituents like COD, TKN, and nitrate, a set of regression equations was developed using extensive data sets of historical influent and flow measurements. By using these equations, the influent constituents can be related to flow, so the correlation of these constituents with flow is respected during the generation of the influent. The uncertainty in model parameters was described by triangular or truncated normal distributions to ensure generation of non-negative values. The mean and variance of these uncertain model parameters were also determined using reported values in the literature. Meanwhile, the variation in some parameters due to temperature variation was explicitly modeled according to an Arrhenius equation using the values provided in the ASM1 model (Henze et al., 1987).

### ***2.4.3 Analytical methods for estimating the PONC***

In contrast to the Monte Carlo methods (Section 2.4.4) which are numerical, in the analytical methods the uncertain model parameters and inputs are expressed by PDFs and the *performance function* (Equation 2-1) is approximated by writing the first order Taylor series expansion about a specific point (linear estimation of the *performance function*). The two common analytical methods that have been used for estimating the PONC of engineering systems and their application are covered in this section.

## I. MFOSM

Methods like the Mean-value First-Order Second-Moment (MFOSM) method were quite popular especially before the advent of computers capable of dealing with large computational loads. In this method the critical condition of the system where the *performance function* in Equation 2-1 equals zero is defined as a function of uncertain variables (the boundary at which the *performance function* is zero and separates the safe region from the failure region is called the *failure surface* (Figure 2-6)). Then, the linear estimate of the *performance function* is calculated by Taylor series expansion about the mean value of the uncertain parameters. After defining the uncertainty of parameters in terms of an analytical PDF (e.g. Normal, Lognormal, etc) the PONC of the system can be estimated analytically.

To clarify this method, consider the *performance function*  $W(\mathbf{X})$  that is a function of  $K$  stochastic variables or  $W(\mathbf{X}) = W(X_1, X_2, \dots, X_K)^T$ , in which  $\mathbf{X} = (X_1, X_2, \dots, X_K)^T$  is a  $K$ -dimensional column vector of uncertain variables. The first-order approximation of  $W(\mathbf{X})$  using Taylor series expansion with respect to a selected point of uncertain basic variables  $\mathbf{X} = \mathbf{x}_0$  in the parameter space can be approximated as ( $\approx$  sign indicates approximation):

$$W(\mathbf{X}) \approx \omega_0 + \sum_{k=1}^K \left[ \frac{\partial W(\mathbf{X})}{\partial X_k} \right]_{\mathbf{x}_0} (X_k - x_{k0}) \quad \text{Equation 2-2}$$

where  $\omega_0$  equals the value of the *performance function* at  $\mathbf{x}_0$  or  $W(\mathbf{x}_0)$ . The mean and variance of  $W$  by the first-order approximation can be expressed respectively as:

$$\mu_{\omega} = E[W(\mathbf{X})] \approx \omega_0 + \sum_{k=1}^K \left[ \frac{\partial W(\mathbf{X})}{\partial X_k} \right]_{x_0} (\mu_k - x_{k0}) \quad \text{Equation 2-3}$$

$$\sigma_{\omega}^2 = Var[W(\mathbf{X})] \approx \sum_{j=1}^K \sum_{k=1}^K \left[ \frac{\partial W(\mathbf{X})}{\partial X_j} \right]_{x_0} \left[ \frac{\partial W(\mathbf{X})}{\partial X_k} \right]_{x_0} Cov(X_j, X_k) \quad \text{Equation 2-4}$$

In matrix form Equation 2-3 and Equation 2-4 can be expressed as below (Tung et al., 2006).

$$\mu_{\omega} \approx \omega_0 + \mathbf{s}_0^T (\boldsymbol{\mu}_x - \mathbf{x}_0) \quad \text{Equation 2-5}$$

$$\sigma_{\omega}^2 \approx \mathbf{s}_0^T \mathbf{C}_x \mathbf{s}_0 \quad \text{Equation 2-6}$$

where  $\mathbf{s}_0 = \nabla_x \mathbf{W}(\mathbf{x}_0)$  is the column vector of sensitivity coefficients with each element representing  $\partial W / \partial X_k$  calculated at  $\mathbf{X} = \mathbf{x}_0$ . As mentioned in the MFOSM method the expansion point is selected as the mean value of uncertain variables or  $\mathbf{x}_0 = \boldsymbol{\mu}_x$  (Figure 2-6). Substituting  $\mathbf{x}_0$  with  $\boldsymbol{\mu}_x$  the mean and the variance of the *performance function* can be derived according Equation 2-3 and Equation 2-4 respectively. Assuming an appropriate PDF for the random *performance function*  $W(\mathbf{X})$ , the probability of failure can be computed by estimating the corresponding probability at which  $W(\mathbf{X})$  is less than zero (i.e. the load function exceeds the resistance function in Equation 2-1).

This method has been extensively used for the design of different water systems like the design of levees (Lee, 1986), open-channels (Huang, 1986), and culverts (Yen et al., 1980).

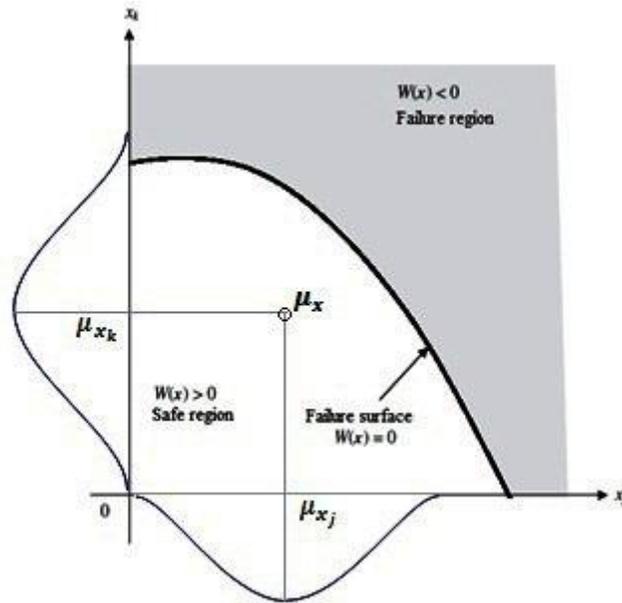
The study of Tung et al. (2006) on probabilistic design of sewer pipes can serve as a straightforward example to better clarify the terms of the performance function in Equation 2-1. In their study the maximum peak flow that enters the sewer pipe represents the load to the system (Equation 2-7) and the sewer flow-carrying capacity represents the resistance of the system (Equation 2-8).

$$L(X_L) = \lambda_L \times C \times A \quad \text{Equation 2-7}$$

$$R(X_R) = \frac{0.463}{n} \times \lambda_c \times D^{8/3} \times S^{1/2} \quad \text{Equation 2-8}$$

As indicated above, the load to the system is formulated using a simple rainfall-runoff model where  $L(X_L)$  is the surface runoff,  $\lambda_L$  is a correction factor,  $C$  is rainfall intensity, and  $A$  is the area contributing to runoff. The resistance function is formulated using Manning's equation where  $R(X_R)$  is the sewer flow-carrying capacity,  $n$  is Manning's roughness coefficient,  $\lambda_c$  is a correction factor,  $D$  is the pipe diameter, and  $S$  is the pipe slope.

Although the application of this method does not require large computational effort, it has some serious shortcomings. For example, Tung et al. (2006) argue that when estimating the PONC, the concern is often those points that put the system at critical conditions. As a result, the expansion point for Taylor series should be located on the *failure surface* where the system is at the threshold of failure (Figure 2-6).



**Figure 2-6** Failure surface for a performance function with two uncertain parameters

The MFOSM method uses the first-order representation of the original *performance function* to provide a linear estimation of the original function. Therefore in case the *performance function* which defines the critical state of the system is highly nonlinear, the linear approximation will not be accurate enough (Tung et al., 2006). It was also noted that the MFOSM method should not be used for cases in which a highly accurate estimation of failure probability is required.

## **II. AFOSM**

The Advanced First-Order Second Moment or AFOSM method is an improvement to the MFOSM method but it still keeps the simplicity of the first-order approximation. The difference is that the expansion point for the first-order Taylor series is located on the *failure surface* where the *performance function*  $W(\mathbf{X})=0$  (Figure 2-6). Therefore the main difference between the MFOSM method and the AFOSM is that in the MFOSM method the linearization of the performance function is done around the mean value of

uncertain variables (which may not be necessarily located at the failure surface ( $\mu_x$  in Figure 2-6)). Conversely, in the AFOSM method the linearization is performed around a point located at the failure surface.

The AFOSM method has been applied to various water systems engineering problems, including storm sewers (Melching and Yen, 1986), groundwater pollutant transport (Sitar et al., 1987), and water quality modeling (Han et al., 2001; Maier et al., 2001).

In the area of WWTP design, Sharma et al. (1993) used the AFOSM method to calculate the PONC of BOD and TSS for a hypothetical WWTP. They used simple equations to model the treatment processes in the WWTP and assumed only the influent BOD, TSS, and flow as uncertain variables. These random variables were assumed to be independent and their variability was characterized by three lognormal distributions whose parameters were estimated using historical data. The two performance functions for the effluent TSS and BOD concentrations were formulated using two empirical functions as below:

$$W(X_{TSS}) = 10 - C_0 \left( 1 - \frac{6.79}{Q} + \frac{10.28}{Q^2} \right) \quad \text{Equation 2-9}$$

$$W(X_{BOD}) = 15 - B_0 + 0.57 \frac{C_0}{Q} + \frac{673.8}{Q} \quad \text{Equation 2-10}$$

where,  $W(X_{TSS})$  is the performance function for TSS concentration,  $C_0$  the influent TSS concentration,  $Q$  the influent flow, and  $B_0$  the influent BOD concentration.

They also compared the PONC estimated by the AFOSM method to that estimated by Monte Carlo simulation using 10000 simulations. They concluded that the AFOSM method can produce approximately the same results with much less computational effort. Furthermore, the dependency of design variables such as aeration tank volume, sizes of the

primary settling tank and secondary clarifier to the probability of compliance can easily be established. This can help engineers design a system with a desired probability of compliance. Despite the small computational effort in comparison with Monte Carlo simulation, this application of the analytical methodology suffers from several shortcomings. First, they used only a set of very simple equations under steady state condition to model the treatment processes taking place in a WWTP. Second, the characterization of uncertainty was only limited to the variability of influent BOD, TSS, and flow with no provision for characterizing other sources of uncertainty (e.g. model structure uncertainty). Third, it was assumed that the uncertain variables (BOD, suspended solid, and flow) are independent from each other, whereas in reality that is not expected to be the case.

Overall, the AFOSM method has some general shortcomings. For example, it uses the same linearization technique (i.e. Taylor expansion) as the MFOSM method. Therefore, its application to highly non-linear systems may not provide accurate results. Difficulties in finding the design point due to the presence of multiple local minima, and slow convergence for systems with a very small PONC are shortcomings of this method (Yen et al., 1986).

#### ***2.4.4 Monte Carlo Method for estimating the PONC***

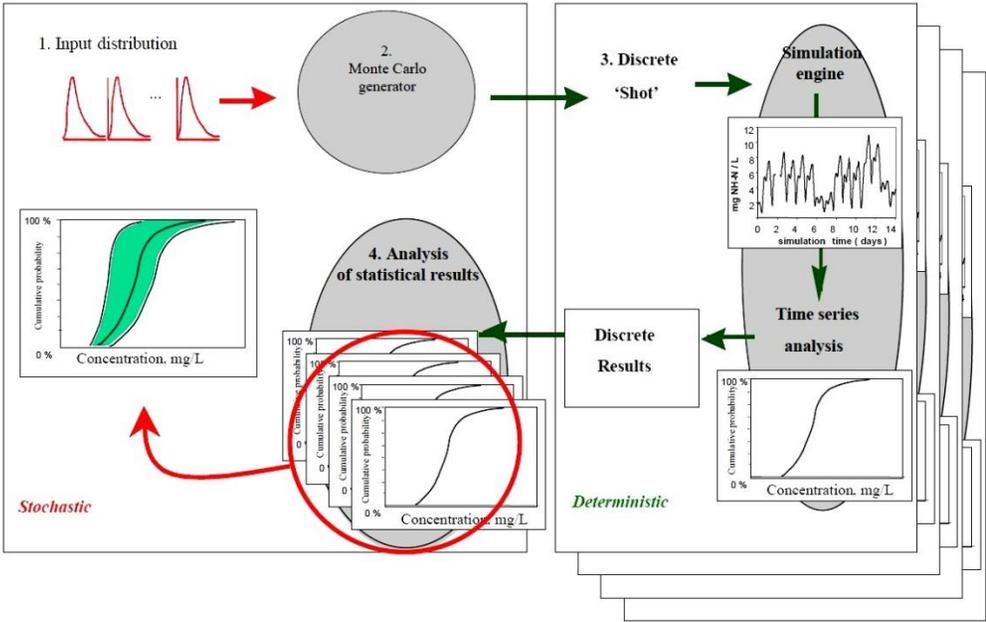
As the issue of uncertainty in the design of civil infrastructures becomes increasingly recognized, proper assessment of the probabilistic behavior of engineering systems is essential. Making too restrictive assumptions when characterizing the uncertainty of the inputs to a system (e.g. imposing a specific type of statistical distribution), or simplifying the complex behavior and the interaction of uncertain inputs for estimating the PONC (e.g. linearization of the *performance function* in the MFOSM and AFOSM methods) may produce inaccurate estimates for the PONC in a complex and non-linear system. In such

cases, Monte Carlo simulation is the only viable alternative to provide numerical estimations of the stochastic features of the system response. In Monte Carlo simulation, the system response of interest is repeatedly evaluated under various system model input and parameter sets generated from known or assumed probabilistic laws. It offers a practical approach to uncertainty analysis because the random behavior of the system response can be duplicated probabilistically (Tung et al., 2006).

USACE (1996) developed a practical risk-based approach for flood damage reduction studies using structural flood prevention alternatives (e.g. Dam, Levee, etc.). In their study three types of models were used for deriving the probability distribution of incurred damage due to flooding events. Uncertainty-based design or simulation approaches have also been applied to water treatment plants and sewer systems (Gupta and Shrivastava, 2006; Willems, 2008). Korving et al. (2009) proposed a methodology for risk-based design and rehabilitation of sewer systems. In their approach the probability of failure of overflow with different sizing of the systems is estimated first, and then, depending on the amount and the intensity of overflow, the total cost, which comprises the cost of construction and the damage due to overflow, is calculated. The optimum sizing of the system is calculated by finding a sizing which corresponds to the minimum expected total cost (total cost is a random variable; therefore its expected value is minimized).

The application of Monte Carlo simulation in conjunction with process-based models for estimating the PONC in WWTP systems has already been reported in several studies (Rousseau et al., 2001; Bixio et al., 2002; Benedetti et al., 2006a; Cierkens et al., 2011; Martin et al., 2012). Rousseau et al. (2001) proposed a methodology for assessing the concentration CDFs by combining the Monte Carlo simulation with the ASM1 model under dynamic state conditions. The variability of influent was characterized by generating a 6-month time series for each Monte Carlo trial and the uncertainties in model parameters were described by triangular or truncated normal distributions. The mean and variance of

uncertain model parameters were determined using available values in the literature. In addition to their uncertainty propagation scheme which could be used for the PONC calculation corresponding to a specific effluent standard, they also proposed the calculation of concentration-duration-frequency curves for different effluent standards. This analysis gives insight into the duration of noncompliance events which may have different impacts on the environment and entail different legal responsibilities and penalties. Bixio et al. (2002) applied the same uncertainty propagation scheme to an upgrade of a conventional WWTP towards more stringent effluent standards on nutrients. The steps for implementing the procedure can be summarized in four steps: 1) expressing the uncertainty/variability of inputs in terms of PDFs, 2) using Monte Carlo simulation to randomly sample a value from each uncertain variable to provide the process-based model with required parameters and inputs, 3) running the process-based model for each randomly-selected set of inputs (shot), 4) storing the results and continuing steps 1 to 3 until the specified number of model iterations is completed (Figure 2-7).



**Figure 2-7** Monte Carlo simulation framework for the probabilistic design of WWTPs (Bixio et al., 2002)

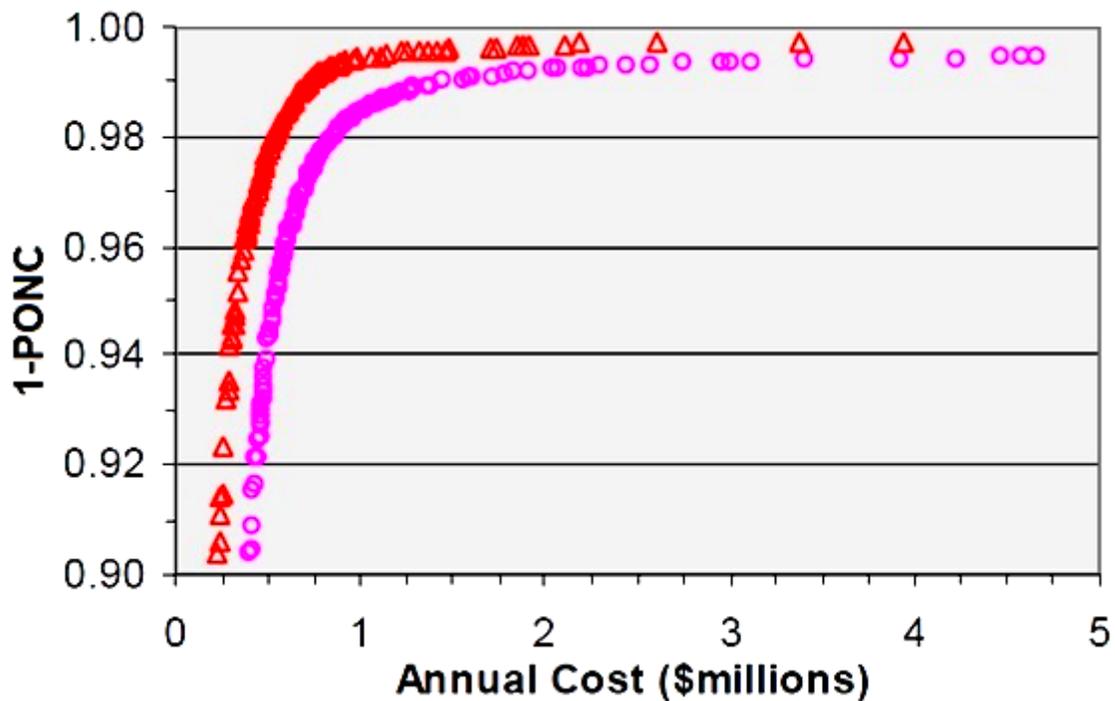
Compared to the ATV-DVWK-A 131E (a traditional method for design, see Section 2.1 for details), the proposed analysis tool showed that with a smaller sizing (which would result in saving 1.2 million euro) the WWTP was still capable of meeting the effluent standards with a high probability of compliance. Since the plant performance was assessed under different 6-month dynamic flow and loading scenarios including both dry and wet weather flow (DWF and WWF) conditions, the results are expected to be more representative of the plant performance compared to the work of Martin et al. (2012) in which the performance of different design alternatives was evaluated under a 28-day dynamic flow and loading time series which did not include any WWF conditions.

The influent generation method used in that study was developed in the Risk Assessment Project (RAP) software package (Rousseau et al., 2001) which simplifies the task of uncertainty analysis in the WEST simulation environment (Vanhooren et al., 2003). In the RAP package the generation of influent time series is achieved by dividing the different input variables into dependent and independent ones. The time series of the independent variable(s) (flow was usually considered as an independent variable) is randomly generated according to a distribution defined by the user. Then the corresponding values of different influent constituents are generated using regression equations that relate the dependent variables to the independent ones and a random component generated according to the statistical distribution assigned to each dependent variable. One of the main shortcomings of this method is that for a specific variable the correlation in time or the autocorrelation is not respected when the sequence of the independent variables are randomly generated according to their corresponding statistical distribution. Moreover, the occurrence of a first flush event is determined by considering only the randomly-generated time series of flow and checking for certain conditions at each time step. Therefore, the effect of rainfall events and their intensities on the characteristics of the flow and influent compositions are not considered.

In another effort Huo et al. (2006) applied the Monte Carlo simulation to the ASM1 model under steady and dynamic state conditions. The uncertainty of ASM1 model parameters was characterized in terms of PDFs. The statistical distribution of 19 ASM1 model parameters was developed from published literature values by Cox (2004) using Bayesian statistics. Most of the parameters were assumed to have lognormal distributions in order to generate only positive values. Meanwhile time series models were used to characterize the variability of the influent to the WWTP. Despite the characterization of model parameter uncertainties and the generation of dynamic influent time series to represent the temporal variation in flow and load, there might be a huge difference between the simulated effluent and the real one as in their study the modeling was limited to the biological processes in bioreactors and no modeling tool was used for simulating the performance of the secondary clarifier (i.e. the efficiency of secondary clarifier was simply assumed to be 99.75 percent).

To derive the relationships between the maximum probability of compliance and cost, Doby (2004) used stochastic optimization for the design of a WWTP. Model parameter uncertainties were characterized by truncated normal distributions, whose parameters were estimated from those available in the literature. The variability of the flow was taken into account by repeating the recorded historical data (a 5-year period) over the estimated life time of the project (e.g. 20 years). The distributions of certain critical ratios (e.g. TKN/COD or P/TKN) and concentrations of other influent constituents were also estimated from available historical data and their relationships with the daily flow. The steady state Water Research Commission model (Section 2.1.3) was used to simulate the biological processes of the WWTP system. An empirical cost function was developed to relate the sizing of design variables with the total cost which includes the capital cost of the bioreactor, the aeration, and the pumping costs. To develop a trade-off curve between cost and the probability of compliance, a tight constraint was put on the volume of the bioreactor (since it was assumed that the final cost is mainly a function of bioreactor size) and then an objective function was defined for the maximization of the probability of compliance with

the effluent standards. For different total cost values a separate optimization using the GA method was carried out with the aim to maximize the probability of compliance with the effluent regulations. The design variables were limited to four variables: internal recycle ratio, sludge recycle, volume fraction of the anoxic tank relative to the total bioreactor volume, and total volume of the bioreactor. The relationships between cost and maximum achievable probability of compliance showed that building a more expensive plant does not necessarily correspond with a higher probability of meeting the effluent standards.



**Figure 2-8** The relationship between the probability of compliance (1-PONC) and annual cost for two design alternatives (Adapted from Doby, 2004) ○ Alternative 1 △ Alternative 2

Although it was claimed that the variability of influent composition and flow was taken into account, the fact that a steady state model was used for the simulation of the WWTP means that the dynamics of the system could not be modeled realistically. Actually, there is a fundamental error in the way that the daily variability of the influent and flow was

characterized and used as inputs to a steady state model (Sharma, 1993; Doby, 2004) for estimating the corresponding effluent variability. This can be explained as follows:

In steady state models, the state variables (e.g. the concentration of substrate or sludge in the bioreactor) are estimated as a function of fixed inputs, model parameters, and sizing of the plant (e.g. volume of the bioreactor). Moreover, it is assumed that the SRT is longer than the minimum SRT required for the utilization of the substrate or nitrification to take place (Henze et al., 2008). Therefore, characterizing the daily variation of flow and influent in terms of a PDF or even time series, and estimating the corresponding daily effluent of the system using a steady state model is a mistake in the first place, even when all parameters of the system are assumed to be fixed (i.e. the daily effluent in a specific day is not only affected by the influent in that day). In other words, the daily effluent of the system can only be estimated by a steady state model if the SRT of the system is at least 3 to 5 days<sup>1</sup>. Moreover, the design of the system was also limited to the bioreactor and its corresponding design variables whereas in reality there are important interactions among the various units of a WWTP.

## **2.5 Cost Analysis**

Calculating the cost of WWTP is a complex task as the available cost functions (see Table 2-3 for examples of cost functions) have usually been developed for a specific region and reflect the planning and building procedures that are used in that specific region (Bode and Lemmel, 2001; Benedetti et al., 2006b).

---

<sup>1</sup>According to Henze et al. (2008) the theoretical minimum SRT under ideal conditions should be at least 3 days.

Gillot et al. (1999) divided the total cost of a typical WWTP (oxidation ditch) into three main categories including the investment cost, fixed operating cost and variable operating cost. The contribution of each treatment unit cost to the total investment cost can be calculated as a function of either the global characteristics of the plant (e.g.  $I = AQ^n$  where  $I$  is investment cost (€),  $Q$  is the flow rate ( $\text{m}^3/\text{h}$ ) and  $A$  and  $n$  are constant parameters (Hernandez-Sancho et al., 2011)) or the size of the different treatment units (e.g. the volume of the bioreactors). The fixed operating cost can be calculated using a set of cost functions that relate different fixed operating costs to the plant's size or the unit size of the different treatment units. To calculate the variable operating costs, simulation data in conjunction with a set of cost functions reported in literature can be used to better take into account the dynamics of the system. It should be noted that the calculation of the costs requires some skill. Incorporation of site-specific data (if available as in the case of WWTP upgrade projects) into the development or modification of cost functions is highly recommended as the cost functions reported in the literature have been usually developed using regional data at a specific period of time (Balmer and Mattsson, 1994; Gillot et al., 1999).

Considering the scarcity of input data required for the application of different cost functions, difficulties in reflecting local cost factors and a constant need to update the cost functions or their parameters, some researchers have advocated the development and application of software for cost analysis (Harris et al., 1982; McGhee et al., 1983; Pineau et al., 1985). One of the software packages of this type is CAPDET (Harris et al., 1982) which was originally developed by the US Army Corps of Engineers for the design and cost evaluation of WWTPs. The CAPDET software is capable of calculating the total cost of a plant using basic global parameters (e.g. basic influent and flow values) and the process configuration. To simplify the task of providing the input data for the calculation of cost, default values based on recorded data sets are suggested for use in case of limited availability of data. Moreover, the flexibility of the software allows users to easily change the default values of some parameters to better reflect the site-specific conditions into the

calculation of cost. However, care must be taken in cases where the CAPDET software is to be applied for cost calculation of WWTPs in conditions significantly different than those of the US. For example, in the study of Pineau et al. (1985) on the applicability of the CAPDET software to cost estimation of WWTPs in Canada, it was found that the software was capable of predicting the construction cost within  $\pm 20$  percent of the actual cost while the operation and maintenance costs were within  $\pm 30$  percent of the observed costs. Replacing the default values for some of the inputs (e.g. overflow rate, oxygen transfer coefficients, etc) with local values and increasing the unit costs for excavation, concrete walls and slabs by certain percentages were among the factors that resulted in the prediction of cost within the expected accuracy levels (Pineau et al., 1985).

**Table 2-3:** Examples of total cost function (Doby, 2004)

Note: All costs are in US dollars

Bioreactor	$C_{bioreactor} = \frac{1.15 \left[ 0.012 \left( (1 - f_{unaerated}) V \right)^{0.83} + 0.0072 \left( f_{unaerated} V \right)^{0.83} \right]}{\sum_{i=1}^n \frac{1}{(1+r)^i}}$ <p> <i>f<sub>unaerated</sub></i> : fraction of unaerated volume  <i>r</i>: interest rate (e.g. 6% )  <i>V</i>: Volume (m<sup>3</sup>)         </p>
Secondary clarifier	$C_{clarifier} = \frac{3105 \left( \frac{Q}{155.9 e^{-3.6K}} \right)^{0.91}}{\sum_{i=1}^n \frac{1}{(1+r)^i}}$ <p> <i>Q</i>: Influent flow rate (mgd)  <i>K</i>: A coefficient that equals 0.6214 if a primary clarifier is used and 0.4691 without primary clarifier  <i>r</i>: interest rate  <i>n</i>: project life time (year, e.g. 20)         </p>
Pumping	$C_{pumping} = \left[ 5 + \frac{1043.94 (aQ)^{1.852}}{(100)^{1.852} D^{4.8655}} \right] \times [101.31D - 231.5] + 0.000257 (sQ)^{1.852} + 23.881$ <p> <i>a</i>: internal recirculation ratio  <i>s</i>: return sludge recycle ratio  <i>Q</i>: influent flow rate (mgd)  <i>D</i>: Pipe diameter (inch)         </p>
Aeration	$C_{aeration} = 25.38 V_{aerobic}$ <p> <i>V</i>: Aerobic bioreactor volume (m<sup>3</sup>)         </p>
Effluent Penalties	$C_{effluent} = \max \left[ (S_{effluent} - S_{standard}), 0 \right] \times 1000000$ <p> <i>S<sub>effluent</sub></i>: Effluent concentration (mg/L)  <i>S<sub>standard</sub></i>: Effluent standard (mg/L)         </p>

## 2.6 Problem statement

In the previous sections different methodologies for the design of WWTPs were reviewed. In this section the main shortcomings and research gaps of current design methods are summarized (Table 2-4). In traditional design methods (i.e. using design guidelines) uncertainty is not taken into account in an explicit manner. To account for the various sources of uncertainty safety factors are applied in terms of safety coefficients, overestimation of influent characteristics or assuming more stringent effluent standards. In this approach there is no clear method for using these safety factors and subjective judgement of engineers could lead either to an over-sized or an under-sized WWTP. Moreover, it should be emphasized that building a more expensive plant does not always corresponds with a higher probability of meeting the effluent standards.

In model-based optimization methods design engineers use *design values* (like the traditional methods) to characterize the influent, effluent standards, and model parameters. They generally select conservative values so that uncertainty is handled in an implicit manner. The problem with this approach is that there is no clear way for determining the proper combination of design input values that is representative of the WWTP system. Moreover, although some constraints are put on the effluent values, the final design does not guarantee compliance with the effluent standards because in these methodologies a steady state condition is assumed whereas in reality the WWTP is a dynamic system. Moreover, they do not tackle the problem within an uncertainty framework. Therefore, the PONC cannot be estimated in a realistic manner.

**Table 2-4** Comparison of different design approaches

<b>Design Approach</b>	<b>Model Type</b>	<b>Handling the Effect of Uncertainty</b>	<b>Degree of Compliance to the Effluent Standards</b>
Design Guidelines	Steady state dynamic model for checking the performance	Application of safety factors	No
Model-based Design	Steady state dynamic model for checking the performance	Making conservative assumptions and imposing constraints on the design variables	No
Robust Optimal Design	Steady state	Imposing constraints on the design variables, and maximizing the robustness index	No
Probability-based Design	Steady and dynamic state	Explicit characterization of uncertainty/variability and minimizing the PONC	Yes

In the robust optimal design methodology for the design of WWTPs, it is assumed that the model equations are a good-enough representation of reality and the issue of model-structure uncertainty is not considered. Although it is explicitly accepted that the parameters are uncertain, the robustness measures are not subjected to uncertainty, as the robustness is calculated at the nominal values of the uncertain parameters. Therefore, even if one could reasonably characterize the uncertain parameters in terms of PDFs using historical records and years of experience, there is no way to incorporate this added knowledge into the analysis. Also, it should be emphasized that the notion of system robustness presented by Uber et al. (1991a) does not include the probabilistic concept of performance failure or the probability of non-compliance with effluent standards.

The only design methodology that provides the design engineers with a quantitative PONC is the probability-based design method, but the characterization of the different types of uncertainties have lacked mathematical rigor in the approaches published so far. For example, in the reviewed papers model parameter uncertainties are characterized in terms of

PDFs whose parameters are estimated using expert opinions or reported values in the literature. Although expert opinion may be the only option when the available information on model parameters is scarce or not available, there are statistical methods which can be used to incorporate the available historical information into the characterization of uncertainty and updating the subjective PDFs of the parameters. Moreover, there are serious shortcomings in the influent flow and composition generation (Section 2.4.4).

## **2.7 Objectives**

Considering the shortcomings of the previous studies on the design of WWTPs, the current study aims to develop a model-based probabilistic method for the design of WWTPs. The relevant sources of uncertainty considered in this study are model parameter uncertainty and the influent variability. Different methodologies have been developed to properly take their effect on the calculated PONC into account. The main objectives of this PhD study can be summarized as follows:

**i. Develop and select a set of models for design and simulation of WWTPs considering influent variability and model parameter uncertainty**

To simulate the behavior of a WWTP several types of model will be used. Dynamic process-based models will be used to simulate the various processes taking place in a WWTP. Design models will be used to come up with preliminary designs. Other types of models such as statistical models will also be used to explicitly characterize the variability and uncertainty of the inputs to the models (e.g. time series models for the generation of rainfall and the influent temperature profile).

**ii. Develop a methodology for quantitative model-based assessment of the probability of non-compliance with effluent standards**

Using the set of modeling tools for simulating the various processes in a WWTP and the characterization of input and model parameter uncertainty/variability, a methodology will be proposed to propagate these variability/uncertainties and estimate their effect on the effluent variability/uncertainty and consequently the probability of non-compliance.

**iii. Develop a procedure for probability-based design that uses these tools and procedures and allows a transition from today's design methods**

In this step the set of models selected in i) is integrated with the methodology in ii) to develop a probabilistic design procedure for WWTPs applicable for design engineers and practitioners. The procedure should allow an easy transition from current design practice.

**iv. Demonstrate the feasibility of the probabilistic design procedure by applying it to a real case study**

The proposed probabilistic design method should be applicable to real design projects so the practitioner could potentially adapt it and benefit from the added knowledge and benefits that come with its applications. To better demonstrate the applicability and requirements of the design method (e.g. required data types, modeling tools, etc), the proposed probabilistic design method will be applied to the upgrade of the Eindhoven WWTP in the Netherlands.



## **CHAPTER 3: PROPOSED PROBABILISTIC DESIGN METHODOLOGY**

This section aims to provide a general outline for the proposed probabilistic design method of WWTPs. Figure 3-1 depicts the proposed methodology which is comprised of three main steps including: 1) Generation of a set of pre-designs with different levels of safety, 2) Preliminary evaluation and screening of pre-designs, 3) Quantification of the probability of non-compliance (PONC) with the effluent standards and total cost.

As indicated in Figure 3-1, the first step in the proposed design methodology is the generation of a set of pre-designs generated using a steady state design tool (i.e. a steady state model or a design guideline). The initial sizing of a plant is thus not significantly different from conventional steady state-based design methods in which the proper values are selected for design inputs (e.g. influent flow, COD concentration, SVI, etc) and the size of different treatment units is calculated using a set of equations and/or rules. However, in contrast to the steady state design methods in which single values are selected for design inputs to calculate the size of different treatment units, a range of values reflecting different levels of safety are assigned to design inputs and a set of values is calculated for the different treatment units. Section 3.1 explains the details of the generation of a set of pre-designs using a steady state design tool. These pre-designs will be further evaluated in the second part of the proposed probabilistic design methodology as shown in Figure 3-1 (i.e. Preliminary evaluation and screening of pre-designs).

After generating a set of pre-designs in step 2), design alternatives that either have a very poor performance in terms of effluent quality or that do not yield a significant difference in the performance under a typical dynamic flow and loading scenario are identified. To identify these alternatives, a two-step screening procedure based on the statistical

distribution of the design outputs (i.e. the size of different treatment units) and dynamic simulation of the design alternatives is proposed (details in Section 3.2). This allows selection of a handful of design alternatives for which PONC will be calculated in the third step of the proposed probabilistic design method.

The third step of the proposed probabilistic design method constitutes procedures used for the model-based calculation of PONC considering the variability of influent time series and the parametric uncertainty in the dynamic model of each of the selected design alternatives. Synthetic generation of dynamic influent time series, uncertainty characterization and random generation of model parameters, different methods for propagating the effect of influent variability and model parameter uncertainty, and in the end calculation of PONC and the total cost for the selected design alternatives are covered in Section 3.3.

The details of the proposed probabilistic design method are explained in this chapter. Chapter 4 focuses on the influent generation. The application of the proposed design method to an actual case study and its corresponding results can be found in Chapter CHAPTER 5:.

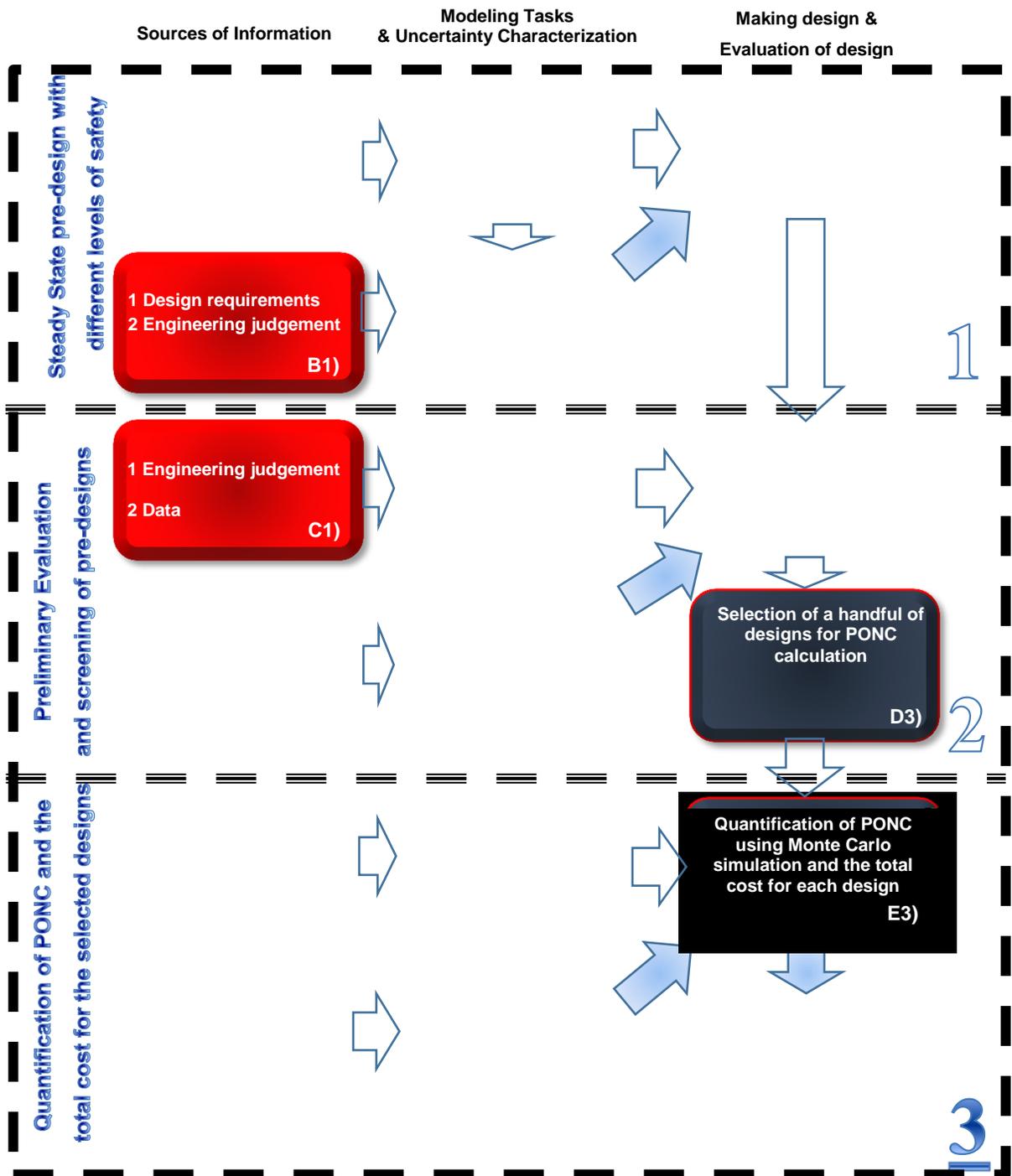


Figure 3-1 Steps for the proposed probabilistic design of WWTPs



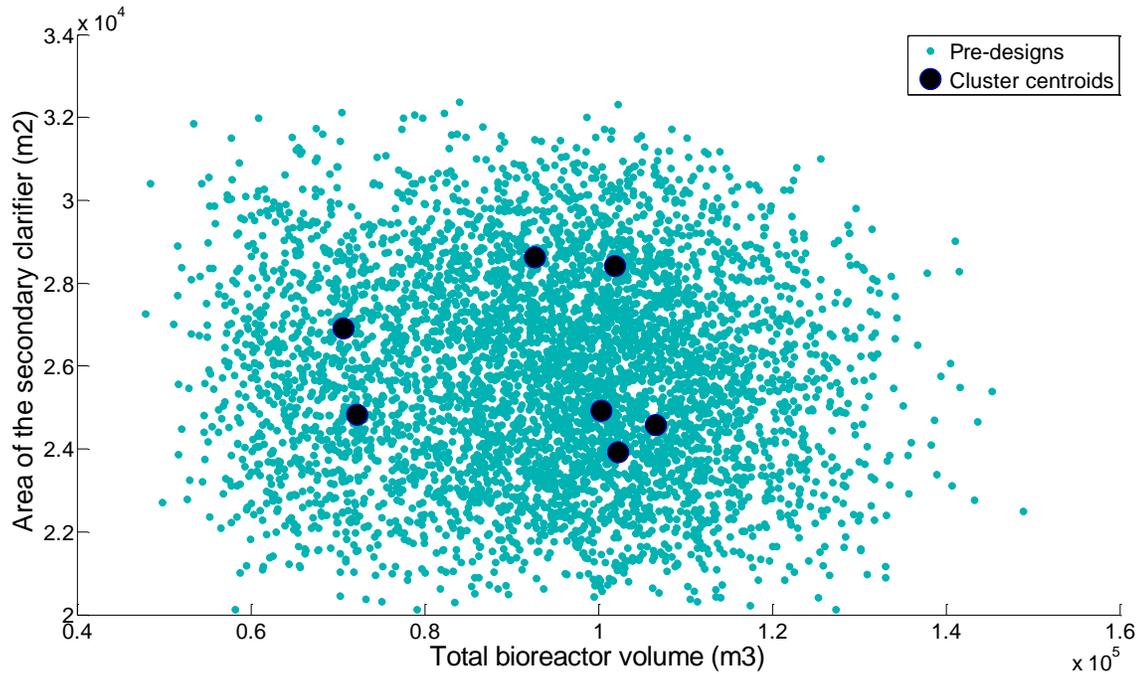








cluster centroids be selected as representative design alternatives for which the PONC will be calculated. Figure 3-4 shows the distribution of the area of the secondary clarifier versus the total bioreactor volume corresponding to a hypothetical set of pre-designs generated according the procedure in Section 3.1.2. The black dots in Figure 3-4 represent the cluster centroids that were determined based on a k-means clustering method with 7 clusters.



**Figure 3-4** Generated pre-designs and cluster locations for a k-means clustering with 7 clusters

Although the k-means clustering can be used for selecting a handful of designs that are dissimilar in size, the designers may be interested in selecting a specific design alternative that does not necessarily correspond to a cluster centroid. For example, neither of the cluster centroids in Figure 3-4 (i.e. the black dots) is located at the boundaries of the design output region. Therefore, if the cluster centroids are selected for the calculation of PONC, the designers will not have any idea regarding the performance of those designs that could

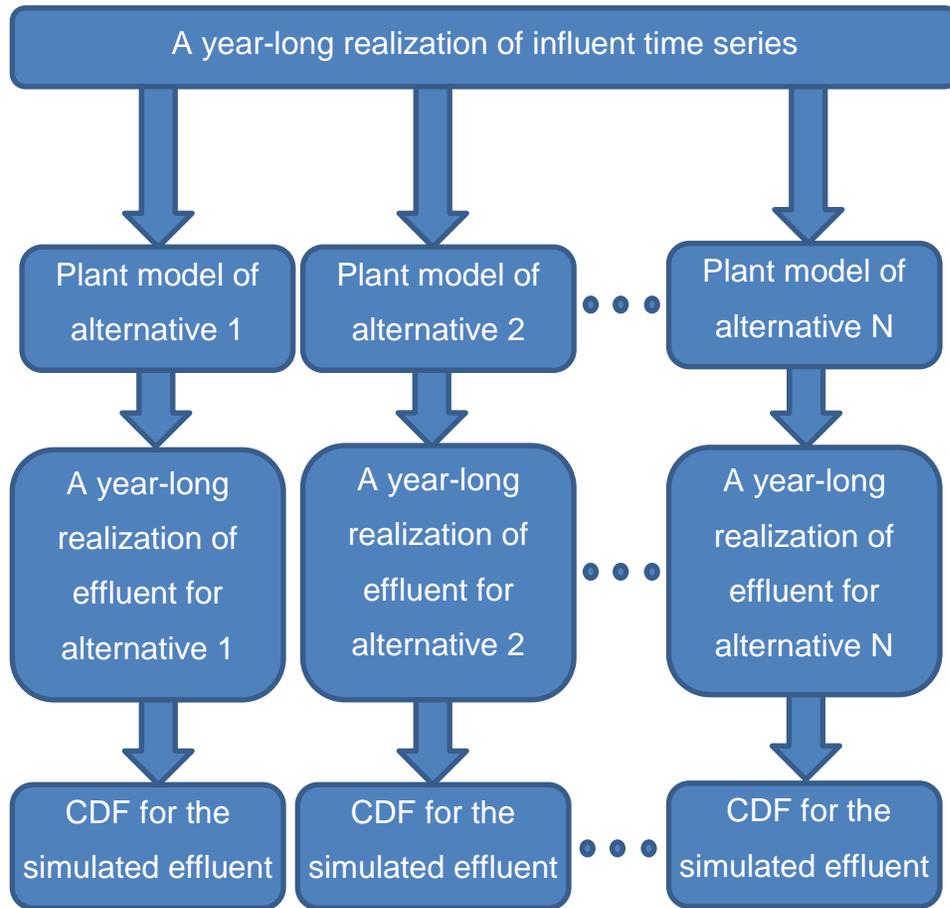
be significantly smaller or larger than the average of pre-designs. In addition, some of the selected cluster centroids might have a comparable performance under dynamic flow and loading conditions and including them in the final analysis for the calculation of PONC would only impose computational load without providing worthwhile information. Therefore, it should be emphasized that the k-clustering can be used as a preliminary tool for identifying a handful of designs. Some designs other than the cluster centroids could be added to the final selected designs after doing an evaluation of their performance under typical dynamic flow and loading conditions.

For the dynamic evaluation of the design alternatives, in this screening phase a typical year-long influent time series is used as input, generated using the influent generator presented in Chapter CHAPTER 4:. The performance of each design alternative is evaluated for the generated influent time series using a dynamic simulation model. Deciding on the type of dynamic model depends on several factors including previous modeling experience, availability of data, as well as the wastewater constituents whose effluent concentrations are of interest. The important point is that the dynamic model should be able to provide answers on the typical performance of designs regarding the removal of those effluent constituents for which the compliance to the effluent standards has to be assessed.

To compare the performance of the different design alternatives in removing pollutants, the cumulative distribution function (CDF) corresponding to the simulated effluent of each design alternative is constructed (Figure 3-5). The CDFs will be used as a means for eliminating those design alternatives with a poor performance in terms of effluent quality. In addition, the comparison of the CDFs can serve as a tool for eliminating design alternatives having the same treatment performance.

It is important to realize that the CDFs representing the performance of the design alternatives under the generated year-long influent time series will be different from the

CDFs by which the PONC will be calculated in the last step of the probabilistic design methodology (next section). For the latter, the inter-annual variability of the influent time series will also be considered.



**Figure 3-5** Calculation of effluent CDFs for the initial design alternatives

### 3.3 Quantification of PONC using Monte Carlo simulation

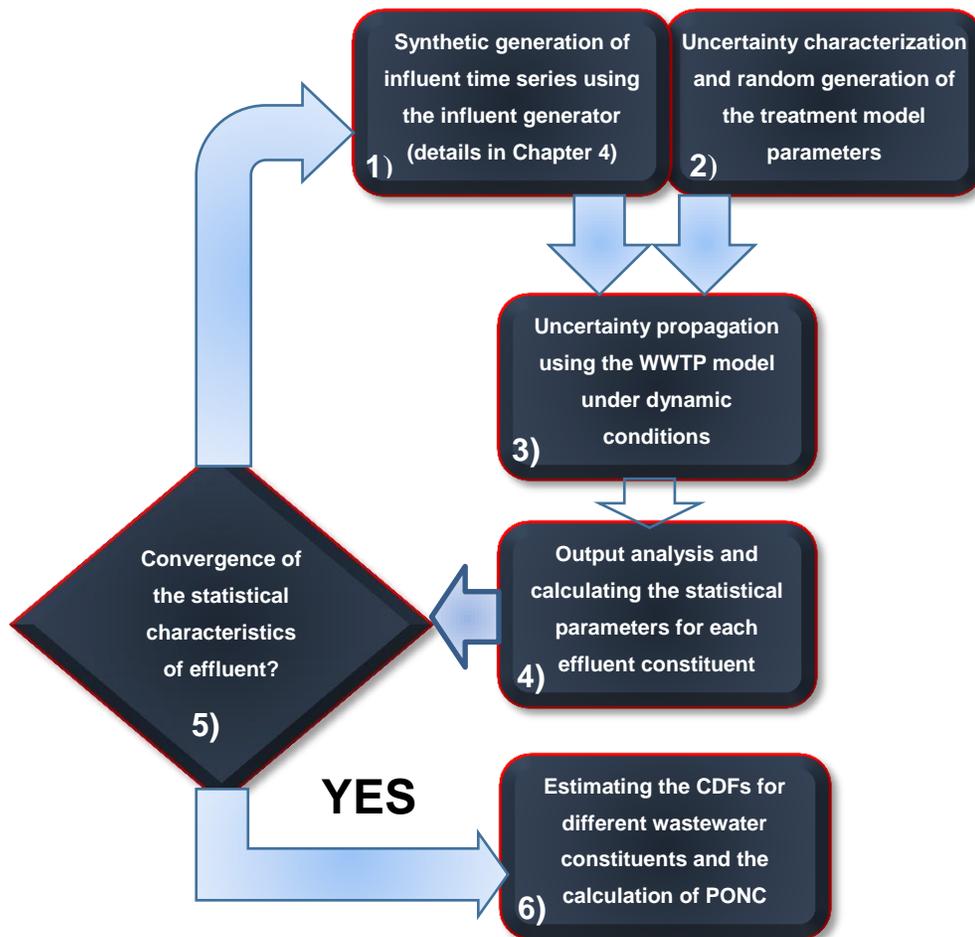
In the proposed methodology PONC will be estimated by performing uncertainty analysis using Monte Carlo simulation. Monte Carlo simulation is a computationally intensive method by which the system response of interest is repeatedly assessed under various model

factor sets generated from known or assumed probabilistic laws. It offers a practical approach to uncertainty analysis because the random behavior of the system response can be duplicated probabilistically (Tung et al., 2006).

Prior to setting up a Monte Carlo simulation, the relevant sources of uncertainty and variability that will affect the distribution of effluent wastewater compositions must be identified and taken into account in the calculation of PONC. In the proposed methodology the important sources of uncertainty and variability that are considered are: 1) the uncertainty in the parameters of the model used for simulating the treatment performance of designs under dynamic conditions, and 2) the variability in influent time series (including both the annual and inter-annual influent variability). The explicit characterization of uncertainty/variability in the performance of the technical components of the WWTP system is beyond the scope of this study (e.g. the uncertainty/variability in the functioning of sensors or pumps). Model structure uncertainty or uncertainty about the relationship among the different variables of a system (Beck, 1987) as well as evaluation of different model structures for selecting the optimum model structure for dynamic simulation of WWTPs are not tackled in this study.

Figure 3-6 illustrates the procedure that is used for the calculation of PONC considering model parameter uncertainty as well as influent variability. As indicated, the effect of influent variability is taken into account by synthetic generation of different realizations of influent time series (step (1) in Figure 3-6, briefly presented in Section 3.3.1 and discussed thoroughly in Chapter CHAPTER 4:). The effect of uncertainty in the treatment plant model is considered by characterizing the uncertain model parameters using different probability distribution functions and random sampling of parameters (step (2) in Figure 3-6 and Section 3.3.2). The dynamic WWTP model is run repeatedly using different realizations of influent time series and model parameters to propagate the effect of influent variability and model parameter uncertainty to the effluent (step (3) in Figure 3-6 and

Section 3.3.3). The simulated effluent time series must be aggregated to the same temporal resolution as the effluent standards (step (4) in Figure 3-6 and Section 3.4). The convergence of the Monte Carlo simulation is assessed by calculating the effluent CDFs and analyzing the changes for increasing numbers of simulations in the values of representative percentiles of different effluent CDFs (step (5) in Figure 3-6 and Section 3.4). Once convergence is achieved, the simulations are stopped and the corresponding PONC and the total cost are calculated for each design, step (6) in Figure 3-6.

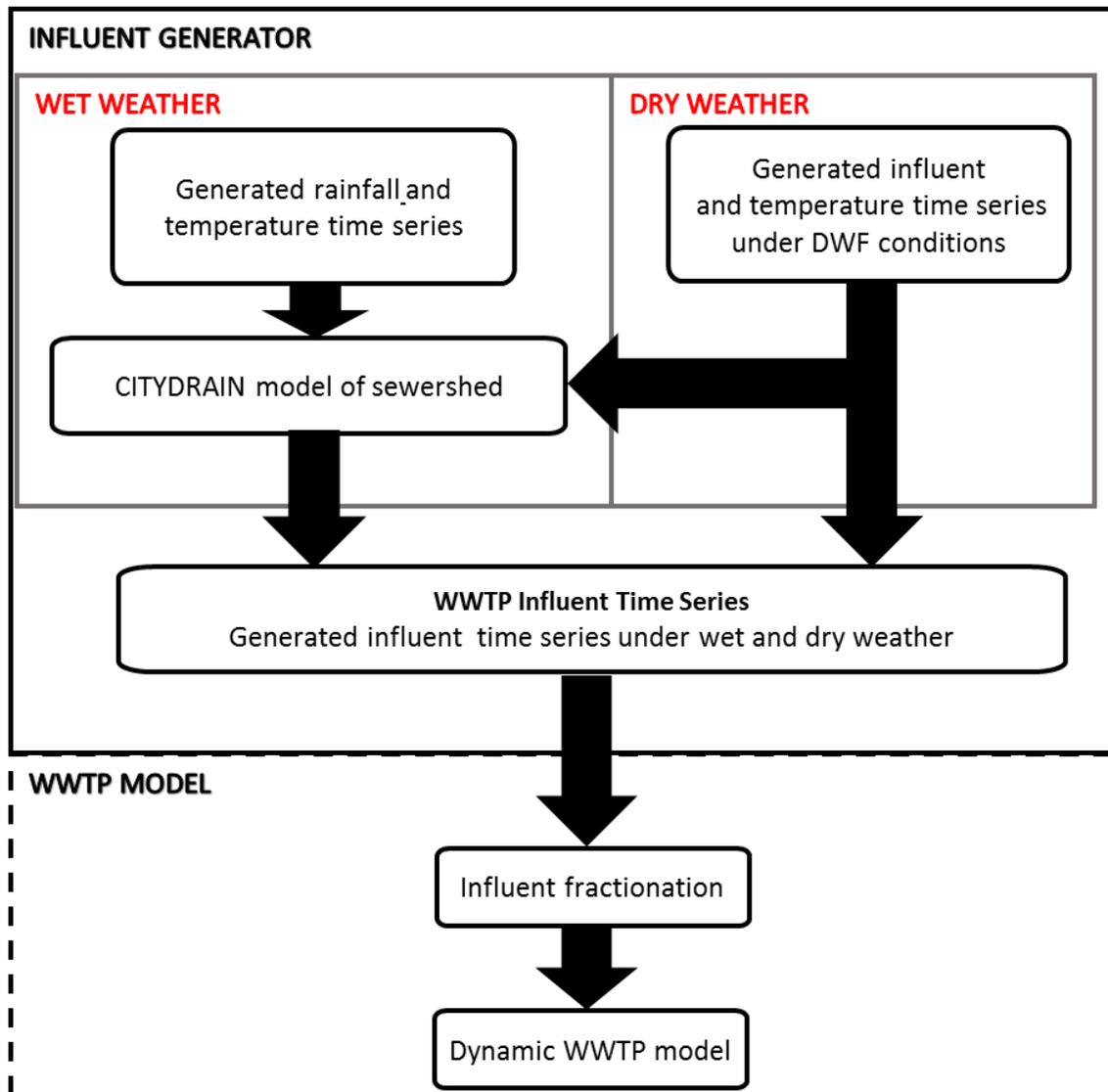


**Figure 3-6** Calculating PONC for a design alternative

### ***3.3.1 Synthetic generation of influent time series***

One of the major sources of uncertainty/variability that both plant designers and operators have to deal with is the dynamics of the influent. The importance of proper representation of influent time series stems from the fact that each year the influent time series to a plant is unique (although the main statistical properties could be the same). Therefore, even in cases of measured influent time series (which are available but are usually not long), synthetic influent generation is crucial in order to assess the performance of the plant for different realizations of dynamic flow and loading conditions that happen in different years. In this study an influent generator model comprising of a set of statistical and conceptual modeling tools was developed for synthetic generation of influent time series. The general structure of the developed influent generator is explained in this section and more details are given in Chapter CHAPTER 4:.

Two types of statistical models are used in the developed influent generator, one for the synthetic generation of rainfall series and one for the time series describing the influent during dry weather flow (DWF) conditions. These two time series (e.g. rainfall and influent in DWF conditions) serve as stochastic inputs to a conceptual model of the sewershed in order to generate the influent time series during both wet weather flow (WWF) and DWF conditions (Figure 3-7).



**Figure 3-7** Schematic of the proposed influent generator (detailed in Chapter CHAPTER 4:).

**I. Synthetic generation of influent time series in DWF conditions**

A combination of Fourier series analysis and a multivariate autoregressive time series model (Neumaier and Schneider, 2001) is used for the synthetic generation of the influent during DWF conditions. First, different Fourier series estimates are used to capture the average daily periodic behavior of the wastewater constituents and then a multivariate

autoregressive model is used to model the stochastic behavior of the different influent wastewater constituents (e.g. flow, BOD, COD, etc).

Ideally, the parameters of the periodic multivariate time series model should be extracted from measured influent time series. However, in case of limited available data, the periodic patterns of the influent time series could be estimated based on expert opinion or analyzing the influent time series of a similar sewershed.

## ***II. Synthetic generation of rainfall time series***

Realistic generation of rainfall time series is crucial as it is one of the most important factors that affect the dynamics of the influent of a combined sewage system (CSS) during WWF conditions. In the current study, a two-state Markov chain model originally proposed by Richardson (1981) is used for stochastic generation of the daily rainfall time series and then two successive time series disaggregation techniques will be applied to convert the daily rainfall time series into a rainfall time series with 15-minutes temporal resolution (to better capture the effect of sub-hourly variations in rainfall). The main objective of this procedure is to better extract and incorporate the important statistics of the rainfall records (e.g. the average rainfall amount, the number of wet days, seasonal variation of both frequency and intensity of rainfall, etc.) into the synthetically-generated rainfall time series.

The data required for calculating the rainfall generator model parameters include (usually available) long term daily rainfall time series for synthetic generation of rainfall time series, and some hourly rainfall time series for estimating the parameters of the time series disaggregation model.

## ***III. Synthetic generation of influent time series in WWF conditions***

Synthetic generation of the influent time series during WWF conditions is relatively more complicated than the generation of the influent time series during DWF conditions. The

difficulty arises as various phenomena occur during WWF conditions and as the availability of measured data is usually limited for these periods. Hence, using a purely statistical model may result in significant discrepancies between the simulated and observed time series. Therefore, a combination of statistical modeling techniques and a conceptual model was proposed to generate the time series of the influent during WWF conditions. The Matlab-based CITYDRAIN (Achleitner et al., 2007) model was selected as the conceptual model as it takes into account the basic phenomena that govern the amount and dynamics of the influent and also requires only a small number of parameters whose values or ranges of values can be inferred from the basic information of the sewershed.

However, one should be aware of the fact that modeling the influent time series in WWF conditions using a conceptual model may not lead to reliable results unless the model is calibrated and the effect of different sources of uncertainties on the outputs (e.g. flow and other pollutants) are taken into account. To this aim, a Bayesian framework was used to update the probability distribution functions, initially assigned to the parameters of the CITYDRAIN model (i.e. estimating the posterior distribution of parameters using their prior distribution and the measured data). It should be noted that the proposed Bayesian approach is not only capable of capturing the effect of model parameter uncertainty, but also of capturing the effect of other sources of uncertainties that could result in some discrepancies between the simulated influent time series and the observed one.

Some of the CITYDRAIN model parameters that have a physical meaning (e.g. area contributing to runoff) can be derived from available reports and studies on the sewershed. Determining a range of values to those parameters which are either difficult to measure or do not have a physical meaning can be done based on previous modeling studies and/or expert opinion. However, it should be noted that when Bayesian calibration is used for updating the initial ranges of uncertainty, one should use concurrently-measured rainfall and effluent time series from the sewershed (influent to the WWTP).

Once the uncertainty ranges of the CITYDRAIN model parameters are updated, synthetic influent time series can be generated for a desired number of years considering the variability in the inputs of the CITYDRAIN model (i.e. rainfall and influent time series in DWF conditions). Synthetic generation of influent time series using the explained influent generator can be performed as follows:

- 1) Synthetic generation of the 15-minute time series of rainfall for one year.
- 2) Synthetic generation of the 15-minute time series of the influent in DWF conditions for one year.
- 3) Sampling a point from the posterior distribution of the CITYDRAIN model parameters.
- 4) Inputting the generated time series 1) and 2) and the parameters sampled in 3) and running the CITYDRAIN model for one year.
- 5) Repeating 1) to 4) for a desired number of years.

### ***3.3.2 Uncertainty characterization and random generation of WWTP model parameters***

Characterization of model parameter uncertainty can be done using different types of distribution functions (e.g. triangular or truncated normal distributions). However, it is recommended to use the uniform distribution function in situations where little information is available on the value of model parameters in order to avoid the under-estimation of uncertainty in model outputs (Freni and Mannina, 2010).

Determining PDFs for characterizing the uncertainty in WWTP model parameters (e.g. lower and upper limits for uniform distributions or mean and standard deviation for truncated normal distributions) is crucial for assessing the effect of model parameter uncertainty on the model outputs (e.g. effluent time series) which in turn affects the calculated PONC. The parameters of the different PDFs can be determined based on lab analysis of wastewater, previous modeling studies, as well as expert opinion.

Once the uncertainty in model parameters is characterized using different distributions, a random vector of model parameters must be sampled and used in each Monte Carlo run. The two commonly-used sampling methods used in the Monte Carlo simulations of WWTPs include the random sampling (RS) and Latin Hypercube sampling (LHS) that are explained in the following section. For the proposed probabilistic design either of the sampling methods can be used. However, as it is explained in the following sections, the implementations of different sampling methods require different settings and inputs which may affect the decision of users on selecting a method (see part II, Section 5.3.3 for the selected sampling method for the current case study).

### *I. RS and LHS sampling*

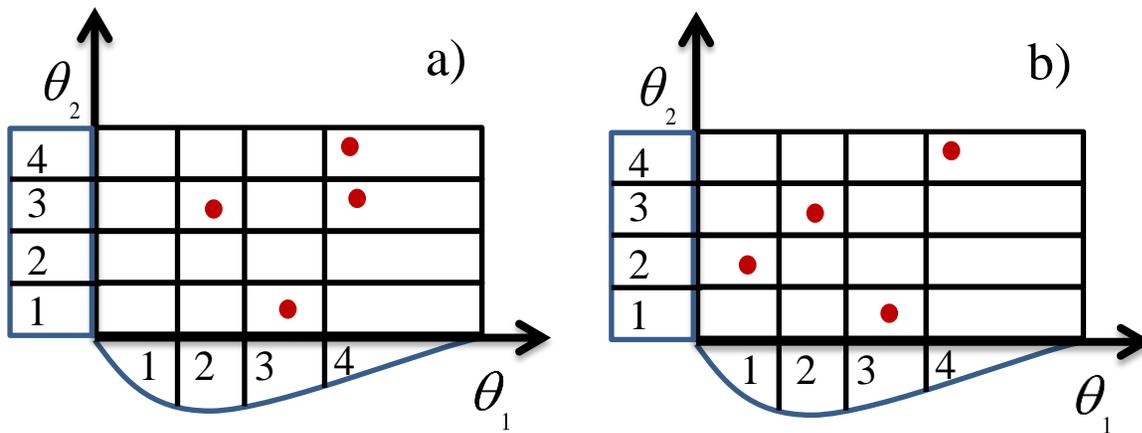
In the RS procedure, at each Monte Carlo run, a vector of model parameters is randomly sampled from the joint distribution of parameters. The sampling of parameters at each Monte Carlo run is independent from the previous ones. Therefore, in this sampling approach, the coverage of the entire support of distributions (used for the characterization of model parameter uncertainty) might not be guaranteed, unless a large-enough number of Monte Carlo simulations is performed.

An alternative sampling to the RS method for exploring the support of different parameter distributions is the LHS method. In the LHS method, the support of distributions is divided into  $N$  sub-intervals (e.g.  $N=4$  in Figure 3-8) with equal probabilities and a value is randomly generated in each sub-interval. Compared to the RS method in which different samples are generated by random sampling directly from the entire range of distributions, in LHS, random sampling is performed in each sub-interval and all sub-intervals are sampled.

Figure 3-8 illustrates the result of generating 4 vectors of parameters in a 2-dimensional parameter space generated using the RS and LHS sampling methods. As indicated in a), in this particular realization of 4 samples, generated according to the RS method, no value is sampled from sub-interval (1) of parameter  $\theta_1$  and sub-interval (2) of parameter  $\theta_2$ .

However, the sampling result based on the LHS method indicates that the generated values include representatives from all sub-intervals.

In general, the application of LHS could reduce the number of sampled values required to reach convergence of the outputs distributions (Tung and Yen 2005). However, there could be some cases where LHS sampling would not lead to a more rapid convergence of output distribution compared to the RS sampling as the convergence also depends on the complexity of the model and its parameters.



**Figure 3-8** (a) Schematic of a RS and (b) LHS procedures

In addition, in the RS method, the sampling of parameters at each Monte Carlo run is independent from the next one and in each run the convergence of the output distributions can be checked to determine whether more simulations are required or not. In contrast, in the LHS method, the number of Monte Carlo runs should be determined and samples generated before running any simulations. Therefore, if the selected number of Monte Carlo simulations turns out to be insufficient, the users cannot simply add more samples (like in the RS method) unless the consistency of the LHS procedure is insured.

A possible solution to increase the size of samples in the LHS method is the replicated LHS method (McKay et al., 1979) in which instead of generating  $N$  number of samples using the LHS,  $k$  number of LHS designs with  $n$  number of samples each, is generated ( $N=k \times n$ ).

After the termination of each Monte Carlo simulation with  $n$  samples, the convergence of the output distributions is checked, and if more simulations are required, another  $n$  samples is generated using the LHS and Monte Carlo simulation continues using the newly-generated samples. The efficiency of the repeated LHS depends on the appropriate choice of  $n$  as selecting it too large may not result in significant reductions of model runs and a value that is too small could result in inadequate coverage of the entire parameter space (Benedetti et al., 2011).

## ***II. Introducing correlations between parameters***

One of the important factors in Monte Carlo simulations that could affect some of the statistical properties of the simulated output distributions is proper incorporation of possible correlation structures in the sampling of uncertain parameters. Different methods presented in the literature can be used to introduce a desired correlation structure among the sampled values (Iman and Conover, 1982; Tung and Yen, 2005). However, some of the methods suffer from certain shortcomings and their application depends on the validity of a set of assumptions regarding the marginal distribution of the parameters (Tung and Yen, 2005).

One of the commonly-used methods for introducing correlation among uncertain parameters is the method of Iman-Conover (Iman and Conover, 1982). Being independent from the type of marginal distributions, applicability to any sampling scheme (e.g. RS or LHS), and relatively simple implementation have been mentioned as the main advantages of this method (Iman and Davenport, 1982). The different steps to introduce a correlation structure (defined by a correlation matrix  $S$ ) to  $n$  sets of parameters with  $r$  dimensions (stored in matrix  $X$  with  $n$  rows and  $r$  columns), sampled using a sampling scheme (e.g. RS or LHS) from their corresponding marginal distributions can be explained as follows (Mildenhall, 2005):

- 1) Generate a  $n \times 1$  column vector of scores in which each score is randomly generated from a standard normal distribution (i.e. vector  $v$  with zero mean and standard deviation of 1)
- 2) Generate a  $n \times r$  matrix by copying  $v$  vector  $r$  times (matrix  $M$  ).
- 3) Randomly permute the elements in each column of matrix  $M$
- 4) Calculate the correlation matrix of  $M$  (i.e.  $E = n^{-1}M^T M$  and  $M^T$  is the transpose of  $M$  ).
- 5) Calculate matrix  $F$  using the Cholesky decomposition of matrix  $E$  (i.e.  $E = F^T F$  ).
- 6) Calculate matrix  $C$  using the Cholesky decomposition of matrix  $S$  (i.e.  $S = C^T C$  ).
- 7) Calculate matrix  $T$  as a function of  $M$ ,  $F$ , and  $C$  (i.e.  $T = MF^{-1}C$  ).
- 8) Re-arrange the elements corresponding to each column of matrix  $X$  so that each column has the same rank ordering as its corresponding column in matrix  $T$  .

### 3.3.3 Propagation of uncertainty and variability using Monte Carlo simulation

As mentioned at the beginning of this section, Monte Carlo simulation is used for propagation of uncertainty and calculation of PONC. Depending on the way the two different types of uncertainties (i.e. uncertainty due to lack of knowledge and inherent variability (see 2.4.2 for details)) are treated in uncertainty analysis, two types of Monte Carlo simulation can be applied (i.e. one dimensional Monte Carlo simulation and two-dimensional Monte Carlo simulation). In this chapter these two types of Monte Carlo simulations and also a new method based on the uncertainty analysis at two specific vectors of parameter values (i.e. one corresponding to the nominal parameter values and the other one to a worst-case vector of model parameters) are explained.

### ***I. One-dimensional Monte Carlo simulation***

In one-dimensional Monte Carlo simulation no distinction is made between uncertainty and variability. Therefore, if one-dimensional Monte Carlo simulation is selected to propagate the relevant sources of uncertainty/variability that affect the effluent distribution, the following steps should be followed to obtain the effluent distribution:

- 1) Random generation of a vector of uncertain model parameters (generated using random sampling from the marginal PDFs of uncertain model parameters).
- 2) Synthetic generation of a year-long influent time series (generated by the influent generator).
- 3) Running the dynamic simulation of a design alternative using the parameter values in 1) and the influent time series in 2).
- 4) Repeating 1-3  $N$  times until convergence of the effluent distribution (see Section 3.4.1 for details on the convergence test).

Because the two types of uncertainties are lumped together (uncertainty in model parameters and variability in the influent time series), a sample of uncertain model parameters (step (2) in Figure 3-6) and a year-long realization of influent time series (step (1) in Figure 3-6) are input simultaneously to the dynamic model of each design alternative in each Monte Carlo run.

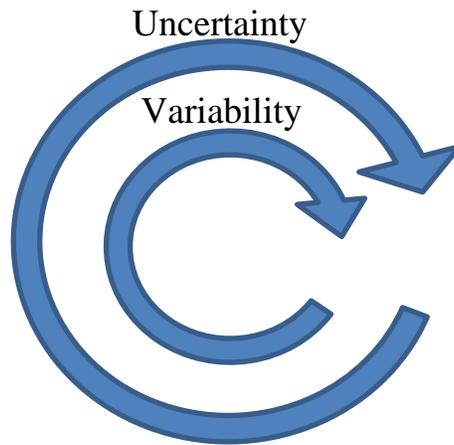
After termination of the Monte Carlo simulations, the  $N$  year-long effluent time series (typically after some time series aggregations, explained in 3.4) are mixed and a single effluent distribution is estimated based on  $N$  years of simulated effluent time series.

In general, one of the main problems that may arise by mixing the two different types of uncertainty is that the information regarding the contribution of each type of uncertainties is lost and the result may become technically difficult to interpret (Wu and Tsang, 2004). Therefore, it is recommended to separate the two sources of uncertainty unless the

contribution of uncertainty is negligible compared to that of variability, or the other way around (Frey, 1993). In other words, if a one-dimensional Monte Carlo simulation is to be used properly for uncertainty analysis, the effect of either uncertainty or variability must be negligible (Merz and Thielen, 2005).

## ***II. Two-dimensional Monte Carlo simulation***

In contrast to one-dimensional Monte Carlo simulation, two-dimensional Monte Carlo simulation is comprised of two loops (Figure 3-9) which allow variability and uncertainty to be considered separately (Frey and Rhodes, 1996).



**Figure 3-9** Schematic of a two-dimensional Monte Carlo simulations

First, a vector of model parameters is sampled from the marginal distribution of uncertain model parameters and then the performance of a design alternative is evaluated repeatedly using dynamic simulation with different realizations of influent time series generated using the explained influent generator (this is the only source of variability considered, see the introduction of Section 3.3). The number of simulations with a fixed vector of model parameters but different influent files continues until convergence is achieved for the different statistics of the effluent distribution (see Section 3.4 for details on convergence test). The same procedure is then repeated for different alternative vectors of uncertain

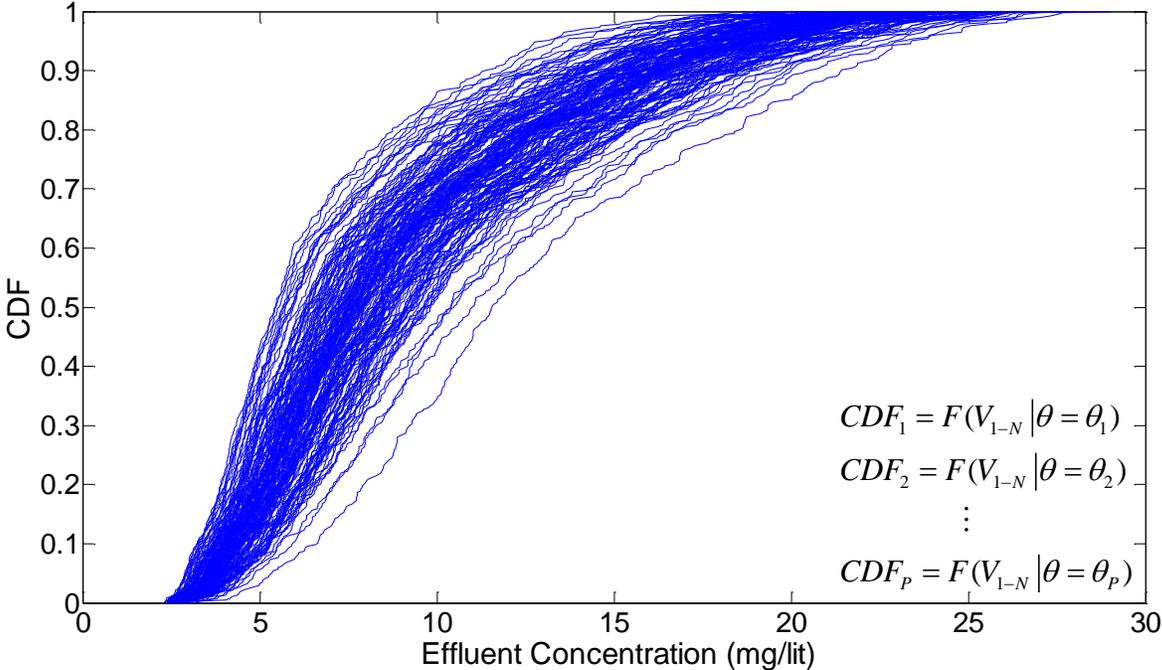
model parameters until convergence is achieved for the distribution of effluent in the uncertainty dimension (after  $P$  alternative sets of model parameters).

The implementation of a two-dimensional Monte Carlo simulation for calculating the effluent distribution can be summarized as follows:

- 1) Random generation of a vector of uncertain model parameters (generated using random sampling from the marginal probability distribution functions of uncertain model parameters)
  - a. Synthetic generation of a year-long influent time series (generated by the influent generator)
  - b. Running the dynamic simulation of a design alternative using the parameter values in 1) and the influent time series in a.
  - c. Repeating a-b until convergence of the effluent distribution (convergence in the variability dimension (i.e.  $N$  simulations))
  - d. End of variability loop
- 2) Going back to 1) and repeating a-d until convergence of the effluent distributions in uncertainty dimension (i.e.  $P$  times))
- 3) End of uncertainty loop

The result of the two-dimensional Monte Carlo simulation would be a cloud of CDF distributions (rather than a single CDF for one-dimensional Monte Carlo) comprising of different effluent CDF with different levels of confidence. Figure 3-10 shows a typical cloud of effluent CDF resulting from a two-dimensional Monte Carlo simulation. Overall there are  $P$  CDFs and each CDF has been generated by fixing the uncertain model parameters at a particular vector of parameter values (i.e.  $\theta = \theta_i$ ) and running the dynamic model of the plant under  $N$  different year-long realizations of the influent time series (i.e.  $V_{1-N}$ ).

It should be noted that each CDF in Figure 3-10 is calculated from the  $N$  year-long effluent time series (after some time series aggregations, explained in 3.4) calculated by running the dynamic model of the plant under  $N$  different year-long realizations of influent time series and a fixed vector of model parameters sampled from the distribution of model parameters. Therefore, the total number of simulations required to generate a cloud of CDFs equals the required number of influent realizations multiplied by the required number of sampled parameter vectors (i.e  $N \times P$ ).



**Figure 3-10** A cloud of effluent CDFs resulting from a two-dimensional Monte Carlo simulation

The results of a two-dimensional analysis could be used for identifying a set of actions to reduce the total uncertainty, determining the value of information, and assessing the performance of a system under a potential change or system modification (e.g. assessing the probabilistic performance of a plant by changing the dimensions of certain treatment units).

However, the two-dimensional Monte Carlo simulation is very computationally-expensive compared to one-dimensional Monte Carlo (Hoffman and Hammonds, 1994) and its application in the context of the current study with a real-world case study (almost 2 hours calculation of a 1 year simulation) was not feasible. The computational burden associated with two-dimensional Monte Carlo simulations could be alleviated by resorting to more efficient computing methods like cluster computing (Claeys et al., 2006; Benedetti et al., 2008), but this was beyond the scope of this study.

### ***III. Pragmatic Monte Carlo method***

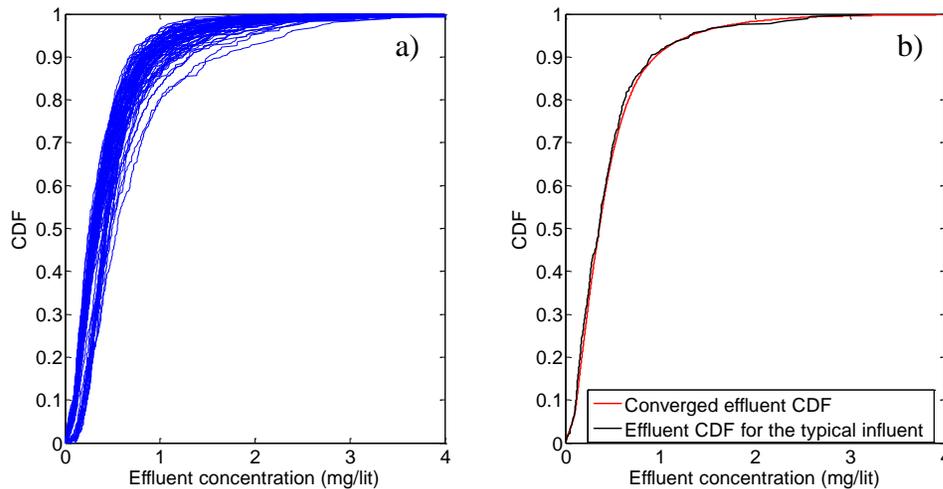
Given the shortcomings of the one-dimensional Monte Carlo simulation method and the huge computational load associated to the two-dimensional Monte Carlo method, a novel approach is proposed to calculate the PONC without mixing variability and uncertainty (as is the case for one-dimensional Monte Carlo) and with a reasonable number of simulations. The proposed method that is called the pragmatic Monte Carlo method is a type of worst-case analysis in which the effect of uncertainty is acknowledged but it is not modelled explicitly. The effect of uncertainty is considered by selecting parameter values in such a way that the calculated risk (in this study PONC) will be conservative (European Commission, 2002; Montgomery, 2009).

In the proposed pragmatic method the effect of parameter uncertainty of the model parameters is evaluated at two particular vectors of model parameters. One vector of model parameters corresponds to the nominal/calibrated values which were obtained from previous studies on the plant. The other vector of parameters represents a “worst case” scenario that would result in a conservative effluent CDF compared to other alternative vectors of parameters. The methodology for selecting a “worst case” vector of parameters is based on repetitive dynamic simulations of the plant model using one typical year-long influent time series and different vectors of model parameters, sampled from their PDFs.

The determination of a typical influent time series and the procedure for identifying a “worst case” vector of model parameters are explained below.

### Determination of a typical year-long influent time series

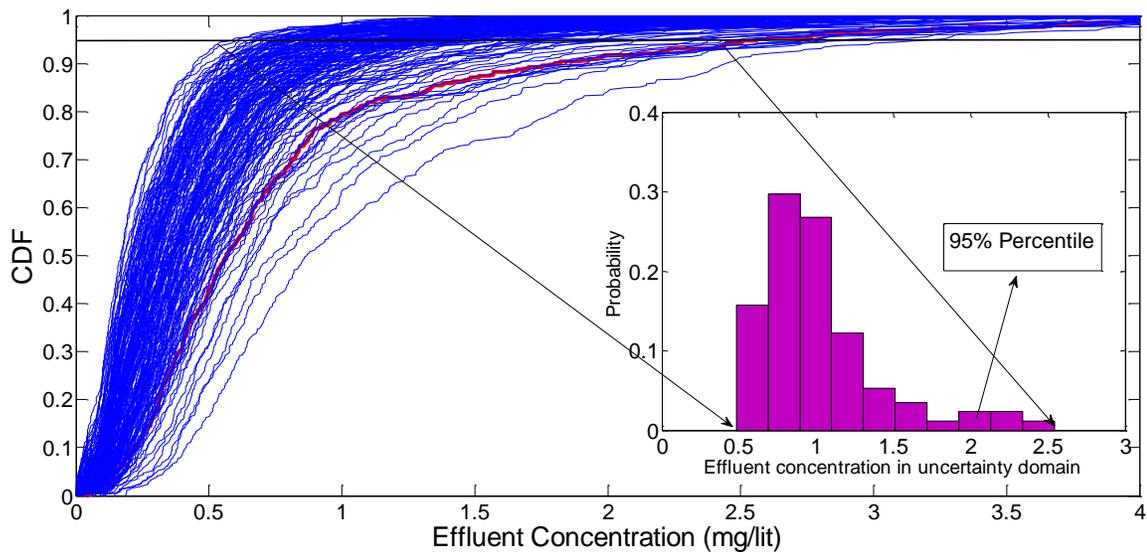
To identify a typical year-long influent time series, the parameters of the plant model are fixed at their nominal/calibrated values and the model is run repeatedly under different realizations of influent time series (generated using the influent generator presented in Chapter CHAPTER 4:). After termination of the simulations ( $N$  simulations until convergence of the effluent distribution, see Section 3.4 for convergence test), the converged CDF (the CDF in red, Figure 3-11 b) is compared with the individual CDFs calculated using different year-long influent time series (the CDFs in blue, Figure 3-11 a). The CDF which has the closest statistical properties to the converged CDF will be identified (the CDF in black, Figure 3-11 b) and its corresponding year-long influent time series is selected as the typical influent time series that will be used in the next step of the analysis. The closest CDF to the converged CDF is identified by calculating the sum of absolute differences between certain percentiles of each individual year-long CDF and the converged one (i.e. 5<sup>th</sup>, 50<sup>th</sup>, 95<sup>th</sup>, etc) and selecting the one with the smallest difference (the CDF in black, Figure 3-11 b).



**Figure 3-11** Identifying a typical year-long influent time series

Determination of a “worst case” vector of model parameters

A cloud of CDFs is generated by one-dimensional Monte Carlo simulation using the typical year-long influent time series obtained in the previous section and different parameter sets sampled at each Monte Carlo run (Figure 3-12). It should be noted that the CDF curves in Figure 3-12 are different than those obtained by running a two-dimensional Monte Carlo simulation (Figure 3-10). In the two-dimensional Monte Carlo simulation explained in the previous section, each CDF was generated by running the plant model for  $N$  number of year-long realizations of influent time series with model parameters fixed at a sampled vector of parameters, whereas in Figure 3-12, each CDF is generated by running the plant model for only a single year-long realization of influent time series with model parameters fixed at a sampled vector of parameters.



**Figure 3-12** CDF cloud generated for the effluent concentrations using a single year-long influent realization and different parameter values sample using a one-dimensional Monte Carlo simulation

After generating such a cloud of CDF graphs, a vector of model parameters representing a “worst case” can be identified by selecting a high level of confidence and identifying a CDF

that corresponds to that level of confidence given the effluent standard limits. For example, in Figure 3-12, if the effluent standard is 2mg/l and the level of confidence is 95%, the vector of parameters that has resulted in the CDF with 95% (i.e. the CDF in red) can be selected to identify the “worst case” vector of model parameters.

It should be noted that the 95% level of confidence for a specific effluent concentration depends on the distribution of effluent concentrations in the uncertainty dimension. Different effluent distributions in the uncertainty dimension can be calculated by drawing different (horizontal) lines perpendicular to the CDF axis, and extracting the intersect points of that line to the CDF curves. The effluent distribution in the uncertainty dimension in Figure 3-12 was constructed using the intersect points between the horizontal black line and the CDF curves. According to the constructed distribution, there is a 95% confidence that the effluent concentration would not exceed 2 mg/l. The CDF in red corresponds to the CDF from which the effluent concentration of 2 mg/l is derived and the vector of model parameters that has resulted to this CDF, constitutes the “worst case” vector of model parameters.

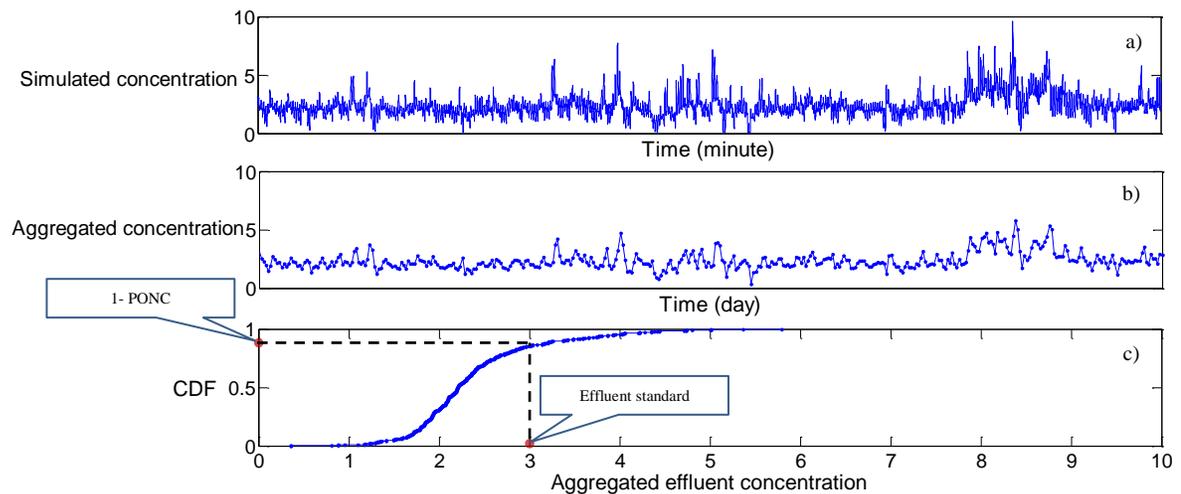
Once the two vectors of model parameters (one representing nominal values and the other one representing a “worst case”) are identified, two one-dimensional Monte Carlo simulations with fixed parameter values (one at the nominal/calibrated values and the other one at the “worst case” parameter values) and different realizations of influent time series will be performed to calculate two alternative values for PONC. This represents the pragmatic Monte Carlo method to explicitly deal with both variability and uncertainty.

### **3.4 Output analysis**

As mentioned in the previous section, the number of Monte Carlo simulations is determined by checking the convergence of the effluent distributions. Convergence of the effluent

distribution can be determined by checking some of the statistical properties (e.g. 5%, 50%, 95%) of the effluent distributions.

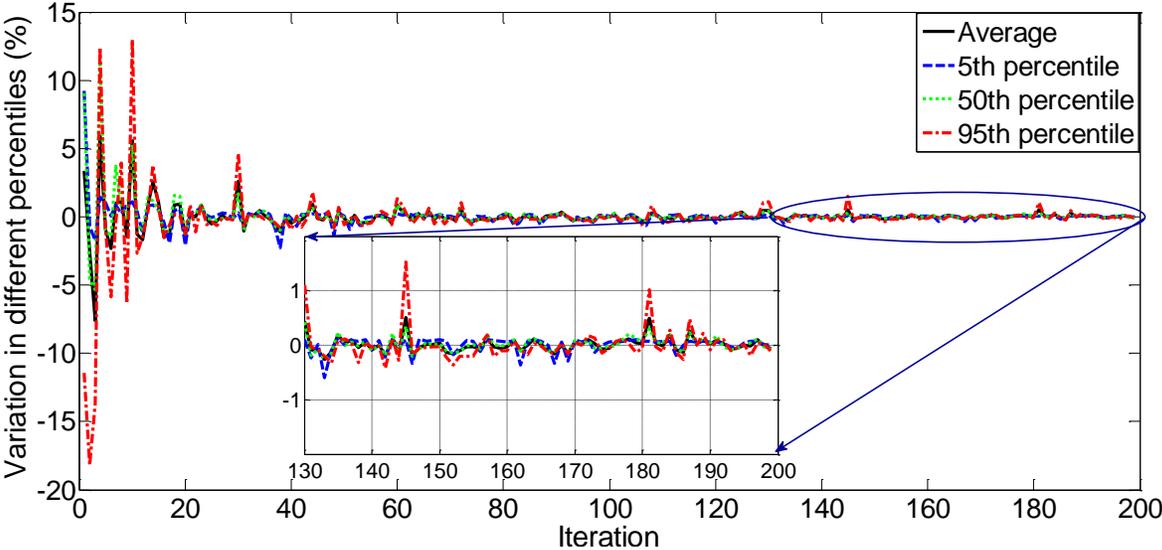
Prior to any convergence checking, the simulated effluent time series may require some time series aggregation before deriving any statistics from it. The need for time series aggregation arises when the temporal resolution of the simulated effluent time series is different than the one by which compliance to a specific effluent standard is measured. For example, if the simulated effluent time series have a temporal resolution of 15 minutes and the compliance to effluent standards is measured based on flow-proportional daily-aggregated concentration values, then the effluent time series (with 15 minutes temporal resolution) need to be aggregated to daily effluent time series (Figure 3-13). Once the simulated effluent time series is aggregated, the convergence of effluent distributions can be evaluated for the three types of uncertainty propagation methods explained in Section 3.3.3.



**Figure 3-13** Aggregation of the simulated concentration time series ( $\Delta t=15$ ) into daily values and calculation of PONC using the empirical CDF (bottom)

**3.4.1 Convergence test for one-dimensional Monte Carlo simulation**

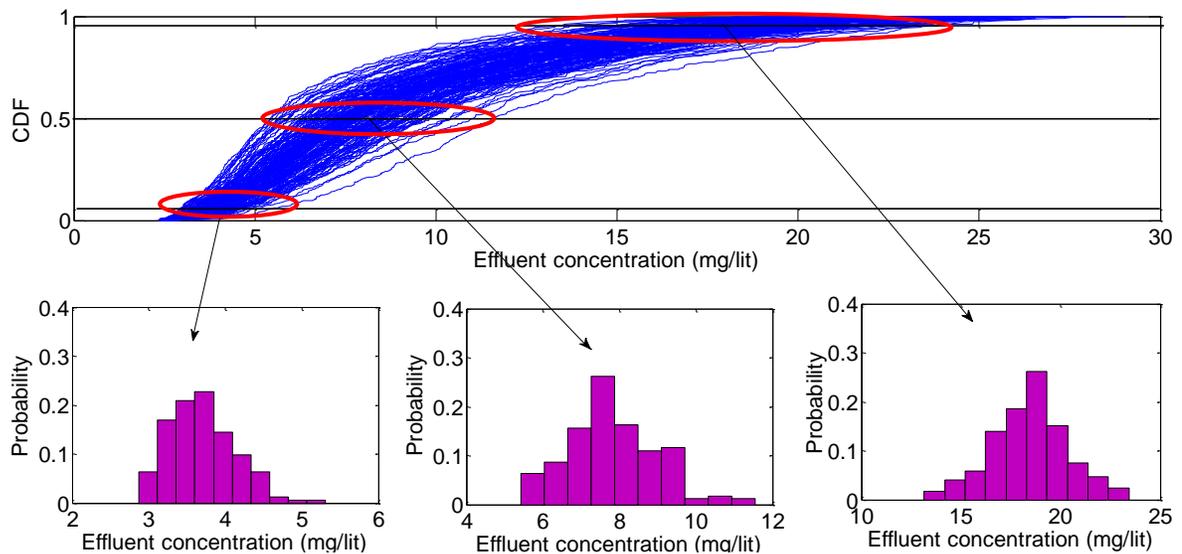
The convergence of the effluent distribution is evaluated by checking the variation in some of the statistical properties of the effluent data (i.e. 5<sup>th</sup>, 50<sup>th</sup>, and 95<sup>th</sup> percentiles), obtained after *N* Monte Carlo runs. The Monte Carlo simulations continue until the variation in some of the statistics of effluent distributions drops below a certain threshold. Figure 3-14 shows the percentage change (i.e. the relative difference between the value of a certain percentile, calculated in simulation *i* and *i+1*, expressed in percentage;) in the values of 5<sup>th</sup>, 50<sup>th</sup> and 95<sup>th</sup> percentiles as a function of Monte Carlo runs. As indicated, the percentage change for all statistics drops below 1% after 145 iterations (145 years of dynamic simulation).



**Figure 3-14** Percentage change in the values of the average, 5<sup>th</sup>, 50<sup>th</sup> and 95<sup>th</sup> percentiles as a function of Monte Carlo iterations

### 3.4.2 Convergence test for two-dimensional Monte Carlo simulation

The convergence of a two-dimensional Monte Carlo should be tested in both variability and uncertainty dimensions. Figure 3-15 shows a typical cloud of CDFs resulting from a two-dimensional Monte Carlo simulation. As explained in Section 3.3.3, each CDF curve is generated by fixing the model parameter values at a sampled vector of parameters and running the dynamic model of plant under  $N$  realizations of influent time series. The convergence of each CDF (i.e. convergence in the variability dimension) is evaluated in the same way as for a one-dimensional Monte Carlo simulation (Figure 3-14). To check the convergence in the uncertainty dimension, the distributions of effluent concentrations corresponding to the 5<sup>th</sup>, 50<sup>th</sup> and 95<sup>th</sup> percentiles are constructed (using the intersect points between the black lines and the CDFs in Figure 3-15) and a similar analyses depicted in Figure 3-14 is performed on to the effluent distributions of the 5<sup>th</sup>, 50<sup>th</sup> and 95<sup>th</sup> percentiles in the uncertainty dimension (the three distributions at the bottom of Figure 3-15).



**Figure 3-15** Convergence of a two-dimensional Monte Carlo simulation

## 3.5 Calculation of PONC and total cost of design alternatives

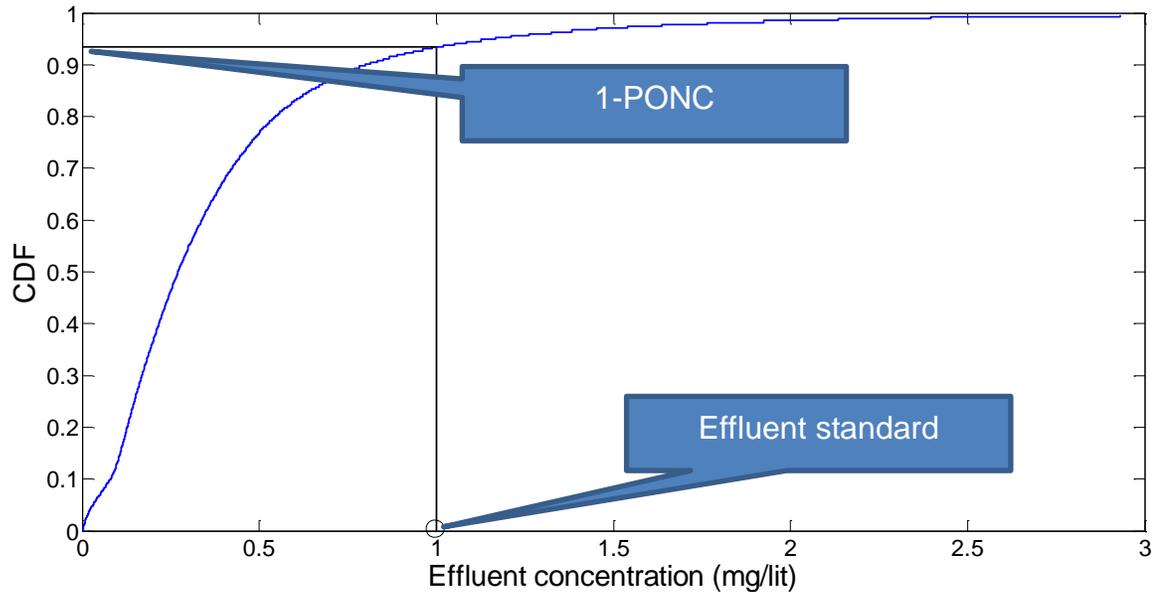
### 3.5.1 Calculation of PONC

In this study two types of analysis are performed for calculating the PONC corresponding to different design alternatives. The first scheme for propagation of uncertainty/variability is to perform a one-dimensional Monte Carlo simulation without distinguishing between the two types (see *I* in Section 3.3.3). The second scheme for calculation of PONC is according to the explained pragmatic Monte Carlo method (see *III* in Section 3.3.3 ) which is comprised of two one-dimensional simulations. The full two-dimensional method could not be applied because of the excessive computational load.

After the convergence of the effluent distribution, the PONC corresponding to a specific effluent standard can be calculated based on the effluent CDF. Figure 3-16 illustrates the calculation of PONC for an effluent standard of 1 mg/l, using the effluent CDF of a design alternative. As indicated, the CDF value corresponding to an effluent concentration of 1mg/l equals the compliment of the PONC. Therefore, the PONC in Figure 3-16 can be calculated by subtracting the CDF value corresponding to the effluent standard of 1mg/l (i.e. 0.92) from 1, which in this example would be around 0.08.

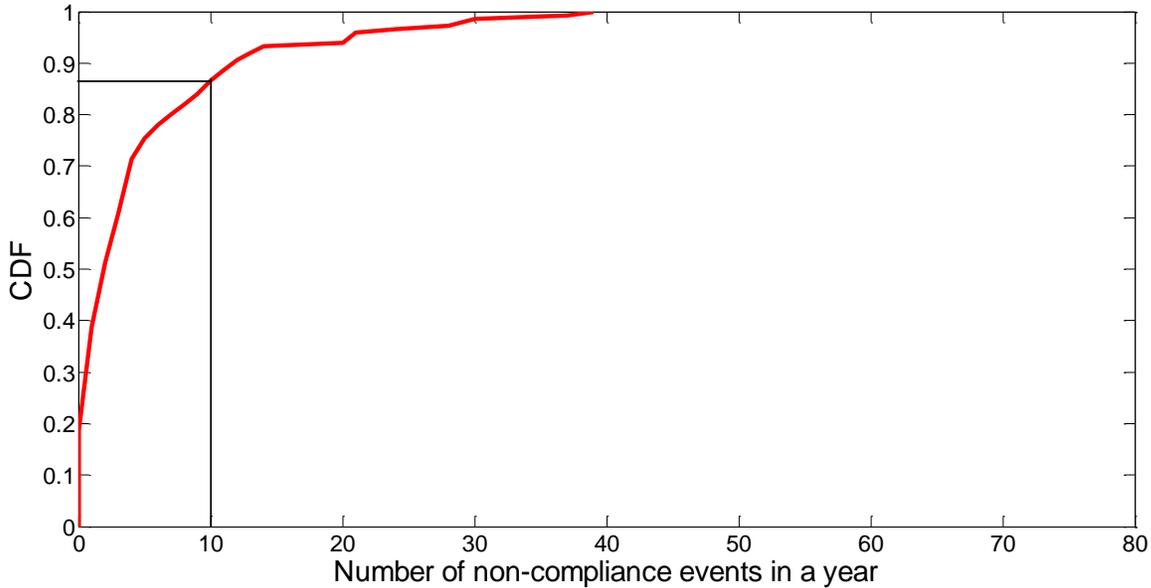
The explained PONC represents the expected ratio of non-compliance events to the total number of events. For example, if the effluent standards are measured based on daily concentration values (each day constitutes either a compliance or non-compliance event), the expected number of days in non-compliance with the effluent standard can be calculated by multiplying the calculated PONC with the total number of days in a year (i.e.  $0.08 \times 365 = 29$ ). However, the number of days in non-compliance could vary significantly for different years depending on the different realizations of the influent time series as well as the set of model parameters. For example, according to Figure 3-16 in which the expected

number of days in non-compliance is 29, it will be very likely to have some years with less than 29 and some years with more than 29 non-compliance events.



**Figure 3-16** Calculation of PONC corresponding to an effluent standard

In some effluent standards it is allowed to have a certain number of non-compliance events in a year and the designers might be interested in knowing the probability of having a certain number of events in a year rather than the expected number of non-compliance events. Calculating the probability of a specific number of days in non-compliance requires the estimation of a discrete probability distribution form the Monte Carlo simulation outputs. To this end, a CDF can be calculated based on the number of days in non-compliance that can be obtained after the termination of each year-long dynamic simulation ( $N$  numbers). Figure 3-17 illustrates a CDF for the number of non-compliance events that may occur in a year. As indicated, the probability of having equal or less than 10 non-compliance events in a year is 0.87. In other words, the probability of having more than 10 non-compliance event in a year equals 0.13.



**Figure 3-17** CDF for the number of non-compliance events in a year based on  $N$  years of simulation

### 3.5.2 Calculation of total cost

After calculation of the PONC values, the total cost associated with each design alternative is to be calculated and presented to decision makers. Calculating the cost of a WWTP is a complex task as the available cost functions have usually been developed for a specific region and reflect the planning and building procedures that are used in that specific region (Bode and Lemmel, 2001; Benedetti et al., 2006b). At this stage of analysis, designers may use specific cost functions suitable for the region in which a plant is to be built. For example, a company may have already developed their in-house cost functions that are based on regional conditions and previous experience in building WWTPs and use them for estimating the total cost. Therefore, there is no limitation on the type of cost functions that can be used for calculating the total cost, as long as the time and location-specific factors affecting the total cost are taken into account (see 5.2.3 for the cost analysis tool selected for the current case study).

At the end, the calculated total cost values associated with different design alternatives can be plotted against their corresponding PONC to help designers identify the design alternatives that have a reasonable cost while meeting the effluent standards with a tolerable PONC.

It should be noted that the scale of the plant is an important factor that needs to be considered when applying cost functions, as some cost calculations cannot be directly related to process unit dimensions, but rather as a percentage of the construction costs. For example, an administration or laboratory costs calculated as a fixed percentage of the construction cost (which in turn is calculated as a function of global plant variables, e.g. flow rate, through some power law relationships) are only valid for certain plant sizes. Therefore, these costs could be underestimated for small plants in which the ratio of certain costs to the construction cost is higher than their corresponding values in larger plants.



# CHAPTER 4: INFLUENT GENERATOR FOR PROBABILISTIC MODELING OF NUTRIENT REMOVAL WASTEWATER TREATMENT PLANTS

Submitted to the journal of Environmental Modelling and Software<sup>1</sup>

## Abstract

The availability of influent wastewater time series is crucial when using models to assess the performance of a wastewater treatment plant (WWTP) under dynamic flow and loading conditions. Given the difficulty of collecting sufficient data, synthetic generation could be the only option. In this paper a hybrid of statistical (a Markov chain-gamma model for stochastic generation of rainfall and two different multivariate autoregressive models for stochastic generation of air temperature and influent time series in dry conditions) and conceptual modeling techniques is proposed for synthetic generation of influent time series. The time series of rainfall and influent in dry weather conditions are generated using two types of statistical models. These two time series serve as inputs to a conceptual sewer model for generation of influent time series. The application of the proposed influent generator to the Eindhoven WWTP shows that it is a powerful tool for realistic generation of influent time series and is well-suited for probabilistic design of WWTPs as it considers both the effect of input variability and total model uncertainty.

---

<sup>1</sup> Authors: Mansour Talebizadeh<sup>1\*</sup>, Evangelia Belia<sup>2</sup>, and Peter A. Vanrolleghem<sup>1</sup>

<sup>1</sup>modelEAU, Département de génie civil et de génie des eaux, Université Laval, 1065 av. de la Médecine, Québec (QC) G1V 0A6, Canada (Email : mansour.talebizadehsardari.1@ulaval.ca; Peter.Vanrolleghem@gci.ulaval.ca)

<sup>2</sup>Primodal Inc., 145 Aberdeen, Québec, QC G1R 2C9, Canada (Email: belia@primodal.com)

**Keywords** Bayesian estimation; probabilistic design; uncertainty analysis; urban hydrology; wastewater composition; weather generator

## 4.1 Introduction

One of the major sources of uncertainty/variability that both plant designers and operators must deal with is the dynamics of the influent (Belia et al., 2009). The recent advances in mathematical modeling and improved computational power have enabled researchers to better understand the performance of different WWTP design alternatives (Hao et al., 2001; Salem et al., 2002; Hyland et al. 2012) and/or evaluate control strategies under dynamic flow and loading conditions. However, the application of mathematical models used for simulating the performance of a WWTP could be misleading unless, among other reasons, models are provided with representative influent time series. One of the problems that arise in this regard is the scarcity or even lack of long-term influent data. To remedy this problem, some researchers have proposed models for synthetic dynamic influent time series scenarios (Bechmann et al., 1999; Gernaey et al., 2011; Martin and Vanrolleghem, 2014). The development of a tool capable of generating dynamic influent time series that is representative of the climate and characteristics of the sewershed could have several applications. Synthetically-generated influent time series can serve as input to a dynamic model of a plant for checking the performance of different configurations, sizings, as well as devising an optimum control strategy regarding the treatment of wastewater (Benedetti et al., 2006a; Devisscher et al., 2006; Guerrero et al., 2011; Ciggin et al., 2012). In addition, realistic generation of different realizations of influent time series is one of the most important component of studies that take into account the issue of uncertainty in design and operation of WWTPs (Rousseau et al., 2001; Bixio et al., 2002; Martin et al., 2012; Talebizadeh et al., 2014).

Various approaches have been adopted by different researchers for influent generation (for a review see Martin and Vanrolleghem (2014)). One of the simplest approaches in synthetic

generation of influent time series is the application of stochastic or regression models with or without periodic components (Capodaglio et al., 1990; Martin et al., 2007; Rodríguez et al., 2013). However, these models may have a poor performance especially during wet weather flow conditions as different complex processes affect the dynamics of the influent. Indeed, such statistical models do not consider the underlying processes that govern the generation and the dynamics of the influent. Langeveld et al. (2014) proposed an empirical model for predicting the influent pollutant time series as a function of influent flow for both dry and wet weather flow conditions (simulation of pollutant time series as a function of flow was also adopted by Rousseau et al. (2001) and Bixio et al (2002)). Although the proposed model could be used for prediction of pollutant influent time series, it requires a stochastic input (i.e. influent flow) generator if it is to be used for generating different realizations of influent time series.

To consider the underlying phenomena, some researchers have advocated the use of detailed conceptual and/or physically-based models (Hernebring et al., 2002; Temprano et al., 2007). The application of these complex models might be useful for certain purposes, (e.g. evaluating the performance of different operating strategies in a sewer system). However, in cases in which the overall behavior of the influent time series is of interest (i.e. the overall variation of influent time series, not all the different phenomena that have resulted in that time series), they might not be very useful as they require very detailed information on the sewer system and running them for a large number of times could be computationally expensive. In addition, even if a detailed sewershed model proves to have a good performance regarding the simulation of the influent time series under a given set of inputs, it cannot be called an influent generator unless a procedure is available for the generation of different realizations of stochastic inputs (e.g. rainfall time series, wastewater time series in dry weather flow (DWF) conditions).

Some researchers have proposed parsimonious conceptual models as an alternative to the complex mathematical models that require detailed information and data (Gernaey et al., 2011). In these models a conceptual view of the main phenomena and interactive processes are formulated in terms of mathematical equations. Despite successful application of these models (at least in giving an overall view of the system), the performance of these models to a great extent depends on the proper choice of model parameters. Since some of the parameters may not have a clear physical meaning they are usually estimated through model calibration. In cases in which there is no measured data available for model calibration, only a rough estimate or a range of values could be inferred from the values reported in literature. In addition, it is almost impossible to have a complete similarity between the model output(s) and the observed values owing to the inextricable uncertainties (e.g. input data uncertainty and/or model structure uncertainty) in any modeling exercise (Belia et al., 2009; Freni and Mannina, 2010).

Given the importance of the issue of uncertainty, several studies have been conducted that consider its effect on both water quality and quantity prediction in urban drainage modeling (Freni et al., 2009; Dotto et al., 2012). However, in these studies, only the effect of model uncertainty under a set of historical rain events (wet weather flow, WWF, conditions) has been considered (i.e. the time series of rainfall and also the contribution of wastewater in DWF conditions were assumed known a priori). In this study not only are we interested in the effects of model uncertainty, but also in the variability of rainfall and influent time series in DWF conditions which significantly affect both the amount and the dynamics of the influent loads.

Considering the shortcomings of the previous studies, this study aims to develop an influent generator which is capable of producing dynamic influent time series of flow and traditional wastewater component concentrations (TSS, COD, TN, TP, NH<sub>4</sub>) with 15-min temporal resolution (15-minutes temporal resolution was assumed to be enough for capturing sub-

daily time variations of the influent which could affect the operating parameters and the performance of WWTPs). The proposed methodology will enable users to generate dynamic influent time series that have the same statistical properties as the observed ones using a set of statistical and conceptual modeling tools that only require basic information on climate and characteristics of the sewershed under study. It should be noted that the proposed influent generator is capable of considering the effect of uncertainty in model parameters on the generated influent time series whether the uncertainty can be reduced using observed data (e.g. for the current study) or not (uncertainty in model parameters is characterized by a range of values, determined through expert elicitation or the data from similar sewersheds). In the current study, the variability in inputs (captured by generating different realizations of rainfall and influent time series in DWF conditions, explained in Section 4.2.1 and Section 4.2.2, respectively) as well as the uncertainty in model parameters (explained in Section 4.2.4) on the generated dynamic influent time series are other important issues that will be covered.

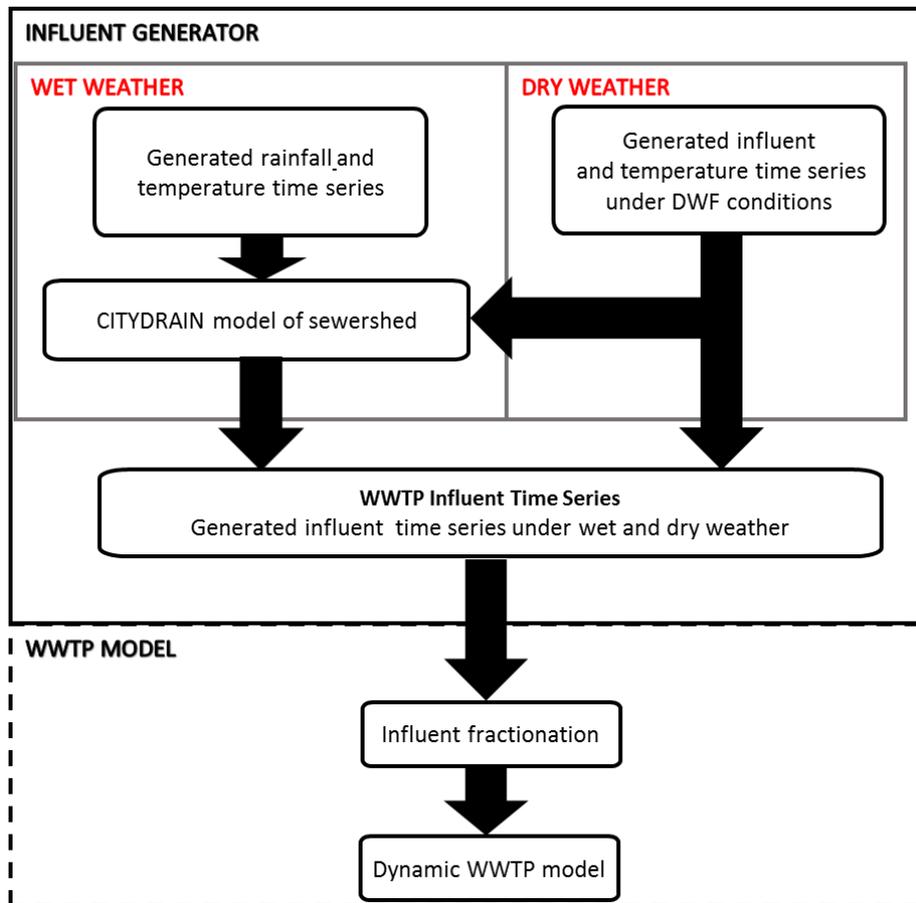
## **4.2 Methodology**

In this paper, a hybrid of statistical and conceptual modeling tools is proposed for synthetic generation of influent time series considering both model parameter uncertainty and input variability. A two-state Markov chain-gamma model (Richardson, 1981) in conjunction with two time series disaggregation methods were used for the stochastic generation of rainfall time series with a high temporal resolution (i.e. 15-minute). To generate the influent time series during DWF conditions, taking into account the daily periodic variation, auto-correlation, and cross-correlation in time, a multivariate time series models was developed and its parameters were estimated using the methodology proposed by Neumaier and Schneider (2001). The proposed stochastic model is expected to be superior compared to previous attempts in the generation of influent, as in previous studies the diurnal variation of the influent under DWF conditions was modeled using only univariate time series

models (Martin et al., 2007), or by multiplying the daily average influent values to a set of coefficients representing the normalized dynamics of the influent at different times of a day with or without addition of a noise term to the generated time series (Achleitner et al., 2007; Langergraber et al., 2008; Gernaey et al., 2011). The problem resulting from the application of univariate time series models for synthetic generation of influent time series is that the cross-correlation structure that exists among different wastewater constituents may not be respected, as the different constituents are generated independently from the others.

In DWF conditions, the influent time series is generated using multivariate time series models. Whereas in WWF conditions, the outputs of the two statistical models used for the generation of the rainfall and influent time series in DWF conditions are input to a conceptual model for modeling the influent time series in WWF conditions (Figure 4-1). In this study the CITYDRAIN model (Achleitner et al., 2007) was selected as the conceptual model owing to its flexibility and parsimony. The CITYDRAIN model of the sewershed under study is calibrated using measured influent data through a Bayesian calibration procedure to account for the total model uncertainty (uncertainty stemming from both model parameters and the distribution of error, i.e. the difference between the observed and simulated time series).

Finally, different realizations of the influent time series can be generated by running the calibrated CITYDRAIN model using a realization of a generated time series of rainfall and a realization of influent under DWF conditions (i.e. the two stochastic input time series).



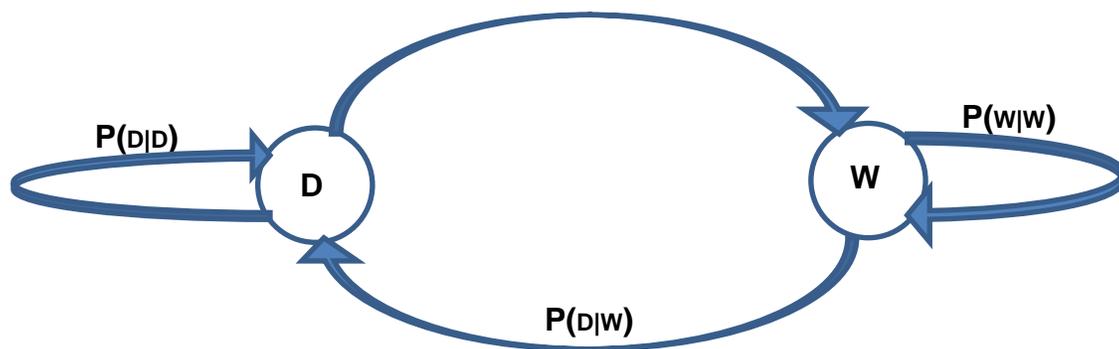
**Figure 4-1** Schematic of the proposed influent generator

Depending on the model to be used for modelling the processes inside the WWTP, an influent fractionation block must be added to convert the generated traditional wastewater composition (COD, TSS, etc) into the state variables of the selected biological models, e.g. the ASM models. Therefore, influent fractionation should be one of the components of WWTP modeling, not a component of the influent generator as different WWTP models may have different state variables.

## 4.2.1 Weather and air temperature generators

### I. Daily Weather generator

Realistic generation of rainfall time series is crucial as it is one of the most important factors that affect the dynamics of the influent. In this study a stochastic model proposed by Richardson (1981) was used for the synthetic generation of daily rainfall and air temperature time series. According to this method the sequence of dry and wet days is generated using a two-state Markov chain model with parameters  $P(W|W)$  and  $P(W|D)$  which represent the probability of having a wet day at day  $t$  given a wet day at day  $t-1$  and the probability of having a wet day at time  $t$  given a dry day at time  $t-1$  respectively (Figure 4-2).



**Figure 4-2** Schematic of a two-state Markov chain with the two states being wet (W) or dry (D) and four transitions between them

The other two parameters of the transition matrix needed for the generation of dry and wet days (i.e.  $P(D|D)$  the probability of having a dry day at day  $t$  given a dry day at day  $t-1$  and  $P(D|W)$  the probability of having a dry day at day  $t$  given a wet day at day  $t-1$ ) can be calculated using Equation 4-1 and Equation 4-2.

$$P(D|D) = 1 - P(W|D) \quad \text{Equation 4-1}$$

$$P(D|W) = 1 - P(W|W) \quad \text{Equation 4-2}$$

Once the transition probabilities have been estimated from climate data, the sequence of wet and dry days can be generated and the amount of rainfall in each wet day is generated by sampling from a gamma probability distribution (Equation 4-3) where  $x$  is the depth of daily rainfall,  $\alpha$  and  $\beta$  are the two parameters of the distribution (estimated from the measured rainfall time series), and  $\Gamma(\alpha)$  represents the gamma function evaluated at  $\alpha$ .

$$f(x) = \frac{(x/\beta)^{\alpha-1} \exp(-x/\beta)}{\beta \Gamma(\alpha)} \quad \text{Equation 4-3}$$

It should be noted that the seasonal variation of the daily rainfall generator parameters (Markov chain transition probabilities i.e.  $P(W|D)$  and  $P(W|W)$ ) as well as the parameters of the gamma distribution (i.e.  $\alpha$  and  $\beta$ ) were taken into account by fitting different Fourier series models on the parameter values derived from rainfall records. To do so, each year with rainfall records was divided into 26 two-week time spans and then the transition probabilities were estimated by dividing the number of wet days preceded by a dry day by the total number of days (for estimating  $P(W|D)$ ) and also dividing the number of wet days preceded by a wet day by the total number of days. Moreover, the parameters of the gamma distributions were calculated for each two-week time span using the maximum likelihood method. Once the parameters of the Markov chain-gamma model are estimated for each two-week time span, different Fourier series are fitted on the estimated parameters to provide a value for each day of the year.

The time series of minimum and maximum air temperature are generated conditioned on the state of the day (i.e. wet or dry) using a multivariate linear first-order time series model (Matalas, 1967). Starting points are a time series of daily maximum and minimum temperature values.

The seasonal variation in mean and standard deviation of maximum and minimum daily temperature values for dry and wet days are captured (in two Fourier series models) and subtracted from the data to derive the residual time series of maximum and minimum temperature (Equation 4-4 and Equation 4-5).

$$Y_i^d(j) = \frac{X_i^d(j) - \bar{X}_i^d(j)}{\sigma_i^d(j)} \quad \text{for dry days} \quad \text{Equation 4-4}$$

$$Y_i^w(j) = \frac{X_i^w(j) - \bar{X}_i^w(j)}{\sigma_i^w(j)} \quad \text{for wet days} \quad \text{Equation 4-5}$$

In the above equations  $\bar{X}_i^d(j)$  and  $\sigma_i^d(j)$  are the mean and standard deviation for a dry day,  $\bar{X}_i^w(j)$  and  $\sigma_i^w(j)$  are the mean and standard deviation for a wet day, and  $Y_i(j)$  is the residual component for transformed variables (i.e.  $j=1$  for maximum temperature, and  $j=2$  for minimum temperature). Once the residual time series are derived, a multivariate time series model as proposed by Matalas (1967) is fitted on the residual time series of maximum and minimum temperatures (Equation 4-6).

$$Y_i(j) = AY_{i-1}(j) + B\varepsilon_i(j) \quad \text{Equation 4-6}$$

In the above equation  $Y_i(j)$  is a  $2 \times 1$  matrix for day  $i$  whose elements are residuals of maximum temperature ( $j=1$ ) and minimum temperature ( $j=2$ ).  $Y_{i-1}(j)$  is a  $2 \times 1$  matrix of the previous day's residuals,  $\varepsilon_i$  is a  $2 \times 1$  matrix of independent random components (the noise term is assumed to be a normal, independent, and identically distributed (i.i.d) variable with zero mean and unit variance), and  $A$  and  $B$  are  $2 \times 2$  matrices whose elements are derived according to (Equation 4-7) and (Equation 4-8):

$$A = M_1.M_0^{-1} \quad \text{Equation 4-7}$$

$$B.B^T = M_0 - M_1.M_0^{-1}.M_1^T \quad \text{Equation 4-8}$$

where the subscript -1 denotes the inverse of matrix and  $M_0$  and  $M_1$  are defined as:

$$M_0 = \begin{bmatrix} 1 & \rho_0(1,2) \\ \rho_0(1,2) & 1 \end{bmatrix} \quad \text{Equation 4-9}$$

$$M_1 = \begin{bmatrix} \rho_1(1,1) & \rho_1(1,2) \\ \rho_1(2,1) & \rho_1(2,2) \end{bmatrix} \quad \text{Equation 4-10}$$

where  $\rho_0(j,k)$  is the correlation coefficient between variables  $j$  and  $k$  on the same day where  $j$  and  $k$  may be set to 1 (maximum temperature) or 2 (minimum temperature).  $\rho_1(j,k)$  is the correlation coefficient between variable  $j$  and  $k$  lagged one day with respect to variable  $j$ .

The above weather generator is suited for random generation of daily rainfall and temperature. However, in this study we need to generate rainfall time series with a finer temporal resolution than daily resolution (15-min temporal resolution, comparable to the temporal resolution of rainfall in the BSM influent model (Gernaey et al., 2011)). Some methodologies have been proposed for random generation of hourly rainfall time series based on historical hourly rainfall data (Pattison, 1965; Rodriguez-Iturbe et al., 1987). However, long-term hourly rainfall data may not be available in every region and using a limited hourly rainfall record for random generation of long-term hourly rainfall time series may result in misrepresentation of the inter-annual variability in rainfall.

In this study the proposed Richardson-based weather generator (used for daily rainfall generation) was combined with two time series disaggregation techniques. Daily rainfall time series is first generated using the Richardson (1981) method and then two time series disaggregation models, including a daily-to-hourly model (Koutsoyiannis and Onof, 2001)

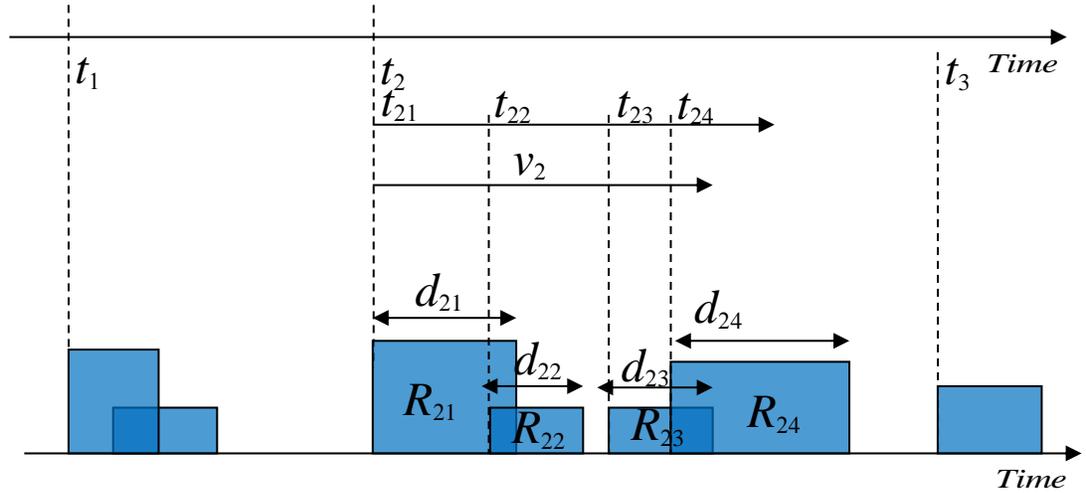
and an hourly-to-15-minutes model (Ormsbee, 1989) are applied for generation of long-term rainfall time series with 15-minute temporal resolution.

## ***II. Daily to hourly rainfall time series disaggregation***

The time series disaggregation method proposed by Koutsoyiannis and Onof (2001) is used in this study to disaggregate the synthetic daily rainfall time series (i.e. generated using the previously explained Richardson-based weather generator) into hourly rainfall. The proposed disaggregation method combines the Bartlett-Lewis stochastic rainfall model (Rodriguez-Iturbe et al., 1987; Rodriguez-Iturbe et al., 1988) with an adjusting algorithm so that the total amount of hourly rainfall in each day becomes consistent with its corresponding daily value. A general description of the Bartlett-Lewis model can be summarized as the following (Koutsoyiannis and Onof, 2001):

- 1) Storm origins ( $t_1, t_2, t_3$  in Figure 4-3) occur according to a Poisson process with rate  $\lambda$ .
- 2) Arrival times of the contributing cells in a storm ( $t_{21}, t_{22}, t_{23}$  for Storm2 in Figure 4-3) occur according to a Poisson process with rate  $\beta$ .
- 3) Cell arrival terminates after time  $v_1$  ( $v_2$  for Storm2 in Figure 4-3) that is exponentially distributed with parameter  $\gamma$ .
- 4) Each cell has a duration that is exponentially distributed with parameter  $\eta$ .
- 5) Each cell has a uniform intensity ( $R_{21}, R_{22}, R_{23}, R_{24}$  for Storm2 in Figure 4-3) coming from an exponential distribution  $\mu$ .

The parameters of the Bartlett-Lewis rainfall model can be calculated from hourly rainfall records (Rodriguez-Iturbe et al., 1987; Rodriguez-Iturbe et al., 1988; Koutsoyiannis and Onof, 2001) and then the model can be used for synthetic generation of hourly rainfall time series. However, the hourly rainfall time series generated using the Bartlett-Lewis model should be adjusted so that the sum of hourly rainfall time series in each day becomes consistent with its corresponding daily value.



**Figure 4-3** Schematic of Bartlett-Lewis rainfall model

In the proposed methodology, a simple adjusting procedure known as the proportional adjusting (Koutsoyiannis and Onof, 200) is used for adjusting the hourly rainfall time series. According to this procedure the initially generated hourly rainfall values ( $X_s$ ) are adjusted to new values ( $X'_s$ ) using Equation 4-11.

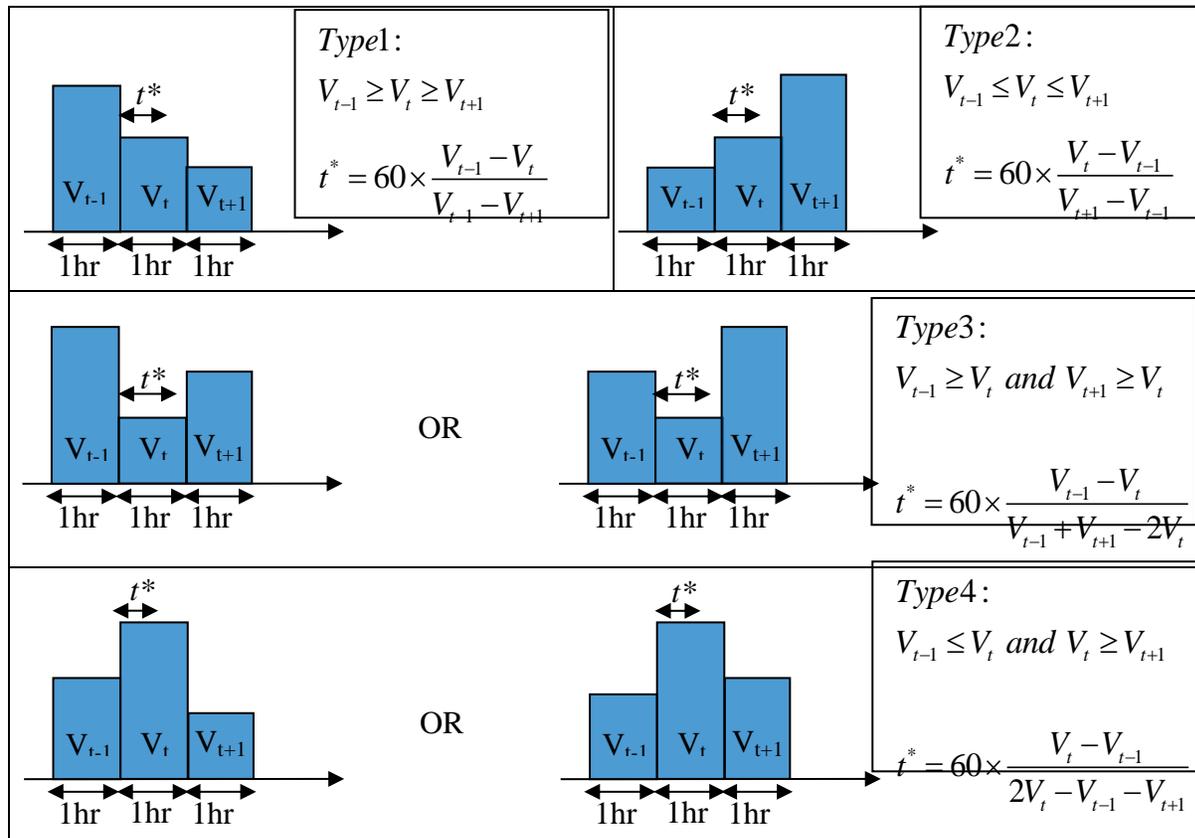
$$X'_s = X_s \left( \frac{Z}{\sum_{j=1}^k X_j} \right) \quad (s = 1, 2, 3, \dots, k) \quad \text{Equation 4-11}$$

where  $Z$  is the amount of daily rainfall (generated using explained Richardson-based rainfall generator), and  $k$  is the number of hourly rainfall values within a day.

### **III. Hourly to 15-minute rainfall time series disaggregation**

The disaggregated hourly rainfall time series in the previous section is further disaggregated to 15-minute rainfall time series using the empirical time series disaggregation procedure proposed by Ormsbee (1989). According to this empirical model, four types of rainfall patterns are identified (Figure 4-4) for each 3-hour sequence of hourly rainfall. After determining the type of sequence, the amount of rainfall at the central hour of each 3-hour

rainfall sequences is disaggregated into a time series with a desired temporal resolution (15 minutes here).



**Figure 4-4** Four types of rainfall patterns for a three-hour rainfall sequence (Ormsbee (1989))

The probability distribution function,  $F(t)$  of sub-hourly rainfall intervals of the central hour rainfall ( $V_t$ ) is calculated using the time parameter  $t^*$ , and the amounts of rainfall in the first and third hour ( $V_{t-1}$  and  $V_{t+1}$ ) of each 3-hour rainfall sequence (Equation 4-12).

$$F(t) = \begin{cases} \frac{V_{t-1}t}{V_t^*} - \frac{(V_{t-1}-V_t)t^2}{2V_t^*t^*} & \text{for } 0 \leq t < t^* \\ \frac{(V_t+V_{t-1})t^*}{2V_t^*} + \frac{V_t(t-t^*)t^2}{V_t^*} - \frac{(V_t-V_{t+1})(t-t^*)^2}{2V_t^*(60-t^*)} & \text{for } t^* \leq t \leq 60 \end{cases} \quad \text{Equation 4-12}$$

For the proposed methodology the central hour rainfall ( $V_t$ ) is disaggregated into 4 intervals and the portion of each interval is calculated by multiplying the probability of each interval to the central hour rainfall.

#### ***IV. Bioreactor temperature***

The explained Richardson-based weather generator is suited for the generation of daily air temperature which could serve as an input for modeling the temperature effect of the influent or in the bioreactors of a WWTP. Bioreactor temperature is of particular interest as it affects the rate of many biological processes taking place in the bioreactors (Antoniou et al., 1990). Whereas, based on the influent temperature a heat balance could be constructed around the bioreactors of the WWTP to calculate the bioreactor temperature (see Gillot and Vanrolleghem (2003) for details), preference is often given to directly input the bioreactor temperature in the WWTP model (e.g. Gernaey et al. (2014)).

To estimate the bioreactor temperature a simple linear regression model between the concurrently measured daily air and bioreactor temperatures is used. Once the parameters of the linear regression model (which calculates the daily bioreactor temperature as a function of daily air temperature) have been estimated, it can be used to convert the generated daily air temperatures (generated using the Richardson-based weather generator) to daily bioreactor temperatures. It should be noted that the variation in bioreactor temperature is not solely function of air temperature (Gillot and Vanrolleghem, 2003). However, calculating the daily bioreactor temperature as a function of daily air temperature would capture the seasonal variation of bioreactor temperature time series.

To further disaggregate the daily bioreactor temperature time series into a time series with 15-minute temporal resolution, the average normalized pattern representing the diurnal variation of bioreactor temperature is estimated by a Fourier series and multiplied to the daily bioreactor temperature values to obtain a bioreactor temperature time series with 15-minute temporal resolution.

#### ***4.2.2 Influent generation in DWF conditions***

The influent time series in DWF conditions usually shows specific periodic patterns which can be mainly attributed to the socio-economic fabric of society and also to the physical characteristics of the wastewater collection system. To mimic these variations in time, it is common practice to estimate representative values (e.g. multiplying flow per person to the total population for estimating flow) for flow and loads and then multiply them to a set of normalized coefficients reflecting diurnal, weekly and seasonal time variation of the influent time series (Gernaey et al., 2011; Flores-Alsina et al., 2014). Moreover, Gernaey et al. (2011) proposed to add a noise term to the deterministic influent profile in order to avoid generating the same influent time series in subsequent days. In this study, the effect of rainfall on the contribution of infiltration (rainfall induced infiltration) is not considered explicitly. Rather, the application of a multivariate auto-regressive model (Neumaier and Schneider, 2001) with periodic components is proposed.

To estimate the parameters of the proposed time series model, an influent time series measured during DWF conditions is to be extracted and analyzed for estimating the parameters of the multivariate auto-regressive model. First, the seasonal (e.g. associated to groundwater infiltration) and diurnal periodic components of flow and other wastewater constituents are to be estimated using different Fourier series approximations (depending on the underlying expected periodic patterns, e.g. a bimodal periodic pattern for flow in urban sewersheds) and removed from the original influent time series to calculate the residual

time series. The zero-mean residual time series of influent flow and composition is then further standardized to have an influent time series with a zero mean and unit standard deviation. The parameters of the multivariate autoregressive model in Equation 4-13 (i.e.  $p, A_l, C$ ) are then estimated through a stepwise least square algorithm proposed by Neumaier and Schneider (2001).

$$v_t = \sum_{l=1}^p A_l \times v_{t-l} + \varepsilon_t \quad \text{Equation 4-13}$$

In Equation 4-13,  $v_t$  is an m-dimensional vector (i.e. for our application m=5 which corresponds to the flow and the four wastewater compositions) containing the generated influent component at time  $t$ ,  $p$  is the order of the auto-regressive model ( $p$  is to be selected based on Schwarz's (1978) Bayesian Criterion SBC, based on the fitting results). More details can be found in Neumaier and Schneider (2001),  $A_1, \dots, A_p$  are the coefficient matrices of the auto-regressive model, and  $\varepsilon_t$  is a noise term generated from an uncorrelated zero-mean multivariate normal distribution with covariance matrix  $C$  (i.e.  $\varepsilon_t \sim N(0, C)$ ). Different realizations of the residual influent time series can be generated using this time series model and converted to the original scale depending on the mean and standard deviation of the original influent time series.

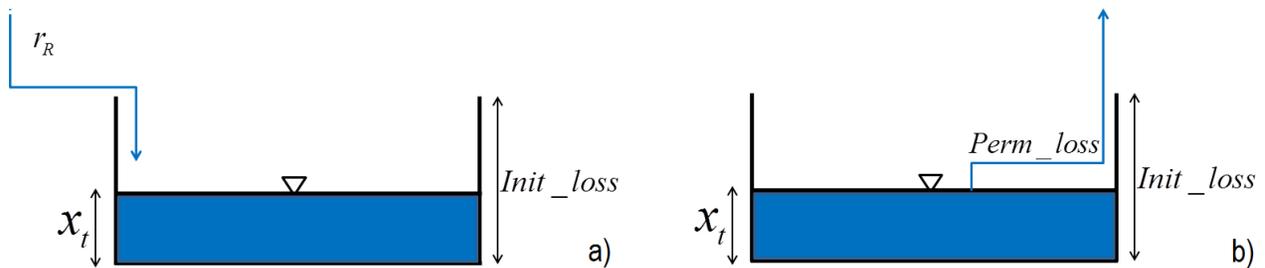
### 4.2.3 *Influent generation in WWF conditions*

Synthetic generation of the influent time series during WWF conditions is relatively more complicated than the generation of the influent time series during DWF conditions. Difficulties arise as various phenomena are occurring during WWF conditions and as the availability of measured data is usually scarce for these periods. Hence, using a purely statistical model may result in significant discrepancies between simulated and observed time series. Therefore, we used a combination of statistical modeling techniques and a conceptual model to generate the time series of the influent during WWF conditions. The

CITYDRAIN model (Achleitner et al., 2007) was selected as the conceptual model as it is open source (inside Matlab) and it takes into account the basic phenomena that govern the amount and dynamics of the influent. Also, it requires the estimation of only a small number of parameters whose values or ranges of values can be inferred from the basic information of a sewershed.

### I. Flow

CITYDRAIN calculates the amount of effective rainfall by adopting the concept of a virtual basin (Achleitner et al., 2007). According to this concept (Figure 4-5), effective rainfall is calculated by subtracting the initial loss from total rainfall and then multiplying it with the runoff coefficient. Permanent losses like evapo-transpiration are considered only in dry periods to mimic an emptying process of the virtual basin (Equation 4-14).



**Figure 4-5** Schematic of virtual basin. a) filling of the basin in wet periods where the amount of spilled water is multiplied by runoff coefficient for calculating the amount of effective rainfall, b) emptying process with a fixed rate (permanent loss) in dry periods (see Equation 4-14).

$$\begin{cases} h_e = \text{Max}(r_R - (\text{Int\_loss} - x_t) \times \text{Runoff\_coeff}, 0) & \text{Wet periods} \\ \frac{dx}{dt} = -\text{Perm\_loss} & \text{Dry periods} \end{cases} \quad \text{Equation 4-14}$$

In the above equation,  $h_e$  (mm/t) represents the effective rainfall,  $r_R$  (mm/t) the total rainfall,  $Int\_loss$  (mm) the initial loss,  $Runoff\_coeff$  (---) the runoff coefficient, and  $Perm\_loss$  (mm/t) the permanent loss (overall three parameters).

The height of the effective rainfall is then multiplied by the fraction of sewershed area which contributes to the generation of runoff to calculate flow. The routing method, proposed by (Motiee et al. (1997) that is based on a simplified form of the Muskingum flow routing equations (Roberson et al., 1995) is then used for routing flow and pollutants inside the sewer system.

## ***II. Composition***

For the generation of pollutant time series in WWF conditions, CITYDRAIN uses a rather simplistic approach in which a fixed pollutant concentration is imposed to the system:

$$\begin{cases} C(t) = C \text{ if } h_e > 0 \\ C(t) = 0 \text{ if } h_e = 0 \end{cases} \quad \text{Equation 4-15}$$

where,  $C(t)$  is the generated pollutant concentration in time,  $C$  is a model parameter representing the concentration in WWF conditions, and  $h_e$  is the effective rainfall. Given the importance of the influent time series in WWF conditions, a more appropriate conceptual model was used for simulating the accumulation-wash off processes corresponding to the particulate concentrations. To this aim, a new block was developed and implemented in CITYDRAIN to generate the pollutant concentration time series in WWF conditions. Equation 4-16 shows the mathematical formulation of the selected accumulation-wash off model (Kanso et al., 2005).

$$\left\{ \begin{array}{l} \text{Accumulation model: } \frac{dM_{(t)}}{dt} = K_a \times (m_{\text{lim}} \times S_{\text{imp}} - M_{(t)}) \\ \text{Wash off model: } \frac{dM_{(t)}}{dt} = -W_e \times I_{(t)}^w \times M_{(t)} \end{array} \right. \quad \text{Equation 4-16}$$

where,  $M_{(t)}$  is: the available pollutant mass on the sewershed surface at time  $t$  (kg),  $K_a$  is the accumulation coefficient (1/day),  $m_{\text{lim}}$  is the maximum accumulated mass (kg/ha),  $S_{\text{imp}}$  is the impervious area (ha),  $I_{(t)}$  is the rainfall intensity (mm/hr),  $W_e$ , and  $w$  are calibration parameters to be estimated using observed rainfall and influent data. In case of the availability of influent data, the CITYDRAIN model parameters can be calibrated (i.e. their uncertainty reduced through Bayesian calibration (see Section 4.2.4)). However, if there is no measured influent data, the uncertainty in the parameters (i.e. parameters in Equation 4-14 and Equation 4-16) will remain and the uncertainty on the generated influent time series could be larger. Other parameters like the sewershed area or maximum conveyance capacity of the sewer system that have physical meaning can be obtained from general information on the sewershed.

### **III. Model setup**

In the CITYDRAIN model, the components of a sewershed system are modeled by a set of sewershed blocks and depending on the availability of data and level of heterogeneity in the sewershed, users may choose different numbers of blocks for modeling the entire sewershed. However, it should be noted that increasing the number of blocks will result in an increase in the number of model parameters which in turn could cause difficulties in parameter estimation (e.g. an unrealistic number of simulations (Martin and Ayesa, 2010)) when the model is to be calibrated using the measured flow and water quality data. In the modeling step, different CITYDRAIN configurations should be tested and a decision made on the best one.

#### 4.2.4 Bayesian model calibration of the CITYDRAIN sewer model

As explained in the previous section, the dynamics of the influent time series under WWF conditions is modeled using the CITYDRAIN model. However, one should be aware of the fact that modeling the influent time series under WWF conditions using a conceptual model may not lead to reliable results unless the model is calibrated and the effect of different sources of uncertainties on the model outputs (i.e. flow and other pollutants) are taken into account. To this aim, a Bayesian estimation framework was used to update the initial ranges of values (i.e. in a Bayesian parameter estimation method, the initial probability distributions or prior distributions reflect the initial knowledge on the value of uncertain model parameters) that were assigned to the parameters of the CITYDRAIN model (i.e. estimating the posterior distribution of parameters using their prior distribution and the measured data on flow and pollutant concentrations see Table 4-5 for the uncertain model parameters)). In general, the posterior distribution of parameters using Bayes' theorem can be formulated by Equation 4-17.

$$h(\theta | Data) = \frac{f(Data | \theta) p(\theta)}{f(Data)} \quad \text{Equation 4-17}$$

where  $h(\theta | Data)$  is the posterior distribution,  $p(\theta)$  is the prior distribution,  $f(Data)$  is merely a proportionality constant so that  $\int h(\theta | Data) = 1$ , and  $f(Data | \theta)$  constitutes the likelihood function which measures the likelihood that the data correspond to the model outputs with parameter set  $\theta$ . Assuming homoscedastic uncorrelated Gaussian error (i.e. having normal distribution with the same variance and no correlation in time), the likelihood function can be formulated according to Equation 4-18 (Bates and Campbell, 2001; Marshall et al., 2004).

$$f(Data|\theta) = (2\pi\sigma^2)^{-n/2} \prod_t^n \exp\left\{-\frac{[Data_t - R(x_t; \theta)]^2}{2\sigma^2}\right\} \quad \text{Equation 4-18}$$

where  $n$  is the number of observations,  $\sigma^2$  is the variance of the residual error (i.e. the difference between model predictions and observed values which equals the measurement error if it is assumed that the model is perfectly representing reality),  $Data_t$  is the observed variable at time  $t$ ,  $x_t$  is the set of inputs at time  $t$ ,  $\theta$  is the set of model parameters and  $R(x_t; \theta)$  represents the model output as a function of  $x_t$  and  $\theta$ .

A specific form of Markov chain Monte Carlo (MCMC) sampler known as differential evolution adaptive Metropolis or DREAM (Vrugt et al., 2008) is used to efficiently estimate the posterior distribution of the CITYDRAIN model parameters that are involved in the generation and routing of flow and pollutants in WWF conditions, given the time series of flow and influent composition. It should be noted that the proposed Bayesian approach is not only capable of capturing the effect of model parameter uncertainty, but also of capturing the effect of other sources of uncertainties (model structure, input, etc) that could result in some discrepancies between the simulated influent time series and the observed series.

#### 4.2.5 Synthetic influent generation

Once the uncertainty ranges of the CITYDRAIN model parameters are updated, synthetic influent time series for a desired number of years considering the variability in the inputs of the CITYDRAIN model (i.e. rainfall and influent time series in DWF conditions) and also the total uncertainty can be obtained as follows:

1. Synthetic generation of the 15-minute time series of rainfall for one year
2. Synthetic generation of the 15-minute time series of the influent under DWF conditions for one year
3. Sampling a point from the posterior distribution of the CITYDRAIN model parameters
4. Inputting the generated time series 1) and 2) and the parameters sampled in 3) and running the CITYDRAIN model for one year
5. Repeating 1) to 4) for a desired number of years

In this study, the contribution of the noise term (i.e. characterized using a normal distribution with zero mean and standard deviation of Sigma) to the output is treated as a source of variability. This decision is based on the assumption that the main part of the difference between the simulated and observed signals is due to an actual fluctuation of the influent time series (an instance of variability) or some random measurement error (an instance of uncertainty).

Obviously during the DWF conditions, the influent time series is generated using the statistical model that is explained in Section 4.2.2 and the CITYDRAIN model has no effect on the generated influent time series.

### **4.3 DATA AND CASE STUDY**

The Eindhoven WWTP with a design capacity of 750000 population equivalent (PE) is the third largest WWTP in the Netherlands. The sewershed served by the Eindhoven WWTP has a total area of approximately 600km<sup>2</sup> and comprises of three main sub-sewersheds called Nuenen/Son, Eindhoven Stad, and Riool-Zuid. The influent data used in this study comprised of sensor data for flow rate, ammonia (measured using an ion-selective sensor) soluble COD, total COD, and TSS (the latter three measured using an UV/VIS-based

sensor) in the period of September 2011 to September 2012 at the outlet of the Nuenen/Son, Eindhoven Stad, and Riool-Zuid sub-sewersheds (entrance point to the treatment plant).

The long-term daily rainfall data and also rainfall data with finer temporal resolution provided by KNMI (Royal Netherlands Meteorological Institute De Bilt, The Netherlands) and Waterschap De Dommel (Boxtel, The Netherlands) were used for estimating the parameters of the weather generator proposed in this paper. Table 4-1 shows a summary of the data types and also their applications in the development of the proposed influent generator.

**Table 4-1** Summary of the data type and their applications

Types of data	Application
Rainfall (1951-2013)	<ul style="list-style-type: none"> <li>✓ Estimation of the rainfall generator parameters (i.e. <math>P(W D)</math>, <math>P(W W)</math>, <math>\alpha</math>, <math>\beta</math>)</li> <li>✓ Calibration of the CITYDRAIN model</li> </ul>
Minimum and maximum air temperature (1951-2013)	<ul style="list-style-type: none"> <li>✓ Calibration of the air temperature generator</li> </ul>
Bioreactor temperature (2011-2012)	<ul style="list-style-type: none"> <li>✓ Calibration of the regression model for the generation of bioreactor temperature as a function of air temperature</li> </ul>
Influent data in DWF conditions (2011-2012)	<ul style="list-style-type: none"> <li>✓ Calibration of the DWF generator</li> </ul>
Influent data in WWF conditions (2011-2012)	<ul style="list-style-type: none"> <li>✓ Calibration of the CITYDRAIN model</li> </ul>

## 4.4 RESULTS AND DISCUSSION

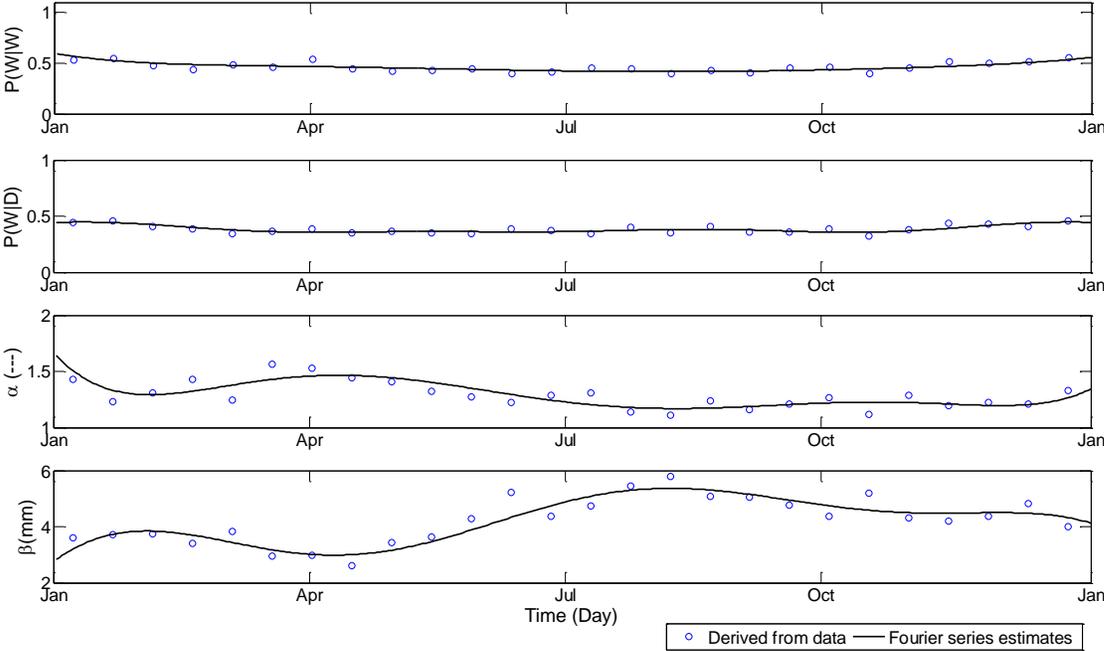
This section presents the outputs and some discussion on the results of different components of the proposed influent generator. As explained in the methodology section, the parameters of the statistical models used for synthetic generation of rainfall, air and bioreactor temperature, as well as the multivariate auto-regressive time series models (used for the generation of influent time series in DWF conditions) were calibrated using the historical weather data and observed influent time series in DWF conditions. The parameters of the CITYDRAIN model (the conceptual model for modeling influent time series under WWF conditions) were estimated using the Bayesian estimation framework. Once the parameters of the statistical and the conceptual models were calibrated, different realizations of influent time series would be generated by running the CITYDRAIN model with different realizations of the rainfall and influent time series under DWF conditions (stochastic inputs) and different sets of parameters sampled from the posterior distribution of the CITYDRAIN model parameters.

The performance of the weather generator and the influent generator under DWF conditions were evaluated by comparing the statistical properties of the generated time series with those of the historical time series. The results corresponding to the Bayesian calibration of the CITYDRAIN model are explained and at the end a 7-day snapshot of a generated one year influent time series is presented and discussed.

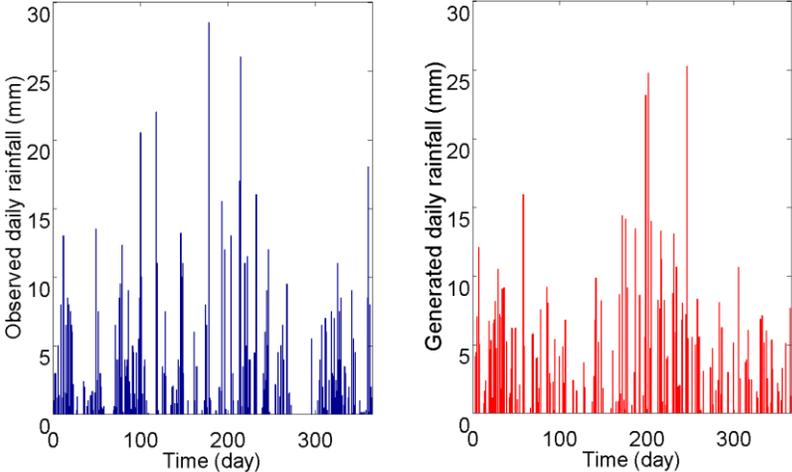
### 4.4.1 *Synthetic generation of rainfall*

The parameters of the statistical Markov-gamma model were estimated using different Fourier series models fitted on parameters values derived from the recorded rainfall data (Figure 4-6) in the studied Eindhoven catchment and then different realizations (with a

random seed used for each year long generation of rainfall time series) of rainfall time series were generated (Figure 4-7).



**Figure 4-6** Seasonal variation in the Markov chain-gamma model parameters



**Figure 4-7** A year-long realization of daily rainfall (right) versus an observed one (left)

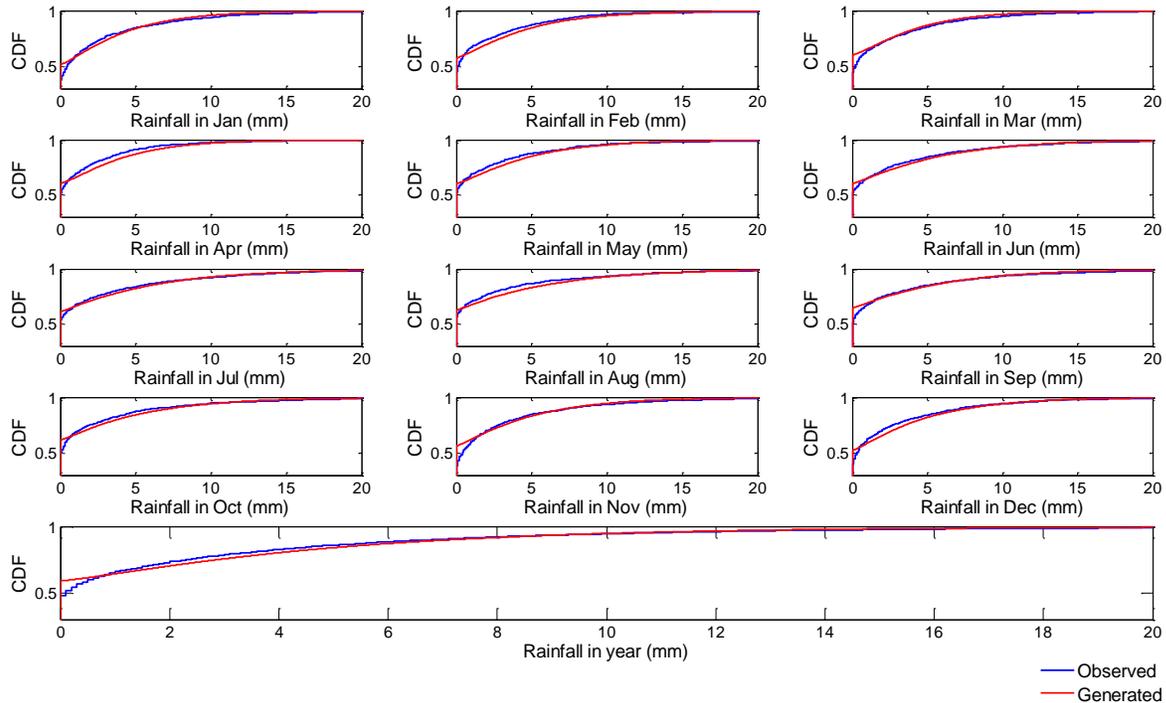
To evaluate the performance of the Markov-gamma model for realistic generation of rainfall time series, the CDF of the observed rainfall time series (CDF curves include both wet and dry days) corresponding to different seasons were compared with those of the generated rainfall time series ( Figure 4-8 for the CDF comparison and Figure A1 for a q-q plot between the simulated and observed rainfall time series).

The observed CDFs in Figure 4-8 are constructed using the daily rainfall records between 1951 to 2013 and the generated CDFs correspond to 1000 years of synthetic rainfall time series, generated using the explained Markov chain-gamma model whose parameters (depicted in Figure 4-6) were estimated from the daily rainfall records (i.e. from 1951 to 2013).

The results indicate that not only are the basic yearly statistics (i.e. average and variance) of the generated rainfall time series consistent with the recorded rainfall time series, but also the seasonal variations in rainfall intensity and frequency of wet days are respected. As indicated in Table 4-2, the differences between the observed average rainfall and the simulated average rainfall (based on the rainfall data from 1951 to 2013 and 1000 years of synthetic rainfall time series, respectively) in the different months are below 10%, except for the months March and April in which the differences are -16.5% and 15% respectively. Nevertheless, all the differences between the average simulated and average observed rainfall values are below 20% which is acceptable according to the work of Richardson (1981) in which the differences between average simulated and observed values in some months are above 20%.

The discrepancies between certain percentiles of the simulated and observed rainfall distributions (Figure 4-8 or Figure A1) could be associated to the difference between rainfall generator parameters derived from observed data (sample parameters, i.e. blue circles in Figure 4-6) and those that are estimated using different Fourier time series and used for the rainfall generator, i.e. solid lines in Figure 4-6). In addition, the difference

between the length of observed and generated rainfall time series (i.e. 62 years of observed data compared to 1000 years of generated rainfall data) could be another reason for the difference between the extreme percentiles (see Figure A1).



**Figure 4-8** Cumulative distribution function of daily rainfall in the studied Eindhoven catchment

Moreover, Table 4-3 shows that the hourly time series of rainfall which was generated using the time disaggregation method (i.e. disaggregation of daily to hourly time series) has the same statistical characteristics as the observed one. Overall, the synthetic generation of rainfall in which the statistical properties of the time series is respected across different time scales is a significant improvement compared to the rainfall generation in for instance the BSM influent generator (Gernaey et al., 2011) in which there is no clear way for extracting and incorporating the statistical properties of available recorded rainfall data into synthetic rainfall time series generation. In addition, the flexibility of the proposed rainfall generator

allows users to define different scenarios reflecting future changes in the precipitation regime (e.g. due to climate change (Chen et al., 2010)) and its effect on the influent time series (e.g. what would happen if the amount of rainfall or the number of wet/dry days increases by 20% in specific seasons, a feature that is not available in previous rainfall generators, e.g. the rainfall generator proposed by Gernaey et al. (2011)).

**Table 4-2** Average rainfall amount and number of wet days for the Eindhoven catchment

Month	Amount of Rainfall <sup>1</sup> (mm)		Average number of Wet Days <sup>2</sup>	
	Observed	Generated	Observed	Generated
January	72.3	67.0	16	14
February	52.0	57.0	12	11
March	63.4	54.4	13	12
April	44.1	51.9	12	11
May	58.3	60.9	12	12
June	68.0	68.4	12	11
July	74.7	73.5	12	11
August	64.6	71.0	11	11
September	67.9	62.1	12	10
October	62.0	65.0	12	11
November	71.1	66.4	15	12
December	70.0	74.0	14	14
Annual	768	772	152	141

<sup>1</sup>The average amount of total rainfall in different months for observed (i.e. rainfall data from 1951 to 2013) and generated rainfall time series (i.e. 1000 years of rainfall data, generated using the proposed rainfall generator)

<sup>2</sup>The average number of wet days in different months for observed (i.e. rainfall data from 1951 to 2013) and generated rainfall time series (i.e. 1000 years of rainfall data, generated using the proposed rainfall generator)

**Table 4-3** Basic statistics of hourly rainfall data for the Eindhoven catchment

Statistics	Unit	Observed Value	Simulated Value
Mean	mm	0.08	0.08
Standard deviation	mm	0.60	0.60
Lag 1 auto-correlation <sup>1</sup>	---	0.33	0.36
Fraction of dry hours	---	0.92	0.94

<sup>1</sup>Correlation between the amount of rainfall at time  $t$  and  $t - 1$

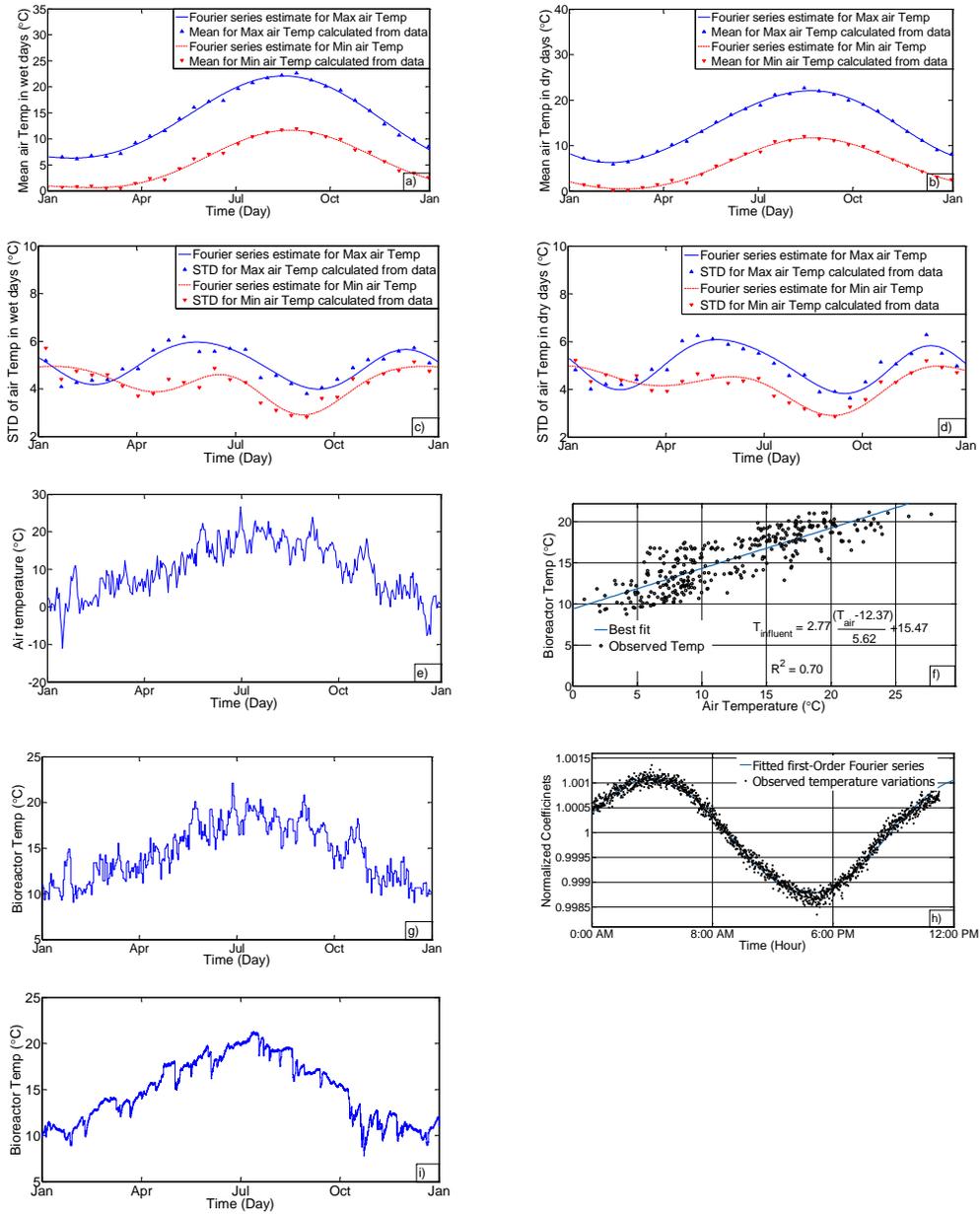
#### 4.4.2 Synthetic generation of air and bioreactor temperature

The seasonal variation in the mean maximum air temperature, the mean minimum air temperature, the standard deviation of maximum air temperature, and the standard deviation of minimum air temperature for dry and wet days, captured using Fourier series models are illustrated in Figure 4-9a to Figure 4-9d. As indicated, the mean values for both maximum and minimum air temperatures have an upward trend from the winter until the midst of summer (when they reach their maximum values), followed by a downward trend until they reach their minimum values in the winter again. However, comparing Figure 4-9a and Figure 4-9b, representing the seasonal variations for wet and dry days suggest that there is no significant difference between the seasonal variation when the state of day (wet or dry) is taken into account. In other words, it can be concluded that for the case study of this research, the variation in the mean maximum and the mean minimum temperatures is mostly a function of the seasons of the year rather than the state of the day. As explained in Section 4.4.3, a multivariate linear first-order model was fitted on the residual time series of maximum and minimum air temperatures for synthetic generation of maximum and minimum air temperatures and in the end the generated air time series were converted to their original values through Equation 4-4 and Equation 4-5 using the seasonal mean and standard deviation values illustrated in Figure 4-9a to Figure 4-9d.

The daily temperature of the bioreactor was generated through a linear regression model which relates the daily average bioreactor temperature to the daily average air temperature. Figure 4-9e and Figure 4-9g show random generation of an air and bioreactor temperature time series for one year. The linear model in Figure 4-9f which was developed using the concurrently measured air and bioreactor temperature for one year (i.e. September 2011 to September 2012, illustrated in Figure 4-9i) shows that the average bioreactor temperature can be estimated reasonably ( $R^2 = 0.70$ ) as a linear function of air temperature. It should be noted that the effect of the state of day (i.e. dry or wet) on air temperature (although for the current case study was not significant) which in turn affects the bioreactor temperature has been taken into account in random generation of air temperature. Moreover, an attempt to use two different regression models depending on the state of day (i.e. one regression model for dry days and another one for wet days) did not result in any improvement in the prediction of bioreactor temperature as a function of air temperature.

The average diurnal variation of bioreactor temperature in Figure 4-9h was extracted by fitting a first order Fourier series estimate to the normalized bioreactor temperature variations which in turn was used for converting the daily bioreactor temperature time series into a time series with 15-minute temporal resolution.

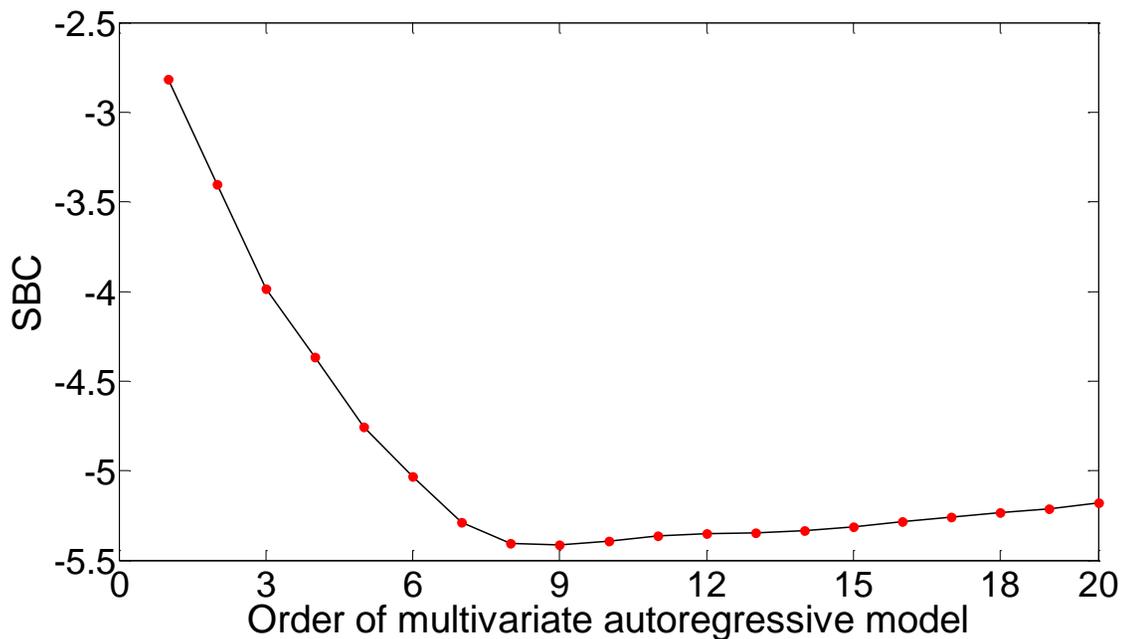
Despite the fact that the diurnal variation pattern in Figure 4-9f clearly shows a periodic behavior in time (which corresponds to the diurnal variation of bioreactor temperature), there is no significant difference between the highest and lowest temperature throughout a day (i.e. the highest temperature is only around 1.001 times the daily average bioreactor temperature and the lowest temperature is around 0.9985 times the daily average bioreactor temperature). Therefore, in practical applications (at least for the case study in this research), the diurnal temperature variation can be ignored.



**Figure 4-9** Random generation of air and bioreactor temperature for one year for the Eindhoven WWTP: Seasonal variation in mean maximum and mean minimum air temperatures in wet days (a), in dry days (b), Seasonal variation in standard deviations of maximum and minimum air temperatures in wet days (c), in dry days (d), Randomly generated daily air temperature (e), Linear regression model between daily air and bioreactor temperatures (f), Bioreactor temperature time series with 15-minute temporal resolution (g), Average diurnal variation of bioreactor temperature (h), Observed historical bioreactor temperatures used in the analysis (i)

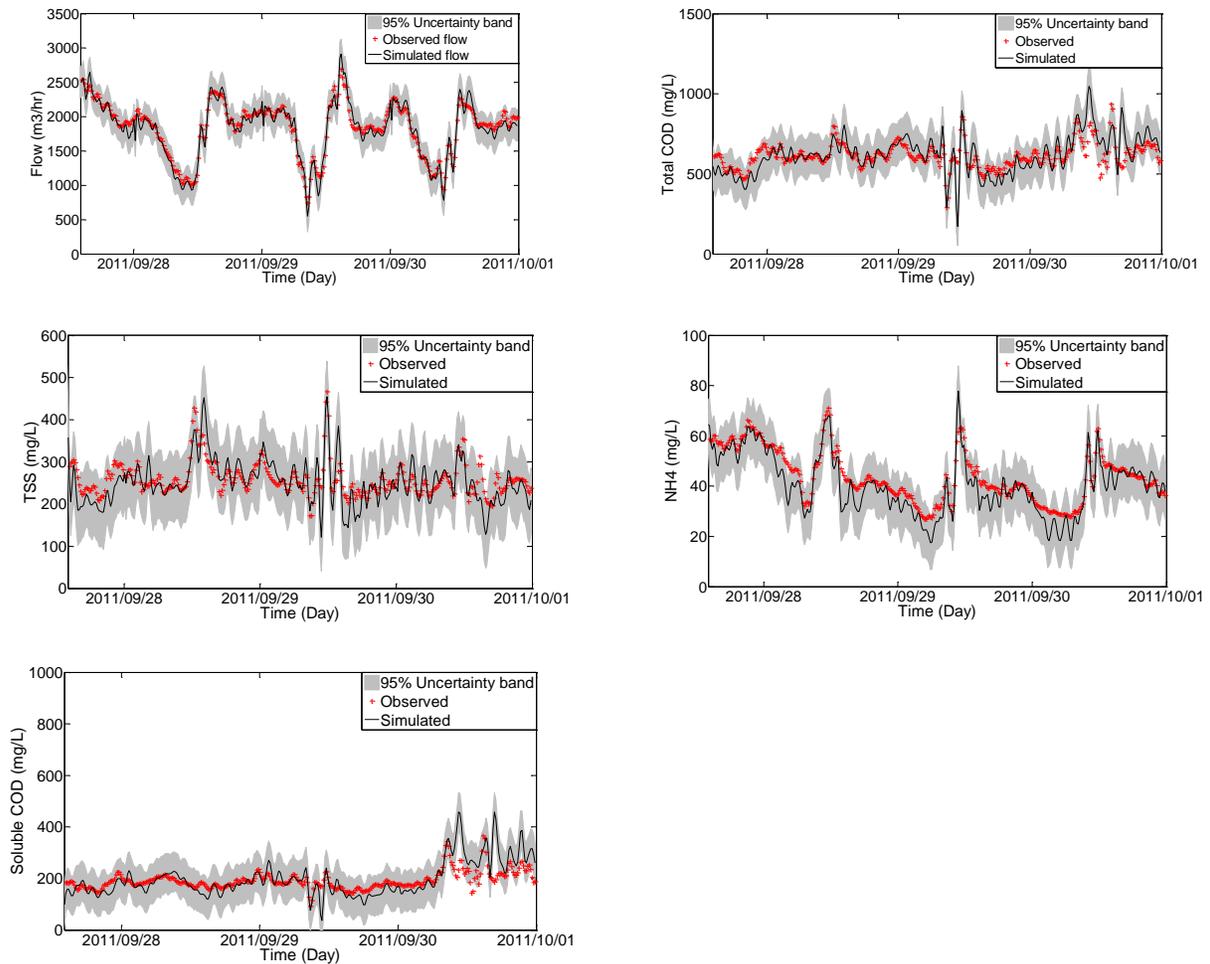
#### 4.4.3 Multivariate auto-regressive model for DWF generation

Influent data corresponding to 82 dry days were analyzed for estimating the parameters of the multivariate auto-regressive time series model for DWF generation (Table A2). As explained, the order of the multivariate auto-regressive model was determined based on the Schwarz Bayesian Criterion (SBC) criterion and the parameters were estimated according to a specific least square algorithm proposed by Neumaier and Schneider (2001). Figure 4-10 shows the variation of the SBC criterion for different model orders, ranging from 1 to 20 ( $p$  in Equation 4-13). As indicated, the SBC criterion reaches its minimum value at 9, which was thus selected as the order of the multivariate autoregressive model. This means that the value of the influent time series at time  $t$  is simulated as a function of the last 9 influent values antecedent to time  $t$ .



**Figure 4-10** Variation of SBC with order of multivariate time series model

Figure 4-11 shows a continuous 3-day DWF influent time series with the results corresponding to the most likely simulated multivariate auto-regressive model. The uncertainty band was generated through random generation of the noise term (i.e.  $p, A_i$  in Equation 4-13 were fixed and the noise term was generated from  $\varepsilon_t \sim N(0, C)$ ). It should be noted that the water quality data, except for ammonia, did not exhibit the strong diurnal variation that is typically observed in other catchments (Martin and Vanrolleghem, 2014).



**Figure 4-11** Observed and most likely simulated influent time series under DWF conditions

Note: The continuous 3-day influent time series belongs to 82 days of DWF data used for estimating the parameters of the multivariate auto-regressive model

One of the main advantages of the proposed multivariate time series model over univariate time series models (Martin et al., 2007) or the DWF generator in the BSM influent generator (Gernaey et al., 2011) is that not only are the auto-correlation structures in time respected but also the cross-correlation structures. Table 4-4 shows the correlation matrix for the randomly generated and observed influent time series under DWF conditions.

**Table 4-4** Correlation matrix for the generated and observed influent time series in DWF

Generated influent time series						Observed influent time series					
	Flow	Soluble COD	Total COD	TSS	NH <sub>4</sub>		Flow	Soluble COD	Total COD	TSS	NH <sub>4</sub>
Flow	1.00					Flow	1.00				
Soluble COD	-0.11	1.00				Soluble COD	-0.12	1.00			
Total COD	-0.04	0.77	1.00			Total COD	-0.06	0.77	1.00		
TSS	0.06	0.32	0.80	1.00		TSS	0.05	0.33	0.81	1.00	
NH <sub>4</sub>	-0.43	-0.04	-0.06	-0.04	1.00	NH <sub>4</sub>	-0.46	0.00	-0.02	-0.03	1.00

Depending on the type of the model that is to be used for modeling the treatment processes inside a WWTP (e.g. ASM models (Henze et al., 2000)), the wastewater constituents in Table 4-4 can be further converted to WWTP model state variables. However, as illustrated in Figure 4-1, influent fractionation should be considered part of WWTP modeling as different WWTP models may have different state variables (Martin and Vanrolleghem, 2014).

#### **4.4.4 CITYDRAIN model calibration and synthetic influent generation**

As explained in the methodology section, the CITYDRAIN model with three catchment blocks representing the main sub-sewersheds in the Eindhoven sewershed was used for modeling the dynamics of the influent time series during WWF conditions. The decision on

the number of catchment blocks was made based on the information obtained from previous studies as well as the measured influent data that were used for model calibration (Schilperoort, 2011).

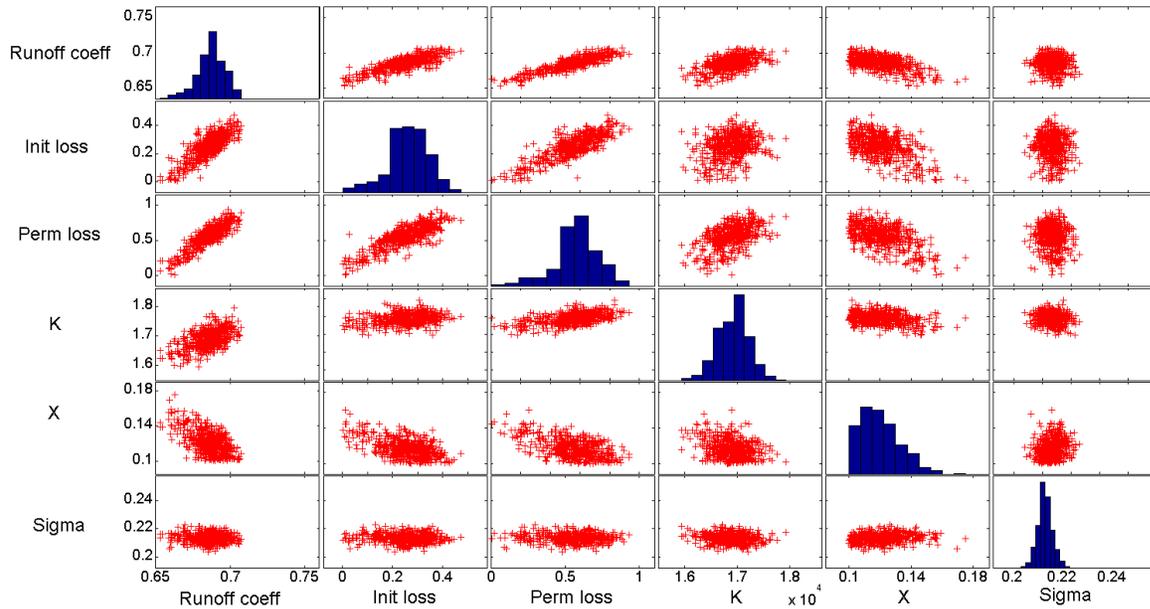
Uniform distributions representing the initial knowledge on parameters were selected as prior distributions (Table 4-5) and their corresponding posterior distributions were estimated by sampling from Equation 4-18 (i.e. 12000 samples which required around 10 hours of computation) using the DREAM sampler (as indicated in Table 4-5, 11 parameters were estimated). Figure 4-12 and Figure 4-13 show the posterior distributions of the CITYDRAIN model after calibrating the model for flow and TSS time series in WWF conditions (three days of dry weather simulation were used as the warm-up period to set the initial conditions of the system).

**Table 4-5** Prior distribution of parameters and the values for the maximum likelihood function

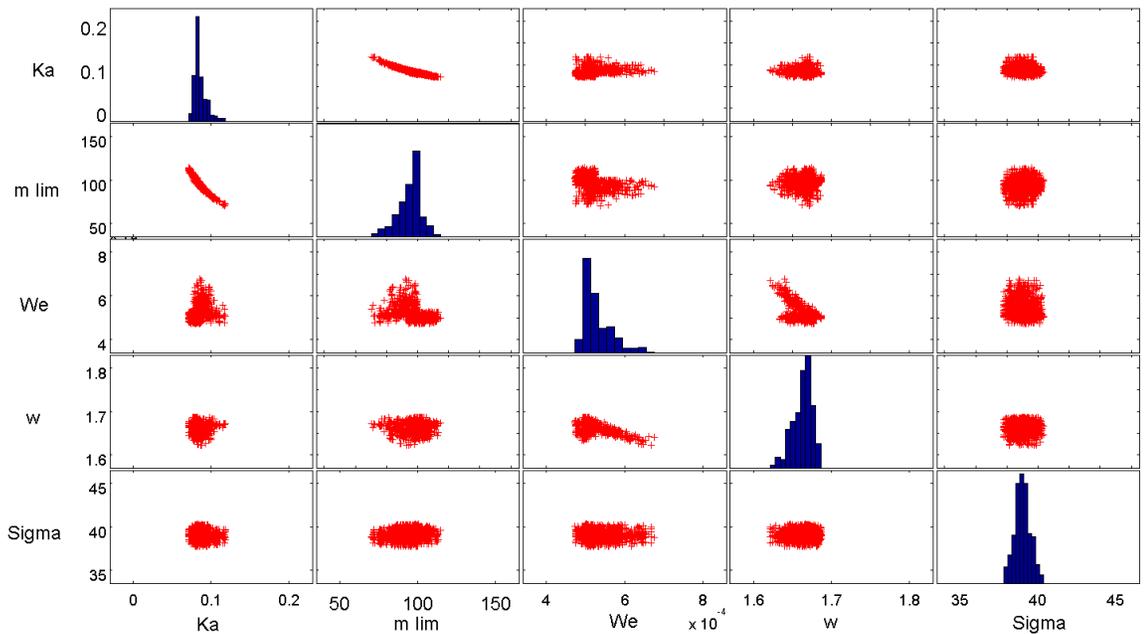
Parameter	Unit	Lower limit	Upper limit	Values corresponding to the maximum likelihood
Runoff coeff	---	0.6	0.9	0.69
Init loss	mm	0	2	0.3
Perm loss	mm/day	0	2	0.57
K (Muskingum coeff)	second	8000	20000	16 869
X (Muskingum coeff)	---	0.1	0.4	0.12
Sigma (for flow)	m <sup>3</sup> /hr	0.05	2	0.22
Ka	1/day	0.001	2	0.08
m lim	Kg/ha	0.001	120	100
We	---	0.0004	0.002	0.0005
w	---	1.5	2	1.667
Sigma (for TSS)	g/m <sup>3</sup>	20	70	38.9

As indicated in Figure 4-12 and Figure 4-13, some correlation among the parameters of the CITYDRAIN model exists. For example, the parameters that affect the generation of effective rainfall (i.e. runoff coefficient, initial loss, and permanent loss) are correlated, meaning that different combinations of these parameters could result in approximately the same amount of effective rainfall given the same inputs and values for other parameters. However, given the narrow ranges of values obtained for the marginal posterior distribution of parameters that affect the amount and dynamics of flow (i.e. Runoff coeff, Init loss, Perm loss, K, and X in Figure 4-12), the uncertainty band for flow relating to the *total model uncertainty* is mainly affected by the standard deviation of the residual error (i.e. Sigma in Figure 4-12) and not by the uncertainty of the CITYDRAIN model parameters.

The parameters that affect the accumulation of a pollutant (i.e.  $m_{lim}$ , and  $K_a$ ) and those that affect the wash-off of TSS are also correlated (Figure 4-13). Given the different correlation structures that exist among some parameters it is very important to sample from the joint posterior distribution of parameters to properly propagate the effect of parameter uncertainties to the outputs.



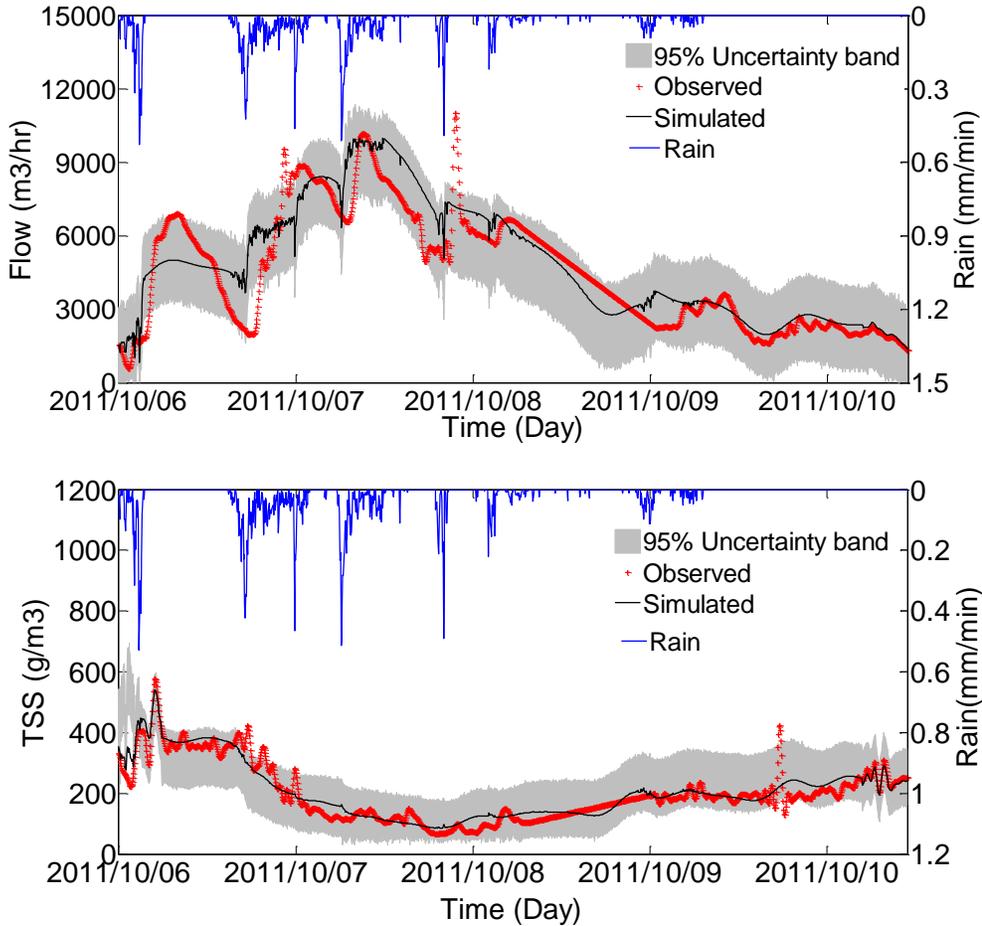
**Figure 4-12** Posterior distribution of parameters for flow calibration where, runoff coeff, init loss, and perm loss are respectively the runoff coefficient, initial loss (mm), permanent loss (mm/day) parameters in the virtual basins model that is used in the CITYDRAIN model,  $K$  (sec) and  $X$  are the routing parameters used in the Muskingum method, and  $\Sigma$  is the standard deviation of the residual error. The blue histograms represent the marginal posterior distribution of the individual parameters and the red scatter plots represent relationships corresponding to various combinations of parameters (Equation 4-14)



**Figure 4-13** Posterior distribution of parameters used for TSS calibration where Ka is the accumulation coefficient (1/day), m\_lim is the maximum accumulated mass (kg/ha), We, and w are the calibration parameters (Equation 4-16)

To consider the effect of *total model uncertainty* (i.e. including model parameter uncertainty and the standard deviation of noise (i.e. Sigma for flow and TSS in Figure 4-12 and Figure 4-13) on the outputs of CITYDRAIN model, a Monte Carlo simulation was performed by sampling from the joint *posterior distribution* of parameters and running the model for 1000 times for the rainfall time series between 6<sup>th</sup> and 10<sup>th</sup> of October 2011 (i.e. the rainfall time series concurrent with the observed flow and TSS concentration time series used for CITYDRAIN calibration). Figure 4-14 illustrates the 95% uncertainty band for flow and TSS. This uncertainty band was constructed by selecting the 2.5 and 97.5 percentiles of the cumulative distribution of flow and TSS as the lower and upper limits of uncertainty of simulation with the rainfall time series shown in the figure (The sudden increase in TSS concentration between the 9<sup>th</sup> and 10<sup>th</sup> of October is most likely due to measurement error as there is no rainfall on that date to justify the observed surge). The figure also presents the observed and the best simulated time series. The latter corresponds

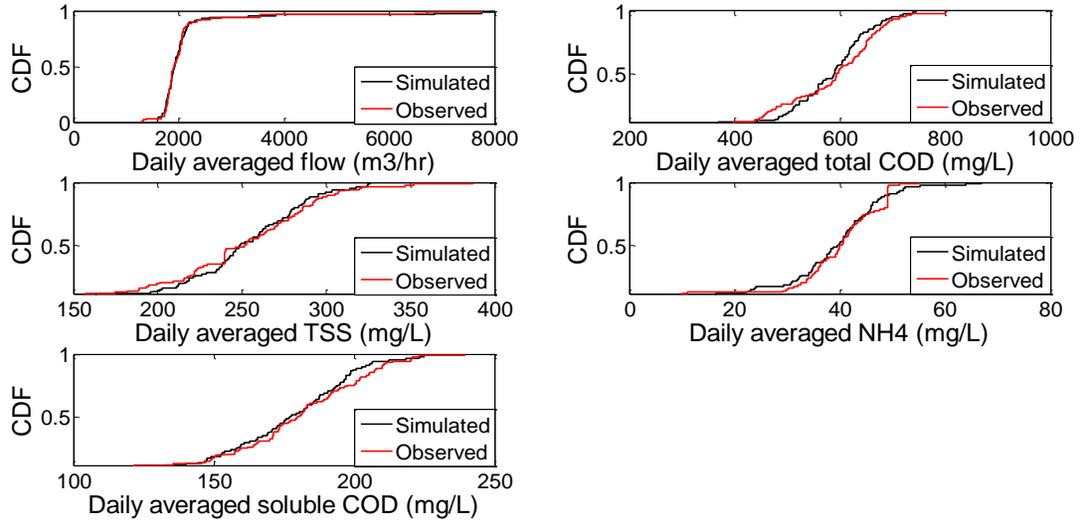
to the simulations obtained with the set of parameters that has the highest *likelihood function* value.



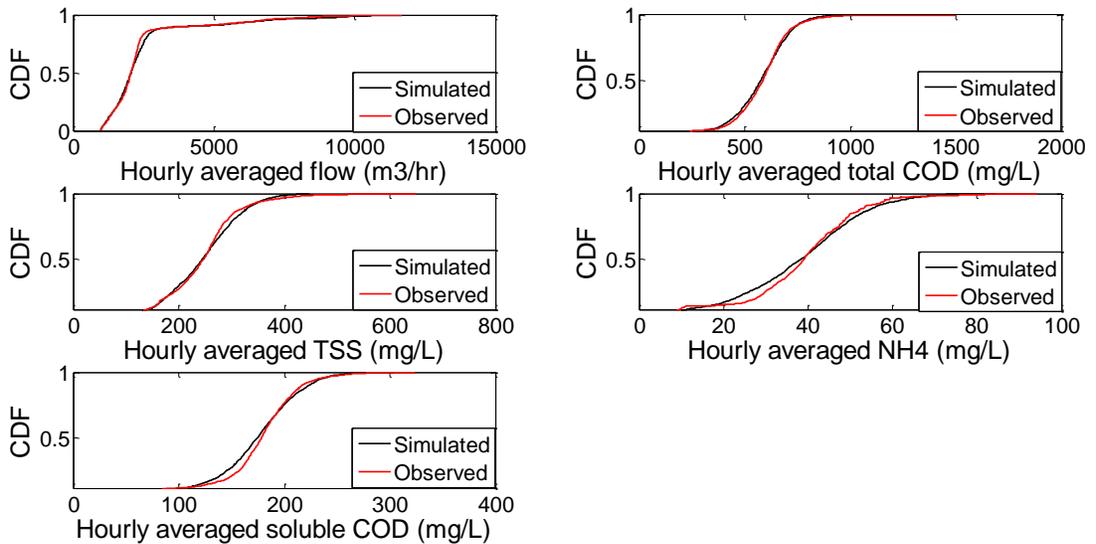
**Figure 4-14** Uncertainty bands for flow (top) and TSS concentration (bottom) in a 4-day wet weather period (Rain series in blue and maximum likelihood simulation in black)

To further analyze the statistical properties of the simulated influent time series during both DWF and WWF conditions, the cumulative distribution function (CDF) of the simulated and observed influent flow and pollutant concentrations were compared in Figure 4-15 and Figure 4-16. The simulated and observed influent time series with 15-minute temporal resolution were averaged (using the flow and concentration time series) to construct the corresponding daily and hourly influent series. Figure 4-15 and Figure 4-16 show that the

influent generator has excellent performance when it comes to predicting the daily and hourly influent flow and pollutant concentration values (the comparison between the simulated and observed load values (results not shown) also indicated an excellent performance of the influent generator).



**Figure 4-15** CDFs of daily-averaged influent flow and concentration of influent pollutants

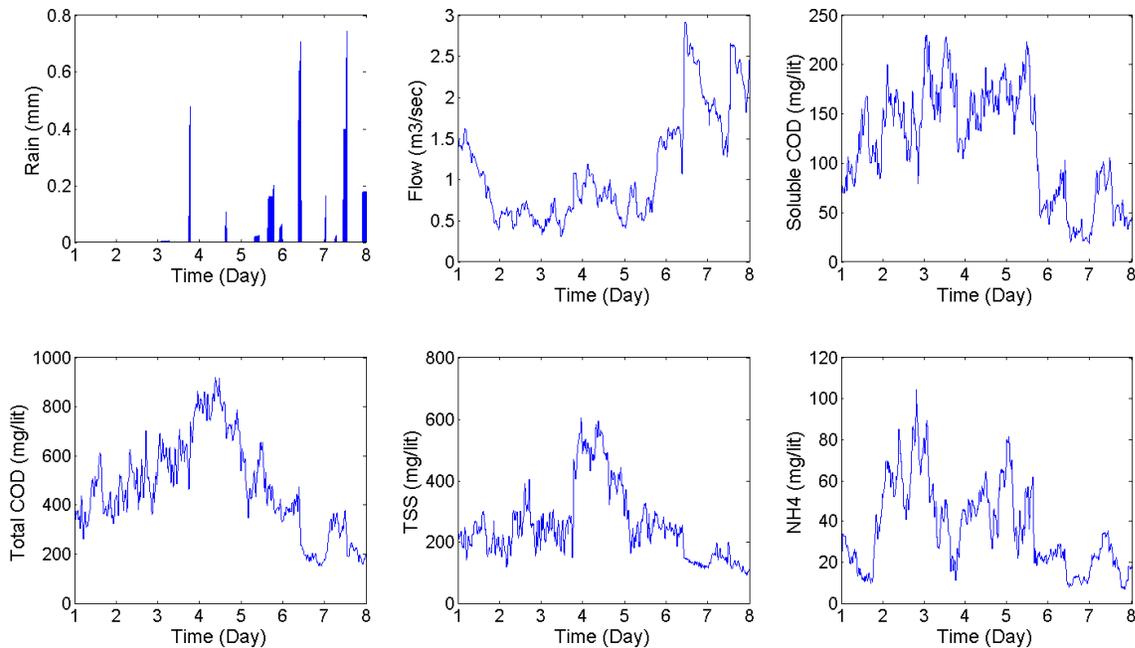


**Figure 4-16** CDFs of hourly-averaged influent flow and concentration of influent pollutants

It can be concluded from Figure 4-15 and Figure 4-16 that the statistical properties of the simulated time series were very similar to the properties of the observed series when the CITYDRAIN model is fed with the observed rainfall time series.

#### 4.4.5 Synthetic generation of influent time series

As explained in the methodology section, synthetic generation of a one year influent time series with 15-minute temporal resolution is thus possible by sampling from the joint posterior distribution of the CITYDRAIN model parameters (one vector of CITYDRAIN parameters for each year) and running the model with the synthetically-generated rainfall and DWF influent time series (both with 15-minute temporal resolution). The latter two series are generated using the proposed rainfall and DWF generators respectively.



**Figure 4-17** A 7-day realization of rainfall and influent time series (flow and composition)

Figure 4-17 shows a 7-day snapshot of a generated one year influent time series. During the hours of the first day the time series of flow has a descending trend as the runoff produced by rainfall event just before the first day (not depicted in Figure 4-17) exits the sewer system and the flow time series reaches its DWF conditions with a typical periodic pattern (the second day in Figure 4-17). During the last hours of the third day another rainfall event occurs and the flow time series increases while the time series of soluble COD and ammonia drop due to dilution of wastewater by runoff. However, during the same period of time there is a sudden increase in the total COD and TSS concentrations due to the wash-off of particulate material. After the wash-off of the particulates during the last hours of the fourth day, the concentrations of total COD and TSS drop due to the dilution of the wastewater with runoff.

## **4.5 CONCLUSION**

In this paper a combination of statistical and conceptual modeling tools was proposed for synthetic generation of dynamic influent time series of flows and pollutant concentrations with 15-minute temporal resolution. The rainfall generator is capable of considering observed annual and seasonal rainfall regimes and keeping the consistency of the generated rainfall time series across different temporal resolutions. For dry weather conditions, comparison between observed and simulated influent time series for the Eindhoven case study confirmed the capability of the proposed multivariate auto-regressive model in generating realistic influent time series for flow and pollutants composition. Moreover, long-term generation of influent time series under dry and wet weather conditions could be achieved by running a constructed CITYDRAIN model of the sewershed using the generated stochastic inputs (i.e. rainfall and influent time series in DWF condition). Further, uncertainty could be captured by sampling different vectors of the model parameters from the posterior distribution obtained after Bayesian parameter estimation on the basis of the case study data.

Overall, the proposed influent generator provides a clear and coherent tool to incorporate general and easy-to-obtain information on the physical characteristics of a sewershed as well as climate conditions of the region into the synthetic generation of the influent flow and composition of a treatment plant. If there are no observed data for calibrating the parameters of the proposed influent generator, a range of values should be assigned to the uncertain model parameters based on expert elicitation or transfer of information from similar sewersheds. The flexibility of the presented influent generator allows users to define different scenarios reflecting the projected change in climate and the characteristics of the sewershed (e.g. population growth, change in pervious area) and evaluate their effect on the generated influent time series and the treatment plant to be designed.

## CHAPTER 5: APPLICATION OF A NEW PROBABILISTIC DESIGN METHOD FOR WASTEWATER TREATMENT PLANTS- CASE STUDY: EINDHOVEN WWTP

To be submitted to the journal of Environmental Modeling and Software<sup>1</sup>

**Abstract:** The primary goal of wastewater treatment plants (WWTPs) is to remove pollutants from wastewaters so as to reach a set of effluent standards under a set of environmental, cost, and regulatory constraints. To design a WWTP according to these criteria, design engineers usually make the initial sizing of the plant using design guidelines or a set of modeling tools under steady state conditions. In these approaches the effect of different sources of uncertainties are taken into account in an implicit manner through the application of safety factors and/or selection of conservative values for design inputs. In this study, the application of a set of statistical and process-based dynamic modeling tools is proposed to explicitly characterize the uncertainty/variability in the input time series and model parameters and propagate these into the uncertainty in the model outputs (i.e. effluent wastewater composition and costs). The probability of non-compliance (PONC) can be calculated for the applicable effluent standards. The proposed probabilistic methodology provides the design engineers with a concerted framework to utilize and incorporate into the design of WWTPs the available and future information on the characteristics of the sewershed and the climate conditions, as well as the latest advances in dynamic modeling. Moreover, the calculated PONC can be used as an objective criterion

---

<sup>1</sup> Authors: Mansour Talebizadeh<sup>1\*</sup>, Evangelia Belia<sup>2</sup>, and Peter A. Vanrolleghem<sup>1</sup>

<sup>1</sup>modelEAU, Département de génie civil et de génie des eaux, Université Laval, 1065 av. de la Médecine, Québec (QC) G1V 0A6, Canada (Email : mansour.talebizadehsardari.1@ulaval.ca; Peter.Vanrolleghem@gci.ulaval.ca)

<sup>2</sup>Primodal Inc., 145 Aberdeen, Québec, QC G1R 2C9, Canada (Email: belia@primodal.com)

for comparing different design alternatives and help designers avoid the application of overly-conservative safety factors.

**Keywords:** Design under uncertainty; Monte Carlo simulation; Stochastic generation; Water Resource Recovery Facility.

## 5.1 Introduction

Wastewater treatment plants (WWTPs) are complex engineered systems whose failure to meet performance requirements can have detrimental effects on public health and the environment. Ideally, a WWTP system should be designed such that it is capable of coping with the dynamic flow and loading conditions so that the treated wastewater meets the effluent standards. However, in practice the initial sizing of a WWTP is performed mostly by assuming steady state conditions. Most design methods are based on the assumption of steady state conditions (whether the design is based on a specific design guideline, e.g. ATV (2000) or a steady state model (Ekama, 2009)). To make a design, representative values are to be selected for design inputs (e.g. influent flow and concentrations and the required effluent standards, as well as operational, kinetics, and stoichiometric parameters) and the dimensioning of the WWTP is done according to a set of experience-based rules or a set of empirical and/or process-based equations (Corominas et al., 2010; Flores-Alsina et al., 2012).

Keeping in mind that the design guidelines are based to a great extent on the invaluable knowledge that has been gained from years of experience on design and operation of WWTPs and pilot plants, they have certain shortcomings. For example, in design guidelines the uncertainty is taken into account by applying safety factors and/or making conservative assumptions regarding the design inputs which could result in a very conservative design (McCormick et al., 2007; Flores-Alsina et al., 2012). Considering the aggregated effect of

different sources of uncertainty by applying safety factors may result in double counting of a specific source of uncertainty. For example, in the ATV design guideline, a safety factor is applied to the calculated minimum required solids retention time (SRT) to take into account the variation in the maximum growth rate of microorganisms (caused by certain inhibition substances in the wastewater or temperature), as well as the variation in influent nitrogen load. The reason behind multiplying the calculated minimum SRT by a safety factor is to insure that SRT would be large enough for proper nitrification and hence the reduction of effluent ammonia concentration to the required effluent standard. However, on top of the mentioned safety factor, another safety factor is applied to the required ammonia concentration in the effluent (i.e. a safety factor less than 1) to guard against negative factors that could cause non-compliance with the effluent standards. The problem of double counting of sources of uncertainty could lead to an overly-conservative and possibly expensive design without necessarily providing a worthwhile benefit (Doby, 2004).

Meanwhile, the inflexibility of design guidelines hinders the design engineers to better incorporate readily-available or easy-to-obtain information on climate conditions, basic sewershed characteristics, and process dynamics into their decision on the final sizing of a WWTP. For example, according to the Metcalf and Eddy design guideline (Tchobanoglous et al., 2003) the variation of influent pollutant concentrations is taken into account by multiplying the average influent concentration of certain pollutants with a so-called peaking factor to insure that the plant is able to reach its desired treatment objectives under different dynamic flow and loading conditions. However, the variation in flow and pollutant concentrations depends on the characteristics of the sewershed and local climate conditions of the region. Therefore, relying on the peaking factors that have been developed based on the analysis of numerous sewersheds in a specific region or country could result in oversizing of some treatment units (Corominas et al., 2010).

Considering the increasingly stringent effluent limits (Vanrolleghem, 2011) that are defined in terms of both value and frequency (e.g. the daily mean concentration of a pollutant in the effluent should not exceed a certain concentration more than N times in a year) and/or to be met with data averaged over a relatively short period (e.g. like in Germany where compliance to effluent standards are usually measured by random grab or 2 hour composite samples (ATV, 2000)), the application of dynamic models provides insight in the dynamics of the effluent and meeting the effluent standards.

Contrary to previous decades during which the application of dynamic models was in its infancy and limited to the academic and research domain, these models have become quite popular in the wastewater industry and are now being used to assess the performance of WWTPs under dynamic conditions as well as determining the optimal operating conditions of plants. In addition, dynamic models enable (provided reliable models are available) designers and plant operators to test the performance of recent innovative treatment technologies whose applications are not as widespread as conventional treatment systems like activated sludge.

Moreover, dynamic models in conjunction with the Monte Carlo simulation procedure provide a tool to propagate uncertainties in model inputs and parameters to model output and hence to provide a probabilistic evaluation of meeting the effluent standards. Several researchers have reported the application of Monte Carlo simulation along with a dynamic WWTP model to assess the effect of different sources of uncertainty on the effluent (Rousseau et al., 2001; Bixio et al., 2002; Huo et al., 2006; Sin et al., 2009; Belia et al., 2012; Benedetti et al., 2013). However, in most of the studies of this kind, the main emphasis was laid on the uncertainty of model parameters while the characterization of the influent variability lacked statistical rigor (e.g. using one year-long influent time series for capturing the effect of influent variability in Benedetti et al. (2013), or not considering the

effect of rainfall time series on synthetically-generated influent time series in Rousseau et al. (2001).

Considering the shortcomings of the conventional design methods and the lack of a rigorous approach for considering the effect of relevant sources of uncertainty in dynamic simulation of WWTPs, this chapter deals with the application of the new probabilistic design method (outlined in Chapter CHAPTER 3:) to the design of an actual WWTP that provides users with a quantitative criterion to measure the degree of compliance of a proposed design to the effluent standards in terms of probability.

## **5.2 Proposed probabilistic design method**

The different steps of the proposed probabilistic design method, illustrated in Figure 5-1, were explained in detailed in Chapter CHAPTER 3:. In this section, the design methodology is briefly explained with more emphasis on its application to the case study of the current research work (see Section 5.2.4 for details on the case study). The main steps of the probabilistic design include 1) Steady state pre-designs with different levels of safety, 2) Probabilistic evaluation and screening of the pre-designs, and 3) Quantification of PONC and the total cost for the selected designs. For the case study of the current research (Section 5.2.4), certain choices on particulate methods and tools for the probabilistic design methodology are made in this section.

### ***5.2.1 Steady state pre-designs with different levels of safety***

The first step in the proposed probabilistic design methodology consists of the generation of a set of pre-designs with different levels of safety, reflected in the value of design inputs corresponding to different pre-designs (see Section 3.1.2). For this step, a common tool that

can be used is based on the assumption of steady state, e.g. a design guideline by which the size of different treatment units is calculated using a set of equations based on: (I) the mass balance equations of the system under steady state conditions, (II) empirical equations as well as (III) experience-based rules (Talebizadeh et al., 2012).

Although the sizing of a WWTP that is made according to a design guideline could have a desirable performance in meeting the effluent standards, the set of design inputs that would result in such a sizing as well as its corresponding PONC is not known a priori (i.e. it is not known which combination of design inputs would results in a design, capable of meeting the effluent standards with a tolerable level of PONC).

In order to explore different combinations of design inputs and their effect on the sizing of the plant, a uniform distribution with proper parameters (i.e. lower and upper limits) is to be assigned to each design input. Note that the effluent standards are fixed values and there is no uncertainty attached to them. Different sizings are generated by sampling different values from the distribution of all design inputs.

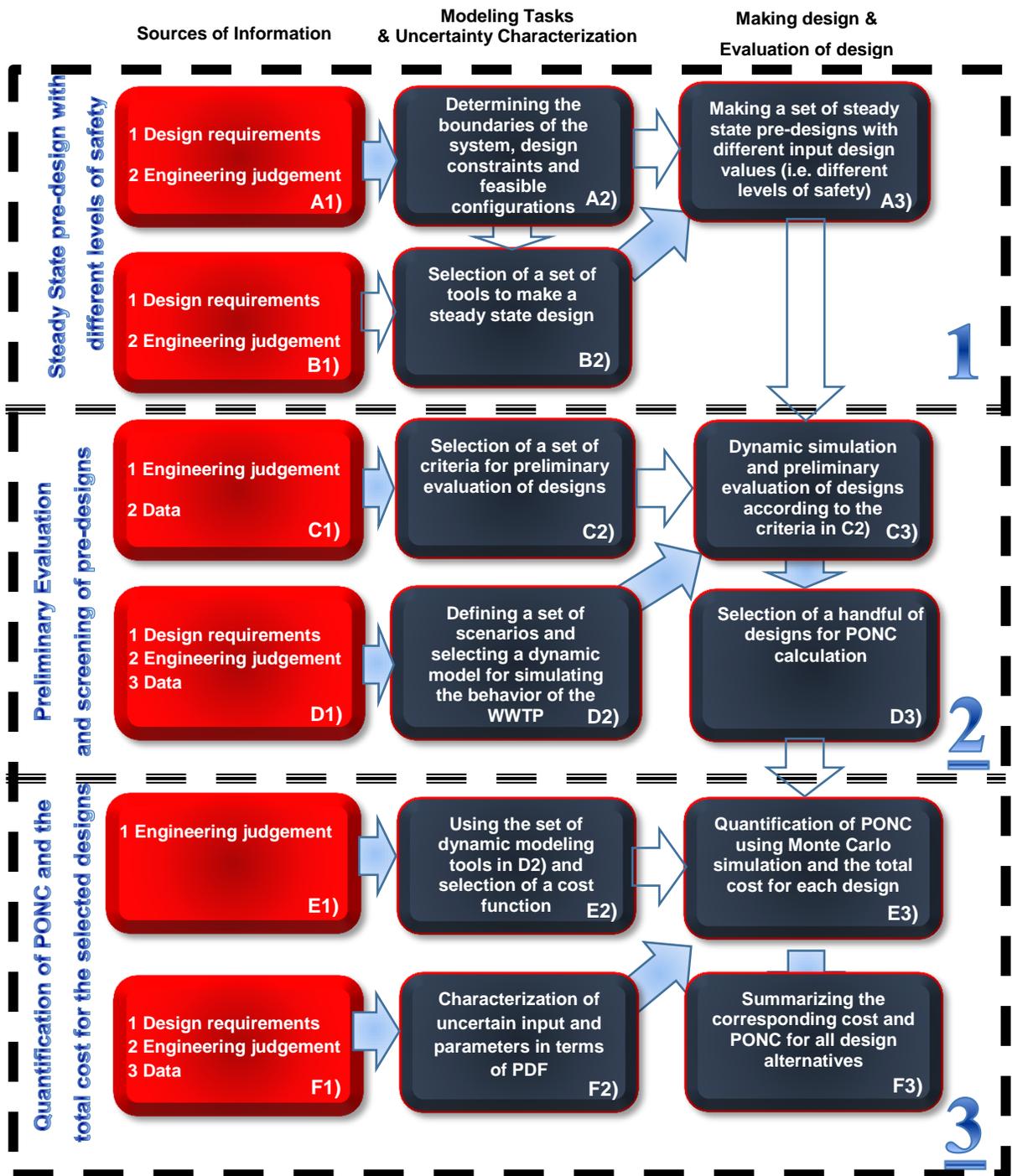


Figure 5-1 Steps for the proposed probabilistic design of WWTPs

### ***5.2.2 Preliminary Evaluation of and screening of design alternatives***

As mentioned in Section 3.2, a preliminary evaluation of pre-designs generated in the first step of the probabilistic design method should be performed so as to retain only a handful of designs that are dissimilar in size and performance in pollutant removal.

In the proposed methodology k-means clustering (Hartigan, 1979) is applied for selecting a handful of design alternatives. K-means clustering method can be used for partitioning a data set with  $n$  data points into  $k$  clusters according to a similarity measure (Wagstaff et al., 2001). The number of clusters is a parameter that should be determined prior to the clustering of data sets. Once the number of centroids is selected the clustering of a dataset can be implemented as below:

- 1) Randomly select  $k$  data point as cluster centroids
- 2) Compute the distance of each data point to the centroids
- 3) Assign each data point to the cluster with the closest centroid
- 4) Calculate the average of the data points in each cluster to obtain  $k$  new centroids
- 5) Repeat 2) through 4) until cluster assignments do not change

The explained k-means clustering algorithm with a handful of clusters is performed on the feasible design alternatives and the centroid of each cluster is selected as a design alternative for further analysis under a number of dynamic loading scenarios, deemed to be important to evaluate by designers (D2 in Figure 5-1).

For the scenario analysis, a year-long influent time series must be generated using the influent generator in Chapter CHAPTER 4: and the performance of each design alternative is evaluated under the generated influent time series using a process model. The decision on the type of dynamic model for simulation of designs under dynamic conditions could

depend on several factors including previous modeling experience, availability of data, as well as the wastewater constituents whose effluent concentrations are of interest. In case of an upgrade projects, the parameters of the process model is set based on previous plant simulation and measurements. Whereas in the case of a greenfield design project, designers may set the values of plant model parameters to default values. The important point is that the dynamic model should be able to provide answers on the typical performance of designs regarding the removal of those effluent constituents for which the compliance to the effluent standards has to be assessed.

To compare the performance of the different design alternatives the cumulative distribution function (CDF) corresponding to the simulated effluent of each design alternative is constructed (Figure 3-5). The CDFs are used as means for filtering those design alternatives that have a poor performance in terms of effluent quality. In addition, the comparison of the CDFs can serve as a tool for removing those design alternatives that have the same treatment performance and calculating their PONC would impose a large computational load without adding any worthwhile information.

### ***5.2.3 Quantification of PONC and Total Cost***

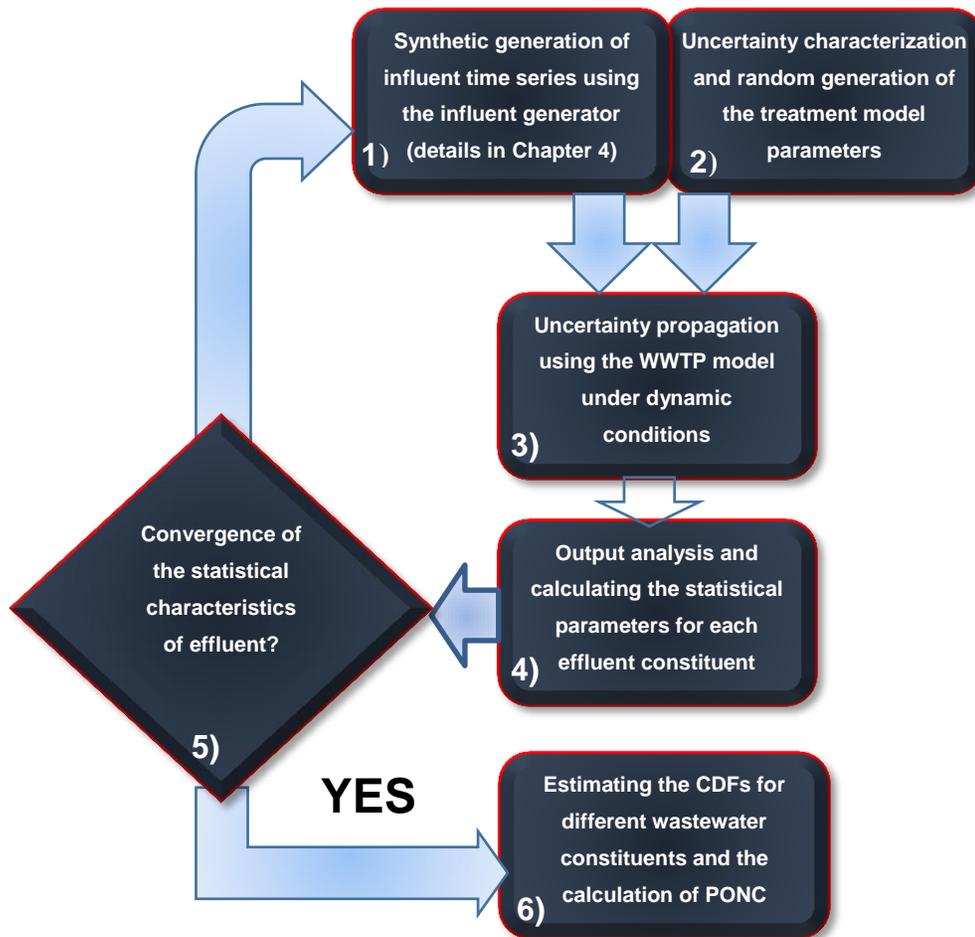
The quantification of PONC constitutes the most innovative part of the proposed probabilistic design method. In this section the PONC will be estimated for the design alternatives that have passed the preliminary evaluation (those design alternatives that have a good performance in terms of effluent quality under the year-long generated influent time series and are not considered redundant in that their performance are different from other design alternatives). The details of the methodology for PONC calculations are already explained in Section 3.3. This section provide a summary of the main steps of the PONC calculation to set the stage for Section 5.3 which contains the results of applying the

proposed methodology to the case study of this research work (see Section 4.3 for details on the case study).

As discussed in Section 3.3.1, the proposed influent generator in Chapter CHAPTER 4: is used for synthetic generation of influent time series (step (1) in Figure 5-2). The main reason for using the influent generator in Chapter CHAPTER 4: was that it provides a realistic view of the dynamics of influent time series using the basic local climate data and characteristics of the sewershed under study. Therefore, its application to the probabilistic design of WWTPs is expected to result in realistic calculation of PONC. The uncertainty in model parameters is characterized by a set of PDFs reflecting the knowledge on the value of model parameters that could be acquired based on a combination of lab analysis of wastewater, previous modeling studies, as well as expert opinion. Once the uncertainty in model parameters is characterized, different realizations of vector of model parameters can be sampled according the selected sampling method (step (2) in Figure 5-2, explained in I, Section 3.3.2).

The effect of model parameter uncertainty and influent variability can be propagated using different Monte Carlo simulation methods (step (3) in Figure 5-2, explained in Section 3.3.3). However, as mentioned in part II of s

Section 3.3.3, the application of two-dimensional Monte Carlo simulation for uncertainty propagation is not feasible without resorting to more efficient computing methods. Therefore, two practically feasible schemes for propagation of uncertainty and variability (i.e. the one-dimensional Monte Carlo simulation with mixing uncertainty and variability (I in Section 3.3.3), and the pragmatic Monte Carlo method (III in Section 3.3.3) are used.



**Figure 5-2** Calculating PONC for a design alternative

In order to facilitate running the dynamic model of each design alternative with different values of model parameters as well as different realizations of the influent time series, the Tornado kernel of the WEST software (Claeys et al., 2006) is used. Tornado is an advanced kernel for modeling and virtual simulations that offers users a great degree of flexibility as well as execution time reduction compared to the WEST software (Claeys et al., 2006; Benedetti, 2006; Benedetti et al., 2008). Tornado also allows users to couple it with several programming languages which could be very useful for post processing of model outputs.

The simulated effluent time series must be aggregated to the same temporal resolution as the one according to which compliance to effluent standards is measured (step (4) in Figure 3-6, explained in Section 3.4). The convergence of the Monte Carlo simulation is assessed by calculating the effluent CDFs and analyzing the changes for increasing numbers of simulations in the values of representative percentiles corresponding to different effluent CDFs (step (5) in Figure 3-6, explained in Section 3.4). Once convergence is achieved, the simulations are stopped and the corresponding PONC and the total cost are calculated for each design alternative (step (6) in Figure 3-6, explained in Section 3.5).

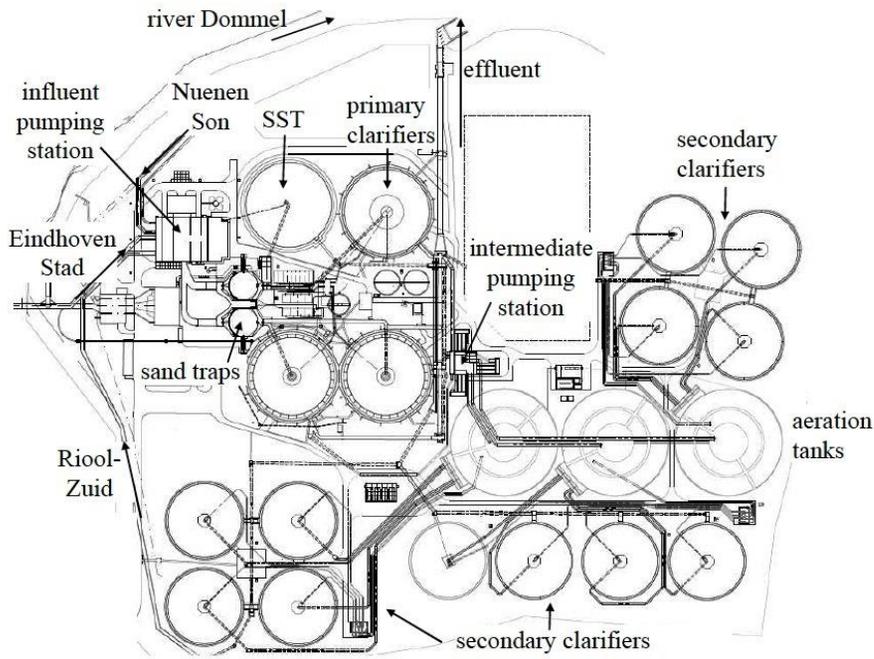
The CAPDET software (Harris et al., 1982; Hydromantis Inc, 2014) is used as a tool for calculating the total cost of each design alternative. The CAPDET software is capable of calculating the total cost of a plant using the basic global parameters (e.g. basic influent and flow values) and process configuration. To simplify the task of providing the input variables for the calculation of cost, default values based on the recorded data sets are suggested which can be used in case of limited availability of data. Moreover, the flexibility of the software allows users to easily change the default values of some parameters to better reflect the site-specific conditions into the calculation of cost. At the end, the calculated total cost values associated to different design alternatives is plotted against their corresponding PONC to better illustrate the relationship between these two indices which can help designers to better identify the design alternatives that have a reasonable cost while meeting the effluent standards with a tolerable PONC.

#### ***5.2.4 Data and case study***

The Eindhoven WWTP with a design capacity of 750,000 population equivalent (PE) is the third largest WWTP in the Netherlands. Between 2003 and 2006 the Eindhoven WWTP underwent an upgrade project to comply with new nutrient removal standards (Schilperoort, 2011). The upgrade of the WWTP was deemed necessary mainly due to more stringent

effluent limits (i.e. Table 5-1) and also the need to increase the hydraulic capacity from 20,000 m<sup>3</sup>/hr to 26,250 m<sup>3</sup>/hr (Belia et al., 2012).

Wastewater entering the plant is screened by 25mm and 6 mm bar screens and de-gritted by two parallel grit chamber (400 m<sup>2</sup> each) before further primary treatment in three primary clarifiers (9,565 m<sup>3</sup>, and the surface area of 2,828 m<sup>2</sup> each). The plant capacity from the influent pumping station up to the primary clarifiers is 350,000 m<sup>3</sup>/h. However, the permissible flow rate to the bioreactors for secondary treatment is 26,250 m<sup>3</sup>/h. During higher flow rates, the surplus flow is sent to a storm storage tank (SST) with a volume of 8,750 m<sup>3</sup>. The biological treatment comprises of 3 activated sludge tank with anaerobic, anoxic, and aerated zones. Each individual activated sludge tank is connected to 4 secondary clarifiers and the final effluent is discharged to the Dommel River (Figure 5-3).



**Figure 5-3** Aerial view and the schematics of of the Eindhoven WWPT (Schilperoort, 2011)

In this study the data and information representative of conditions prior to the upgrade of the Eindhoven WWTP were used to make a design using the proposed probabilistic design method in Chapter CHAPTER 3:. The general information and data regarding the desired capacity of the plant, effluent requirements, as well as the basic characteristics of the connected sewershed were acquired from the available reports and studies on the Eindhoven WWTP (Schilperoort, 2011; Belia et al., 2012).

<b>Table 5-1</b> Effluent standards			
Effluent composition	Effluent limit	Unit	Sampling for measuring compliance
Total nitrogen (TN)	10	mg/l	Annual mean
NH <sub>4</sub> -N	2	mg/l	Flow-proportional daily mean
BOD <sub>5</sub>	20	mg/l	Flow-proportional daily mean
COD	125	mg/l	Flow-proportional daily mean
TSS	10	mg/l	Annual mean

Influent data that were used for the development of the influent generator of the Eindhoven WWTP (presented in detail in Chapter CHAPTER 4:) are related to sensor data of flow, ammonia (measured using an ion-selective sensor at the entrance point to the treatment plant), soluble COD, total COD and TSS (the latter 3 measured using an UV/VIS-based sensor with a frequency of 1 per minute) in the period from September 2011 to September 2012 at the outlet of the sewersheds. Long-term daily rainfall data and also rainfall data with finer temporal resolution provided by KNMI (Royal Netherlands Meteorological Institute) and Waterschap De Dommel were used for estimating the parameters of the weather generator.

## 5.3 RESULTS

This section presents the results of applying the proposed probabilistic design methodology to the case study of this research. The results corresponding to the three main steps of the proposed methodology are covered in the following sections: 1) Steady state pre-designs with different levels of safety, 2) Preliminary evaluation and screening of pre-designs, and 3) Quantification of PONC and the total cost for the selected designs.

### 5.3.1 *Steady state pre-designs with different levels of safety*

In the current research work (see Section 5.2.4 for details on the case study) the German “Standard ATV-DVWK-A 131E, Dimensioning of Single-Stage Activated Sludge Plants” (ATV, 2000) was used as a steady state design tool (the actual design was based on the ATV and STOWA design guidelines). Depending on the available information and data a uniform uncertainty range was assigned to each input of the ATV design guideline and the design outputs were generated by Monte Carlo simulation. The ranges of uncertainty depicted in Table 5-2, were derived using the information obtained from the previous studies on the Eindhoven WWTP (i.e. Schilperoort, 2011; Belia et al. 2012), ATV (2000) design guideline recommendations, effluent standard (imposed by regulations), and expert opinion. An important variation among the size of the process units of the different design alternatives was found, reflecting the effect of the range in ATV design inputs on the dimensions of the generated design alternatives. For example, sampling a high value of BOD concentration (i.e. close to  $150\text{ mg/l}$ , the upper limit for concentration of BOD in the primary effluent in Table 5-2) with a high daily flow (i.e. close to  $180,000\text{ m}^3/\text{day}$ ) would result in a large BOD load in the primary effluent, which in turn result in a large total bioreactor volume.

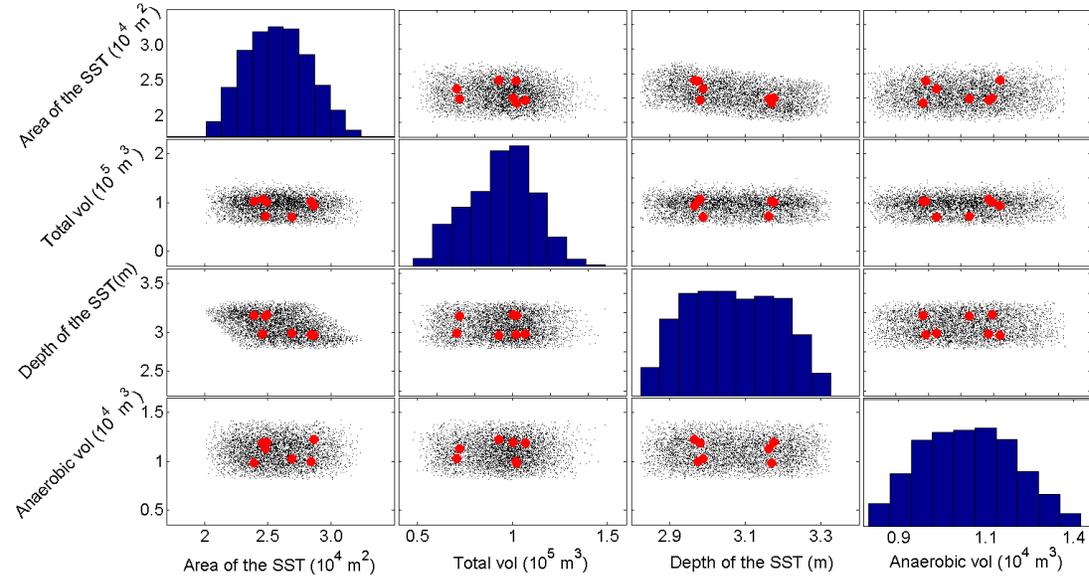
**Table 5-2** Range of values assigned to the ATV design guideline (uniform distribution)

ATV design inputs	Lower limit	Upper limit	Units
<b>Influent constituents</b>			
Concentration of COD in the primary effluent <sup>1,2</sup>	200	400	mg / l
Concentration of Nitrogen in the primary effluent <sup>1,2</sup>	30	50	mg / l
Concentration of Phosphorous in the primary effluent <sup>1,2</sup>	5	7	mg / l
Concentration of BOD in the primary effluent <sup>1,2</sup>	100	150	mg / l
Concentration of TSS in the primary effluent <sup>1,2</sup>	60	100	mg / l
Daily flow (i.e. flow value used for the calculation of design BOD load) <sup>1,2</sup>	150,000	180,000	m <sup>3</sup> / day
Max. hourly dry weather flow rate as 2 hour mean <sup>1,2</sup>	20,000	26,250	m <sup>3</sup> / hr
Max. hourly wet weather flow rate as 2 hour mean <sup>1,2</sup>	45,000	65,000	m <sup>3</sup> / hr
Percentage of inert particulate COD in particulate COD <sup>3</sup>	0.2	0.35	---
Percentage of inert soluble COD in COD <sup>3</sup>	0.05	0.1	---
Percentage of inorganic TSS in the total TSS <sup>3</sup>	0.2	0.3	---
Ratio of nitrogen in the biomass to the BOD concentration in the primary effluent <sup>3</sup>	0.04	0.05	N / BOD
<b>Safety factors</b>			
Safety factor for nitrification <sup>3</sup>	1.45	1.5	---
Safety factor applied to the effluent inorganic nitrogen <sup>3</sup>	0.6	0.8	---
Safety factor applied to the effluent phosphorous <sup>3</sup>	0.6	0.7	---
<b>Required parameters</b>			
Sludge Volume Index (SVI) <sup>1,2</sup>	100	120	l / kg
<b>Operation parameters</b>			
Permitted sludge volume loading rate on SC <sup>3</sup>	300	350	l / m <sup>2</sup> .hr
Sludge thickening time in the SC <sup>3</sup>	1.5	2	hr
Minimum contact time in anaerobic tanks <sup>1,2</sup>	0.9	1.1	hr
<b>Effluent concentrations</b>			
Total nitrogen concentration in the effluent <sup>4</sup>	10	10	mg / l
Phosphorous concentration in the effluent <sup>4</sup>	1	1	mg / l
1: Expert opinion 2: Previous studies on the Eindhoven WWTP 3: ATV design guideline 4: Effluent standards			

In this study the dimensions (i.e. the total volume of the bioreactors (anoxic + aerated), the anaerobic volume, the area and depth of the secondary clarifier (SC)) of 5000 pre-designs were generated (Figure 5 3) by random sampling of 5000 sets of ATV inputs from the ranges of values in Table 5 2 and making a design for each sampled set of inputs. Performing a formal convergence test on the distribution of design outputs (to check the

sufficiency of the number of sampling) was not necessary as the computational time for generating a large number of pre-designs (using the ATV design rules and equations coded inside Matlab2013) was very short and increasing the number of samplings (i.e. more than 5000 samples) did not result in any significant change in the distribution of outputs.

As explained in Section 5.2.2, a k-means clustering was performed on the generated ATV outputs to reduce the number of design alternatives that are to be evaluated under dynamic conditions. In this study the number of clusters was set to 7 as it turned out that 7 design alternatives would result in a good balance between selecting enough number of alternatives that are representative of the space of the design outputs and maintaining the computational load at a manageable level (i.e. setting the number of clusters to a value more than 7 would increase the computation time and identification of cluster centroids that are close to those corresponding to 7 clusters. Therefore, it was decided to set the number of clusters to 7, and add more designs manually after preliminary evaluation of 7 clusters (to insure proper coverage of the design outputs).



**Figure 5-4** Distribution of the generated 5000 pre-designs and the centroids locations corresponding to the k-means clustering with 7 centroids (i.e. the red dots)

The histograms in Figure 5-4 represent the distribution of design outputs that were generated according to the ATV design guidelines, while the red dots represent cluster centroids that were calculated using the k-means clustering method.

The scatter plots shown in Figure 5-4 were used to better evaluate the location of the cluster centroids (the 7 red dots) in the space of the design outputs (the design outputs are summarized in Table 5-3). Figure 5-4 suggests that the data points (design outputs) are more concentrated in certain regions of the output space (i.e. higher bins close to the average value of each design outputs or darker regions in the scatter plots). Since the distance of data points to their corresponding cluster centroids plays an important role in the calculated location of cluster centroids, it is not unexpected to see that the cluster centroids (the red dots in Figure 5-4) are located in regions with a high concentration of data points. While, designers may be interested in checking the performance of specific design alternatives that not necessarily correspond to the centroids (e.g. design alternatives that are close to the maximum and minimum ranges of design outputs) were added in this study to better cover the entire design space.

**Table 5-3** Dimensions of the preliminary selected design alternatives (cluster centroids)

Design alternatives	Total volume (m <sup>3</sup> )	Anaerobic volume (m <sup>3</sup> )	Depth of the secondary clarifier <sup>1</sup> (m)	Area of the secondary clarifier (m <sup>2</sup> )
A	70 650	10 250	3.0	26 900
B	72 200	11 250	3.2	24 850
C	92 700	12 200	3.0	28 650
D	100 350	11 950	3.2	24 950
E	101 900	9 950	3.0	28 450
F	102 300	9 850	3.2	23 950
G	106 650	11 850	3.0	24 600

<sup>1</sup>Having only two values for depth of the secondary clarifier (i.e. 3.0 and 3.2) is due to the small ranges of values (2.8-3.3) that have been calculated in Figure 5-4, which in turn is the result of the range of values assigned to the inputs of the ATV design guideline (Table 5-2).

### 5.3.2 Preliminary evaluation and screening of pre-designs

The performance of the seven design alternatives in Table 5-3 was evaluated under a year-long generated influent time series using the calibrated influent generator in Chapter CHAPTER 4: (i.e. influent model parameters fixed at the maximum likelihood vector of model parameters) and the WEST software<sup>1</sup> with the ASM2d model (Henze et al., 1999) for simulating the bioprocesses and the secondary settling model by Bürger et al. (2011).

The simulated effluent time series were then aggregated to mean daily values, and their corresponding CDFs were constructed. Figure 5-5 shows the CDF graphs corresponding to the effluent time series for four constituents (COD, NH<sub>4</sub>-N, TSS and TN).

The CDFs for the effluent COD are similar to each other as all of the design alternatives are capable of easily meeting the COD effluent standards (i.e. daily mean of 125 mg/l according to Table 5-1). It should be noted that the compliance of TSS and TN to the effluent standards are measured using yearly mean concentration values (Table 5-1). Therefore, the mean daily concentration values corresponding to TSS and TN in Figure 5-5, cannot be directly compared to the effluent standard values for TSS and TN. Nevertheless, the CDFs in Figure 5-5 can be used as a tool for a comparative analysis regarding the performance of different designs in TSS or TN removal based on the CDFs of daily mean concentration values.

---

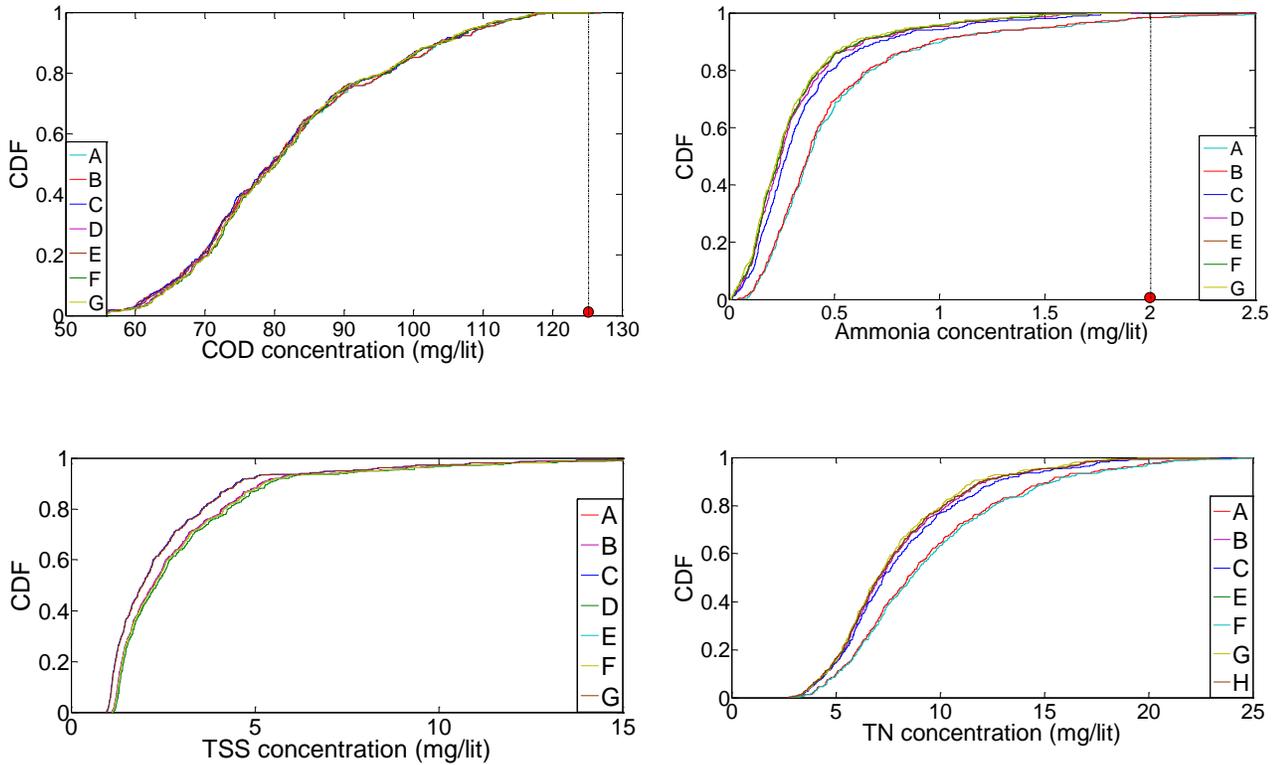
<sup>1</sup> The WEST model of the Eindhoven WWTP, illustrated in Figure A2, was provided by Youri Amerlinck from Ghent University. The parameters of the model reflecting the different treatment unit sizes in Table 5-3 were adjusted to create a dynamic model for each design alternative.

As indicated in Table 5-3, there is no noteworthy variation in the different design alternatives as far as the calculated depth and area of the secondary clarifier. All of the secondary clarifiers have a favorable performance for the TSS removal (i.e. even if the compliance to the TSS effluent standard (i.e. 10 mg/l according to Table 5-1) were to be measured based on daily mean concentration values, all of the design alternatives would meet the effluent standards more than 90% of days).

On the contrary, the effluent ammonia and total nitrogen (TN) CDFs show nitrogen removal that depends on the design alternative. As indicated, the effluent ammonia and TN concentrations for Alternative A and Alternative B are consistently higher compared to the other design alternatives. Comparing the effluent ammonia CDFs for Alternatives A and B suggests that these two design alternatives have an almost similar performance in ammonia removal. Moreover, the CDFs of Alternative C, D, E, F, and G suggest that there is no significant difference in the performance of ammonia removal among these five design alternatives. Based on the shape of the CDFs, two different groups of design alternatives (one group containing Alternative A and Alternative B and the other group containing Alternative C, D, E, F, and Alternative G) can be distinguished. Therefore, two design alternatives namely Alternative A and Alternative G from Table 5-3 that correspond to two cluster centroids were selected for further analysis to make sure that the PONC would not be calculated for alternatives that have similar treatment performance.

As mentioned in the methodology section, the design alternatives for which the PONC can be calculated do not need to be limited to cluster centroids (Table 5-3) as the designers might be interested in the calculation of PONC for other design alternatives. Given the importance of the total volume on the performance of nitrogen removal (as it provides more capacity for the system to insure the minimum SRT), it was decided to select three more designs from the 5000 design alternatives that were generated according to the ATV design guideline. To this end, two design alternatives with total volumes that are respectively 16%

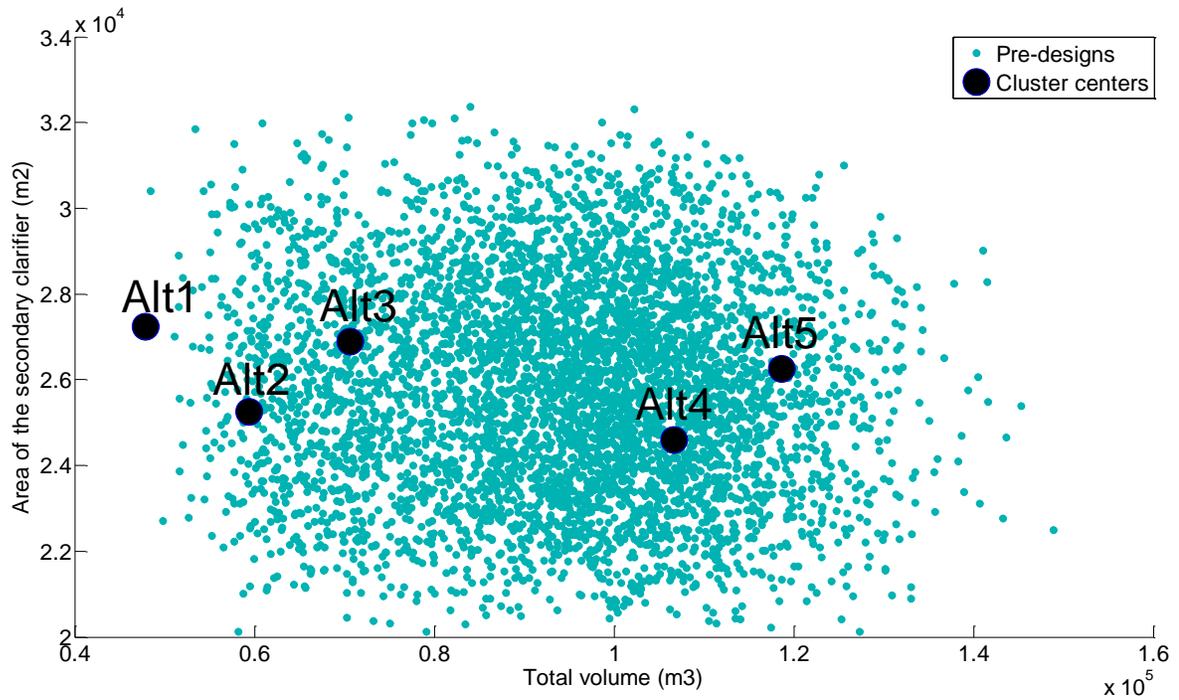
and 32% smaller (Alt2 and Alt1 in Table 5-4) and a third design alternative with total volume approximately 68% larger than the total volume of Alternative E (Alt5 Table 5-4) were selected for further analysis. Alt3 and Alt4 in Table 5-4 refer to Alternatives A and G of Table 5-3. The location of the final design alternatives in the space of design outputs are illustrated in Figure 5-6.



**Figure 5-5** CDF graphs for the effluent of design alternatives

**Table 5-4** Dimensions of the final design alternatives

Design alternatives	Total volume (m <sup>3</sup> )	Anaerobic volume (m <sup>2</sup> )	Depth of the secondary clarifier (m)	Area of the secondary clarifier (m <sup>2</sup> )
Alt1	47 850	12 200	3.1	27 250
Alt2	59 400	11 100	3.0	25 250
Alt3 (centroid cluster A)	70 650	10 250	3.0	26 900
Alt4 (centroid cluster G)	106 650	11 850	3.0	24 600
Alt5	118 700	9 500	3.1	26 250
Actual design	79 200	11 200	2.5	21 696



**Figure 5-6** Location of final design alternatives in the space of design outputs

### 5.3.3 Quantification of PONC and the total cost for the selected designs

#### I. Synthetic generation of influent time series

Different realization of the influent time series were generated using the proposed influent generator in Chapter CHAPTER 4:. The results for the application of the proposed influent generator for generating different realizations of the influent time series are already reported in Section 4.4. The generated influent time series are stored in different files and depending on the uncertainty propagation schemes that are used (i.e. one-dimensional with influent variability and model parameter uncertainty mixed, and the pragmatic Monte Carlo simulations), different dynamic influent files are inputted to the model.

## ***II. Uncertainty characterization and random generation of WWTP model parameters***

In this study the uncertainty in the model parameters was characterized (i.e. (2)) in Figure 3-6 which is thoroughly explained in 3.3.2) by assigning uniform distributions to uncertain model parameters. A combination of expert opinion and modeling experience was used for assigning proper uncertainty ranges to uncertain model parameters. The lower and upper limits of uncertainty for each parameter were calculated using a nominal value multiplied by a percentage of the nominal value for each uncertain model parameter. The relative uncertainty that was expressed in terms of a certain percentage is based on the work of Brun et al. (2002) in which the uncertainty of the ASM2d model parameters is characterized by uniform distributions. Random sampling (RS) with no-correlation (see Section 3.3.2 for details on different sampling methods) was selected for sampling from the distribution of uncertain model parameters. The choice of random sampling of model parameters was based on the study of Hauduc et al. (2011) in which no strong correlation was reported between the parameters of the ASM2d model.

Table A3 gives the lower and upper limits of uniform distributions used for uncertainty characterizations of the WWTP's uncertain model parameters, the nominal vector of model parameters (based on previous modeling studies of the case study), as well as the “worst case” vector of model parameters (obtained using the methodology explained in part III of Section 3.3) that were used in the pragmatic Monte Carlo simulations. It should be noted that the “worst case” vector of model parameters in Table A3 corresponds to 95% confidence for the NH<sub>4</sub> effluent standards (i.e. 2 mg/l in Table 5-1) and a higher than 95% for other pollutant concentrations. Therefore, the “worst case” vector of model parameters in Table A3 not only represents a difficult condition for NH<sub>4</sub> removals but also for other pollutants as well (more details can be found in part III of Section 3.3).

### ***III. Propagation of uncertainty and output analysis***

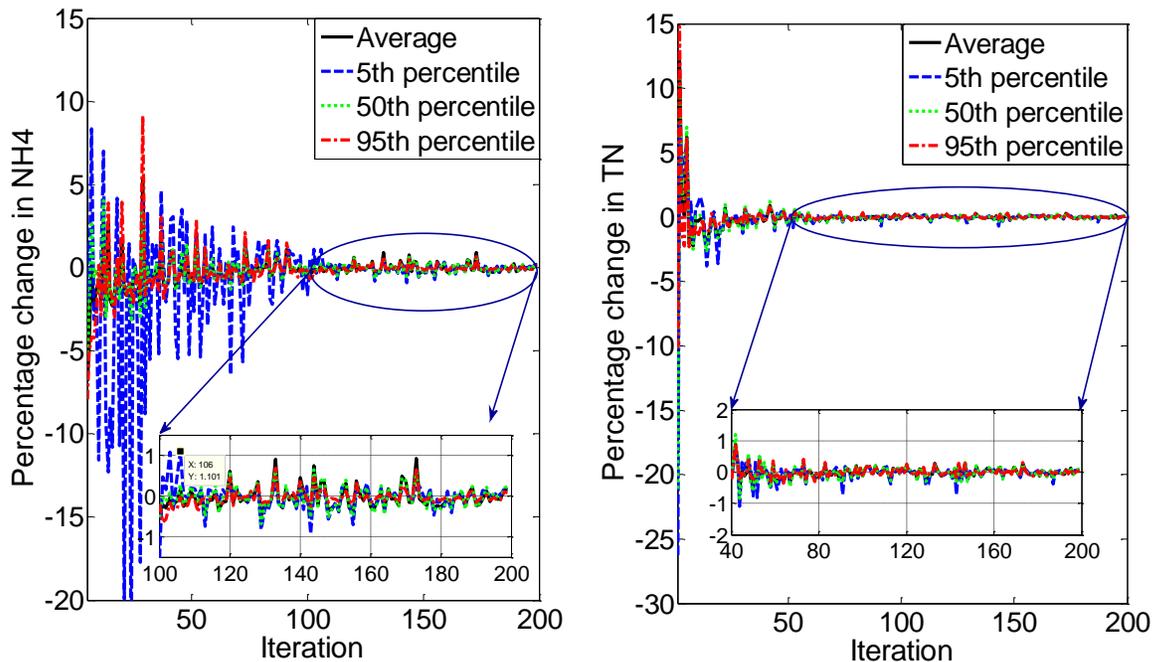
As mentioned in part III of Section 5.2.3, two uncertainty propagation methods including the one-dimensional Monte Carlo simulation (for more details see *I*, Section 3.3) and the pragmatic Monte Carlo simulation (more details can be found in *III*, Section 3.3) were applied for propagation of uncertainty and variability.

As mentioned in Section 3.4, the simulated effluent time series obtained using dynamic simulation may require some time series aggregation prior to the convergence test and calculation of PONC. Therefore, the simulated effluent time series of BOD, COD, and NH<sub>4</sub> (with 15-minute temporal resolution) were aggregated to daily values and the time series of effluent TN and TSS to yearly values so as to match the sampling method used for measuring compliance to the effluent standards (see Table 5-1 for the effluent standards, imposed by regulations).

Monte Carlo runs should continue until the convergence of the statistical distributions of different effluent constituents (i.e. step 5) in Figure 5-2. See Section 3.4 for details on the convergence of the effluent distributions). In this study, the fluctuation in the average, 5<sup>th</sup>, 50<sup>th</sup>, 95<sup>th</sup> percentiles was evaluated at each Monte Carlo run, and simulations continued until the relative changes in the values of these four statistics drop below 1% (Figure 3-14).

Figure 5-7 shows the fluctuations in the percentage change values (i.e. the relative difference between the value of a certain percentile, calculated in simulation *i* and *i*+1, expressed in percentage) for average, 50<sup>th</sup>, 5<sup>th</sup>, and 95<sup>th</sup> percentiles of the effluent NH<sub>4</sub> and TN concentrations (aggregated to mean daily and yearly values, respectively). As indicated in Figure 5-7, the percentage changes corresponding to the different percentiles drop below 1% (a small threshold value for determining convergence) after 108 Monte Carlo runs for NH<sub>4</sub> distribution and after 40 runs for the TN distribution. The same analysis for the other

effluent constituents showed that running 108 simulations was enough to achieve convergence for other effluent constituents (i.e. BOD, COD, and TSS).



**Figure 5-7** Fluctuations in percentage change values of the average, 5<sup>th</sup>, 50<sup>th</sup>, and 95<sup>th</sup> percentiles for the effluent NH<sub>4</sub> and TN distributions

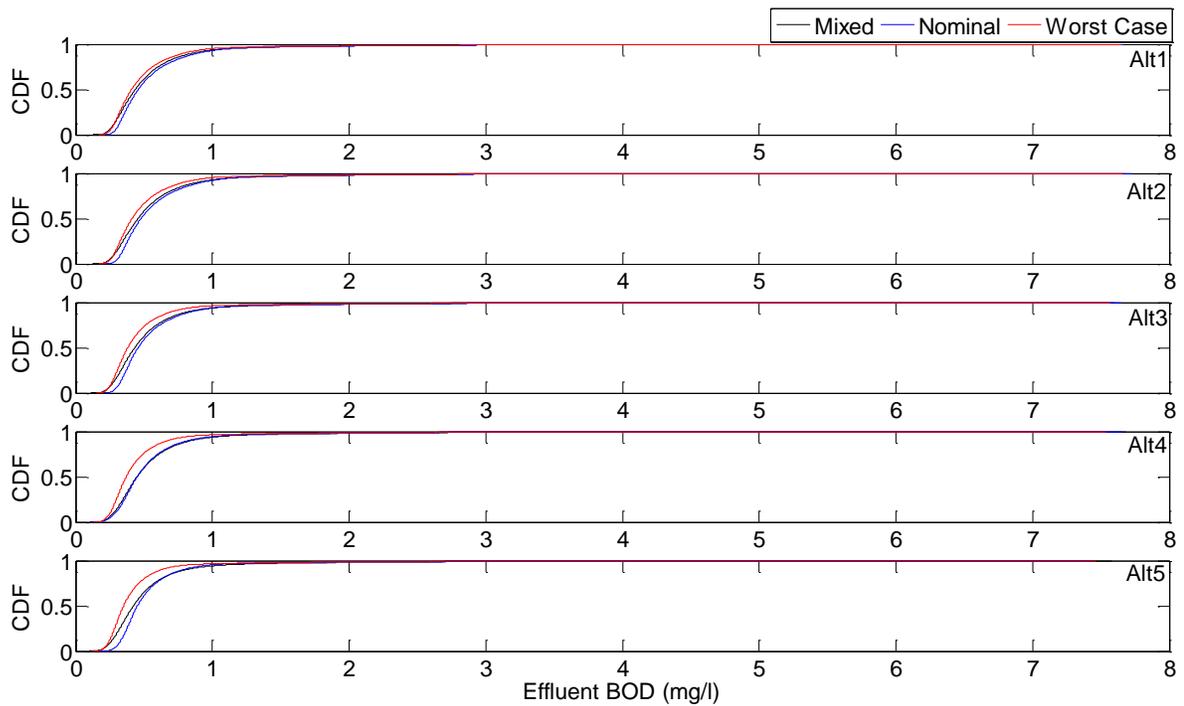
Once the convergence of the effluent distributions was insured, the CDFs of the different effluent constituents were derived and their corresponding PONC values were calculated given the effluent standards of Table 5-1.

#### **IV. PONC values for BOD, COD, and TSS**

Figure 5-8 and Figure 5-9 show three CDFs (i.e. Mixed, for one-dimensional Monte Carlo simulation, detailed in part I of Section 3.3.3; Nominal and Worst Case for the pragmatic Monte Carlo simulation, detailed in part III of Section 3.3.3) calculated from running the one-dimensional Monte Carlo simulation and the pragmatic Monte Carlo simulation corresponding to the design alternatives in Table 5-4. The CDFs in Figure 5-8 and

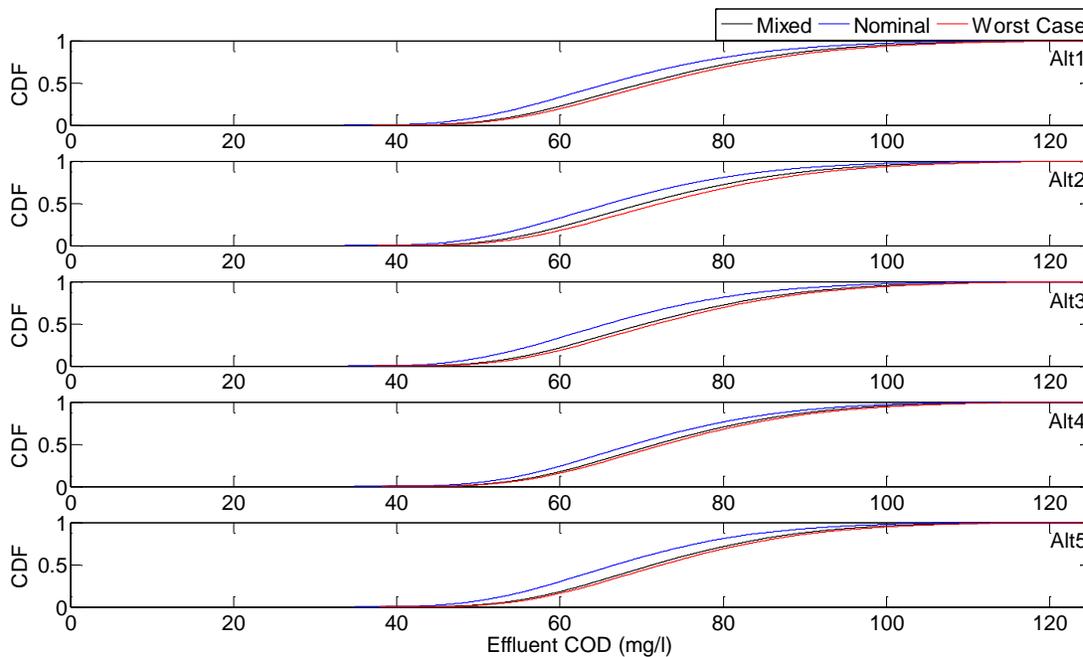
Figure 5-9 corresponding to the effluent BOD and COD distributions were constructed using 108 years of daily aggregated values (i.e. overall 108×365 concentration values).

The black CDF corresponds to the result of the one-dimensional Monte Carlo simulation, in which no distinction is made between the propagation of uncertainty and variability (i.e. Mixed). The blue and red CDFs correspond to the results of the pragmatic Monte Carlo simulation, in which the distribution of aggregated effluent values is calculated at the nominal (i.e. blue CDFs) and a “worst case” vector of model parameters (i.e. the CDFs in red).



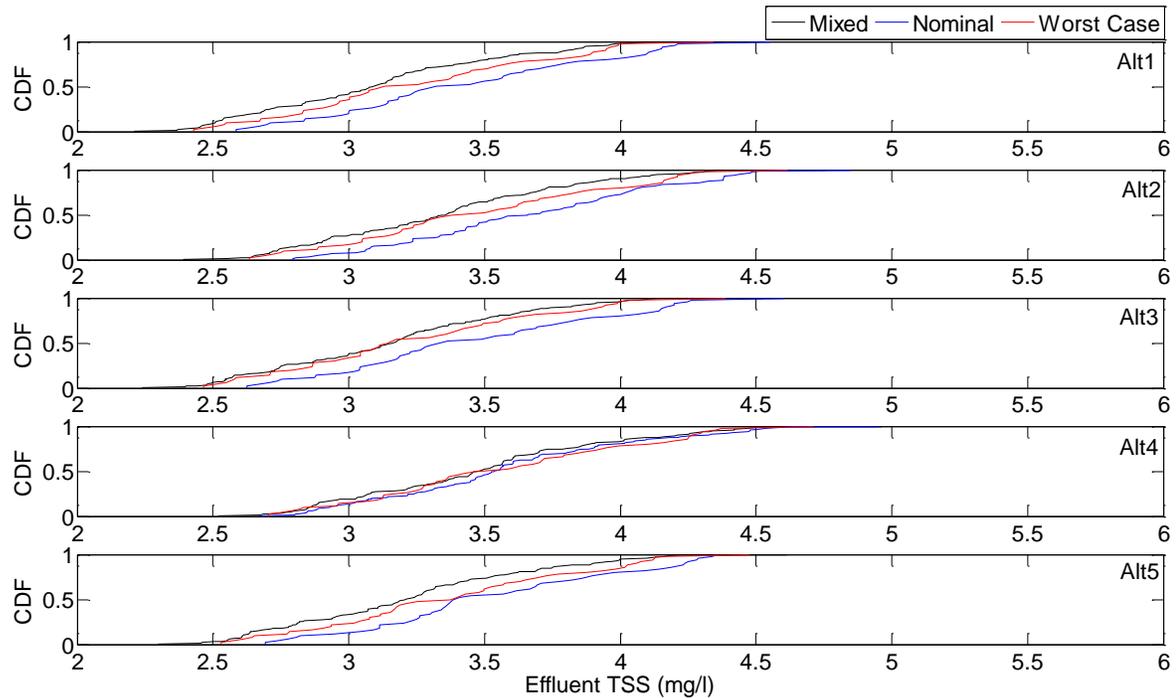
**Figure 5-8** Daily mean effluent BOD distributions for 5 different design alternatives for one-dimensional Monte Carlo (i.e. influent variability and parameter uncertainty mixed) and the pragmatic Monte Carlo simulation (i.e. influent variability with model parameters set at the nominal and “worst case” parameter sets)

All of the design alternatives have almost the same performance in removing carbonaceous constituents (measured in terms of BOD and COD) with zero or negligible PONC (PONC values are reported in Table 5-5) for the effluent standards of 20 mg/l and 125 mg/l, respectively (see Table 5-1). This comes as no surprise since plants that are designed for biological nitrogen removal have a higher sludge residence time (SRT) than the ones that are designed for biological removal of carbonaceous material only. Therefore the system designed for nitrogen removal will usually be large enough for removing carbonaceous constituents. In addition, the absence of a significant difference between the nominal and “worst case” CDFs could be attributed to the selection of the “worst case” vector of model parameters based on the  $\text{NH}_4$  removal (see part II of Section 5.3.3 for the “worst case” vector of model parameters).



**Figure 5-9** Daily mean effluent COD distributions for 5 different design alternatives (Alt1-Alt5) for one-dimensional Monte Carlo (i.e. influent variability and parameter uncertainty mixed) and the pragmatic Monte Carlo simulation (i.e. influent variability with model parameters set at the nominal and “worst case” parameter sets)

Unlike the CDFs for the effluent BOD and COD, the CDF curves related to the effluent TSS concentrations in Figure 3-10 uses the yearly-averaged effluent TSS concentrations as the effluent standards for TSS are specified in yearly-averaged effluent values (Table 5-1). Again, all of the design alternatives are capable meeting the effluent standard concentration for TSS (i.e. 10 mg/l, see Table 5-1) with zero PONC values (Table 5-5), based on the three CDFs for each design.



**Figure 5-10** Yearly mean effluent TSS distributions for 5 different design alternatives (Alt1-Alt5) for one-dimensional Monte Carlo (i.e. influent variability and parameter uncertainty mixed) and the pragmatic Monte Carlo simulation (i.e. influent variability with model parameters set at the nominal and “worst case” parameter sets)

#### ***V. PONC values for NH<sub>4</sub> and TN***

The largest difference among the design alternatives is in their nitrogen removal performance. As illustrated in Figure 5-11, the CDF for effluent ammonia is directly affected by the total size of the bioreactor (Figure 5-11, top), i.e. an increase in the value of

the bioreactor volume results in a reduction of PONC (see Table 5-5 for tabulated PONC values). Comparing the CDFs corresponding to the design alternatives suggests that the three CDFs tend to move to the left (hence a smaller PONC) as bioreactor volumes increase.

For example, based on the one-dimensional Monte Carlo simulation, the calculated PONC for alternative Alt1 with the smallest bioreactor volume equals 0.0018, whereas the PONC value for Alt5 with the largest bioreactor equals 0.0006 (three times smaller PONC). To better communicate the concept of PONC, the PONC values for those effluent constituents whose standards are specified in daily average values (COD, BOD, and  $\text{NH}_4$ ) were multiplied by 365 to calculate the expected number of days in non-compliance throughout a year (Table 5-5). For the one-dimensional Monte Carlo simulation explained in *I*, Section 3.3.3 in which variability and uncertainty are mixed, the difference between the expected number of  $\text{NH}_4$  standards non-compliance days for alternative Alt5 (largest bioreactor volume) and alternative Alt1 (smallest bioreactor volume) is around 17 days.

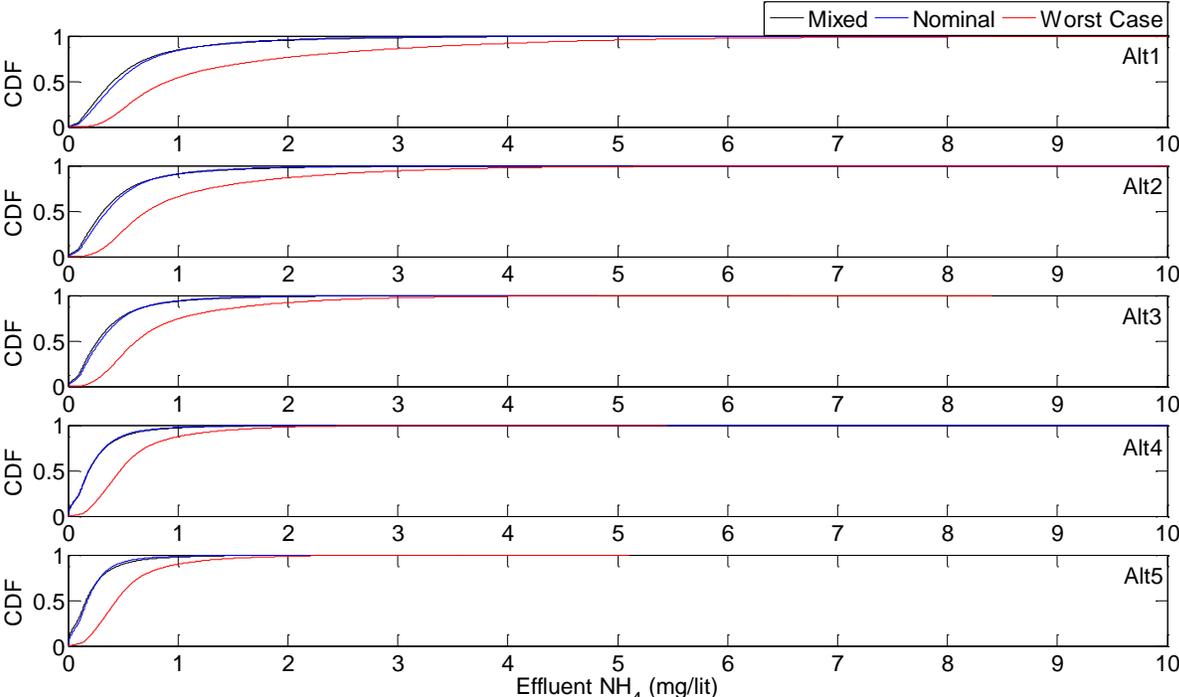
A comparison between the PONC values calculated using the pragmatic Monte Carlo procedure confirms the effect of the bioreactor volume on the values of PONC 15 and 81 days difference between the expected number of non-compliance days for Alt1 and Alt5 calculated at the nominal and “worst case” values for the uncertain model parameters.

As indicated in Figure 5-11, the shapes of the CDFs corresponding to the one-dimensional Monte Carlo simulation (see *I*, Section 3.3.3) are very close to the CDFs resulting from running the pragmatic Monte Carlo simulation at the nominal model parameter values. It might be difficult to provide a solid argument for justifying the similarity between the two CDFs (resulting from running a one-dimensional with variability and uncertainty mixed and the ones resulting from the pragmatic Monte Carlo method with consideration of influent variability but fixing model parameters at their nominal values) as the assumptions

underlying the two methods are different (see Section 3.3.3). For example, attributing the similarity of the two CDFs to a larger importance of influent variability compared to model parameter uncertainty is not a justifiable argument because if the effect of model parameter uncertainty were negligible, there would not be a significant difference between the effluents CDFs when the model is run with the same influent time series but different model parameter sets. Whereas, the results for the pragmatic Monte Carlo simulation illustrates that fixing uncertain model parameters at two different model parameter sets (i.e. nominal and “worst case”) and running the dynamic model of the plants with the same realizations of influent time series would result in a significant difference between the two calculated CDFs corresponding to the nominal and “worst case” parameters sets (e.g. see the difference between the blue CDF and the red one for Alt1 in Figure 5-11).

Comparing the two CDFs corresponding to the pragmatic Monte Carlo simulation provides useful information regarding the response of a design to model parameter uncertainty. According to Figure 5-11 (and Figure 5-12), as the total bioreactor volume increases, the difference between the blue and red CDFs and hence the calculated PONCs tend to decrease (see Figure 5-11, Figure 5-12, and Table 5-5). For example, for Alt1 the calculated PONC using the pragmatic Monte Carlo simulation at the nominal and “worst case” vector of parameters equals 0.0414 (i.e. 15.1 expected number of days in non-compliance) and 0.2360 (i.e. 86.4 expected number of days in non-compliance in a year) respectively. In contrast, for Alt5 the PONCs corresponding to the nominal and “worst case” vector of parameters equals 0.0007 (i.e. 0.2 expected number of days in non-compliance) and 0.0140 (i.e. 5.1 expected number of days in non-compliance in a year) respectively. Therefore, the difference between the expected number of days in non-compliance corresponding to the nominal vector of parameters (corresponding to the most likely parameter values) and the “worst case” vector of parameters (corresponding to a possible but conservative set of model parameter values) for Alt1 and Alt 5 are 71.3 and 4.9 days respectively.

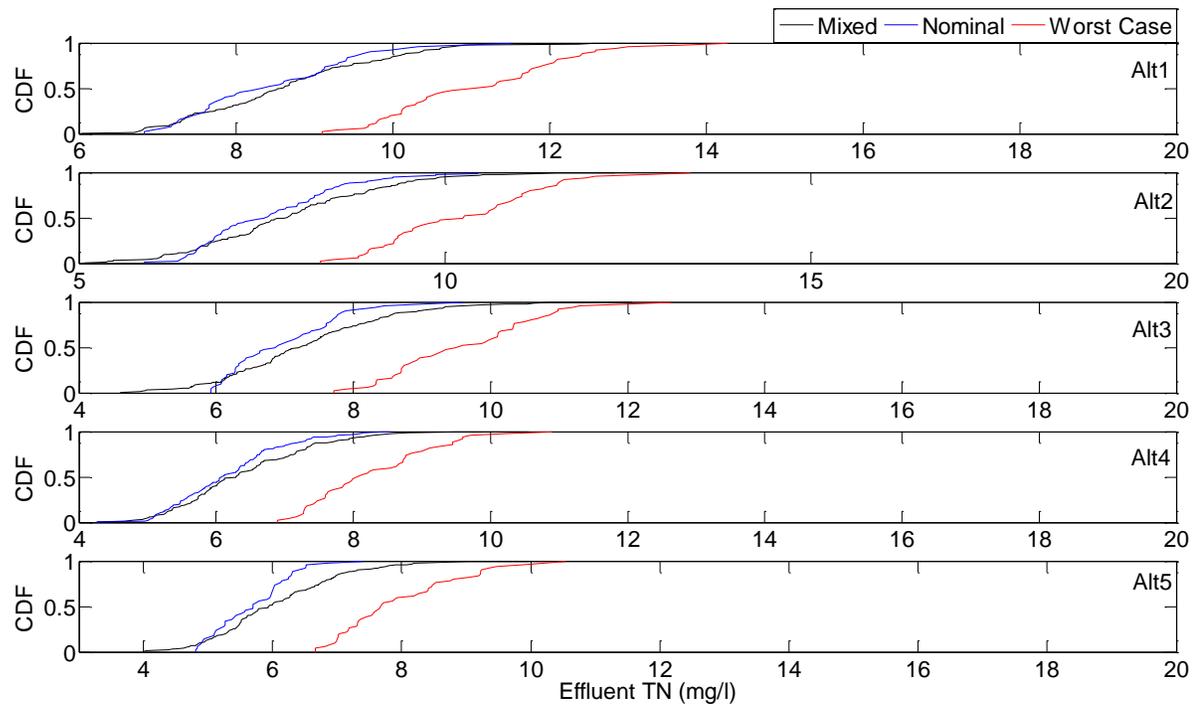
In other words, increasing the total bioreactor volume not only improves the performance of a design in coping with the influent variability (the PONC values for Alt5 are smaller than the PONC values for Alt1 for both nominal and “worst case” vectors of parameters) but also makes the system less sensitive to the value of the model parameters or in other words to the unknown process characteristics (71.3 and 4.9 days difference between the two cases for Alt1 and Alt5, respectively).



**Figure 5-11** Daily mean effluent  $\text{NH}_4$  distributions for 5 different design alternatives (Alt1-Alt5) for one-dimensional Monte Carlo (i.e. influent variability and parameter uncertainty mixed) and the pragmatic Monte Carlo simulation (i.e. influent variability with model parameters set at the nominal and “worst case” parameter sets)

Comparing the CDFs for TN corresponding to different designs also reveals the effect of total bioreactor volume on the calculated PONC values. As illustrated in Figure 5-12, the PONC values for Alt1 (with the smallest bioreactor volume), calculated at the nominal and “worst case” vector of model parameters (i.e. CDFs in blue) equals 8% and 80%,

respectively. On the other hands, the PONCs for Alt5 (with the largest bioreactor volume) for the nominal and “worst case” vector of model parameters are 0% and 2%. This implies that an increase in the total bioreactor volume will reduce PONCs (i.e. 8% for Alt1 compared to 0% for Alt5, calculated at the nominal vector of model parameters, and 80% for Alt1 compared to 2% for Alt5, calculated at the “worst case” vector of model parameters) similar to the  $\text{NH}_4$  results. The difference between the PONCs calculated for the nominal and “worst case” vector of model parameters is reduced (i.e.  $80-8 = 72\%$  difference compared to  $4-0 = 4\%$  difference between the PONCs corresponding to the nominal and “worst case” vector of model parameters, compared).



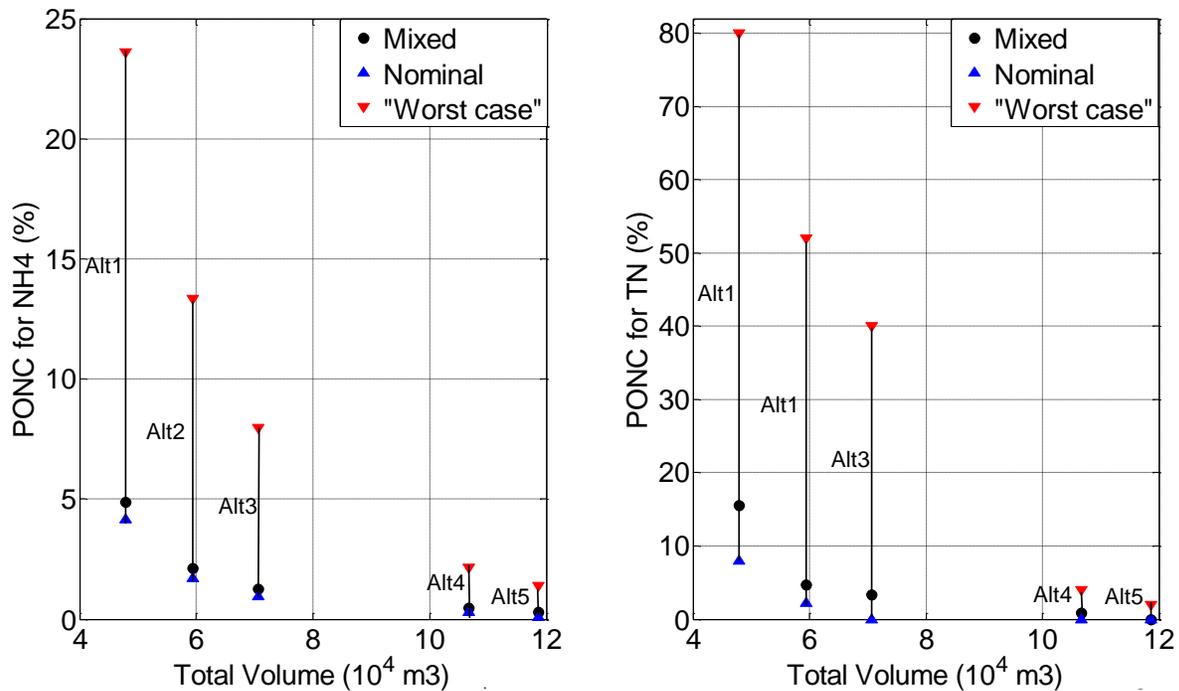
**Figure 5-12** Yearly mean effluent TN distributions for 5 different design (Alt1-Alt5) alternatives for one-dimensional Monte Carlo (i.e. influent variability and parameter uncertainty mixed) and the pragmatic Monte Carlo simulation (i.e. influent variability with model parameters set at the nominal and “worst case” parameter sets)

**Table 5-5** PONC values for different design alternatives calculated using one-dimensional Monte Carlo simulation (i.e. Mixed) and the pragmatic Monte Carlo simulation (i.e. Nominal and “Worst Case”)

<b>Mixed</b>										
Alternatives	BOD		COD		NH <sub>4</sub>		TN		TSS	
	PONC	Days <sup>1</sup>	PONC	Days <sup>1</sup>	PONC	Days <sup>1</sup>	PONC	Percent <sup>2</sup>	PONC	Percent <sup>2</sup>
Alt1	0.0000	0.0	0.0018	0.6	0.0488	17.8	0.1552	15.5%	0.000	0.0%
Alt2	0.0000	0.0	0.0018	0.6	0.0211	7.7	0.0467	4.6%	0.000	0.0%
Alt3	0.0000	0.0	0.0009	0.5	0.0126	4.6	0.0333	3.3%	0.000	0.0%
Alt4	0.0000	0.0	0.0014	0.3	0.0046	1.7	0.0081	0.8%	0.000	0.0%
Alt5	0.0000	0.0	0.0006	0.2	0.0028	1.1	0.0000	0.0%	0.000	0.0%
<b>Nominal</b>										
Alternatives	BOD		COD		NH <sub>4</sub>		TN		TSS	
	PONC	Days <sup>1</sup>	PONC	Days <sup>1</sup>	PONC	Days <sup>1</sup>	PONC	Percent <sup>2</sup>	PONC	Percent <sup>2</sup>
Alt1	0.0000	0.0	0.0030	1.1	0.0414	15.1	0.0800	8.0%	0.000	0.0%
Alt2	0.0000	0.0	0.0026	1	0.0170	6.2	0.0225	2.2%	0.000	0.0%
Alt3	0.0000	0.0	0.0020	0.7	0.0094	3.4	0.0000	0.0%	0.000	0.0%
Alt4	0.0000	0.0	0.0010	0.4	0.0029	1.0	0.0000	0.0%	0.000	0.0%
Alt5	0.0000	0.0	0.0008	0.3	0.0007	0.2	0.0000	0.0%	0.000	0.0%
<b>Worst Case</b>										
Alternatives	BOD		COD		NH <sub>4</sub>		TN		TSS	
	PONC	Days <sup>1</sup>	PONC	Days <sup>1</sup>	PONC	Days <sup>1</sup>	PONC	Percent <sup>2</sup>	PONC	Percent <sup>2</sup>
Alt1	0.0000	0.0	0.0060	2.2	0.2360	86.4	0.8000	80%	0.000	0.0%
Alt2	0.0000	0.0	0.0042	1.5	0.1334	48.7	0.5200	52%	0.000	0.0%
Alt3	0.0000	0.0	0.0027	1	0.0796	29.0	0.4	40%	0.000	0.0%
Alt4	0.0000	0.0	0.0035	1.3	0.0215	7.8	0.0200	4%	0.000	0.0%
Alt5	0.0000	0.0	0.0000	0.0	0.0140	5.1	0.0400	2%	0.000	0.0%
<sup>1</sup> Expected number of days with non-compliance event in a year (i.e. PONC×365)										
<sup>2</sup> Expected percentage of years with non-compliance events										

To better explore the relationship between the total volume of the bioreactors corresponding to the different design alternatives and PONC, the PONC values for NH<sub>4</sub> and TN corresponding to different design alternatives were plotted against their total bioreactor volume values (Figure 5-13).

As indicated, the NH<sub>4</sub> PONC values for all of the design alternatives, calculated using the pragmatic Monte Carlo simulation at the nominal vector of model parameters are below 5%. However, the NH<sub>4</sub> PONC values for Alt1, Alt2, and Alt3 calculated using the pragmatic Monte Carlo simulation at the “worst case” vector of model parameters (i.e. corresponding to a possible but conservative vector of model parameters) are very high (i.e. 86.4, 78.7, and 29 expected days of non-compliance in a year, respectively (Table 5-5)). which may render them unacceptable due to their poor performance in NH<sub>4</sub> removal. In contrast to alternatives Alt1, Alt2 and Alt3, alternatives Alt4 and Alt5 have a zero PONCs at the nominal vector of model parameters and small values at the “worst case” vector of model parameters (i.e. 7.8 and 5.1 expected days of non-compliance in a year corresponding to Alt4 and Alt5, respectively). In addition, the small difference between the two PONCs (i.e. calculated at the nominal and “worst case” vector of model parameters) of Alt4 and Alt5 implies the robustness of these two alternatives to the value of the uncertain model parameters. The same behaviors can be observed for the TN PONC values for the different designs (i.e. too high PONC values for Alt1, Alt2 and Alt3, and small PONCs for Alt4 and Alt5, calculated at the “worst case” vector of model parameters).



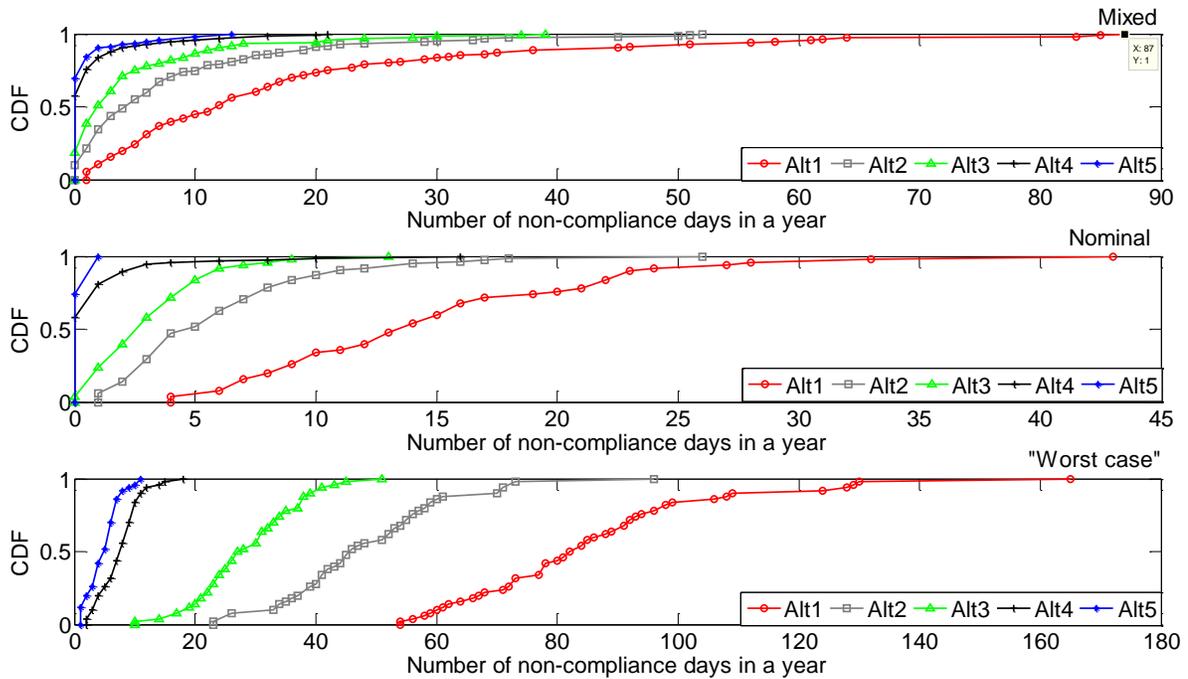
**Figure 5-13** Relationship between PONC values (calculated using one-dimensional Monte Carlo simulation, i.e. Mixed and the pragmatic Monte Carlo simulation, i.e. Nominal and “Worst Case”) and the total bioreactor volumes corresponding to the final design alternatives in Table 5-4.

## VI. Probability of a certain number of non-compliance days

As explained in Section 3.5, the expected number of days of non-compliance with the effluent standards that are based on daily average values can be calculated by multiplying PONC by 365 (Table 5-5). However, if the designers are interested not only in the expected number of days of non-compliance, but also in the probability of having a certain number of days of non-compliance in a year, a discrete distribution should be calculated using the simulated effluent time series (see Section 3.5 for details).

Figure 5-14 shows the discrete distributions representing the probability of the number of non-compliance days with the NH<sub>4</sub> effluent standards in a year for different design alternatives according to the one-dimensional Monte Carlo simulation (i.e. influent variability and model parameter uncertainty mixed) as well as the pragmatic Monte Carlo

simulation (i.e. with the vector of model parameters fixed at the nominal and “Worst case” values). The range of values on the x-axis were adjusted to better illustrates the discrete CDFs calculated from one-dimensional (i.e. Mixed) and the pragmatic Monte Carlo simulations (i.e. Nominal and “Worst case”).



**Figure 5-14** CDFs for the number of non-compliance days with NH<sub>4</sub> standard (i.e. daily mean NH<sub>4</sub> concentrations more than 2 mg/l) in a year calculated using one-dimensional Monte Carlo simulation (i.e. Mixed) and the pragmatic Monte Carlo simulation (i.e. Nominal and “Worst case”).

In addition to PONC values (i.e. Table 5-5), the probabilities associated to a certain number of non-compliance days can be used as another index for making the final decision on the optimum design alternative. For example, if designers are interested in a design alternative whose probability of having less than 10 days of non-compliance is smaller than 20% (regardless of the uncertainty procedure (i.e. mixed, nominal and “worst case”)), then Alt4 and Alt5 are the only alternatives that meet these conditions (Table 5-6).

**Table 5-6** Probability of having a certain number of non-compliance days with the NH<sub>4</sub> effluent standard in a year calculated using one-dimensional Monte Carlo (i.e. Mixed) and the pragmatic Monte Carlo simulations (i.e. Nominal and “Worst Case”)

<b>Mixed</b>				
Design alternative	Pr(X≥N=5 d)	Pr(X≥N=10 d)	Pr(X≥N=15 d)	Pr(X≥N=20 d)
Alt1	75.6%	55.2%	40.0%	26.7%
Alt2	44.7%	25.3%	14.7%	8.7%
Alt3	24.7%	13.3%	6.7%	6.0%
Alt4	8.0%	4.0%	1.6%	0.8%
Alt5	6.7%	2.2%	0.0%	0.0%
<b>Nominal</b>				
Design alternative	Pr(X≥N=5 d)	Pr(X≥N=10 d)	Pr(X≥N=15 d)	Pr(X≥N=20 d)
Alt1	100%	66.7%	40.0%	24.0%
Alt2	48.2%	13.0%	4.7%	1.2%
Alt3	16.0%	2.0%	0.0%	0.0%
Alt4	4.0%	0.0%	0.0%	0.0%
Alt5	0.0%	0.0%	0.0%	0.0%
<b>“Worst Case”</b>				
Design alternative	Pr(X≥N=5 d)	Pr(X≥N=10 d)	Pr(X≥N=15 d)	Pr(X≥N=20 d)
Alt1	100%	100%	100%	100%
Alt2	100%	100%	100%	100%
Alt3	100%	100%	96.0%	86.0%
Alt4	74.0%	16.0%	2.0%	0.0%
Alt5	48.0%	4.0%	0.0%	0.0%

It should be noted that selecting the best design alternative depends on how the decision makers weigh the different outputs of the proposed probabilistic method (i.e. PONC, total cost, and the probability of certain number of days of non-compliance) in their decision making process. As discussed throughout this section, the calculated PONCs (Table 5-5) and the probability of having a certain number of non-compliance days in a year (Table 5-6)

corresponding to the explained uncertainty propagation methods have different meanings. The following section provides a summary of the probabilistic indices calculated according to the proposed probabilistic design method.

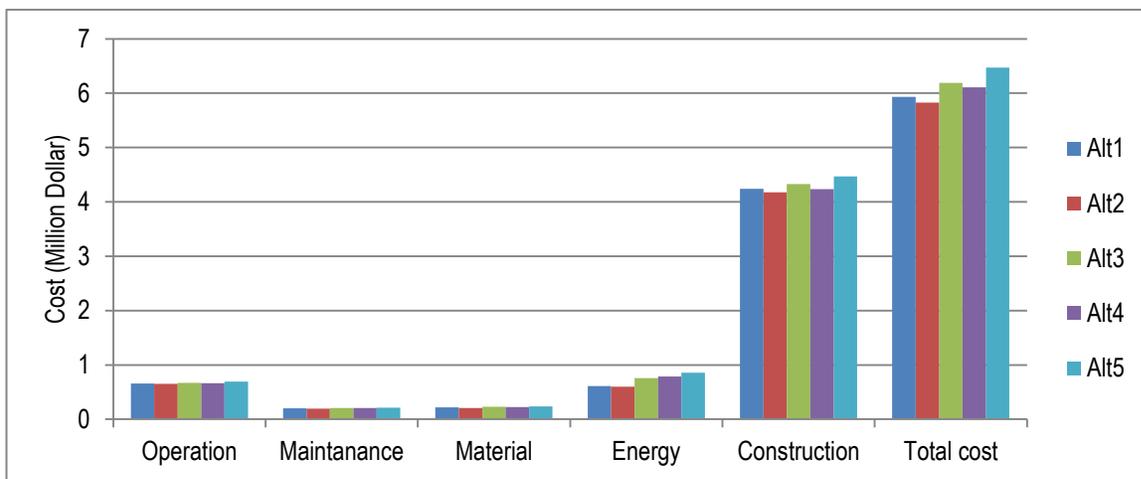
### ***VII. Summary of the probabilistic indices***

For each design alternative, three PONC values corresponding to the one-dimensional Monte Carlo (mixed influent variability and model parameter uncertainty) and the pragmatic Monte Carlo simulation at the nominal and “worst case” vector of model parameters were calculated for each design alternative (Table 5-5). As mentioned, the CDFs resulting from the one-dimensional Monte Carlo simulation might be difficult to interpret as the effect of influent variability and model parameter uncertainties are lumped. However, if it is assumed that the effect of model parameter uncertainty on the calculated CDFs is negligible compared to influent variability, its application may be justifiable. The PONC values calculated using the pragmatic Monte Carlo with model parameters fixed at the nominal values correspond to the most likely behavior of the plant. Finally, the PONCs calculated using the pragmatic Monte Carlo method with model parameters fixed at a “worst case” vector of model parameters, correspond to a possible (but less likely compared to the nominal vector of model parameters) condition in which the plant faces a very difficult circumstance. Therefore, the calculated PONC values are usually larger compared to the case of nominal model parameters (Table 5-5).

The PONC values in Table 5-5 represent the expected number of times that effluent concentration exceeds the effluent standards (Table 5-1). This could be used as a useful index for comparing the expected behavior of the design alternatives regarding different pollutant removals. Table 5-6 illustrates the probability of having a certain number (rather than the expected) of days of non-compliance in a year.

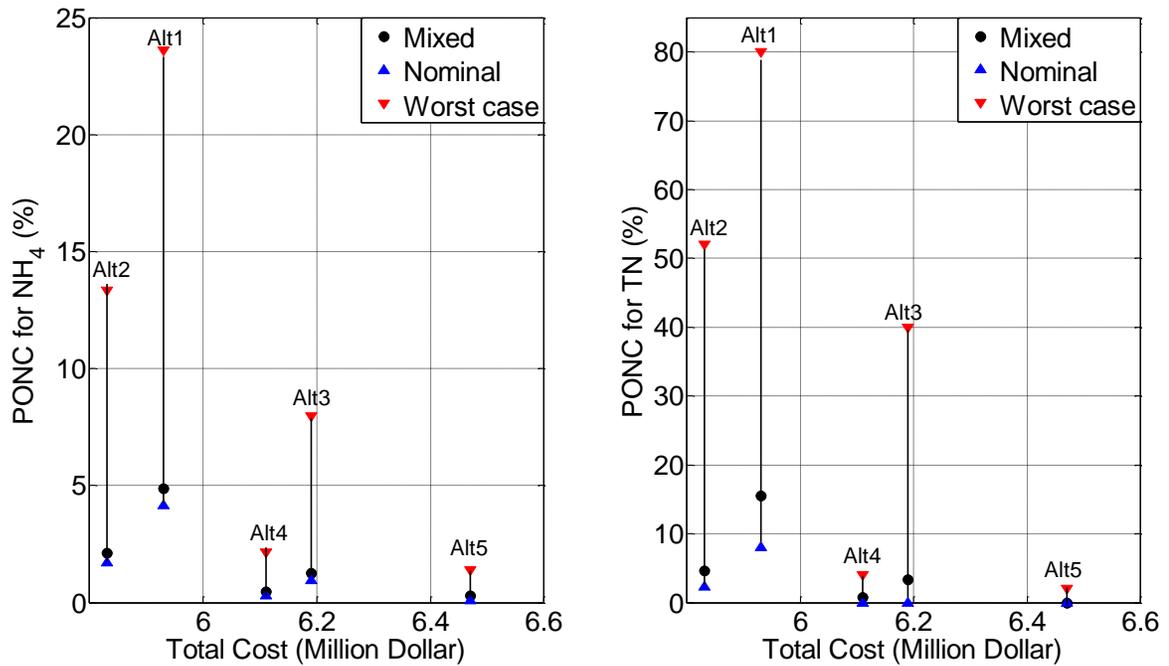
### VIII. Cost calculation

As mentioned in Section 5.2.3, the CAPDET software was used for calculation of the total cost corresponding to the different design alternatives. The calculated costs in Figure 5-15 are based on the costing database for 2013 in the United States with 8% interest rate and 40 years for the life time of the project. As indicated in Figure 5-15, differences in the total costs calculated using the CAPDET software are not very significant and a more detailed cost calculation might be required considering the regional conditions in which a WWTP is to be build (please see Section 2.5 and 3.5.2 for issues associated with cost calculations).



**Figure 5-15** Total cost breakdown for the final design alternatives in Table 5-4 calculated using CAPDET

Figure 5-16 illustrates the relationship between the PONC values and the corresponding total cost for the different design alternatives. Plotting the variation of PONC against the total cost could help designers identify those regions in design space for which the ratio of reduction in PONC to the increase in the total cost is at its highest and the effluent standards are met with a tolerable PONC. For example, if designers were interested in a  $\text{NH}_4$  PONC value of less than 5%, a TN PONC value of less than 10%, and a total cost of less than 6 million dollars for the “worst case” vector of model parameters, Alt4 would be selected as the best design alternative.



**Figure 5-16** Relationship between PONC values (calculated using one-dimensional Monte Carlo simulation, i.e. Mixed and the pragmatic Monte Carlo simulation, i.e. Nominal and “Worst Case”) and the total cost corresponding to the final design alternatives in Table 5-4.

## 5.4 CONCLUSION AND RECOMMENDATIONS

In this study a probabilistic design method was presented for the design of wastewater treatment plants (WWTP). First, a set of pre-designs with different levels of safety was generated under steady state conditions using the ATV design guideline. A k-means sampling method with a handful number of clusters (i.e. in these study 7 clusters) was performed on the generated pre-designs and the cluster centroids were selected as preliminary designs. The performances of the preliminary designs (i.e. corresponding to the cluster centroids) were further evaluated under dynamic conditions with the nominal/calibrated dynamic model of designs and a year-long dynamic influent time series.

Based, on the resulting effluent CDFs, some of the preliminary designs that have similar performance were eliminated and more designs that did not necessarily correspond to a cluster centroids were added to the final designs for which the probability of non-compliance (PONC) were calculated.

The uncertainty in the parameters of a dynamic WWTP model was characterized using uniform distribution functions. Moreover, an influent generator was developed for generating different realizations of influent time series. Two different Monte Carlo simulation schemes including the one-dimensional Monte Carlo (i.e. with influent variability and model parameter uncertainty mixed) and the pragmatic Monte Carlo simulation (i.e. with influent variability evaluated at the nominal and “worst case” vector of model parameters) were used for propagating the effect of influent variability and model parameters uncertainty to the effluent distributions.

Depending on the sampling method, used for measuring the compliance to effluent standards (i.e. daily average for BOD, COD,  $\text{NH}_4$ , and yearly average for TSS, and TN), the simulated effluent time series (with a 15-minute temporal resolution) were aggregated to daily and yearly values and the convergence of Monte Carlo simulations were checked at each Monte Carlo run. Once the convergence of the statistical properties of the effluent time series were ensured, three CDFs and hence three PONC values (i.e. one PONC calculated using the results of the one-dimensional Monte Carlo and two others from the results of pragmatic Monte Carlo simulation evaluated at the nominal and “worst case” vector of model parameters) corresponding to different effluent constituents were calculated for each design alternative.

The PONC values calculated using the one-dimensional Monte Carlo distributions might be difficult to interpret as the effect of influent variability and model parameter are mixed in the propagation of uncertainty. The PONC values calculated using the pragmatic Monte

Carlo, at the nominal vector of model parameters (representing the most likely values for model parameters assuming a perfect model calibration) represent the most likely PONCs for the different design alternatives. In addition, the PONC corresponding to the pragmatic Monte Carlo simulation, at the “worst case” vector of model parameters represent a very unfavorable condition for pollutant removal.

The results indicate that all of the selected design alternatives meet the effluent standards for carbonaceous wastewater constituents (i.e. BOD, COD) and the TSS with a negligible PONC. However, there is a significant difference between the  $\text{NH}_4$  and TN PONC values calculated for the different design alternatives. Plotting  $\text{NH}_4$  and TN PONC values against their corresponding bioreactor volume reveals that not only the increasing the total bioreactor volume will result in reduction of PONCs, but also in reduction of difference between the PONCs calculated at the nominal and “worst case” vector of model parameters. In other words, increasing the size of total bioreactor volume will make a design and hence its corresponding PONC less sensitive to the values of model parameters.

Using the PONC values, the designers can calculate the expected number of days that a design alternative is in non-compliance with the effluent standards. However, if calculating the probability of having equal or less than a certain number is of interest (i.e. the probability of having 10 or less than 10 days of not meeting the effluent standards), then a discreet CDF should be constructed on the number of non-compliance events.

In the end, the PONC, the probability of having a certain number of non-compliance events in a year, as well as the total cost values calculated using the explained one-dimensional and pragmatic Monte Carlo simulation methods are provided for each design alternatives.

Calculating PONC as a quantitative measure of safety for each design alternative helps designers better understand and compare the performance of different design alternatives.

Even in projects in which the sizing of a WWTP should be consistent with a specific design guideline, the proposed probabilistic design can be used as a tool for selecting proper values for safety factors and other inputs that are required for dimensioning the different units of a WWTP. It should be noted that any calculated PONC depends on the validity of the assumptions that have been made. For example, in this study it is assumed that there is no interruption or failure in the performance of the technical components of the WWTP (e.g. pumps or sensors). Considering the impact of failure in technical components on PONC requires explaining their performance in terms of probability and doing a comprehensive reliability analysis. Moreover, the calculation of PONC in light of future climate change scenarios, post design evaluation of risk registry, as well as the effect of using different plant configurations, treatment technologies, and plant control strategies constitute the main research subjects that could be tackled in future studies.

It should be noted that experienced designers may be able to make an optimum design using the traditional method of design. However, the application of the proposed probabilistic design can provide some insights that the traditional design methods cannot. The advantage of the proposed probabilistic design method can be summarized as follows. It allows to:

- 1) reduce the effect of subjectivity in selecting proper design input values, especially in situations where the designers do not have enough experience (e.g. not enough knowledge on the effect of different process configurations on treatment performance);
- 2) provide more insight regarding the compliance with the effluent standards for sub-daily effluent standards (e.g. hourly average concentrations);
- 3) provide an explicit quantitative measure of compliance to the effluent standards;

- 4) assist designers in identifying the limits of a specific treatment technology or process configuration as well as the design regions where the increase in certain wastewater treatment plant's process units would not result in a significant reduction of PONC.



## **CHAPTER 6: GENERAL CONCLUSIONS AND PERSPECTIVES**

In this study a novel probabilistic design method that is based on the explicit characterization of relevant sources of uncertainty and variability was developed. The different sources of information, the developed modeling and analysis tools, as well as the new uncertainty analysis techniques for the propagation of influent variability and model parameter uncertainty are explained in detail. To better illustrate the feasibility of the novel design method, it was applied to the design of the Eindhoven wastewater treatment plant (WWTP) upgrade. Moreover, an attempt was made to shed some light on the rather abstract concept of probability and the importance of the probabilistic design approach in terms of how designers can benefit from outcomes of the proposed design approach in selecting the optimum design configuration and sizing of WWTPs.

This section provides a summary of the main findings and contributions to the field of probabilistic design of WWTPs. At the end, the author proposes a list of research topics that could be tackled in future studies concerning probabilistic design of WWTPs.

### **6.1 Conclusions**

#### ***6.1.1 A practical framework for the probabilistic design of WWTPs***

One of the main advantages of the proposed probabilistic method compared to the previous studies in this field is the applicability of the method to actual (and non-actual) design projects. According to the proposed method, the initial sizing of a plant is not significantly different from current design methods that are based on the assumption of steady state conditions. This makes the method quite comfortable for use by experienced designers that know this approach for years.

However, in the proposed method a set of steady state designs with different levels of safety is made and then the Probability Of Non-Compliance (PONC) corresponding to each design alternative is calculated using dynamic simulation and the explained methods of analysis.

Including PONC as an additional quantitative performance criterion in the selection of the optimum design alternative reduces the risk of selecting improper values for design inputs that would result in either undersized or oversized plants. In other words, the calculation of PONC for a set of potential design alternatives reduces the effect of subjectivity - that could be significant depending on the experience of the designers - in selecting proper design inputs.

Overall, the proposed probabilistic design method can be thought of as an integrated design method which incorporates different types of knowledge into the design of WWTPs:

- ✓ The knowledge based on years of experience in the design and operation of WWTPs (i.e. the knowledge that is reflected in the design guidelines and used in the probabilistic design method for the initial sizing of a set of design alternatives, e.g. the ranges of parameters for steady state design inputs)
  
- ✓ The knowledge that is based on the detailed understanding of treatment processes and is reflected in the current state of mathematical modeling (i.e. the application of dynamic mathematical models for simulating the performance of the different design alternatives)
  
- ✓ The knowledge on the quantification of uncertainties and the PONC (i.e. the statistical analysis that is used for explicit characterization of relevant sources of uncertainties).

Therefore, it can be concluded that the proposed probabilistic design method enriches and complements the previous design methods in which subjectivity plays an important role in the final design of the plant (both in the selection of steady state design inputs and the definition of dynamic flow and loading scenarios that are to be used for checking the performance of a design alternative using dynamic simulation).

### ***6.1.2 Development of an influent generator using the basic characteristics of sewershed and climate data***

One of the most innovative outcomes of this research work is the development of a dynamic influent generator for the synthetic generation of influent time series. Given the importance of the influent variability on the performance of WWTPs, a substantial amount of effort was dedicated to the development of an influent generator that is capable of producing different realizations of influent time series and being tuned for a particular plant.

Comparing the outputs of the influent generator with the observed data of a case study indicated that the developed tool is capable of producing different realizations of the influent time series that have the same statistical properties as the observed ones. Besides, the correlation between the different outputs of the influent generator is also respected during the synthetic generation of the different influent constituents, which is imperative for realistic generation of influent time series.

Moreover, the flexibility of the proposed tool allows user to easily incorporate the projected change in climate conditions or the overall change in the characteristics of the connected sewershed into the generated influent time series. This is an essential feature given the lifespan of WWTPs (30 years or more).

### ***6.1.3 Rigorous calculation and communication of PONC concept***

Proper treatment of relevant sources of uncertainty and variability and considering their effect on the PONC constitutes an important contribution to the field of probabilistic design of WWTPs. Contrary to the previous works in which the effect of influent variability was considered in a haphazard manner (i.e. using only a single influent time series or a limited number of dynamic flow and loading scenarios), the current study enables designers to simultaneously consider the effect of both influent variability and parametric uncertainty.

In addition, the rigorous calculation of PONC, an attempt was made to clarify the rather abstract meaning of PONC. It was demonstrated how the PONC can be used to estimate the expected number of non-compliance events in a year and how increasing the total cost of a plant would affect the PONC.

### ***6.1.4 Application of the proposed methodology to a real case study***

The application of the proposed probabilistic to the design of the upgrade of the Eindhoven WWTP illustrates how the different components of the design method (i.e. different sources of information, characterization of uncertainty and modeling, creating designs and evaluation of design alternatives) are put into practice to create a design.

## **6.2 Perspectives**

### ***6.2.1 Calculation of conditional PONC in view of future anthropological and climate change***

According to the proposed method, the calculation of PONC requires the explicit characterization of different sources of uncertainty in terms of probability distribution functions whose parameters are calculated using the available measured data (e.g. rainfall data, influent data, etc). However, in cases where a significant change in the anthropological or climate condition of the catchment is projected, characterization of some of the relevant sources of uncertainty/variability should be performed in view of future change(s). Fortunately, the flexibility of the developed influent generator in this study allows the designers to incorporate the effect of major changes in climate (e.g. an increase in the frequency and/or amount of rainfall) or holistic changes in the characteristics of the sewershed (e.g. an increase in the average flow or concentration of certain pollutants) into the synthetic generation of influent time series which in turn affects the calculated PONC (See Chapter CHAPTER 4: for detail on the proposed influent generator). Calculating the PONC in view of future projected changes constitutes an interesting and multidisciplinary research field which could be tackled in future studies.

### ***6.2.2 Including the uncertainty in the performance of technical components of WWTPs***

In this study it was assumed that there was no interruption in the proper functioning of the technical components of the WWTP system. However, in reality there may be some malfunctioning or a cessation of performance in the components of WWTPs which could result in non-compliance events. Defining the reliability of technical components of a WWTP system in terms of failure probabilities and considering their combined effect on the

PONC constitutes another interesting field of future studies. The result of such studies could help designers increase the reliability of the system and determine an optimum redundancy capacity for the different components of a plant.

### ***6.2.3 Devising improved control strategies to reduce the risk of non-compliance***

Efficient control of WWTPs plays an important role in meeting the effluent standards in a cost-efficient manner. The operators of a WWTP may opt to use different control strategies to improve the performance of the plant during different periods of the year in which there could be significant differences in the temperature and/or dynamic influent load entering the WWTP. The developed tools in this study provide an integrated modeling framework that considers the interaction between the local climate, sewershed, and WWTP. Such a modeling framework can be used to devise a model-based flexible control strategy that could result in a significant reduction in PONC. Therefore, the application of the developed tools for analyzing the effect of different control strategies on PONC for existing plants is suggested as another subject matter for future studies.

## REFERENCES

- Achleitner S., Möderl M., Rauch W. (2007). CITY DRAIN © – An open source approach for simulation of integrated urban drainage systems. *Environmental Modelling & Software*, 22(8), 1184-95.
- Antoniou P., Hamilton J., Koopman B., Jain R., Holloway B., Lyberatos G., Svoronos S.A. (1990). Effect of temperature and pH on the effective maximum specific growth rate of nitrifying bacteria. *Water Research*, 24(1), 97-101.
- ATV (2000). Standard ATV-DVWK-A 131E, Dimensioning of Single-Stage Activated Sludge Plants. German Association for Water, Wastewater and Waste, Hennef, Germany, pp. 57.
- Balmer P., Mattsson B. (1994). Wastewater treatment plant operation costs. *Water Science and Technology*, 30(4), 7-15.
- Bates B.C., Campbell E. P. (2001). A Markov Chain Monte Carlo Scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. *Water Resources Research*, 37(4), 937-947.
- Bechmann H., Nielsen M.K., Madsen H., Kjølstad Poulsen N. (1999). Grey-box modelling of pollutant loads from a sewer system. *Urban Water*, 1(1), 71-8.
- Beck M. B. (1987). Water quality modeling: a review of the analysis of uncertainty. *Water Resources Research*, 23(8), 1393-1442.

Belia E., Amerlinck L., Benedetti B., Johnson B., Sin G., Vanrolleghem P.A., Gernaey K.V., Gillot S., Neumann M.B., Rieger L., Shaw A., Villez K. (2009). Wastewater treatment modelling: Dealing with uncertainties. *Water Science and Technology*, 60(8), 1929-1941.

Belia E., Benedetti L., Cierkens K., De Baets B., Flaming T., Nopens I., Weijers S. (2012). Translating safety factors to sources of variability and uncertainty; Model-based design evaluation for the Eindhoven WWTP. In: *Proceedings 85<sup>th</sup> Annual WEF Technical Exhibition and Conference (WEFTEC2012)*. New Orleans, LA, USA, September 29 - October 3, 2012.

Belia E., Johnson B., Benedetti L., Bott C.B., Martin C., Murthy S., Neumann M.B., Rieger L., Weijers S., Vanrolleghem P.A. (2013). Uncertainty evaluations in model-based WRRF design for high level nutrient removal - Literature review and research needs. *Water Environment Research Foundation*, Alexandria, VA, USA. pp. 46.

Benedetti L. (2006). *Probabilistic Design and Upgrade of Wastewater Treatment Plants in the EU Water Framework Directive Context*. PhD Thesis, Ghent University, Belgium, pp. 304.

Benedetti L., Bixio D., Vanrolleghem, P.A. (2006a). Assessment of WWTP design and upgrade options: Balancing costs and risks of standards' exceedance. *Water Science and Technology*, 54(6-7), 371-378.

Benedetti L., Bixio D., Vanrolleghem P.A. (2006b). Benchmarking of WWTP design by assessing costs, effluent quality and process variability. *Water Science and Technology*, 54(10), 95-102.

- Benedetti L., Bixio D., Claeys F., Vanrolleghem P.A. (2008). Tools to support a model-based methodology for emission/immission and benefit/cost/risk analysis of wastewater systems that considers uncertainty. *Environmental Modelling & Software*, 23(8), 1082-1091.
- Benedetti L., Claeys F., Nopens I., Vanrolleghem P.A. (2011). Assessing the convergence of LHS Monte Carlo simulations of wastewater treatment models. *Water Science and Technology*, 63(10), 2219-2224.
- Benedetti L., Belia E., Cierkens K., Flameling T., De Baets B., Nopens I., Weijers S. (2013). The incorporation of variability and uncertainty evaluations in WWTP design by means of stochastic dynamic modeling: The case of the Eindhoven WWTP upgrade. *Water Science and Technology*, 67(8), 1841-1850.
- Bixio D., Parmentier, G., Rousseau D., Verdonck F., Meirlaen J., Vanrolleghem P.A. (2002). A quantitative risk analysis tool for design/simulation of wastewater treatment plants. *Water Science and Technology*, 46(4), 301-307.
- Board G.L.U.M.R. (2004). *Recommended Standards for Wastewater Facilities*, Health Education Services Division, Albany, NY, USA.
- Bode H., Lemmel P. (2001). International product cost comparison in the field of water management. *Water Science and Technology*, 44(2-3), 85-93.
- Brun R., Kühni M., Siegrist H., Gujer W., Reichert P. (2002). Practical identifiability of ASM2d parameters-Systematic selection and tuning of parameter subsets. *Water Research*, 36(16), 4113-4127.

- Bürger R., Diehl S., Nopens I. (2011). A consistent modelling methodology for secondary settling tanks in wastewater treatment. *Water Research*, 45(6), 2247-2260.
- Claeys F., Chtepen M., Benedetti L., Dhoedt B., Vanrolleghem P.A. (2006) Distributed virtual experiments in water quality management. *Water Science and Technology*, 53(1), 297-305.
- Capodaglio A., Zheng S., Novotny V., Feng X. (1990). Stochastic system identification of sewer-flow models. *Journal of Environmental Engineering*, 116(2), 284-98.
- Chen J., Brissette F.P., Leconte R. (2010). A daily stochastic weather generator for preserving low-frequency of climate variability. *Journal of Hydrology*, 388(3-4), 480-90.
- Cierkens K., Belia E., Benedetti L., Weijers S., Flameling T., De Baets B., Nopens I. (2011). Comparison of implicit and explicit uncertainty analyses in the design of a WWTP upgrade. In: *Proceedings of the 8th IWA Symposium on Systems Analysis and Integrated Assessment (WATERMATEX2011)*. San Sebastián, Spain, June 20-22, 2011.
- Ciggin A.S., Rossetti S., Majone M., Orhon D. (2012). Effect of feeding regime and the sludge age on the fate of acetate and the microbial composition in sequencing batch reactor. *Journal of Environmental Science and Health, Part A*, 47(2), 192-203.
- Claeys F., De Pauw D.J.W., Benedetti L., Vanrolleghem P.A. (2006). Tornado: A versatile and efficient modelling and virtual experimentation kernel for water quality systems. In: *Proceedings of the International Congress on Environmental Modelling and Software (iEMSs2006)*. July 9-12, 2006, Burlington, VT, USA.

- Corominas, L., Flores-Alsina, X., Muschalla, D., Neumann, M.B., Vanrolleghem, P.A. (2010). Verification of WWTP design guidelines with activated sludge process models. In: Proceedings of the 83<sup>rd</sup> Annual Water Environment Federation Technical Exhibition and Conference (WEFTEC2010). October 2-6, 2010, New Orleans, LA, USA.
- Copp, J.B. (2002). The COST Simulation Benchmark: Description and Simulator Manual. Office for Official Publications of the European Community, Luxembourg. ISBN 92-894-1658-0. pp. 154.
- Cox C.D. (2004). Statistical distributions of uncertainty and variability in activated sludge model parameters. *Water Environment Research*, 76(7), 2672-2685.
- Daigger G.T., Grady C.P.L.Jr. (1995). The use of models in biological process design. In: Proceedings of the Water Environment Federation 68<sup>th</sup> Annual Conference and Exposition (WEFTEC1995). Miami Beach, FL, USA, October 21-25, 1995, pp. 501-510.
- Daigger G.T, Crawford G.V., O'Shaughnessey M., Elliott M. (1998). The use of coupled refined stoichiometric and kinetic/stoichiometric models to characterize entire wastewater treatment plants. In: Proceedings of the Water Environment Federation 71<sup>th</sup> Annual Conference and Exposition (WEFTEC1998). Orlando, FL, USA, October 3-7, 1998, pp. 617-628.
- Devisscher M., Ciacci G., Benedetti L., Bixio D., Thoeye C., Gueldre G.D., Marsili-Libelli S., Vanrolleghem P.A. (2006). Estimating costs and benefits of advanced control for wastewater treatment plants: The MAGIC methodology. *Water Science and Technology*, 53(4-5), 215-223.

- Doby T. (2004). Optimization of Wastewater Treatment Design under Uncertainty and Variability. PhD Thesis, North Carolina State University, USA, pp. 194.
- Doby T., Loughlin D., de los Reyes 3rd F., Ducoste J. (2002). Optimization of activated sludge designs using genetic algorithms. *Water Science and Technology*, 45(6), 187-198.
- Dotto C. B. S., Mannina G., Kleidorfer M., Vezzaro L., Henrichs M., McCarthy D. T., Freni G., Rauch W., Deletic A. (2012). Comparison of different uncertainty techniques in urban stormwater quantity and quality modelling. *Water Research*, 46(8), 2545-58.
- Ekama G.A. (2009). Using bioprocess stoichiometry to build a plant-wide mass balance based steady-state WWTP model. *Water Research*, 43(8), 2101-2120.
- Ekama G.A., Marais G.V.R., Siebritz I.P., Pitman A.R., Keay G.F.P., Buchan L., Gerber A., Smollen M. (1984). Theory, Design and Operation of Nutrient Removal Activated Sludge Processes. Water Research Commission, Pretoria, South Africa, pp. 241.
- Ekama G.A., Wentzel M.C. (2004). A predictive model for the reactor inorganic suspended solids concentration in activated sludge systems. *Water Research*, 38(19), 4093-4106.
- European Commission (2002). Guidance document on risk assessment for birds and mammals under council directive 91/414/EEC. SANCO/4145/2000. Brussels, [http://ec.europa.eu/food/plant/protection/evaluation/guidance/wrkdoc19\\_en.pdf](http://ec.europa.eu/food/plant/protection/evaluation/guidance/wrkdoc19_en.pdf)

- Flores-Alsina X., Corominas L., Neumann M.B., Vanrolleghem P.A. (2012). Assessing the use of activated sludge process design guidelines in wastewater treatment plant projects: A methodology based on global sensitivity analysis. *Environmental Modelling & Software*, 38(), 50-58.
- Flores-Alsina X., Saagi R., Lindblom E., Thirsing C., Thornberg D., Gernaey K. V., Jeppsson U. (2014a). Calibration and validation of a phenomenological influent pollutant disturbance scenario generator using full-scale data. *Water Research*, 51(), 172-85.
- Flores-Alsina X., Arnell M., Amerlinck Y., Corominas L., Gernaey K.V., Guo L., Lindblom E., Nopens I., Porro J., Shaw A., Snip L., Vanrolleghem P.A., Jeppsson U. (2014b). Balancing effluent quality, economic cost and greenhouse gas emissions during the evaluation of (plant-wide) control/operational strategies in WWTPs. *Science of the Total Environment*, 466-467(), 616-624.
- Freni G., Mannina G., Viviani G. (2009). Uncertainty assessment of an integrated urban drainage model. *Journal of Hydrology* 373(3), 392-404.
- Freni G., Mannina G. (2010). Bayesian approach for uncertainty quantification in water quality modelling: The influence of prior distribution. *Journal of Hydrology*, 392(1), 31-9.
- Frey H.C. (1993). Separating variability and uncertainty in exposure assessment: Motivations and method. In: *Proceedings of the 86<sup>th</sup> Annual Meeting & Exhibition, Air & Waste Management Association, Denver, CO, USA, June 13–18, 1993.*

- Frey H.C., Rhodes D.S. (1996). Characterizing, simulating, and analyzing variability and uncertainty: An illustration of methods using an air toxics emissions example. *Human and Ecological Risk Assessment*, 2(4), 762-797.
- Fries M.K., Johnson B., Daigger G., Murthy S., Bratby, J. (2010). Dynamic and steady state modeling for the design of the enhanced nutrient removal facility at the Blue Plains Advanced Wastewater Treatment Facility. In: *Proceedings of the 83<sup>rd</sup> Annual Water Environment Federation Technical Exhibition and Conference (WEFTEC2010)*. New Orleans, LA, USA, October 2-6, 2010.
- Gernaey K.V., Petersen B., Parmentier G., Bogaert H., Ottoy J.P., Vanrolleghem P.A. (2000). Application of dynamic models (ASM1) and simulation to minimize renovation costs of a municipal activated sludge wastewater treatment plant. In: *Proceedings of the 1<sup>th</sup> World Congress of the International Water Association*, Paris, France, 3–7 July 2000.
- Gernaey K.V., Flores-Alsina X., Rosen C., Benedetti L., Jeppsson U. (2011). Dynamic influent pollutant disturbance scenario generation using a phenomenological modelling approach. *Environmental Modelling & Software*, 26(11), 1255-1267.
- Gernaey K.V., Jeppsson U., Vanrolleghem P.A. and Copp J.B. (2014) *Benchmarking of Control Strategies for Wastewater Treatment Plants*. Scientific and Technical Report No 23. IWA Publishing, London, UK.
- Gillot S., De Clercq B., Defour D., Simoens F., Gernaey K.V., Vanrolleghem P.A. (1999). Optimisation of wastewater treatment plant design and operation using simulation and cost analysis. In: *Proceedings of the 72<sup>nd</sup> Annual Water Environment Federation Technical Exhibition and Conference (WEFTEC1999)*. New Orleans, LA, USA, October 9–13 1999.

- Gillot S., Vanrolleghem P.A. (2003). Equilibrium temperature in aerated basins-comparison of two prediction models. *Water Research* 37(), 3742-48.
- Han K.Y., Kim S.H., Bae D.H. (2001). Stochastic water quality analysis using realibility method. *Journal of the American Water Resources Association*, 37(3), 695-708.
- Hao X., van Loosdrecht M.C.M., Meijer S.C.F., Qian Y. (2001). Model-based evaluation of two BNR processes-UCT and A2N. *Water Research*, 35(12), 2851-60.
- Hartigan J. A., Wong M.A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society*, 28(1), 100-108.
- Harris R.W., Cullinane M.J., Sun, P. (1982). *Process Design & Cost Estimating Algorithms for the Computer Assisted Procedure for Design and Evaluation of Wastewater Treatment Systems (CAPDET)*. United States Army Engineer Waterways Experiment Station, Vicksburg, MS, USA, pp. 729.
- Hauduc H., Rieger L., Ohtsuki T., Shaw A., Takács I., Winkler, S., Heduit A., Vanrolleghem P.A., Gillot, S. (2011). Activated sludge modelling: development and potential use of a practical applications database. *Water Science and Technology*, 63(10), 2164-2182.
- Henze M., Grady C.P.L., Gujer W., Marais G.R., Matsuo T. (1987). *Activated Sludge Model no. 1*. IAWPRC Scientific and Technical Report no. 3. International Association of Water Quality, London, UK, pp. 33.

- Henze M., Gujer W., Mino T., Matsuo T., Wentzel M.C., van Loosdrecht M. (1999). Activated Sludge Model no. 2d, ASM2d. *Water Science and Technology*, 39(1), 165-182.
- Henze M., Gujer W., Mino T., van Loosdrecht M. (2000). Activated Sludge Models ASM1, ASM2, ASM2d and ASM3, IWA Task Group on Mathematical Modelling for Design and Operation of Biological Wastewater Treatment, IWA Publishing, London, UK, pp. 121.
- Henze M., van Loosdrecht M., Ekama G., Brdjanovic D. (2008). *Biological Wastewater Treatment: Principles, Modelling and Design*. IWA Publishing, London, UK.
- Hernandez-Sancho F., Molinos-Senante M., Sala-Garrido R. (2011). Cost modelling for wastewater treatment processes. *Desalination*, 268(1), 1-5.
- Hernebring C., Jönsson L., Thorn U., Møller A. (2002). Dynamic online sewer modelling in Helsingborg. *Water Science and Technology*, 45(4-5), 429-36.
- Hoffman F.O., Hammonds J.S. (1994). Propagation of uncertainty in risk assessments: the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. *Risk Analysis*, 14(5), 707-712.
- Huang K.Z. (1986). Reliability analysis of hydraulic design of open channel. In *Stochastic and Risk Analysis in Hydraulic Engineering*, ed. by B.C. Yen, Water Resources Publications, Littleton, CO, USA, pp. 59–65.

- Huo J., Jiang Y., Seaver W.L., Robinson R.B., Cox C.D. (2006). Statistically based design of wastewater treatment plants (WWTPs) using Monte Carlo simulation of Activated Sludge Model No.1 (ASM1). In: Proceedings of the 2006 World Environmental and Water Resources Congress. Omaha, NE, USA, May 21-25, 2006, pp 21-30.
- Hyland K.C., Dickenson E.R., Drewes J.E., Higgins, C.P. (2012). Sorption of ionized and neutral emerging trace organic compounds onto activated sludge from different wastewater treatment configurations. *Water Research* 46(6), 1958-1968.
- Iman R.L., Conover W.J. (1982). A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics-Simulation and Computation*, 11(3), 311-334.
- Iman R.L., Davenport, J.M. (1982). Rank correlation plots for use with correlated input variables. *Communications in Statistics-Simulation and Computation*, 11(3), 335-360.
- Kanso A., Tassin B., Chebbo G. (2005). A benchmark methodology for managing uncertainties in urban runoff quality models. *Water Science and Technology*, 51(02), 163-70.
- Korving H., Van Noordwijk J.M., van Gelder P.H.A.J.M., Clemens F.H.L.R. (2009). Risk-based design of sewer system rehabilitation. *Structure and Infrastructure Engineering*, 5(3), 215-227.
- Koutsyiannis D., Onof C. (2001). Rainfall disaggregation using adjusting procedures on a Poisson cluster model. *Journal of Hydrology*, 246(1-4), 109-122.

- Gupta, A.K., Shrivastava R.K. (2006). Uncertainty analysis of conventional water treatment plant design for suspended solids removal. *Journal of Environmental Engineering*, 132(11), 1413-1421.
- Langergraber G., Alex J., Weissenbacher N., Woerner D., Ahnert M., Frehmann T., Halft N., Hobus I., Plattes M., Spring V. (2008). Generation of diurnal variation for influent data for dynamic simulation. *Water Science and Technology* 57(9), 1483.
- Langeveld J.G., Schilperoot R.P.S., Rombouts P.M.M., Benedetti L., Amerlinck Y., de Jonge J., Flameling T., Nopens I., Weijers S. (2014). A new empirical sewer water quality model for the prediction of WWTP influent quality. In: *Proceedings of the 13th IWA/IAHR International Conference on Urban Drainage*. Sarawak, Malaysia, September, 7-12, 2014.
- Lee H.L. (1986). Hydraulic uncertainties in flood levee capacity. *Journal of Hydraulic Engineering*, 112(10), 928-934.
- Maier H.R., Lence B.J., Tolson B.A., Foschi R.O. (2001). First-order reliability method for estimating reliability, vulnerability, and resilience. *Water Resources Research*, 37(3), 779-790.
- Marais G.R., Ekama G. (1976). The activated sludge process. part I: steady state behavior. *Water SA*, 2(4), 164-200.
- Marshall L., Nott D., Sharma A. (2004). A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling. *Water Resources Research*, 40(2), W02501, doi:10.1029/2003WR002378.

- Martin C., Eguinoa I., McIntyre N. R., García-Sanz M., Ayesa E. (2007). ARMA models for uncertainty assessment of time series data: Application to Galindo-Bilbao WWTP. In: Proceedings of the 7<sup>th</sup> International IWA Symposium on Systems Analysis and Integrated Assessment in Water Management (WATERMATEX 2007). Washington DC, USA, May 7-9, 2007.
- Martin C., Ayesa E. (2010). An Integrated Monte Carlo Methodology for the calibration of water quality models. *Ecological Modelling*, 221(22), 2656-2667.
- Martin C., Neumann M. B., Altimir J., Vanrolleghem P. A. (2012) A tool for optimum design of WWTPs under uncertainty: Estimating the probability of compliance In: Proceedings International Congress on Environmental Modelling and Software (iEMSs2012). Leipzig, Germany, July 1-5, 2012.
- Martin C., Vanrolleghem P.A. (2014). Analyzing, completing, and generating influent data for WWTP modelling: A critical review. *Environmental Modelling & Software* 60(), 188-201.
- Matalas N.C. (1967). Mathematical assessment of synthetic hydrology. *Water Resources Research*, 3(4), 937-945.
- McGhee T.J., Mojgani P., Vicidomina F. (1983). Use of EPA's CAPDET program for evaluation of wastewater treatment alternatives. *Journal of the Water Pollution Control Federation*, 55(1), 35-43.

- McCormick J.F., Johnson B., Turner A. (2007). Analyzing risk in wastewater process design: Using Monte Carlo simulation to move beyond conventional design methods. In: Proceedings of the 80<sup>th</sup> Annual WEF Technical Exhibition and Conference (WEFTEC2007). San Diego, CA, USA, October, 15-17, 2007.
- McKay M.D., Beckman, R.J., Conover, W.J. (1979). A comparison of three methods of selecting values of input variables in the analysis of output from a computer code. *Technometrics*. 21(2), 239–245.
- Melching C.S., Yen B.C. (1986). Slope influence on storm sewer risk. In *Stochastic and Risk Analysis in Hydraulic Engineering*, ed. by B.C. Yen, Water Resources Publications, Littleton, CO, USA, pp. 66–78.
- Merz B., Thielen A.H. (2005). Separating natural and epistemic uncertainty in flood frequency analysis. *Journal of Hydrology*, 309(1), 114-132.
- Mildenhall S.J. (2005). The report of the Research Working Party on Correlations and Dependencies among all Risk Sources: Part 1—Correlation and aggregate loss distributions with an emphasis on the Iman-Conover method. Casualty Actuarial Society (CAS), Available:  
<https://www.casact.org/pubs/forum/06wforum/06w107.pdf>
- Montgomery V. (2009) *New Statistical Methods in Risk Assessment by Probability Bounds*. PhD Thesis, Durham University, UK, pp. 152.
- Motiee H., Chocat B., Blanpain O. (1997). A storage model for the simulation of the hydraulic behaviour of drainage networks. *Water Science and Technology* 36 (8-9), 57-63.

- Neumaier A., Schneider T. (2001). Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transactions on Mathematical Software*, 27(1), 27-57.
- Ormsbee L. (1989). Rainfall disaggregation model for continuous hydrologic modeling. *Journal of Hydraulic Engineering*, 115(4), 507-25.
- Pattison A. (1965). Synthesis of hourly rainfall data. *Water Resources Research* 1(4), 489-498.
- Pineau M., Cote P., Villeneuve J.P. (1985). Estimation of wastewater treatment costs: Evaluation of the CAPDET model for Canadian conditions. *Canadian Journal of Civil Engineering*, 12(3), 483-493.
- Refsgaard J.C., Van Der Sluijs J.P., Højberg A.L., Vanrolleghem P.A. (2007). Uncertainty in the environmental modelling process - A framework and guidance. *Environmental Modelling & Software*, 22(11), 1543-1556.
- Richardson C.W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, 17(1), 182-190.
- Rivas A., Ayesa E., Galarza A., Salterain A. (2001). Application of mathematical tools to improve the design and operation of activated sludge plants. Case study: The new WWTP of Galindo-Bilbao. Part I: Optimum design. *Water Science and Technology*, 43(7), 165.
- Rivas A., Irizar I., Ayesa E. (2008). Model-based optimisation of wastewater treatment plants design. *Environmental Modelling & Software*, 23(4), 435-450.

- Roberson J.A., Cassidy J., Chaudhry M.H. (1995). *Hydraulic Engineering*. John Wiley Sons, Inc, New York.
- Rodríguez J.P., McIntyre N., Díaz-Granados M., Achleitner S., Hochedlinger M., Maksimović Č. (2013). Generating time-series of dry weather loads to sewers. *Environmental Modelling & Software*, 43, 133-143.
- Rodriguez-Iturbe I., Cox D.R., Isham V. (1987). Some models for rainfall based on stochastic point processes. *Proc. R. Soc. London, Ser. A*. 410(), 269-288.
- Rousseau D., Verdonck F., Moerman O., Carrette R., Thoeye C., Meirlaen J., Vanrolleghem P.A. (2001). Development of a risk assessment based technique for design/retrofitting of WWTPs. *Water Science and Technology*, 43(7), 287-294.
- Seber G.A.F (1984). *Multivariate Observations*. John Wiley & Sons, Hoboken, NJ, USA, pp. 692.
- Salem S., Berends D., Heijnen J., van Loosdrecht M. (2002). Model-based evaluation of a new upgrading concept for N-removal. *Water Science and Technology*, 45(6), 169-76.
- Schilperoort R. (2011). *Monitoring as a Tool for the Assessment of Wastewater Quality Dynamics*. MSc Thesis Water Management Academic Press, Delft, the Netherlands. pp. 320.
- Sharma M., Hall W., McBean E. (1993). Reliability analyses of activated sludge systems in attaining effluent standards. *Canadian Journal of Civil Engineering*, 20(2), 171-179.

- Schwarz G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.
- Sin G., Gernaey K.V., Neumann M.B., van Loosdrecht M., Gujer W. (2009). Uncertainty analysis in WWTP model applications: A critical discussion using an example from design. *Water Research*, 43(11), 2894-2906.
- Sitar N., Cawfield J.D., Der Kiureghian A. (1987). First-order reliability approach to stochastic analysis of subsurface flow and contaminant transport. *Water Resources Research*, 23(5), 794–804.
- Talebizadeh M., Martin C., Neumann M.B., Vanrolleghem P.A. (2012). Global sensitivity analysis of the ATV design guideline: Which are the most important inputs for a WWTP design? In: *Proceedings 85th Annual WEF Technical Exhibition and Conference (WEFTEC2012)*. New Orleans, LA, USA, September 29 - October 3, 2012.
- Talebizadeh M., Belia E., Vanrolleghem P.A. (2014). Probability-based design of wastewater treatment plants. In: *Proceedings International Congress on Environmental Modelling and Software (iEMSs2014)*. San Diego, CA, USA, June 15 - 19, 2014.
- Tang, C.C., Brill J.E.D., Pfeffer, J.T. (1987). Comprehensive model of activated sludge wastewater treatment system. *Journal of Environmental Engineering*, 113(5), 952-969.
- Temprano J., Arango Ó., Cagiao J., Suárez J., Tejero I. (2007). Stormwater quality calibration by SWMM: A case study in Northern Spain. *Water*, SA 32(1), 55-63.

- Tchobanoglous G., Burton F., Stensel H.D. (2003). *Metcalf & Eddy Wastewater Engineering: Treatment and Reuse*. McGraw Hill, New York, USA. pp. 1848.
- Tung Y.K., Yen B.C. (2005). *Hydrosystems Engineering Uncertainty Analysis*. McGraw-Hill, New York, USA. pp. 285.
- Tung Y.K., Yen B.C., Melching C.S. (2006). *Hydrosystems Engineering Reliability Assessment and Risk Analysis*. McGraw-Hill, New York, USA. pp. 512.
- Uber J.G., Brill E.D., Pfeffer J.T. (1991a). Robust optimal design for wastewater treatment. I: General approach. *Journal of Environmental Engineering*, 117(4), 425-437.
- Uber J.G., Brill E.D., Pfeffer J.T. (1991b). Robust optimal design for wastewater treatment. II: Application. *Journal of Environmental Engineering*, 117(4), 438-456.
- USACE (1996). *Risk-Based Analysis for Flood Damage Reduction Studies*, Department of the Army, Washington, DC, USA. pp. 63.
- Vanhooren H., Meirlaen J., Amerlinck Y., Claeys F., Vangheluwe H., Vanrolleghem P.A. (2003). WEST: Modelling biological wastewater treatment. *Journal of Hydroinformatics*, 5(1), 27-50.
- Vanrolleghem P.A. (2011). Resolving the tensions over the treatment plant's fence: Diverse approaches around the world. Lecture held at the WEF/IWA Nutrient Recovery and Management Conference 2011. Miami, FL, USA, January 9-12, 2011.

- Vrugt J.A., ter Braak C.J. F., Clark M.P., Hyman J.M., Robinson B.A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research*, 44(12), W00B9.
- Wagstaff K., Cardie C., Rogers S., Schroedl S. (2001). Constrained k-means clustering with background knowledge. In: *Proceedings 18<sup>th</sup> International Conference on Machine Learning*. Williamstown, MA, USA, June 26 – July 1, 2001.
- Walker W.E., Harremoes P., Rotmans J., van der Sluijs J.P. (2003). Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, 4(1), 5-17.
- Wentzel M.C., Ekam G.A., Dold P.L., Marais, G. (1990). Biological excess phosphorus removal-steady state process design. *Water SA*, 16(1), 29-48.
- Willems, P. (2008). Quantification and relative comparison of different types of uncertainties in sewer water quality modeling. *Water Research*, 42(13), 3539-3551.
- Wu F.C., Tsang Y.P. (2004). Second-order Monte Carlo uncertainty/variability analysis using correlated model parameters: Application to salmonid embryo survival risk assessment. *Ecological Modelling*, 177(3), 393-414.
- Yen B.C., Ang A.H.S. (1971). Risk analysis in design of hydraulic projects. In: *Proceedings of First International Symposium on Stochastic Hydraulics*, University of Pittsburgh, Pittsburgh, PA, USA. 694–701.

Yen B.C., Cheng S.T., Tang W.H. (1980). Reliability of hydraulic design of culverts. In: Proceedings International Conference on Water Resources Development, IAHR Asian Pacific Division Second Congress. Taipei, Taiwan, May 12-14, 1980, 991–1001.

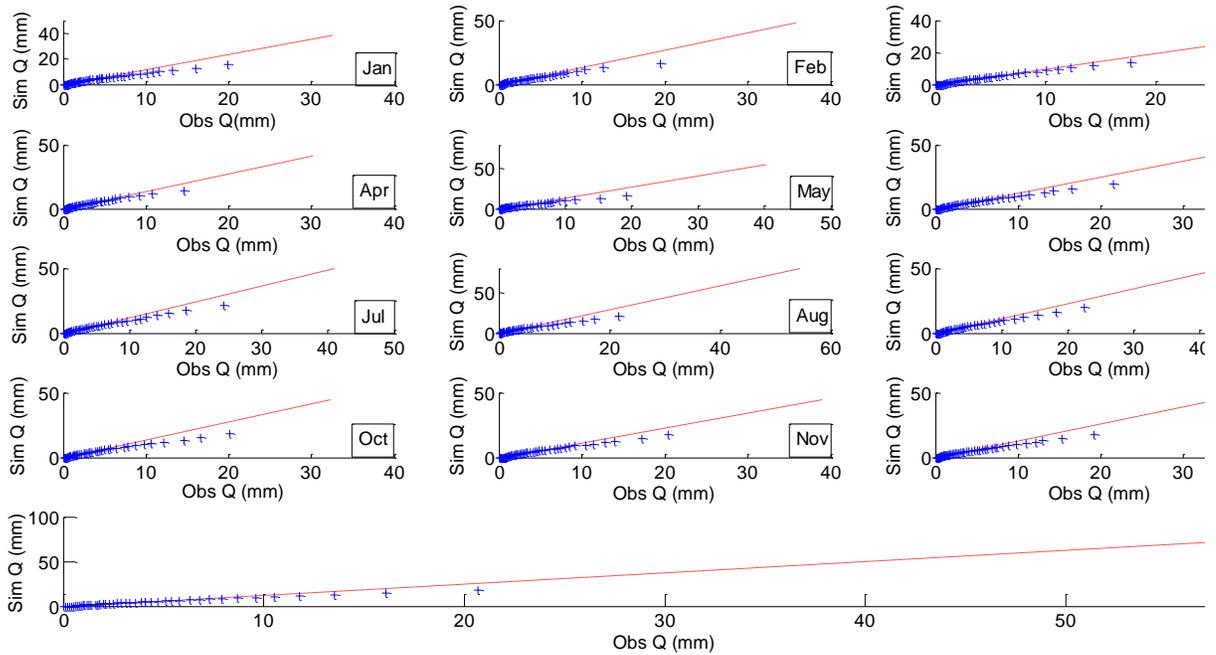
## APPENDIX

**Table A1.** Estimated parameters of air temperature model (i.e. Equation 4-6)

$\begin{pmatrix} T_{\max} \\ T_{\min} \end{pmatrix}_t = A \times \begin{pmatrix} T_{\max} \\ T_{\min} \end{pmatrix}_{t-1} + B \times N(0,1)_t$			
<b>A</b>		<b>B</b>	
0.79	0.05	0.56	-0.06
0.34	0.52	0.28	0.52

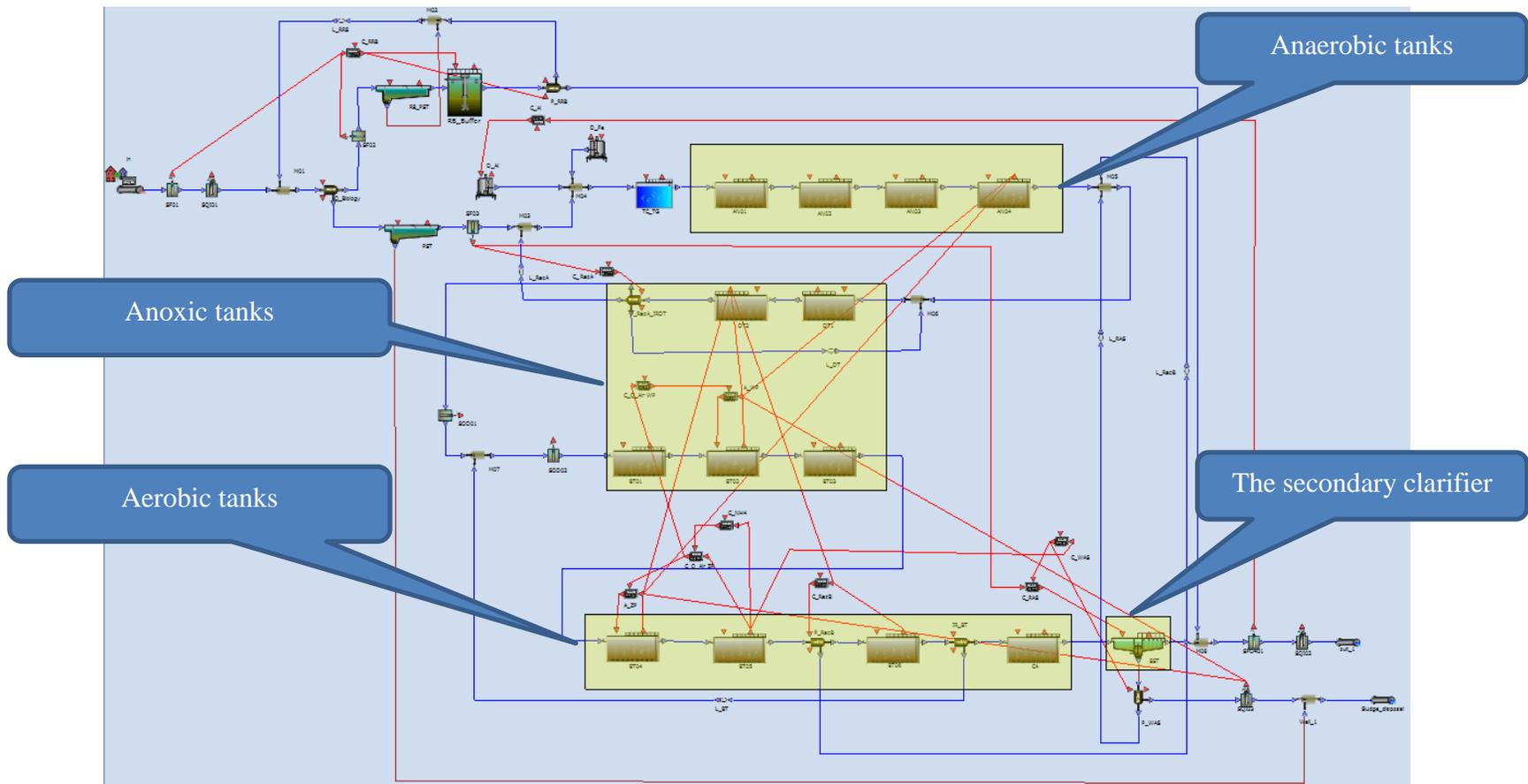
**Table A2.** Estimated parameters of influent model in DWF conditions (i.e. Equation 4-13)

$\begin{pmatrix} Flow \\ Solube\_COD \\ Total\_COD \\ TSS \\ NH_4 \end{pmatrix}_t = \sum_{i=1}^{i=9} A_i \times \begin{pmatrix} Flow \\ Solube\_COD \\ Total\_COD \\ TSS \\ NH_4 \end{pmatrix}_{t-1} + \varepsilon_t$									
<b>A1</b>					<b>A2</b>				
1.62	0.05	-0.09	0.03	-0.18	-0.90	-0.13	0.18	-0.09	0.23
0.00	2.06	-0.90	0.67	0.00	-0.02	-1.88	1.80	-1.33	-0.01
-0.02	0.26	1.10	0.36	-0.03	0.03	-0.51	0.05	-0.73	0.04
-0.02	0.38	-0.57	2.00	-0.01	0.04	-0.78	1.25	-1.84	0.02
-0.08	0.01	-0.04	0.04	1.58	0.14	0.00	0.06	-0.05	-0.81
<b>A3</b>					<b>A4</b>				
0.07	0.12	-0.16	0.10	-0.06	0.27	-0.01	0.03	-0.03	-0.04
0.02	0.81	-1.34	0.99	0.04	-0.01	0.17	0.23	-0.12	-0.05
0.00	0.37	-0.59	0.51	-0.01	-0.02	-0.05	0.40	-0.04	-0.04
0.00	0.59	-0.94	0.79	-0.01	-0.02	0.01	0.01	0.29	-0.02
-0.05	0.00	-0.06	0.04	0.13	-0.05	-0.02	0.06	-0.05	0.14
<b>A5</b>					<b>A6</b>				
-0.10	-0.05	0.06	-0.04	0.03	-0.13	0.05	-0.10	0.08	0.05
0.00	-0.39	0.66	-0.57	0.03	0.00	0.16	-0.65	0.55	0.02
0.00	-0.17	0.25	-0.28	0.04	0.01	0.13	-0.43	0.24	0.04
0.00	-0.40	0.58	-0.47	0.00	0.02	0.29	-0.44	0.15	0.05
0.04	0.00	-0.04	0.04	-0.01	0.01	0.03	0.00	-0.02	-0.12
<b>A7</b>					<b>A8</b>				
0.12	-0.01	0.04	-0.02	-0.04	-0.13	-0.11	0.13	-0.08	0.01
0.00	0.02	0.24	-0.21	-0.03	0.01	-0.21	0.12	-0.11	0.00
-0.01	-0.02	0.18	-0.06	-0.04	0.00	-0.17	0.08	-0.18	-0.01
-0.02	-0.05	0.06	0.08	-0.04	-0.01	0.01	-0.02	-0.13	0.00
-0.02	-0.02	0.02	0.00	0.06	0.01	0.14	-0.17	0.14	-0.06
<b>A9</b>									
0.04	0.00	0.01	-0.03	-0.03					
-0.01	0.11	-0.03	0.03	-0.02					
-0.01	0.02	0.08	0.04	-0.01					
-0.01	0.00	0.01	0.12	0.01					
0.00	-0.11	0.09	-0.07	0.05					



**Figure A1** q-q plot for observed and simulated rainfall quantiles in different month and the entire year

Note: the q-q plots were generated by plotting 1 to 99 quantiles of observed rainfall (Obs Q) data against their corresponding quantiles in simulated rainfall series (Sim Q). The red dotted lines represent the locations where the corresponding quantiles are equal.



**Figure A2:** Eindhoven WWTP's model configuration in WEST

**Table A3:** Upper and Lower limits, Nominal, and “Worst Case” of uncertain model parameters

Model parameters	Symbol	Lower limit	Upper limit	Nominal	“Worst Case”
Nitrogen content of biomass	i_N_BM	0.0560	0.0840	0.07	0.058237
Nitrogen content of soluble substrate S_F	i_N_S_F	0.0240	0.0360	0.03	0.03089
Nitrogen content of inert soluble COD S_I	i_N_S_I	0.0080	0.0120	0.01	0.032892
Nitrogen content of inert particulate COD X_I	i_N_X_I	0.0160	0.0240	0.02	0.022718
Nitrogen content of particulate substrate X_S	i_N_X_S	0.0320	0.0480	0.04	0.033631
Phosphorus content of biomass	i_P_BM	0.0160	0.0240	0.02	0.01887
Phosphorus content of soluble substrate	i_P_S_F	0.0050	0.0150	0.01	0.006429
Phosphorus content of inert soluble COD	i_P_S_I	0.0000	0.0000	0	0
Phosphorus content of inert particulate COD	i_P_X_I	0.0050	0.0150	0.01	0.006793
Phosphorus content of particulate substrate	i_P_X_S	0.0050	0.0150	0.01	0.011516
TSS to biomass ratio	i_TSS_BM	0.8550	0.9450	0.9	0.899784
TSS to X_I ratio	i_TSS_X_I	0.7125	0.7875	0.75	0.766168
TSS to X_S ratio	i_TSS_X_S	0.7125	0.7875	0.75	0.72006
Conversion factor for BOD/COD	F_BOD_COD	0.6175	0.6825	0.65	0.64
Saturation coefficient for acetate	K_A	2.0000	6.0000	4	5.193756
Saturation coefficient for alkalinity	K_ALK	0.0500	0.1500	0.1	0.141013
Saturation coefficient of autotrophs for alkalinity	K_ALK_AUT	0.2500	0.7500	0.5	0.740657
Saturation/inhibition coefficient for growth on S_F	K_F	2.0000	6.0000	4	4.968208
Inhibition coefficient for X_PP storage	K_IPP	0.0100	0.0300	0.02	0.01858

Maximum ration of X_PP/X_PAO	K_MAX	0.2720	0.4080	0.34	0.322486
Saturation coefficient for ammonium	K_NH	0.0250	0.0750	0.05	0.048037
Saturation coefficient of autotrophs	K_NH_AUT	0.5000	1.5000	1	0.064486
Saturation/inhibition for nitrate	K_NO	0.4000	0.6000	0.5	0.555457
Saturation/inhibition coefficient for oxygen	K_O	0.1000	0.3000	0.2	0.278227
Saturation/inhibition coefficient for autotrophs for oxygen	K_O_AUT	0.4000	0.6000	0.5	0.551884
Saturation coefficient for phosphorous	K_P	0.0050	0.0150	0.01	0.005136
Saturation coefficient for PHA	K_PHA	0.0050	0.0150	0.01	0.013511
Saturation coefficient for poly-phosphate	K_PP	0.0050	0.0150	0.01	0.01178
Saturation coefficient for phosphorous in PP	K_PS	0.1000	0.3000	0.2	0.144605
Saturation coefficient for particulate COD	K_X	0.0500	0.1500	0.1	0.062501
Saturation coefficient for fermentation on S_F	K_fe	2.0000	6.0000	4	4.247837
Rate constant for storage of PHA	Q_PHA	1.5000	4.5000	3	4.220613
Rate constant for storage of PP	Q_PP	0.7500	2.2500	1.5	0.898529
Maximum rate for fermentation	Q_fe	1.5000	4.5000	3	3.057867
Reference temperature of the activated sludge	Temp_Ref	Fixed	20	20	20
Decay rate	b_AUT	0.0750	0.2250	0.15	0.224382
Rate constant for lysis and decay	b_H	0.3200	0.4800	0.4	0.418392
Rate constant for lysis of X_PAO	b_PAO	0.1000	0.3000	0.2	0.118721
Rate constant for lysis of X_PHA	b_PHA	0.1000	0.3000	0.2	0.296328
Rate constant for lysis of	b_PP	0.1000	0.3000	0.2	0.218855

X_PP					
Rate constant for P precipitation	k_PRE	0.5000	1.5000	1	0.799492
Rate constant for P redissolution	k_RED	0.3000	0.9000	0.6	0.453273
Hydrolysis rate constant	k_h	1.5000	4.5000	3	2.855829
Maximum growth rate	mu_AUT	0.8000	1.2000	1	0.862154
Maximum growth rate on substrate	mu_H	3.0000	9.0000	6	4.388589
Maximum growth rate	mu_PAO	0.5000	1.5000	1	1.013913
Anoxic reduction factor for decay of autotrophs	n_NO_AUT_d	0.1650	0.4950	0.33	0.386134
Reduction factor for denitrification	n_NO_Het	0.6400	0.9600	0.8	0.849516
Anoxic reduction factor for decay of heterotrophs	n_NO_Het_d	0.4000	0.6000	0.5	0.427666
Anoxic hydrolysis reduction factor	n_NO_Hyd	0.4800	0.7200	0.6	0.683991
Active PAO organism under anoxic conditions	n_NO_PAO	0.4800	0.7200	0.6	0.544545
Anoxic reduction factor for decay of PAO, PP, PHA	n_NO_P_d	0.2640	0.3960	0.33	0.265146
Anaerobic hydrolysis reduction factor	n_fe	0.3200	0.4800	0.4	0.417526
Temperature correction factor for K_X	theta_K_X	Fixed		0.896	0.896
Temperature correction factor for Q_PHA	theta_Q_PHA	Fixed		1.041	1.041
Temperature correction factor for K_Q_PP	theta_Q_PP	Fixed		1.041	1.041
Temperature correction factor for Q_fe	theta_Q_fe	Fixed		1.072	1.072
Temperature correction factor for b_AUT	theta_b_AUT	Fixed		1.116	1.116
Temperature correction factor for b_PAO	theta_b_PAO	Fixed		1.072	1.072
Temperature correction factor for b_PHA	theta_b_PHA	Fixed		1.072	1.072

Temperature correction factor for Q_PP	theta_b_PP	Fixed		1.072	1.072
Temperature correction factor for k_h	theta_k_h	Fixed		1.041	1.041
Temperature correction factor for maximum growth rate of autotrophs	theta_mu_AUT	Fixed		1.111	1.111
Temperature correction factor for maximum growth rate of heterotrophs	theta_mu_H	Fixed		1.072	1.072
Temperature correction factor for mu_PAO	theta_mu_PAO	Fixed		1.041	1.041
Yield for autotrophic biomass	Y_AUT	0.2280	0.2520	0.24	0.248352
Yield for heterotrophic biomass	Y_H	0.5938	0.6563	0.625	0.638402
Yield coefficient (biomass/PHA)	Y_PAO	0.5938	0.6563	0.625	0.600379
PHA requirement for PP storage	Y_PHA	0.1900	0.2100	0.2	0.190819
PP requirement per PHA stored	Y_PO	0.3800	0.4200	0.4	0.396938
Fraction of inert COD in particulate substrate	f_S_I	0.0000	0.0000	0	0.040943
Fraction of inert COD in particulate COD generated in biomass lysis	f_X_I	0.0500	0.1500	0.1	0.147731
Fraction TSS/COD	F_TSS_COD	0.71	0.78	0.75	0.76
SVI	SVI	100	140	120	125
Non-settable fraction of suspended solids	f_ns	0.0011	0.0034	0.00228	0.00271