



# Computational Intelligence Techniques for Control and Optimization of Wastewater Treatment Plants

---

**Peter Kern**

**Student No. 69250040**

**Doctoral Dissertation**

Department of Electronic Engineering  
National University of Ireland Maynooth

**February 2016**

Supervisor: Prof. Dr. Seán McLoone (Maynooth University)  
External Examiner: Prof. Dr. Gustaf Olsson (Lund University)  
Internal Examiner: Dr. Robert Lawlor (Maynooth University)

## Abstract

The development of novel, practice-oriented and reliable instrumentation and control strategies for wastewater treatment plants in order to improve energy efficiency, while guaranteeing process stability and maintenance of high cleaning capacity, has become a priority for WWTP operators due to increasing treatment costs. To achieve these ambitious and even contradictory objectives, this thesis investigates a combination of online measurement systems, computational intelligence and machine learning methods as well as dynamic simulation models. Introducing the state-of-the-art in the fields of WWTP operation, process monitoring and control, three novel computational intelligence enabled instrumentation, control and automation (ICA) methods are developed and presented. Furthermore, their potential for practical implementation is assessed. The methods are, on the one hand, the automated calibration of a simulation model for the Rospe WWTP that provides a basis for the development and evaluation of the subsequent methods, and on the other hand, the development of soft sensors for the WWTP inflow which estimate the crucial process variables *COD* and *NH<sub>4</sub>-N*, and the estimation of WWTP operating states using Self-Organising Maps (SOM) that are used to determine the optimal control parameters for each state. These collectively, provide the basis for achieving comprehensive WWTP optimization. Results show that energy consumption and cleaning capacity can be improved by more than 50%.

## Acknowledgements

Writing a PhD thesis is probably always a demanding task. For me, it was a pretty rough time as well and I am glad and happy that I was surrounded by people, who supported me more than I could have ever imagined.

First of all I would like to thank my two supervisors, Prof. Dr. Seán McLoone from Queen's University Belfast and Prof. Dr. Michael Bongards from Cologne University of Applied Sciences: Thank you for giving me the possibility to contact you day and night, for your professional advice and your constructive reviews.

Dr. Christian Wolfs support was amazing: Thanks Christian for sharing your knowledge and spending hours with me in front of my monitor. Oliver Trauer's exceptional skills in writing software programs were incredibly helpful: Oliver, thank you so much for your advice and your support, which improved my work a lot.

Furthermore I would like to thank Dr. Anne-Kathrin Hillenbach for the given support. Without you, this work could not have been done or at least could not have been finished in time. I am happy to hand in that work and even happier to have you at my sight. Thanks to the staff of the National University of Ireland Maynooth, who made the administration processes smooth and easy. Last, but not least thanks to Thomas Schiller for the help formatting of the thesis.

## Table of Contents

Abstract .....	I
Acknowledgements .....	II
List of Figures.....	VII
List of Tables.....	X
Nomenclature .....	XII
1 Introduction.....	- 1 -
1.1 Aims and Scope of the Thesis.....	- 3 -
1.2 Objectives and Contributions.....	- 4 -
1.3 Outline.....	- 4 -
1.4 Publications.....	- 5 -
2 Wastewater Treatment .....	- 9 -
2.1 Wastewater.....	- 9 -
2.1.1 Sources and Types of Wastewater.....	- 9 -
2.2 Biochemical Wastewater Treatment .....	- 10 -
2.2.1 Activated Sludge .....	- 11 -
2.2.2 Nitrogen Removal.....	- 11 -
2.2.3 Carbon Removal (Organic Matter Removal) .....	- 15 -
2.2.4 Phosphate Removal .....	- 16 -
2.3 Wastewater Treatment Infrastructure.....	- 17 -
2.3.1 Sewer System .....	- 17 -
2.3.2 Wastewater Treatment Plants .....	- 18 -
2.3.3 Sludge Treatment.....	- 24 -
2.4 Summary .....	- 25 -
3 Wastewater treatment plant control .....	- 27 -
3.1 WWTP Instrumentation and Actuators.....	- 27 -
3.1.1 Actuators.....	- 28 -
3.1.2 Process Variables / Measurement Values.....	- 29 -

3.1.3	Measurement Principles .....	- 32 -
3.2	WWTP Control .....	- 38 -
3.2.1	Aeration Control .....	- 40 -
3.2.2	Recirculation Control .....	- 46 -
3.2.3	Phosphate Precipitation Control .....	- 46 -
3.2.4	Filtrate Water Control.....	- 48 -
3.2.5	Return Sludge Control and Excess Sludge Control.....	- 49 -
3.3	Summary .....	- 49 -
4	WWTP Modelling.....	- 51 -
4.1	General wastewater treatment plant models .....	- 51 -
4.1.1	Modeling of the hydraulics.....	- 51 -
4.1.2	Modeling biological processes .....	- 53 -
4.1.3	Activated Sludge Models.....	- 55 -
4.1.4	Activated Sludge Model No. 1 .....	- 57 -
4.2	Modelling of the Rospe WWTP.....	- 60 -
4.2.1	Rospe WWTP .....	- 60 -
4.2.2	WWTP Model .....	- 65 -
4.3	Model Calibration .....	- 69 -
4.3.1	Calibration Prerequisites.....	- 69 -
4.3.2	Modell Initialization .....	- 70 -
4.3.3	Parameter optimization.....	- 71 -
4.4	Results and Discussion.....	- 74 -
4.4.1	Conclusion and consequences for further use .....	- 76 -
5	Virtual COD and NH <sub>4</sub> -N inflow measurements.....	- 77 -
5.1	Introduction.....	- 77 -
5.2	Installed measurement probes .....	- 79 -
5.3	Measurement Campaign at Rospe WWTP .....	- 81 -
5.3.1	Data Description .....	- 81 -
5.3.2	Data Preparation .....	- 86 -

5.3.3	Generation of input data sets .....	- 89 -
5.4	Description of mathematical methods.....	- 90 -
5.4.1	Stepwise backward elimination .....	- 90 -
5.4.2	Regression Methods .....	- 92 -
5.4.3	Classification Methods .....	- 95 -
5.5	Results.....	- 98 -
5.5.1	Sensitivity Analysis .....	- 98 -
5.5.2	Regression Results COD .....	- 101 -
5.5.3	Regression Results $NH_4-N$ .....	- 103 -
5.5.4	Classification Results .....	- 104 -
5.6	Comparison to standard reference samples.....	- 108 -
5.6.1	Calculation of the Reference Standard .....	- 108 -
5.6.2	Back Transformation from Classes to Concentrations .....	- 111 -
5.6.3	Direct Comparison of Regression and Classification Results .....	- 111 -
5.6.4	2h mean Comparison of Regression and Classification Results.....	- 112 -
5.7	Conclusion .....	- 115 -
6	Application and Testing of Model-Based inflow estimation .....	- 116 -
6.1	Introduction to feed stream estimation.....	- 116 -
6.2	Methodology limitations .....	- 119 -
6.2.1	Effects based on retention time .....	- 119 -
6.2.2	Drifting of the state vector.....	- 121 -
6.3	Application of feed stream estimation to the Rospe model .....	- 122 -
6.3.1	Model adaption.....	- 122 -
6.3.2	Data generation and model stabilization.....	- 122 -
6.3.3	Parameterization of the Feed Stream Algorithm .....	- 123 -
6.4	Optimization Possibilities .....	- 124 -
6.5	Results of the feed stream estimation.....	- 125 -
6.6	Conclusion .....	- 128 -

7	Self-Organizing Map based operation regime estimation for state based control of Wastewater Treatment Plants .....	- 130 -
7.1	Introduction.....	- 130 -
7.2	Self-Organizing Maps .....	- 130 -
7.2.1	SOM Initialization and Training.....	- 131 -
7.2.2	SOM Validation.....	- 132 -
7.3	Clustering the SOM .....	- 133 -
7.4	Operational State-Based controller design.....	- 134 -
7.4.1	Implementation of the online Controller .....	- 134 -
7.5	Test Case 1: Modified BSM1 Model .....	- 135 -
7.5.1	Description of the Simulation Model .....	- 135 -
7.5.2	Model initialisation.....	- 136 -
7.5.3	Data generation and variable selection.....	- 136 -
7.5.4	Development of the fitness function.....	- 137 -
7.5.5	SOM Controller development CASE 1 .....	- 138 -
7.5.6	Results of test case 1.....	- 140 -
7.6	Test Case 2: Rospe Plant.....	- 142 -
7.6.1	Data generation and input variable selection Case 2 .....	- 142 -
7.6.2	SOM development and optimization.....	- 144 -
7.6.3	SOM Controller development for Case 2 .....	- 151 -
7.6.4	Development of a fitness function for Case 2 .....	- 152 -
7.6.5	Results and discussion for Case 2.....	- 153 -
7.7	Conclusion .....	- 155 -
8	Conclusion .....	- 156 -
	References .....	- 157 -
	Appendix .....	- 164 -

## List of Figures

FIGURE 1-1: IMPLEMENTATION OF THE URBAN WASTE WATER TREATMENT DIRECTIVE (UWWTD) (EUROPEAN ENVIRONMENT AGENCY, 2014).....	- 2 -
FIGURE 2-1: SIMPLIFIED REPRESENTATION OF NITROGEN REMOVAL IN WWTPS .....	- 12 -
FIGURE 2-2: CARBON AND NITROGEN BEHAVIOUR DURING NITRIFICATION IN AN IDEAL BATCH REACTOR.....	- 14 -
FIGURE 2-3: LAYOUT OF A TYPICAL WASTEWATER TREATMENT PLANT (EBEL, 2009).....	- 19 -
FIGURE 2-4: DOWNSTREAM DENITRIFICATION .....	- 21 -
FIGURE 2-5: UPSTREAM DENITRIFICATION.....	- 21 -
FIGURE 2-6: CASCADED DENITRIFICATION.....	- 22 -
FIGURE 2-7: SIMULTANEOUS DENITRIFICATION .....	- 22 -
FIGURE 2-8: INTERMITTENT DENITRIFICATION (ID) .....	- 23 -
FIGURE 2-9: ALTERNATING DENITRIFICATION (AD).....	- 24 -
FIGURE 3-1: O <sub>2</sub> -MEASUREMENT WWTP HEINZENBERG.....	- 28 -
FIGURE 3-2: TYPICAL POLLUTIONS OF PROBES .....	- 31 -
FIGURE 3-3: MEASUREMENT TEST OF DIFFERENT $NH_4-N$ ONLINE MEASUREMENT DEVICES AT THE ROSPE WWTP (2004).....	- 32 -
FIGURE 3-4: PRINCIPLE OF UV/VIS SPECTROMETRY .....	- 33 -
FIGURE 3-5: SIGNIFICANT REGIONS OF A TYPICAL UV/VIS ABSORPTION SPECTRUM (NON COMPENSATED) .....	- 34 -
FIGURE 3-6: SIGNIFICANT REGIONS OF A TYPICAL UV/VIS ABSORPTION SPECTRUM (COMPENSATED).....	- 35 -
FIGURE 3-7: SCHEMATIC ISE (WTW GMBH, 2007) .....	- 37 -
FIGURE 3-8: AERATION $NH_4-N$ CASCADE CONTROLLER .....	- 39 -
FIGURE 3-9: CONTROLLER SHELL MODEL.....	- 43 -
FIGURE 3-10: BEHAVIOUR $O_2$ -PI-CONTROLLER IN THE BEGINNING OF NITRIFICATION.....	- 44 -
FIGURE 3-11: AERATION CONTROL - INTERACTION WITH OTHER CONTROL LOOPS .....	- 45 -
FIGURE 3-12: PROCESS WATER CONTROL .....	- 48 -
FIGURE 4-1: STIRRED TANKS IN SERIES.....	- 52 -
FIGURE 4-2: GROWTH BEHAVIOUR - MONOD KINETIK .....	- 54 -
FIGURE 4-3: BIOLOGICAL STAGES OF THE ROSPE WWTP .....	- 61 -
FIGURE 4-4: BIOLOGICAL STAGES OF THE ROSPE WWTP .....	- 62 -
FIGURE 4-5: RESULTS OF A GRID SEARCH FOR THE OPTIMUM RECIRCULATION FLOW BASED ON $NH_4-N$ RMSE.-	- 63 -
FIGURE 4-6: RESULTS OF A GRID SEARCH FOR OPTIMUM RECIRCULATION FLOW BASED TO $NO_3-N$ RMSE.....	- 63 -
FIGURE 4-7: ANALYSERS AT ROSPE WWTP A: HACH-LANGE AMTAX AND B: HACH-LANGE NITRATAX ...	- 64 -
FIGURE 4-8: ROSPE MODEL OVERVIEW.....	- 65 -
FIGURE 4-9: INFLOW FRACTIONING.....	- 66 -
FIGURE 4-10: SECONDARY CLARIFIER.....	- 67 -
FIGURE 4-11: SECONDARY TREATMENT / BIOLOGICAL TREATMENT.....	- 68 -
FIGURE 4-12: APPLIED CALIBRATION PROCEDURE.....	- 71 -
FIGURE 4-13: COMPARISON OF DIFFERENT OPTIMIZATION RESULTS (BLUE: MEASURED, RED: SIMULATED) .....	- 75 -
FIGURE 4-14: PARAMETER OPTIMIZATION PARETO FRONT.....	- 76 -

FIGURE 5-1: INSTALLED MEASUREMENT PROBES: (A) ION-SELECTIVE AMMONIUM SENSOR, (B) TURBIMAX CUS51D TURBIDITY PROBE, (C) E+H STIPSCAN SPECTROMETRIC SENSOR, (D) TRIOS UV/VIS SPECTROSCOPIC  $COD/NO_3$  ANALYZER ..... - 81 -

FIGURE 5-2: SCATTERPLOT ATU VS. FNU (SCALED) ..... - 82 -

FIGURE 5-3: COMPARISON OF FNU AND ATU (SCALED)..... - 82 -

FIGURE 5-4: OVERVIEW OF RELEVANT MEASUREMENT VALUES ..... - 83 -

FIGURE 5-5: DISTRIBUTION OF MEASUREMENT VALUES IN THE MEASUREMENT CAMPAIGN ..... - 89 -

FIGURE 5-6: SEQUENCE OF STEPWISE BACKWARD ELEMINATION (SBE) ..... - 91 -

FIGURE 5-7: SBE ALGORITHM IN PSEUDO-CODE ..... - 92 -

FIGURE 5-8: PREDICTING  $COD$  – THE MOST SIGNIFICANT INPUT IS ON THE RIGHT ( $SAC_{254}$ ) AND THE LEAST SIGNIFICANT IS ON THE LEFT ( $COND.$ ) ..... - 99 -

FIGURE 5-9:  $COD$  - COEFFICIENT OF DETERMINATION ( $R^2$ ) FOR REGRESSION MODELS WITH DIFFERENT NUMBERS OF INPUT, WITH THE SELECTION ORDER GIVEN IN FIGURE 5-8..... - 100 -

FIGURE 5-10: PREDICTING  $NH_4-N$  – THE MOST SIGNIFICANT INPUT IS ON THE RIGHT ( $SAC_{254}$ ) AND THE LEAST SIGNIFICANT IS ON THE LEFT ( $COND.$ ) ..... - 100 -

FIGURE 5-11:  $NH_4-N$  - COEFFICIENT OF DETERMINATION ( $R^2$ ) FOR REGRESSION MODELS WITH DIFFERENT NUMBERS OF INPUT, WITH THE SELECTION ORDER GIVEN IN FIGURE 5-10..... - 101 -

FIGURE 5-12: REGRESSION RESULTS DATASET  $U_{1,COD}$  (50 RUNS MONTE CARLO SIMULATION)..... - 101 -

FIGURE 5-13: REGRESSION RESULTS DATASET  $U_{2,COD}$  (50 RUNS MONTE CARLO SIMULATION)..... - 102 -

FIGURE 5-14: PREDICTION USING SVR FOR  $U_{2,COD}$ (MEASURED: BLUE, PREDICTED: RED) ..... - 102 -

FIGURE 5-15: REGRESSION RESULTS DATASET  $U_{3,COD}$  (50 RUNS MONTE CARLO SIMULATION) ..... - 103 -

FIGURE 5-16: REGRESSION RESULTS DATASET  $U_{1,NH4}$  (50 RUNS MONTE CARLO SIMULATION)..... - 103 -

FIGURE 5-17: REGRESSION RESULTS DATASET  $U_{2,NH4}$  (50 RUNS MONTE CARLO SIMULATION)..... - 104 -

FIGURE 5-18: REGRESSION RESULTS DATASET  $U_{3,NH4}$  (50 RUNS MONTE CARLO SIMULATION)..... - 104 -

FIGURE 5-19: COMPARISON ON MEAN INFLOW VALUES ( $NH_4-N$ ) (BLUE CS, RED CSF) ..... - 110 -

FIGURE 5-20: COMPARISON OF MEAN INFLOW VALUES ( $COD$ )(BLUE CS, RED CSF) ..... - 110 -

FIGURE 5-21: RF CLASSIFICATION 2H MEAN  $U_{3,COD}$ (BLUE: MEASURED, RED: PREDICTED)..... - 113 -

FIGURE 5-22: SUPPORT VECTOR REGRESSION 2H MEAN  $U_{2,COD}$ (BLUE: MEASURED, RED: PREDICTED) ..... - 114 -

FIGURE 5-23: PARTIAL LEAST SQUARES REGRESSION 2H MEAN  $U_{1,NH4}$  (BLUE: MEASURED, RED: PREDICTED)- 114 -

FIGURE 5-24: LDA 2H MEAN  $U_{2,NH4}$  (BLUE: MEASURED, RED: PREDICTED) ..... - 114 -

FIGURE 6-1: MODEL BASED INFLOW ESTIMATION PRINCIPLE (EBEL, 2009) ..... - 117 -

FIGURE 6-2: MODEL BASED INFLOW ESTIMATION PRINCIPLE USING AN ADDITIONAL SURROGATE MODEL (EBEL, 2009) ..... - 118 -

FIGURE 6-3: STEP RESPONSE OF THE ROSPE MODEL TO A  $NH_4-N$  STEP FROM 9.5  $MG/L$  TO 15  $MG/L$  (SCALED) ...- 120 -

FIGURE 6-4: ALTERNATIVE FEED STREAM ESTIMATION FOR DIMENSION REDUCTION (RED: ORIGINAL FEED STREAM ESTIMATION, BLUE: MODIFIED FEED STREAM ESTIMATION) ..... - 124 -

FIGURE 6-5: INFLOW ESTIMATION (ORIGINAL VERSION) COMPARISON VARIABLES  $NO_3-N$  AND  $NH_4-N$  USING TWO STEPS OF 120 MINUTES (OPTIMIZATION HORIZON 240 MINUTES)..... - 126 -

FIGURE 6-6: INFLOW ESTIMATION FOR  $COD$  AND  $NH_4-N$  USING 2 STEPS A 120 MINUTES (240 MINUTES OPTIMIZATION HORIZONT) AND  $NH_4-N$  AND  $NO_3-N$  AT TANK D4 FOR COMPARISON..... - 126 -

FIGURE 6-7: INFLOW ESTIMATION (MODIFIED VERSION) COMPARISON VARIABLES  $\text{NO}_3\text{-N}$  AND  $\text{NH}_4\text{-N}$  USING 360 MINUTE'S OPTIMIZATION HORIZON ..... - 127 -

FIGURE 6-8: INFLOW ESTIMATION (MODIFIED VERSION) FOR  $\text{COD}$  AND  $\text{NH}_4\text{-N}$  USING 360 MINUTES OPTIMIZATION HORIZON AND  $\text{NH}_4\text{-N}$  AND  $\text{NO}_3\text{-N}$  AT TANK D4 FOR COMPARISON ..... - 128 -

FIGURE 7-1: ADJACENT NEURONS ON RECTANGULAR (LEFT) AND HEXAGONAL (RIGHT) GRIDS ON SOMs (BMU – RED, ADJACENT NEURONS – GREEN, CORNER NEURONS – VIOLET, NOT ADJACENT NEURONS – WHITE) ..... - 133 -

FIGURE 7-2: OPERATION REGIME CONTROLLER (ORG) WORKING PRINCIPLE ..... - 134 -

FIGURE 7-3: DESIGN OF THE MODIFIED BENCHMARK SIMULATION MODEL No. 1 ..... - 135 -

FIGURE 7-4: BSM1 DRY WEATHER INFLOW SCENARIO ..... - 136 -

FIGURE 7-5: AMMONIUM DEGRADATION IN AN AERATED BIOREACTOR UNDER THE ASSUMPTION THAT ALL VARIABLES EXCEPT AMMONIUM AND OXYGEN ARE CONSTANT ..... - 137 -

FIGURE 7-6: COMMUNICATION PRINCIPLE BETWEEN MATLAB AND THE JAVA SOMTOOLBOX ..... - 138 -

FIGURE 7-7: RECORDED STATES OVER A SIMULATION PERIOD OF THREE DAYS ..... - 141 -

FIGURE 7-8: SMOOTHED DATA HISTOGRAM OF THE SOM CLUSTERED WITH WARD LINKAGE ALGORITHM (LIGHTER COLOURS INDICATE AREAS WITH HIGHER DATA DENSITY ON THE MAP) ..... - 141 -

FIGURE 7-9: OPTIMIZATION RESULTS OF THE TOPOLOGY ERROR FOR DATA SET DS1 ..... - 145 -

FIGURE 7-10: OPTIMIZATION RESULTS OF THE QUANTIZATION ERROR FOR DATA SET DS1 ..... - 145 -

FIGURE 7-11: OPTIMIZATION RESULTS OF THE TOPOLOGY ERROR FOR DATA SET DS2 ..... - 146 -

FIGURE 7-12: OPTIMIZATION RESULTS OF THE QUANTIZATION ERROR FOR DATA SET DS2 ..... - 146 -

FIGURE 7-13: DETERMINATION OF THE OPTIMAL NUMBER OF CLUSTERS USING SILHOUETTE CRITERION FOR TWO DIMENSIONAL SOM WEIGHTS USING WARD LINKAGE CLUSTERING. .... - 147 -

FIGURE 7-14: WARD LINKAGE CLUSTERS FOR DATA SET DS1 WITH DE NORMALIZED WEIGHT VECTORS ..... - 147 -

FIGURE 7-15: WARD LINKAGE CLUSTERS FOR DATA SET  $DS_7$  WITH DE NORMALIZED WEIGHT VECTORS AND DIURNAL CYCLE OF DAY 1, DAY8 AND DAY 13 ..... - 148 -

FIGURE 7-16: DETERMINATION OF THE OPTIMAL NUMBER OF CLUSTERS USING SILHOUETTE CRITERION FOR THREE DIMENSIONAL SOM WEIGHTS USING WARD LINKAGE CLUSTERING. .... - 149 -

FIGURE 7-17: SOM NEIGHBOR WEIGHT DISTANCES  $DS_2$  (DARKER COLORS INDICATE BIGGER DISTANCES IN THE INPUT SPACE BETWEEN NEURONS) ..... - 149 -

FIGURE 7-18: WARD LINKAGE CLUSTERS FOR DATA SET DS3 WITH DE NORMALIZED WEIGHT VECTORS ..... - 150 -

FIGURE 7-19: WARD LINKAGE CLUSTERS FOR DATA SET DS3 WITH DE-NORMALIZED WEIGHT VECTORS INCLUDING DIURNAL CYCLE OF DAY 1 FROM TWO ANGLES ..... - 151 -

FIGURE 7-20: DETERMINATION OF THE OPTIMAL NUMBER OF CLUSTERS USING WARD LINKAGE AND SILHOUETTE CRITERION AS QUALITY MEASURE FOR THREE DIMENSIONAL DATA SET DS3 ..... - 152 -

FIGURE 7-21: SOM NEIGHBOUR WEIGHT DISTANCES FOR 30X30 MAP USING DS3 ..... - 152 -

FIGURE 7-22: FITNESS RESULTS CONTROLLER OPTIMIZATION ..... - 154 -

FIGURE 7-23: OPERATION REGIMES OF THE ROSPE PLANT BETWEEN DAY 12 AND DAY 14 ..... - 154 -

## List of Tables

TABLE 3-1: TYPICAL TECHNICAL PROPERTIES OF ACTUATORS .....	29 -
TABLE 3-2: TYPICAL WWTP PROCESS VARIABLES AND THEIR RELEVANCE FOR CONTROL APPLICATIONS.....	30 -
TABLE 3-3: OVERVIEW OF MAIN CONTROL LOOPS IN WWTPS .....	39 -
TABLE 4-1: MATRIX REPRESENTATION OF MODELS (HENZE ET AL., 2000).....	55 -
TABLE 4-2: ASM1 PROCESSES .....	58 -
TABLE 4-3: ASM1 DEFAULT PARAMETER VALUES (JEPPSSON, 1996) .....	59 -
TABLE 4-4: WWTP ROSPE - BASIC DATA .....	60 -
TABLE 4-5: WWTP ROSPE - LIMIT / DESIGN VALUES .....	61 -
TABLE 4-6: SECONDARY CLARIFIER SETTLING PARAMETER .....	67 -
TABLE 4-7: DENITRIFICATION AND NITRIFICATION TANK VOLUMES.....	69 -
TABLE 4-8: SMS-EGO PARAMETERS.....	74 -
TABLE 4-9: RMSE COMPARISON OF DIFFERENT MODEL PARAMETER SETS.....	75 -
TABLE 5-1: OVERVIEW OF MEASUREMENT VALUES.....	85 -
TABLE 5-2: COD AND NH <sub>4</sub> -N CLASSES .....	89 -
TABLE 5-3: MLP OPTIMIZATION - OPTIMAL NUMBER OF NEURONS FOR THE DATASETS .....	94 -
TABLE 5-4: SVR OPTIMIZATION - OPTIMAL PARAMETER FOR DATASETS.....	95 -
TABLE 5-5: LDA OPTIMIZATION - OPTIMAL NUMBER OF EIGENVECTORS FOR THE DATASETS.....	97 -
TABLE 5-6: RF OPTIMIZATION - OPTIMAL NUMBER OF DECISION TREES FOR CLASSIFICATION AND REGRESSION FOR EACH DATASET.....	97 -
TABLE 5-7: SVM OPTIMIZATION - OPTIMAL PARAMETER FOR DATASETS.....	98 -
TABLE 5-8: MEDIAN (20 REPETITIONS) CLASSIFICATION RESULTS FOR COD TEST DATA -NMCR[%].....	105 -
TABLE 5-9: CONFUSION MATRICES DATASET U <sub>1,COD</sub> .....	106 -
TABLE 5-10: CONFUSION MATRICES DATASET U <sub>2,COD</sub> .....	106 -
TABLE 5-11: CONFUSION MATRICES DATASET U <sub>3,COD</sub> .....	106 -
TABLE 5-12: MEDIAN (20 REPETITIONS) CLASSIFICATION RESULTS FOR NH <sub>4</sub> -N TEST DATA -NMCR[%] .....	107 -
TABLE 5-13: CONFUSION MATRICES DATASET U <sub>1,NH4</sub> .....	107 -
TABLE 5-14: CONFUSION MATRICES DATASET U <sub>2,NH4</sub> .....	108 -
TABLE 5-15: CONFUSION MATRICES DATASET U <sub>3,NH4</sub> .....	108 -
TABLE 5-16: CLASS MEAN MEASUREMENT VALUES.....	111 -
TABLE 5-17: NMSE DIRECT COMPARISON OF REGRESSION AND CLASSIFICATION RESULTS [x100%] .....	112 -
TABLE 5-18: NMSE COMPARISON RESULTS FOR TEST DATA: VIRTUAL 2H-COMPOSITE SAMPLES [x100%]..	113 -
TABLE 6-1: PARAMETERIZATION OF THE FEED STREAM ALGORITHM.....	123 -
TABLE 6-2: PARAMETERIZATION OF THE KRIGING MODEL .....	123 -
TABLE 6-3: PARAMETERIZATION OF THE PSO ALGORIHTM .....	124 -
TABLE 6-4: CONFIGURATION OF THE MODIFIED FEED STREAM ESTIMATION .....	125 -
TABLE 7-1: GA SOM PARAMETERS CASE 1 .....	139 -
TABLE 7-2: GA SOM OPTIMIZATIN RESULTS CASE 1 .....	139 -

TABLE 7-3: SOM PREDICTION MODEL CLUSTERING EVALUATION USING SILHOUETTE FUNCTION FOR THREE CLUSTERS ..... - 140 -

TABLE 7-4: OPERATION REGIME OXYGEN SET POINTS CASE 1..... - 141 -

TABLE 7-5: COMPARISON OF THE PLANT PERFORMANCE CASE 1..... - 142 -

TABLE 7-6: CONTROL SCENARIOS TO CREATE DIFFERENT OPERATING CONDITIONS FOR SOM TRAINING..... - 143 -

TABLE 7-7: CHARACTERISTICS OF DATA SET DS3 ..... - 143 -

TABLE 7-8: INTERPRETATION OF SOM CLUSTERS FOR DATA SET DS2..... - 150 -

TABLE 7-9: OPERATION REGIME OXYGEN SET POINTS CASE 2..... - 154 -

TABLE 7-10: COMPARISON OF DIFFERENT SET POINTS FOR THE ROSPE PLANT..... - 155 -

## Nomenclature

Abbreviation	Definition
$FO_C$	flux of oxygen utilised
$Q_R$	Recirculation
$Q_{RS}$	Return Sludge
$Q_{ES}$	Excess sludge
$U_{ion}$	Potential of the electrodes
$a_{ion}$	The active concentration of the target ion
$K_{ion,interf}$	dimensionless coefficient which describes the influence of other ions
$a_{interf}$	activity of these ions
$Q_{PA}$	Flocking Agent flow
$Q [l / s]$	Inflow
$C_P$	Phosphate concentration
$\beta [Fe / molP]$	Beta-Value
$p [kg / m^3]$	Flocking Agent density
$m_{ME} [kgFe / kgPA]$	effective metal content
AD	Alternating Denitrification
AD	Anaerobic Digestion
ASM1	Activated Sludge Models No. 1 , auch ASM2, ASM 3
ATU	Absorptiometric Turbidity Units
BMU	nearest weight vector
BOD	Biological Oxygen Demand
BSM1	Benchmark Simulation Model No.1
C:N:P ratio	Carbon:Nitrogen:Phosphorous Ratio
CDT	Clarifier Detention Time
CL	Closed Loop – feed back
COD	Chemical Oxygen Demand
CS	Composite Sample
CSF	Composite Sample considering the Flow
CSTR	Continuously stirred tank reactors
Da	Dalton
DIN	German industry norm
DT	Decision Trees
DWA	Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall e. V. – German Association for Water Wastewater and Waste
EBPR	Enhanced Biological Phosphate Removal
EQ	Effluent Quality

## Nomenclature

F	Faraday constant
FF	Feed Forward
FL	Fill level
<i>FNU</i>	Formazin Nephelometric Units
GA	Genetic Algorithm
GECO►C	Gummersbach Environmental Computing Center
GMP	Good Modelling Practice
IAWPRC	International Water Association on Water Pollution Research and Control
IAWQ	International Association on Water Quality
ICA	Instrumentation and Control Applications
ID	Intermittend Denitrification
IE3	International Efficiency class 3
IEC	International Electrotechnical Commission
ISE	Ion-Selective Electrode
IWA	International Water Association
LDA	Linear Discriminant Analysis
LIBSVM	Support Vector Machine Toolbox
LSR	Least squares regression
MA	Moving Average
MAD	Median Absolute Deviation
MCR	Misclassification Rate
MFM	magnetic flow meter
MID	Magnetic inductive flow meter
MLP	Multi-Layer Perceptrons
MLR	Multivariate linear regression
MOOA	Multi-Objective Optimization Algorithms
N	Nitrogen
NMCR	Normal Misclassification Rate
NMSE	Normalized Mean Squared Error
NS	Neighbourhood Size
$O_2$	Oxygen
OHOs	Ordinary Heterotrophic Organisms
OL	Open Loop – feed forward
PAOs	Phosphorus Accumulating Organisms
PC	Predictive Control
PE	Population Equivalents
PI or PID controller	Proportional Integral Differential controller
PLC	Programmable Logic Controller

## Nomenclature

PLS	Partial Least Squares
PVC	Polyvinyl chloride
PW	Process water
PWM	Pulse Width Modulation
R	gas constant
$R^2$	coefficient of determination
RBF	Radial Basis Function
Ref	Reference electrode
RF	Random Forest
RMSE	Root Mean Squared Error
SBE	stepwise backward elimination algorithm
SCADA	Supervisory Control and Data Acquisition
SD	Simultaneous Denitrification
SIMBA	Matlab Toolbox for
SLR	Surface Loading Rate
SOM	Self-Organizing Maps
SP	Set Point
<i>SRT</i>	sludge age
SVI	Sludge Volume Index
SVM	Support Vector Machines
SVR	Support Vector Regression
SWT	Storm Water Tanks
T	Temperature in Kelvin
TOC	Total Organic Carbon
TS	Total Solids
TSS	Totally Suspended Solids
UB	Upper Bound
UV/VIS	UV and visible light
UWWTD	Urban Wastewater Treatment Directive
WHG	Wasserhaushaltsgesetz
ws	window size
WWTP	Wastewater treatment plants

## 1 Introduction

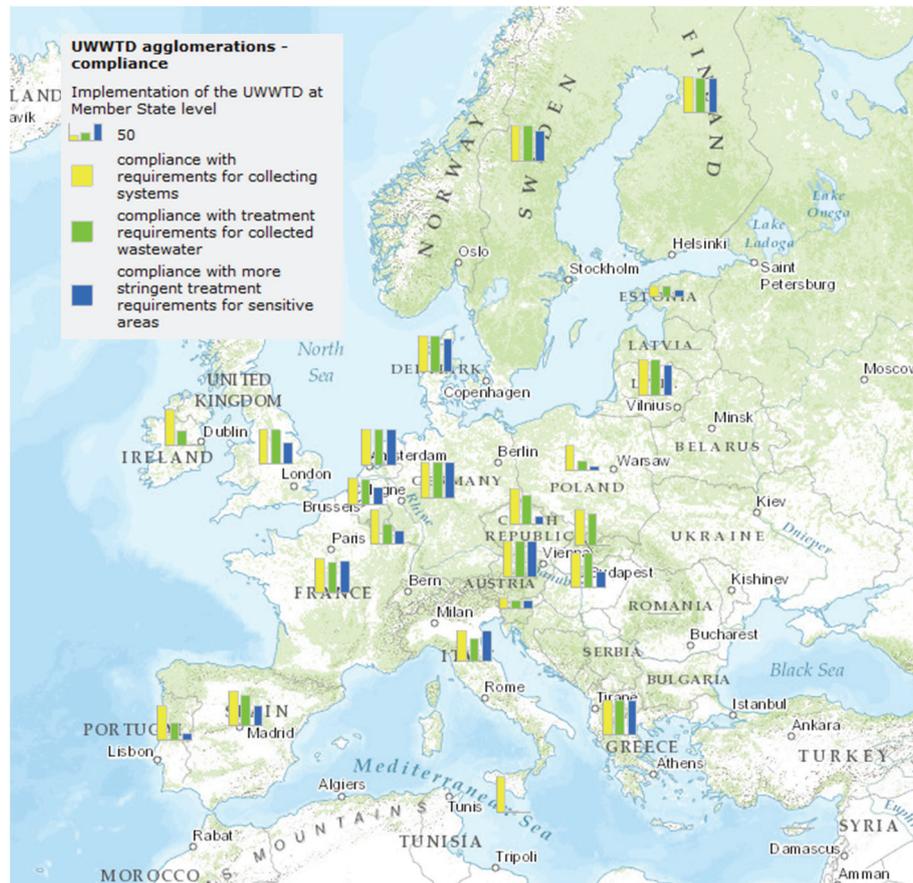
*“To summarize the summary of the summary: People are a problem.”*

*Douglas Adams, The Restaurant at the End of the Universe (1980)*

In their natural state, rivers and water bodies have the impressive capability to clean themselves due to the microorganisms living in them. As is the case with most things, this capability has its limits. Problems arise when it is overburdened and the carrying capacity of the water body is exceeded. In preindustrial times most of the pollutions or nutrients originated from plants and animals living in the surrounding areas and, for example, flushed from the riverbanks during rain events. The situation today is different. Nowadays, water bodies suffer from the contamination by different sources that are mainly artificial. The agricultural industry uses high amounts of artificial fertilizers which are in parts flushed into the rivers during rain events. This effect is further enhanced by the fact that many surfaces are sealed or have a reduced absorption capacity due to soil compaction. Besides the diffuse inputs, the water bodies are burdened by point sources, like sewer systems or wastewater treatment plants (WWTP). The basic concept of most sewer systems is to gather the wastewater of an “unnatural” large agglomeration of people and lead it to a single location. Besides high amounts of pollutants and nutrients concentrated at a single point, the wastewater reaches the water body in a significantly shorter time. Especially in areas with combined sewers flushing surges appear in case of rain events. It is therefore not surprising that small water bodies are not able to cope. The described effects lead to eutrophication, oxygen depletion and eventually river death. The consequences are not only a problem for the river and the organisms living in it, but also for the people in the area. Diseases like cholera, dysentery, typhoid fever, hepatitis etc. occurred in the past as a direct result of the use of contaminated water. These effects are well known from the past and still happen today in many countries without proper treatment facilities for wastewater. This alone explains the high importance of wastewater treatment prior to the release into a water body.

On a European level this issue is addressed by the European Council in its Urban Wastewater Treatment Directive 91/271/EEC (UWWTD) (Directive, 1991, p. 271), which was adopted in 1991 and implemented by 31<sup>st</sup> December 1998. The UWWTD requires the collection and treatment of waste water in agglomerations of more than 2000 population equivalents (PE) and more advanced treatment for agglomerations of more than 10,000 PE. Furthermore, the member states have to ensure certain limit values for discharge concentrations as well as a minimum percentage of reduction of Chemical Oxygen Demand (*COD*), Biological Oxygen Demand (*BOD*), Total Suspended Solids (*TSS*), phosphorus (*P*) and nitrogen (*TNb*).

Figure 1-1 gives an overview of the implementation of the UWWTD in the different member states of the EU. While the implementation in the Netherlands, Austria and Germany reaches already high levels, the envisaged objectives are not met in other member states.



**Figure 1-1: Implementation of the Urban Waste Water Treatment Directive (UWWTD) (European Environment Agency, 2014)**

Following the subsidiarity principle of the EU, member states have to implement national law which fulfils the guidelines of the UWWTD. For example, in Germany this is done in the water resources act (WHG – Wasserhaushaltsgesetz) (Deutschland and Bundesministerium der Justiz, 2009). With the stringent requirements on effluent values of WWTP other problems for the operators arise. Demographic change, or more precisely rural-urban migration, amplifies this problem. Currently many plants in rural areas are under loaded while urban plants are operated at their load limits.

Whilst many developments in the area of advanced control strategies for WWTP have been made in the last number of years, it is difficult to find success stories of control systems which outperform conventional feedforward-feedback controllers in the long run (Åmand et al., 2013). From personal experience in the GECO►C Research Group of developing control systems for more than 10 WWTP over the last decade, I am convinced that it is very challenging to build up a complex system which is superior to conventional systems in the long run. The real problem is that the effort required to develop

advanced control systems and to adapt them during operation is often too high to justify the benefit. Nevertheless, I am convinced that in line with the ongoing development in sensor technology, automation systems and instrumentation, advanced control systems, which today run often only for test and academic purposes, will become more and more interesting for operators and become state of the art eventually.

## 1.1 Aims and Scope of the Thesis

The aim of this thesis is to develop novel instrumentation and control strategies for wastewater treatment plants in order to improve energy efficiency while guaranteeing process stability and maintainence of high cleaning capacity. To achieve these ambitious and even contradictory objectives, a combination of online measurement systems, computational intelligence and machine learning methods as well as dynamic simulation models was used. Furthermore, the focus of the thesis is on the development and assessment of practice-oriented and realisable as well as reliable solutions that do not require sophisticated, and often maintenance intensive, instrumentation, control and automation (ICA) equipment. The overall goal is to work with what is available and to make the best of it. The benefit of such an approach is the high acceptance by the plant operator of these solutions as well as the easy transfer to other WWTP.

Analysing what is available at WWTP that can be used for instrumentation and control, two things stand out:

- (1) The degree of online instrumentation in the inflow and in the bioreactors is normally good and allows sufficient monitoring of the processes, but it would be advantageous to measure additional process variables in the WWTP inflow to allow for faster control response times.
- (2) Large volumes of operational and process data are collected and available at most WWTP, but very seldom used for simulation, optimization and control purposes. Important process knowledge is wasted.

Therefore, the scope of the thesis is to set the scene by introducing the state of the art in the fields of WWTP operation, process monitoring and control and to subsequently develop and present three novel methods and to assess their potential for practical implementation. Those are, on the one hand, the automated calibration of a simulation model for the Rospe WWTP that provides a basis for the development and evaluation of the subsequent methods, and on the other hand, the development of soft sensors for the WWTP inflow which estimate the crucial process variables *COD* and *NH<sub>4</sub>-N*, and the estimation of WWTP operating states using Self-Organising Maps (SOM) that are used to determine the optimal control parameters for each state. These collectively, provide the basis for achieving comprehensive WWTP optimization.

## 1.2 Contributions

Based on the defined aims and scope, the thesis makes major contributions to the field of WWTP instrumentation and control in terms of simulation model development, process data analysis and advanced but operationally robust process control. The major contributions are as follows:

**(1) Development of a full scale simulation model for the Rospe WWTP**

The development of the simulation model is a significant contribution as it provides a challenging test bed for the development and evaluation of the novel ICA methods. The simulation model represents the real WWTP with its main hydraulic and biological characteristics and thus ensures that the developed ICA strategies can be successfully implemented in practice and are of high relevance to practitioners.

**(2) Automated calibration of a full scale WWTP using the non-deterministic multi-objective optimization method SMS-EGO**

The calibration of dynamic WWTP simulation models is a major challenge due to the fact that many model parameters are extremely difficult to measure, which normally requires their estimation based on experience. The proposed automated calibration method is a novel and sophisticated solution to this problem. By optimizing the model calibration with respect to two separate fitness functions rather than a single combined fitness function (i.e. the sum of the RMSE of  $NH_4-N$  and  $NO_3-N$ ), a more target-oriented calibration is possible. Furthermore, model calibration is much faster and available to a broader user base due to its easy applicability.

**(3) Development of a virtual COD and  $NH_4-N$  measurement system for the WWTP inflow**

When it comes to model calibration as well as to advanced control strategies, online measurements of  $COD$  and  $NH_4-N$  in the WWTP inflow are absolutely necessary. Unfortunately, online instrumentation for these parameters is expensive and maintenance intensive, which is why most WWTP do not have them in the inflow. The development of a virtual measurement system, so-called soft sensors is a solution. Based on available process data from standard instrumentation, machine learning methods, both regression and classification methods, were configured and trained to estimate  $COD$  and  $NH_4-N$  in the WWTP inflow. In order to evaluate the performance of the soft sensors an extensive measurement campaign was conducted at the Rospe WWTP. Furthermore, a toolbox for data pre-processing and sensitivity analysis using a stepwise backward elimination algorithm was developed to improve the machine learning results.

**(4) Evaluation of a model-based inflow estimation method for  $COD$  and  $NH_4-N$**

An alternative to the method described in the third contribution is to use an already existing fully calibrated simulation model to retrospectively estimate the inflow to WWTP, as proposed in Ebel (2009). This involves simulating the response of the WWTP model to different concentration of  $COD$  and  $NH_4-N$  applied in discrete steps. Based on a performance function that compares the simulated concentrations measured in the biological stages with the actual values over a defined

estimation horizon, estimating the inflow concentrations can be formulated as an optimisation problem. As this optimization is high dimensional and complex Kriging surrogate models of the performance function are used to speed up the estimation process. An evaluation of the methodology for the Rospe WWTP model highlights that it is severely impacted by variable retention times as well as blending of different wastewater and recirculation streams in a WWTP, with the result that it has limited utility.

**(5) Identification of the operating regimes of a WWTP**

A good control method is one that adapts to changing process states, such that gives the optimal response under all operating conditions. In reality, there is mostly one control strategy for a control loop that is not adapted to varying process conditions but more a compromise as the different process states are unknown. Using Self-Organising Maps (SOM) process data is analysed and the information mapped onto a two-dimensional feature space, in order to visualise the different operating regimes of a WWTP. This information can then be used for the development of optimal controllers for each operating regime.

**(6) Development and optimization of a state-based controller for different WWTP operating regimes**

The information on WWTP operating regimes generated using SOM was used to develop a state-based aeration controller, that chooses the optimal oxygen concentration set point for the nitrification tanks with regard to energy consumption and cleaning efficiency. In order to identify these set points a Genetic Algorithm (GA) was used. Results show that a significant improvement in plant operation can be achieved.

### 1.3 Outline

The remainder of the thesis is organised as follows:

Chapter 2 provides the necessary background to the field of wastewater treatment using activated sludge with a focus on describing the whole process chain and the related challenges. The chapter begins with a definition for wastewater, and its different types and sources followed by an overview of biological wastewater treatment and biochemical processes. Then, wastewater infrastructure from the collection to the final treatment and different plant designs are introduced and briefly explained. The chapter concludes with a summary, highlighting the resulting challenges for control design and implementation.

Chapter 3 describes the main areas of WWTP control, introducing the most common instrumentation and control loops from the point of view of an automation engineer. This chapter, thus sets the context for all following development chapters. Initially an overview of typical online instrumentation, actuators and the challenges for control are presented. Then the typical control loops and different strategies for WWTP control are described, and finally a summary of the state-of-the-art control strategies is presented.

As modelling and simulation are important to the development and testing of novel optimization and control methodologies, chapter 4 is devoted to WWTP modelling. The basics of the modelling of hydraulic and biological processes are introduced, followed by the presentation of different activated sludge models and their matrix representation. Then the development of the Rospe WWTP simulation model is presented. A detailed description of the plant and the modelling process are provided and the related modelling challenges are thoroughly described. Finally, the chapter introduces a novel automated model calibration procedure that uses a multi-objective optimization algorithm that allows faster and more accurate model calibration compared to manual approaches based on expert knowledge. The resulting calibrated Rospe WWTP model is used for the experiments conducted in chapter 6 and 7.

Chapter 5 describes the development of virtual *COD* and *NH<sub>4</sub>-N* sensors starting off with a description of the particularities of the inflow measurement of a WWTP, followed by a detailed description of the measurement campaign conducted at the Rospe WWTP plant, whose goal was to get a comprehensive data basis for the application of machine learning methods. The necessary data pre-processing steps as well as the mathematical background for applied regression and classification methods are introduced and then a detailed evaluation of regression and classification algorithms presented for a number of virtual sensor scenarios.

Chapter 6 presents an alternative to the soft sensor approach of chapter 5, where a model-based numerical-based inflow estimation for *COD* and *NH<sub>4</sub>-N* is applied and tested using the Rospe WWTP simulation model. The chapter begins with an introduction to the inflow estimation method developed by (Ebel, 2009), followed by an investigation of the limitations in relation to the Rospe WWTP model. An extension of the estimation method by Ebel (2009) using longer estimation horizons with constant inflow concentrations is then introduced. Results are then presented for both implementations and their major limitations highlighted.

Chapter 7 presents a state-based control method based on operation regime estimation on available process data. The chapter begins with an introduction to the concept of operating regime decomposition, followed by a short description of self-organizing maps (SOM) and SOM clustering as a means of achieving this. After the fundamentals are explained, an operational state controller design is presented. The chapter then concludes with an evaluation of the proposed methodology on a standard benchmark model (BSM1) and the Rospe plant model.

Chapter 8 summarizes the contributions of the thesis and provides an outlook on possible future work.

## 1.4 Publications

Bongards M, Schaefer S, Ebel A, Kern P. 2006. Wastewater Treatment Plant Improvement by Smart Sensors and Computational Intelligence (WAPSCIENCE). Workshop: Environmental Research for SMEs – Technology Challenges and Market Opportunities in Waste Water Treatment: Brussels

- Bongards, M., Hilmer, T., Kern, P., 2007. Online-Konzentrationsmessung in Kanalnetzen - Technik und Betriebsergebnisse. Forschungsbericht der Fachhochschule Köln 173–176.
- Kern, P., Wolf, C., Bongards, M., Gaida, D., 2010. Zustandsbasierte Regelung von Kläranlagen, in: Forschungsbericht Der Fachhochschule 2010. Joh.Heider Verlag GmbH, Köln, pp. 144–147.
- Ludwig, T., Kern, P., Bongards, M., Wolf, C., 2011. Simulation and optimization of an experimental membrane wastewater treatment plant using computational intelligence methods. *Water Science and Technology* 63, 2255–2260. doi:10.2166/wst.2011.135
- Kern, P., Wolf, C., Bongards, M., Oyetoyan, T.D., McLoone, S., 2011. Self-Organizing Map based operating regime estimation for state based control of wastewater treatment plants, in: 2011 International Conference of Soft Computing and Pattern Recognition (SoCPaR). Presented at the 2011 International Conference of Soft Computing and Pattern Recognition (SoCPaR), pp. 390 – 395. doi:10.1109/SoCPaR.2011.6089275
- Ludwig, T., Gaida, D., Keysers, C., Pinnekamp, J., Bongards, M., Kern, P., Wolf, C., Sousa Brito, A.L., 2012. An advanced simulation model for membrane bioreactors: Development, calibration and validation. *Water Science and Technology* 66, 1384–1391. doi:10.2166/wst.2012.249
- Kern, P., Wolf, C., Gaida, D., Bongards, M., McLoone, S., 2014. COD and NH<sub>4</sub>-N estimation in the inflow of Wastewater Treatment Plants using Machine Learning Techniques, in: *Automation Science and Engineering (CASE)*, 2014 IEEE International Conference on. IEEE, pp. 812–817.
- Kern, P., Ebel, A., Wolf, C., Bongards, M., McLoone, S., 2014. Simulation based feed stream estimation of wastewater treatment plants using a surrogate model, in: *Wastewater and Biosolids Treatment and Reuse: Bridging Modeling and Experimental Studies*, Dr. Domenico Santoro, Trojan Technologies and Western University Eds, ECI Symposium Series. [http://dc.engconfintl.org/wbtr\\_i/32](http://dc.engconfintl.org/wbtr_i/32)

Kern, P., Bongards, M., Gros, S., Sterger, O., Rudolph, P., 2014. Energieoptimierung von Kläranlagen durch Internet-gestützte Analyse und Simulation. WWT-Online, Modernisierungsreport 2013/2014.

Gahr, A., Wolf, C., Kern, P., 2014. Wasser und Abwasser in Megastädten der Zukunft. atp edition 56, 28–37.

Kern, P., Bongards, M., Cronrath, A., 2015. Wie den Stromverbrauch auf Kläranlagen reduzieren? gwf-Wasser|Abwasser, Netzwerk Wissen, Abwasser, ISSN 0016-3651, B5399, pp.745–756.

## 2 Wastewater Treatment

This chapter aims to give an introduction into the field of wastewater treatment using activated sludge. The common process of wastewater treatment using nitrification and denitrification and the associated process engineering is widely known and well described in literature. The first articles about nitrification were published by Sergei Winogradsky in 1890 (Winogradsky, 1890). For this reason this chapter focuses on the aspects of the treatment process that are relevant for automation and optimization purposes. After a short introduction to the topic, the description follows the water from the origin of the wastewater, over the sewer system, through the WWTP to the connected water body.

### 2.1 Wastewater

In general wastewater is water whose quality was negatively affected by human activity. This means that it can be contaminated with a wide range of substances. Despite this diversity of possible contaminants, the typical biological treatment process using nitrification and denitrification focusses on the main pollution with carbon, nitrogen and phosphorus. In addition to that, different mechanical steps remove particulate matter.

#### 2.1.1 Sources and Types of Wastewater

While wastewater is different from region to region the sources are mainly similar. The following list gives an overview of typical sources:

- Domestic Wastewater
- Industrial Wastewater
- Municipal Water
- Agricultural Wastewater
- Infiltration Water
- Surface Water

**Domestic wastewater** is the most relevant type for this thesis and is basically water which was used by households or similar facilities and can be further divided into two groups: 1. Black water, which is mainly affected by faeces and urine from people or animals. 2. Grey water, which is polluted from bathing, washing, cooking and similar activities. In the majority of all cases black and grey water are not separated and reach the sewer system as a mixture. Important for the treatment process is the fact that these activities lead to a certain composition of pollutants / nutrients in the wastewater. This fact can be used by intelligent control systems, e.g. for the prediction of measurement values (soft-sensors, see chapter 5). Particularly interesting is the relationship between carbon, nitrogen and phosphorus, because of its importance for the biological processes during nitrification and denitrification, which will be described in section 2.2.

**Industrial wastewater** is water polluted during industrial processes like fabrication, cleaning or cooling. The composition of industrial wastewater can be completely different from domestic wastewater and depends highly on the type of industry discharger. This means that every source has its own characteristics. While one industrial discharger releases high amounts of *COD* (e.g. food processing industry) into the water, others are characterized by high nitrate values. If the production process is organized in batches, the composition can change significantly from day to day. These imponderables complicate the application of control and optimization techniques. Furthermore, industrial wastewater can be contaminated with substances that can only be measured in a laboratory.

The term **municipal wastewater** refers to water from a city and means a mixture of domestic and industrial wastewater.

**Agricultural wastewater** is water used on farms. Typical contaminants are faeces, fertilizers, pesticides etc. For treatment plants only the agricultural wastewater from point sources, like live breeding facilities are of interest. Contaminated water in fields is usually not collected and flows directly into the groundwater or water body.

**Infiltration water** is a special kind of wastewater. In a narrow sense it could be argued that it is not really wastewater. Infiltration water is water (usually groundwater) which seeped into the sewer system. In the sewer it is then mixed with the other types of wastewater and transported to the WWTP. Therefore, it is treated in the WWTP, but mainly leads to a dilution of the other types of wastewater. During optimization or controller development it has to be considered, because the amount varies over the year and changes with the ground water level.

**Surface wastewater** is urban run-off water from streets, parking lots and basically all sealed surfaces adjacent to a sewer system. Therefore, it can be contaminated with oil, fuel, rubber, exhaust, street abrasion, etc. For the treatment plant and the control strategy it becomes relevant during rain events when the WWTP is connected to a combined sewer as the influent to the plant increases significantly.

## 2.2 Biochemical Wastewater Treatment

This subsection gives a short introduction into the biochemical processes of wastewater treatment using activated sludge. A detailed description of the complex processes can be found, for example in *Biological Wastewater Treatment* (Henze et al., 2008). The biological processes in common wastewater treatment plants focus typically on the removal of carbon, nitrogen and phosphorus. Although the processes are considered individually, they are closely linked and influence each other. Other substances than the ones mentioned before, such as pharmaceutical compounds or heavy metals, which can be partly removed during the treatment process are not targeted specifically. The removal of these substances can

be considered a useful and desirable side effect. In many cases the removal is rather a physical effect than a biochemical where substances are bound to the sludge and disposed of in this way.

### 2.2.1 Activated Sludge

Activated sludge is a widespread technique used in wastewater treatment. Using this technique wastewater is combined with microorganisms to build up flocs. The resulting combination of wastewater and biological mass is often referred to as mixed liquor. Using activated sludge offers different advantages, two major ones are:

1. The sludge settles easily, which simplifies the clarification process.
2. Mixed liquor is pumpable, which makes it possible to control biomass concentration in the reactors.

In all activated sludge plants, the sludge has to be separated from the water in the final treatment also known as clarification. The separated biomass is then pumped back in the bioreactors. Due to the fact that the biomass is fed continuously with fresh wastewater, the biomass grows. This makes it necessary to separate a part of the pumped-back sludge, the so called excess sludge, and lead it to the sludge treatment. Besides these advantages different operating problems can occur during the treatment process. Typical problems are bulking sludge and floating sludge. In both cases separation does not work properly. **Bulking Sludge** occurs when certain microorganisms, the so called filamentous bacteria (Lee and Lin, 2007), gain a growth advantage, which leads to slow settling conditions. Sludge is considered bulked sludge if the Sludge Volume Index (SVI)<sup>1</sup>, is greater than  $150\text{ml} / \text{g}$  (Gujer, 2007, p. 327). **Rising Sludge / Floating Sludge** occurs if denitrification appears during the final treatment which produces carbon dioxide gas and is furthermore facilitated by the growth of certain bacteria, such as *Nocardia*. The result is a layer of floating sludge on the clarifier. Bulking sludge and rising sludge problems can both lead to sludge outflow and result in violation of effluent values.

### 2.2.2 Nitrogen Removal

Nitrogen removal is one of the relevant processes in the context of WWTP control. It is composed of two main steps: 1. Nitrification and 2. Denitrification. Nitrification is the process in which ammonium is converted to nitrate under aerobic conditions, whereas denitrification is the process in which nitrate is converted to elemental nitrogen under anoxic conditions. Figure 2-1 shows a simplified representation of the main nitrogen removal process in the bioreactors of a WWTP.

---

<sup>1</sup> Sludge Volume Index: The quotient of the settled volume of sludge and the amount of suspended solids in ml/g. The lower the SVI, the better the settling properties.

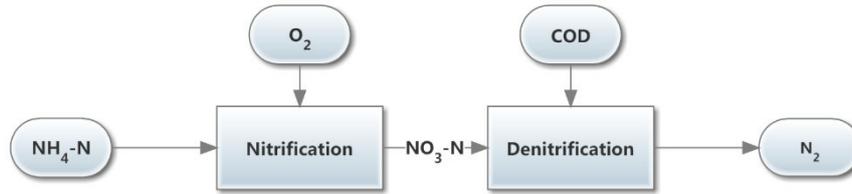


Figure 2-1: Simplified representation of Nitrogen Removal in WWTPs

In order for this to work, urea one of the main substances in urine, is transformed in a previous step by urea hydrolysis into ammonium. This process mainly occurs in the sewer system.

### 2.2.2.1 Nitrification

Nitrification is mediated in two steps: 1. Conversion from ammonium to nitrite. 2. Conversion from nitrite to nitrate. This is done by two special kinds of autotrophic organisms (ammonium oxidizing organisms (ANOs) and nitrite oxidizing organisms (NNOs)). The stoichiometric reaction for the conversion of ammonium to nitrite is shown in equation (2.1):



The conversion from nitrite to nitrate is shown in equation (2.2):



#### Oxygen consumption during Nitrification

Both reactions involve oxygen, which has to be provided by the plants' blowers and which is therefore a cost factor for the plant operator. Therefore, the oxygen consumption in  $mgO_2 / mgN$  for the nitrogen conversion has to be calculated. Considering only the left side of equation (2.1) it can be seen that  $\frac{3}{2}O_2$  are needed for the conversion of one  $NH_4^+$  ion. Considering further that nitrogen ( $N$ ) has an atomic mass of  $14da$  and oxygen ( $O_2$ ) an atomic mass of  $32da$ , the following oxygen consumption can be calculated for the conversion of ammonium to nitrite (2.3):

$$\frac{3}{2} \cdot \frac{32da [O_2]}{14da [N]} = 3.42 \frac{mgO_2}{mgN} \quad (2.3)$$

Dealing in the same manner with equation (2.2), the oxygen consumption for the conversion from nitrite to nitrate can be calculated as:

$$\frac{1}{2} \cdot \frac{32da [O_2]}{14da [N]} = 1.14 \frac{mgO_2}{mgN} \quad (2.4)$$

Combining the output of equations (2.3) and (2.4), the total oxygen consumption per mg nitrogen is:

$$2 \cdot \frac{32da[O_2]}{14da[N]} = 4.57 \frac{mgO_2}{mgN} \quad (2.5)$$

### **Nitrification Control**

Nitrification is significantly more difficult to establish than carbon removal, because autotrophic biomass grows approximately five times slower than heterotrophic biomass. For the same reason carbon is consumed more quickly than ammonium. This is why  $NH_4 - N$  is commonly used as a reference value for oxygen controllers. For domestic wastewater it could be said in simple terms: If nitrogen removal is working sufficiently, carbon removal should not be a problem (On the condition that the wastewater has a common C:N:P ratio (Bever, 2002, p. 4)).

Figure 2-2 shows the basic behaviour of carbon ( $BOD_5$ ) and Nitrogen ( $NH_4 - N$ ,  $NO_3 - N$ ) as it would be in an ideal batch reactor during nitrification. After the oxygen ( $O_2$ ) is turned on, the heterotrophic organisms start to consume carbon, while autotrophic biomass starts to convert ammonium. Because the first process is significantly faster, carbon will be completely consumed after a while and only nitrification continues. On full-scale plants nitrification often starts only after carbon is completely consumed. The reason is that the oxygen is consumed immediately by the heterotrophic organisms and blowers are not able to provide enough oxygen to reach the required level for nitrification. This effect is especially important for intermittently operated plants, because it leads to a strong non-linearity in the system. Until the carbon is consumed, oxygen concentration will not rise and subsequently it rises very rapidly. The time to this 'jump' depends on the carbon concentration in the reactor. Due to the control deviation (in oxygen concentration), a PI or PID controller will react by increasing the integral part, which leads to overshooting, when carbon is consumed. A detailed description of this effect is provided in chapter 3.

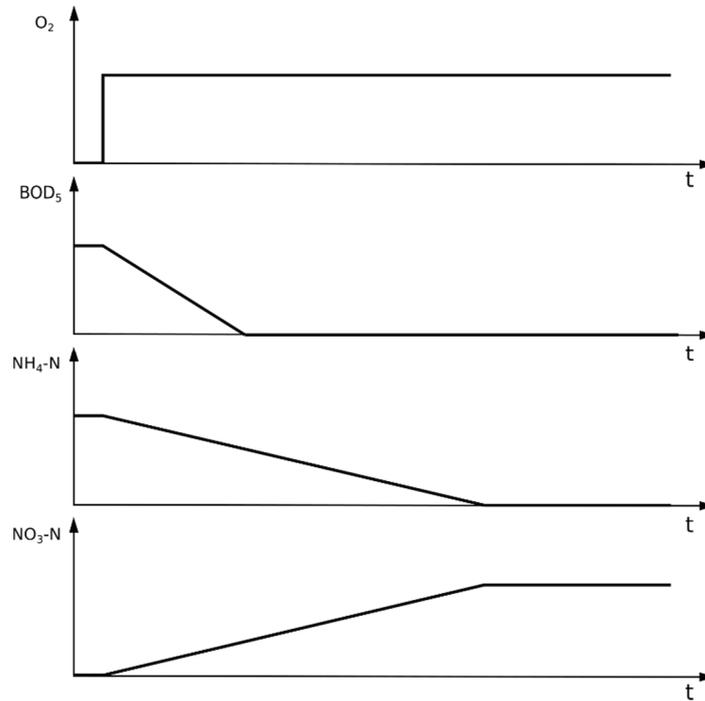


Figure 2-2: Carbon and Nitrogen behaviour during Nitrification in an ideal Batch Reactor

### 2.2.2.2 Denitrification

The denitrification process takes place in an anoxic environment and is the process in which nitrate is converted to nitrogen gas ( $N_2$ ). Denitrification is performed by so-called ordinary heterotrophic organisms (OHOs). The typical stoichiometric reaction, omitting carbon consumption, for nitrate conversion is shown in equation (2.6):



#### Oxygen Recovery by Denitrification

Besides the elimination of nitrogen, denitrification has the benefit that it recovers oxygen. This means that a part of the oxygen used for nitrification can be used in the form of nitrate for the degradation of carbon. The oxygen equivalent of nitrate is  $2.86 \text{ mgO}_2 / \text{mgNO}_3 - N$ , which means that  $1 \text{ mgNO}_3 - N$  has the same electron accepting capacity as  $2 \text{ mgO}_2$ . This oxygen equivalent can be calculated as follows: During nitrification ammonium donates  $8e^- / \text{mol}$  changing  $N$  from an  $e^-$  state of -3 to 5. During denitrification, nitrate accepts  $5e^- / \text{mol}$  changing  $N$  from an  $e^-$  state of 5 to 0. Considering equation (2.5), the oxygen equivalent can be calculated as:

$$\frac{5}{8} \cdot 4.57 \frac{\text{mgO}_2}{\text{mgN}} = 2.86 \frac{\text{mgO}_2}{\text{mgN}} \quad (2.7)$$

which means that in theory 63% of the oxygen can be recovered:

$$\frac{2.86 \text{ mgO}_2}{4.57 \text{ mgO}_2} = 63\% \quad (2.8)$$

In practice a recovery rate of approximately 50% is realistic (Henze et al., 2008, p. 109). This means that even for WWTPs, where nitrogen removal is not necessary, it is useful to enforce denitrification to save energy by using the oxygen recovered during denitrification for carbon degradation.

However, in practice denitrification is desired for nitrogen removal and a problem is actually a lack of carbon. The carbon problem is usually addressed by a different plant design (described in section 2.3.2.3) or by the use of an additional external carbon source.

### 2.2.3 Carbon Removal (Organic Matter Removal)

Historically, carbon was the first substance targeted in wastewater treatment and up to the mid-seventies plants were built only for organic matter removal (Ludzack and Ettinger, 1962). These plants were continuously aerated until the sludge was stabilized. This process needed high amounts of energy, which was less relevant due to significantly lower energy costs (EnergieAgentur.NRW, 2015). For plants applying nitrogen removal, carbon has to be available in sufficient concentrations for denitrification, but has to be removed sufficiently before the clarified water leaves the plant. This complicates process engineering and control of the plant. During the process carbon is used mainly for two processes:

1. Growth of heterotrophic biomass.
2. Endogenous respiration of the sludge.

For the control of the plant two aspects have to be considered:

1. Having sufficient *COD* available during the denitrification process
2. Achieving sufficient *COD* removal to bring the effluent within regulatory limits.

Due to the fact that sources of carbon or organic matter are versatile, a stoichiometric description is not suitable. Instead, it is preferable to describe the carbonaceous oxygen demand using process kinetics. Equation (2.9) calculates the daily flux of oxygen utilised ( $FO_C$ ).

$$FO_C = FS_{bi} \left[ (1 - f_{cv} Y_{Hv}) + (1 - f_H) b_H \frac{Y_{Hv} f_{cv} SRT}{(1 + b_H SRT)} \right] \quad (2.9)$$

(Henze et al., 2008, p. 60)

- $FO_C$  - Daily flux of oxygen utilised [ $mgO_2/d$ ]  
 $FS_{bi}$  - Daily flux of influent biodegradable COD [ $mgCOD/d$ ]  
 $Y_{HV}$  - VSS yield of OHOs [ $mgVSS/mgCOD$ ]  
 $f_{cv}$  - COD to VSS ratio of sludge [ $mgCOD/mgVSS$ ]  
 $f_H$  - Unbiodegradable fraction of OHOs [ $mgCOD/mgCOD$ ]  
 $b_H$  - Specific rate of endogenous mass loss of OHOs [ $d^{-1}$ ]  
 $SRT$  - Sludge retention time - sludge age [ $d$ ]

A closer look at equation (2.9) shows that it consists of the sum of two terms. The first part  $FS_{bi}(1 - f_{cv}Y_{HV})$  represents the oxygen demand for the growth of the heterotrophic biomass (OHOs) and is independent of the sludge age ( $SRT$ ). The second term  $FS_{bi}(1 - f_H)b_H \frac{Y_{HV}f_{cv}SRT}{(1 + b_H SRT)}$  represents the endogenous respiration of the biomass. From the fact that  $SRT$  is included in the second term, it can be seen, that the endogenous respiration depends on the sludge age.

#### 2.2.4 Phosphate Removal

In the past, phosphate was mainly removed using chemical processes. This involves adding iron or aluminium salts such as  $FeSO_4$  or  $AlCl_3$  to the wastewater. The dissolved phosphate compounds are converted as a result into undissolved phosphates, which can then be removed from the plants with the sludge. The most commonly used process is simultaneous precipitation, where the precipitant is dosed into the aeration tanks. Alternatives are pre-precipitation, in which the precipitant is dosed into the inflow for the purpose of preliminary treatment, or post-precipitation, in which an additional reaction tank is required (Mudrack et al., 1991, pp. 265–266). Today, biological phosphate removal is also used so that phosphate elimination usually occurs through a combination of chemical and biological precipitation processes (Stier and Bundesinstitut für Berufsbildung, 2003, pp. 173–174). Although the elimination of phosphates is not explicitly part of this work, it is necessary to briefly discuss the basic processes involved as it is relevant due to the close relationships between the individual processes. In anaerobic conditions, it is possible for the phosphate to redissolve back into the wastewater so that the phosphate bound in the biomass is released again. Therefore, it is important to also take this into account during the development of the oxygen controller because otherwise it cannot be guaranteed that the amount of phosphate will not rise to exceed the imposed limits.

##### 2.2.4.1 Biological Phosphate Removal and Enhanced Biological Phosphate Removal (EBPR)

In the biological elimination of phosphates, the phosphate is stored within microorganisms and removed from the plant via the excess sludge. This is where there is the already mentioned danger of phosphate being redissolved back into the wastewater because the microorganisms can release the phosphate again. In general, microorganisms require phosphate to build their cell structure. It is known that phosphate

uptake can be increased under stress conditions (change between aerobic and anaerobic conditions). For this reason, an additional upstream anaerobic tank is often utilised during biological phosphate removal in which return sludge is brought together with the fresh inflowing wastewater. In the case of intermittently operated plants with a low load, it is possible to biologically remove the phosphate without an additional tank using the oxygen control to achieve similar conditions as with an additional tank. Typical phosphate consumption during this process is around  $0.02 \text{ mgP/mgVSS}$ . This corresponds to phosphate removal of up to 25% for normal wastewater. Enhanced Biological Phosphate Removal (EBPR) is based on specifically promoting certain microorganisms – so-called Phosphorus Accumulating Organisms (PAOs). These organisms store phosphate in the form of polyphosphates and can thus achieve phosphate consumption of up to  $0.15 \text{ mgP/mgVSS}$  (Henze et al., 2008).

In terms of the regulation of oxygen, this means avoiding forced anaerobic conditions in tanks in those plants where Bio-P occurs and there is no aerated tank downstream. Otherwise, it is difficult to avoid the increased use of chemical precipitation.

## 2.3 Wastewater Treatment Infrastructure

### 2.3.1 Sewer System

The purpose of the sewer system is the fast transport of wastewater from a producer to a treatment facility. Usually it is an underground network of pipes without significant instrumentation (Koch, 2002). Two main kinds are common:

#### 2.3.1.1 Sanitary Sewers

This kind of sewer system is dedicated only to wastewater. Rainwater is collected separately in a runoff drainage system. These systems have the advantage that the connected treatment facility only has to deal with pure wastewater, where fluctuations in the inflow are only generated by wastewater producers. Furthermore, the inflow usually follows a diurnal cycle, which is easily predictable and can be used for control optimization purposes. However, in practice it is hard to avoid infiltration water (2.1.1), especially if the system is older. For this reason dilution effects have to be expected, when groundwater levels rise. Generally, these systems have the disadvantage that a second system for runoff water has to be installed. Therefore, the costs are significantly higher than for combined sewers.

#### 2.3.1.2 Combined Sewers

Combined sewers transport wastewater and runoff water. On the one hand, the system is much cheaper compared to sanitary sewers, because only one underground system has to be installed. On the other hand, it has several disadvantages:

1. The system has to be designed to be able to carry the additional runoff water.
2. In case of rain events, peak flows appear.

3. The connected treatment plant has to be designed in a way that it is able to treat higher inflows.
4. Additional storm water tanks have to be integrated into the system to buffer wastewater in case of strong rain events.

### **2.3.1.3 Storm Water Tanks (SWT)**

SWTs are hydraulic capacities (typically concrete tanks) integrated in combined sewer systems to temporarily store high amounts of water in cases of rain events, to avoid an overflowing of the sewer network or overloading of the connected WWTP. The control principle is simple: A flow meter (typically a magnetic flow meter - MFM) and a controllable throttle valve is integrated in the sewer system. An additional controller, typically a PLC, controls the valve so that only the desired amount of water can pass. Surplus water is led into the SWT. After a rain event the water in the SWT is pumped back into the sewer. SWTs are typically sized such that only in the case of very strong rain events will their capacity be exceeded. If this happens the water is usually discharged into the adjacent river. Under these extreme conditions the wastewater is already diluted to the point where it is no longer harmful to the receiving water body. Furthermore, sedimentation occurs in SWTs, so that coarse dirt and particulate matter are held back part wise. For better control, SWT level as well as the discharged water is measured. Using the measured data and throttle valves SWTs offer several opportunities for optimization:

1. Early measurement of the amount of water in the sewer system, which can be used for feed-forward plant control (early aeration, if peak loads are expected).
2. Controlling the inflow to the plant.

A more complex application is the better distribution of the water to the different SWTs in a sewer system during rain events. Various research projects have dealt with the optimization of sewer systems using SWT e.g.(Nickolaus et al., 2005; Trauer, 2010; Wolf, Merkel, 2015). A detailed description of the KANNST research project, which deals with these aspects and additionally considers the COD and NH<sub>4</sub>-N loads can be found in the doctoral dissertation “Water in Society” (Hilmer, 2008).

### **2.3.2 Wastewater Treatment Plants**

Treatment plants are the main components of wastewater treatment systems. This is where the actual purification process (see subsection 2.2) is carried out in individual stages. These plants deploy mechanical, biological and chemical purification processes. Following a variety of mechanical purification steps, the biological purification of carbon, nitrogen and phosphorus is then carried out in the main reactor. This is followed by the final treatment in which the activated sludge is mechanically separated from the wastewater through sedimentation. A final purification stage is optional and is only utilised in cases where there are particularly stringent requirements placed on the quality of the effluent. Figure 2-3 shows the basic structure of a typical treatment plant. The following subsections describe the relevant stages and the possibilities for intervention in each case.

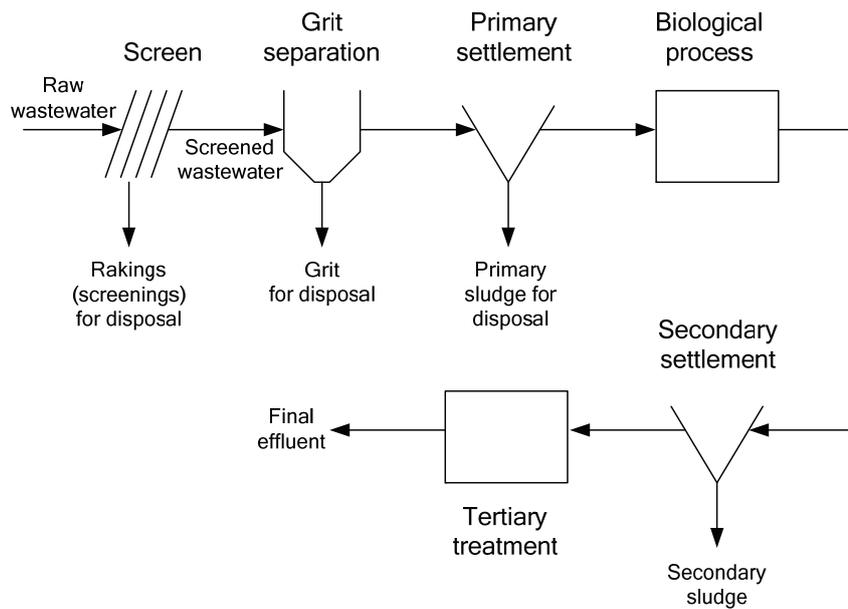


Figure 2-3: Layout of a typical wastewater treatment plant (Ebel, 2009)

### 2.3.2.1 Inflow Gate / Screen and Grit separation

The inflow grate performs the task of keeping coarse dirt out of the treatment plant in order to, on the one hand, avoid mechanical damage and, on the other hand, reduce the solid content. The grates can be subdivided into coarse and fine grates. Coarse grates have openings between 10 mm and 50 mm, while the openings on fine grates are between 2 mm and 10 mm. Typically, automatic grates are used today e.g. step grates. These systems detect how blocked the grates have become, automatically clean them and transfer accumulated waste directly into a container. (Teichmann, 1997)

The traditional method for recognising the extent to which the grates have become blocked is to measure the water level in front of the grates. If the water level exceeds a certain level, the automatic cleaning process is initiated. Improved systems measure the water level before and after the grate and use the difference between them as the defining criterion. A new approach is the use of image recognition methods to assess the extent to which the grate has become blocked. Tests conducted by the GECO►C Research Group have demonstrated that this is possible in principle but remains difficult because the majority of the dirt collects below the water level (Trauer et al., 2011).

For the control of treatment plants, it is conceivable that water level sensors could be used at the grates to detect hydraulic impact at an early stage.

The grate is usually followed by a grit chamber. This has the job of trapping other solids that are being carried by the water.

### *2.3.2.2 Primary Treatment / Primary Settlement*

The primary treatment tank or clarifier is part of the mechanical purification process. The goal is to free the wastewater of any suspended solids that could not be removed in the previous stages. It is constructed so that the wastewater flows slowly through the tank. In rectangular tanks, scrapers at the bottom and also usually at the surface of the tank are used to remove the deposited or floating material. The discharged material, which generally has a high carbon content, is concentrated and fed into the digestion tanks at the treatment plant as so-called raw sludge.

In activated sludge plants, primary treatment tanks are operated with a clarifier detention time (CDT) of between 0.5 h and 1 h or a surface loading rate (SLR) of between 4 m/h and 2.5 m/h. It can be assumed that practically all of the material that can settle out will have been separated after approximately two hours (Stier and Bundesinstitut für Berufsbildung, 2003, p. 107).

From a control engineering perspective, two aspects of the primary treatment tank are interesting.

1. They have a hydraulic capacity, or dead time, which control systems can take advantage of when measurements are taken during the inflow of wastewater.
2. Large amounts of carbon are removed during the primary treatment.

While the goal of older plants was to allow as much material to settle out as possible, the clarifier detention time is today kept shorter to retain some carbon for the denitrification process (Stier and Bundesinstitut für Berufsbildung, 2003, p. 108). It is possible in some plants for part of the wastewater to bypass the primary treatment and feed into the bioreactors so that additional carbon is present for denitrification. If this option is available, it must also be integrated into the control system for nitrogen removal.

### *2.3.2.3 Secondary Treatment*

Secondary treatment is the biochemical purification stage. There are six types of reactor layouts or configurations that are typically used in activated sludge plants. Every reactor layout has specific characteristics, as well as advantages and disadvantages.

In practice, complexity of the plants often increases because many of them, especially the medium and large sized plants, have multiple lanes. The main types, their relevant characteristics and their specific intervention and control possibilities will be described below. A detailed description of the control loops, as well as the general control options, will follow in Chapter 3 (Wastewater treatment plant control).

### Downstream Denitrification

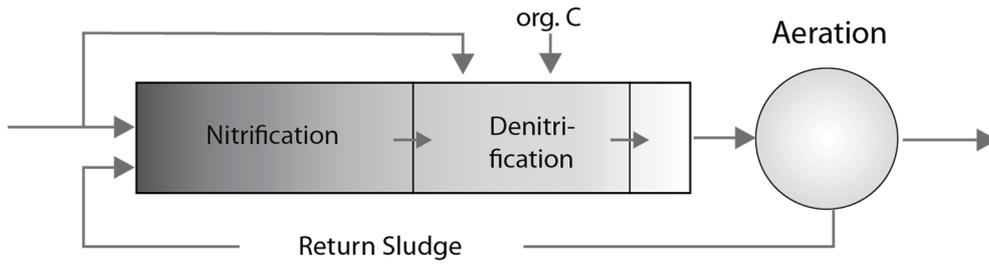


Figure 2-4: Downstream Denitrification

Downstream denitrification (Figure 2-4) is actually a chemical degradation process in which the wastewater is firstly subjected to a nitrification and then a denitrification process. The benefit here is that almost all of the ammonium ( $NH_4 - N$ ) is converted to nitrate ( $NO_3 - N$ ) during the nitrification process. The nitrate is then largely degraded during the denitrification process. In this manner, it is possible to ensure that there is hardly any nitrogen present in the effluent from the plant. The disadvantage is that the carbon is largely consumed by the heterotrophic biomass during nitrification and is thus no longer available for the denitrification process. This is doubly disadvantageous:

1. Carbon is lacking for the subsequent denitrification and must be added externally.
2. Dissolved oxygen is also consumed during carbon degradation, which could otherwise be recovered in part from the nitrate during a functioning denitrification process (see subsection 2.2.2.2).

In this type of plant, the main control options are in the areas of oxygen concentration and carbon dosing.

### Upstream Denitrification

Upstream denitrification (Figure 2-5) is one of the most widely used types due to the benefits it offers.

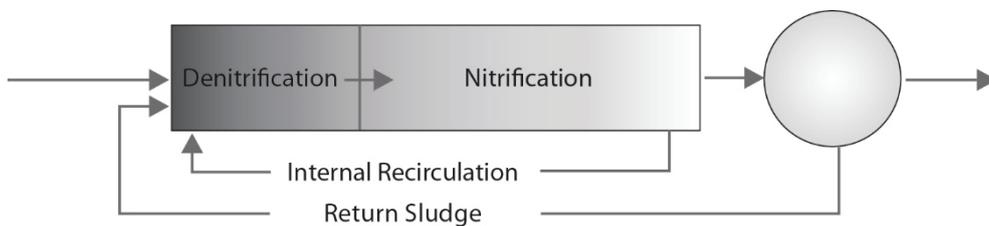


Figure 2-5: Upstream Denitrification

In upstream denitrification, the process steps of nitrification and denitrification are interchanged. This means that the nitrate that is created from ammonium in the second nitrification tank needs to be pumped back to the start of the plant with the activated sludge (recirculation  $Q_R$ ). It is then combined with fresh carbonaceous wastewater in the denitrification tank. The advantage of this arrangement is that no supplementary source of carbon is necessary. The disadvantage is that it is impossible to avoid some of

the nitrate produced during nitrification remaining in the water and entering the final treatment stage. In addition, energy is required for the internal recirculation process.

In this arrangement of tanks, the ammonium concentration ( $NH_4 - N$ ) can be influenced by the oxygen concentration and the nitrate concentration ( $NO_3 - N$ ) by the recirculation volume  $Q_R$ .

### Cascaded Denitrification

Cascaded denitrification is similar to upstream denitrification, although it does not involve a recirculation process. It begins with a denitrification tank into which return sludge ( $Q_{RS}$ ) and fresh wastewater are fed and is followed by a nitrification tank. This is followed by a further denitrification tank and so on (Figure 2-6). Part of the inflow is fed into each of the denitrification tanks in order to provide them with fresh carbon. In principle, this system consists of a cascade of multiple upstream denitrification plants.

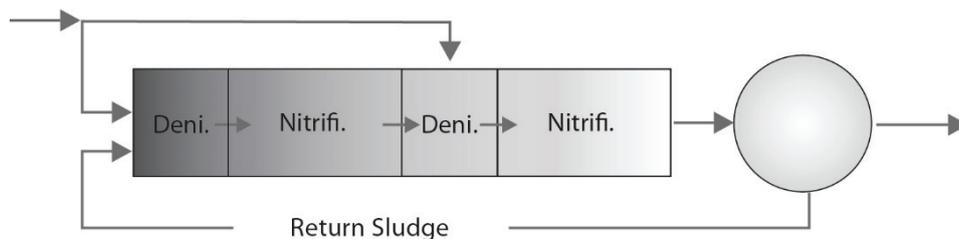


Figure 2-6: Cascaded Denitrification

The advantages of this arrangement are that neither a recirculation process nor any externally added carbon are required. In addition, biological phosphate removal is promoted as a consequence of the anaerobic, aerobic and anoxic zones. The possibilities for intervention here relate to the distribution of the inflowing wastewater and the oxygen concentrations in the individual aerobic tanks. In practice, it is also not unusual for individual zones to be aerated only intermittently or as required.

### Simultaneous Denitrification (SD)

In the case of simultaneous denitrification, the tanks are split into individual aerobic and anoxic zones in which the different processes take place (Figure 2-7).

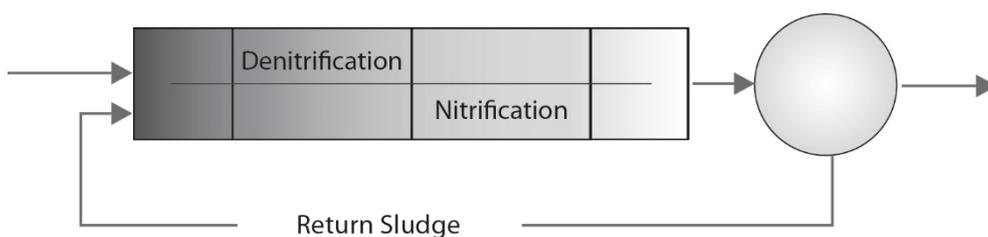


Figure 2-7: Simultaneous Denitrification

As there is no separation between the individual zones, relatively large tank volumes are required. The advantages are that the tanks have a simple design and no recirculation process is required. This type of system is suitable for smaller plants. The control options are limited here to the oxygen concentration in the individual zones of the tank.

### Intermittent Denitrification (ID)

In intermittent denitrification, both processes run successively in the same tank (Figure 2-8). This type of plant is still very widespread not least because it was very easy to convert continuously aerated plants that were originally used for the removal of carbon for the purposes of nitrogen degradation.

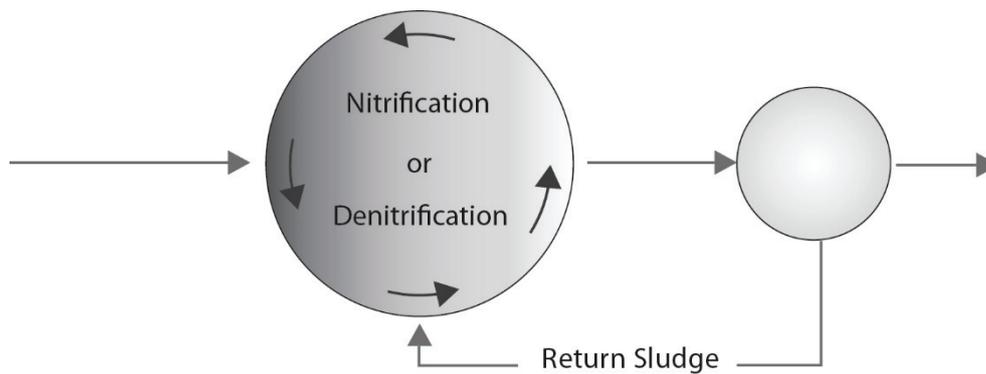


Figure 2-8: Intermittent Denitrification (ID)

From the perspective of control technology, this type of plant is one of the most interesting because, alongside the oxygen concentration during the nitrification phase, it is also possible to control the different phases themselves. For example, it is possible during peak loads to quickly switch to the nitrification phase.

### Alternating Denitrification

In alternating denitrification, the water is fed alternatively into two tanks. In addition, the two tanks are also (intermittently) aerated on a rotating basis (Figure 2-9). Here, the wastewater being treated flows out of the tank into which it was originally fed and into the other tank and from there to the final treatment tanks.

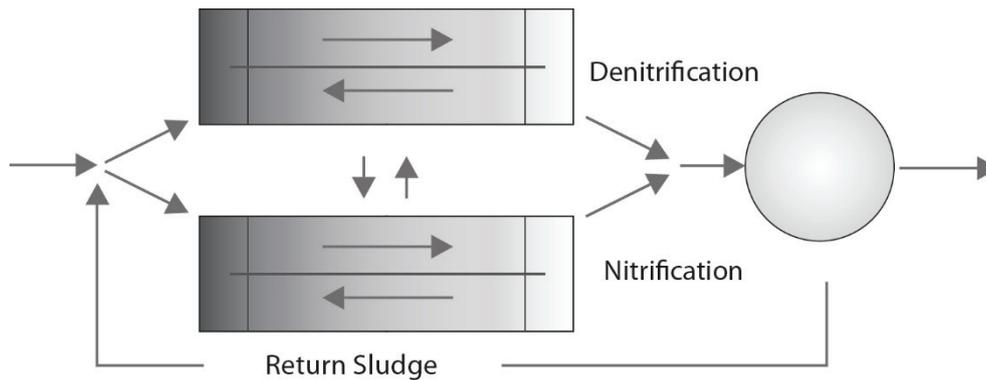


Figure 2-9: Alternating Denitrification (AD)

The main advantage for operators is that two tanks can be operated with one blower and investment costs are saved as a result.

#### 2.3.2.4 Final Treatment

The role of the final treatment tank is to separate the water and sludge, as well as to store the activated sludge or biomass when it is raining. Typical versions of final treatment tanks are hopper bottomed tanks or round tanks with scrapers. The separation process is based on the sedimentation of the sludge. It is for this reason that the most important variable for the separation efficiency of the final treatment tank is the sludge volume index  $SVI[mg/l]$ , which describes the ratio of the sludge volume to the dry matter content  $TSS$  and the surface area  $A_{FT}$  of the final treatment tank. Here, it is assumed that smaller compacter flakes settle better than larger lighter flakes. The deposited sludge is then fed back into the bioreactors as return sludge ( $Q_{RS}$ ). As the biomass grows constantly, part of the sedimentary sludge is separated and fed into the sludge treatment process as excess sludge ( $Q_{ES}$ ).

From a control perspective, the final treatment tank is a storage system for biomass. The TSS content in the aeration tanks can be controlled using the volume of return sludge. It is thus conceivable, for example, that the volume of return sludge can be increased during peak loads, thus increasing the amount of active biomass in the reactors.

#### 2.3.3 Sludge Treatment

Sludge treatment is another major aspect in the area of wastewater treatment. During wastewater treatment sludge is generated at different process steps. Primary sludge is mechanically separated from the wastewater during the primary treatment and is characterized by high amounts of  $COD$  due to the high fat and grease content. The second source is excess sludge. Because the biomass is fed constantly it grows continuously. Thus, for an operator the sludge is a waste disposal problem. It has to be stabilized, which means that the organic content has to be reduced until the biological activity is reduced

to a minimum. This has to be done to avoid unpleasant odour and to make the sludge storable. After this process the sludge is dewatered and burned or used as fertilizer.

Different methods are used for sludge stabilization:

- Aerobic sludge stabilization

This method is often used on small-scale WWTPs, which do not have a digester because of financial reasons. Basically, the sludge is aerated so that the microorganisms are in a constant hunger condition, where they are forced to use all available resources to avoid starvation. This method has the major disadvantage that a lot of energy is needed for the aeration. Furthermore it is only possible in the same tank as the nitrification, if the plant is only lightly loaded.

- Anaerobic sludge digestion

The Anaerobic Digestion (AD) of sludge serves two major purposes: (1) Carbon and nitrate removal as well as (2) energy production. In order to achieve these goals, sludge is treated in digesters under an anaerobic environment to produce biogas which mainly consists of methane ( $CH_4$ ), carbon dioxide ( $CO_2$ ), hydrogen ( $H_2$ ) and hydrogen sulphide ( $H_2S$ ). The biogas is then burned in cogeneration units to produce electrical and thermal energy or upgraded to natural gas quality, which is in turn supplied to the gas grid. The AD process in WWTP digesters is normally operated at constant temperature in the mesophilic range between 35 and 42°C and at a TS concentration between 5 and 10%. Although, there exist AD designs with multiple digesters to increase process efficiency, the single reactor design is most common on WWTP due to lower costs. The state of online instrumentation and control of AD plants is still in its infancy, which is why only temperature, biogas yield and composition are commonly measured on most AD plants for sludge treatment. The same applies to the use of control systems. As process conditions are very stable, due to the fact that sludge composition is mostly stable as well, online process control is often considered to be not necessary.

Generally, it can be said, that sludge treatment is one of the major costs of wastewater treatment (Friedrich et al., 2006). This work does not focus particularly on sludge treatment, but never the less during plant optimization or controller development, sludge treatment has to be considered.

## 2.4 Summary

This overview of the complete wastewater system from wastewater collection to final treatment in WWTP illustrates its high complexity, which is the main reason why wastewater treatment optimization and control is challenging requiring novel but also sufficiently robust solutions. In particular, the multiple biological treatment processes of nitrification, denitrification as well as carbon and phosphate

removal that take place simultaneously and/or successively all require different specific optimal process conditions and maybe even specifically adapted plant designs as shown in section 2.3.2.3. In addition to that the major challenges in wastewater treatment in terms of degree of water purification and energy consumption are introduced and briefly described. Constantly rising demands for higher treatment standards as wells as rising energy costs create a high demand for practice-oriented optimization and control solutions.

Thus, this introduction to wastewater collection and treatment and the respective processes is an important basis for WWTP simulation using the Activated Sludge Model No. 1 (Henze et al., 1999) introducing the relevant biological process variables and their influence at a higher level. Furthermore, the development of process control strategies has to consider general plant designs and process conditions as well in order to meet the objectives.

### 3 Wastewater treatment plant control

This chapter gives an overview of WWTP control as well as the related instrumentation and automation. The aim of this chapter is to provide a summary from the point of view of an automation engineer as a basis for the following chapters. The main focus is on presenting typical control loops and the associated design and operation challenges.

From several years of experience in the field of controller development for WWTPs with the GECO►C research group, we learned that the development of the “perfect controller” from a control theory viewpoint is usually not the optimal goal but instead the development of a control system which is operationally robust and tailored to the requirements of a specific plant. A similar opinion was expressed by Olsson et al. in 2014 who observed that: *“The control of wastewater treatment systems is certainly not limited by the available control theory. Rather, the challenge is to have a comprehensive understanding of the process and its limitations, the control authority of the actuators, the reliability of and information from the sensors and also data management and monitoring strategies.”* (Olsson et al., 2014). Apart from controllers, as the main components, the complete control loop consists of many additional parts. Sensors are used to measure different control variables, further hardware is required to implement the controllers (nowadays typically a PLC or an industrial PC) and the actuators (blowers, valves, pumps, etc.), which have to be controlled. In particular in the area of online-sensors, major developments have come to market in recent years, which offer new opportunities especially for smaller WWTPs.

This chapter describes briefly the measurement and machine technology in use today, as well as their development in recent years and the associated changes in the control and optimization of wastewater treatment plants. Then typical control loops, as well as their procedural specificities from the perspective of an automation engineer, are described. Finally, the ammonium ion-selective measurement, and the spectrometric measurements of Chemical Oxygen Demand, two particularly important developments in the context of this work, are described in detail.

#### 3.1 WWTP Instrumentation and Actuators

While many plants were operated with only a few or even no online sensors just a few years ago, they play a much more important role in the operation of plants today. The first instrumentation and control applications (ICA) were introduced around 1970 (Olsson et al., 2014). The proportion of investment costs accounted for by instrumentation in treatment plants was 10%-15% in 2005 and the forecasts then predicted it would reach 30% by 2010 (Olsson et al., 2005, p. 36). This development clearly demonstrates the increasing importance of ICA.

### 3.1.1 Actuators

Actuators are generally transducers or motor elements that convert the control signals from controllers into mechanical movements or other physical variables. In treatment plants, these actuators are mainly pumps, compressors or valves. Just as in the area of instrumentation, there have been far-reaching developments in the area of actuators in the last few years. In particular, power electronics can be sourced for significantly lower prices than was the case a few years ago. This makes it possible to now employ frequency converters in significantly greater numbers and above all in smaller plants. Continuous controllers are thus becoming more and more important in comparison to simple 2-point and 3-point controllers.

However, the existing machine technology is the limiting factor in many areas of application. Typical examples of these limitations are pumps or compressors without frequency converters or equipment that can only be adjusted to a few different performance levels. Figure 3-1 shows an example of an oxygen control system. In this case, the desired oxygen setpoint was 1.8 mg/l. However, the blowers only had three different setting levels and could only be changed a maximum of every 5 minutes.

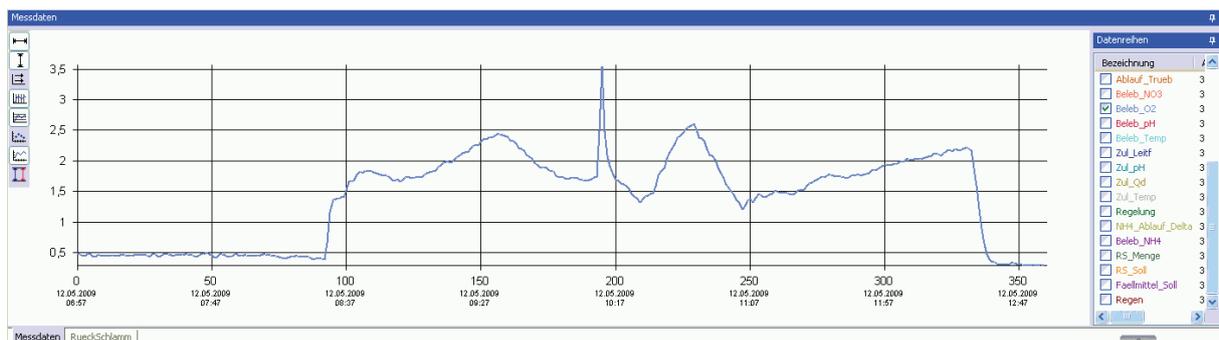


Figure 3-1: O2-Measurement WWTP Heizenberg

Many compressors can also only be operated within certain limits (e.g. between 30 Hz and 55 Hz). During periods of very low load, it is often particularly desirable to be able to operate the compressors at even lower outputs or lower frequencies. In many cases, the required controllability for maintaining good operational control is thus inadequate.

At the same time electric motors have become significantly more efficient in the last few years. This can be seen in the energy efficiency classes introduced by the International Electrotechnical Commission (IEC) in 2008 and defined in the IEC 60034-30 (Deutsche Elektrotechnische Kommission, 2009) standard. Many manufacturers already offer extremely energy-saving premium efficiency motors with efficiency class IE3. However, some optimisations consequently lead to problems. One example is wastewater pumps that can be operated very efficiently at a range of speeds. Lower flow velocities can tend to promote blockages to the impeller when it is operated at low speeds.

These diverse framework conditions need to be taken into account when optimizing plant control systems. In addition, the typical response times of the control loop, which often reacts relatively slowly due to its compensatory times, must also be taken into account. Table 3-1 shows typical actuators and example response times, as well as standard types of control signals. It is often possible to influence a great number of process variables with relatively few actuators and these will be presented in the following section.

**Table 3-1: Typical technical properties of actuators**

Actor	Signaltype	Response Time – (including Process until measurable)
<b>Compressors</b>	continuous	2-10 min
<b>Wastewater / Sludge Pumps</b>	continuous	<1min
<b>Dosage Pumps</b>	discrete / PWM / continuous	>10 min
<b>Valves</b>	discrete / continuous	-
<b>Gate Slider</b>	Discrete / continuous	-

### 3.1.2 Process Variables / Measurement Values

It is possible to manage a great number of process variables in a treatment plant with the aid of modern SCADA systems. These variables can be roughly divided into three categories:

1. Archive variables – variables that are gathered but only recorded.
2. Monitoring variables – variables that are used for monitoring processes.
3. Control variables – variables that are actively used for controlling processes.

The largest proportion of these variables belong to category 1. Typical examples are e.g. runtimes and frequencies of pumps and motors or status information on measurement devices and assemblies. In addition, legislators require that certain variables are recorded and archived, such as the *pH value* and the conductivity of the inflow into treatment plants.

Variables in the second category are typically measurement values that become important when the values deviate strongly from the expected values. Although they are relevant for monitoring the processes, they are not required for controlling the plant during normal operation. Examples of this type of variable are flow measurements, ammonium and nitrate measurements in the primary treatment or the final treatment and turbidity measurements at the effluent of a treatment plant. These variables are usually visualised in SCADA systems as trends. They are used by operators to evaluate the process and to determine when a manual intervention may be necessary. Process variables in the third category (control variables) are actively integrated into the control loop. Typical examples include oxygen,

ammonium and nitrate measurements in the aeration tanks. They only account for a small proportion of the total number of variables recorded. In order to ensure the memory requirements do not become too large, data is often aggregated after certain time intervals.

A comprehensive overview of typical measurement variables, although without taking into account nutrient parameters, and information on the optimal arrangement of measurements can be found in the DWA M256 information sheet (DWA, 2011). The various measurement values are subdivided according to their measurement location into the categories A: essential, B: expedient and C: not expedient (see subsection 5.1).

In general, it is desirable to record the highest possible number of relevant variables online. However, only some of them can be measured online with an appropriate amount of effort and cost. Measurements can be distinguished as physical, chemical, biological and, in special cases, discrete process variables. Discrete process variables are e.g. states such as aeration on/off or performance levels. For control purposes, it is also relevant whether the variable can be determined online or only in a laboratory. Table 3-2 contains a brief overview of those process variables relevant to control and monitoring together with their characteristics:

**Table 3-2: Typical WWTP process variables and their relevance for control applications**

Name	Unit	Measurement principle	Location	Application / Control Loop	Relevance
<b>Flow</b>	$m^3 / d$	magnetic inductive	Inflow / Outflow / internal Flows	Return Sludge, Aeration, Recirculation	high
<b>Temperature</b>	°C	resistant	Inflow / Outflow / Biology	Monitoring / Advanced Control: Aeration	medium
<b>Conductivity</b>	$\mu S / cm$	resistant	Inflow	Monitoring	low
<b>pH-Value</b>	-	ion-selective	Inflow	Monitoring	low
<b>Oxygen</b>	$mg / l$		Biology	Aeration	high
<b><math>NH_4-N</math></b>	$mg / l$	ion-selective	Biology	Aeration	high
<b><math>NO_3-N</math></b>	$mg / l$	optical / ion-selective	Biology	Recirculation, Aeration,	high
<b>COD</b>	$mg / l$	chemical / spectrometric	Inflow / Outflow / Biology	Aeration	medium
<b><math>PO_4-P</math></b>	$mg / l$	chemical	Biology / Outflow	Chemical Precipitation / Aeration	high

### Typical Problems / Challenges

Depending on their type and manufacturer, sensors require regular maintenance. Despite careful maintenance, there are a series of recurring problems that typically occur:

- Drift
- Offsets
- Breakdowns
- Lower sensitivity against concentration changes

These problems are also made worse in certain measurement locations e.g. at the inflow into the plant due to heavy contamination, biofouling, high deviations in the temperature and concentrations and unfavourable flow conditions (Graner et al., 2005), whereby they can have very different service lives depending on the plant and installation location. Figure 3-2 a-c shows typical contamination experienced by different ion-selective  $NH_4 - N$  sensors. The sensor in image a) was installed to take measurements directly in the sewerage system operated by the Aggerverband, while images b) and c) show the impact on sensors installed in the inflow into the industrial treatment plant in Emmerich. The sensor in image a) still functioned perfectly, while the membranes in the sensors in images b) and c) were destroyed.

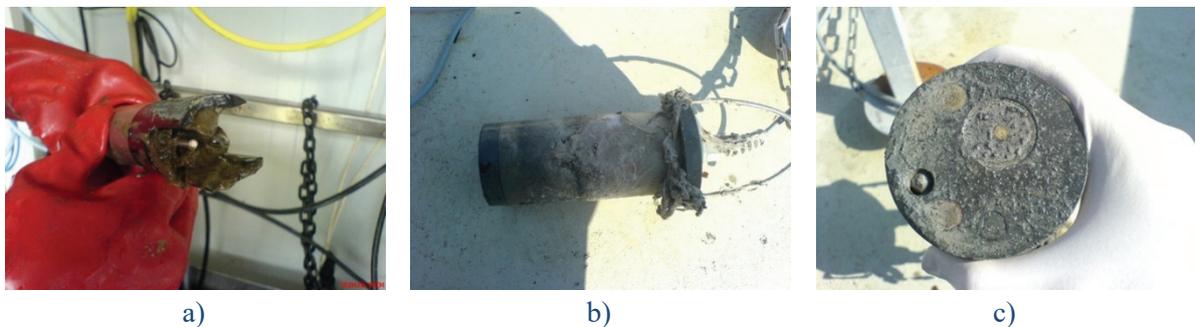


Figure 3-2: Typical pollutions of probes

Figure 3-3 shows measurement values from a measurement device test with two ion-selective measurement devices (Hydrion, WTW), as well as a chemical analyser (Danfoss). It is possible to clearly recognise some of the described effects. Hydrion demonstrates a typical drift towards the end of the measurement. In a direct comparison between the Danfoss measurement and the WTW measurement, it can also be seen that the Danfoss device displays considerably less dynamism. It is also evident that the Danfoss device experienced a breakdown on 26<sup>th</sup> december and showed a measurement value of  $0\text{ mg/l}$ . The Danfoss membrane was cleaned on the 27<sup>th</sup> December, which led to a discernible jump in the measurement value.

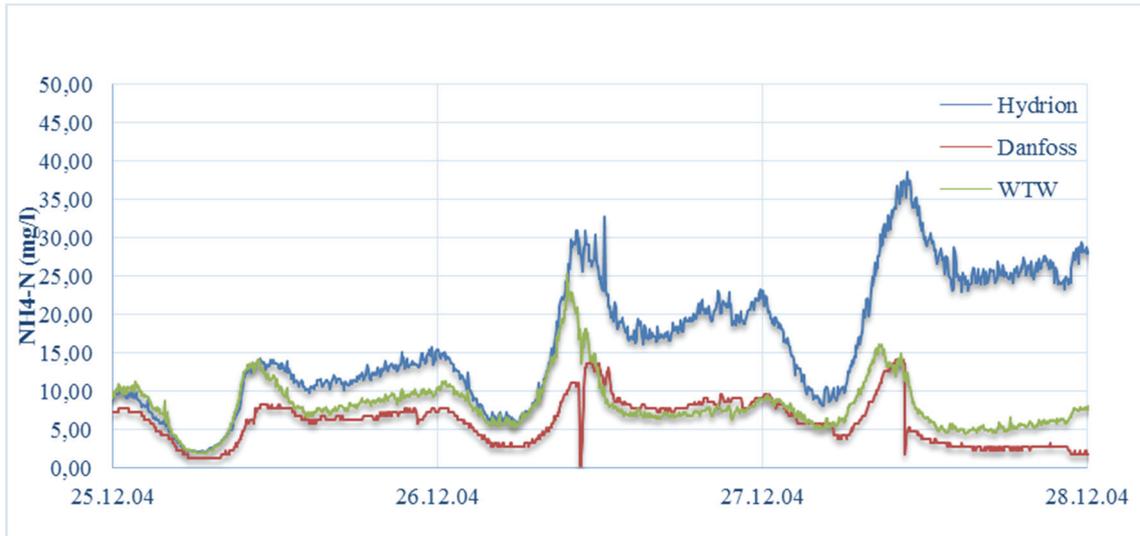


Figure 3-3: Measurement Test of different  $NH_4-N$  online measurement devices at the Rospe WWTP (2004)

The described effects invariably lead to incorrect measurements and are difficult to avoid in practice. Therefore, a proposed system must always be designed to guarantee the safe operation of the plant despite these errors. One possibility is described in section 3.2.1.

### 3.1.3 Measurement Principles

This section gives an introduction to the measurement principles of the most important probes used for this thesis, namely the UV/VIS spectrometry and ion-selective measurements as they represent state-of-the-art technologies. Technologies for other measurements such as *flow*, *pH values*, and *conductivity* are well established and therefore not reviewed here (Koppe and Stozek, 1999). The technical principle of these relatively complex measurement systems, compared to standard *conductivity* or *temperature* gauges, has to be kept in mind, when further analysis of measurement data is presented in chapter 5. The level of precision of these measurements is an important technical limitation for the prediction quality of state variables.

#### 3.1.3.1 Spectrometric measurement probes (UV/VIS spectrometry)

The sensor of a spectrometric (UV/VIS) probe consists of four main components:

- A light source (xenon or deuterium lamp)
- An optical diffraction grating
- A detector
- A computation unit (typically an embedded linux or windows system)

UV/VIS spectrometry uses ultraviolet and visible light (STIP-scan 200 nm-680 nm / Trios ProPS-WW 190 nm-360 nm) to determine certain substances in a liquid sample. During the measurement a special xenon or deuterium lamp shines through the sample. After the light

passes the sample, the light is refracted by a prism or in most cases by a diffraction grating and split into different wavelengths. Figure 3-4 shows the principle of UV/VIS spectrometry:

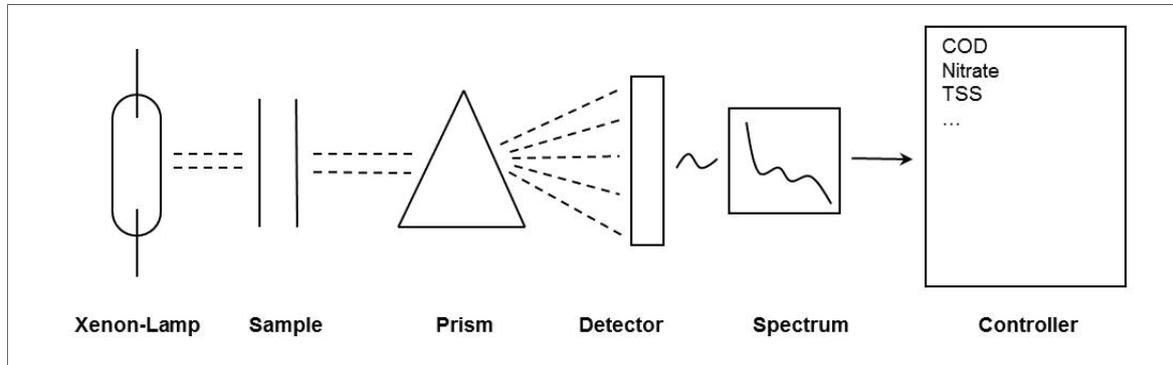


Figure 3-4: Principle of UV/VIS spectrometry

UV/VIS spectrometry is based on the fact that many compounds have unique absorption patterns. Based on the Beer-Lambert law (3.1) it is possible to relate an absorption pattern to specific concentrations of compounds in a liquid sample.

$$A(\lambda) = -\log_{10} \left( \frac{I}{I_0} \right) = \varepsilon(\lambda) \cdot c \cdot l \quad (3.1)$$

Where  $A$  is the absorption,  $I_0$  the intensity of the incident light,  $I$  the intensity of transmitted light,  $\varepsilon$  the extinction coefficient,  $c$  the concentration and  $l$  the length the light travels through the sample. Due to the fact that these are spectrometric data, the values depend on the wavelength  $\lambda$ . In contrast to measurement systems which only use the absorption of one wavelength, the spectrometric measurement systems use several or all measured wavelengths. To calculate the concentration of one compound, an absorption coefficient  $a_i$  for each wavelength  $\lambda_i$  is determined. The concentration is calculated by a linear combination of  $a$  and  $A$  (3.2).

$$c = \sum_{i=1}^n (a_i A_i) \quad (3.2)$$

To determine the absorption coefficients  $a_i$  for  $i = 1 \dots n$  mathematical methods like multiple linear regression analyses are used. A typical machine learning method often used to deduce chemical parameters from UV/VIS measurements is Partial Least Squares Regression (PLS) (Langergraber et al., 2003). Typical detectors used in UV/VIS spectrometers, such as the Zeiss MMS (Carl Zeiss Microscopy GmbH, 2013) are able to detect 256 different wavelengths. The determination of all 256 coefficients is complex due to the high number of input / output value pairs needed to properly capture the underlying process information. To reduce the number of coefficients, some companies use a subset of wavelengths and laboratory measurements for the calculation.

Figure 3-5 shows typical absorption areas in the UV/VIS spectrum. For wastewater the area from 200 nm to 250 nm is used to measure nitrogen, from 250 nm to 380 nm to measure carbon and above 380 nm (the area of visible light) to measure turbidity, totally suspended solids (TSS) and colour.

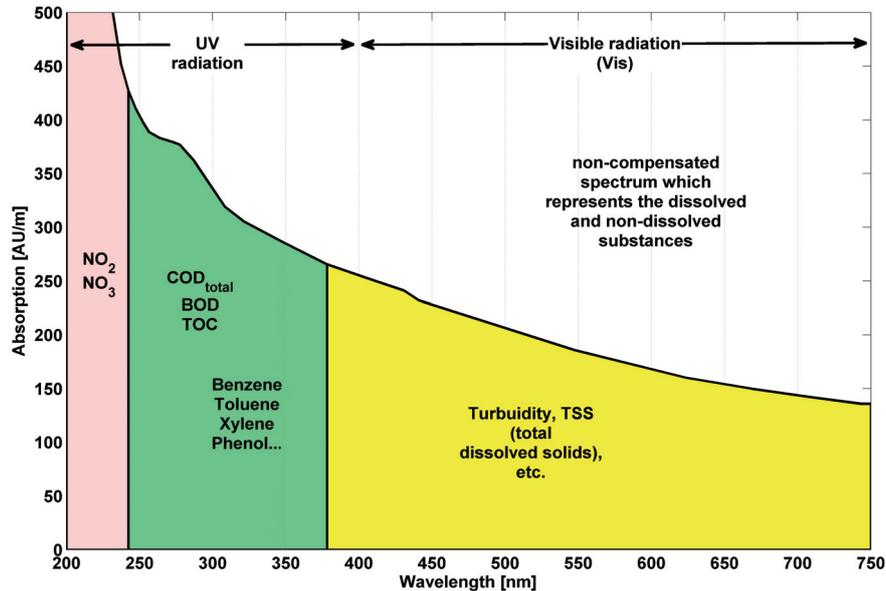


Figure 3-5: Significant regions of a typical UV/VIS absorption spectrum (non compensated)

The measurement in Figure 3-5 is not yet compensated against turbidity which results in errors when calculating the other parameters. In reality the measurement data at all wavelengths are impacted by turbidity, which makes it more complicated to differentiate between turbidity and the concentration values of the measured substances. Figure 3-6 shows the turbidity compensated absorption spectrum for the same measurement. For this compensation the absorption at 350 nm, which represents mainly the turbidity is subtracted. In direct comparison it becomes obvious that in this case the turbidity has a considerably higher influence on the absorption than other concentration values (note that the scaling of the y-axes is different).

In contrast to nitrate, which is a relatively easy to measure substance and has a very specific absorption area (about 210 nm), COD is much more complex to measure. Due to the fact that COD is a sum parameter which represents the sum of all organic compounds and not a specific compound, there are no generally valid wavelengths to calculate COD. For this reason a special calibration (a set of coefficients  $a_i$ ), depending on the specific kind and composition of wastewater, is necessary.

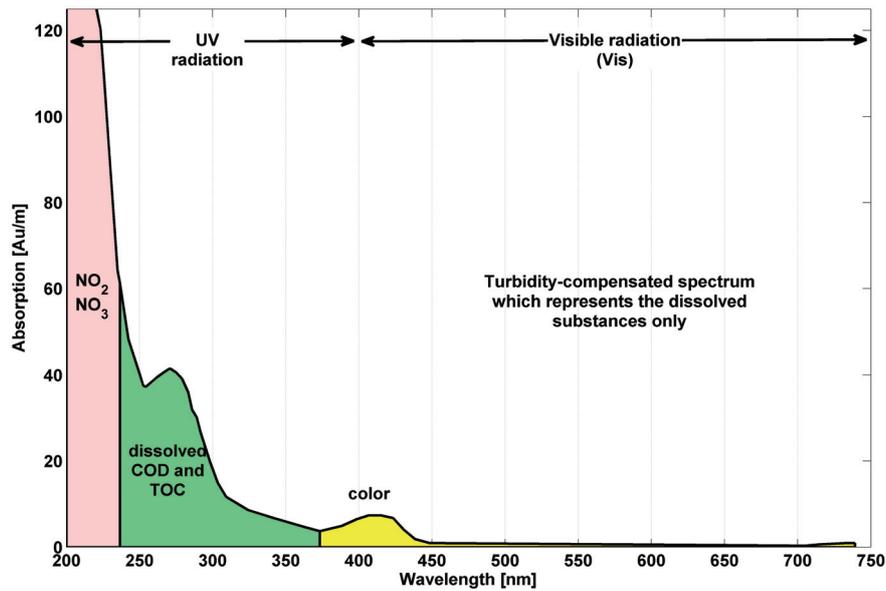


Figure 3-6: Significant regions of a typical UV/VIS absorption spectrum (compensated)

The measurement of *COD* differs from the classical approach described above, where usually turbidity is compensated and known wavelengths are used. The *COD*-analysis uses turbidity and additional measurement values as well. This is possible because of the specific composition of the wastewater. In simple terms it means, when wastewater looks much polluted or has a strong colour, it can be assumed that it contains a high amount of *COD* etc. This rule is mostly valid, even if the colour or the dirty look has no chemical connection to the *COD* concentration. The application of the method is limited, if the composition of the wastewater often changes substantially. An example could be a construction site near to the WWTP, where the water is polluted with mud. The mud itself won't contain high amounts of *COD*, but it will change the turbidity and colour of the water, which will lead to wrong results.

### 3.1.3.2 Ion-selective probes

Ion-selective in-situ sensors are a relatively new technique in the area of wastewater treatment. The first commercially available devices were brought to the market about 10 years ago. The advantages compared to classic analysers are:

- smaller size
- no sample preparation
- small delay time (<2 min )
- low price

However, there are also some disadvantages in comparison to classical analysers:

- higher maintenance requirements (calibration is necessary in intervals between a week and a month)
- measurement errors by cross-sensitivities (ammonium to potassium, nitrate to chloride)
- tendency to drift (offset)
- membrane ageing (loss of sensitivity in measurement values)
- adversely impacted by several substances (membrane poisoning)

Today many operators install ion-selective probes instead of analysers. Especially small plants, which were in the past only equipped with basic measurement probes, can now operate ion-selective online ammonium and nitrate measurement probes.

A typical ion-selective sensor system is composed of four components:

- An ion-selective electrode (ISE)
- A reference electrode (Ref)
- A temperature sensor
- An amplifier and controller unit

The combination of an ISE and a reference electrode is called a measuring chain (Figure 3-7). The ISE is composed of a tube filled with an electrolyte, an inner electrode and a PVC membrane at the end. The membrane contains so called ionophores, special molecules which are able to transport specific ions. These ionophores make the membrane permeable for specific ions. Due to the different chemical concentrations in the electrolyte and the medium, an osmotic pressure arises at the membrane. This pressure causes a movement of charge which builds up an electrical gradient over the membrane until equilibrium between the two forces is reached. Since only specific ions (depending on the ionophores) are able to permeate the membrane, the electric potential at the equilibrium depends on the target ion.

The purpose of the reference electrode (Figure 3-7) is to establish a stable connection to the liquid medium. It builds up an electric potential which does not depend on the medium to be measured.

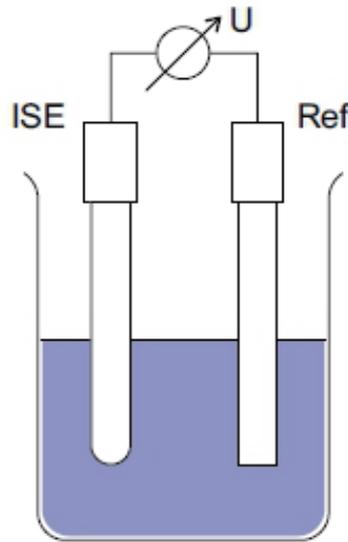


Figure 3-7: Schematic ISE (WTW GmbH, 2007)

The difference between these potentials  $U_{ion}$  (potential of the electrodes) can be measured and is described by the Nernst equation (Cammack et al., 2006):

$$U_{ion} = U_{ion}^0 \pm S \cdot \log(a_{ion}) \quad (3.3)$$

Where  $U_{ion}^0$  is a fixed voltage depending on the measurement system (standard potential of the electrode at a reference point),  $S$  is the slope which describes how much the signal changes depending on the concentration of the target ion (typically 58 mV-59 mV for ammonium or nitrate (single charged ions) with a new membrane) and  $a_{ion}$  is the active concentration of the target ion. The activity takes into account the influence by other ions in the medium that cause matrix effects. The slope is defined as:

$$S = \frac{RT}{zF} \quad (3.4)$$

where  $R$  is the gas constant ( $8.314J \cdot K^{-1} \cdot mol^{-1}$ ),  $T$  the temperature in Kelvin,  $z$  the valence of the ion and  $F$  the Faraday constant ( $96485C \cdot mol^{-1}$ ). Here it becomes obvious that the temperature has a major influence on the measurement. For this reason every ion-selective probe measures the temperature. In practice ISEs are influenced by other ions (cross sensitivity). For this reason the equation has to be extended:

$$U_{ion} = U_{ion}^0 \pm S \cdot \log(a_{ion} + K_{ion,interf} \cdot a_{interf}) \quad (3.5)$$

where  $K_{ion,interf}$  is a dimensionless coefficient which describes the influence of other ions and  $a_{interf}$  is the activity of these ions. By rearranging the equation, the active concentration  $a_{ion}$  can be calculated as:

$$a_{ion} = 10^{(U_{ion} - U_{ion}^0) \cdot S^{-1}} - K_{ion,interf} \cdot a_{interf} \quad (3.6)$$

In practice, it has to be considered that the active concentration  $a_{ion}$  is influenced by the amount of other ions present in the medium. There are two ways to calibrate the probe: 1. An offset calibration which corrects  $U_{ion}^0$ . A change of  $U_{ion}^0$  is what can be seen as drift. 2. A two point calibration to correct the slope  $S$ . Current electrodes are able to keep the slope  $S$  stable over a longer period of time, which is why some manufacturers, such as Hach, claim that two point calibrations are not necessary anymore because the quality of the membranes has improved substantially. Personal application experience has found that in practice the slope decreases over time. This effect was strongly present when the probes were placed in the WWTP inflow stream.

In chapter 4, results using these measurement systems are presented and discussed.

### 3.2 WWTP Control

The control loops utilised in treatment plants differ significantly in some relevant areas from traditional control loops in process engineering systems. For example, the volume of inflow and its composition is constantly changing. This causes complications for the control system. The biochemical processes that take place in the plant are not linear and there are interactions between the individual processes. Furthermore, the number of manipulatable variables is very low in relation to the complexity of the system. In wastewater treatment plants, controllers are used to control internal flows, dose chemicals and regulate oxygen levels.

In general, three different types of controllers can be found in wastewater treatment plants:

1. Control without feedback (open loop – feed forward) (OL)
2. Control with feedback (closed loop – feed back) (CL)
3. Predictive control (PC)

Control systems in the first category are now only utilised for simple applications such as pumps. In the area of oxygen regulation, closed loop (CL) controllers are primarily used, whereby a typical application is a  $NH_4 - N$  cascade controller. In rarer cases, predictive controllers are also utilised but their development costs mean they can usually only be financed as part of research and development projects.

One example is the wastewater treatment plant in Hildesheim, which was optimised with the aid of an online simulation model (Rosenwinkel, 2003).

**Calculation of the setpoint and maintaining the setpoint**

It is common in most controllers for the maintenance of the oxygen setpoint to play a subordinate role. It is more important to define the optimal oxygen setpoint. Figure 3-8 shows the basic structure of an oxygen  $NH_4 - N$  cascade controller.

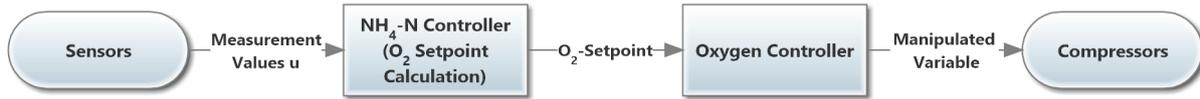


Figure 3-8: Aeration  $NH_4-N$  Cascade Controller

The important point is to differentiate here between the two parts:

- Calculation of the setpoint
- Maintaining the setpoint

As the status of the wastewater treatment plant is constantly changing (biology, inflow, load, etc.), it is necessary to calculate the appropriate oxygen setpoint in each situation. The ammonium concentration is mostly used for this purpose. Although the aim is to maintain the setpoint, deviations of 10 to 30% do not endanger the stability of the plant. In the aeration tanks, the oxygen content can vary by a similar amount depending on the flow conditions so that just by measuring the oxygen in the tanks at different positions similar variations in the concentration can be observed. Similar local dynamic effects can also be expected for the measurement of the concentrations of other substances.

**Typical control loops / main control loop**

Depending on the process technology, there are a series of typical control loops used in treatment plants, which can be described as the main control loop. Table 3-3 provides an overview of these control loops, the operation types in which they are used and the relevant controller type categories.

Table 3-3: Overview of main control loops in WWTPs

Control Loop	Used for Operation Type	Controller Type
<b>Aeration</b>	all types	CL, FF
<b>Recirculation</b>	upstream denitrification	OL, CL
<b>Return Sludge</b>	all types	OL, CL
<b>Phosphate Precipitation</b>	all types	OL, CL, PC
<b>Filtrate Water</b>	all types	OL, CL
<b>Excess Sludge</b>	all types	OL, CL

In general, every one of these controllers can be designed as an OL, CL or PC system. Table 3-3 shows an overview based on practical considerations and empirical values. In many cases, the effort and costs involved in additional measurement technology is not proportionate to the achievable results. Therefore,

many plants do not have a inductive flow measurement system for the recirculation so that only an OL control system is possible for this variable.

Although every one of these controllers is firstly considered individually, they sometimes have a strong influence on each other. A challenge for control system developers is thus to recognise and avoid as best as possible any potential negative effects on other control loops without having any major negative consequences on their own control objective. In the following sections, the control loops and their most important interactions with other control loops will be briefly described.

### 3.2.1 Aeration Control

#### General Notions concerning Aeration Control

Control of the oxygen content in the aeration tanks is usually the largest energy consumer in a wastewater treatment plant: The compressors or blowers require the most electrical energy in the plant and thus represent one of the largest running costs. Any reduction in aeration will substantially decrease operating costs. For this reason, considerably more effort is invested in the control of oxygen in many cases than in the development or optimisation of the other controllers.

In practice, an oxygen concentration of  $2 \text{ mg / l}$  is frequently set. Although this is a safe option, it is far from the optimal solution in many cases. Therefore, the question arises: *What is the right oxygen concentration?*

The answer to this question is not simple, it depends to a great degree on the operating and process conditions. The following factors play a role here:

- the load placed on the plant
- the inflow dynamics (strong forces make higher concentrations necessary for safety reasons)
- the distribution of the oxygen in the tank (avoidance of anoxic zones)
- the design of the plant
- the control objective (saving energy vs. lower effluent concentrations)

This list covers only a selection of the typical factors; there are often other individual criteria in practice. This is also the reason why efficient controllers for wastewater treatment plants generally need to be individually customized. Approaches for calculating the optimal  $O_2$  setpoint are, for example, the method described by A. Niet (Niet et al., 2011), which attempts to minimise the sum of  $NH_4 - N$  and  $NO_3 - N$  or Lindberg (Lindberg and Carlsson, 1996) which take the nonlinear oxygen transfer function into account to design a nonlinear oxygen controller. As treatment plants are never in a steady state in reality, this calculation remains a challenge in operational practice.

**Continuous oxygen control**

The standard system used today for regulating the oxygen value consists of two or three cascade controllers:

1. An  $NH_4$  controller with a fixed  $NH_4$  setpoint that calculates an oxygen setpoint based on its control deviation.
2. An  $O_2$  controller that calculates the airflow with the aid of the  $O_2$  control deviation (or calculates the speed of the compressor directly)
3. An airflow controller that maintains the desired flow of air by controlling the valves on the air lines based on its control deviation.

Every single one of these control loops has individual non-linearities so that they must be setup and optimised separately. Controller 3 is often utilised in systems with highly non-linear control elements e.g. prism valves in the air lines (Olsson et al., 2005). Today, regulating valves are sometimes omitted to save energy and the airflow is controlled directly via the compressor. This avoids the otherwise inevitable energy losses at the valves. In these cases, the third controller is omitted.

**Feed Forward Control**

In order to be able to guarantee that the effluent limits can be maintained even when the inflow has a highly dynamic nature, it is necessary to select a correspondingly low  $NH_4 - N$  setpoint. Alternatively, additional predictive components can be introduced into the control system. These utilise measurement values from the wastewater treatment plant inflow or from the upstream sewage system. The changes to the  $NH_4$  setpoint can thus be adjusted according to the measurement of the inflow volumes. Better results are achieved when  $NH_4 - N$  or  $COD$  measurement values from the inflow are used. These methods make it possible to use higher  $NH_4$  setpoints when the plant is subject to lower loads and vice versa. In practice, these predictive components enable the  $O_2$  setpoint to be increased at an early stage when a sudden increase in load is recognised in the inflow or in the sewage system. This means that large energy savings can be made without making any major sacrifices to operational safety. However, safe operation in this case necessitates relatively high maintenance costs as the measurement location itself in the inflow or sewage system can quickly lead to disruptions or the breakdown of the measurement technology. This is one of the most important motivations for the development of virtual sensors that will be presented in Chapter 5 (Virtual  $COD$  and  $NH_4 - N$  measurements).

**Intermittent oxygen control**

Intermittent  $O_2$  controllers are required in certain plants for procedural reasons. Older plants often only have one aeration tank. They date back to the time when the primary objective was to eliminate  $COD$ , which required basic and continuous aeration. These plants were later converted by the introduction of

intermittent operation for nitrification and denitrification. From a control engineering perspective, these plants are very interesting because they offer more possibilities for intervention:

- $O_2$  setpoint
- Length of the aerated or non-aerated phase
- Start time of each phase.

Simple controllers work with fixed  $NH_4 - N$  switching points so that, for example, the aeration is switched on at  $4 \text{ mg / l } NH_4 - N$  and then switched off again at  $2 \text{ mg / l } NH_4 - N$ . The challenge here is to decide whether longer nitrification phases with lower  $O_2$  concentrations or shorter phases with higher  $O_2$  concentrations are better for the specific plant. Depending on the load experienced by the plant, longer nitrification phases are better in terms of energy consumption but mean that the length of the denitrification phases are no longer sufficient. Within the scope of this work, controllers were developed that could determine both the  $NH_4 - N$  removal rate:

$$NH_4 - N_{\text{removal rate}} = \frac{dNH_4 - N}{dt} \quad (3.7)$$

and the  $NO_3 - N$  removal rate:

$$NO_3 - N_{\text{removal rate}} = \frac{dNO_3 - N}{dt} \quad (3.8)$$

The optimal  $O_2$  setpoint is the value that minimises the ratio between the  $O_2$  concentration and ammonium removal  $NH_4 - N_{\text{removal rate}}$ , that is:

$$O_2(\text{optimal Setpoint}) = \min \left( \frac{O_2}{NH_4 - N_{\text{removal rate}}} \right) \quad (3.9)$$

However, this approach has a number of limitations in practice:

- The removal rates are difficult to determine and change greatly because the inflow acts as a disturbance variable.
- The  $O_2$  concentration is not the optimal key parameter because it gives no information on either the volume of air provided or the energy consumption. However, this information is often not available.
- If the load experienced by the plant is high, the nitrification phases must be kept short to provide sufficient time for the denitrification phases. The required length of the denitrification time can be estimated based on the nitrate removal  $NO_3 - N_{\text{removal rate}}$ .

The change in the length and starting points of the nitrification and denitrification phases provide good opportunities for the dynamic control of plant operation based on the load. In many cases, the desired

$O_2$  setpoint cannot be adjusted dynamically enough (see subsection 3.1.1). In order to compensate for this, the nitrification phase can be started, for example, earlier if there is a sudden increase in load so that sufficient capacity for the nitrification is made available.

There are a variety of approaches for calculating the  $NH_4$  concentrations at which the aeration should be switched on and off. The most important are the statutory effluent values and the safety-related considerations. It is possible in normal operation e.g. to select a  $NH_4 - N$  switching point at the end of the denitrification phase that is higher than the statutory  $NH_4 - N$  effluent limits because a combination of denitrification and nitrification takes place in the final treatment tank. In the case of a sudden and sharp increase in load shortly before or after reaching the switching limit, e.g. a rain event, there is however a risk that the effluent limits could be breached.

### Variable zones

A special feature of intermittent control is variable zones. These are tanks or zones that can be aerated depending on the load experienced by the plant. The principle is similar to the intermittent operation described above, although it is implemented as an additional component in other plant designs.

### Safe operation

A more important aspect than saving energy for the operator is the safe operation of the plant, which includes in particular the safe observance of limit values. As measurement devices do not always work correctly (see subsection 3.1), it is necessary to recognise any possible incorrect measurement and to build in fallback strategies. The controllers developed by the GECO►C Research Group (Betz, 2015) employ a strategy involving multiple fall-back levels. Figure 3-9 shows a representation of these as a shell model.

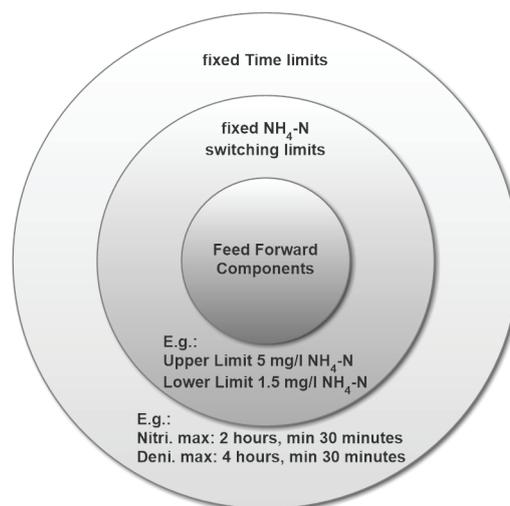


Figure 3-9: Controller shell model

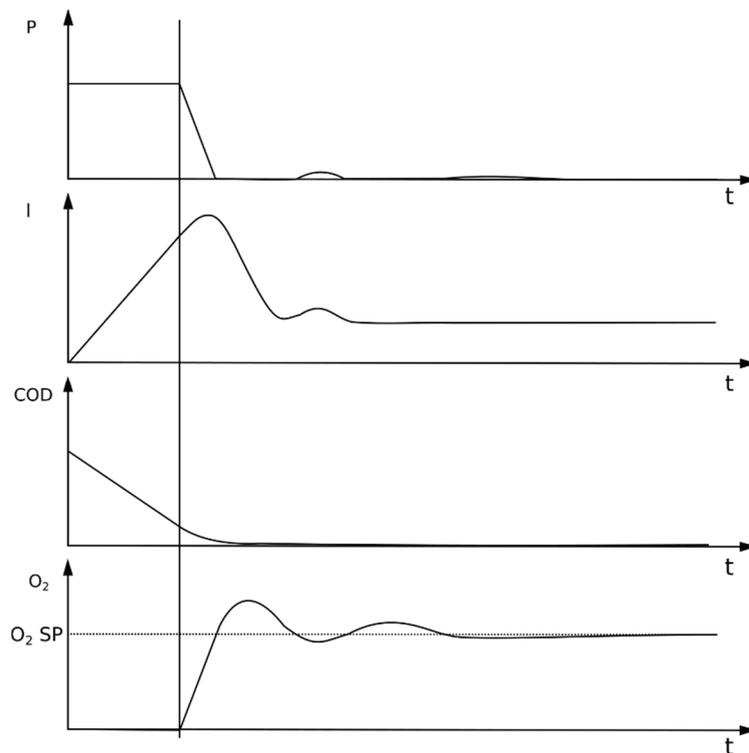
In normal operation, the controller is located in the innermost shell. The switching limits are calculated according to the described method. However, if these limits exceed the stated limits on the middle shell

then they are ignored and overridden. The same principle is valid for the middle shell. For example, there is an automatic changeover to the nitrification phase after four hours of denitrification even if the lower  $NH_4 - N$  limit has not yet been achieved. In this manner, it is possible to guarantee continued safe operation even if a measurement device breaks down overnight when the personnel are absent.

### **$O_2$ -Control**

The  $O_2$  controller is generally just one component in a cascade control system (see Figure 3-8). Although many different types of system, e.g. fuzzy control (Bongards et al., 2005), have been successfully implemented in the last few years, the most common controller is still the PI or PID controller. One reason for this is certainly their high degree of availability in standard development environments for control systems. While the first stages of a cascade control system are partially implemented in the SCADA system, particularly those that use CPU-intensive methods, the  $O_2$  controller is mostly implemented independently on the PLC itself. The practical limits of this controller will be briefly described using the example of intermittent control.

The control deviation is at its maximum at the start of the nitrification phase because the actual value is  $0 \text{ mg/l}$ .



**Figure 3-10: Behaviour  $O_2$ -PI-Controller in the beginning of nitrification**

At the beginning of the nitrification phase, oxygen is consumed at a very high rate and this continues until the  $COD$  has been largely used up. The  $O_2$  concentration in the tanks during this period mostly remains at  $0 \text{ mg/l}$ . Afterwards, the  $O_2$  concentration rises quickly. The start of this rise is marked in

Figure 3-10 by a vertical line. During the first phase, the I-term of the controller rises so that although the P-term falls when the setpoint (SP) is reached, the I-term remains high. This leads to an overshoot in the  $O_2$  concentration. The time it takes for the consumption of oxygen up to the beginning of the nitrification phase is dependent on the  $COD$  concentration, meaning this phase can take a variable amount of time and thus displays highly non-linear behaviour. This effect is not very disruptive in normal operation but it does cause unnecessary energy consumption for the aeration. PI or PID controllers with different settings for different phases of the nitrification can provide assistance here but these types of solution are not widely used in practice.

### Interactions with other controllers

As mentioned at the start of this chapter, there are dependencies between the different control loops that need to be taken into account. Figure 3-11 shows the most relevant dependencies for the oxygen controller.

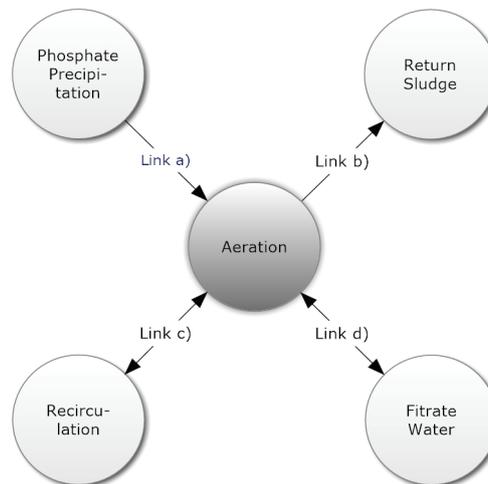


Figure 3-11: Aeration Control - Interaction with other control loops

#### Link a)

- Phosphate is stored in the biomass during aeration. This means less precipitant is required.
- If anaerobic states occur in phases without aeration, there is a risk of redissolution.

#### Link b)

- High volumes of degradable biomass in the return sludge require oxygen so that more air needs to be provided.

#### Link c)

- High levels of recirculation in combination with high oxygen concentrations can lead to oxygen carryover from the nitrification phase to the denitrification phase, which endangers nitrate removal. The removal of high levels of nitrate during the denitrification phase is usually achieved through greater recirculation.

Link d)

- Filtrate water from sludge treatment increases the ammonium concentration, which leads to higher oxygen consumption.
- High levels of aeration lead to lower levels of ammonium. These are generally beneficial situations for adding filtrate water to the activated sludge.

### 3.2.2 Recirculation Control

Recirculation control is a special case and is only required in plants with upstream denitrification. It is responsible for transferring the nitrate formed during the nitrification phase into the inflow of the denitrification phase (see section 2.2.2.2). In general, two typical approaches are used for recirculation control:

- Inflow-proportional control
- Nitrate-based control

Inflow-proportional control is the simplest approach. In this case, the volume of recirculated water is controlled at a fixed rate to the inflow. Therefore, this is a type of non-feedback control as it is not regulated by feedback on the volume of nitrate. This has the disadvantage that in the case of already high inflow volumes and correspondingly shorter retention times in the tanks, the recirculated water shortens the retention times even further.

In nitrate-based control, the  $NO_3 - N$  concentration of the effluent from the aeration tank is used as the key parameter. The control principle is as follows: If the  $NO_3 - N$  concentration increases, the volume of recirculated water is increased so that a higher proportion of nitrate flows through the denitrification stage.

#### Interactions between the recirculation and other controllers

Changes to the recirculation have an effect on oxygen control if nitrate-based control is used. In the case of higher  $O_2$  concentrations, there is an increased amount of  $NO_3 - N$  which leads to higher levels of recirculation. In turn, the shorter retention times can cause poorer  $NO_3 - N$  degradation, whereby most oxygen controllers respond by reducing the  $O_2$  setpoint.

### 3.2.3 Phosphate Precipitation Control

A variety of different iron salts are utilised as precipitants for the removal of phosphate. A diverse range of products based on iron or aluminium are available on the market. The goal of all control systems is to maintain the phosphate limit values in the effluent and at the same time keep the precipitant costs to a minimum. The following approaches are described by the DWA (ATV-DVWK, 2001):

- Continuous dosing

- Dosing according to a schedule
- Dosing according to the flow of wastewater
- Dosing according to the phosphate load
- Dosing according to the phosphate concentration

During continuous dosing, a dosing pump runs permanently delivering a constant dose. In order to safely remove the phosphate, overdosing of the precipitant is carried out in most cases. Dosing according to a schedule is the simplest improvement, where the dosing quantity is adjusted based on empirical values depending on the time of day. In dosing according to the flow of wastewater, the precipitant is dosed within a controlled range in proportion to the flow of wastewater. Fixed values are used above and below this range. Control based on the phosphate load is significantly more efficient because it utilises a phosphate measurement. This involves multiplying the wastewater volume by the  $PO_4-P$  concentration to calculate the required precipitant flow  $Q_{PA}$  (example for iron-based precipitant):

$$Q_{PA} = Q \cdot C_p \cdot f \cdot \beta \cdot \frac{55.8}{30.9} \cdot \frac{1}{\rho} \cdot \frac{1}{m_{ME}} \quad (3.10)$$

using the wastewater volume  $Q[l/s]$ , the phosphorus concentration  $C_p$ , the safety factor  $f$ , the Beta value  $\beta[\text{Fe/mol P}]$ , the density of the precipitant  $\rho[\text{kg}/\text{m}^3]$  and the active metal content of the precipitant  $m_{ME}[\text{kg Fe/kg PA}]$ . It is often the case that the precipitant is hyperstoichiometrically dosed so that a safety factor  $f$  of over 1 is selected. The value  $55.8/30.9$  represents the ratio of the molecular weight of iron to phosphorus.

Control according to the phosphate concentration corresponds to a traditional feedback control solution, where the phosphate concentration is measured after the dosing point.

### Biological Phosphate Removal and Interactions with Other Controllers

As described in section 2.2.4, the targeted application of aeration makes it possible to store phosphate in the biomass and to remove it from the plant in the form of excess sludge. It is for this reason that a value of under 1 can be selected in the calculation of the safety factor  $f$  in many plants. This can lead to significant savings in the amount of precipitant used. A problem experienced in biological phosphate precipitation is so-called phosphate redissolution. Here, the biomass releases the phosphate again which occurs especially under anaerobic process conditions. As this process occurs very quickly, it can lead to high phosphate peaks and endanger the effluent limit values. In intermittently operated plants, this effect can be avoided by preventing anaerobic states through briefly aerating the wastewater during denitrification.

### 3.2.4 Filtrate Water Control

Filtrate or process water (PW) is created during the dewatering of sewage sludge. Depending on the type of sludge and the concentration method used, the process water is subject to varying levels of contamination. The most heavily contaminated PW comes from digested sludge. It is characterised by a high  $NH_4 - N$  concentration and a comparatively low  $COD$  concentration. This process water needs to be fed back into the plant. It is often the case that this is performed continuously throughout the day. It is more sensible to dose the process water during phases when the plant is experiencing low loads so that the plant is not overburdened. A typical mistake that is often made is to dose the process water at night. Although this is when the plant is experiencing the least load and there are low  $NH_4 - N$  concentrations, there is a lack of carbon for denitrification so that the nitrate created cannot be degraded (see section 2.2.2.2). An improvement can be made by dosing according to a schedule because the times of extremely high or low load are often known and it is possible to plan accordingly. In addition, it is sensible to integrate the fill level (FL) of the process water tank into the control system. As the tank's capacity is limited, the control system must take this into account and dose higher volumes before there is a danger that no more capacity is available. A tried-and-tested practical solution developed by the GECO►C research group is the following relatively simple formula (3.11).

$$PW(FL) = \begin{cases} 10 & \text{if } 0 < FL < 25 \\ 10 + FL \cdot 1.12 & \text{if } 25 < FL < 75 \\ 66 + FL \cdot 1.40 & \text{if } 75 < FL < 100 \end{cases} \quad (3.11)$$

Based on the graphical representation in Figure 3-12, it is easy to recognise that the dose of PW increases as the FL increases. The slope of the curve can be individually adapted to the plant where required.

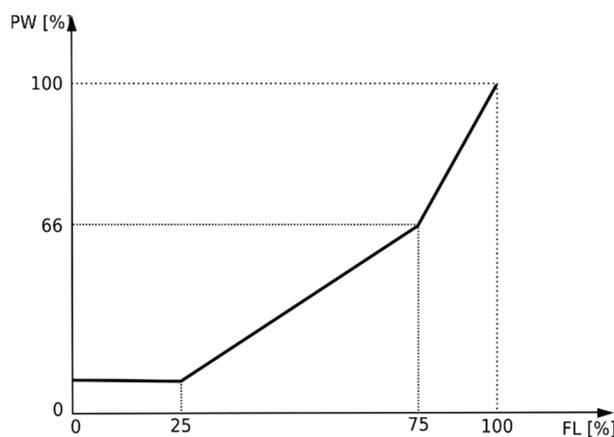


Figure 3-12: Process water control

### 3.2.5 Return Sludge Control and Excess Sludge Control

All activation plants with a final treatment tank must feed the separated biomass (see section 2.3.2.4) back into the bioreactors. Therefore, the volume of active biomass in the reactors is directly controlled by the volume of return sludge. Different processes have become established for controlling the return sludge:

- Inflow-proportional control
- Control based on the volume of sludge in the aeration and/or final treatment stages
- $NO_3 - N$  -based control

The inflow-based process is always utilised if the plant is not capable of measuring the dry matter content ( $TSS$ ) in the aeration tank or the sludge level in the final treatment tank. In other cases, the volume of sludge can be controlled so that a constant amount of dry matter is retained in the aeration tank or a predetermined sludge level in the final treatment tank is not exceeded.

The nitrate degradation in the aeration tank can also be used as a control variable: If it is insufficient, more biomass or activated sludge is pumped into the aeration tank.

Advanced control systems also use the volume of return sludge to ensure there is more active biomass in the aeration tanks during peak times or to partially store the biomass in the final treatment tanks during phases of low load. As the associated pumping processes can take many hours, it is necessary to be able to estimate the behaviour of the plant and particularly the variations in the levels of inflow in advance.

Part of the return sludge is discharged as excess sludge. It is common to divert a fixed proportion or to use the dry matter content as a guide value. The goal is to maintain the active biomass in the plant at a constant level. In addition, it is possible in this way to control the age of the sludge in the plant, which should remain within certain limits depending on the type of plant.

## 3.3 Summary

The chapter shows that wastewater treatment plant control is much more than the well-established aeration control but rather a comprehensive optimization of several control loops including recirculation, phosphate precipitation, filtrate water dosage and return sludge as well (plant-wide control). It becomes evident that all those control loops influence the biological treatment process and thus each other to a great extent, which is why the optimization of a single control loop is mostly not constructive. Instead the overall operating and process conditions at different treatment stages of the plant need to be considered.

Nevertheless, the overview of available online instrumentation for process monitoring clearly illustrates that this is easier said than done. In particular, cross-sensitivities, sensor drift and high maintenance, depending on the type of instrumentation used, turn the proper assessment of the operating state of a

plant into a challenge. Based on these limitations on the instrumentation side, the control side needs to rely on sufficiently robust and financially feasible online instrumentation, in order to come up with practice-oriented optimization and control solutions that achieve a high acceptance by plant operators.

Therefore, this chapter sets the scene for the developed soft-sensor solution in section 5 as well as for the SOM- and state-based control in section 7 by providing insight into and justification for the line of thought that these novel methods are based on.

## 4 WWTP Modelling

This chapter provides a general description of the modelling of wastewater treatment plants and especially the modelling of the Rospe treatment plant with the help of the Activated Sludge Model No. 1 (ASM1) (Henze et al., 2000). Initially, an introduction to the basics of modelling wastewater treatment plants is given. The plant and the implementation of the model in Matlab® Simba is then described. This is followed by a description of the model calibration. The calibration is initially carried out manually and then a quicker and more efficient method using the multi-parameter optimisation process SMS-EGO (Zitzler and Thiele, 1998) is presented.

Modelling of wastewater treatment plants is a combination of models of multiple processes. It is initially necessary to model the hydraulic conditions in the different reactors. The biological processes, process kinetics and stoichiometry then need to be modelled.

### 4.1 General wastewater treatment plant models

#### 4.1.1 Modeling of the hydraulics

Although the hydraulic conditions are a relevant factor in the behaviour of a plant, they are highly simplified in many cases due to their complexity. It is common practice to view reactors as continuously stirred tank reactors (CSTR) or as zero order systems. It is assumed here that the concentrations are the same across the entire reactor. In order to model plants in which large changes in concentration in the direction of flow occur, such as oxygen ditches or plug-flow systems, first order systems are sometimes used. In general, it is possible to differentiate between three types of reactors for the hydraulic description: batch reactors, stirrer tanks and plug-flow reactors. For larger plants, stirrer tank reactors are primarily used.

##### 4.1.1.1 Batch reactors

Batch reactors are tanks in which the process is completed as a batch job. This means that they are filled, the desired process is completed and then the tank is emptied. Assuming that nothing flows into or out of the tank during the process – i.e. the volume remains constant – then the change in concentration can be described as follows:

$$\frac{dc}{dt} = r \left[ \frac{mg}{l \cdot s} \right] \quad (4.1)$$

The concentration is represented by  $c$  and the conversion rate by  $r$ . In wastewater treatment, batch reactors are popular solutions for small wastewater treatment plants or highly contaminated industrial wastewater. It is important to note here that the process often already starts during the long filling phase and therefore there is a continuous change in volume in many cases.

#### 4.1.1.2 Continuously stirred tank reactors (CSTR)

Continuously stirred tank reactors are the most commonly used type of reactor for the modelling of wastewater treatment plants. As the name already suggests, it is assumed that they are completely mixed and thus the concentrations of the individual substances are the same throughout the entire reactor. The change in concentration  $\frac{dc}{dt}$  can be described as follows:

$$\frac{dc}{dt} = \frac{Q(c_{in} - c_{out})}{V} + r \quad \left[ \frac{mg}{l \cdot s} \right] \quad (4.2)$$

In contrast to batch reactors, the inflowing and outflowing load is calculated and added to the conversion rate  $r$ . This simple approach is based on the assumption that the volume remains constant. In practice, there is often a small change in volume because the fill level in the relevant tanks can change by a few cm if there are high levels of inflow e.g. in the event of rain. This effect is intensified if the tanks are not directly connected but the medium passes through a pump.

#### 4.1.1.3 Stirred tanks in series

In principle, stirred tanks in series are the standard solution for most plants and are simply a series of linked stirred tanks. In this process, every stirred tank behaves as a CSTR as described in section 4.1.1.2.

$$\frac{dc_1}{dt} = \frac{Q(c_{in1} - c_{out1})}{V} + r \quad \left[ \frac{mg}{l \cdot s} \right] \quad (4.3)$$

$$\frac{dc_2}{dt} = \frac{Q(c_{in2} - c_{out2})}{V} + r \quad \left[ \frac{mg}{l \cdot s} \right] \quad (4.4)$$

The outflowing concentration  $c_{out1}$  is thus identical to the inflowing  $c_{in2}$  (Figure 4-1).

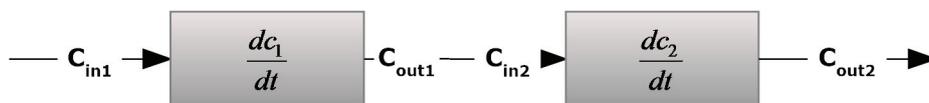


Figure 4-1: Stirred tanks in series

#### 4.1.1.4 Plug-flow Systems

In plug-flow systems, it is assumed that the water passes through the tank like a plug. There is thus no complete mixing of the water. This results in a concentration gradient in the direction of flow (z-axis) in the tank. The mathematical description of the change in concentration in a plug-flow is as shown in equation (4.5):

$$\frac{\delta c}{\delta t} = \frac{Q}{A} \cdot \frac{\delta c}{\delta z} + r \left[ \frac{g}{m^3 \cdot d} \right] \quad (4.5)$$

The reduction in concentration over time is represented here by  $\frac{\delta c}{\delta t}$ , while  $\frac{\delta c}{\delta z}$  represents the reduction in concentration along the axis  $z$  and the area is represented by  $A$ . In direct comparison, stirred tanks in series reduce the concentration in steps, while plug-flow reactors reduce the concentration continuously. As it is not really possible to mix a stirred tank perfectly, the cascade effect is also present in the stirred tank to some extent.

#### 4.1.2 Modeling biological processes

The following subsections provide an introduction to the mathematical description of biological processes. This consists of a description of the kinetics, meaning the rate of conversion, and the stoichiometry, meaning the actual conversion of the materials.

##### 4.1.2.1 Kinetics / Monod equation

Through his experiments on the growth of bacterial cultures in discontinuously stirred tanks at the start of the 1940s, the French scientist Monod recognised parallels with the Michaelis-Menten theory due to the non-linear relationship between the specific growth rate and the limiting substrate and formulated analogous relationships for bacterial growth (Gerber, 2009). The so-called Monod equation is today a widely used method for describing the rate of conversion. Equation (4.6) shows an example of a typical Monod equation. The Monod equation is very precise for pure cultures and simple substrates (Gerber, 2009).

$$\mu = \mu_{\max} \cdot \frac{S}{S + K_h} \quad (4.6)$$

The growth rate is represented here by  $\mu$ , the maximum growth rate by  $\mu_{\max}$ , the substrate concentration by  $S$  and the so-called Monod constant by  $K_h$ , which gives the substrate concentration at which 50% of the maximum growth rate  $\mu$  is achieved. It is noticeable in Figure 4-2 that the growth rate increases sharply at low substrate concentrations and then always flattens out when it nears the maximum growth rate  $\mu_{\max}$ .

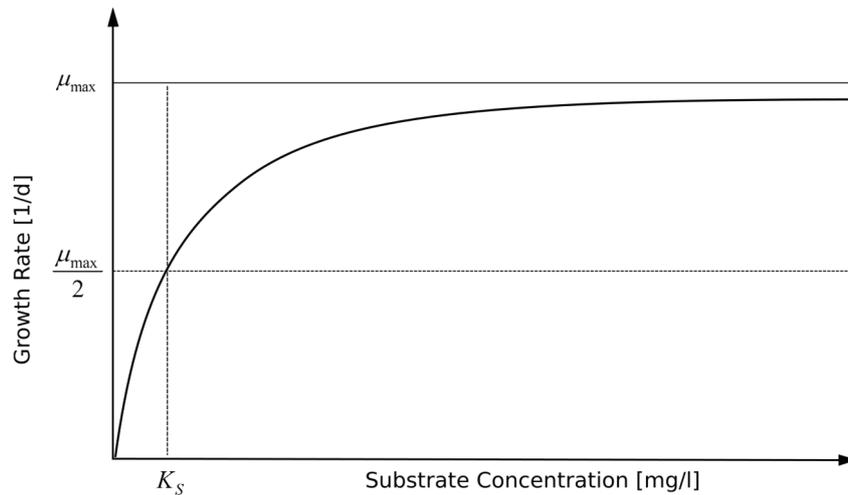


Figure 4-2: Growth behaviour - Monod kinetics

Furthermore, a differentiation is also made between growth kinetics and inhibition kinetics. In growth kinetics, the conversion rate rises with an increased amount of substrate, while it falls in inhibition kinetics. A typical example for growth kinetics in the area of wastewater treatment is nitrification. In this case, the conversion rate increases along with the availability of ammonium.

#### 4.1.2.2 Stoichiometry

Stoichiometry describes the conversion of materials. This is formulated with the help of chemical equations. Stoichiometry does not only describe the reactions but also gives additional information about the precise proportions of the materials. An example of a stoichiometric equation can be found, for example, in the description of the denitrification process in section 2.2.2.2 (equation (2.6)).

#### 4.1.2.3 Matrix representation

In order to illustrate the complex dynamic reaction model more closely, a matrix notation has become established based on the work of Peterson (Peterson, 1965). The so-called Peterson Matrix describes a combination of the kinetics and stoichiometry. Table 4-1 shows an example of the aerobic growth and decay of a group of microorganisms. The table is structured as follows:

- The first row describes the different components involved in the processes with their indices  $i$ .
- The leftmost column describes the different processes with their indices  $j$ .
- The rightmost column describes the kinetic expressions / process rates  $\rho_j$  in the corresponding row.
- The lower left corner describes the stoichiometric coefficients and the lower right corner the kinetic parameters.
- The elements in the matrix are the stoichiometric coefficients  $v_{ij}$ .

In order to calculate the reaction rate of an individual component, follow the relevant column for the component downwards and multiply the stoichiometric coefficient for each process  $j$  by the relevant

process rate  $\rho_j$ . This is repeated for every process and the individual results are then added together. In the case of  $X_B$ , this gives, for example, the process rate  $r_{X_B}$  as follows:

$$r_{X_B} = \frac{\hat{\mu}S_s}{K_s + S_s} X_B - bX_B \quad (4.7)$$

The system process rate is calculated as follows:

$$r_i = \sum_j v_{ij} \rho_j \quad (4.8)$$

All organic constituents are expressed as equivalent amounts of *COD*. Therefore oxygen is expressed as negative *COD*.

**Table 4-1: Matrix representation of models (Henze et al., 2000)**

Component →	<i>i</i>	1	2	3	Process Rate, $\rho_j$ [ML <sup>3</sup> T <sup>-1</sup> ]
<i>j</i> Process ↓		$X_B$	$S_s$	$S_O$	
1 Growth		1	$-\frac{1}{Y}$	$-\frac{1-Y}{Y}$	$-\frac{\hat{\mu}S_s}{K_s + S_s} X_B$
2 Decay		-1		-1	$bX_B$
Observed Conversion Rates ML <sup>-3</sup> T <sup>-1</sup>		$r_i = \sum_j r_{ij} = \sum_j v_{ij} \rho_j$			Kinetic Parameters:
Stoichiometric Parameters: Trow growth yield: <i>Y</i>	Biomass [M(COD)L <sup>-3</sup> ]	Substrate [M(COD)L <sup>-3</sup> ]	Oxygen (negative COD) [M(-COD)L <sup>-3</sup> ]	Maximum specific growth rate: $\hat{\mu}$ Half-velocity constant: $K_s$ Specific decay rate: $b$	

A major advantage of this representation is the easy and fast recognition of the fate of each component. Looking at the rows all components involved in the corresponding process are visible, likewise looking at the columns gives an overview of all processes influencing the corresponding component.

A detailed description of the matrix representation used for the description of the ASM models can be found in Activated Sludge Models ASM1, ASM2, ASM2d and ASM3 (Henze et al., 2000).

### 4.1.3 Activated Sludge Models

In 1982 the International Water Association on Water Pollution Research and Control (IAWPRC) established the Task Group on Mathematical Modelling for Design and Operation of Activated Sludge Processes. The aim was to develop a model with minimum complexity. The result was the Activated Sludge Model No. 1 (ASM1) (Henze et al., 2000). In the following years, further models with different

extensions were developed and published. In addition, many research groups developed their own models based on the ASM models. The following list gives a short overview of the main properties / differences of the official ASM versions published in journals of the IWA:

- Activated Sludge Model No. 1 (ASM1) – The ASM1 considers processes for carbon oxidation, nitrification and denitrification. It was finally published in the IAWPRC Scientific and Technical Report No. 1 (Henze et al., 1986).
- Activated Sludge Model No. 2 (ASM2) – The ASM2 is an extension of the ASM1 with higher complexity due to more components. It was developed out of the need for a model that is able to simulate phosphorus removal. To make this possible, biomass in the ASM2 has a cell internal structure. Furthermore, two chemical processes were added, which are needed to simulate phosphorus precipitation. It was first published 1995 in IAWQ Scientific and Technical Report No. 3 (Gujer et al., 1995)
- Activated Sludge Model No. 2d (ASM2d) – The ASM2d is a minor extension of the ASM2. It extends the ASM2 model by two additional processes which account for the fact that phosphorous accumulating organisms (PAOs) can be considered as two fractions which can use cell internal organic products for the denitrification process. The ASM2d was published 1999 in Water Science and Technology (Henze et al., 1999).
- Activated Sludge Model No. 3 (ASM3) – The ASM3 is the most advanced model and was designed to fulfill the needs of engineering specialists as well as scientists. It combines the core functionalities of many different models in order to eliminate different weaknesses of the ASM1 and to make it possible to connect additional modules such as modules for biological phosphorus removal. A comparison between the ASM1 and the ASM3 model can be found in (Henze, 2000, p. 105). The ASM3 was published in 1999 in Water Science and Technology (Gujer et al., 1999).

#### *4.1.3.1 Choice of model (ASM1)*

The ASM1 model was selected for this work. ASM1 is the first and oldest model in the ASM series of models. Although a series of deficiencies are known, which are described in IWA Scientific Report No. 9 (Henze, 2000, p. 103), they are outweighed by the advantages it offers over the other models for this work.

The model is used in this piece of work for the purpose of optimising the aeration process. The ASM1 model is stable, simple and appropriate for this task. Phosphate is neither taken into account nor required for this task. Furthermore, it is not recorded at the Rospe wastewater treatment plant. In comparison to other models, there are also a low number of parameters and processes in ASM1, which makes it easier to calibrate. Last but not least, the good availability of already published implementations is also one of the arguments for using this model.

#### 4.1.4 Activated Sludge Model No. 1

As described in subsection 4.1.3.1, the ASM1 model was used for the development of the model for the Rospe wastewater treatment plant. For this reason, this subsection describes ASM1 in more detail. ASM1 has a total of eight processes and 13 variables. Accordingly, the state vector  $\mathbf{x}$  is formed as follows:

$$\mathbf{x} = [S_I, S_S; X_I, X_S, X_{B,H}, X_{B,A}, X_P, S_O, S_{NO}, S_{NH}, S_{ND}, X_{ND}, S_{ALK}] \quad (4.9)$$

In general, there is a distinction made between dissolved substances ( $S$ ) and particulate substances ( $X$ ). The model recognises two types of biomass: The heterotrophic biomass ( $X_{B,H}$ ) and the autotrophic biomass ( $X_{B,A}$ ). Organic matter is subdivided here into four fractions: soluble inert ( $S_I$ ), soluble readily biodegradable ( $S_S$ ), particulate inert ( $X_I$ ) and particulate readily biodegradable ( $X_S$ ). Nitrogen is represented by nitrate and nitrite ( $S_{NO}$ ), ammonium and ammonia ( $S_{NH}$ ), soluble biodegradable organic nitrogen ( $S_{ND}$ ) and particulate biodegradable organic nitrogen ( $X_{ND}$ ). In addition there is oxygen ( $S_{NO}$ ), alkalinity ( $S_{ALK}$ ) and, as a intermediary product of the process, particulate products arising from biomass decay ( $X_P$ ). Phosphate is not taken into account in ASM1. Using these variables, the model employs eight processes (Table 4-2).

##### 4.1.4.1 ASM1 Model Parameters

The ASM1 model has a total of 19 parameters, which describe the stoichiometry (5 parameters) and kinetics (14 parameters). Default values for 20°C and 10°C at a neutral pH value have been provided by the IWA Task Group to assist with calibrating the model. These act as a good starting point for a calibration and were used as such in calibration of the Rospe model. However, the range covered by the individual parameters is large. When viewing the parameters in the 10°C column and comparing them to the parameters for 20°C, it is noticeable that only seven of the kinetic parameters are influenced by the change in temperature. These are marked in bold in Table 4-3. The last column of Table 4-3 shows published values from relevant literature (Jeppsson, 1996) for the different parameters. If the published values for the half-saturation coefficient (hsc) for heterotrophs  $K_S$ , which lies between 5 and 225, is taken as an example, it clearly demonstrates how large the range can be.

Table 4-2: ASM1 Processes

No	Name	Process
1	Aerobic growth of heterotrophs	$P_1 = \left( -\frac{S_L}{Y_H} + X_{BH} - \frac{S_O - S_{O-H}}{Y_H} - S_{NH}i_{XB} - S_{ALK}i_{XB} - \frac{i_{XB}}{14} \right) \hat{\mu}_H \left( \frac{S_S}{K_S + S_S} \right) \left( \frac{S_O}{K_{O,H} + S_O} \right) X_{B,H}$ (4.10)
2	Anoxic growth of heterotrophs	$P_2 = \left( \frac{S_S + X_{B,H}}{Y_H} - \frac{S_{NO} - S_{NO-H}}{2.86Y_H} - S_{NH}i_{XB} + \frac{S_{ALK} - S_{ALK}Y_H}{14 \cdot 2.86Y_H} i_{XB} - \frac{i_{XB}}{14} \right) \left( \frac{S_S}{K_S + S_S} \right) \left( \frac{K_{O,H}}{K_{O,H} + S_O} \right) \left( \frac{S_{NO}}{K_{NO} + S_{NO}} \right) \eta_g X_{B,H}$ (4.11)
3	Aerobic growth of autotrophs	$P_3 = \left( X_{B,A} - \frac{S_O}{Y_A} - \frac{S_{O-H}}{Y_A} + \frac{S_{NO}}{Y_A} - S_{NH}i_{XB} - \frac{S_{ALK}i_{XB}}{14} - \frac{S_{ALK}}{7Y_A} \right) \hat{\mu}_A \left( \frac{S_{NH}}{K_{NH} + S_{NH}} \right) \left( \frac{S_O}{K_{O,A} + S_O} \right) X_{B,A}$ (4.12)
4	Decay of heterotrophs	$P_4 = (X_S - X_S f_P - X_{B,H} + X_P f_P + X_{ND}i_{XB} - X_{ND}f_P i_{XP}) (b_H X_{B,H})$ (4.13)
5	Decay of autotrophs	$P_5 = (X_S - X_S f_P - X_{B,H} + X_P f_P + X_{ND}i_{XB} - X_{ND}f_P i_{XP}) (b_A X_{B,A})$ (4.14)
6	Ammonification of soluble organic nitrogen	$P_6 = \left( S_{NH} - S_{ND} - \frac{S_{ALK}}{14} \right) (k_d S_{ND} X_{B,H})$ (4.15)
7	Hydrolysis of entrapped organics	$P_7 = (S_S - X_S) \left( k_b \frac{X_s / X_{B,H}}{K_X + (X_s / X_{B,H})} \right) \left[ \left( \frac{S_O}{K_{NH} + S_{NH}} \right) + \eta_h \left( \frac{K_{O,H}}{K_{O,H} + S_O} \right) \left( \frac{S_{NH}}{K_{O,H} + S_{NO}} \right) \right] X_{B,H}$ (4.16)
8	Hydrolysis of entrapped organic nitrogen	$P_8 = (S_{ND} - X_{ND}) \left( k_b \frac{X_s / X_{B,H}}{K_X + (X_s / X_{B,H})} \right) \left[ \left( \frac{S_O}{K_{NH} + S_{NH}} \right) + \eta_h \left( \frac{K_{O,H}}{K_{O,H} + S_O} \right) \left( \frac{S_{NH}}{K_{O,H} + S_{NO}} \right) \right] X_{B,H} \left( \frac{X_{ND}}{X_S} \right)$ (4.17)

The reaction rates for the individual substances can also be calculated accordingly. The complete Peterson matrix, as described in Table 4-1, can be found in the Appendix.

Table 4-3: ASM1 Default parameter values (Jeppsson, 1996)

Parameters	Symbol	Unit	@10°C	@20°C	literature
<b>Stoichiometric parameters</b>					
Heterotrophic yield	$Y_H$	g cell COD formed (g COD oxidized) <sup>-1</sup>	0.67	0.67	0.38-0.75
Autotrophic yield	$Y_A$	g cell COD formed (g N oxidized) <sup>-1</sup>	0.24	0.24	0.07-0.28
Fraction of biomass yielding particulate products	$f_p$	dimensionless	0.08	0.08	-
Mass N/mass COD in biomass	$i_{XB}$	g N(g COD) <sup>-1</sup> in biomass	0.086	0.086	-
Mass N/mass COD in products from biomass	$i_{XP}$	g N(g COD) <sup>-1</sup> in endogenous mass	0.06	0.06	-
<b>Kinetic parameters</b>					
Heterotrophic max specific growth rate	$\hat{\mu}_H$	day <sup>-1</sup>	<b>6.0</b>	<b>3.0</b>	0.6-13.2
Heterotrophic decay rate	$b_H$	day <sup>-1</sup>	<b>0.62</b>	<b>0.20</b>	0.05-1.6
Half-saturation coefficient (hsc) for heterotrophs	$K_S$	g COD m <sup>-3</sup>	20.0	20.0	5-225
Oxygen hsc for heterotrophs	$K_{O,H}$	g O <sub>2</sub> m <sup>-3</sup>	0.20	0.20	0.01-0.2
Nitrate hsc for denitrifying heterotrophs	$K_{NO}$	g NO <sub>3</sub> -N m <sup>-3</sup>	0.50	0.50	0.1-0.5
Autotrophic max. specific growth rate	$\hat{\mu}_A$	day <sup>-1</sup>	<b>0.80</b>	<b>0.30</b>	0.2-1.0
Autotrophic decay rate	$b_A$	day <sup>-1</sup>	<b>0.2</b>	<b>0.1</b>	0.05-0.2
Oxygen hsc for autotrophs	$K_{O,A}$	g O <sub>2</sub> m <sup>-3</sup>	0.40	0.40	0.4-2.0
Ammonia hsc for autotrophs	$K_{NH}$	g NH <sub>3</sub> -N m <sup>-3</sup>	1.0	1.0	-
Correction factor for anoxic growth of heterotrophs	$\eta_g$	dimensionless	0.80	0.80	0.6-1.0
Ammonification rate	$k_a$	m <sup>3</sup> COD(g day) <sup>-1</sup>	<b>0.08</b>	<b>0.04</b>	-
Max. specific hydrolysis rate	$k_h$	g slowly biodegradable COD (g cell COD day) <sup>-1</sup>	<b>3.0</b>	<b>1.0</b>	-
Hsc for hydrolysis of slowly biodeg. substrate	$K_x$	g slowly biodegradable COD (g cell COD) <sup>-1</sup>	<b>0.03</b>	<b>0.01</b>	-
Correction factor for anoxic hydrolysis	$\eta_h$	dimensionless	0.40	0.40	-

## 4.2 Modelling of the Rospe WWTP

### 4.2.1 Rospe WWTP

The Rospe plant is a municipal WWTP which treats a fraction of the wastewater of the German city of Gummersbach. The plant is designed for 34,000 population equivalents (*PE*) and is currently connected to approximately 18,000 *PE*. The wastewater is mostly municipal wastewater. Approximately 15,900 *PE* of the 18,000 *PE* are wastewater from households. In 2014, the Rospe plant treated 3,537,940  $m^3$  wastewater.

The Rospe plant is continuously operated with upstream de-nitrification. Mechanical purification is carried out in the form of a 6 mm grate followed by a sand filter. Afterwards, the water passes through a horizontal flow primary treatment tank. The biological stage consists of a total of 18 tanks into which the wastewater flows successively. The structure of the plant is laid out in 3 lanes. Finally, there is the final treatment stage that consists of three lanes connected in parallel. Table 4-4 gives an overview of the different volumes.

Table 4-4: WWTP Rospe - Basic Data

Procedural Stage	Size	Unit
Volume Primary Treatment	550	$m^3$
Volume Biology (total)	7,215	$m^3$
Volume Nitrification	3,207	$m^3$
Volume Denitrification	4,008	$m^3$
Volume Final Treatment	5,207	$m^3$

From the volumes in Table 4-4 and the total amounts being treated, it is possible to roughly calculate the following retention times for the individual stages:

- Primary treatment approx. 1.5 hours
- Biology approx. 18 hours
- Final treatment approx. 13 hours

It is important to note here that the plant has different internal flows, such as two recirculation lines (see Figure 4-4) and the return sludge. Table 4-5 shows the statutory effluent limits for the plant.

Table 4-5: WWTP Rospe - Limit / Design Values

Substance / Entity	Value	Unit
Chemical Oxygen Demand (COD)	60	mg/l
Ammonium Nitrate ( $NH_4-N @ T > 6^\circ C$ )	8	mg/l
Total Nitrogen ( $N_{Tot}$ )	18	mg/l
Total Phosphate ( $P_{Tot}$ )	2	mg/l

#### 4.2.1.1 Biological Treatment

Figure 4-3 shows the 18 tanks in the biological stage of the Rospe wastewater treatment plant. Lane one can be seen on the left of the picture, lane two in the middle and lane three on the right.



Figure 4-3: Biological Stages of the Rospe WWTP

However, the distribution of these lanes is misleading because the wastewater flows through all of the tanks in succession. An overview of the precise distribution of the tanks is given in Figure 4-4. The dotted line in Figure 4-4 shows the flow path. With respect to the process technology, the plant utilises upstream denitrification. The first nine tanks are denitrification tanks, followed by eight nitrification tanks. The last tank (Deni 4) is again a denitrification tank. There are two paths for the recirculation (Figure 4-4 blue dotted line) required for upstream denitrification. The outer recirculation (1) leads from the last tank (Deni 4) back to the first tank, while the second recirculation (2) is an additional internal loop. In terms of modelling the plant, a reduction in the complexity to two or three tanks would appear appropriate at first glance. However, this is not possible due to the two recirculation paths. The red (denitrification tanks) and green (nitrification tanks) areas represent the volumes that will be merged together in the model into one tank each respectively. In this manner, it is possible to reduce the model from 18 to 7 tanks without significantly changing the representation of the process technology.

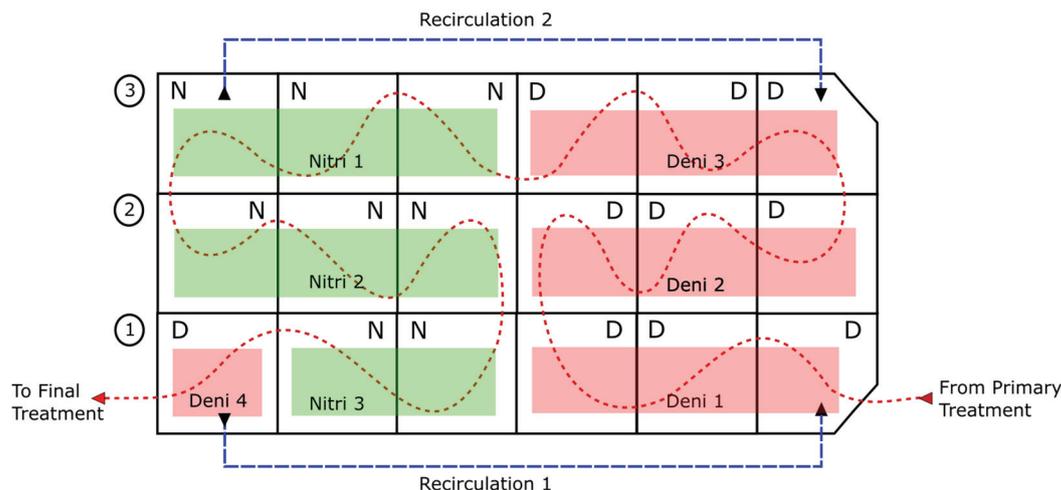


Figure 4-4: Biological Stages of the Rospe WWTP

#### 4.2.1.2 Calculation of recirculation flow

The recirculation volumes at the Rospe plant are neither controlled nor measured. This poses a severe problem for the calibration of the simulation model as the recirculation flow has a strong impact on the plant hydraulics and the respective process behaviour. Thus, the recirculation flow rate needs to be estimated realistically so that it matches the power of the installed pumps as well as the concentration ranges of the process variables. In this case the influence of the recirculation flow on  $NH_4 - N$  and  $NO_3 - N$  concentrations is analysed. The plant has two recirculation pumps of type KSB Amaline P260-501/014. The type description describes a propeller pump with a propeller diameter of 501 mm, a nominal speed of  $260 \text{ min}^{-1}$  and a 4-Pol 1.3 kW asynchronous motor. The pumps for the recirculation are operated via a frequency converter. According to the display, recirculation pump 1 (D1→D4) is operated at  $300 \text{ min}^{-1}$  @ 1.02 kW and recirculation pump 2 (N1→D3) at  $290 \text{ min}^{-1}$  @ 0.92 kW. As the pumps only have a small height difference to overcome, it can be assumed that the counterpressure is primarily created by the recirculation lines themselves. Based on the pump data sheet (KSB, 2011), it is difficult to estimate the flowrate for this output because the setting range is very large. However, flow rates of up to 350 l/s or  $30240 \text{ m}^3/\text{d}$  are possible for the pump speed. In order to identify a realistic flow rate in the two recirculation lines, a grid search with different flow volumes for each pump was carried out. This took into account the realistically achievable flow volumes. Figure 4-5 shows the resulting difference in  $NH_4 - N$  concentration between the measured and simulated concentrations based on the estimated recirculation flow (RMSE). It can be seen that the best value of around  $0.88 \text{ mg/l}$  lies in the upper left section. This is between  $15000 \text{ m}^3/\text{d}$  and  $27000 \text{ m}^3/\text{d}$  for recirculation volume 1. The best result for recirculation volume 2 is between  $35000 \text{ m}^3/\text{d}$  and  $40000 \text{ m}^3/\text{d}$ . These are relatively high values. It can be seen in Figure 4-6 that the optimal value for the lowest  $NO_3 - N$  concentration RMSE is a little lower.

As the  $NH_4 - N$  limit value in the biology is more relevant for wastewater treatment plants, the following values were selected as a compromise:

- recirculation volume 1: 20000 m<sup>3</sup>/d
- recirculation volume 2: 30000 m<sup>3</sup>/d

The higher flow rate in the second recirculation line corresponds to the expectation that there should be a higher flow rate because this line is significantly shorter than the first recirculation line.

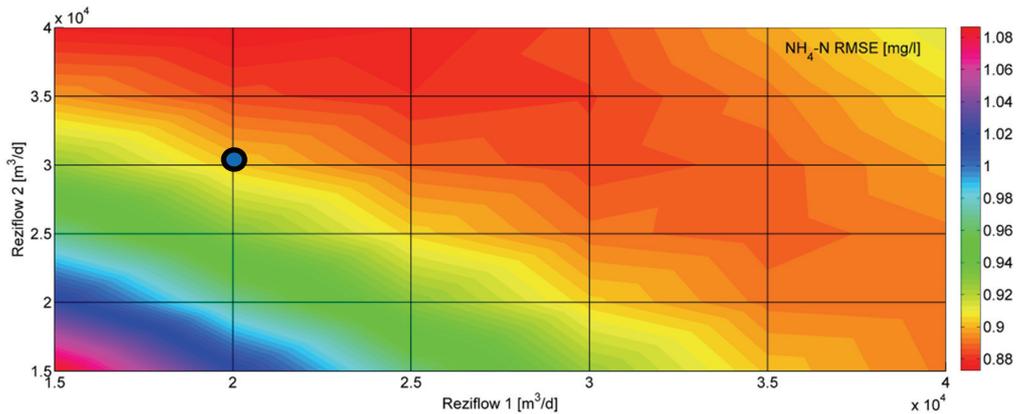


Figure 4-5: Results of a grid search for the optimum recirculation flow based on  $NH_4-N$  RMSE

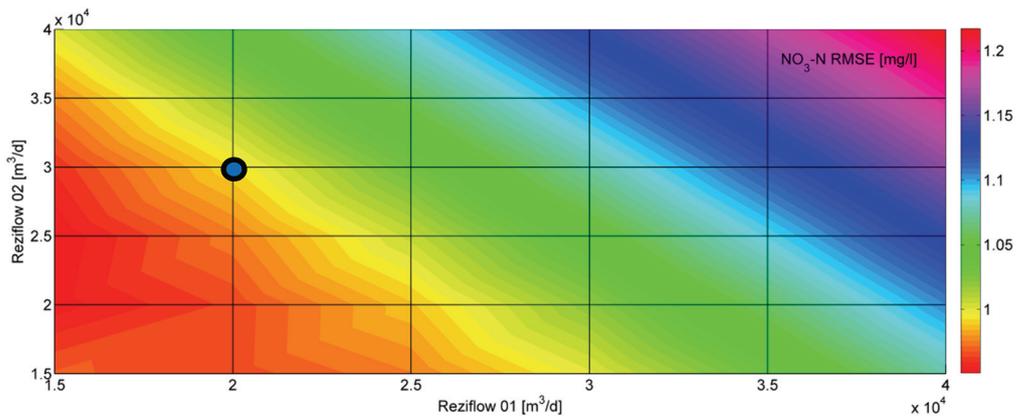


Figure 4-6: Results of a grid search for optimum recirculation flow based to  $NO_3-N$  RMSE

#### 4.2.1.3 Excess Sludge Treatment

The treatment of raw sludge and excess sludge at the Rospe plant is a special feature because the plant does not have a digestion tower. For this reason, a pipeline for the generated sludge has been laid between the Rospe plant and the central wastewater treatment plant at Gummersbach Kruppenohl. All of the excess sludge generated is handled in Kruppenohl. In order to partially balance the system, process water is pumped back to the Rospe plant and added again to the process. This additional load has not been taken into account in the model for a variety of reasons: 1. There are no laboratory samples about the load of

$NH_4 - N$  or  $COD$  in the process water. 2. The process water is added together with the water from the primary treatment underground to the biological stage (Deni1). It is thus not possible to measure the inflow of process water in a practical way.

#### 4.2.1.4 Phosphorus Removal

Chemical precipitation is used in Rospe to remove phosphate. However, the plant does not have any online measurement for phosphate which is why the precipitant is dosed continuously. A description of this process can be found in section 3.2.3.

#### 4.2.1.5 Relevant plant instrumentation

The Rospe plant has two main quantities that are reliably measured,  $NH_4 - N$  (Hach Lange Amtax) (Figure 4-7 A) and  $NO_3 - N$  (Figure 4-7 B), both taken at the effluent from the biological stage. This is supplemented by three oxygen sensors in the effluent from each of the nitrification tanks (Nitri 1, Nitri 2 and Nitri 3). In addition, the plant takes  $TSS$  measurements in the effluent from the biological stage. These measurements are the most important for the model because they are required for calibration.

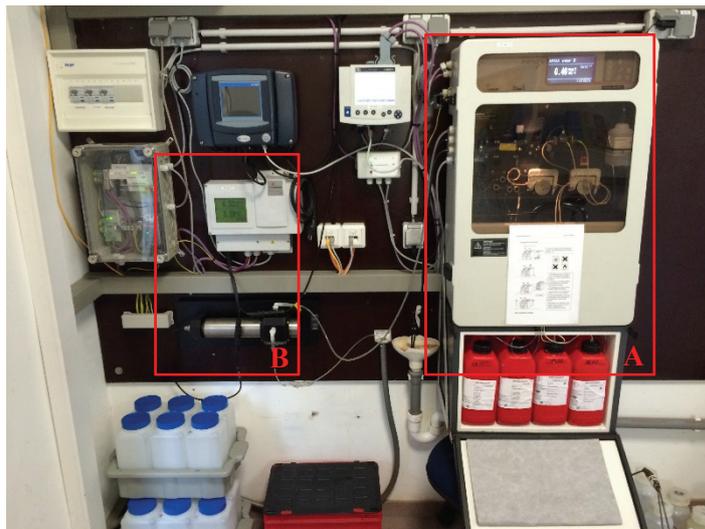


Figure 4-7: Analysers at Rospe WWTP A: Hach-Lange AMTAX and B: Hach-Lange NITRATAX

The plant also possesses other measurement technology but this is largely utilised for monitoring purposes. The  $PO_4 - P$  value and the  $COD$  value are not recorded.

#### 4.2.1.6 Control Loops

The main control loops at the Rospe wastewater treatment plant are:

- Oxygen controller
- Return sludge or  $TSS$  controller

Both controllers in the model differ from those installed in the real plant. The controllers in the real plant are fuzzy controllers that are set individually. As the aim of this work is not to simulate the fuzzy controllers but rather to model the process conditions, they were replaced in the model by PI controllers which use the actually measured values ( $O_2$  and  $TSS$ ) during the measurement campaign as setpoints.

The real oxygen controller consists of a fuzzy controller that uses the three measured oxygen values from the tanks Nitri 1, Nitri 2 and Nitri 3 and  $NH_4 - N$  and  $NO_3 - N$  values from the effluent of the biological stage as inputs.

The  $TSS$  controller or the controller for the volume of return sludge also takes the form of a fuzzy controller. In the model, it was implemented as a PI controller with a fixed setpoint because the  $TSS$  content at the plant is almost stable at 3.5 g/l. Excess sludge is discharged at a fixed ratio to the return sludge.

As the plant does not have an inductive flow measurement system in the recirculation line, it is operated with fixed values. In practice, the recirculation volume fluctuates to a certain extent because the outlets from the recirculation lines are installed a few centimetres below the surface of the water and thus the water level in the tanks can influence the counterpressure in the recirculation lines. As these effects are in the single-digit percentage range, they have not been considered further in this work.

#### 4.2.2 WWTP Model

The Simba<sup>®</sup> toolbox from the company ifak in the software Matlab<sup>®</sup> / Simulink was utilised for implementing the treatment plant model.

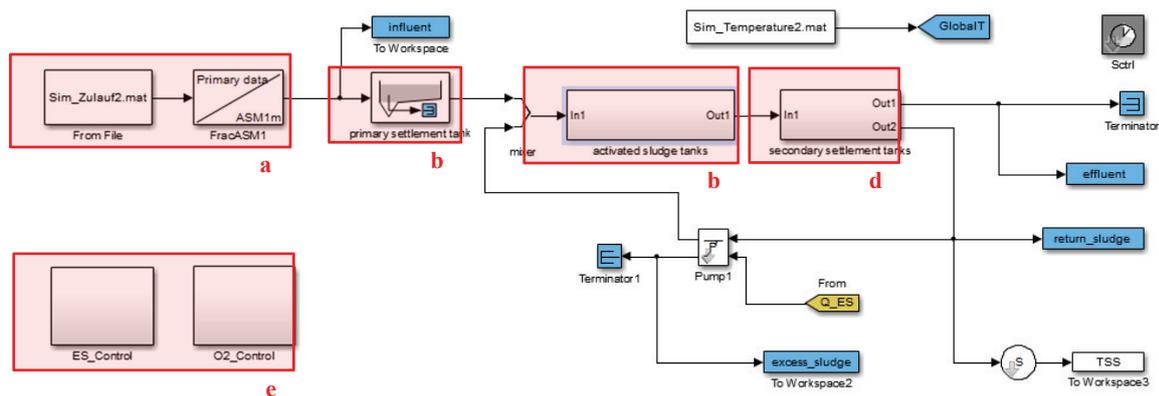


Figure 4-8: Rospe Model Overview

Figure 4-8 shows an overview of the model. The model is split into five main areas a) inflow and inflow fractioning, b) primary treatment, c) biological wastewater treatment, d) final treatment and e) control loops. The following subsections describe the model and highlight the differences to the implementation in the real wastewater treatment plant.

#### 4.2.2.1 Model of the inflow and inflow fractioning

The data from a measurement campaign at the Rospe treatment plant in 2012 was used as the inflow data for the model. For this purpose, additional measurement devices for  $NH_4 - N$ ,  $COD$ , turbidity and a variety of other parameters were installed. After eliminating any incorrect or incomplete data, an inflow set covering 17 days was created. A detailed description of the installation and the measurement campaign can be found in section 5.2 and section 5.3.

As there were no precise investigations into the individual  $COD$  fractions, standard values for inflow fractioning in municipal wastewater treatment plants were used (see Figure 4-9):

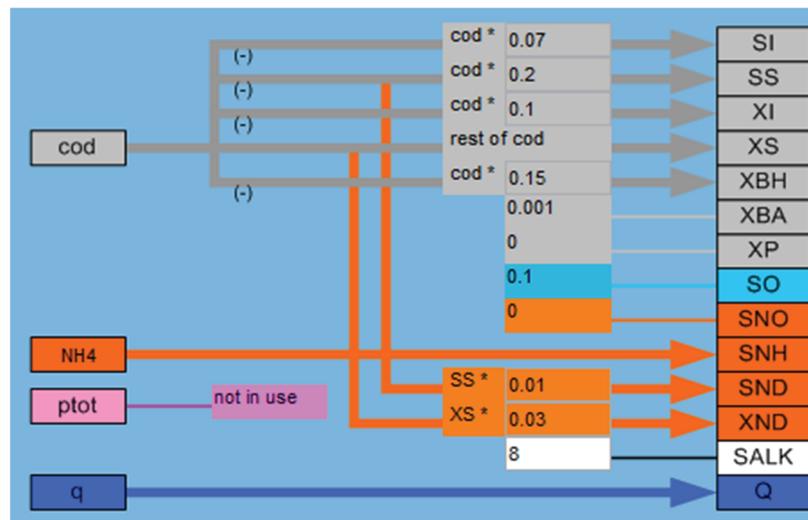


Figure 4-9: Inflow Fractioning

For this fractioning data, it is assumed that the nitrogen is completely present in the form of  $NH_4 - N$ . The  $COD$  is primarily present as particulate readily biodegradable  $COD$  ( $X_S$ ). A further 15% of the  $COD$  is heterotrophic biomass ( $X_{B,H}$ ) and 20% is soluble readily biodegradable  $COD$  ( $S_S$ ).

#### 4.2.2.2 Primary and Secondary Clarification

For the primary treatment an adaptation of the simple primary clarifier without biological processes developed by Otterpohl (Otterpohl and Freund, 1992) and part of the Simba toolbox was chosen. The model considers the clarifier as a completely mixed reactor, which models the buffering behaviour of a clarifier. The elimination of particulate fractions of  $COD$  and  $N$  is calculated using the hydraulic retention time and the proportion of particulate fractions in the  $COD$ . The primary clarification tank of the Rospe plant has a volume of  $V = 550.2m^3$  and the ratio of particulate  $COD$  to total  $COD$  is:

$$f_x = \frac{COD_{part.}}{COD_{tot.}} = 0.7 \quad (4.18)$$



### 4.2.2.3 Secondary Treatment

Figure 4-11 shows the model for the biological stage. In contrast to the real plant with 18 tanks, the model only comprises 7 tanks. A detailed explanation for this simplification can be found in subsection 4.2.1.1.

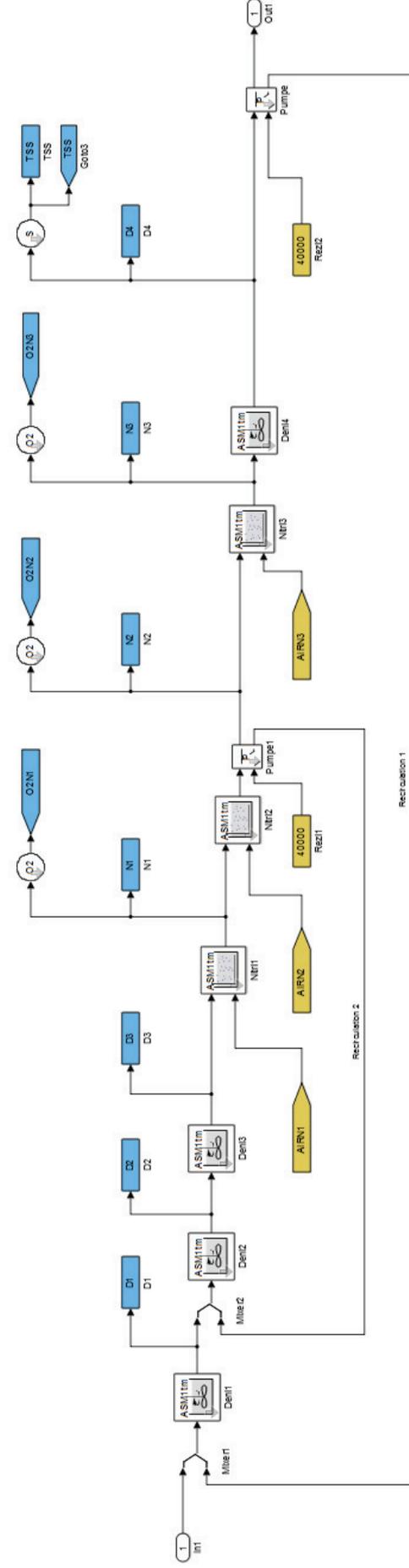


Figure 4-11: Secondary Treatment / Biological Treatment

The secondary treatment tanks consist of three denitrification tanks followed by three nitrification tanks and a final denitrification tank at the outflow from the biological treatment. Models based on the ASM1 model were selected for the nitrification and denitrification tanks. These represent fully mixed reactors. The ASM1 models used were slightly modified by the manufacturer and are thus designated as ASM1m. A relevant change relates to situations in which all ammonium is nitrified or stored while there is a low load in the wastewater treatment plant. The original ASM1 model has no limit in this case, which can lead to concentrations below zero. The analysis was carried out on a case-by-case basis for process 1 (aerobic growth of heterotrophs) and process 2 (anoxic growth of heterotrophs). A detailed description of the modifications including the resulting process equations can be found in the SIMBA Users Guide (ifak system GmbH, 2009, pp. 120–123).

The sizes of the tanks are taken from the sum of the individual tanks in the real plant and are shown in Table 4-7.

**Table 4-7: Denitrification and nitrification tank volumes**

Deni-Tanks	m <sup>3</sup>	Nitri-tanks	m <sup>3</sup>
Deni 1	1129	Nitri 1	1288
Deni 2	1094	Nitri 2	1288
Deni 3	1129	Nitri 3	855
Deni 4	437		

The specific parameters required for the nitrification tanks were selected as follows: Oxygen transfer rate  $\alpha = 0.6$ , specific oxygen input  $R_{air} = 16 [g/(m^3 \cdot m)]$ , oxygen saturation concentration  $DO_{sat.} = 8.637 [g/m^3]$ , and immersion depth  $d_{im.} = 5.8 [m]$ . The modelling of aeration control is described in subsection 4.2.1.6. This consists of PI controllers, which simulate the real  $O_2$  concentrations measured during the measurement campaign. The precise process conditions during the campaign are thus reproduced. For simulations outside of this time period, fixed setpoints for the controllers were adopted that correspond to standard operating values. This was the case, for example, during the ramping up of the model when fixed inflow values were used.

### 4.3 Model Calibration

A numeric model adapted to the data measured during the field studies was used for calibration or adjustment of the model. Individual calibration processes or protocols were defined for many of the models used. The IWA Task Group on Good Modelling Practice (GMP) worked on the development of a standard protocol for model calibration, based on their results they published the Unified Protocol (GILLOT, 2012). The correct design of the model and its calibration go hand in hand because diverse parameters must be set in both. Therefore, selecting the correct dimensions for the internal flows in the model is actually part of the calibration but this process could also be part of the design of the model. The calibration of the Rospe model presented in the following sections relates only to the settings for the biological and biochemical processes. The dimensions for the internal flows were already presented in section 4.2.1.2.

#### 4.3.1 Calibration Prerequisites

In general, the prerequisite for a calibration is a correctly designed model based on process technology and the sufficient availability of measured process values. The following list provides an overview of the most important prerequisites for successful model calibration:

- Complete representation of the process technology
  - Tank sizes
  - Lines
  - Pumps
  - Internal flows
  - Control loops
  - Disturbance variables (e.g. process water)
- Sufficient availability of process measurement technology:
  - Oxygen ( $O_2$ )
  - Ammonium ( $NH_4 - N$ )
  - Nitrate ( $NO_3 - N$ )
- Availability of inflow measurement values (measurement campaign) (based on ASM1)
  - Chemical Oxygen Demand ( $COD$ ) (Inflow)
  - Ammonium ( $NH_4 - N$ ) (Inflow)
  - Flow ( $Q$ )
  - Temperature

Even modelling the process technology is a challenge for many plants, as can be seen in the example of the recirculation volumes and the dosing of process water at the Rospe wastewater treatment plant (section 4.2.1.2). In addition, other measurements in the individual tanks, as well as laboratory experiments to identify the individual fractions, would be desirable. However, these measurements are complex and costly to implement in practice and are thus not generally available.

#### 4.3.2 Model Initialization

In order to adapt the internal states of the model that represent the biology to the real circumstances, the model is initialised in two steps. In the first step, the model is ramped up (run-in) for several hundred days at constant inflow values, normally using averages from real measurement values. The *TSS* builds up and the internal states become adapted during this period. In practice, this represents the development of the “biology”. In the case of the Rospe model, 300 days were selected until the internal states were stable. There was then a second phase in which the model was further ramped up using dynamic inflow values. In this phase, the internal states adapt further. For the Rospe model, the data recorded by the measurement technology over 17 days was used a total of 10 times to create a simulation period of 170 days. One difficulty is that the achievement of a stable state is dependent on the model parameters. If they deviate from the real biology, the stable state of the model deviates from reality. In order to overcome this problem, it is necessary for the state vector to be considered to assess whether the individual fractions are realistically represented. At the same time, the RMSE between the measured

and simulated process values can also be used. The first adjustments to the parameters are then carried out on this basis. Next, the model is stabilised further based on the adjusted parameters. This process is repeated until a satisfactory stable state is achieved. This is therefore an iterative process. Based on this starting state, the fine tuning of the parameters can now be carried out. An overview of this process is shown in Figure 4-12:

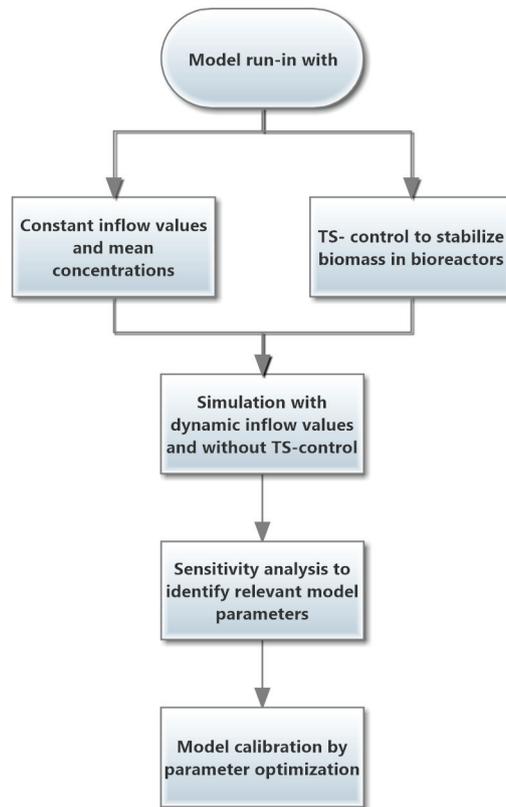


Figure 4-12: Applied calibration procedure

### 4.3.3 Parameter optimization

The ASM1 Model features 19 parameters (see section 4.1.2.1 and section 4.1.4.1), which can be adjusted during the calibration. The modified version used (see section 4.2.2.3 and ifak system GmbH, 2009) was supplemented by an additional 6 parameters. In order to limit the search space, only the five most important parameters that have a relevant influence on the simulation model were selected. The parameter optimization was carried out in two ways. Firstly, a manual adjustment of the parameters was carried out. The RMSE between the measured and simulated  $NH_4 - N$  ( $S_{NH}$ ) time series in tank D4 and the RMSE between the measured and simulated  $NO_3 - N$  ( $S_{NO}$ ) time series in tank D4 were used as a fitness function – which was possible due to the available measurement technology.

#### 4.3.3.1 *Manual parameter optimization and parameter selection*

While the fully automated optimisation only considers the error values, the manual parameter optimization also allows the precise consideration of the state vector and the adjustment of parameters in line with knowledge about the processes. However, this is a time-consuming iterative process due to the complexity of the model, where each individual parameter must be adjusted one at a time. The influence or sensitivity of each of the individual parameters can thus be assessed correspondingly. On the basis of relevant published literature, such as the ASM1 Report (Henze, 2000), it is however possible to limit the number of relevant parameters. The following parameters were thus selected for the Rospe model:

- $k_a$  which describes the release of ammonium from soluble organic nitrogen and is therefore crucial to determine ammonium production correctly (Henze, 2000, p. 22).
- $\mu_A$  which is according to the ASM1 report (Henze, 2000, p. 19) described as the most critical parameter characterizing the growth of the autotrophic biomass.
- $b_H$  which is according to the ASM1 report (Henze, 2000, p. 19) very important for the prediction of sludge production and oxygen requirements.
- $K_{NH}$  which provides information on the relationship between the specific nitrification rate and the pseudo-steady state ammonium nitrogen concentration (Henze, 2000, p. 19).
- $K_S$  which describes biomass growth, but is not a critical parameter as the model is not very sensitive to it (Henze, 2000, p. 20). Nevertheless, it was included to allow for optimized biomass growth kinetics.

#### 4.3.3.2 *Automated parameter optimization using SMS-EGO algorithm*

The S-Metric-Selection-based Efficient Global Optimization (SMS-EGO) algorithm was selected for the automated optimization because it can use multiple fitness functions at the same time. Therefore, the  $NH_4 - N$  RMSE and the  $NO_3 - N$  RMSE were each minimized. For a better understanding, the algorithm will be briefly described in the following section.

#### **S-Metric-Selection-based Efficient Global Optimization (SMS-EGO)**

The optimization of highly complex nonlinear and thus, computationally intensive problems requires a high number of fitness function evaluations (Simpson et al., 2001, pp. 129–150). In the case of real world problems, these fitness functions are often represented by dynamic simulation models consisting of an ODE system, which increases evaluation time significantly. Furthermore, most optimization problems have more than one objective at a time, which creates the need for multi-objective optimization algorithms (MOOA). This poses two additional challenges: (1) The evaluation of several fitness functions  $f$  with underlying simulation models is even more time consuming as the computation time

grows linearly with  $f$  ; (2) A fair assessment of the different fitness functions during each optimization run needs to be guaranteed, i.e. by equally distributing the optimization solutions along a Pareto Front.

SMS-EGO addresses these two problems by using Kriging surrogate models (Matheron, 1963, pp. 1246–1266) to allow for faster fitness function evaluation and by using the  $S$ -Metric (Zitzler and Thiele, 1998, pp. 292–301) to look for promising solutions along the Pareto Front, calculating hypervolumes of fitness values from  $f_1$  to  $f_N$ .

The use of Kriging models for optimization purposes is well established as only a limited number of fitness function evaluations  $y = f(\mathbf{X})$  for the design sites  $x_1 \dots x_n$  are required to get a reasonable approximation  $\hat{y} = \mu + \varepsilon(\mathbf{X})$  where  $\mu$  denotes the mean of all design sites  $\mathbf{X}$  and  $\varepsilon$  the error which is determined by the covariance and the process variance as described by Matheron in 1963 (Matheron, 1963, pp. 1246–1266). While a fast computation of a Kriging model is possible for a low number of design sites ( $< 100$ ), computation time increases exponentially with the number of design sites  $O(N^n)$ . Therefore, the design sites to update the Kriging model need to be chosen carefully.

In order to achieve a high surrogate model accuracy where it is most needed, the Kriging model is updated calculating the  $S$ -Metric along the Pareto Front for the optimal solutions generated by each run of the MOOA. The principle of the  $S$ -Metric is to calculate the hypervolume of each potential optimal solution  $x_i^*$  which is spanned by the corresponding fitness values  $\hat{y}_{i,1} \dots \hat{y}_{i,N}$  and the distance to the fitness values of the surrounding solutions  $x_{i-1}^*$  and  $x_{i+1}^*$ . Thus, greater hypervolumes represent solutions that are more balanced with regard to the  $N$  optimization objectives and that are more widely spread over the Pareto Front, whereas small hypervolumes represent biased solutions that tend to optimize one objective more than the other or solutions that are very close to other optimal solutions.

The idea behind SMS-EGO is to use that information on the potential solutions and choose the ones associated with a high hypervolume to update the Kriging model while neglecting the others with a low hypervolume. Consequences are that the optimization algorithm has a strong focus on balanced solutions along the Pareto Front and that extreme solutions favoring one objective over the others are not considered. Due to the wide spread of the solutions over the interesting area of the Pareto Front by the  $S$ -Metric, the possibility of a local search is minimized. Nevertheless, this evaluation criterion narrows the search area considerably. In order to compensate for this effect, Ponnweiser et al. used additional penalties to compensate that effect (Ponweiser et al., 2008, pp. 784–794).

### SMS-EGO parameter setting and application

Table 4-8 shows the selected parameters for the SMS-EGO algorithm. The upper bound (UB) and lower bound (LB) describe the relevant limits within which the algorithm is allowed to set the parameters. The budget for the initial design describes the size of the initial population. The maximum budget of objective evaluations is the maximum number of simulation runs.

Table 4-8: SMS-EGO Parameters

Parameter	Value
Budget for the initial design (11*dim-1)	54
Maximum budget of objective evaluations	70
Type of smoothing using	none
Lower Bound (LB) [ $\mu_A, k_a, b_H, K_{NH}, K_S$ ]	[0.3, 0.8, 0.2, 1.0, 20]
Upper Bound (UB) [ $\mu_A, k_a, b_H, K_{NH}, K_S$ ]	[1.0, 0.04, 1.6, 2.0, 225]

The fitness function  $f$  utilises the initialised Rospe model run for the time period of 17 days for which inflow values are available. The goal of the optimization is to minimize the two RMSE values:

$$f_1 = \min \left( \frac{1}{T} \sum_{t=0 \text{ days}}^{t=17 \text{ days}} \sqrt{(S_{NHm}(t) - S_{NHs}(t))^2} \right) \quad (4.19)$$

$$f_2 = \min \left( \frac{1}{T} \sum_{t=0 \text{ days}}^{t=17 \text{ days}} \sqrt{(S_{NOm}(t) - S_{NOs}(t))^2} \right) \quad (4.20)$$

## 4.4 Results and Discussion

Table 4-9 gives an overview of the most important optimization results. The best results of an RMSE of 0.85 mg/l for  $NH_4 - N$  and 0.98 mg/l for  $NO_3 - N$  were achieved with SMS-EGO. However, these were only a little better than the manually optimized values of 0.9031 for  $NH_4 - N$  and 0.9924 mg/l for  $NO_3 - N$ . The high RMSE value when using the default parameters was particularly noteworthy. The results for the default parameters are thus only given for reference purposes.

Table 4-9: RMSE comparison of different model parameter sets

Parameter set	$\mu_A, k_d, b_H, K_{NH}, K_S$	NH <sub>4</sub> -N RMSE [mg/l]	NO <sub>3</sub> -N RMSE [mg/l]
Default	0.8, 0.08, 0.62, 1.0, 20.0	1.8625	6.7556
Manually	1.0, 0.8, 1.42, 1.65, 216.0	0.9031	0.9924
SMS-EGO	0.918, 0.04, 1.599, 1.146, 156.76	0.8512	0.9776

When examining Figure 4-13, it can be seen that the simulated values were less dynamic than the measured values in all cases. If the results of the manual and automatic optimizations are considered, it is noticeable that the NH<sub>4</sub> – N peaks, in particular, are not accurately modelled.

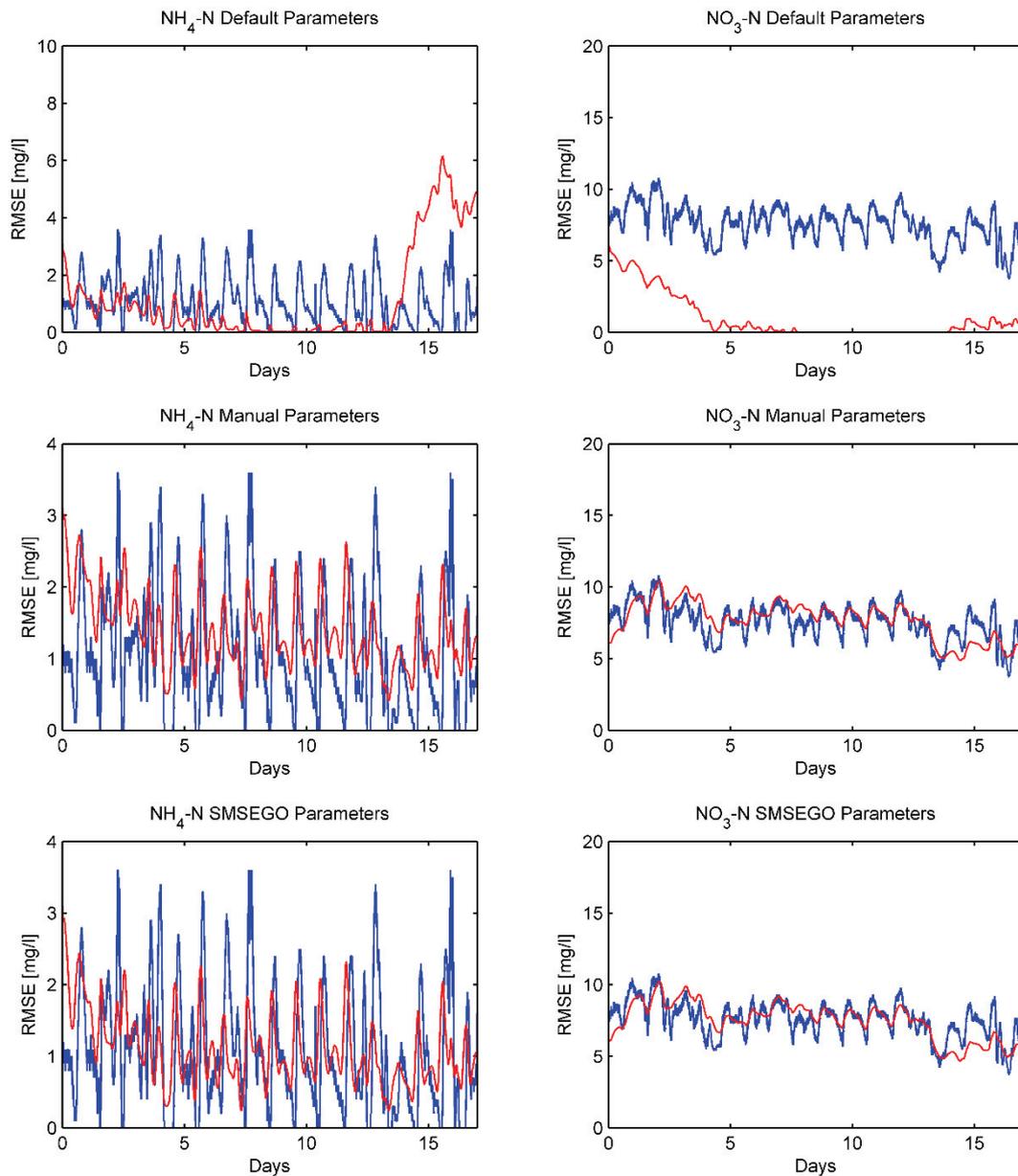


Figure 4-13: Comparison of different optimization results (blue: measured, red: simulated)

SMS-EGO not only calculates an optimal value but also a set of optimal results according to the Pareto principle, i.e. where for each result, there will be no other result that would not be worse in at least one of the optimization criteria. These results are illustrated in Figure 4-14 in the form of a Pareto Front.

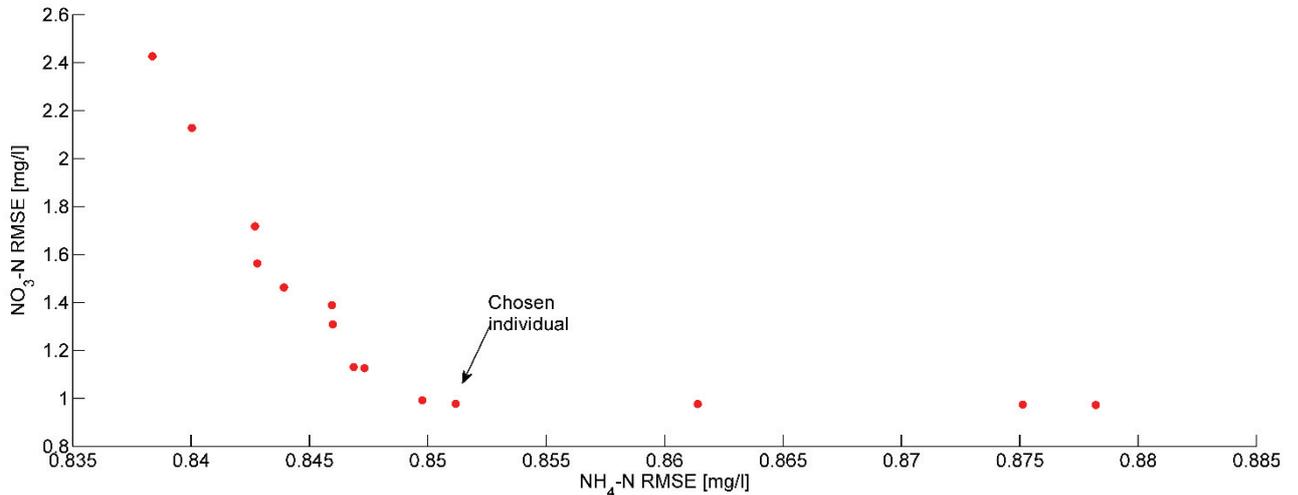


Figure 4-14: Parameter optimization Pareto Front

An individual parameter or set of parameters was selected that had the smallest possible  $NO_3 - N$  RMSE value.

#### 4.4.1 Conclusion and consequences for further use

The results generated and the model created are applied further in the following chapters. Although the design of the model from the completion of the measurement campaign through learning the processes in the real plant and modelling the process technology through to the parameter optimization was completed with the maximum reasonable effort and the greatest of care, the simulated results could only approximate the real conditions to a certain degree. In the case of the Rospe plant, this was especially noticeable in the dynamics of the  $NH_4 - N$  values. There are a diverse range of reasons for this problem: the measurement errors across all of the measurement devices used accumulate, the plant was only subjected to a low load so that the  $NH_4 - N$  value was often at the lower limit of the measurement range, etc. and various internal flows could only be estimated. In order to improve these results further, it is thus necessary to once again significantly increase the work involved. This would mean completing complex laboratory experiments, using additional online measurement devices and would result in high additional costs. However, this would go beyond the scope and justifiable budget available for this piece of work. The extent to which the results influence the model-based inflow estimation and what this means for its practical application will be discussed in the next chapter.

## 5 Virtual COD and NH<sub>4</sub>-N inflow measurements

### 5.1 Introduction

When considering the water purification process of a WWTP, the chemical oxygen demand (*COD*) and ammonium (*NH<sub>4</sub> - N*) concentration, in combination with phosphate (*PO<sub>4</sub> - P*) are the most important substances in the inflow of municipal wastewater treatment plants (WWTPs) to assess the degree of pollution. Therefore, inline measurement systems for *COD* and *NH<sub>4</sub> - N* are needed as they are not only able to provide sufficiently accurate (+- 10 to 20 %) and timely information about the incoming wastewater stream, but also offer many possibilities for advanced optimization and control strategies. Nevertheless, existing inline measurement systems are expensive and come with high maintenance costs, so that their application in small and medium-sized WWTP is mostly not financially feasible. Thus, the need for cost-efficient but also robust monitoring solutions is high. This chapter addresses this need and presents a virtual *COD* and *NH<sub>4</sub> - N* monitoring solution using Machine Learning to estimate the two process variables based on existing and widely available inflow inline measurement systems.

The measurement of *NH<sub>4</sub> - N* and *COD* concentrations in the inflow of WWTPs provides several advantages:

- (1) It is possible to optimize the control strategy for WWTPs by knowing in advance the amount of *NH<sub>4</sub> - N* and *COD* that flows into the bioreactors. In particular, it is possible to increase the oxygen (*O<sub>2</sub>*) concentration, when a load peak is detected in the inflow giving the nitrification more time to treat ammonium.
- (2) It enables a comparison of inflow and effluent and hence, the evaluation of plant efficiency.
- (3) To calibrate WWTP simulation models based on, for example, the activated sludge model (ASM) (Henze et al., 2000), it is essential to have measurements of these process variables in the WWTP inflow.

In guideline DWA-M 256-1 (DWA, 2011) the DWA<sup>2</sup> classifies WWTP measurement devices in three classes:

- A: essential
- B: expedient
- C: not expedient

For the WWTP inflow only the benefits of pH-value, conductivity and temperature are evaluated. The fact, that other measurements are not considered by the DWA working group KA-13.3 (process

---

<sup>2</sup> DWA – German Water Association

measurement systems for wastewater treatment plants) reflects that today only these measurements are considered to be standard in the inflow of most WWTPs. Although *COD* and *NH<sub>4</sub>-N* are important for inflow monitoring and aeration control purposes, they are rarely used. Furthermore, the working group categorized pH-value as class A, conductivity as class B and temperature as class C. While this might be a reasonable categorization for simple control systems, there is more than meets the eye that can be learned from C-type measurements such as temperature. For example, a sudden drop in temperature is a typical sign of a rain event (at least in colder regions), which can be detected and used by sophisticated state-of-the-art control systems to detect high inflow volumes and peak loads of *COD* and *NH<sub>4</sub>-N*. This shows that measurement values which are allegedly not expedient can be used by intelligent control systems to improve process efficiency. Nevertheless, one of the consequences of such DWA recommendations is also that most plants in Germany are not equipped with probes for *NH<sub>4</sub>-N* and *COD*. Together with the high maintenance effort and high prices for the measurement systems, which, depending on the operating principle of the measurement systems, are between €3,500 for ion-selective<sup>3</sup> probes and €20,000 for chemical analyzers<sup>4</sup> for *NH<sub>4</sub>-N*, as well as between €10,000<sup>5</sup> and €25,000<sup>6</sup> for *COD*, installation of these systems is often not financially feasible. Another reason is the harsh environment in the inflow. Raw Inflow Water can be characterized by the following attributes:

- strong variations in concentration
- strong temperature variations
- high fat / grease content
- high flow rates
- variations in *pH-value*

These facts lead to significantly higher maintenance costs in comparison to operation within the bioreactors. The probes have to be cleaned, calibrated and maintained in relatively short intervals (Graner et al., 2005). Furthermore, ion-selective membranes tend to wear out much faster in the inflow of WWTP than in bioreactors.

Therefore, the development of a new method to estimate important inflow concentrations based on existing inline-measurement systems using Machine Learning offers a solution to many if not to all issues with existing, highly complex and costly inline measurement systems.

The main contributions of this chapter are:

---

<sup>3</sup> WTW VARiON® Plus 700 IQ ca. €3500

<sup>4</sup> Hach-Lange AMTAX €19000

<sup>5</sup> Trios – Probs €11.500

<sup>6</sup> S::CAN spectro::lyser ca. €22.000

- (1) A measurement campaign in the inflow of the Rospe WWTP for two months to gather real-time *COD* and *NH<sub>4</sub>-N* concentrations.
- (2) Development of Matlab functions for fully automated preprocessing of measurement data.
- (3) Calibration, application and performance assessment of eight Machine Learning methods for the estimation of *COD* and *NH<sub>4</sub>-N* in the WWTP inflow.
- (4) Determination of the optimal combination and financially reasonable combination of probes to estimate the inflow variables *COD* and *NH<sub>4</sub>-N* sufficiently accurately.
- (5) Development of a software estimator for *COD* and *NH<sub>4</sub>-N* that can replace costly and high maintenance inline measurement systems using different regression and classification methods.
- (6) Development and evaluation of a method to compare classification and regression results for inflow measurements.

The remainder of this chapter is organized as follows: Section 2 describes the probes used for the measurement campaign. Section 3 gives a detailed overview of the measurement campaign and the related data. This includes data preparation (outlier detection, sensitivity analysis, class division and generation of datasets) necessary for the investigation. Section 4 gives an overview of the applied mathematical methods including the determination of the related optimal parameters for each method and dataset. In section 5 the results of the different regression and classification methods are presented. The final section concludes which method gives the best prediction of each inflow variable and which combination of probes is practically manageable and financially reasonable.

## 5.2 Installed measurement probes

To carry out this study, a set of measurement probes were installed at Rospe WWTP operated by the Aggerverband<sup>7</sup> and a measurement campaign over two months was conducted in 2012.

The standard instrumentation in the inflow of the Rospe WWTP is a magnetic induction flow meter (MID), a pH-probe, a conductivity sensor and a temperature sensor. In addition to this, the following measurement probes for *COD*, *NH<sub>4</sub>-N* and turbidity were installed in the intake canal:

- ISEmax CAS40D (ion-selective measurement probe) (Figure 5-1 (a))
- Turbimax CUS51D (turbidity measurement probe 860 nm) (Figure 5-1 (b))
- STIP-scan (spectrometric measurement probe 200 nm-680 nm) (Figure 5-1 (c))
- Trios ProPS-WW (spectrometric measurement probe 190 nm-360 nm) (Figure 5-1 (d))

---

<sup>7</sup> Aggerverband - A local water association managing the water-related tasks of the Agger river basin

ISEmax CAS40D is an ion-selective measurement probe for ammonium and nitrate produced by Endress+Hauser. For the measurement campaign it is solely equipped with electrodes for ammonium measurement. It is able to measure ammonium between 0.1 mg/l and 1000 mg/l. To compensate a cross sensitivity between ammonium-ions ( $NH_4^+$ ) and potassium-ions ( $K^+$ ), the probe is additionally equipped with an electrode for potassium. Figure 5-1 (a) shows a picture of the CAS40D probe.

The Turbimax CUS51D is an immersion sensor produced by Endress+Hauser which measures turbidity as well as solid contents. The optics integrated in the sensor head allow measurements according to three different measurement principles namely the 90° scattered light, 135° back-scattered light and four-beam pulsed-light method (Endress+Hauser, 2013a). Furthermore, the sensor uses a wavelength of  $\lambda=860\text{ nm}$  to avoid interference with diluted light absorbing substances. This measurement principle conforms to the German industry norm for turbidity measurements (DIN EN ISO 7027) (Wasserbeschaffenheit: Bestimmung der Trübung, 2000). The unit of the measurement values is formazin nephelometric units (FNU) (Mike Sadar, 2004; Oregon Water Science Center, 2013) and is factory calibrated to 0-4,000 FNU. Figure 1(b) shows a picture of the CUS51D probe.

The Endress+Hauser STIP-scan CAM74/CAS74 is a spectral measurement probe for nitrate, *CODeq*, *BODeq*, *TOCeq*, *SAC*, total solids, sludge volume, sludge index and turbidity (*ATU*). It measures the absorption spectrum between 200 nm and 680 nm using a xenon lamp as light source (Endress+Hauser, 2013). For the measurement campaign it is used as a reference measurement for *COD*. Furthermore, it is used to measure the spectral absorption coefficient at 254 nm (*SAC<sub>254</sub>*), *SAC<sub>433</sub>* and turbidity in *ATU*. Figure 5-1(c) shows a picture of the CAM74/CAS74.

The Trios ProPS-WW is a spectrometric measurement probe and is comparable to the E&H STIP-scan. Instead of the xenon lamp it uses a deuterium lamp and measures the absorption spectrum between 190 nm and 360 nm. It is also able to measure several parameters including *COD*, nitrate, turbidity and *SAC<sub>254</sub>*. For the measurement campaign it was used as a backup measurement for *COD*. Figure 5-1(d) shows a picture of the ProPS-WW probe.



Figure 5-1: Installed measurement probes: (a) ion-selective ammonium sensor, (b) Turbimax CUS51D turbidity probe, (c) E+H StipScan spectrometric sensor, (d) Trios UV/vis spectroscopic COD/NO<sub>3</sub> analyzer

### 5.3 Measurement Campaign at Rospe WWTP

#### 5.3.1 Data Description

The analysis is based on measurement values recorded in March 2012. Measurements were conducted every three minutes. The full dataset consisted of 14,380 samples; after data cleansing 9,843 samples remained. This resulted in approximately three weeks of usable data. Due to the fact that different sampling intervals were used by data loggers, it was necessary to use interpolation to fit the data to the same time grid. Figure 5-4 and Table 5-1 give an overview of the most important characteristics of the recorded data. Note that there are three data gaps in the data: 22<sup>th</sup>-27<sup>th</sup> February; 5<sup>th</sup>-7<sup>th</sup> March; and 13<sup>th</sup>-

14<sup>th</sup> March. At these limited periods one or more important measurement devices malfunctioned. For this reason those segments of data were unusable and omitted. Overall the data exhibits cyclic daily variations, but has no strong long term tendency or trend. Turbidity appears twice in the data sets. The first measurement is from the Turbimax CUS51D turbidity probe and measures reflected light at 860 nm. The unit is FNU as described in section 5.2. The second measurement is taken by the STIP-scan CAM74/CAS74 spectrometric probe. It measures in ATU (Absorptiometric Turbidity Units), which means it uses transmitted light rather than reflected light. For the measurement campaign the ATU probe was not calibrated, which leads to significant differences in the absolute values of the CUS51D and the Stip-scan probe. Due to the fact that these values are used as input values for regression and classification models the absolute values are not important. For comparison purposes both were scaled to have zero mean and standard deviation of 1. Figure 5-2 and Figure 5-3 show a comparison of the scaled FNU and ATU values.

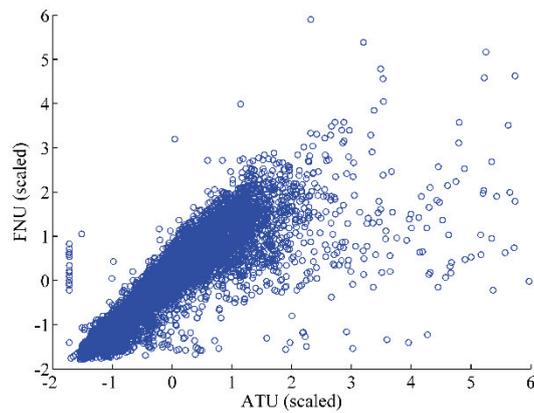


Figure 5-2: Scatterplot ATU vs. FNU (scaled)

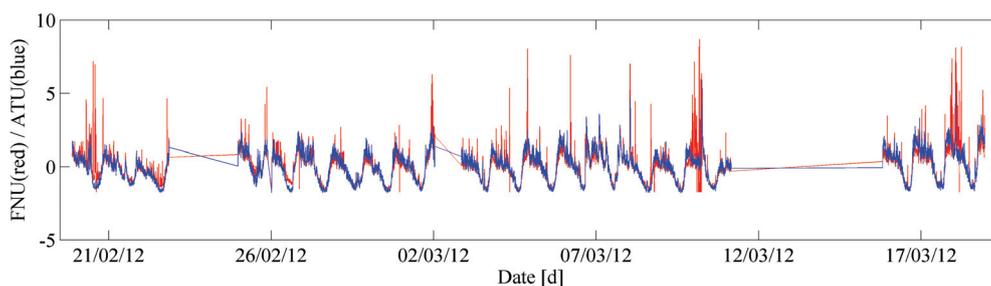


Figure 5-3: Comparison of FNU and ATU (scaled)

It becomes obvious that both values are comparable. In areas with low noise, e.g. between 26<sup>th</sup> February and 2<sup>nd</sup> March the correlation is  $\rho = 0.92$ . For a practical application the FNU values are of more interest, because of the low cost measurement probe. The data depicted in Figure 5-4 and Table 5-1 represent a typical inflow situation in spring.  $NH_4 - N$  is most of the time below 20 mg/l and COD is mainly below 400 mg/l. These relatively low values are caused by dilution and probably infiltration into the sewer system. During summertime with little rain much higher values are expected. Obviously a

dynamic with a one day period can be identified in all data. This is expected due to variations in the inflow of a WWTP. Usually the load follows the daily water consumption of the households connected to the sewer system. For this reason the load rises in the morning, reaches its maximum at midday, stays at a high level until the early evening and decreases into the night. During the night the load is at its minimum. Basically all measurement values depicted in Figure 5-4 show this behaviour. Although one might expect  $NH_4 - N$  and  $COD$  to behave in a similar fashion, taking a closer look at the progression of the curves reveals significant differences. While  $COD$  stays at its maximum  $NH_4 - N$  shows a short peak before settling at a slightly lower level. The flow ( $Q$ ) deviates from this behaviour when rain events occur. It can be observed that concentrations of  $COD$  and  $NH_4 - N$  are slightly lower when the flow is high. This effect can be explained by dilution. While the sum of substances in the sewer stays at the same level, the volume of water rises.

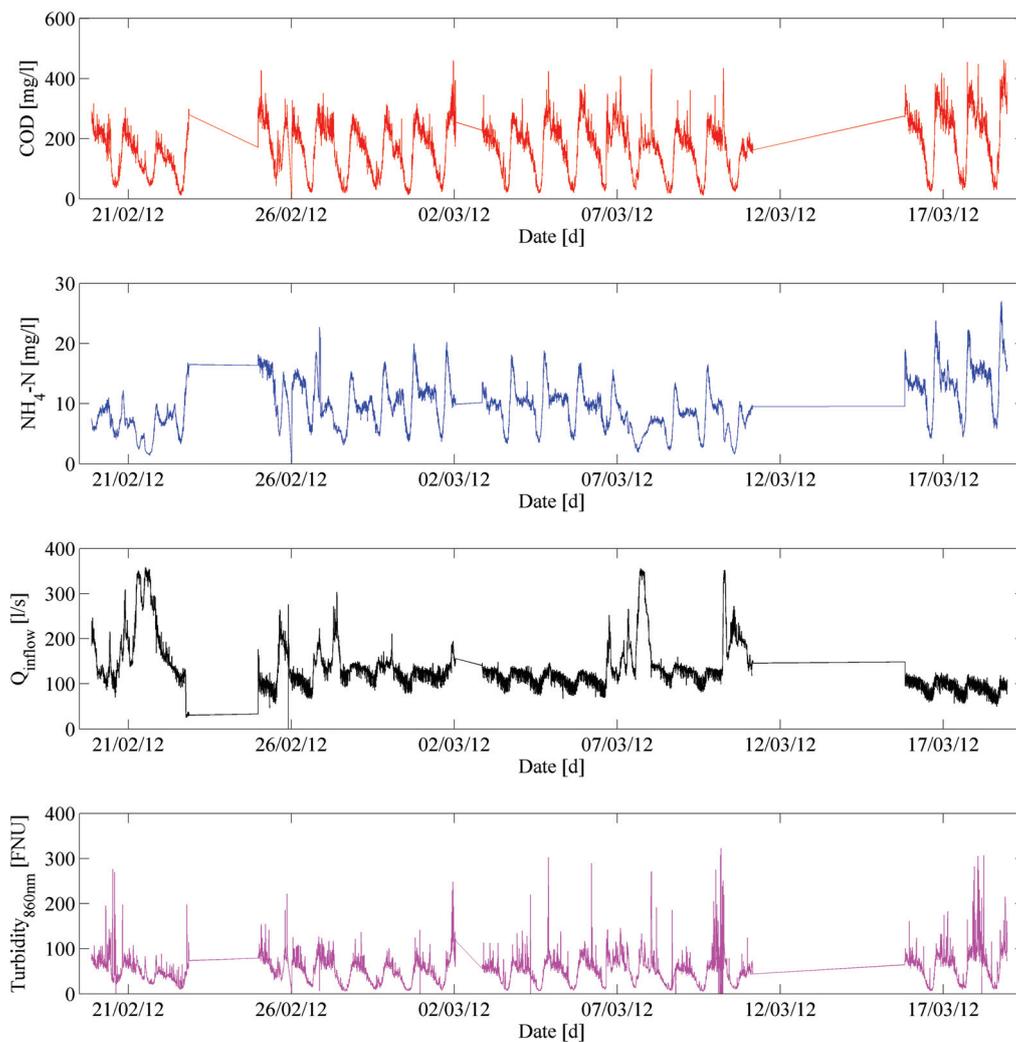


Figure 5-4: Overview of relevant measurement values

Table 5-1 gives an overview of the sensors used to collect the data, including the sensor range, and the min- and max-values and standard deviation of the data recorded. As already described the measurement campaign was conducted in spring. The mean concentration of *COD* is 169.2 mg/l and the mean of *NH<sub>4</sub> - N* is 27 mg/l. Both concentrations are relatively low, while the mean flow is 134.3 l/s. Even when the rain events are not considered, this is still a high inflow. Looking at the flow in Figure 5-4 it can be seen that during the night it stays at approximately 80 l/s. One might conclude from this, that this flow is caused by infiltration into the sewer system.

Table 5-1: Overview of measurement values

Parameter	Probe	Sensor Range	Min Value	Max Value	Mean	Standard Deviation
<i>COD</i>	STIP-scan	0 to 1000 mg/l	3.1	462.0	169.2	85.6
<i>NH<sub>4</sub> - N</i>	ISEmax CAS40D	0 to 1000 mg/l	0.1	27.0	9.63	4.1
<i>FNU<sub>860</sub></i>	Turbimax CUS51D	0 to 9999 FNU	0.0	322.9	53.4	31.0
<i>SAC<sub>25.6</sub></i>	STIP-scan	1 to 250 m <sup>-1</sup>	0.47	88.4	32.4	14.9
<i>ATU</i>	STIP-scan	1 to 250 m <sup>-1</sup>	0	34.4	8.0	4.47
<i>SAC<sub>43.3</sub></i>	STIP-scan	1 to 250 m <sup>-1</sup>	0.02	10.8	3.19	1.98
<i>pH</i>	WTW Sensolyt	2 to 12	6.9	8.5	7.4	0.22
<i>Conductivity</i>	WTW TetraCon	1 uS/cm to 2 S/cm	0.0	1095	412	105.0
<i>Temperature</i>	WTW Sensolyt	0 to 40 °C	5.3	9.6	7.6	0.81
<i>Flow ( Q )</i>	E+H	0 to 500 l/s	0.0	357.6	134.3	53.7

### 5.3.2 Data Preparation

To prepare the data for further analysis, the following five steps were applied:

- Selection of a suitable period of time by hand
- Global outlier elimination
- Local outlier elimination
- Data were transformed to have zero mean and a standard deviation of one
- Randomization of datasets

#### Selection of suitable time period

The first step of data preparation was to find a suitable period of time, where all measurement probes were operating correctly. In the first weeks, in particular, there were several failures due to clogging of the sensors. For the first experiments, three weeks in March 2012 were selected for analysis.

#### Global outlier elimination

The global outlier detection algorithm takes all  $N$  input values  $\mathbf{u} \in \mathbb{R}$  into account and detects global outliers using the median absolute deviation ( $MAD$ ). The filter is non-causal and for this reason not suitable for online filtering.

First the median  $\tilde{u} \in \mathbb{R}^N$  is calculated:

$$\tilde{u} = \text{median}(\mathbf{u}) \quad (5.1)$$

In a second step the median centered vector  $\tilde{\mathbf{u}} \in \mathbb{R}^N$  is calculated by removing the median  $\tilde{u}$  from the input data  $\mathbf{u}$ .

$$\tilde{\mathbf{u}} = [u_0 - \tilde{u}, \dots, u_k - \tilde{u}, \dots, u_{N-1} - \tilde{u}]^T \quad (5.2)$$

Based on  $\tilde{\mathbf{u}}$  the new vector  $\tilde{\mathbf{x}} \in \mathbb{R}^N$  is calculated as follows

$$\begin{aligned} & \text{for each } k = 0, 1, 2, \dots, N-1 \\ & \text{if } |\tilde{u}_k| < \varphi \cdot MAD(\tilde{\mathbf{u}}) \\ & \quad \tilde{x}_k = \tilde{u}_k \\ & \text{else} \\ & \quad \tilde{x}_k = MA(\tilde{\mathbf{u}}, ws) \\ & \text{end} \end{aligned} \quad (5.3)$$

where  $\tilde{\mathbf{x}}$  is composed of the valid values of the original  $\tilde{\mathbf{u}}$  and the moving average  $MA(\tilde{\mathbf{u}}, ws)$  replaces the outliers of  $\tilde{\mathbf{u}}$ . The moving average window size  $ws \in \mathbb{N}$  has to be larger than the biggest data gap

$\tilde{u}$  to allow for complete coverage of missing or implausible data points. Threshold  $\varphi$  determines filter sensitivity and was set to  $\varphi = 3$  in accordance with standard practice. The *MAD* is defined as

$$\begin{aligned} MAD(\tilde{\mathbf{u}}) &= 1.4826 \cdot \text{median}(\hat{\mathbf{u}}) \\ \hat{\mathbf{u}} &= \left[ |\tilde{u}_0 - \tilde{u}|, \dots, |\tilde{u}_k - \tilde{u}|, \dots, |\tilde{u}_{N-1} - \tilde{u}| \right]^T \text{ for } k = 0, \dots, N-1 \end{aligned} \quad (5.4)$$

while the moving average (*MA*) is given by

$$\begin{aligned} MA(\tilde{\mathbf{u}}, ws) &:= \frac{1}{ws+1} \left( \tilde{u} \left( k - \frac{ws}{2} \right) + \tilde{u} \left( k - \frac{ws}{2} + 1 \right) + \dots + \tilde{u} \left( k + \frac{ws}{2} \right) \right) \text{ for } \text{mod}(ws, 2) = 0 \\ \text{and} & \\ MA(\tilde{\mathbf{u}}, ws) &:= \frac{1}{ws} \left( \tilde{u} \left( k - \frac{ws}{2} \right) + \tilde{u} \left( k - \frac{ws}{2} + 1 \right) + \dots + \tilde{u} \left( k + \frac{ws}{2} \right) \right) \text{ for } \text{mod}(ws, 2) \neq 0. \end{aligned} \quad (5.5)$$

In the final step the median  $\tilde{u}$  is added again, so that the corrected vector is  $\mathbf{x} \in \mathbb{R}^N$ :

$$\mathbf{x} = \left[ \tilde{u}_0 + \tilde{u}, \dots, \tilde{u}_k + \tilde{u}, \dots, \tilde{u}_{N-1} + \tilde{u} \right]^T \quad (5.6)$$

This filter is used to eliminate global outliers which do not fit to the complete data set, but cannot be detected by a local filter, because they persist over a longer period of time. A temporary pollution of a sensor can lead to such outliers.

### Local outlier detection

The local outlier detector is an online outlier detection and removal filter. The filter is an extension of the well-known median filter. The observed data point  $u_k$  is compared to the median  $\tilde{u}$  of the present and past data points. If the distance is greater than a predefined threshold  $T_k \in \mathbb{R}$  the value is replaced by a more reasonable value (Menold et al., 1999). This filter is causal, which makes it suitable for online filtering. First a window of the present and past data points is defined

$$\mathbf{w} = (u_{k-wz+1}, u_{k-wz+2}, \dots, u_k), \quad (5.7)$$

where  $ws \in \mathbb{R}$  is the size of the window  $\mathbf{w}$ . After calculating the median  $\tilde{u}$  of  $\mathbf{w}$ , the distance  $d_k$  between the current data point  $u_k$  and the median  $\tilde{u}$  is calculated:

$$d_k = |\tilde{u} - u_k| \quad (5.8)$$

Whether  $u_k$  is valid or an outlier, is determined using a threshold  $T_k$  calculated using the  $MAD$  times a factor  $\varphi$ . The outlier corrected vector  $x$  is calculated as

$$\begin{aligned} & \mathbf{Algorithm\ localOutlierDetection} (x, u, \tilde{u}, w, \varphi) \\ & 1: \text{Set } T_k = \varphi \cdot MAD(w) \\ & 2: \text{For the number of samples } 0 \leq k \leq N - 1 \\ & 3: \quad d_k = dist(\tilde{u}, u_k) \\ & 4: \quad \text{If } d_k < T_k \\ & 5: \quad \quad \text{Set } x_k = u_k \cup x_k = \tilde{u} \\ & 6: \quad \text{End if} \\ & 7: \text{End for} \end{aligned} \tag{5.9}$$

A detailed description of this filter can be found in Menold et al. (1999).

### Normalization and splitting

In the last step the mean was removed from the input data and the data was scaled to have a standard deviation of one. After this step the data was split into training, validation and test data.

To keep this analysis consistent with real world application, the test data used was chronologically after the training and validation data. This is done considering the fact, that a newly developed measurement system based on these analysis would always have to use a model based on historical data.

### Generation of Data Classes

For the application of classification methods the data was divided into different classes. Each class represents data which is in a certain band (e.g. between 100 mg/l and 200 mg/l). The number of classes as well as the different spans were determined manually considering the following aspects:

- typical concentrations of the substance in the inflow
- even distribution of data points in each class
- practical applicability (easy to understand boundaries for an operator)
- complexity (minimal number of required classes to maintain inflow dynamics)

Figure 5-5 shows the distribution of the values of the complete measurement campaign. It can be seen that  $COD$  has two maxima at 50 mg/l and at 230 mg/l. This effect can be traced back to a long high load period during the day and a shorter low load period during the night. In general the data is relatively evenly distributed. Hence, for simplicity  $COD$  was divided into 5 classes (1-5) of equal width with the exception of class 5 which represents all high values over 300 mg/l.  $NH_4 - N$  data is concentrated around 10 mg/l (see Figure 5-5). Consequently, class ranges were made smaller around the maximum to achieve a better distribution of data points per class. Even with this step the classes in the middle still

have significantly more data points than the classes at the edges. A further division of the classes was not made because of the low practical advantages which arise from more small classes in the middle.

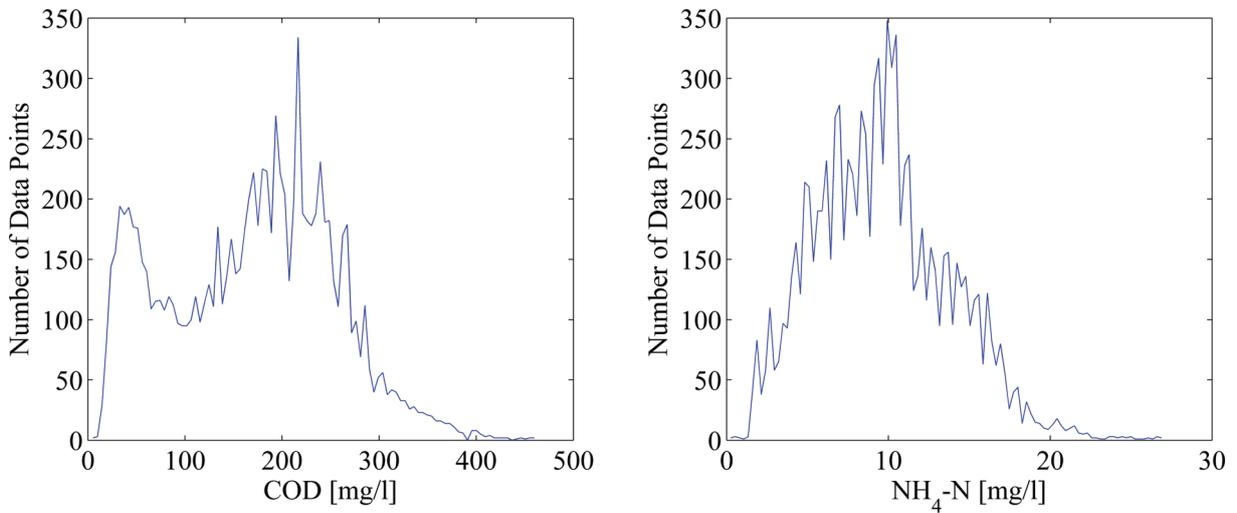


Figure 5-5: Distribution of measurement values in the measurement campaign

Table 5-2 shows the number of samples and ranges of the different classes. For *COD* it is noticeable that class 1 and 4 have the most data points, while for *NH<sub>4</sub>-N* class 2 and 3 are over-represented. The number of data points in class 3 is slightly more than fifty times higher than in class 5. This sample distribution over all classes mainly depends on the inflow load of the plant, which not only changes from plant to plant but also with the seasons of the year. Thus, in order to keep the analysis transferable to other plants, no weighting was applied during classification.

Table 5-2: *COD* and *NH<sub>4</sub>-N* classes

Class	<i>COD</i> [mg/l]	Number of Data Points in <i>COD</i> Classes	<i>NH<sub>4</sub>-N</i> [mg/l]	Number of Data Points in <i>NH<sub>4</sub>-N</i> Classes
1	$0 \geq COD < 100$	2501	$0 \geq NH_4 - N < 5$	1306
2	$100 \geq COD < 150$	1359	$5 \geq NH_4 - N < 8$	2302
3	$150 \geq COD < 200$	2064	$8 \geq NH_4 - N < 15$	5109
4	$200 \geq COD < 300$	3393	$15 \geq NH_4 - N < 20$	1002
5	$300 \geq COD < \infty$	526	$20 \geq NH_4 - N < \infty$	124

### 5.3.3 Generation of input data sets

For the practical implementation and application of the estimation methods at a WWTP, it is important to investigate which in-line measurements in the WWTP inflow are necessary and also financially feasible to achieve sufficiently good results for *COD* and *NH<sub>4</sub>-N*, estimation. Therefore, three different input data sets **U** were generated.

- $\mathbf{U}_1 = (\mathbf{u}_1, \dots, \mathbf{u}_n)$  contains all  $n = 9$  input variables except the respective target variable  $\mathbf{x}$  and thus represents the optimal version assuming the WWTP is fully-equipped with state-of-the-art inline-measurement systems.
- $\mathbf{U}_2 = (\mathbf{u}_Q, \mathbf{u}_{pH}, \mathbf{u}_{cond}, \mathbf{u}_T)$  contains only the standard inflow in-line measurements (Flow rate, pH-value, conductivity and temperature). This represents the majority of WWTPs and is at the same time the version, which requires the least investment costs.
- $\mathbf{U}_3 = (\mathbf{u}_{FNU}, \mathbf{u}_{pH}, \mathbf{u}_Q)$  contains the variables *FNU*, *pH-value* and flow rate which were determined to have significant influence on predictions based on the backward elimination sensitivity analysis described in 5.4.1.

While several other combinations of input variables were tested, these three datasets are the most interesting combinations for practical application. The first dataset is a quasi-reference for the best result that can be achieved using all input variables and can be used as a benchmark. The second one, in contrast, shows what is possible without investment into additional in-line probes and the third one represents a pragmatic trade-off between investment costs and prediction results (only an additional turbidity probe has to be installed).

## 5.4 Description of mathematical methods

The following subsection describes the mathematical methods used in the analysis. With regard to the length of this chapter only the most important methods for the chapter are explained in detail, while standard methods are described in an abbreviated form.

### 5.4.1 Stepwise backward elimination

To identify relevant measurement variables for *COD* and *NH<sub>4</sub>-N* prediction, a stepwise backward elimination algorithm (SBE) was implemented in Matlab®. Figure 5-6 depicts the working principle of the algorithm.

After the data set  $\mathbf{U}$  is loaded and normalized, data rows are randomly permuted and  $\mathbf{U}$  split into four parts: Input training data  $\mathbf{U}_T$ , input validation data  $\mathbf{U}_V$ , target training data  $\mathbf{x}_T$  and target validation data  $\mathbf{x}_V$ . The ratio between training and validation data is 70:30. In this case, the data in  $\mathbf{U}$  was not used in chronological order to minimize influence of the temporal location of the validation data on the input variable selection process.

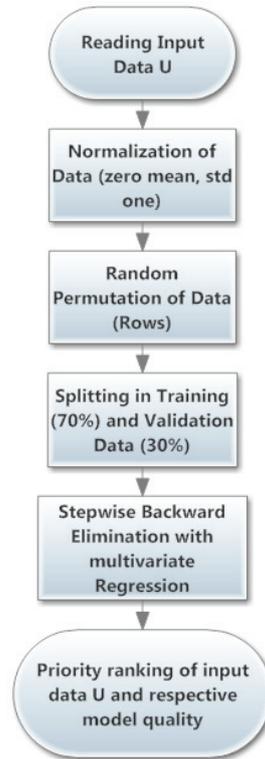


Figure 5-6: Sequence of Stepwise Backward Elimination (SBE)

The actual SBE algorithm (Figure 5-7) consists of two loops. In an inner loop the SBE algorithm uses the training data  $\mathbf{U}_T$  to generate a new input data set  $\mathbf{U}_{T_i}$  omitting data column  $i$  (i.e. omitting the  $i$ -th input). The resulting data set has  $n - 1$  columns, where  $n$  is the number of data columns in  $\mathbf{U}_T$ .  $\mathbf{U}_{T_i}$  is then used to train a regression model, which is evaluated by calculating the RMSE of the prediction on the validation data  $\mathbf{U}_V$  and  $\mathbf{x}_V$ . This is repeated  $n$  times until each input column has been omitted once. After a complete run of the inner loop the best regression model, which is the model where the input column with the least impact on performance was omitted, is determined. In a second outer loop this least important column is removed permanently and the whole procedure is repeated until only one column remains. This last remaining column has the most impact and hence the highest information content about the target value. To ensure the robustness of the results the procedure is repeated several times and the distribution of results analysed. The following pseudo code provides a more detailed description of the SBE variable selection procedure:

```
U = readInputData();
U = normalize(U);
For jj=1 to nTests
{
    U = randPermutate(U); //Rows
    UT = setTrain_Data(U, 70%);
    xT = setTargetTrain_Data(U);
    UV = setValid_Data(U, 30%);
    xV = setTargetValid_Data(U);
    For(kk = 1 to nDataColumns)
    {
        i = number of DataColumns;
        while(i > 1)
        {
            regModel= buildRegressionModel(UT, xT);
            xP = computeModelResults(regModel,UV);
            RMSE[i]= calcRMSE(xP, xV);
            UT.DeleteColumn[i]
            UV.DeleteColumn[i];
            i= i-1;
        }
        UT.DeleteLeastImpactColumn(where min(RMSE));
        UV.DeleteLeastImpactColumn(where min(RMSE));
    }
}
```

Figure 5-7: SBE algorithm in Pseudo-Code

The main advantage of this algorithm is that it generates a priority ranking of the different input columns, similar to a sensitivity analysis. Furthermore, irrelevant input data can be easily identified. Starting with a big data set with a high number of probable irrelevant data, the quality of the regression model will barely decrease until the first relevant input variable is removed. This becomes obvious in the results section. Another advantage is that the SBE doesn't depend on a certain regression algorithm, so that different algorithms can be easily implemented and compared.

### 5.4.2 Regression Methods

Different regression methods were applied to predict the *COD* and *NH<sub>4</sub>-N* concentrations. The following section gives a short overview of the different methods and the reasons for their selection.

### Multivariate linear regression (MLR) / Least squares regression (LSR)

Multi-linear regression, often referred to as least squares regression or ordinary least squares, is one of the simplest linear regression models. One major advantage is, that it is independent of the underlying distribution (Draper, 1998, pp. 135–136). MLR seeks to minimize the sum of squared residuals

$$(\mathbf{x} - \mathbf{U}\mathbf{b})^T (\mathbf{x} - \mathbf{U}\mathbf{b}), \quad (5.10)$$

with respect to the regression parameter vector  $\mathbf{b}$ . Here  $\mathbf{U}$  are the input values and  $\mathbf{x}$  are the target values.

### Partial Least Squares Regression (PLS)

In partial least squares (PLS) the original predictor data matrix  $\mathbf{U}$  and the predicted variables  $\mathbf{x}$  are both projected into a new space in which the covariance between the projected  $\mathbf{U}^*$  and  $\mathbf{x}^*$  (so called  $\mathbf{U}^*$  scores and  $\mathbf{x}^*$  scores) is maximal. The  $\mathbf{U}^*$  scores may contain less predictors than  $\mathbf{U}$ ; the number of predictors is chosen by the user. Furthermore, all predictors in the  $\mathbf{U}^*$  scores are orthogonal to the preceding predictors in the  $\mathbf{U}^*$  scores (Geladi and Kowalski, 1986). For this analysis MATLAB's function *plsregress* is used which implements the SIMPLS algorithm (de Jong, 1993).

### Neural Networks (MLP)

Feedforward artificial neural networks consist of multiple layers of neurons that are fully connected from one layer to the next. Going beyond standard linear perceptrons, so-called Multi-Layer Perceptrons (MLP) are very well suited for learning and mapping of non-linear relationships in highly complex data structures (Cybenko, 1989). Furthermore, MLPs have several desirable properties like universal function approximation capabilities, good generalization properties and the availability of robust efficient training algorithms (Haykin, 1999). For the regression problem at hand a three layer feed forward MLP is used (i.e. with one hidden layer). MLP training is performed using a Levenberg-Marquardt training algorithm and cross-validation to prevent over-fitting.

### MLP Optimization

To optimize the network, different numbers of neurons were tested for each dataset. Each number of neurons was tested several times to ensure robustness using a Monte Carlo algorithm to randomize the training data. Table 5-3 shows the results of the MLP optimization.

Table 5-3: MLP Optimization - optimal number of neurons for the datasets

Dataset	Number of Neurons
$U_{1,COD}$	4
$U_{2,COD}$	50
$U_{3,COD}$	15
$U_{1,NH4}$	20
$U_{2,NH4}$	100
$U_{3,NH4}$	30

### Support vector regression (SVR)

The most commonly used form of Support Vector Regression (SVR) is called  $\varepsilon$ -SVR and was introduced by (Vapnik, 1998). The basic concept behind this SVR is to describe the relationship between input and target variables  $(\mathbf{u}, \mathbf{x})$  using a high dimensional linear function  $\hat{f}$ . This function is defined by minimizing the distance between each predicted and each real value  $|\hat{x}_i - x_i|$  to be not greater than  $\varepsilon$ . Due to the fact, that it might be difficult to find a solution for  $\hat{f}$ , so that all points  $\mathbf{x}$  are within  $\varepsilon$ , an additional, so called slack variable  $\xi$  is introduced. Thus,  $\varepsilon$  is transformed into a soft margin around  $\hat{f}$  that can be described as follows

$$\begin{cases} x_i - ((\mathbf{w} \cdot \mathbf{u}_i) + b) \leq \varepsilon - \xi_i \\ x_i - ((\mathbf{w} \cdot \mathbf{u}_i) + b) \leq -\varepsilon + \xi_i \\ \text{for } \xi_i \geq 0, \varepsilon \geq 0 \end{cases} \quad (5.11)$$

where  $\mathbf{w}$  describes the gradient and  $b$  the intercept of  $\hat{f}$ . In order to determine  $\mathbf{w}$  and  $b$ , the following equation (3.23) needs to be minimized

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + c \sum_{i=1}^N \xi_i \quad (5.12)$$

The trade-off parameter  $c$  is introduced to determine the influence of deviations larger than  $\varepsilon$ . The greater  $c$ , the lower the number of deviations larger than  $\varepsilon$  that are still tolerated.

As the minimization problem from (5.12) can be solved more easily using its dual form, a standard method with Lagrangian multipliers  $\eta_i, \eta_j, \alpha_i, \alpha_j$  is used to create the following dual formulation using the kernel function  $k$

$$\begin{aligned} \max_{\mathbf{a} \in \mathbb{R}^N} \quad & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \eta_i \eta_j \alpha_i \alpha_j \cdot k(\mathbf{u}_i, \mathbf{u}_j) \\ \text{subject to} \quad & \sum_{i=1}^N \eta_i \alpha_i = 0 \quad \text{and} \quad 0 \leq \alpha_i \leq c \quad i = 1, \dots, N \end{aligned} \quad (5.13)$$

where the dual multipliers have to satisfy positivity constraints (3.25).

$$\text{all } \alpha, \eta \geq 0 \quad (5.14)$$

The solution to the dual optimization problem then gives the necessary values for  $\mathbf{w}$  and  $\mathbf{b}$ . In order to allow the SVR algorithm to work for non-linear data sets, it is assumed that a solution for a linear function  $\hat{f}$  exists in a high dimensional feature space  $F$ . As the calculation of a solution for (5.13) in  $F$ , can be very difficult, so called kernel-functions are used to implicitly calculate the dot products needed to determine a solution. For the SVR on the *COD* and *NH<sub>4</sub> - N* data sets an RBF kernel was used as it is perfectly suited for a non-linear relation between target values and training data. The linear kernel is a special case of the RBF kernel as proven by Keerthi and Lin (2003).

### SVR Optimization

The SVR models were optimized using a grid search to find the optimal trade off parameter  $c$  and  $\gamma$ , a parameter which defines how far the influence of a single training example reaches. The optimal parameter are depicted in Table 5-4.

**Table 5-4: SVR Optimization - Optimal Parameter for Datasets**

Dataset	$\gamma$	$c$
$\mathbf{U}_{1,COD}$	0.005	1024
$\mathbf{U}_{2,COD}$	0.125	4
$\mathbf{U}_{3,COD}$	0.125	4
$\mathbf{U}_{1,NH4}$	0.25	4
$\mathbf{U}_{2,NH4}$	0.5	8
$\mathbf{U}_{3,NH4}$	1	2

### 5.4.3 Classification Methods

As an alternative to the regression methods, different classification methods were applied to predict the inflow concentrations. To formulate the mapping from  $\mathbf{u}_i$  to  $\mathbf{x}_i$  for all measurements  $i = 1, \dots, N$  as a classification problem, *COD* and *NH<sub>4</sub> - N* concentrations are clustered into  $C = 5$  classes from low, low-normal, normal, normal-high and high inflow concentrations. Thereby, every input sample  $\mathbf{u}_i$  is

associated with a class number  $\kappa_i \in \Theta$  with  $\Theta := \{1, 2, 3, 4, 5\}$  as defined in Table 5-2. This section gives a short overview of the applied classification methods.

In order to evaluate classification performance, the following two error measures as given in Wolf et al. (2013) are used.

$$\text{MCR} := 100 \cdot \left( 1 - \frac{1}{N} \cdot \sum_{i=1}^N 1(\mathbf{u}_i) \right), \text{ where } 1(\mathbf{x}_i) := \begin{cases} 1 & \text{if } f_{\text{classifier}}(\mathbf{u}_i) = \kappa_i \\ 0 & \text{otherwise} \end{cases} \quad (5.15)$$

where  $f_{\text{classifier}} : \mathbf{U} \rightarrow \Theta$  is the mapping function. Given the confusion matrix  $\mathbf{K} := (k_{j,l}) \in \mathbb{R}^{C \times C}$ , with  $\sum_{l=1}^C k_{j,l} = 100$ ,  $j = 1, \dots, C$ , for each classifier, the Normalized MCR (NMCR) is used as an alternative performance measure, that gives equal weighting to each class.

$$\text{NMCR} := 100 - \frac{1}{C} \cdot \sum_{j=1}^C k_{j,j} \quad (5.16)$$

Due to the unbalanced number of samples in the fifth class in particular, the NMCR was used for validation of the classification methods.

### Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a method from the area of machine learning, which is used to find a linear combination of features to separate one or more classes of data. In order to do this LDA searches for a linear transformation  $\mathbf{A} \in \mathbb{R}^{m \times p}$  with  $m \leq p$ , such that the transformed data  $\mathbf{U}^* = \mathbf{A} \cdot \mathbf{U}$  can be better linearly separated than the original feature input vectors  $\mathbf{U}$ . To determine the transformation matrix  $\mathbf{A}$  an optimization problem has to be solved that corresponds to the maximization of the Fisher discriminant criterion:

$$\text{trace}(\mathbf{S}_T^{-1} \cdot \mathbf{S}_B), \quad (5.17)$$

where  $\mathbf{S}_T$  is the total scatter matrix and  $\mathbf{S}_B$  is the between classes scatter matrix (Duda, 2001). This method was chosen for the analysis because of its linear approach.

In this case LDA was used not only for classification but also for dimension reduction. For the latter, different numbers of eigenvectors, meaning different target spaces were tested for each dataset. Each number of eigenvectors was tested several times to ensure robustness using a Monte Carlo algorithm to randomize the training data. Table 5-5 shows the results of the LDA optimization.

**Table 5-5: LDA Optimization - Optimal Number of Eigenvectors for the Datasets**

Dataset	Number of Eigenvectors
$U_{1,COD}$	1
$U_{2,COD}$	4
$U_{3,COD}$	4
$U_{1,NH4}$	4
$U_{2,NH4}$	4
$U_{3,NH4}$	3

### Random Forest (RF)

Random Forest is an algorithm for solving complex classification and regression problems, introduced by (Breiman, 2001). The algorithm is an ensemble of decision trees (DT). Hence, the classification consists of an ensemble of classification trees, where each tree is trained on a bootstrapped sample of the original training data set, and at each new branch the candidate set of variables is a random subset of all variables. RF was chosen as a non-linear classification method as well as regression method from the field of ensemble modelling.

### RF Optimization

For the optimization different numbers of decision trees were tested for each dataset. Each number of decision trees was tested several times to ensure robustness using a Monte Carlo algorithm to randomize the training data. Table 5-6 shows the optimal number of decision trees when RF is used for classification and regression.

**Table 5-6: RF Optimization - Optimal Number of Decision Trees for Classification and Regression for each Dataset**

Dataset	Number of Trees (Classification )	Number of Trees (Regression )
$U_{1,COD}$	3000	50
$U_{2,COD}$	1000	100
$U_{3,COD}$	3000	50
$U_{1,NH4}$	500	100
$U_{2,NH4}$	1000	50
$U_{3,NH4}$	3000	50

## Support Vector Machines (SVM)

Support Vector Machines are a method for solving multi-class classification problems by finding hyperplanes, which separate data sets into classes in a high dimensional feature space. For the classification problem under consideration a C-Support Vector Classification is used with soft margin optimization and a Radial Basis Function Kernel (RBF Kernel) (Cortes and Vapnik, 1995) using the SVM implementation LIBSVM (Chang and Lin, 2001).

For the same reasons as for SVR an RBF kernel was also selected for SVM (section 5.4.2). Furthermore, a grid search is performed to determine the best values of parameters  $c$  and  $\lambda$  for the RBF kernel function according to the training data using the misclassification rate (MCR). Training is performed with a 5-folded cross-validation procedure (one against one) and different pairs of  $c$  and  $\lambda$  values are tested. Finally, the one that yields the best cross-validation accuracy is picked. As suggested by Hsu in 2003, in a first pass an exponentially growing sequences of  $c$  and  $\lambda$  are evaluated to identify interesting regions for a detailed grid search (Hsu et al., 2003).

### SVM Optimization

In the same way as for SVR a grid search to find the optimal parameters was conducted for each dataset separately. The optimal parameters are depicted in Table 5-7.

Table 5-7: SVM Optimization - optimal parameter for datasets

Dataset	$\gamma$	$c$
$U_{1,COD}$	0.0625	64
$U_{2,COD}$	1	1
$U_{3,COD}$	0.5	16
$U_{1,NH4}$	0.25	16
$U_{2,NH4}$	0.25	128
$U_{3,NH4}$	2	8

## 5.5 Results

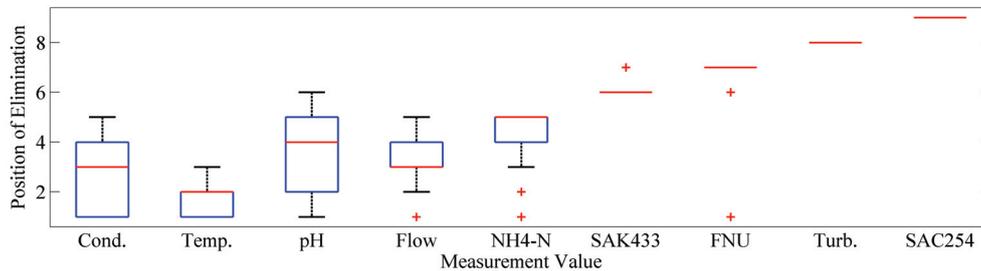
### 5.5.1 Sensitivity Analysis

The idea of the sensitivity analysis is to get a better idea of the influence of the different measured process variables on  $COD$  and  $NH_4 - N$ . For this purpose Stepwise Backward Elimination (SBE) (described in section 5.4.1) was implemented in Matlab and applied to the input variables  $\mathbf{u}$  in order to find which of the input variables are best suited for inclusion in the models  $COD = f_1(\mathbf{u})$  and  $NH_4 - N = f_2(\mathbf{u})$  to estimate  $COD$  and  $NH_4 - N$ . To get robust results SBE was repeated 50 times using different randomly selected data sets generated using Monte Carlo simulation. Because each data

row represents one set of measurements at a certain point of time, it is possible to randomize the data rows. For each repetition the data rows were randomly permuted with the exception of the first run where the data was left in its original order to investigate if the timed measurement sequence has a significant impact on the results.

### 5.5.1.1 COD Analysis Sensitivity Analysis

Figure 5-8 shows a box plot of the order of elimination, which means that the last input value eliminated, in this case  $SAC_{254}$ , has the highest impact on model quality. This was expected as  $COD$  represents the sum of all organic compounds and a common alternative for  $COD$  is  $SAC_{254}$ . Turbidity ( $ATU$ ) has the second largest influence on  $COD$  prediction, which can be explained by the fact that a fraction of the  $COD$  exists in particulate form. The order of elimination of  $FNU$  (measured at 860 nm) and  $SAC_{433}$  was also consistent across most of the simulation runs. While the remaining values show a much higher variation. The reason for this variation lies in their small influence on model quality, hence they are much more sensitive to the randomization of the data. From a practical point of view it is disappointing that *conductivity*, *temperature*, *pH-value* and *flow* were the first values to be eliminated because it means that these variables, that are currently available as standard on most WWTP, do not improve model quality significantly.



**Figure 5-8: Predicting COD – The most significant input is on the right ( $SAC_{254}$ ) and the least significant is on the left (Cond.)**

Figure 5-9 depicts the coefficient of determination ( $R^2$ ) of the different models for the test data. It can be seen that model quality is and stays high up to the point where only three input values remain. It seems that these three input variables ( $FNU$ , *turbidity* and  $SAC_{254}$ ) have the highest impact on model quality. The first small degradation can be seen when only four input variables remain. The fourth value is  $SAC_{433}$  which represents the absorption of a sample at 433 nm, where the colour yellow shows a high absorption as well. This makes sense because with higher pollution of the water (and therefore higher  $COD$ ) the wastewater tends to have a stronger colour. At this point it has to be mentioned that these interpretations are based on the fact that all measurements were taken in municipal wastewater. As already noted in chapter 2, the characteristics of, for example, industrial wastewater in many cases are completely different. For example high  $COD$  values can appear without a particulate fraction and

without significant coloration. Thus, detection and regression analysis is much more difficult in this environment.

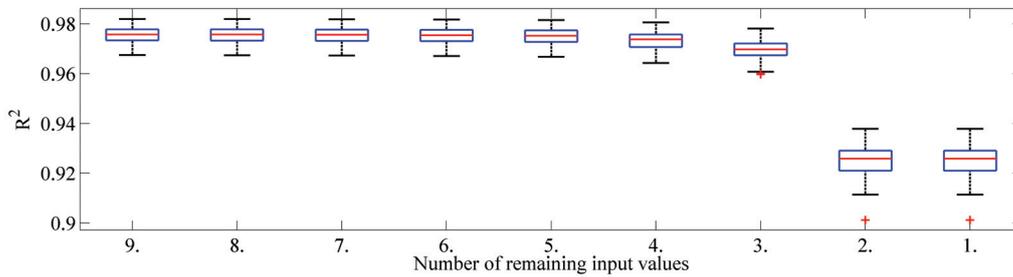


Figure 5-9: COD - Coefficient of determination ( $R^2$ ) for regression models with different numbers of input, with the selection order given in Figure 5-8

### 5.5.1.2 NH<sub>4</sub>-N Analysis

Figure 5-10 shows the order of priority of input variables for the prediction of  $NH_4 - N$ . The prediction of  $NH_4 - N$  is expected to be more difficult than  $COD$  due to the fact that it is present in completely dissolved form and has no specific absorption. This is also supported by the fact that *Turbidity* has no significant impact on model quality. Interesting for the practical application are *pH - value* and *Flow* which were eliminated fourth from last and third from last respectively. The influence of the *pH - value* can be explained by the fact that in the inflow of a WWTP the major part of ammonium is present as fully protonated  $NH_4^+$ . Depending on *pH - value* and *temperature* it can deprotonate to  $NH_3$  (ammoniac) and vice versa. This means that the *pH - value* has an influence on the  $NH_4 - N$  concentration and the  $NH_4 - N$  concentration on the *pH - value*. The correlation between *flow* and  $NH_4 - N$  can be explained by diluting effects (the more water, the less  $NH_4 - N$ ). The  $SAC_{254}$  and  $SAC_{433}$  measurements have the highest influence. In comparison to  $COD$  it becomes obvious that the variation in selection order is much higher, which means that most of the variables have a similar influence on the result, which means in this case only a small influence.

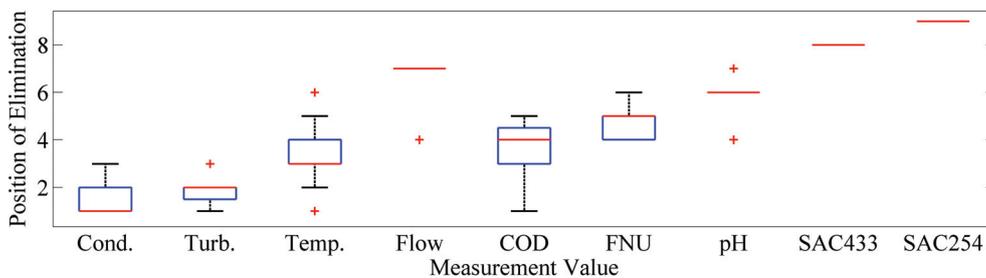


Figure 5-10: Predicting  $NH_4-N$  – The most significant input is on the right ( $SAC_{254}$ ) and the least significant is on the left (*Cond.*)

Looking at Figure 5-11 it can be seen that even with all input variables included the coefficient of determination only reaches  $R^2 = 0.75$ . When *Flow* is omitted from the model  $R^2$  drops to 0.56. Based on these first test results it is not expected that good prediction results can be achieved with other mathematical methods for  $NH_4 - N$ . The later presented results confirm this expectation.

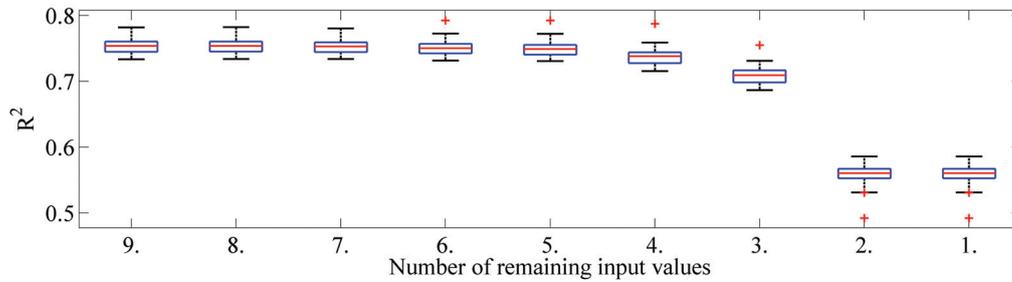


Figure 5-11:  $NH_4-N$  - Coefficient of determination ( $R^2$ ) for regression models with different numbers of input, with the selection order given in Figure 5-10

The results of the sensitivity analysis are used for the following analysis to identify valuable process variables and to create optimized data sets for regression and classification.

### 5.5.2 Regression Results COD

#### Regression Dataset $U_{1,COD}$

Dataset  $U_{1,COD}$  includes all measured input variables except the target variable. Therefore the model errors (Figure 5-12) are, as expected for *COD*, relatively small. Furthermore it is noticeable that the errors of the different regression algorithms are in the same range.

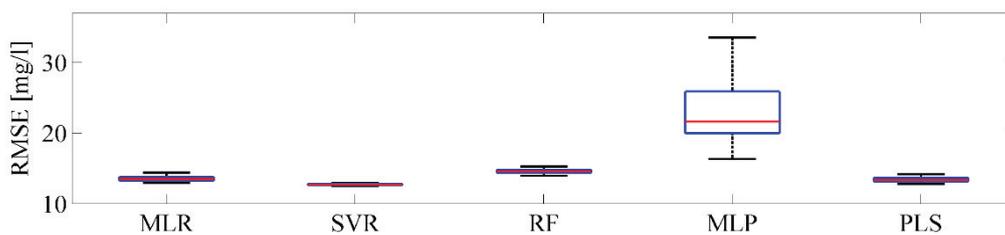


Figure 5-12: Regression Results Dataset  $U_{1,COD}$  (50 runs monte carlo simulation)

An average RMSE of 12 *mg/l* is achieved, which corresponds to an average NMSE of 6 %. The accuracy of these results is absolutely sufficient for the previously mentioned different target applications such as modelling and optimized control. The problem is, that  $SAC_{254}$  is the most important input variable (Figure 5-11).  $SAC_{254}$  probes are in general nearly as expensive as *COD* probes and therefore often used as an alternative to a *COD* probe. So from this point of view there is no major financial advantage for an operator to use this combination of sensors.

### Regression Dataset $U_{2,COD}$

Due to the fact that dataset  $U_{2,COD}$  only includes standard in-line measurements, which are variables with minor impact on  $COD$  prediction (see Figure 5-8), the results show an average RMSE of 66.7 mg/l, which corresponds to an average NMSE of 42 %. As expected these are substantially worse than the results for dataset  $U_{1,COD}$ .

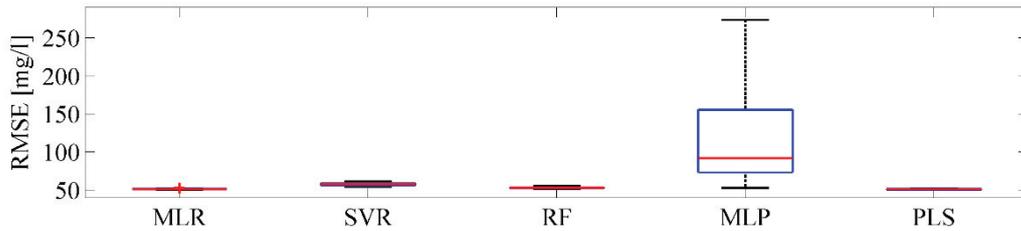


Figure 5-13: Regression Results Dataset  $U_{2,COD}$  (50 runs Monte Carlo simulation)

Apart from the fact that neural networks (MLP) are more sensitive to the randomized input data sets and therefore show a higher error distribution, it is obvious that the linear and non-linear methods yield similar performance. Despite the poor RMSE results these models may still be useful for practical application, for example in the early detection of load peaks. Figure 5-14 shows the SVR  $COD$  prediction results. The dynamics are clearly depicted as well as the range of the  $COD$  values. Therefore, this is an interesting result, due to the fact that it is obviously possible to predict  $COD$  up to a certain level with the measurement equipment installed as standard on WWTP. The level of prediction is sufficient to be a useful input variable for advanced control strategies.

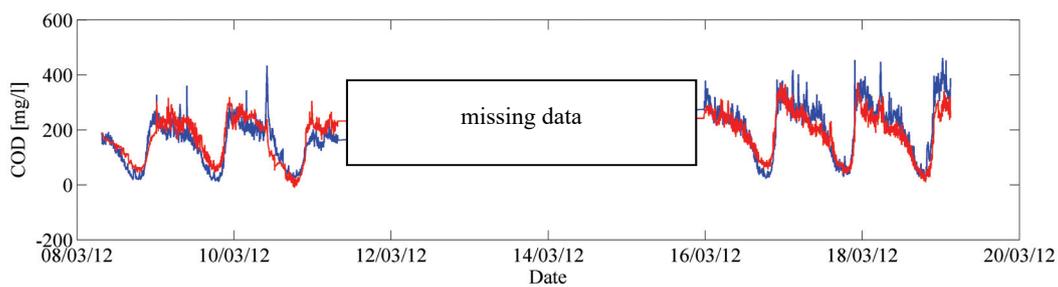


Figure 5-14: Prediction using SVR for  $U_{2,COD}$  (Measured: Blue, Predicted: Red)

### Regression Dataset $U_{3,COD}$

Dataset  $U_{3,COD}$  uses the input variables with the highest information about the target variable  $COD$  determined by the sensitivity analyses (stepwise backward elimination) (see Figure 5-6). Furthermore, all variables measured by the Stip-scan spectrometric probe were eliminated from the dataset (including

$SAC_{254}$  which would lead to significant improvement of the results) to create a dataset that contains only those variables which can be measured by cheaper measurement probes. Therefore, this dataset could be the most important one for the targeted applications. To gather these data, it is only necessary to install an additional turbidity probe in the inflow of a WWTP.

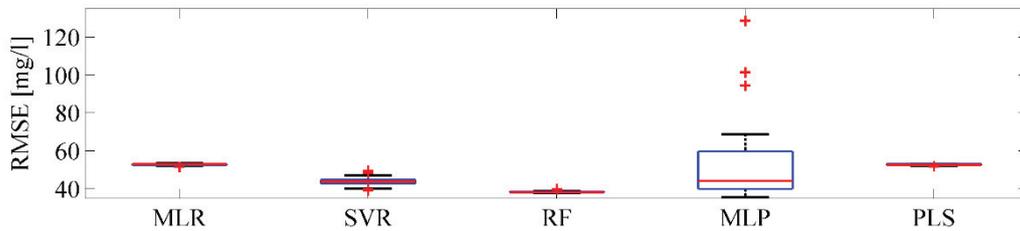


Figure 5-15: Regression Results Dataset  $U_{3,COD}$  (50 runs Monte Carlo simulation)

With an RMSE of 48.8 mg/l the results are significantly better than the results of dataset  $U_{2,COD}$ . Noticeable is the fact that the non-linear methods SVR, RF and MLP perform better than the linear methods MLR and PLS.

### 5.5.3 Regression Results $NH_4-N$

#### Regression Dataset $U_{1,NH4}$

Dataset  $U_{1,NH4}$  contains all measured input variables and offers the best chance for good prediction results. From this point of view it can be seen as a reference for which results can be achieved.

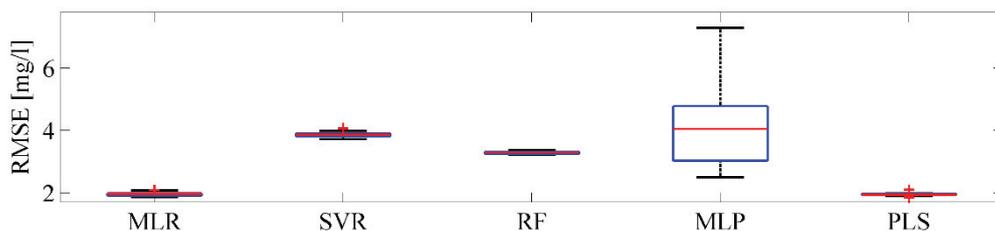


Figure 5-16: Regression Results Dataset  $U_{1,NH4}$  (50 runs Monte Carlo simulation)

In comparison to  $COD$  results  $NH_4-N$  results are much worse. MLR and PLS have a RMSE of 1.96 mg/l and 1.94 mg/l respectively. Interesting is the fact that in this case linear methods (MLR and PLS) perform better than the non-linear methods (RF, MLP and SVR). It can be assumed that there is a linear correlation with  $SAC_{254}$ .

### Regression Dataset U<sub>2,NH4</sub>

The average RMSE for dataset U<sub>2,NH4</sub> is 4.3 mg/l. It is noticeable that the differences in the RMSE between the applied regression methods is negligible. Furthermore, variations are relatively small, which shows that the results do not depend on a certain area in the validation data.

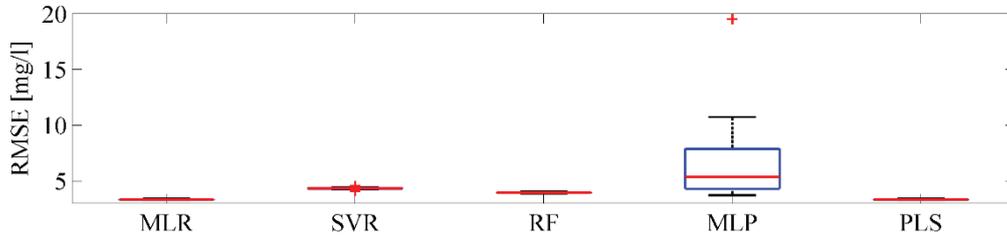


Figure 5-17: Regression Results Dataset U<sub>2,NH4</sub> (50 runs Monte Carlo simulation)

### Regression Dataset U<sub>3,NH4</sub>

Although an improvement of the results was expected by including *turbidity*, the results for dataset U<sub>3,NH4</sub> have higher variations than dataset U<sub>2,NH4</sub>. MLR and PLS show comparable prediction results to dataset U<sub>2,NH4</sub> for dataset U<sub>3,NH4</sub> with an RMSE of 3.58 mg/l and 3.59 mg/l respectively. For this data set it is noticeable that the linear methods MLR and MNR show significantly better results than the non-linear methods.

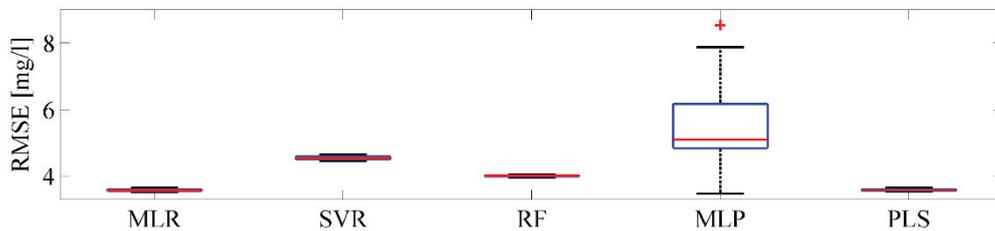


Figure 5-18: Regression Results Dataset U<sub>3,NH4</sub> (50 runs Monte Carlo simulation)

It can be assumed that the differences are caused by the variation of the *turbidity*, because it is the only new variable included. An exception is MLP which has an average RMSE of 5.53 mg/l, which is slightly better than the MLP RMSE for dataset U<sub>2,NH4</sub> with 6.71 mg/l. Nevertheless, these results are strongly influenced by the sensitivity of the MLP to randomization and the corresponding variation of the MLP results.

### 5.5.4 Classification Results

A different approach is to estimate COD and NH<sub>4</sub>-N concentrations in the inflow by classification. Instead of estimating the concentrations continuously they are estimated in different bands (for example

COD class 1 is between 100 mg/l and 200 mg/l). Table 5-8 and Table 5-12 show the NMCR of the applied classification methods for COD and NH<sub>4</sub> – N respectively.

### Different Kinds of Misclassifications

For later interpretation of the confusion matrices it is important to keep in mind that there are different types of misclassifications. Because the classes are on a cardinal scale, it makes a big difference for the utilization of classification results whether the algorithm confuses adjacent classes or non-adjacent classes. From this point of view, it can be distinguished between “severe” misclassifications and “mild” misclassifications. After the back transformation into mg/l values this effect will be considered automatically by applying the NMSE and the RMSE metrics.

### Discussion of COD Classification Results

Similar to the regression results, COD predictions show considerably better results than NH<sub>4</sub> – N predictions and there is good agreement in terms of datasets which give good results. The best prediction results are achieved by Random Forest Regression for U<sub>1,COD</sub> with an average NMCR of approximately 6.13%. For dataset U<sub>2,COD</sub> the best results are achieved by LDA with an average NMSE of 56.9%. The best results for U<sub>3,COD</sub> are achieved by Random Forest with an average NMSE of 47.8%. These results are particularly interesting, as it can be seen, that the additional turbidity measurement improves the classification results significantly compared to U<sub>2,COD</sub>.

Table 5-8: Median (20 repetitions) Classification Results for COD Test Data -NMCR[%]

Dataset	RF <sub>class</sub>	LDA	SVM
U <sub>1,COD</sub>	6.13	11.09	10.05
U <sub>2,COD</sub>	66.93	56.90	73.14
U <sub>3,COD</sub>	47.79	51.86	52.34

Table 5-9, Table 5-10 and Table 5-11 show the confusion matrices for the different classification algorithms and datasets. For datasets U<sub>1,COD</sub> RF achieved a hit ratio over 90% for all classes (Table 5-9a). LDA and SVM performed slightly worse (Table 5-9 b+c), but still achieved a hit ratio between 80% and 99%. In general it can be said that all classification methods performed very well for the prediction of dataset U<sub>1,COD</sub>.

**Table 5-9: Confusion Matrices Dataset U<sub>1,COD</sub>**

a) LDA		predicted					b) SVM		predicted					c) RF		predicted				
given	[%]	1	2	3	4	5	given	[%]	1	2	3	4	5	given	[%]	1	2	3	4	5
	1	<b>99.5</b>	0.53	0.0	0.0	0.0		1	<b>99.6</b>	0.1	0.0	0.3	0.0		1	<b>98.8</b>	0.9	0.0	0.3	0.0
	2	5.4	<b>86.7</b>	7.8	0.0	0.0		2	12.3	<b>69.3</b>	16.3	2.1	0.0		2	3.9	<b>92.8</b>	3.3	0.0	0.0
	3	0.0	4.1	<b>88.5</b>	7.3	0.0		3	0.2	6.1	<b>78.8</b>	14.8	0.2		3	0.2	2.4	<b>94.4</b>	2.9	0.0
	4	0.0	0.0	8.8	<b>85.4</b>	5.7		4	0.0	0.0	4.0	<b>90.9</b>	5.1		4	0.0	0.0	3.3	<b>92.2</b>	4.5
	5	0.0	0.0	0.5	5.8	<b>93.5</b>		5	0.0	0.0	0.3	9.9	<b>89.8</b>		5	0.0	0.0	0.6	4.7	<b>94.8</b>

Table 5-10 shows the results for dataset U<sub>2,COD</sub>. For this dataset LDA performs best. The majority of samples are mapped to the correct or the adjacent class (mild misclassification). SVM has obviously the biggest problems, it tends to map most samples to class 4 and is not able to recognize class 5 at all. While RF has problems with class 5 too, most of the incorrect mappings are to the adjacent classes.

**Table 5-10: Confusion Matrices Dataset U<sub>2,COD</sub>**

a) LDA		predicted					b) SVM		predicted					c) RF		Predicted				
given	[%]	1	2	3	4	5	given	[%]	1	2	3	4	5	given	[%]	1	2	3	4	5
	1	<b>95.6</b>	4.4	0.0	0.0	0.0		1	<b>22.3</b>	3.2	0.0	74.5	0.0		1	<b>84.0</b>	7.3	8.6	0.0	0.0
	2	36.4	<b>35.8</b>	25.3	2.4	0.0		2	0.0	<b>9.6</b>	14.8	75.6	0.0		2	20.5	<b>35.5</b>	30.4	13.6	0.0
	3	6.4	13.9	<b>50.3</b>	29.3	0.0		3	0.0	0.5	<b>15.0</b>	84.5	0.0		3	1.4	13.9	<b>43.1</b>	41.6	0.0
	4	1.1	2.0	40.9	<b>48.4</b>	7.6		4	0.0	0.0	0.5	<b>99.4</b>	0.1		4	0.0	0.8	19.1	<b>79.9</b>	0.2
	5	0.0	0.0	6.7	31.2	<b>62.1</b>		5	0.0	0.0	0.0	100	<b>0.0</b>		5	0.0	0.0	4.1	88.0	<b>7.9</b>

Table 5-11 shows the results for dataset U<sub>3,COD</sub>. For this dataset RF shows again the best results, but the differences in NMSE are relatively small compared to the other datasets. Except for RF class 5 the majority of samples are mapped to the correct classes. Furthermore, the number of severe misclassifications is relatively small for all methods. This shows that besides the comparable NMSE results for dataset U<sub>2,COD</sub> and U<sub>3,COD</sub>, the additional turbidity probes brought notable improvements.

**Table 5-11: Confusion Matrices Dataset U<sub>3,COD</sub>**

a) LDA		predicted					b) SVM		predicted					c) RF		predicted				
given	[%]	1	2	3	4	5	given	[%]	1	2	3	4	5	given	[%]	1	2	3	4	5
	1	<b>88.1</b>	11.2	0.3	0.3	0.1		1	<b>60.5</b>	2.5	23.0	14.0	0.0		1	<b>95.2</b>	4.1	0.4	0.3	0.0
	2	26.5	<b>34.3</b>	22.6	16.6	0.0		2	9.6	<b>17.5</b>	55.7	17.2	0.0		2	16.9	<b>61.7</b>	21.1	0.3	0.0
	3	3.5	0.2	<b>42.1</b>	51.3	3.0		3	0.3	1.6	<b>44.5</b>	52.9	0.7		3	0.2	13.8	<b>68.2</b>	17.7	0.2
	4	1.9	0.0	10.3	<b>74.2</b>	13.6		4	0.0	0.0	17.8	<b>74.8</b>	7.3		4	0.3	0.5	26.9	<b>69.8</b>	2.5
	5	0.0	0.0	0.3	19.8	<b>79.9</b>		5	0.0	0.0	4.9	32.1	<b>62.9</b>		5	0.0	0.0	0.6	70.3	<b>29.1</b>

Generally it can be said that, as long as the number of mild misclassifications are not the majority of the results, even classification results with a high NMCR > 50% can still be considered to be of high value for the practical applications.

### Discussion of NH<sub>4</sub>-N Classification Results

The NMCR distribution of NH<sub>4</sub> – N for the different datasets  $U_{1,NH_4}$ ,  $U_{2,NH_4}$  and  $U_{3,NH_4}$  over all methods is much less diverse than for COD. While the NMSE for COD had a range between 6% and 73%, the NMSE for NH<sub>4</sub> – N results are for all datasets between 44% and 75%. The reason is that NH<sub>4</sub> – N does not correlate as strongly as COD with SAC<sub>254</sub>. The best results with an NMCR of approximately 44% were achieved by LDA and for dataset  $U_{1,NH_4}$ . Clearly more interesting are the results for dataset  $U_{2,NH_4}$  and  $U_{3,NH_4}$  (Table 5-12). The different methods achieve NMCRs comparable to the NMCRs of the COD prediction, which was not expected.

**Table 5-12: Median (20 repetitions) Classification Results for NH<sub>4</sub>-N Test Data -NMCR[%]**

Dataset	RF <sub>class</sub>	LDA	SVM
$U_{1,NH_4}$	55.85	43.99	70.60
$U_{2,NH_4}$	62.4	57.72	75.36
$U_{3,NH_4}$	57.00	64.41	56.15

Only looking at the NMSE the impression is created that the results for  $U_{2,NH_4}$  and  $U_{3,NH_4}$  are not much worse than for  $U_{1,NH_4}$ . Table 5-13 shows the confusion matrices for dataset  $U_{1,NH_4}$ . Looking at the confusion matrices it can be seen, that none of the methods were able to map class 5 correctly. Class 5, with 124 data samples (Table 5-2), is underrepresented and there are even fewer class 5 NH<sub>4</sub> – N samples in the training data. In the future this problem could be addressed by artificially multiplying the number of samples in underrepresented classes. While LDA is not able to map class 5 correctly it maps 99% of the samples to class 4, which is only a mild misclassification. The same applies for the other classes. Most of the misclassifications are mild. SVM and RF have obviously more problems with higher classes. The reason LDA outperforms SVM and RF is that non-linear methods have more difficulty to extrapolating. Due to the fact that the test data samples are chronologically after the training data samples and the values rose during the measurement campaign, the training data contains fewer class 5 samples.

**Table 5-13: Confusion Matrices Dataset  $U_{1,NH_4}$**

a) LDA		predicted						b) SVM		predicted						c) RF		predicted					
given	[%]	1	2	3	4	5	given	[%]	1	2	3	4	5	given	[%]	1	2	3	4	5			
	1	<b>74.4</b>	12.4	13.2	0.0	0.0		1	<b>75.5</b>	24.0	0.5	0.0	0.0		1	<b>95.4</b>	4.6	0.0	0.0	0.0			
2	5.3	<b>23.8</b>	70.9	0.0	0.0	2	5.7	<b>47.9</b>	46.0	0.30	0.0	2	31.4	<b>57.3</b>	11.3	0.0	0.0						
3	0.1	0.6	<b>88.6</b>	10.7	0.0	3	0.0	5.1	<b>93.0</b>	1.9	0.0	3	2.3	23.7	<b>71.8</b>	2.1	0.0						
4	0.0	0.0	42.4	<b>57.6</b>	0.0	4	0.0	5.4	86.8	<b>7.7</b>	0.0	4	0.0	43.1	36.7	<b>20.3</b>	0.0						
5	0.0	0.0	0.9	99.0	<b>0.0</b>	5	0.0	0.0	70.0	30.0	<b>0.0</b>	5	0.0	18.2	5.5	76.4	<b>0.0</b>						

Table 5-14 shows the confusion matrices for dataset  $U_{2,NH_4}$ . RF and SVM have again problems mapping the higher classes correctly; they map nearly all samples below class 4. Again, none of the methods was able to map class 5 correctly. The simpler form of LDA gives it an advantage over the non-linear methods. All methods give the best results for class 3.

**Table 5-14: Confusion Matrices Dataset  $U_{2,NH_4}$**

a) LDA		predicted					b) SVM		predicted					c) RF		predicted				
given	[%]	1	2	3	4	5	given	[%]	1	2	3	4	5	given	[%]	1	2	3	4	5
	1	<b>38.8</b>	26.7	34.5	0.0	0.0		1	<b>55.0</b>	43.4	1.6	0.0	0.0		1	<b>65.2</b>	34.2	0.5	0.0	0.0
	2	17.6	<b>32.8</b>	49.4	0.2	0.0		2	8.2	<b>45.6</b>	46.1	0.0	0.0		2	23.5	<b>48.6</b>	27.9	0.0	0.0
	3	0.6	5.6	<b>88.6</b>	5.1	0.0		3	1.0	12.1	<b>85.0</b>	1.9	0.0		3	4.4	18.1	<b>75.0</b>	2.5	0.0
	4	0.0	2.1	69.1	<b>29.0</b>	0.0		4	0.0	34.6	57.6	<b>7.7</b>	0.0		4	0.2	29.6	54.2	<b>15.9</b>	0.0
	5	0.0	0.0	28.2	71.8	<b>0.0</b>		5	0.0	3.6	84.5	11.8	<b>0.0</b>		5	0.0	0.9	36.6	62.7	<b>0.0</b>

Table 5-15 shows the confusion matrices of dataset  $U_{3,NH_4}$ . Again, there is a strong tendency of all methods to map all samples below class 3. SVM maps nearly all samples in class 3, which is interesting because looking at the NMSE SVM outperforms LDA and RF. The results for dataset  $U_{3,NH_4}$  aren't much better than for  $U_{2,NH_4}$ , while the necessary measurement effort is much higher. This is why dataset  $U_{3,NH_4}$  is considered not interesting for practical application.

**Table 5-15: Confusion Matrices Dataset  $U_{3,NH_4}$**

a) LDA		predicted					b) SVM		Predicted					c) RF		predicted				
given	[%]	1	2	3	4	5	given	[%]	1	2	3	4	5	given	[%]	1	2	3	4	5
	1	<b>54.2</b>	9.7	36.1	0.0	0.0		1	<b>14.0</b>	11.9	74.1	0.0	0.0		1	<b>72.2</b>	25.3	2.4	0.0	0.0
	2	8.2	<b>31.7</b>	60.1	0.0	0.0		2	2.9	<b>32.2</b>	64.9	0.0	0.0		2	33.3	<b>43.7</b>	23.0	0.0	0.0
	3	0.9	5.1	<b>84.2</b>	9.8	0.0		3	0.0	0.3	<b>96.3</b>	3.4	0.0		3	7.7	19.9	<b>70.0</b>	2.3	0.0
	4	0.0	2.3	55.8	<b>41.9</b>	0.0		4	0.0	0.0	99.1	<b>0.9</b>	0.0		4	0.7	45.8	34.9	<b>18.7</b>	0.0
	5	0.0	0.0	23.6	76.4	<b>0.0</b>		5	0.0	0.0	100	0.0	<b>0.0</b>		5	0.0	18.2	57.3	24.5	<b>0.0</b>

However, dataset  $U_{2,NH_4}$  and  $U_{3,NH_4}$  show relatively weak results. Later in section 5.6.3 and 5.6.4 it will become obvious, that the prediction quality is too low for practical application.

## 5.6 Comparison to standard reference samples

### 5.6.1 Calculation of the Reference Standard

Regression and classification results are not directly comparable. While regression methods map to a specific value, classification methods map to a domain. To be able to compare the results a reference standard has to be used. Two characteristics of the inflow are important: 1. The amount of substances

(in this section *COD* is used as example) which flowed into the plant in a certain period of time. 2. The variations of inflow concentrations.

To compare regression and classification results it is possible to use the total amount in grams in time period  $T_p = t_2 - t_1$ :

$$COD_g = \sum_{t=t_1}^{t_2} COD(t) \cdot Q(t) \cdot t_s \left[ \frac{mg}{l} \cdot \frac{l}{s} \cdot s \right] [mg] \quad (5.18)$$

where  $t_s$  is the step size of the measurements and  $COD(t)$  the *COD* concentration at time  $t$  and  $Q(t)$  the flow at time  $t$ . Alternatively it is possible to use the mean concentration in  $T_p$ :

$$\overline{COD} = \left( \sum_{t=t_1}^{t_2} COD(t) \right) \cdot n_s^{-1} \quad (5.19)$$

where  $n_s$  is the number of samples in time period  $T_p$ . Using equation (5.19) has the disadvantage that the flow is not considered.

A major challenge is to find a suitable timespan, which is short enough to maintain the dynamics but long enough to make regression and classification comparable. Because it depends on the application whether the dynamics or the total amount of substance is more important there is no optimum answer to this question. For this reason a two hour composite sample, which is commonly used for measurement campaigns to calibrate activated sludge models (ASM) (Henze et al., 2000) is used as reference.

As already described, the inflow of most plants is only equipped with measurement probes for *pH – value*, *temperature*, *flow* and *conductivity*. This is why measurements for *COD* and *NH<sub>4</sub> – N* are usually done in measurement campaigns over a short period of time (one to several days). During these campaigns samples are taken which are analysed in a laboratory. As described above a typical time interval for these samples is two hours. In order to cover the entire inflow in these two hours composite samples should be used. To generate these samples a continuous amount of wastewater is pumped into a container which is changed every two hours. To get a better representation of the inflow it would be necessary to consider the *flow* too (for example by coupling the pump to the inflow measurement). In practice both are often not possible. For this reason it is common practice to make spot tests every two hours.

For this investigation it is assumed that a standard measurement campaign with composite samples was conducted on the Rospe plant. The *COD* and *NH<sub>4</sub> – N* composite samples  $COD_{CS}$  and  $NH_4 - N_{CS}$

are calculated using the *COD* and *NH<sub>4</sub>-N* online measurements respectively. The following equations show the example of *COD*.

$$COD_{CS} = \frac{1}{n_m} \sum_{t=t_0}^{t_0+n_m \cdot t_i} COD(t) \quad (5.20)$$

where  $n_m$  is the number of measurements for the composite sample  $COD_{CS}$  and  $t_i$  the measuring interval of the online probes. For better comparison samples considering the *flow* ( $Q$ ) are calculated in equation (5.21):

$$COD_{CSF} = \frac{1}{nm \cdot \sum_{t=t_0}^{t_0+n_m \cdot t_i} Q(t)} \sum_{t=t_0}^{t_0+n_m \cdot t_i} Q(t) \cdot COD(t) \quad (5.21)$$

Both equations (5.20) and (5.21) are applicable to the regression as well as the classification results.

Figure 5-19 and Figure 5-20 show the comparison between the composite samples (CS) and the composite samples considering the flow (CSF) for *NH<sub>4</sub>-N* and *COD* respectively. For this particular inflow and during the time of the measurement campaign the differences are negligible. Both measurement values show only minor differences (Figure 5-19 and Figure 5-20).

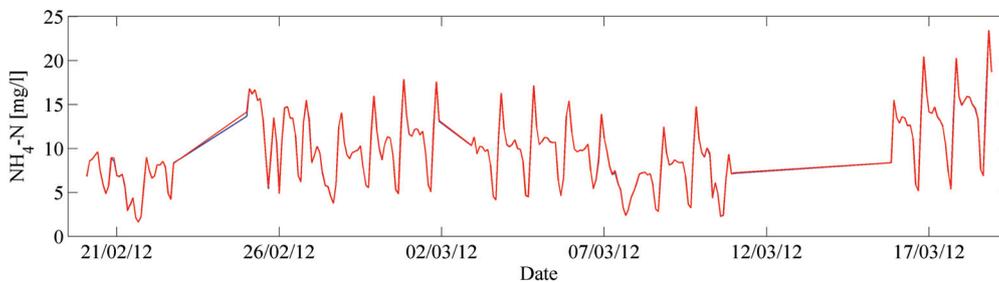


Figure 5-19: Comparison on mean inflow values (*NH<sub>4</sub>-N*) (blue CS, red CSF)

Due to these insignificant differences the simpler CS metric is used in the following comparison of regression and classification.

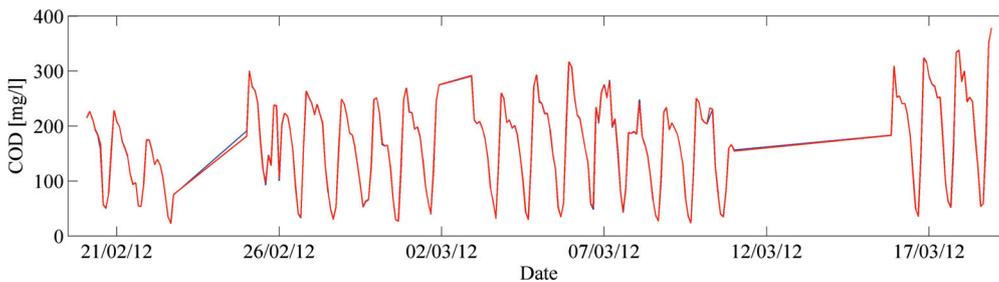


Figure 5-20: Comparison of mean inflow values (*COD*)(blue CS, red CSF)

### 5.6.2 Back Transformation from Classes to Concentrations

For the comparison of classification and regression results, it was necessary to transform the predicted classes back into measured concentrations given in mg/l. Instead of simply using the class mean value  $\bar{C}$ , defined as

$$\bar{C} = \frac{C_{ulim} + C_{llim}}{2} \quad (5.22)$$

where  $C_{ulim}$  and  $C_{llim}$  represent the upper and lower limit of the class respectively, the mean value of all measurement values in a class was used:

$$\bar{C}_{ms} = \left( \sum_{i=1}^N x_i \right) N^{-1} \quad (5.23)$$

where  $N$  is the number of data points in the particular class. The measurement values were taken from the  $COD$  and  $NH_4 - N$  reference measurements.

Table 5-16: Class mean measurement values

Class	$COD$ [mg/l]	$NH_4 - N$ [mg/l]
1	54.6	3.6
2	127.2	6.53
3	176.7	10.9
4	239.9	16.5
5	339.1	22.0

### 5.6.3 Direct Comparison of Regression and Classification Results

After the transformation back into concentration values it is possible to compare regression and classification values directly. Because of the limited number of classes, representing concentration ranges, classification results will most likely perform worse using an error measure like RMSE or NMSE. Therefore a two hour mean, representing a virtual 2h sample, was calculated from all results (classification and regression). The goal is to show that over a period of two hours the differences between classification and regression results are minimal. As these 2h intervals are normally used when performing inflow measurements at WWTP, it is proven to be good enough to preserve the dynamics of the inflow while sufficiently describing the absolute values of the inflow load. Table 5-17 provides a direct comparison of the classification and regression results in terms of NMSE based on these virtual 2h samples.

Table 5-17: NMSE direct comparison of regression and classification results [x100%]

Dataset	RF <sub>class</sub>	LDA <sub>class</sub>	SVM <sub>class</sub>	MLR <sub>reg</sub>	RF <sub>reg</sub>	MLP <sub>reg</sub>	PLS <sub>reg</sub>	SVR <sub>reg</sub>
U <sub>1,COD</sub>	0.074	0.11	0.07	0.02	0.02	0.05	0.02	0.02
U <sub>2,COD</sub>	0.3	1.03	0.35	0.28	0.30	0.90	0.28	0.36
U <sub>3,COD</sub>	0.24	0.5	0.22	0.30	0.16	0.21	0.30	0.21
U <sub>1,NH<sub>4</sub></sub>	0.4	0.6	0.63	0.17	0.47	0.72	0.16	0.65
U <sub>2,NH<sub>4</sub></sub>	0.62	0.73	0.77	0.49	0.86	1.26	0.49	0.83
U <sub>3,NH<sub>4</sub></sub>	0.61	0.86	0.98	0.56	0.70	1.13	0.56	0.90

It is obvious that the regression methods outperform in most cases classification methods, with the exception of dataset U<sub>3,COD</sub> where RF outperforms all other classification methods. Interesting is the fact that RF Classification outperforms RF Regression for dataset U<sub>2,NH<sub>4</sub></sub>. But in this case it has to be considered that both results are very bad. Furthermore it is interesting that SVR is outperformed by RF classification as well as LDA. Important is the fact that for NH<sub>4</sub> - N the additional turbidity probe brings no improvement, while it improves the results for COD significantly. In fact with some methods the NH<sub>4</sub> - N prediction results are even better with the standard measurements used for U<sub>2,NH<sub>4</sub></sub>.

#### 5.6.4 2h mean Comparison of Regression and Classification Results

For the final comparison of classification and regression results, classification results were transformed back to concentrations using equation (5.23). In a second step all results (regression and classification) were transformed according to the described 2h reference standard using equation (5.20). The same was done to the real measured reference values of COD and NH<sub>4</sub> - N. From this point on the results are considered to be virtual 2h composite samples. The final step for the comparison is the application of the NMSE to the different datasets and the reference data respectively.

Table 5-18 gives an overview of the performance of all applied regression and classification methods. The results show that for most cases regression and classification methods achieve similar NMSE values. This effect is obvious for dataset U<sub>1,COD</sub>, U<sub>2,COD</sub> and U<sub>3,COD</sub> in particular. While dataset U<sub>1,COD</sub> achieves good results no matter which method is used, it is the least interesting dataset for practical utilization, due to the fact that in-line SAC<sub>254</sub> probes are in the same price range as in-line COD probes. Significantly more interesting are the results for datasets U<sub>2,COD</sub> and U<sub>3,COD</sub>. For dataset U<sub>2,COD</sub> the best performing classification and overall method is LDA which achieves a NMSE of 0.18, while for dataset U<sub>3,COD</sub> the best classification method with a NMSE of 0.12 is RF<sub>class</sub>.

The overall best result for  $U_{3,COD}$  is achieved by  $RF_{reg}$  with a NMSEs of 0.11. This shows that the turbidity probe improves the overall estimation performance over a period of 2 hours, which could not be seen in Table 5-8. Looking at the  $NH_4 - N$  results only  $U_{1,NH_4}$  is of interest for practical utilization.  $U_{2,NH_4}$  and  $U_{3,NH_4}$  show significant deviations and were not able to follow the dynamics. In particular the height of the peaks could not be predicted.

Table 5-18: NMSE Comparison Results for Test Data: Virtual 2H-Composite Samples [x100%]

Dataset	$RF_{class}$	$LDA_{class}$	$SVM_{class}$	$MLR_{reg}$	$RF_{reg}$	$MLP_{reg}$	$PLS_{reg}$	$SVR_{reg}$
$U_{1,COD}$	0.00	0.01	0.02	0.00	0.01	0.02	0.00	0.00
$U_{2,COD}$	0.24	0.18	1.07	0.23	0.26	0.89	0.23	0.30
$U_{3,COD}$	0.12	0.14	0.39	0.20	0.11	0.16	0.19	0.14
$U_{1,NH_4}$	0.51	0.28	0.48	0.13	0.48	0.70	0.13	0.64
$U_{2,NH_4}$	0.70	0.54	0.64	0.47	0.70	1.20	0.47	0.82
$U_{3,NH_4}$	0.92	0.52	0.77	0.51	0.72	1.15	0.51	0.88

Figure 5-21 shows the comparison results for  $U_{3,COD}$ . The middle part is not considered, due to a data gap. It can be seen that the estimated data not only follows the dynamics of the real measured data but also matches the peaks. Nevertheless, RF has problems in the second half of the validation period, where the  $COD$  levels rise. This can be attributed to the fact that values as high as these were not present in the training data.

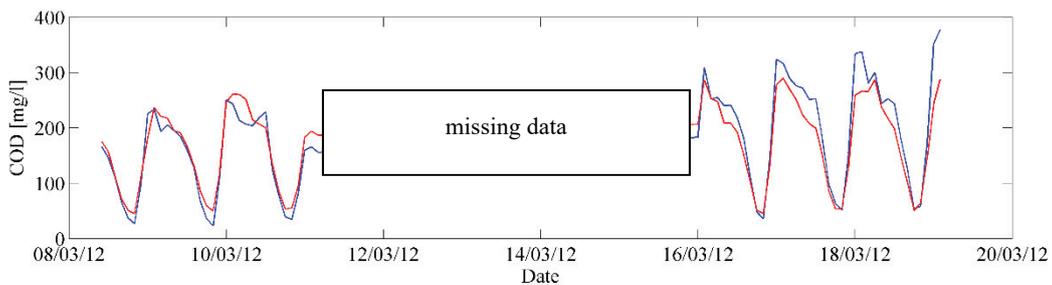


Figure 5-21: RF Classification 2h mean  $U_{3,COD}$  (blue: measured, red: predicted)

Figure 5-22 shows the results for SVR for  $U_{2,COD}$ . While the results are still sufficient in the first half, it is obvious that SVR has problems capturing the peak concentrations. Nevertheless, SVR achieves surprisingly good results in the second half of the validation period, although  $COD$  levels are much higher and values this high are not present in the training data.

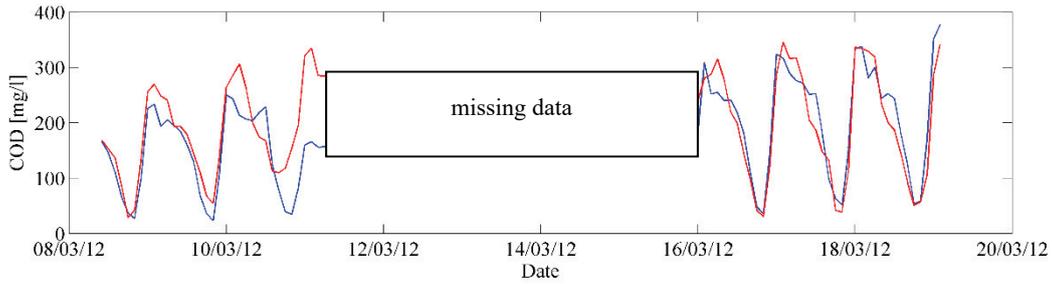


Figure 5-22: Support Vector Regression 2h mean  $U_{2,COD}$  (blue: measured, red: predicted)

The best result for  $NH_4 - N$  presented in Figure 5-23 is achieved for dataset  $U_{1,NH_4}$ . Although  $NH_4 - N$  is considered harder to estimate than  $COD$ , estimation of  $NH_4 - N$  concentrations yields good results. The dynamics are captured well and even peak concentrations are estimated well for the second half of the test data. Estimation for the first half of the test data is slightly worse with peak concentrations underestimated and an offset in predictions relative to the measured values. However, overall these results are of high value for WWTP operation and advanced control.

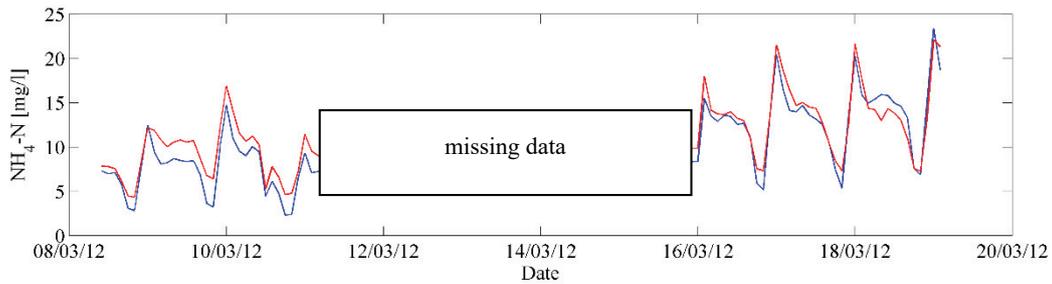


Figure 5-23: Partial Least Squares Regression 2h mean  $U_{1,NH_4}$  (blue: measured, red: predicted)

Figure 5-24 shows the best classification result for dataset  $U_{2,NH_4}$  using LDA with an NMSE of 0.54. It can be seen that the prediction is not able to follow the dynamics or reach the peak values. The reason can be seen in the confusion matrices (Table 5-14). As already described the classification methods tend to map most samples to the middle classes.

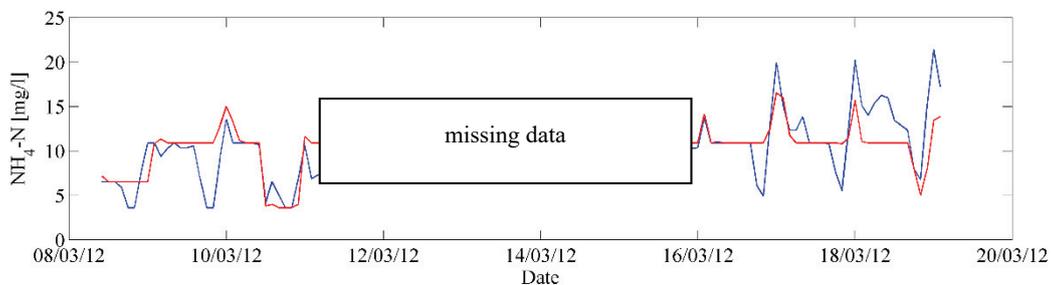


Figure 5-24: LDA 2h mean  $U_{2,NH_4}$  (blue: measured, red: predicted)

In general, it can be seen that for  $U_{2,COD}$  and  $U_{3,COD}$  the non-linear methods perform best, while for  $U_{2,NH_4}$  and  $U_{3,NH_4}$  the linear methods achieve the best results. While regression and classification

performance is similar for easily predictable datasets like  $\mathbf{U}_{1,COD}$ , classification methods achieve better results for datasets that are harder to estimate.

## 5.7 Conclusion

The results show that it is possible to estimate *COD* in the WWTP inflow based on standard measurements with sufficient accuracy for use in optimization and control strategies. Furthermore, it is possible to achieve better prediction results through installation of an additional in-line turbidity probe. The prediction of  $NH_4 - N$  with standard equipment is not feasible due to low estimation accuracy whereas it is possible to achieve sufficiently accurate results, using all input variables. For practical application this means that it is recommended to estimate *COD* based on standard equipment or after installation of an additional in-line turbidity probe, while  $NH_4 - N$  estimation is a valid option for operators who already have an in-line *COD* probe in the WWTP inflow.

In addition, this analysis shows that it is possible to achieve sufficient estimation results using classification instead of regression methods. This not only facilitates the transparent presentation of inflow concentrations to the operator but also allows for specifically adapted training to focus on correct estimation of interesting individual classes.

## 6 Application and Testing of Model-Based inflow estimation

An alternative approach to the methodology presented in chapter 5 to get process data from the WWTP inflow was presented by (Ebel, 2009). While the methodology in chapter 5 pursued the goal of creating soft sensors for important inflow variables, whose data can be used for online control, the idea of a model-based estimation of inflow variables is to determine inflow measurement data retrospectively. The advantage of this approach is, that the optimization of WWTP operation using simulation models requires a broad range of process data representing multiple inflow conditions and, in particular, as many critical operating events as possible. Due to the fact, that inflow measurements are usually done in measurement campaigns over a limited period of time and, as described in chapter 5, permanently installed online measurement devices for *COD* and *NH<sub>4</sub>-N* are not common due to high operating costs. Thus, it is unlikely to capture interesting critical events. Therefore, in order to be able to simulate these scenarios, artificial events are normally generated. However, the composition of these events and their representation of real events have its limitations. In this chapter the retrospective inflow estimation method is applied to the Rospe WWTP model developed in chapter 4 (section 4.2) to generate missing inflow data over a period of several days in order to capture relevant operating states for further optimization and control applications as described in chapter 7. The achieved results clearly show the strong limitations of the method and provide an insight into further necessary improvements of the method.

### 6.1 Introduction to feed stream estimation

The developed methodology requires a fully calibrated simulation model of the WWTP which is used in combination with a non-deterministic optimization method, in this case Particle Swarm Optimization (PSO) (Clerc, 2006a), to estimate the correct feed stream. Basically, the method creates an artificial feed stream  $\mathbf{u} = [u_1, u_2, \dots, u_m]$ , e.g. a combination of *COD* and *NH<sub>4</sub> - N* concentrations and the measured inflow  $Q$ . This artificial feed is then simulated for a period of time  $T_S$  (e.g. 2 hours) using the calibrated model. During that time  $T_S$  the chosen inflow parameters are kept constant. At the end of the simulation period  $T_S$ , the trajectories of the simulated process variables (denoted as  $\mathbf{s}_v(\mathbf{u})$ ),  $v = 1, \dots, n$ , for a set  $V$  of process variables, e.g.  $V = [NH_4 - N, NO_3 - N]$ , at the outflow of the biological treatment, are compared with the trajectories of the measurements  $\mathbf{m}_v$  from the real plant. The result of the comparison of simulated versus measured values is used to create the performance index  $J_v(\mathbf{u}) : E$

$$J_v(\mathbf{u}) = E(\mathbf{m}_v, \mathbf{s}_v(\mathbf{u})) \quad (6.1)$$

where  $E$  is a trajectory matching error metric such as the RMSE. The total performance index is then calculated by aggregation of the individual errors:

$$J(\mathbf{u}) = \sum_{v=1}^n J_v(\mathbf{u}) \quad (6.2)$$

This is repeated by the optimization algorithm using different feed stream combinations for the same period of time  $T_S$  in order to minimize  $J(\mathbf{u})$ . When a minimum is found, this feed stream is used to simulate the model for period  $T_S$  and the process is repeated for the next time period  $T_S + 1$ . The result is a stepwise feed stream (Figure 6-4), which results in the same plant behaviour, as the real feed stream. Therefore, this estimated feed stream can be used in experiments as a surrogate for the real feed stream being a financially feasible approach that does not require extensive long online measurement campaigns. Furthermore, by choosing 2h-feed stream sample intervals the accuracy of the estimation is comparable to the 2h-online fully automated feed stream sampler machines that are commonly used in practice. Figure 6-1 gives an overview of the described methodology.

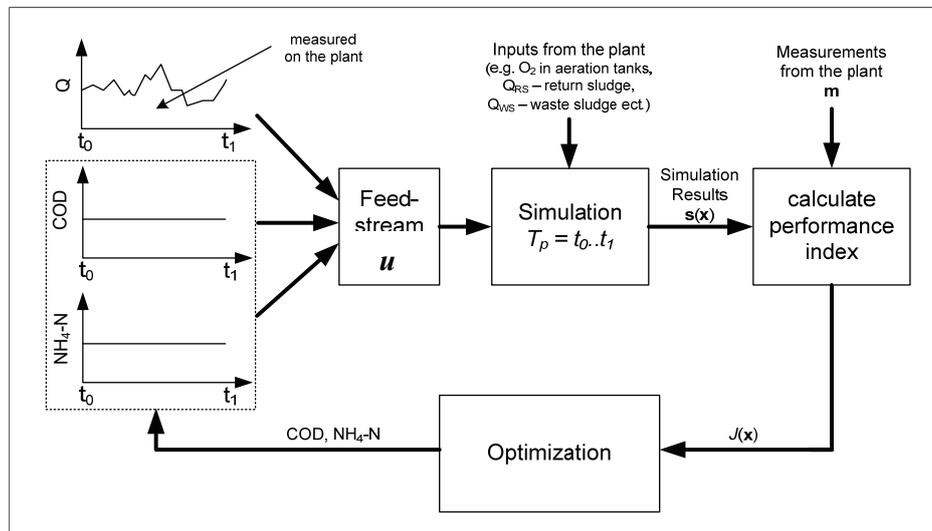


Figure 6-1: Model based inflow estimation principle (Ebel, 2009)

A problem with this approach is the high computational overhead. The dimension  $m$  of the feed stream  $\mathbf{u}$  is directly related to the dimension of the optimization problem. In the basic implementation only two feed stream variables ( $COD$  and  $NH_4-N$ ) need to be estimated because  $Q$  is measured, resulting in a two dimensional optimization problem. However, in order to take account of the dynamics of the WWTP model, and to achieve more accurate inflow estimates, it is necessary to simultaneously optimise two or more time steps  $T_S$  (see Figure 6-2). This is necessary because the current plant output will be a function of the inflow variables over several simulation steps, and not just the current one. However even when time steps  $T_S$  and  $T_{S+1}$  are optimized, which constitute the estimation horizon, only the optimal result for  $T_S$  is used to simulate the next state of the simulation model. This state is then used as the new starting point for the next estimation run with the estimation horizon being  $T_S + 1$  and  $T_S + 2$ . The problem is that estimating two feed stream variables and using  $p$  time steps, the optimization problem

is already  $2p$  dimensional. As the dimension of the optimization problem increases the number of experiments needed also increases rapidly. The experiments with the Rospe model were conducted on a Windows 7 computer, with a i7-2600 processor with four cores running at 3.4 GHz. Using a small simulation step size of one minute for the Rospe WWTP model, it takes up to 120 seconds to simulate a 4 hour simulation (e.g.  $T_S$  and  $T_{S+1}$ ). Using only 60 feed stream combinations to find the optimal one, this is already real time. In order to reduce the number of necessary computer experiments, Ebel (2009) extended the algorithm by introducing a surrogate model (Kriging) to approximate the performance index (see Figure 6-2).

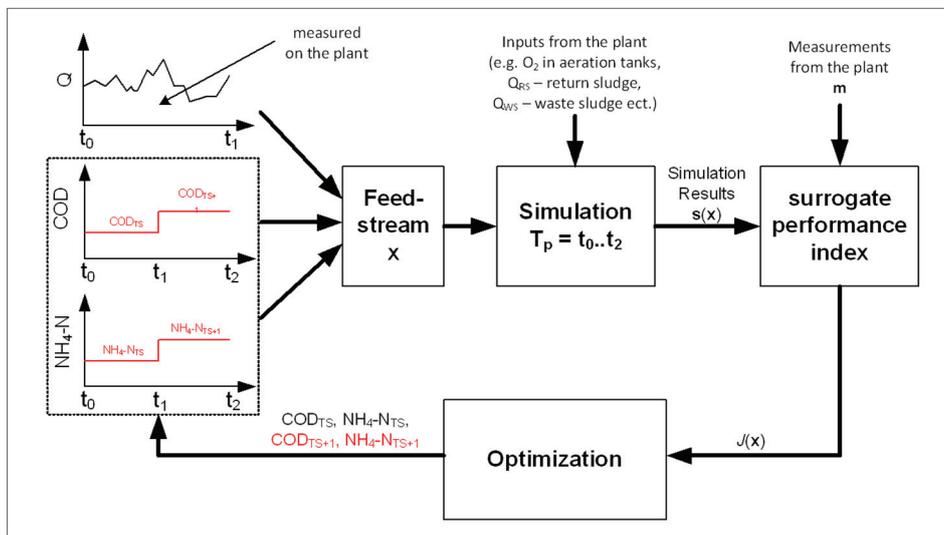


Figure 6-2: Model based inflow estimation principle using an additional surrogate model (Ebel, 2009)

To create the surrogate model, a small number of feed stream concentration vectors are created using Latin Hypercube Sampling (LHS) (McKay et al., 1979) and applied to the full dynamic simulation model. LHS splits each dimension of the search space into intervals of equal length. Then, for each interval in every dimension one value is randomly generated. In the final step, these values from each dimension are randomly brought together in pairs. The result is a sparse sample distribution in the search space. Using these inputs and their corresponding  $J(\mathbf{u})$  values, a Kriging surrogate model  $J_k(\mathbf{u})$ , is built. Applying PSO to  $J_k(\mathbf{u})$  instead of  $J(\mathbf{u})$  yields significantly faster computation of the optimum  $\mathbf{u}^*$ , as the number of expensive WWTP simulation runs is substantially reduced. A detailed description, testing different optimization algorithms, error measures and the number of necessary computer experiments is given in (Ebel, 2009).

## 6.2 Methodology limitations

In order to apply retrospective inflow estimation to the Rospe model, consideration needs to be given to the limitations of the methodology and challenges associated with using it in practice. The following sections describe three scenarios, which can lead to unsatisfactory estimation results when the method is applied to certain plant designs.

### 6.2.1 Effects based on retention time

The Rospe plant is equipped with measurement devices for  $NH_4 - N$  and  $NO_3 - N$  at the end of the biological treatment (tank D4). Due to the fact that these measurement devices are the only ones available, it means that the wastewater has to pass through the primary clarifier with a volume of  $550 \text{ m}^3$  and through the 18 (reduced to 7 in the Rospe model) nitrification and denitrification tanks with a total volume of  $7,215 \text{ m}^3$ . The Rospe plant is continuously operated with upstream de-nitrification. Mechanical purification is carried out in the form of a 6 mm grate followed by a sand filter. Afterwards, the water passes through a horizontal flow primary treatment tank. The biological stage consists of a total of 18 tanks into which the wastewater flows successively. The structure of the plant is laid out in 3 lanes. Finally, there is the final treatment stage that consists of three lanes connected in parallel. Table 4-4 gives an overview of the different volumes.

Assuming a mean inflow of  $128 \text{ l/s}$  ( $11,131 \text{ m}^3/\text{d}$ ), it needs 72 minutes to pass through the primary clarifier. The retention time for the biological treatment is calculated as

$$\frac{7,215 \text{ m}^3}{11,131 \frac{\text{m}^3}{\text{d}}} = 0.65 \text{ d} = 15.6 \text{ h} \quad (6.3)$$

Hence, the combined retention time of the primary clarifier and the biological treatment stages is 16.8 hours. Therefore, if the plant would work like a plug flow system, the water from the feed stream  $u$  would reach the location of the online measurement probes in the system after approximately 17 hours. In reality, this delay is much shorter due to mainly two reasons:

- The internal flow of the plant is higher, because of internal recirculation.
- The return sludge from the final clarification increases the internal flow.
- The tanks are completely mixed (in the model)

Figure 6-3 shows the normalized step response of the Rospe model after the primary clarification (PC) and at the tank D4 to a step in  $NH_4 - N$  from  $9.5 \text{ mg/l}$  to  $15 \text{ mg/l}$ . Using a constant inflow flow of  $128 \text{ l/s}$ . The other inflow parameters were kept constant. Prior to the test the model was stabilized over a period of 200 days with mean inflow values.

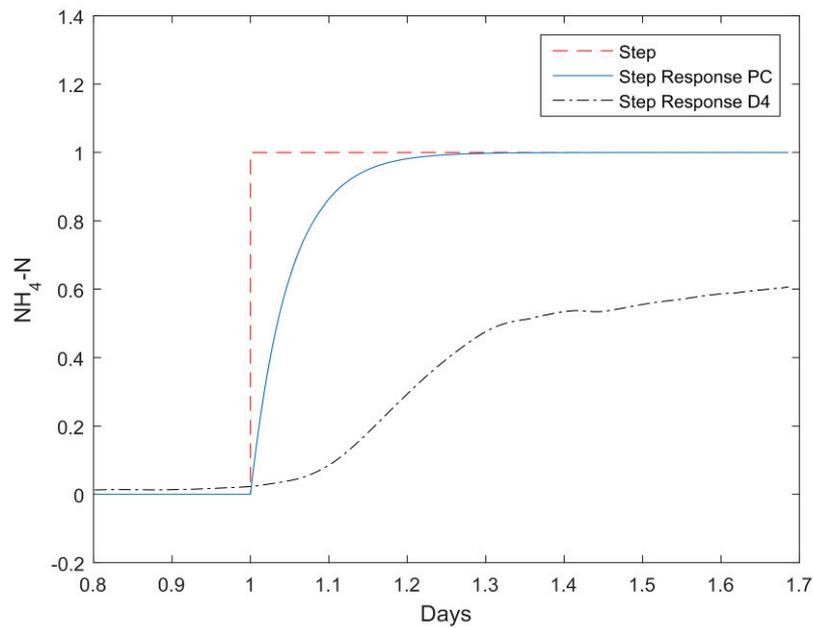


Figure 6-3: Step response of the Rospe model to a  $NH_4-N$  step from 9.5 mg/l to 15 mg/l (scaled)

It can be seen, that the model reacts much faster, than one might expect from the calculated retention times. After approximately 0.1 days (2.4 hours) the  $NH_4-N$  concentration starts to rise at tank D4. After 0.3 days (7.2 hours) 90% of the maximum  $NH_4-N$  concentration is reached. Furthermore, the  $NH_4-N$  degradation decreases the height of the step response in tank D4. Thus, a simulation period for the feed stream of 4 to 6 hours should be sufficient for the water to reach the online measurement devices. Depending on the plant used, two undesired effects can occur when using the described approach for feed stream estimation.

**Undesired hydraulic retention time.** Depending on the size of the primary clarifier and the following tank volumes, the hydraulic retention time cuts off a part of the feed stream which might be stored in the clarifier. Using for example two time steps  $T_S$  and  $T_{S+1}$ , each of two hour duration, time step  $T_{S+1}$  is lost, because the water is still “stuck” in the primary clarifier and has no effect on the performance index  $J(\mathbf{u})$ . To give an example, during the night or in the early morning, many WWTP have a low inflow with low concentrations. During these times, the inflow water stays in the primary clarifier for several hours without reaching the biological treatment stage. For the given example this means, that a four dimensional  $\mathbf{u} = [NH_4 - N_{T_S}, NH_4 - N_{T_{S+1}}, COD_{T_S}, COD_{T_{S+1}}]$  optimization problem has to be solved, although only two of the variables have an effect on the performance index  $J(\mathbf{u})$  which is considering only the concentrations in the biology over the estimation horizon, which is not long enough to see the effect of the  $T_{S+1}$  samples. Thus, section 6.4 describes an alternative approach which avoids this problem and saves computational effort.

**Undesired blending of feed stream steps.** In cases, when the time for the inflow water to reach the measurement devices is longer than the time of one time step and the plant has internal flows, e.g. the recirculation of plants with upstream denitrification, the feed stream of the second and later steps will mix up with the first step. The following example should explain the problem. Assume that only the ammonium concentration in the inflow has to be estimated. Two time steps are created:

$$\begin{aligned} T_S : \quad NH_4 - N &= 0 \\ T_{S+1} : \quad NH_4 - N &= 10 \end{aligned} \quad (6.4)$$

Depending on the inflow  $Q$  a trajectory of  $NH_4 - N$  that lies between 0 mg/l and 10 mg/l is the result after the primary clarification (PC). The start point of the trajectory depends on the  $NH_4 - N$  concentration in the PC at the beginning of  $T_S$ . The trajectory which reaches tank D4 is further deformed. The important aspect is that the trajectory is influenced by  $T_{S+1}$ . Due to the blending of the feed stream, different combinations for  $T_S$  and  $T_{S+1}$  can lead to the same result. In the given example this means that the feed stream is 0 mg/l for  $T_S$ , while the minimum of  $J(\mathbf{u})$  is determined on a blend of 0 mg/l and 10 mg/l. Over time, this optimization will compensate this effect up to a point, when the process values used to calculate  $J(\mathbf{u})$  start to deviate. The influence of these effects depends on the size of the plant, the design, the inflow and the location of the measurement devices used to calculate  $J(\mathbf{u})$ .

While Ebel used the model of an intermittent operated WWTP with cascaded tanks (Odenthal WWTP) for his experiments, the Rospe plant is continuously aerated with an upstream denitrification. Due to the higher internal flows, in the Rospe plant the described negative effects are stronger. Furthermore, the Odentahl WWTP is equipped with  $NH_4 - N$  and  $NO_3 - N$  measurement devices at the outflow of each cascade. Therefore, the delay in measuring the inflow is much shorter.

### 6.2.2 Drifting of the state vector

The reaction of the model to a feed stream  $\mathbf{u}$ , as described in Chapter 4, depends on the internal state of the tanks based on the ASM1 model. This state is defined by the concentration of the different biomass fractions and chemical components, using a state vector  $\mathbf{x}$ :

$$\mathbf{x} = [S_I, S_S; X_I, X_S, X_{B,H}, X_{B,A}, X_P, S_O, S_{NO}, S_{NH}, S_{ND}, X_{ND}, S_{ALK}] \quad (6.5)$$

The state vector  $\mathbf{x}$  changes according to the flow and the availability of nutrients. The processes involved are described in Table 4-1. It is concluded that the model reacts differently to a feed stream  $\mathbf{u}$ , depending on the state vector  $\mathbf{x}$ . This is the reason why models have to reach a steady state, which can then be used as a starting point for future simulations. An internal state which differs from the real state, will result

in a wrong model reaction. For example if  $X_{B,H}$  is too high,  $NO_3 - N$  (model variable  $S_{NO}$ ) will decrease too quickly and too much  $COD$  is consumed. With regard to the feed stream estimation, this effect can become stronger due to a concentration built up. A small error in the feed stream  $\mathbf{u}$  will cause the state vector  $\mathbf{x}$  to deviate from the real vector. In the next simulation step, this will lead to a slightly wrong model reaction. Due to the fact, that the optimization algorithm searches for a minimum using the performance vector  $J(\mathbf{u})$ , the next estimated feed stream  $\mathbf{u}_{T_s+1}$  has to deviate from the real inflow in order to achieve the same model reaction. After a certain amount of steps, depending on the model and measurement quality, it will be necessary to present an inflow stream, which is far from the real inflow stream in order to get the same model reaction. The same effect can be caused if the measurements used to calculate  $J(\mathbf{u})$  are corrupted.

### 6.3 Application of feed stream estimation to the Rospe model

#### 6.3.1 Model adaption

To facilitate the application of the feed stream estimation methodology, the Rospe model that was presented in chapter 4, was extended by the inclusion of an additional “pattern generator” function block, which sits between the inflow block and the primary clarifier. The block bypasses the desired measured variables and replaces the variables which are being estimated with the estimated feed stream  $\mathbf{u}$ . For the estimation of  $NH_4 - N$ , the 10<sup>th</sup> element of the input vector  $S_{NH}$ , the model variable for  $NH_4 - N$ , is replaced by its estimate. Due to the fact that  $COD$  is represented in the model by  $S_I, S_S; X_I, X_S, X_{B,H}, X_{B,A}, X_P, S_{ND}$  and  $X_{ND}$  the fraction block (see Figure 4-9) is used to split the  $COD$  estimate into the necessary fractions. With  $COD$  and  $NH_4 - N$  estimated, only the flow is used from the inflow block.

#### 6.3.2 Data generation and model stabilization

As preparation, the model is stabilized using mean inflow values over a period of 200 days followed by five repetitions of the real dynamic inflow values gathered during the measuring campaign (see chapter 5, section 5.3). After that the model is simulated until day 8 and the model state is saved. Due to the high computational effort associated with conducting the experiments a limited period of time, from day 8 to day 11, is chosen for the inflow estimation evaluation.

In order to eliminate errors from the measurement devices and from the model, instead of using the recorded process data  $\mathbf{m}_v$  from the measurement campaign, the targeted time period from day 8 to day 11 is simulated and the process values  $S_{NH}$  and  $S_{NO}$ , the model variables corresponding to  $NH_4 - N$  and  $NO_3 - N$ , were recorded. To indicate, that these values are from the model, they are designated as

$m_s$ . This is done to ensure, that possible problems are related to the method and the application and not to measurement errors.

### 6.3.3 Parameterization of the Feed Stream Algorithm

The following section gives an overview of the most important parameters used for the different algorithms. Table 6-1 describes the parameters used for the feed stream algorithm itself, as well as the estimated inflow variables and the variables used to calculate the performance.

Table 6-1: Parameterization of the feed stream algorithm

Feed Stream Estimation	
Number of Steps	2
Step Size (minutes)	120
Estimated Feed Stream variables	$COD$ , $NH_4 - N$
Comparison variables for performance calculation	$NH_4 - N @ D4$ , $NO_3 - N @ D4$
Error measure for performance calculation	RMSE
Number of Evaluation runs	30-50
Simulation period (days of the measurement campaign)	8-11 (3 days)

For the Kriging model, that is used as surrogate model for the performance function  $J$ , which is a combination of a regression and correlation model (see Table 6-2), the DACE Matlab toolbox (Lophaven et al., 2002) is used. The trained Kriging model  $J_k(\mathbf{u})$  approximates the performance index  $J(\mathbf{u})$ . The parameters are the standard parameters proposed in (Ebel, 2009). For the regression function a linear model is chosen and for the correlation function a Gaussian function. The number of so called design sites, which are support points for the model, is equal to the number of evaluation runs, because each evaluation run / experiment results in one value of the performance index.

Table 6-2: Parameterization of the Kriging model

Kriging Parameter	
Number of design sites	30-50
Correlation function	Gaussian
Regression function	Zero order polynomial

Particle Swarm Optimization (PSO), a non-deterministic global optimization algorithm, is used to find the minimum of the Kriging approximation  $J_k$ . The implementation used is the one provided in the Matlab PSOT toolbox (Birge, 2003), and the algorithm settings were chosen as given in Table 6.3.

Table 6-3: Parameterization of the PSO algorithm

PSO Parameter	
Population size	100
Maximum number of training epochs	2000

### 6.4 Optimization Possibilities

This section proposes an approach to reduce the dimension of the optimization problem. As described in section 6.2.1, depending on the plant design and the inflow conditions, it can happen that the latest step of the estimated horizon of the feed stream has no effect on the model, because the time is too short for the feed stream to reach the plants’ sensors. To solve this problem, the original feed stream estimation is able to generate several steps  $T_S, T_{S+1}, \dots, T_{S+n}$ . But, as the dimension of the search space is equal to the number of steps multiplied by the number of estimated variables, this, in the worst case, results in a high dimensional search space. This is necessary to consider the hydraulics of the plant. Figure 6-4 shows in red the original feed stream estimation principle for three steps, which for the estimation of two parameters ( $COD$  and  $NH_4-N$ ) results in a six dimensional search space. A simple solution for this problem of high dimensional search spaces is depicted in blue in Figure 6-4. Rather than using three steps over the optimisation horizon, the modified version uses only one step over the  $t_0$  to  $t_3$  horizon, reducing the search space by a factor of three. When the minimum of  $J_k(\mathbf{u})$  is found, the next optimization starts at  $t_1$ . The periods  $t_1-t_2$  to  $t_2-t_3$  allow for the transport delay to the sensors, but without increasing the search space. By employing this modification the retention time problem outlined previously can be addressed. The modified approach also has some advantages with regard to the blending problem, since only one value is selected over the three steps, it is not possible to choose extremely different values from one step to the next.

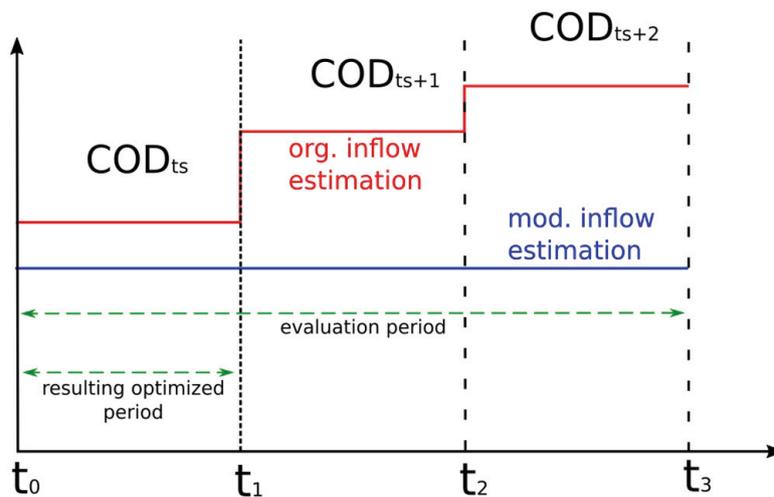


Figure 6-4: Alternative feed stream estimation for dimension reduction (red: original feed stream estimation, blue: modified feed stream estimation)

Table 6-4 shows the parameters for the modified version of the inflow estimation. Using this alternative approach with only one constant concentration over the estimation horizon, the dimensionality of the problem is significantly reduced while considering higher retention times.

**Table 6-4: Configuration of the modified feed stream estimation**

Feed Stream Estimation	
<b>Step Size (minutes)</b>	240-360
<b>Resulting optimized time (minutes)</b>	120
<b>Estimated Feed Stream variables</b>	<i>COD, NH<sub>4</sub>-N</i>
<b>Comparison variables for performance calculation</b>	<i>NH<sub>4</sub>-N @ D4, NO<sub>3</sub>-N @ D4</i>
<b>Error measure for performance calculation</b>	RMSE
<b>Number of Evaluation runs</b>	30-50

## 6.5 Results of the feed stream estimation

To test the feed stream estimation algorithm different settings for the estimation horizon were used varying between 240 and 360 minutes. The resulting optimized inflow step size, which is used, after a minimum is found, is always kept at 120 minutes. The result is an inflow that consists of steps of 120 minutes for *COD* and *NH<sub>4</sub>-N*. Both methods, the original one using several equally long steps, as well as the modified one using only one step with different estimation horizons longer than 120 minutes were tested.

The original version of the inflow estimation is tested with two time steps each 120 minutes in duration (resulting in a 240 minutes estimation horizon) and the modified version uses estimation horizons of 240 and 360 minutes with one step keeping the concentrations constant. Furthermore, different combinations of process variables are tested to calculate the performance index. The process variables used for comparison were always measured in tank D4 of the Rospe plant model, which is the location of the measurement devices in the real plant. In order to determine, if the method is applicable to the real plant without investing in new measurement equipment, no artificially created measurements from the simulation model were considered for the experiments to improve estimation quality. Although the inflow conditions during the tests include no extreme weather events such as rain, the algorithm performed poorly for most experiments. For this reason, the following section shows only excerpt results, which describe the algorithm behaviour.

Figure 6-5 shows the comparison of the process variables for an experiment with the original version of the algorithm using two 120 minutes steps (240 minutes estimation horizon). While the *NH<sub>4</sub>-N* concentrations in tank D4 are in the correct range, they still show a different dynamic behaviour, than the measured concentrations and the estimates are not able to capture the real concentration peaks. However, the *NO<sub>3</sub>-N* concentrations are completely uncorrelated to the measured values. This shows

that the algorithm is clearly not able to keep the  $NO_3-N$  concentration stable which is why it starts with the correct concentration of 9 mg/l  $NO_3-N$  and falls down to 2 mg/l.

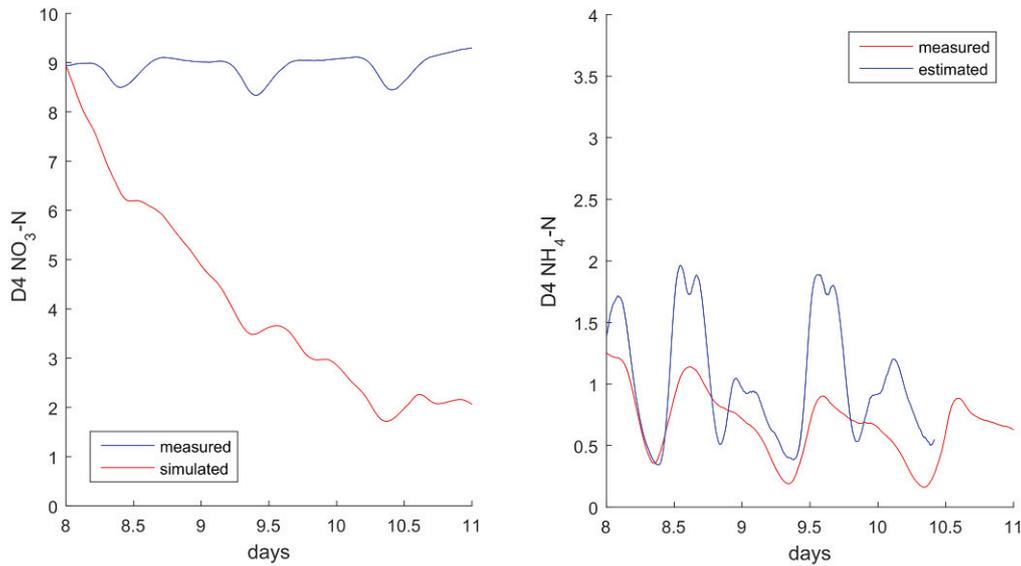


Figure 6-5: Inflow Estimation (original version) comparison variables  $NO_3-N$  and  $NH_4-N$  using two steps of 120 minutes (optimization horizon 240 minutes)

Figure 6-6 shows the corresponding inflow estimation. It is obvious, that the estimated values have no correlation with the real inflow whatsoever. The estimated  $COD$  as well as the estimated  $NH_4-N$  only jump between the boundaries of the optimization algorithm.

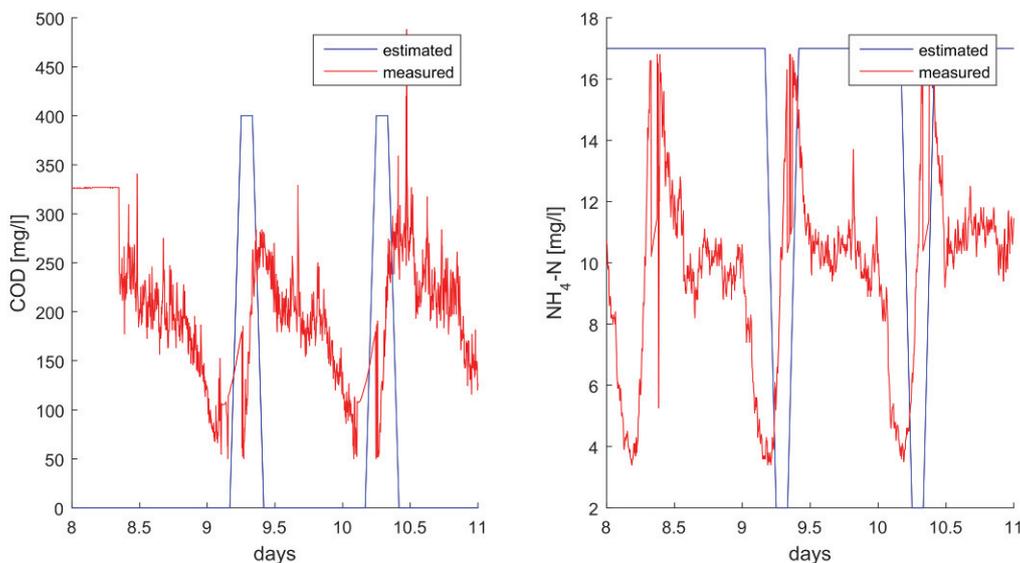
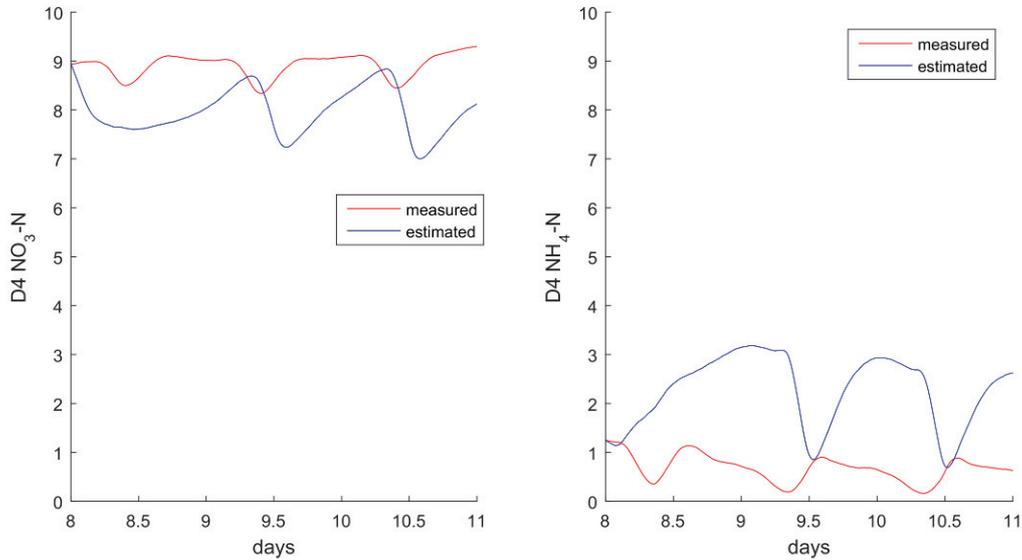


Figure 6-6: Inflow Estimation for  $COD$  and  $NH_4-N$  using 2 steps a 120 minutes (240 minutes optimization horizon) and  $NH_4-N$  and  $NO_3-N$  at tank D4 for comparison

While most of the experiments failed to achieve usable results, one estimation run achieved reasonable results. Figure 6-7 shows the comparison variables of an estimation using the modified version with an

estimation horizon of 360 minutes. While the trajectories still do not fit perfectly, they are in the correct range.



**Figure 6-7: Inflow Estimation (modified version) comparison variables NO<sub>3</sub>-N and NH<sub>4</sub>-N using 360 minute's optimization horizon**

The corresponding feed stream estimation when compared to the previous experiments looks reasonably good. While *COD* is very noisy, the mean of the estimated *COD* is 171.32 mg/l, while the mean value of the measured value is 203.2 mg/l. Moreover, the mean value of the estimated *NH<sub>4</sub>-N* is 10.47 mg/l, while the mean of the estimated *NH<sub>4</sub>-N* is 9.7 mg/l. Due to the high noise on the *COD* estimation the RMSE for *COD* is very poor with 119.4 mg/l. On the contrary, the *NH<sub>4</sub>-N* RMSE is 2.43 mg/l, which is an acceptable result.

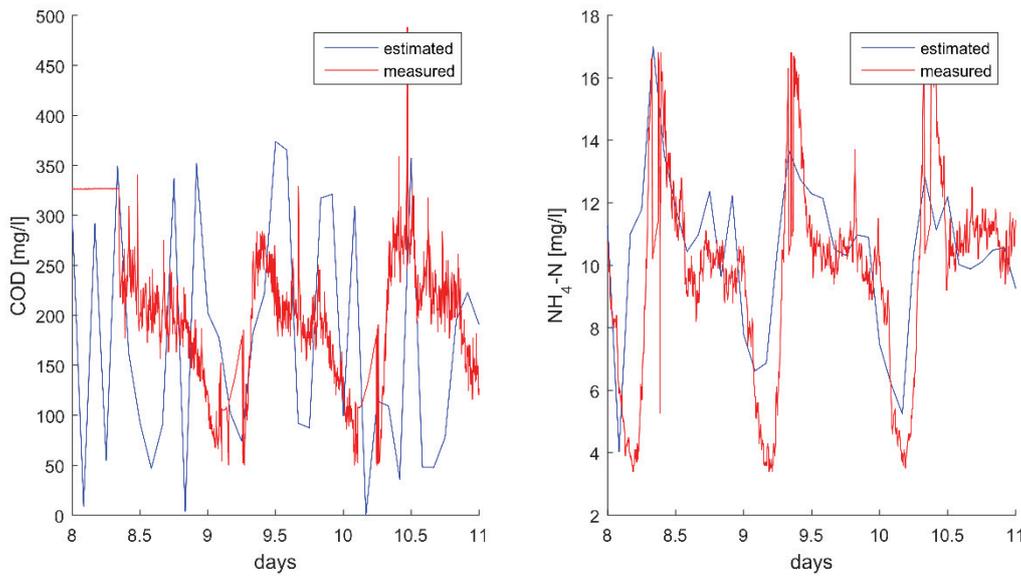


Figure 6-8: Inflow Estimation (modified version) for  $COD$  and  $NH_4-N$  using 360 minutes optimization horizon and  $NH_4-N$  and  $NO_3-N$  at tank D4 for comparison

The better results for the  $NH_4-N$  measurement could have been expected, because the comparison variables both represent forms of nitrogen. The  $COD$  estimation proves to be more difficult, because it indirectly depends on the  $COD$  available for  $NO_3-N$  degradation.

## 6.6 Conclusion

Overall the retrospective inflow estimation methodology performed very poorly for the Rospe WWTP model as most results show no correlation with the real inflow values. Only the inflow estimates of  $NH_4-N$  using an optimization horizon of 6 hours achieved reasonable results. It is clear that the approach is not able to offer a viable alternative to artificial inflow events for this application. The  $NH_4-N$  results of the last experiment show, that an improvement is possible, but further research is needed to address the impact of high and variable retention times and recirculation in order to achieve reasonable results.

The main reason for the poor results on the Rospe WWTP is most likely the variable retention time, which makes an evaluation of the possible estimates based on the measured values during the same estimation horizon nearly impossible. Different optimization strategies considering varying dead times in the model and longer estimation horizons should be tested in future work in order to thoroughly assess the applicability of the method.

Furthermore, there is scope to improve the performance function  $J(\mathbf{u})$ . The high noise on the  $COD$  inflow suggests that the optimization algorithm has problems finding an optimal solution. The reasons for that can be multifaceted. Multiple local minima or an extremely flat performance function require a very high number of experiments in order to allow the Kriging surrogate model to give a sufficiently

accurate representation of the original performance function. Furthermore, the effect of blending needs to be incorporated.

One practical problem, which has to be solved is the high computational effort. The experiments for this evaluation took several weeks of computing time. Therefore, ways have to be found to speed up the simulation time.

With regard to the initial goal of investigating an alternative to the soft sensor methodology developed in chapter 5, it has to be concluded that retrospective inflow estimation in its current form is not a viable proposition. However, it does represent a first step in the right direction. The poor quality of the results show that the dynamics of the plant have to be addressed specifically in order to get good results, whereas the Machine Learning approach considered in chapter 5 seems to be able to achieve this mostly automatically. Therefore, with regard to the aim of this thesis to find and assess possibilities for improved WWTP operation and control, it is obvious that inflow estimation is highly complex and strongly dependent on individual WWTP characteristics.

## 7 Self-Organizing Map based operation regime estimation for state based control of Wastewater Treatment Plants

Optimal control of WWTPs is often hindered by a lack of knowledge about the bio-chemical state of the bioreactors. As many process variables of a WWTP are difficult to measure online, the development of an efficient control strategy is a great challenge. This chapter presents an approach, which combines the use of Self-Organizing Maps (SOM) and a clustering algorithm to identify operational patterns in WWTP process data. These patterns provide a basis for the optimization of controller set points that are well suited to the previously identified operating regimes of the plant. The optimization is performed using Genetic Algorithms (GA). The approach is tested and validated on two test cases using (1) a modified version of the Benchmark Simulation Model No.1 (BSM1) (Alex et al., 2008) and (2) the Rospe model described in section 4.2. The results of this state-based control strategy indicate that the presented methodology is a promising and useful control strategy that is able to address and resolve the individual energy and effluent limit challenges faced by WWTP operators.

### 7.1 Introduction

The development of an optimal controller for wastewater treatment plants has to consider changes in the bio-chemical state of the bioreactors. The ASM1 model describes this state, as a vector with 13 elements, representing the different biomass fractions and chemical components with a 14<sup>th</sup> element, the inflow rate  $Q$ . In practice, online measurement devices for  $O_2$ ,  $NH_4 - N$ ,  $NO_3 - N$ ,  $COD$  and  $Q$  are available for reasonable prices, whereas spectrometric  $COD$  probes are hardly ever used and when they are, they are typically installed in the inflow or the outflow of a plant instead of the bioreactors. This means that only three of 13 elements of the state vector are available to control the process. As described in section 3.2.1 most controllers use only  $NH_4 - N$  or  $O_2$  as input. This shows that only a small fraction of the real state is used, while the other state variables are largely unknown. The idea of operation regime estimation using SOM is that different regions on the trained SOM offer, depending on the process values used to train the map, a better representation of the plant's state, than single process values such as  $NH_4 - N$ .

### 7.2 Self-Organizing Maps

SOM is a special kind of artificial neural network, whose training is unsupervised and which has properties of vector quantization and vector projection algorithms (Kohonen, 1995). Thus, SOM reduces multidimensional data into a much lower-dimensional map space that usually has two dimensions. Matching input data to neurons with similar weight vectors, SOM limits the number of different data samples to the number of neurons on the map. This method is called vector quantization (Kohonen, 1995). In addition, it performs a vector projection, which creates a topology preserving the high dimensional input space by mapping it to the two dimensional output space usually referred to as the

map. This means, that data that are close together in the high dimensional input space are mapped into spatially close areas on the map and elements which are spatially close on the map should be similar in the input space (Kohonen, 1995).

The SOM algorithm can be divided into three parts: (1) architecture, (2) initialization and (3) training. The created map consists of artificial neurons that are randomly created and fitted to the lattice of the map. Each neuron has a vector of weights  $\mathbf{m}_i = [m_{i,1}, \dots, m_{i,n}]$ ,  $i = 1, \dots, K$  where  $n$  is the dimension of the weight / input vector and  $K$  is the number of the map nodes (neurons). Each input vector  $\mathbf{x}_j = [x_{j,1}, \dots, x_{j,n}]$ ,  $j = 1, \dots, P$  where  $P$  is the size of the input dataset that is mapped to exactly one neuron  $K$  during training.

### 7.2.1 SOM Initialization and Training

The initialization is done by randomly initializing the network with uniformly distributed values or by sample initialization with random samples drawn from the training set (Engelbrecht, 2002). During training, each single neuron is activated and the best matching unit (BMU) to the input vector  $\mathbf{x}_j$  is determined by a distance measure. For this work, Euclidean distance is used.

$$d(\mathbf{x}_j, \mathbf{m}_i) = \sqrt{\sum_{k=1}^n (x_{jk} - m_{ik})^2} . \quad (7.1)$$

An adaption rule, known as the Kohonen rule, for the neuron weight  $\mathbf{m}_i$  is defined by:

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + h_{ci} [\mathbf{x}_j(t) - \mathbf{m}_i(t)] \quad (7.2)$$

where  $h_{ci}$  is the neighbourhood of the BMU  $\mathbf{m}_{ci}$  at time  $t$  and it defines the region of influence that the input sample  $\mathbf{x}_j$  has on the SOM. The size of the neighbourhood  $NS$  is altered through the training from an initial size (default  $NS = 3$ ) down to 1. It is during this phase that neuron weights order themselves in the input space consistent with the associated neuron positions. When  $NS \leq 1$  only the BMU is influenced. Using neighbourhood sizes of 1 or below is also called tuning, due to the fact that the map topology is retained but the neurons are better fitted to the input data. For this work the so-called bubble neighbourhood function was used. It allows that all neurons in the neighbourhood  $h_{ci}$  are influenced in the same way. Other neighbourhood functions influence the neurons in the neighbourhood differently depending on the Euclidean distance to the BMU.

Regarding the control application, SOM has several advantages compared to other methods. First, SOM is relatively insensitive to non-equally distributed data, which is important if certain operating regimes appear less frequently. Due to the fact that similar data will always hit the same BMU, all operating conditions are still represented by their BMUs and not masked as done by other clustering algorithms.

For the work at hand two different SOM software packages were used for test cases (1) and (2). The Java SOMToolbox from the Technical University Vienna (Vienna University of Technology, 2010) was used for test case (1) and the Matlab Neural Network toolbox (Hagan et. al, 2002) was used for test case (2) as the integration into the Rospe simulation model has proven to be easier to handle. However, most of the error measures used for optimizing the SOM are not implemented in the Matlab toolbox and had to be developed.

## 7.2.2 SOM Validation

To evaluate the quality of a SOM two main aspects are considered: (1) quantisation quality and (2) topological preservation.

### 7.2.2.1 Quantisation Error

The quantisation quality can be measured using the quantisation error  $E_Q$  which is defined as the sum of Euclidean distances of all input vectors to the weight vectors of their best matching unit BMU (also known as winning neurons) divided by the number of input samples. This is defined as

$$E_Q = \frac{1}{P} \sum_{j=1}^P d(\mathbf{x}_j, \mathbf{m}_{c_j}) \quad (7.3)$$

where  $\mathbf{m}_{c_j}$  is the BMU for the corresponding input sample  $\mathbf{x}_j$ . This means that a small  $E_Q$  represents a better fitting of the neural map to the data. Using this error two aspects have to be taken into consideration. (1) The quantisation error offers no information about the topologic quality of the map, which means that neurons with similar weight vectors can be located in different regions of the map. For the task at hand this plays a minor role due to the fact that the second clustering step is applied to the input space. Therefore the location on the 2-D map is not important. (2) With the size of the map (number of neurons) it is likely that the error will decrease. The reason is, that the likelihood that an input sample is close to its BMU rises with the number of neurons on the map. If the number of neurons is equal or higher as the number of input samples, it would be possible to have a perfectly fitting neuron for each input sample, which would result in a quantisation error  $E_Q = 0$ .

### 7.2.2.2 Topological Error

The topological error  $E_T$  defines the projection quality from the input dimension to the 2-D map.  $E_T$  is calculated as follows. For each input data vector, the nearest weight vector (BMU) and the second

nearest weight vector (second BMU) is computed using Euclidean distance. If they are not adjacent on the SOM-grid, this is counted as a local error. The global topographic error measure is then computed by counting the number of local errors and divided by the overall number of data samples (Uriarte and Martín, 2005). The topographic error is given by

$$E_T = \frac{1}{P} \sum_{j=1}^P u(\mathbf{x}_j), \quad (7.4)$$

where the function  $u(\mathbf{x}_j)$  is 0 if the first and second BMUs of the data vectors  $\mathbf{x}_j$  are adjacent and 1 otherwise. Using this error it has to be kept in mind that depending on the map topology different numbers of neurons are considered adjacent. For example for a rectangular grid 4 neurons are adjacent, while on a hexagonal grid 6 neurons are adjacent. This will lead to smaller errors on maps with hexagonal topology. Figure 7-1 shows the different behaviour of the rectangular and hexagonal maps.

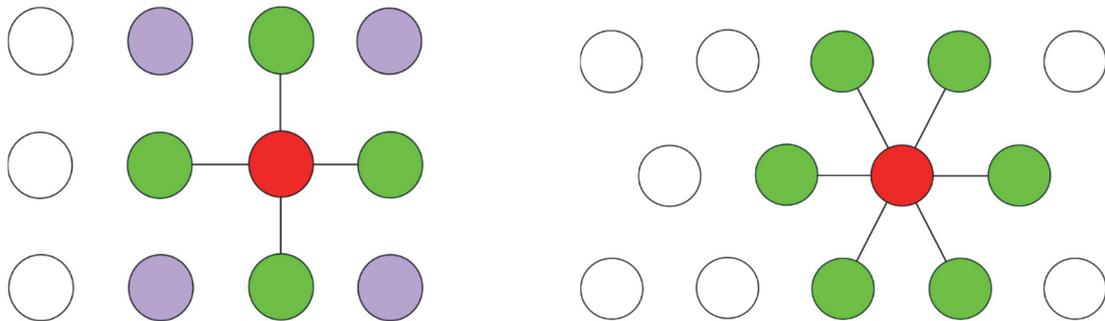


Figure 7-1: Adjacent neurons on rectangular (left) and hexagonal (right) grids on SOMs (BMU – red, adjacent neurons – green, corner neurons – violet, not adjacent neurons – white)

Uriata (Uriarte and Martín, 2005) proposes a methodology which offers a possibility to improve the topological error by extending  $u(\mathbf{x}_j)$  so that it counts the corner neurons as 0.5. For the optimization of the SOM, these errors are used as quality measures.

### 7.3 Clustering the SOM

For the purpose of this work, it is important to create bigger clusters from the SOM to group similar operation regimes together. Different algorithms such as single linkage, complete linkage, centroid linkage, ward linkage and k-means (MacKay, 2003; MacQueen, 1967) can be applied on the SOM to create such bigger clusters. From experiments conducted, ward linkage fared better than other algorithms. Ward linkage aims to keep the variance between the clusters as small as possible. Mathematically, the distance in ward linkage clustering is computed as

$$d(C_1, C_2) = \frac{\|\mathbf{c}_1 - \mathbf{c}_2\|^2}{\frac{1}{n_{C_1}} + \frac{1}{n_{C_2}}} \quad (7.5)$$

where  $C_1$  and  $C_2$  represent clusters 1 and 2,  $c_1$  and  $c_2$  are the centroids of the clusters and  $n_{c1}$  and  $n_{c2}$  are the number of points in the clusters.

## 7.4 Operational State-Based controller design

The basis for the controller is the trained SOM, which is used to identify relevant operating states of a WWTP in order to define optimal control settings accordingly. The first step is to build up a SOM based on the available process data and to optimize the map design using an optimization algorithm, such as Genetic Algorithms (GA). For this optimization, the quantization and the topology error are used as performance measures. The second step is to identify the operating states on the map which is achieved by a clustering algorithm, such as ward linkage. In the third step, optimal control parameters for each of the clusters need to be chosen, which are selected using GA once again. Based on the fully functional SOM, new process data can be used to quickly identify the corresponding operating state and the respective control parameters.

### 7.4.1 Implementation of the online Controller

For the optimization, as well as for the later operation, the controller is developed as a Matlab Simulink module (Matlab S-Function). During simulation, the Simulink module is continuously fed with the current measurable state variables. Figure 7-2 depicts the different steps which are executed in the controller module.

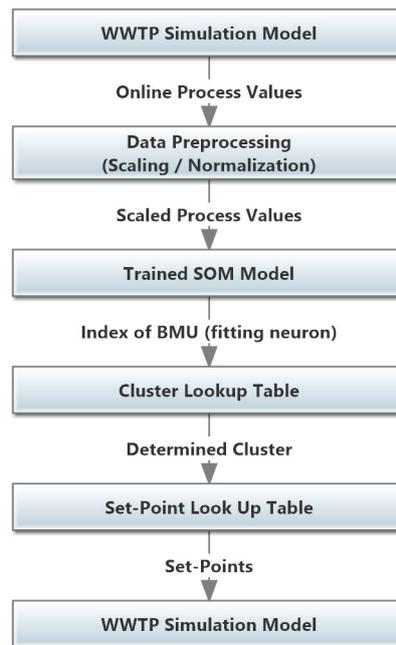


Figure 7-2: Operation regime controller (ORG) working principle

After data pre-processing (data normalization), the scaled process values  $\mathbf{x}$  are presented to a SOM, which has previously been trained on process data covering the full operating range. Using the Euclidean distance measure  $\mathbf{x}$  is compared with the vectors of weight  $\mathbf{m}_i = [m_{i1}, \dots, m_{in}]$  of the SOM and the BMU is determined. In the next step the index  $i$  of the BMU is transferred to a Cluster Lookup Table. The lookup table is created using a clustering algorithm, in this case ward linkage, applied to the vectors of weights of the SOM. The resulting lookup table contains the information on which neuron  $\mathbf{m}_i$  belongs to which cluster  $c_k$ ,  $k=1, \dots, P$  where  $P$  is the number of clusters. The cluster index  $k$  is then transferred to a second lookup table containing controller set-points for WWTP plant oxygen PID controller. During optimization the optimal set points for each operation regime, represented by the cluster, were determined. In the last step, these set points are fed back to the WWTP model.

## 7.5 Test Case 1: Modified BSM1 Model

Test case 1 is introduced to evaluate the method and to conduct a thorough investigation of the possible settings of the SOM using a widely known standard model.

### 7.5.1 Description of the Simulation Model

Test case 1 was conducted using a modified version of the Benchmark Simulation Model 1 (BSM1) (Alex et al., 2008), which represents a typical European WWTP. The model is based on the ASM1 Model, which was also used for the Rospe WWTP model. To simplify the model, the number of tanks was reduced from five to two tanks, but the denitrification and the nitrification tank volumes, as well as the clarifier stage were retained. The simulation model represents a WWTP with two biological reactors (see Figure 7-3). An upstream denitrification tank with a capacity of  $2000 \text{ m}^3$  is followed by a nitrification tank with a capacity of  $4000 \text{ m}^3$  and a clarifier with a surface area of  $1000 \text{ m}^2$ . The nitrate concentration is controlled by internal recirculation.

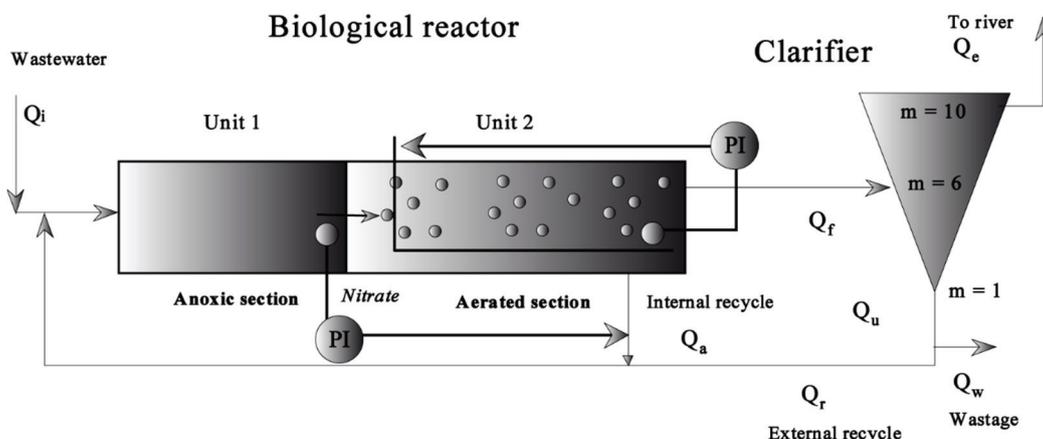


Figure 7-3: Design of the modified Benchmark Simulation Model No. 1

A good starting point for the control of the model is a fixed set point of  $2 \text{ m/l}$  for the oxygen concentration. This value is considered a good compromise between energy consumption and effluent quality for this plant under the given inflow conditions. The IWA Taskgroup on Benchmarking of Control Strategies developed several typical inflow scenarios for the BSM1. Figure 7-4 depicts the dry weather inflow scenario for hydraulic flow  $Q$  and ammonium  $S_{NH}$ . Further inflow scenarios for rain and storm weather are described in the BSM1 technical report (Alex et al., 2008).

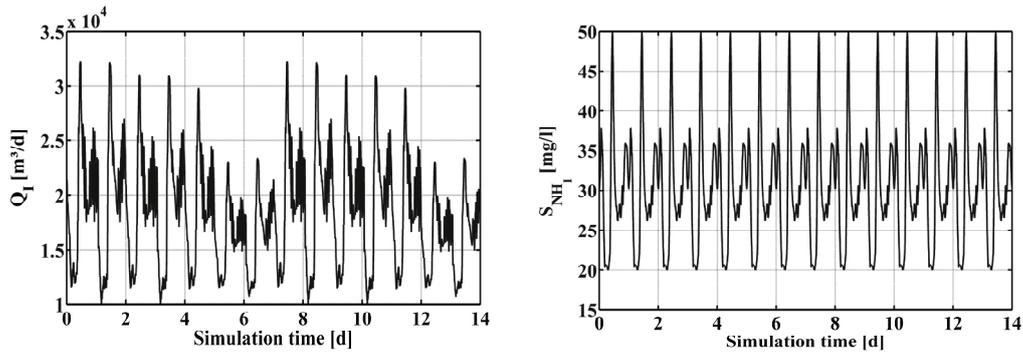


Figure 7-4: BSM1 Dry weather inflow scenario

Due to the fact that the size of the plant is similar to the unmodified BSM1, these scenarios are also applicable for the model and therefore are used as inflow.

### 7.5.2 Model initialisation

To initialize the model, a 100-days of stabilization period under closed-loop control using constant inputs (average dry weather inflow) is simulated. The system is stabilized during this period. The period is particularly important for the growth of the biomass  $X_{B,H}$  and  $X_{B,A}$  and for achieving a reference steady state of the plant for further simulations. Following this, a 14-day simulation period with dry weather inflow is performed to bring the plant to a desired state for testing of the controller strategy.

### 7.5.3 Data generation and variable selection

The data for the SOM clustering is created synthetically using the initialized and fully calibrated simulation model. To make the SOM clustering as generally applicable as possible, three different weather case scenarios were simulated for 14 days each, namely dry, rainy and stormy weather. During these simulation runs, the state vector of the ASM1 for the aerated bioreactor is sampled every 15 minutes for each weather scenario. Due to the fact, that not all 14 process variables of the state vector are relevant for aeration control, the most suitable variables have to be selected. Furthermore, it is important to consider the fact, that some process variables are extremely difficult to measure, which reduces practical applicability of the presented approach. Therefore, selection is performed based on measurability and relevance of the variables for the nitrification process in the aerated bioreactor. The selected variables are  $S_{NH}$ ,  $S_{NO}$ ,  $Q$  and the sum of  $S_S$  and  $X_S$ , which represents the portion of

available degradable  $COD$ . In addition to these state vector variables, an approximation of the oxygen consumption  $O_c$  is used. This is defined as the total airflow into the bioreactor  $Q_{air}$  divided by the oxygen concentration  $S_o$  inside the reactor.

$$O_c = \frac{Q_{air}}{S_o} \quad (7.6)$$

#### 7.5.4 Development of the fitness function

Under the assumption that all state variables except ammonium concentration ( $S_{NH}$ ) and oxygen ( $S_o$ ) are constant, it is possible to plot the general shape of the ammonium degradation function for a certain state. In reality, shape and position of the best area depend on several other state variables. Figure 7-5 shows the area, where ammonium is relatively high, while the oxygen concentration is still adequate ( $<2.5 \text{ mg/l}$ ). It is evident, that there is always a trade-off between oxygen concentration, meaning higher energy consumption, and ammonium degradation. Therefore, the most efficient ratio between  $S_o$  and  $S_{NH}$  to get the most efficient ammonium degradation ( $D_{S_{NH}}$ ) with the least amount of energy can be determined.

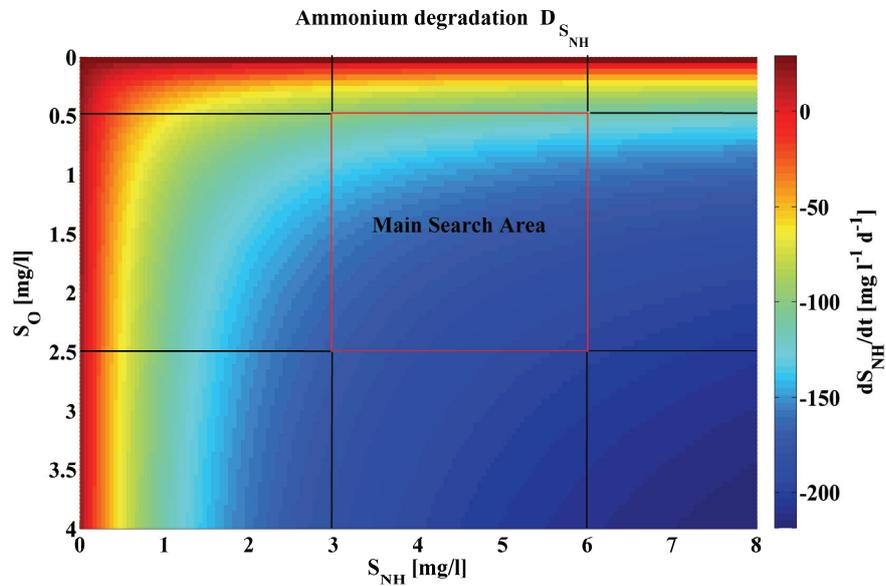


Figure 7-5: Ammonium degradation in an aerated bioreactor under the assumption that all variables except ammonium and oxygen are constant

Using the assumption depicted in Figure 7-5 the following fitness function was developed:

$$f := \min \left( \int_{t_{S_i}}^{t_{E_i}} \frac{Q_{air}}{D_{S_{NH}}} dt \right) \quad (7.7)$$

where  $Q_{air}$  is the air pumped into the bioreactor,  $D_{S_{NH}}$  is the  $S_{NH}$  degradation,  $t_{S_i}$  the start time and  $t_{E_{S_i}}$  the end time of state  $i$ . This function has a minimum where the ratio between airflow and ammonium degradation is optimal. A well dimensioned WWTP should be able to keep the desired effluent values at the point of highest energy efficiency. This means that the fitness function has to be modified for overloaded or under loaded WWTP.

### 7.5.5 SOM Controller development CASE 1

To develop the Controller for the modified BSM1 model, the different functions depicted in Figure 7-2 had to be implemented and optimized.

#### 7.5.5.1 Data pre-processing

As a pre-processing step, the data is scaled between 0 and 1 using the min-max method:

$$x_{scaled} = \frac{(x - x_{min})}{(x_{max} - x_{min})} \quad (7.8)$$

During training of the SOM, Euclidean distance is used to determine the similarity or dissimilarity between the nodes and the input vector. Without normalization, high values from the state variables like flow would have a significantly higher influence on the map.

#### 7.5.5.2 SOM model CASE 1

The SOM model for the given data set is realized using the Java SOMToolbox developed by the Institute of Software Technology and Interactive Systems (Vienna University of Technology, 2010). The Java SOMToolbox offers high performance algorithms for the development of SOMs as well as different visualization methods. The best SOM for the given dataset is realized using Genetic Algorithms (GA) in Matlab to determine the optimal map dimension, learning rate and number of iterations. To integrate both platforms for the optimization steps, an interface between the Matlab GA toolbox (also the simulation model for the online operation) and the Java SOMToolbox was developed. The communication principle is depicted in Figure 7-6.

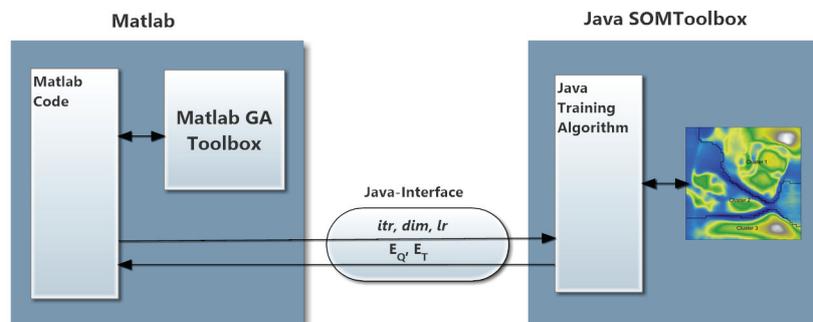


Figure 7-6: Communication principle between Matlab and the Java SOMToolbox

Given that the cost functions are quantization error  $E_Q$  and topology error  $E_T$ , the objective function and boundaries were defined as follows

$$\begin{aligned} Cost_a &:= \min(E_T) \\ Cost_b &:= \min(E_Q) \end{aligned} \tag{7.9}$$

subject to

- $d_1 \leq d \leq d_2$ , where  $d_1$  and  $d_2$  are the min and max map dimensions
- $lr_1 \leq lr \leq lr_2$ , where  $lr_1$  and  $lr_2$  are the min and max learning rates
- $itr_1 \leq itr \leq itr_2$ , where  $itr_1$  and  $itr_2$  are the min and max training iterations

The objective function is formulated as a single objective function for each cost function separately, because the computational effort for a combined multi-objective optimisation can grow significantly. Table 7-1 shows the optimization parameters and the boundaries fed to the GA.

Table 7-1: GA SOM Parameters CASE 1

Boundary	$d[x, y]$	$lr$	$itr$
lower	20	0.5	90000
upper	60	1	150000

Table 7-2 shows that the results for  $E_Q$  and  $E_T$  differ by approximately 10%. Using  $E_Q$  as cost function leads to a map with 3364 neurons (58 x 58), while using  $E_T$  as cost function leads to a smaller map with 2704 neurons (52 x 52).

Table 7-2: GA SOM Optimizatin results CASE 1

Cost function	SOM tuned parameters			Quality factors	
	$d[x, y]$	$lr$	$itr$	$E_Q$	$E_T$
$E_Q$	58,58	0.9561	147560	0.0343	0.1182
$E_T$	52,52	0.9785	131720	0.0377	0.1033

Having a look at the definition of  $E_Q$  and  $E_T$  (section 7.2.2) this result is expected since a higher number of neurons reduces the probability that similar neurons are adjacent. As there is only a 10% difference in  $E_Q$  and the topological error is better, the second configuration with 2,704 neurons was chosen for the controller.

### 7.5.5.3 Clustering of the trained map

After training the map with the optimal parameters as described in section 7.5.5.2, the model generated is further clustered using a ward linkage algorithm in order to create the state prediction model. From several experiments conducted, ward linkage proved better in comparison to complete linkage and k-means. To validate these results, the silhouette algorithm (Kaufman, 1990) is applied on the data and predicted states. The silhouette value for each point on the map is a measure of how similar that point is to points in its own cluster, when compared to points in other clusters. The silhouette value ranges from -1 to +1. A high silhouette value indicates that it is well-matched to its own cluster, and poorly-matched to neighbouring clusters. If most points have a high silhouette value, then the clustering solution is appropriate. If many points have a low or negative silhouette value, then the clustering solution may have either too many or too few clusters. The mean values of the results are given in Table 7-3:

**Table 7-3: SOM prediction model clustering evaluation using silhouette function for three clusters**

Silhouette results for 3 cluster			
Method	Ward Linkage	Complete Linkage	k-means
Mean	0.5011	0.3895	0.3677

### 7.5.6 Results of test case 1

In this section, the results from the experiments are discussed. Figure 7-7 shows the three operational states identified by the SOM prediction model. The model is formulated based on the clustered SOM as shown in Figure 7-8. As discussed the map shows the results of clusters produced by the ward linkage algorithm and the various islands are generated with a Smoothed Data Histogram (SDH) (Pampalk et al., 2002). The SDH identifies clusters by resembling the distribution of the data on the map. Due to the complex character of the plants states, it is a challenge to relate the state to certain measurement values. Looking at Figure 7-7, it becomes obvious that the states follow a daily course. This daily course represents a typical load of a WWTP. From this, it can be argued that the determined states represent the state of the plant. A typical operational pattern of the plant revealed that in the night the plant operates at minimal load, while the load is rated medium in the morning and highest around mid-day. The pattern formed by the clusters (see Figure 7-8) showed a correlation with the operational load. Cluster 1 represents the night load, cluster 2 reveals the transition period and cluster 3 represents the mid-day load.

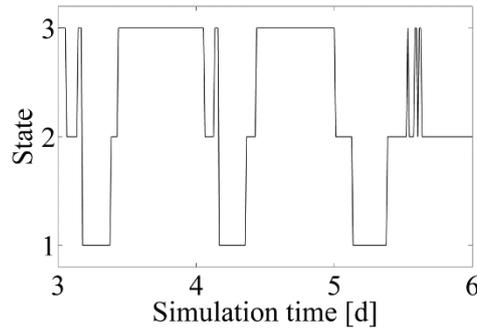


Figure 7-7: Recorded states over a simulation period of three days

From Figure 7-8, it can be seen that the ward linkage algorithm separates the clusters at the same boundaries as the SDH algorithm. The three clusters correspond to the inflow categories of low, medium and high as described in (Ebel, 2009).

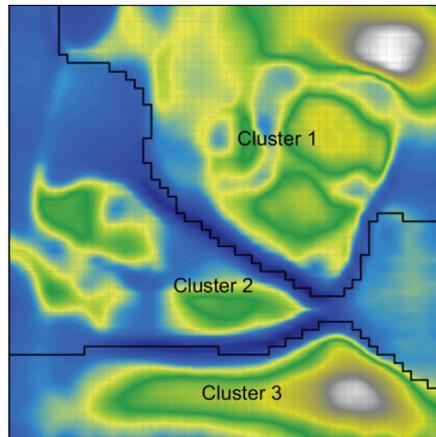


Figure 7-8: Smoothed data histogram of the SOM clustered with ward linkage algorithm (lighter colours indicate areas with higher data density on the map)

In Table 7-4, the Optimized oxygen ( $S_o$ ) set points found by the GA are tabulated. It is noticeable that for cluster 1 during the night a high set point was found. This is because the set point does not directly represent the airflow into the plant, rather the most efficient area.

Table 7-4: Operation Regime oxygen set points Case 1

Operation Regime	Optimized $S_o$ Setpoint [ $mg / l$ ]
1	2.48
2	1.77
3	2.26

Furthermore, the mean value for  $S_o$  of the three operation regimes is  $2.17 \text{ mg/l}$ . This value is, as expected, very close the fixed  $2 \text{ mg/l}$  set point which is considered optimal for the BSM1 plant. Although the effluent values were not included in the fitness function, good results with regard to efficient plant

operation and effluent values were achieved by the presented methodology. Keeping the effluent values and the energy in a reasonable domain shows that the plant is well operated. Otherwise the effluent values would be violated or the overall energy consumption would increase significantly. Table 7-5 gives an overview of the most important effluent parameters as well as the aeration energy. The aeration energy was calculated for the second week of the simulation period as described in the simulation procedure for the BSM1 (Alex et al., 2008). The results show that the aeration energy and the total nitrogen are kept nearly at the same level while yielding a reduction of 3.3% in ammonium in the effluent.

**Table 7-5: Comparison of the plant performance Case 1**

	Fixed $S_o$ Set Point	Optimized $S_o$ Set Point
<b>Total Nitrogen</b>	17.97 mg/l	18.06 mg/l
<b>Ammonium</b>	3.56 mg/l	3.44 mg/l
<b>Aeration Energy</b>	1163 kwh	1184 kwh

## 7.6 Test Case 2: Rospe Plant

The second test of the developed methodology is conducted using the Rospe WWTP model developed in section 4. While the BSM1 model is an artificially generated model with perfect inflow data created by a group of specialists, the Rospe model is a model which is representative of a real WWTP.

### 7.6.1 Data generation and input variable selection Case 2

The selection of input variables is based on measurement devices which are available on the real plant. These are ammonium ( $S_{NH}$ ) and nitrate ( $S_{NO}$ ) at the outflow of tank D4 (see Figure 4-8) as well as the flow ( $Q$ ) in the inflow of the plant. While additional process variables from other tanks, as well as COD measurements would be desirable, creating these variables artificially from the model would contradict the idea of the methodology. For this reason three input data sets were created:

- $DS_1$ : Two dimensional dataset including  $S_{NH}$  and  $S_{NO}$  based on measured plant operating conditions.
- $DS_2$ : Three dimensional dataset including  $S_{NH}$ ,  $S_{NO}$  and  $Q$  based on measured plant operating conditions.

Both datasets consist of 2,448 data samples (10 min sample interval) over the measuring period of 17 days, described in section 5.3. From a practical point of view, it makes sense to use the second dataset  $DS_2$  because the flow in combination with the concentration of  $S_{NH}$  and  $S_{NO}$  it offers better information on the load on the plant. A problem with using these datasets to optimize the SOM regime estimation model and control setpoints for the optimization is the fact, that they only represent operating conditions, which occurred during normal operation of the real plant. These operating conditions are

characterized by relatively low ammonium and relatively high nitrate values. This means that new operating conditions generated during GA optimization, especially when ammonium is medium or high, are not captured by the trained model. For this reason an additional data set  $DS_3$  was created.

- $DS_3$ : Three dimensional dataset including  $S_{NH}$ ,  $S_{NO}$  and  $Q$  based on measured plant operating conditions and additional simulated operating conditions.

The simulated operating conditions were generated by simulating the 17 days of the measurement campaign four times with different fixed oxygen set points for the three nitrification tanks (N1, N2, N3). The different set points for the operating scenarios are depicted in Table 7-6. The set points are chosen by hand, so that they cover the whole operating range from complete ammonium to complete nitrate elimination.

**Table 7-6: Control scenarios to create different operating conditions for SOM training**

Scenario	Fixed $S_o$ Tank N1	Fixed $S_o$ Tank N2	Fixed $S_o$ Tank N3
1	0	0	0
2	0.1	0.1	0.1
3	0.6	0.6	0.5
4	1	1	1

From each simulated scenario 500 data samples were randomly picked and merged into one training data set. Therefore  $DS_3$  contains a total number of 2,500 data samples. Table 7-7 shows the main characteristics of the  $DS_3$ . The min and max values are particularly important, because they show that the complete possible range is covered.

**Table 7-7: Characteristics of data set DS3**

Scenario	$NH_4 - N [mg / l]$	$NO_3 - N [mg / l]$	$Q [m^3 / d]$
<b>min</b>	0	0	29,849
<b>max</b>	11.64	12.02	62,742
<b>mean</b>	3.41	6.03	40,533
<b>Number of samples</b>	2,500		

The necessary data pre-processing was done using the min-max method, scaling the measurement values for each process variable between 0 and 1, as described in section 7.5.5.1.

## 7.6.2 SOM development and optimization

The SOM development for the Rospe WWTP was undertaken using the Matlab neural network toolbox as mentioned previously. Although the Java SOMtoolbox used for Case 1, is a powerful tool offering a wide range of visualization and data processing methods, using it in combination with Matlab for optimization and online simulation is a slow process. The reason is that it has to be called during each simulation step from Matlab. Furthermore, the Matlab neural network toolbox offers a batch training algorithm, which processes the complete input data set in one pass and updates all weights according to the learning rate (see section 7.2.1). Each pass is called an epoche. For this work, this is considerably faster than presenting every sample individually. Due to the fact, that the topology error and the quantization error were not available as Matlab functions custom functions were implemented. The experiments for the Rospe plant were conducted separately for data sets  $DS_1$  and  $DS_2$  and data set  $DS_3$ .

### 7.6.2.1 SOM Optimization for data sets $DS_1$ and $DS_2$

From the experiments in Case 1 it became obvious that the optimum learning rate  $lr$  and the number of iterations  $itr$  are inversely related (i.e. using a small learning rate leads to a higher number of necessary iterations and vice versa). Tuning them individually only leads to minor improvements, while the computation time increases significantly. Therefore, for the experiments in case 2, Matlab's default learning rate  $lr = 0.9$  (which is close to the determined optimal  $lr$  in Case 1) was used and only the map size and number of training epochs were optimized. The optimal map size and number of epochs, was determined by a grid search. The map dimensions were varied between 5 and 30 with an increment of 5 (using a quadratic map both dimensions are kept the same  $dim_1 = dim_2$ ). For each map size, different numbers of training epochs were tested, between 10 and 300 with an increment of 10. These ranges were chosen as a result of preceding tests.

Figure 7-9 shows the topology error  $E_T$  for data set  $DS_1$ . It can be seen that the minimum is at 100 epochs. While the map dimension has a minor influence on  $E_T$ , it rises with dimension size. This behaviour is expected due to the fact that each neurons has only 6 adjacent neuron on a hexagonal map or 4 on a rectangular map (see section 7.2.2.2). Therefore the probability that the first and the second BMU are adjacent decreases with map size.

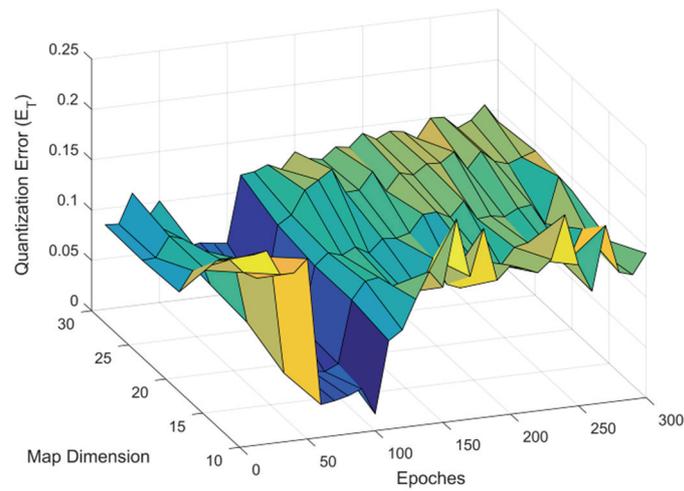


Figure 7-9: Optimization results of the topology error for data set DS1

Figure 7-10 shows the quantization error  $E_Q$  for data set  $DS_1$ . For  $E_Q$  the dimension of the map has a higher influence, while the training shows no significant improvement after 100 epochs.

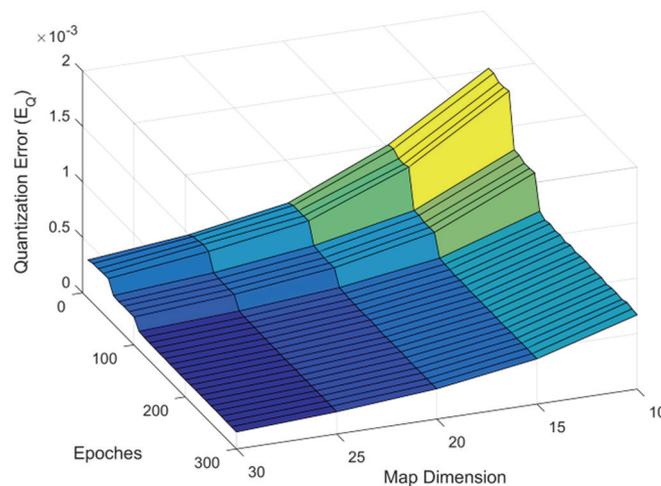


Figure 7-10: Optimization results of the quantization error for data set DS1

While topology preservation is important for visualization and a good indicator of the training success, it plays a minor role for this application, due to the fact that the weight vectors are clustered in the next step, and similar weight vectors will be assigned to the same cluster independent of their location in the SOM. Therefore a map with a dimension of 30x30 (900) neurons and 100 epochs of training was chosen. This map has a  $E_Q = 0.0506$  and a  $E_T = 2.518 \cdot 10^{-4}$ .

Figure 7-11 shows the results for the topology error optimization for data set  $DS_2$ . It shows obviously similar behaviour as  $DS_1$ . The minimum topologic error is at 100 epochs training, while the dimension of the map has a minor influence on the error result.

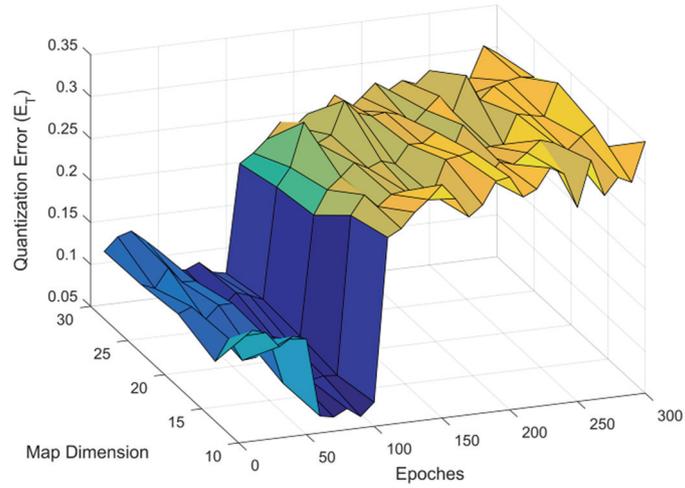


Figure 7-11: Optimization results of the topology error for data set DS2

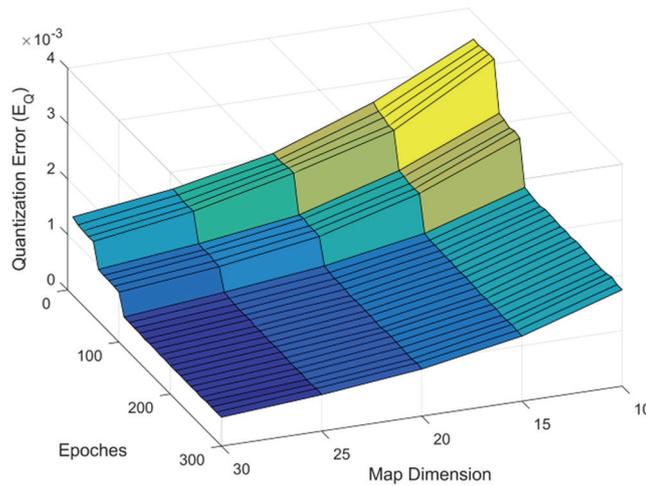


Figure 7-12: Optimization results of the quantization error for data set DS2

The quantization error for  $DS_2$ , shown in Figure 7-12, has similar behaviour to  $DS_1$ . The minimum is reached after 100 epochs of training, while the error decreases with the map dimension (number of neurons). Therefore the same size of 30x30 neurons is chosen. The quantization error for  $DS_2$  is  $E_Q = 0.0698$  and the topology error is  $E_T = 8.9 \cdot 10^{-4}$ .

### 7.6.2.2 SOM Clustering for DS1 and DS2

The optimal number of clusters was determined using the silhouette criterion. For data set  $DS_1$  with two input variables, the best silhouette value of 0.61 is achieved with two clusters, while the silhouette value for three clusters is only slightly smaller with 0.59.

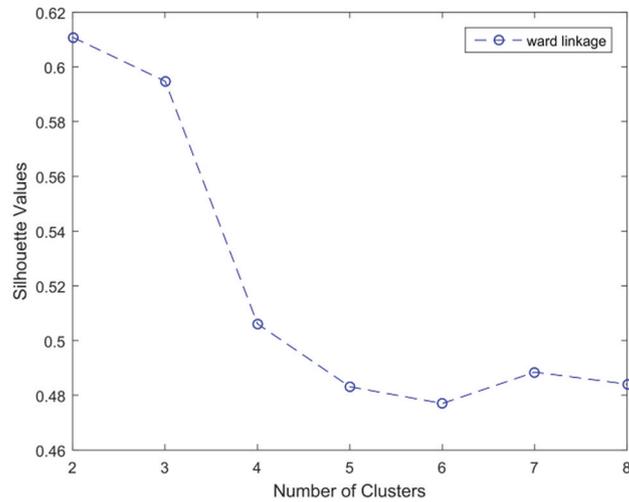


Figure 7-13: Determination of the optimal number of clusters using Silhouette criterion for two dimensional SOM weights using ward linkage clustering.

Under the assumption, that one cluster represents high load, one cluster medium load and one cluster low load conditions, three clusters were chosen over two clusters. Using only two dimensional input data offers the possibility to plot the weight vectors of the map using a scatter plot. Figure 7-14 shows the three clusters generated with ward linkage. The weight vectors of the neurons were de-normalized. Therefore, the axes show the real process values.

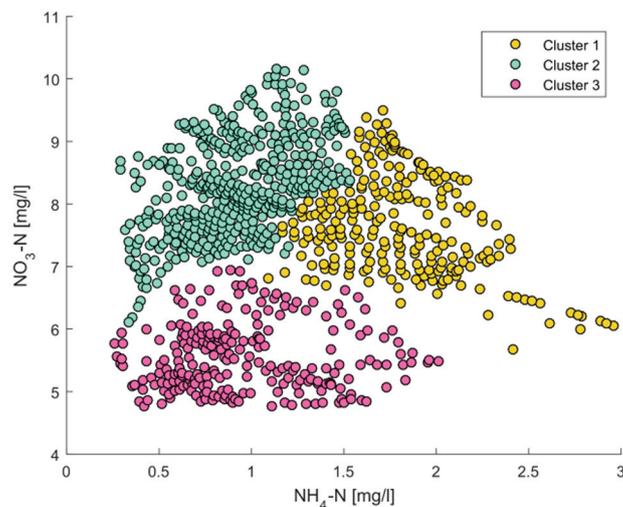


Figure 7-14: Ward linkage clusters for data set DS1 with de-normalized weight vectors

It can be seen, that the neurons are evenly distributed. In this example cluster 1 represents a higher load on the plant, where cluster 3 represents a low load. Cluster 2 shows an operating condition where ammonium is low and nitrate is high. In this this area the controller should decrease the airflow to the plant. Figure 7-15 shows the diurnal cycles of three different days from the measurement campaign on the plant. These three days were taken from different load situations. Day 8 represents a medium load day, which cycles through cluster 1 and 2, day 1 represents a higher load with high ammonium and high nitrate values and day 13 represents a low load situation. It can be seen that most of the time the plant resides in cluster 3.

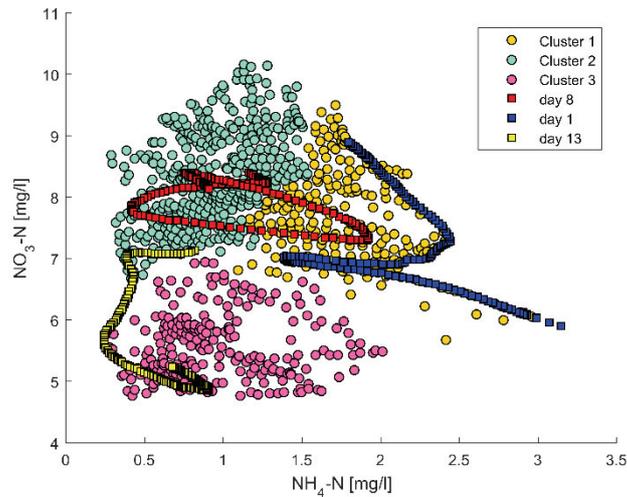


Figure 7-15: Ward linkage clusters for data set  $DS_1$  with de normalized weight vectors and diurnal cycle of day 1, day8 and day 13

Figure 7-16 shows the silhouette values for data set  $DS_2$ . The highest value is achieved for 2 clusters. Using only two operation regimes would end up in a very simple controller. The next best silhouette value is achieved for six clusters. Figure 7-17 shows the neighbour weight distance plot for the SOM trained with  $DS_2$ . The blue dots represent the neurons, while the fields between the neurons represent the distances in the input space (in this case three dimensional). Similar data would be represented by lighter areas separated by darker lines. In this case it is obvious that the map contains more than two areas with similar data. For this reason the second best result of the silhouette criterion with six clusters is chosen.

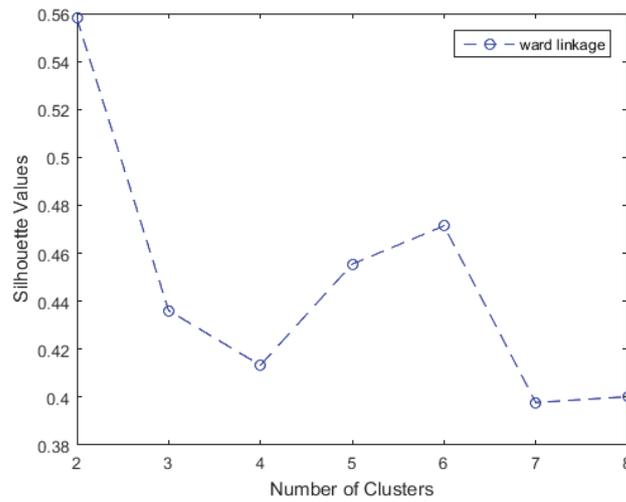


Figure 7-16: Determination of the optimal number of clusters using Silhouette criterion for three dimensional SOM weights using ward linkage clustering.

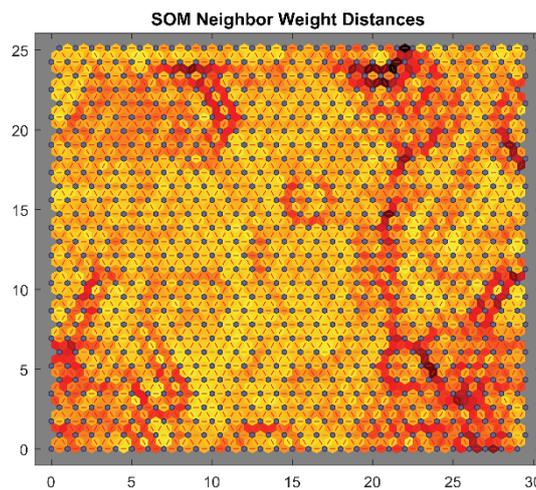


Figure 7-17: SOM Neighbor weight distances  $DS_2$  (darker colors indicate bigger distances in the input space between neurons)

Figure 7-18 shows a 3D scatter plot of the clusters generated from dataset  $DS_2$  using ward linkage clustering. The weight vectors of the neurons were again de-normalized to relate them to the real inputs. It can be seen that while neuron weights are distributed over the whole space, distinct clusters are formed. In particular in the vicinity of clusters 4 and 5 a small area, which is not covered with neurons, because an operation regime with low  $NH_4 - N$ , low  $NO_3 - N$  and low  $Q$  doesn't appear in the data.

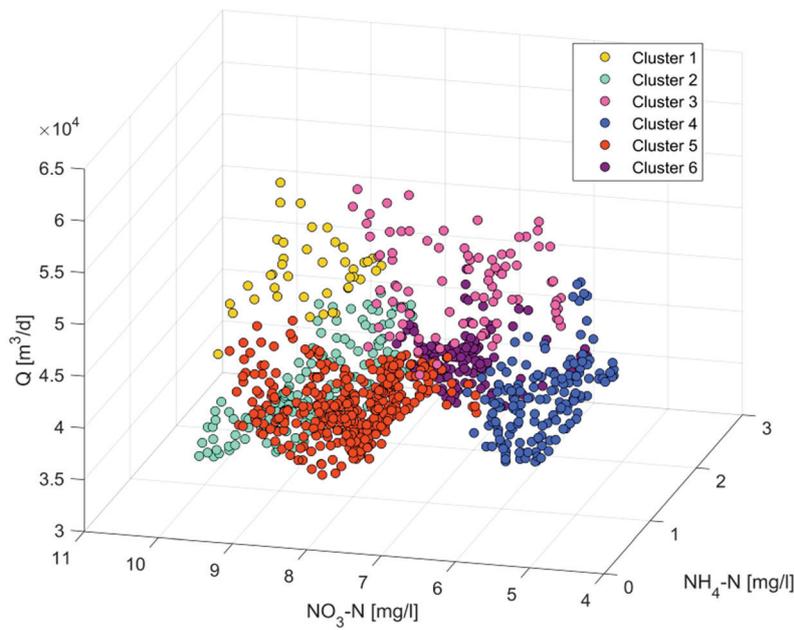


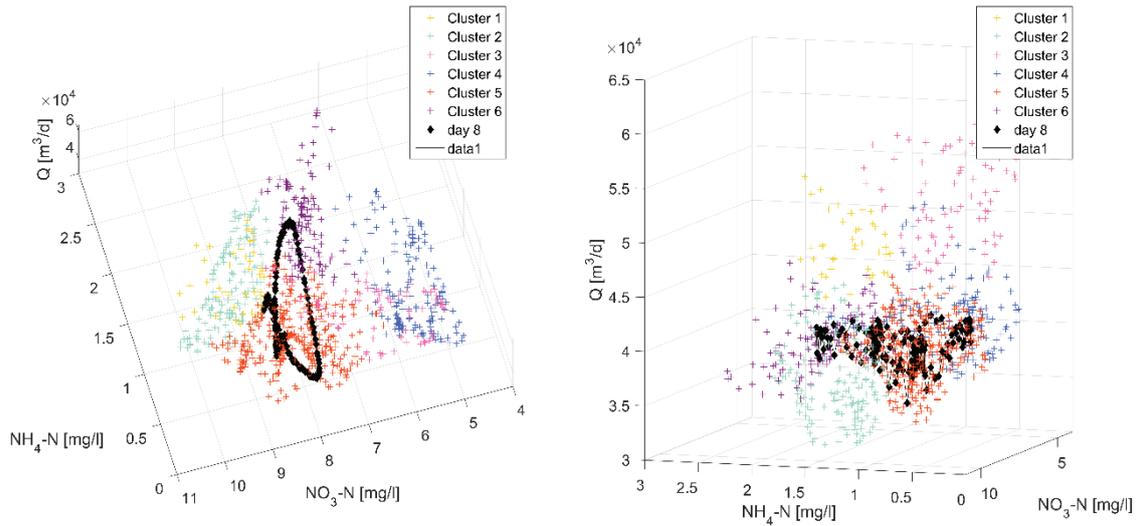
Figure 7-18: Ward linkage clusters for data set DS3 with de normalized weight vectors

Table 7-8 gives a short interpretation of the clusters shown in Figure 7-18. It can be seen, that the clusters generated depict realistic operation conditions. Two clusters 1 and 3 present high inflow conditions. E.g. cluster 1 represent high flow, high ammonium and medium nitrate, a condition typical after a rain peak, when ammonium is pushed into the plant. Cluster 3 is typical for rainy weather with high flow and diluted wastewater, which leads to low ammonium and nitrate.

Table 7-8: Interpretation of SOM Clusters for data set DS2

Cluster	$Q$	$NH_4 - N$	$NO_3 - N$
1	high	high	medium
2	low	high	medium
3	high	low	low
4	low-medium	low	low-medium
5	low	high	low
6	low	medium	high

The other four clusters depict dry weather inflow conditions, which typically occur over the course of a day. Figure 7-19 shows day 8 from two different angles. It is recognizable that day 8 starts and ends in the same area in the 3D space. Over the day the operation regime of the plant is mainly cluster 5 and 6, which are both low inflow clusters (see right side of Figure 7-19). Therefore it can be seen that it was a typical dry weather inflow situation with the typical load peak at noon (Cluster 6).

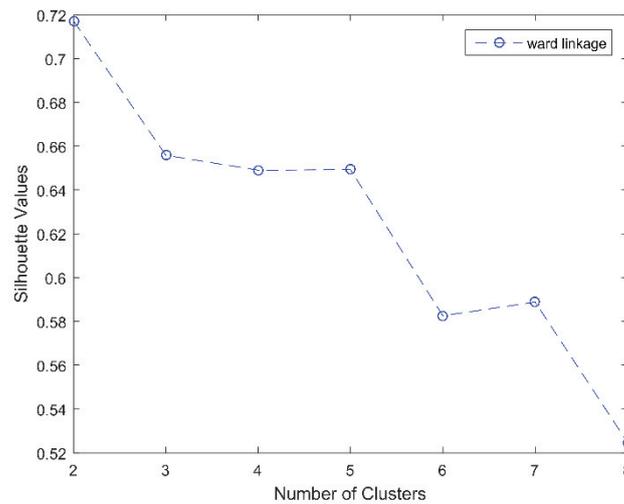


**Figure 7-19: Ward linkage clusters for data set DS3 with de-normalized weight vectors including diurnal cycle of day 1 from two angles**

The results show that, the cluster depict different operation regimes, and that the plant is following a path through the generated clusters over the operation. The next section will show an example for the Rospe plant, of how this decomposition of the operating space can be used for control purposes.

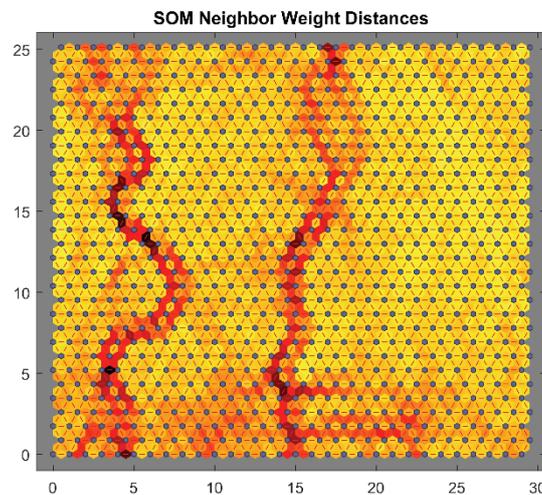
### 7.6.3 SOM Controller development for Case 2

Controller development for the Rospe plant followed the same approach as presented for Case 1. The SOM was trained using dataset  $DS_3$  (see section 7.6.1) with the optimal map dimension determined using the method described in section 7.6.2. A map with 900 neurons (30x30) was chosen with a  $E_T = 0.0521$  and a  $E_Q = 8.24 \cdot 10^{-4}$ . The optimal number of clusters was determined using the silhouette function (Figure 7-20).



**Figure 7-20: Determination of the optimal number of clusters using ward linkage and Silhouette criterion as quality measure for three dimensional data set DS3**

According to the silhouette value, two clusters would be the best result, but looking at the neighbour weight distances in Figure 7-21 it is obvious that at least three clusters are visibly separated by the darker vertical lines. While the silhouette values for 3, 4 and 5 clusters are very similar, 3 clusters with equivalently operating regimes were chosen.



**Figure 7-21: SOM neighbour weight distances for 30x30 map using DS3**

#### 7.6.4 Development of a fitness function for Case 2

The goal of the fitness function for the WWTP Rospe is to save air or energy respectively, without violating the effluent limits. Therefore a simple fitness function is used which integrates the airflow  $Q_{air}$  over the simulation period. This value is divided by the number of days in the simulation period and a fixed number used to reduce the magnitude of the fitness value:

$$f_1 := \min \left( \int_{t_s}^{t_E} \frac{Q_{air}}{d \cdot 4000} dt \right) \quad (7.10)$$

Where  $t_s$  is the start time and  $t_E$  the end time of the simulation period. Violations of the effluent limits for  $NH_4 - N$  (model variable  $S_{NH}$ ) and total nitrogen  $N_{tot}$  are penalized, when :

$$f_2 := \begin{cases} 0 & \text{if } (S_{NH_{max}} < 6.4) \\ ((S_{NH} - S_{NH_{max}}) + 1) \cdot 0.5 & \text{if } (S_{NH_{max}} > 6.4) \end{cases} \quad (7.11)$$

$$f_3 := \begin{cases} 0 & \text{if } (N_{Tot_{max}} < 14.4) \\ ((N_{Tot} - N_{Tot_{max}}) + 1) \cdot 0.5 & \text{if } (N_{Tot_{max}} > 14.4) \end{cases} \quad (7.12)$$

where  $N_{Tot}$  is total nitrogen and  $N_{Tot_{max}}$  the maximum value during the simulation. The limit values for the penalty are 80% of the plant limit values. The complete fitness value is calculated by summing up these three parts:

$$f = f_1 + f_2 + f_3 \quad (7.13)$$

This function has a minimum when the minimal amount of air is used without violating one of the nitrogen effluent limits.

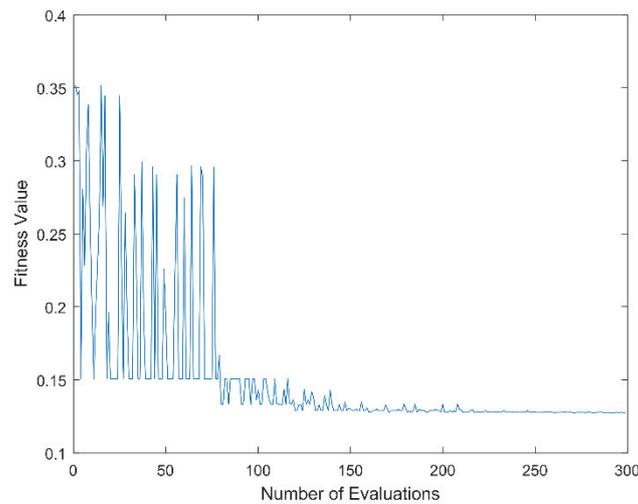
#### 7.6.4.1 Optimization of operation regime oxygen set points

To determine the optimal oxygen set points for the three operation regimes, an online SOM controller was developed in Matlab. Because the complete system was simulated in Matlab, simulations were approximately twice as fast as those conducted via the java interface used for Case 1. This speed-up is important, because the GA runs the model several hundred times with different oxygen setpoints. The full 17 day period from the measurement campaign is used as evaluation period. The determined set-point for an operating regime is used for all three nitrification tanks.

For each evaluation run, the GA writes a new oxygen look-up table and runs the model. At the end of the evaluation run, the fitness value is calculated using the fitness function described in equation (7.13).

#### 7.6.5 Results and discussion for Case 2

Figure 7-22 shows the fitness values for 300 evaluation runs. While in the first 75 the spread is relatively high, it settles down from the 76<sup>th</sup> run forward. The step at the 76<sup>th</sup> evaluation run is caused by the  $NH_4 - N$  penalty function. After 250 runs, there is no further improvement.



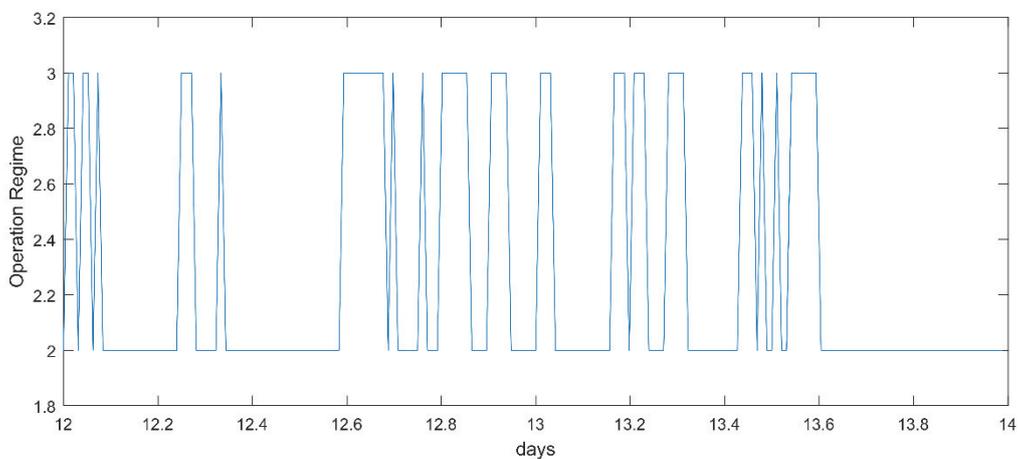
**Figure 7-22: Fitness results controller optimization**

Table 7-4 shows the optimal set points determined by the GA. The set points are very low, but considering that the suggested set points for the aeration of this plant are between 0.5 mg/l and 0.6 mg/l, these low set-points seem reasonable.

**Table 7-9: Operation Regime oxygen set points Case 2**

Operation Regime	Optimized $S_o$ Setpoint [mg / l]
<b>1</b>	0.8235
<b>2</b>	0.145
<b>3</b>	0.168

To compare the performance of the optimized set points against the fixed suggested set points and the set points used on the actual plant, the different settings were simulated. Figure 7-23 shows the operation regimes which occurred during the run used to develop the SOM-based controller.



**Figure 7-23: Operation Regimes of the rospe plant between day 12 and day 14**

It is obvious, that only operation regime 2 and 3 occur. The reason is, that the used fitness function tries to save energy by reducing the air flow. Therefore low  $NH_4 - N$  in combination with high  $NO_3 - N$  values, do not occur. This is why the operating regime doesn't appear in this simulation run.

Table 7-10: Comparison of different set points for the Rospe plant

Operation Regime	Airflow [m <sup>3</sup> / d]	Mean $NH_4 - N$ [mg / l]	Mean $NO_3 - N$ [mg / l]
Measured set points	11,580	0.86	8.2
Fixed [0.5 0.5 0.6]	6,774	2.27	6.83
SOM Controller	3,110	5.9	3.2

Table 7-10 shows the results of the different controllers. It is obvious, that the operator of the plant has a different operation goal than the described fitness function. This is done for technical reasons. The results for the measured set points show that  $NH_4 - N$  is very low, while  $NO_3 - N$  is high. Using the fixed set point, the situation is not as extreme as with the measured set points, but still has the same tendency to keep the  $NH_4 - N$  low. The SOM controller operates at the other end of the legal spectrum. Therefore it reduces the airflow significantly, which leads to higher  $NH_4 - N$  and lower  $NO_3 - N$ . Considering only the airflow, the new controller saves 73% air in comparison to the measured set points and 54% in comparison to the fixes set points. While these savings seem to be impressive, they only show that the plant can be operated differently.

## 7.7 Conclusion

The challenges of controlling a highly disturbed, non-linear system like a WWTP are not trivial. The proposed operating regime estimation and state-based set-point control method introduces new possibilities for control applications in the wastewater treatment. Although the favoured results for test case 1, are very similar to the static  $S_o$  set point controller results, many potential opportunities are created to achieve improved results in this field. An important aspect of this work dealt with operation state identification from process variables of a WWTP. The results proved encouraging with the clustering algorithm employed capturing and successfully categorizing the operating states of the WWTPs. This itself provides a platform for exploring its integration to developing control strategies. The outcome of this study presents a great potential for discovering some level of optimum in controlling the energy usage and at the same time keeping the effluent limits within given by regulations. In conclusion, a framework for state-based data driven controller design has been developed. This strategy is promising and offers great potential for achieving optimum energy control and at the same time compliance with effluent limits.

## 8 Conclusion

The description of the state of the art in WWTP operation, instrumentation and control and its relevance for the environment and society in chapters one, two and three shows that the development of novel operationally robust and practice-oriented WWTP instrumentation and control strategies is challenging but necessary. High treatment and maintenance costs put the WWTP operators under a lot of pressure to optimize their processes and to go beyond conventional control strategies.

Fortunately, the rapid development of new technologies and faster hardware in the area of process automation and industrial IT allows for the implementation of new computationally intensive approaches.

The developed methods and their evaluation on the BSM1 and the Rospe WWTP show that there is high potential for improvement in process optimization and control using computational intelligence methods and dynamic simulation models. The calibration of a WWTP model using the non deterministic multi-objective SMS-EGO algorithm shows that WWTP calibration can be efficiently automated facilitating the use of simulation models for optimization and control strategies significantly.

Furthermore, the lack of inflow measurements that are required for the development of control strategies is addressed by the virtual measurement systems introduced in chapter 5. Thus, crucial process variables in the inflow such as *COD* and *NH<sub>4</sub>-N* can be successfully estimated with sufficient accuracy based on process data from existing online instrumentation. Machine Learning methods have proven to be a valuable tool, although their parametrisation and training is time-consuming and requires expert knowledge. A comparison to an alternative model-based inflow estimation method developed by Ebel (2009) illustrates that Machine Learning achieves much better results while the model-based approach suffers from variable retention times and blending of different wastewater streams inside a WWTP.

Nevertheless, all this information on the treatment process provided by online instrumentation needs to be analysed to extract valuable process information that can be used for control purposes. Therefore, an operating regime estimation using Self-Organising Maps (SOM) was successfully implemented, in order to develop a state-based controller with optimal parameters for each operating regime. Results show that this control strategy achieves major savings in energy while maintaining sufficient cleaning capacity.

All in all, the achieved results prove that the use of computational intelligence methods for instrumentation and control purposes is able to substantially improve WWTP operation. While requiring a high degree of expert knowledge for the development and implementation of the novel methods, the required effort for operation is similar to conventional ICA strategies.

## References

- Alex, J., Benedetti, L., Copp, J., Gernaey, K.V., Jeppsson, U., Nopens, I., Pons, M.N., Rieger, L., Rosen, C., Steyer, J.P., 2008. Benchmark simulation model no. 1 (BSM1). Rep. IWA Taskgroup Benchmarking Control Strateg. WWTPs.
- Åmand, L., Olsson, G., Carlsson, B., 2013. Aeration control – a review. *Water Sci. Technol.* 67, 2374. doi:10.2166/wst.2013.139
- Arens, T., Radke, M., 2008. *Mathematik*. Spektrum, Akad. Verl., Heidelberg.
- ATV-DVWK (Ed.), 2001. *Automatisierung der chemischen Phosphatelimination ATV-DVWK-M 206*, Stand: November 2001. ed, ATV-DVWK-Regelwerk Merkblatt. GFA Ges. zur Förderung der Abwassertechnik, Hennef.
- Bever, J., 2002. *Weitergehende Abwasserreinigung*. Oldenbourg-Industrieverl., München.
- Bongards, M., Hilmer, T., Ebel, A., 2005. *Computational intelligence-based optimisation of wastewater treatment plants*. IWA Publishing 2005.
- Bongards M, Schaefer S, Ebel A, Kern P. 2006. *Wastewater Treatment Plant Improvement by Smart Sensors and Computational Intelligence (WAPSCIENCE)*. Workshop: Environmental Research for SMEs – Technology Challenges and Market Opportunities in Waste Water Treatment: Brussels
- Bongards, M., Hilmer, T., Kern, P., 2007. *Online-Konzentrationsmessung in Kanalnetzen - Technik und Betriebsergebnisse*. Forschungsbericht der Fachhochschule Köln 173–176.
- Brito, R.S., Pinheiro, H.M., Ferreira, F., Matos, J.S., Lourenço, N.D., 2014. In situ UV-Vis spectroscopy to estimate COD and TSS in wastewater drainage systems. *Urban Water J.* 11, 261–273. doi:10.1080/1573062X.2013.783087
- Carl Zeiss Microscopy GmbH, 2013. *Zeiss Spektrometermodule*.
- Clerc, M., 2006. *Particle swarm optimization*. ISTE, London; Newport Beach.
- De Boor, C., 2011. *Matlab 2011b Statistics Toolbox Users Guide*. MathWorks, Natick, MA.
- Deutsche Elektrotechnische Kommission, 2009. *Drehende elektrische Maschinen*. 30, 30. Beuth, Berlin.

- Deutsche Vereinigung für Wasserwirtschaft, A. und A., 2008. DWA-Regelwerk 269: Merkblatt. Prozessmessgeräte für Stickstoff, Phosphor und Kohlenstoff in Abwasserbehandlungsanlagen. DWA, Hennef.
- Deutschland, Bundesministerium der Justiz, 2009. Gesetz zur Ordnung des Wasserhaushalts: Wasserhaushaltsgesetz vom 31. Juli 2009 (BGBl. I S. 2585); WHG. Bundesministerium der Justiz, Bonn.
- Directive, C., 1991. Directive 91/271/EEC Council Directive of 21 May 1991 concerning urban waste water treatment. Off. J. Eur. Communities.
- Dürrenmatt, D.J., Gujer, W., Morgenroth, E., Gujer, W., Gujer, W., Morgenroth, E., Morgenroth, E., 2011. Data mining and data-driven modeling approaches to support wastewater treatment plant operation. ETH, Zürich.
- DWA, 2011. DWA M256-1 Prozessmesstechnik auf Kläranlagen. 1, Allgemeine Anforderungen. DWA, Dt. Vereinigung für Wasserwirtschaft, Abwasser u. Abfall, Hennef.
- DWA, 2001. DWA M256-4 Anforderungen an Einrichtungen zur Messung des pH-Wertes und des Redox-Potentials. GFA, Ges. zur Förderung der Abwassertechnik, Hennef.
- DWA, Lohmann, J., 2001. DWA M256-3 Anforderungen an Leitfähigkeitsmeseinrichtungen. GFA, Ges. zur Förderung der Abwassertechnik, Hennef.
- Ebel, A., 2009. Application of Computational Intelligence Techniques to Modelling, Control and Optimisation of Wastewater Treatment Plants, Doctoral Dissertation, Department of Electronic Engineering, National University of Ireland, Maynooth, Maynooth, 2009
- Endress+Hauser, 2016. Endress+Hauser - STIP-scan - One probe with 8 parameters. <http://www.de.endress.com/de/messgeraete-fuer-die-prozesstechnik/fl%C3%BCssigkeitsanalyse-produkt%C3%BCbersicht/nitrat-csb-bsb-toc-sak-schlamm-sensor-cas74> (accessed 08.20.16).
- EnergieAgentur.NRW, 2015. Strompreise in Deutschland [WWW Document]. URL <http://www.energieagentur.nrw.de/infografik/grafik.asp?TopCatID=3106&CatID=3106&RubrikID=3165> (accessed 2.10.15).
- Engelbrecht, A., 2002. Computational intelligence: an introduction. J. Wiley & Sons, Chichester England ; Hoboken N.J.

- European Environment Agency, 2014. Urban Waste Water Treatment maps [WWW Document]. URL <http://www.eea.europa.eu/data-and-maps/uwwtd/interactive-maps/urban-waste-water-treatment-maps-1> (accessed 8.20.16).
- Gahr, A., Wolf, C., Kern, P., 2014. Wasser und Abwasser in Megastädten der Zukunft. atp edition 56, 28–37.
- Gerber, M., 2009. Ganzheitliche stoffliche und energetische Modellierung des Biogasbildungsprozesses - Doctoral Dissertation. RUB - Ruhr University Bochum.
- Graner, M., Hilmer, T., Bongards, M., 2005. Einsatz ionenselektiver Messgeräte für die Online-Messung von Stickstoffverbindungen auf Kläranlagen – ein Erfahrungsbericht. VDI-Verl., pp. 81–92.
- Gujer, W., 2007. Siedlungswasserwirtschaft: mit 84 Tabellen. Springer, Berlin [u.a.].
- Gujer, W., Henze, M., Mino, T., Matsuo, T., Wentzel, M.C., Marais, G. v. R., 1995. The Activated Sludge Model No. 2: Biological phosphorus removal. Water Sci. Technol., Modelling and Control of Activated Sludge Processes Selected Proceedings of the IAWQ International Specialized Seminar on Modelling and Control of Activated Sludge Processes 31, 1–11. doi:10.1016/0273-1223(95)00175-M
- Gujer, W., Henze, M., Mino, T., von Loosdrecht, M.C.M. van, 1999. Activated sludge model No. 3. I W A Publishing.
- Hagan, M.T., Demuth, H.B., Beale, M.H., 2002. Neural network design, [Nachdr.]. ed. Campus Publ. Service, Boulder, Colo.
- Hastie, T., Efron, B., Johnstone, I., Tibshirani, R., 2003. Least Angle Regression, Forward Stagewise and the Lasso; Oral Presentation at Stanford University; URL: <http://web.stanford.edu/~hastie/TALKS/larstalk.pdf> (accessed 8.20.16).
- Henze, M., 2000. Activated sludge models ASM1, ASM2, ASM2d and ASM3. IWA Pub., London.
- Henze, M., Gary, A., Cameau, Y., Ekama, G.A., Garcia, J., Gebra, C.P., Hooijmans, C.M., Judd, S., Kim, B., von Lier, J.B., Mahmoud, N., Martens, A.M., Morgenroth, E.F., Olsson, G., Rosso, D., Stenstrom, M.K., Takacs, I., von Loosdrecht, M.C.M. van, Wentzel, M.C., Zeeman, G., 2008. Biological wastewater treatment: principles, modelling and design. IWA Pub, London.

- Henze, M., Grady, C.P.L., Gujer, W., Marais, G. v. R., Matsuo, T., 1986. Activated sludge model no. 1. IAWPRC.
- Henze, M., Gujer, W., Mino, T., Matsuo, T., Wentzel, M.C., Marais, G. v. R., Van Loosdrecht, M.C.M., 1999. Activated sludge model No.2D, ASM2D. *Water Sci. Technol., Modelling and microbiology of activated sludge processes* Proceedings of the 4th IAWQ Seminar on Modelling and Microbiology of Activated Sludge Processes 39, 165–182. doi:10.1016/S0273-1223(98)00829-4
- Hilmer, T., 2008. *Water in society: integrated optimisation of Sewerage systems and wastewater treatment plants with computational intelligence tools*. Mensch und Buch, Berlin.
- ifak system GmbH, 2009. SIMBA 6.0 Users Guide.
- Janata, J., 2009. *Principles of chemical sensors*, 2nd ed. ed. Springer, Dordrecht ; New York.
- Jeppsson, U., 1996, Modeling aspects of wastewater treatment processes. Ph.D. Thesis, Department of Industrial Electrical Engineering and Automation. Lund Institute of Technology, I., Industriell elektroteknik och automation. Lunds Tekniska Högskola.
- Kaelin D, R.L., Eugster J, Rottermann K, Bänninger C, Siegrist H., 2008. Potential of in-situ sensors with ion-selective electrodes for aeration control at wastewater treatment plants. *Water Sci. Technol. Water Pollut. Res.* 58, 629–637.
- Kaufman, L., 1990. *Finding groups in data : an introduction to cluster analysis*. Wiley, New York.
- Kern, P., Wolf, C., Bongards, M., Gaida, D., 2010. Zustandsbasierte Regelung von Kläranlagen, in: *Forschungsbericht Der Fachhochschule 2010*. Joh.Heider Verlag GmbH, Köln, pp. 144–147.
- Kern, P., Wolf, C., Bongards, M., Oyetoyan, T.D., McLoone, S., 2011. Self-Organizing Map based operating regime estimation for state based control of wastewater treatment plants, in: *2011 International Conference of Soft Computing and Pattern Recognition (SoCPaR)*. Presented at the 2011 International Conference of Soft Computing and Pattern Recognition (SoCPaR), pp. 390–395. doi:10.1109/SoCPaR.2011.6089275
- Kern, P., Wolf, C., Gaida, D., Bongards, M., McLoone, S., 2014. COD and NH<sub>4</sub>-N estimation in the inflow of Wastewater Treatment Plants using Machine Learning Techniques, in: *Automation Science and Engineering (CASE), 2014 IEEE International Conference on*. IEEE, pp. 812–817.

- Kern, P., Ebel, A., Wolf, C., Bongards, M., McLoone, S., 2014. Simulation based feed stream estimation of wastewater treatment plants using a surrogate model, in: *Wastewater and Biosolids Treatment and Reuse: Bridging Modeling and Experimental Studies*, Dr. Domenico Santoro, Trojan Technologies and Western University Eds, ECI Symposium Series. [http://dc.engconfintl.org/wbtr\\_i/32](http://dc.engconfintl.org/wbtr_i/32)
- Kern, P., Bongards, M., Gros, S., Sterger, O., Rudolph, P., 2014. Energieoptimierung von Kläranlagen durch Internet-gestützte Analyse und Simulation. *WWT-Online, Modernisierungsreport 2013/2014*.
- Kern, P., Bongards, M., Cronrath, A., 2015. Wie den Stromverbrauch auf Kläranlagen reduzieren? *gwf-Wasser|Abwasser, Netzwerk Wissen, Abwasser*, ISSN 0016-3651, B5399, pp.745–756.
- Kohonen, T., 1995. *Self-organizing maps*. Springer, Berlin ;, New York.
- KSB, 2011. KSB Amaline- recirculating pump, data leaflet, URL: <https://www.ksb.com/blob/315496/57c7021f88b6988376ab1f3215feaf17/amaline-leaflet-data.pdf> (accessed 8.20.16)
- Langergraber, G., Fleischmann, N., Hofstädter, F., 2003. A multivariate calibration procedure for UV/VIS spectrometric quantification of organic matter and nitrate in wastewater. *Water Sci. Technol.* 47, 63–71.
- Lee, C.C., Lin, S. (Eds.), 2007. *Handbook of environmental engineering calculations*, 2nd ed., McGraw-Hill handbooks. McGraw Hill, New York.
- Lophaven, S.N., Søndergaard, J., Nielsen, H.B., 2002. DACE - A MATLAB Kriging Toolbox - Technical Report IMM-TR-2002-12.
- Ludwig, T., Kern, P., Bongards, M., Wolf, C., 2011. Simulation and optimization of an experimental membrane wastewater treatment plant using computational intelligence methods. *Water Science and Technology* 63, 2255–2260. doi:10.2166/wst.2011.135
- Ludwig, T., Gaida, D., Keysers, C., Pinnekamp, J., Bongards, M., Kern, P., Wolf, C., Sousa Brito, A.L., 2012. An advanced simulation model for membrane bioreactors: Development, calibration and validation. *Water Science and Technology* 66, 1384–1391. doi:10.2166/wst.2012.249
- Ludzack, F.J., Ettinger, M.B., 1962. Controlling Operation to minimize activated sludge effluent nitrogen. *Journal of the Water Pollution Control Federation* 34 (9), pp. 920-931.

- Lumley, D., 2002. On-line instrument confirmation: how can we check that our instruments are working? *Water Sci. Technol. J. Int. Assoc. Water Pollut. Res.* 45, 469–476.
- MacKay, D.J.C., 2003. *Information theory, inference, and learning algorithms*. Cambridge University Press, Cambridge, UK ; New York.
- MacQueen, J.B., 1967. Some methods for classification and analysis of multivariate observations, *Proceedings of the Fifth Berkeley Symposium on Mathematics, Statistics and Probability*, 1 (Univ. of Calif. Press), pp. 281-297.
- Matheron, G., 1963. Principles of geostatistics. *Econ. Geol.* 58, 1246–1266.
- McKay, M., 2000. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*.
- Mudrack, K., Kunst, S., Mudrack-Kunst, ... , 1991. *Biologie der Abwasserreinigung ; 102 Abbildungen, 16 Tabellen*. G. Fischer, Stuttgart [u.a.].
- Niet, A., Vrugt, M., Korving, H., Boucherie, R.J., 2011. Adaptive model based control for wastewater treatment plants, *Eleventh International Conference on Computing and Control for the Water Industry, CWWI*, 5-7 September, Exeter, UK, pp. 683-688.
- Olsson, G., 2012. ICA and me – A subjective review. *Water Res.* 46, 1585–1624. doi:10.1016/j.watres.2011.12.054
- Olsson, G., Carlsson, B., Comas, J., Copp, J., Gernaey, K.V., Ingildsen, P., Jeppsson, U., Kim, C., Rieger, L., Rodríguez-Roda, I., Steyer, J.-P., Takács, I., Vanrolleghem, P.A., Vargas, A., Yuan, Z., Åmand, L., 2014. Instrumentation, control and automation in wastewater – from London 1973 to Narbonne 2013. *Water Sci. Technol.* 69, 1373. doi:10.2166/wst.2014.057
- Olsson, G., Nielsen, M.K., Zhiguo Yuan, Lynggaard-Jensen, A., Steyer, J.-P., 2005. *Instrumentation, control and automation in wastewater systems*. IWA Pub., London.
- Otterpohl, R., Freund, M., 1992. Dynamic models for clarifiers of activated sludge plants with dry and wet weather flows. *Water Sci. Technol.* 26, 1391–1400.
- Pampalk, E., Rauber, A., Merkl, D., 2002. Using Smoothed Data Histograms for Cluster Visualization in Self-Organizing Maps, in: *In Proceedings of the International Conference on Artificial Neural Networks (ICANN'02)*. Springer, p. 871-876.
- Peterson, E.E., 1965. Chemical reaction analysis. *AIChE J.* 11, 770–955.

- Pözlbauer, G., 2004a. Application of self-organizing maps to a political dataset. Citeseer.
- Pözlbauer, G., 2004b. Survey and comparison of quality measures for self-organizing maps, Proceedings of the Fifth Workshop on Data Analysis (WDA'04), pp. 67-82, Slovakia.
- Ralston, A., Wilf, H.S., 1960. Mathematical methods for digital computers. Wiley.
- Rieger, L., Alex, J., Winkler, S., Boehler, M., Thomann, M., Siegrist, H., 2003. Progress in sensor technology--progress in process control? Part 1: sensor property investigation and classification. Water Sci. Technol. 47, 103–112.
- Rosenwinkel, 2003. Reduzierung der Gesamtemission aus Mischwasserüberlauf und Kläranlagenablauf durch Nutzung der dynamischen Online-Simulation (Abschlussbericht). Universität Hannover - Insitut für Siedlungswasserwirtschaft und Abfalltechnik, Hannover.
- Savitzky, A., Golay, M.J., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36, 1627–1639.
- Simpson, T.W., Poplinski, J.D., Koch, P.N., Allen, J.K., 2001. Metamodels for Computer-based Engineering Design: Survey and recommendations. Eng. Comput. 17, 129–150.
- Stier, E., Bundesinstitut für Berufsbildung, 2003. Handbuch für umwelttechnische Berufe Bd. 3. Bd. 3. Hirthammer, Oberhaching.
- Takacs, I., 1991. A dynamic model of the clarification-thickening process. Water Res. 25, 1263–1271. doi:10.1016/0043-1354(91)90066-Y
- Uriarte, E.A., Martín, F.D., 2005. Topology preservation in SOM. Int. J. Appl. Math. Comput. Sci. 1, 19–22.
- Usman Rehman, M.V., 2015. Effect of sensor location on controller performance in a wastewater treatment plant. Water Sci. Technol. 71 (5) pp. 700-708.
- Vienna University of Technology, 2010. The Java SOMToolbox @ IFS, TU Vienna [WWW Document]. Data Min. Java SOMToolbox. URL <http://www.ifs.tuwien.ac.at/dm/somtoolbox/index.html> (accessed 8.20.16).
- WTW GmbH, 2007. Primer: Ion selective measurement in online analysis. Weilheim.

## Appendix

### Peterson matrix for the ASM1 model

Component →	<i>i</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	Process Rate, $\rho_j$ [ML <sup>-3</sup> T <sup>-1</sup> ]	
<i>j</i> Process ↓		<i>S<sub>I</sub></i>	<i>S<sub>S</sub></i>	<i>X<sub>I</sub></i>	<i>X<sub>S</sub></i>	<i>X<sub>B,H</sub></i>	<i>X<sub>B,A</sub></i>	<i>X<sub>P</sub></i>	<i>S<sub>O</sub></i>	<i>S<sub>NO</sub></i>	<i>S<sub>NH</sub></i>	<i>S<sub>ND</sub></i>	<i>X<sub>ND</sub></i>	<i>S<sub>ALK</sub></i>		
1 Aerobic growth of heterotrophs			$\frac{1}{Y_H}$			1			$\frac{1-Y_H}{Y_H}$						$\mu_H \left( \frac{S_S}{K_S + S_S} \right) \left( \frac{S_O}{K_{O,H} + S_O} \right) X_{B,H}$	
2 Anoxic growth of heterotrophs			$\frac{1}{Y_H}$			1				$\frac{1-Y_H}{2.86Y_H}$					$\mu_H \left( \frac{S_S}{K_S + S_S} \right) \left( \frac{K_{O,H}}{K_{O,H} + S_O} \right) \left( \frac{S_{NO}}{K_{NO} + S_{NO}} \right) \eta_F X_{B,H}$	
3 Aerobic growth of autotrophs							1		$-\frac{4.57}{Y_A} + 1$	$\frac{1}{Y_A}$					$\mu_A \left( \frac{S_{NH}}{K_{NH} + S_{NH}} \right) \left( \frac{S_O}{K_{O,A} + S_O} \right) X_{B,A}$	
4 'Decay' of heterotrophs					$1-f_P$	$-1$		$f_P$							$b_H X_{B,H}$	
5 'Decay' of autotrophs					$1-f_P$		$-1$	$f_P$							$b_A X_{B,A}$	
6 Ammonification of soluble organic nitrogen														$\frac{1}{14}$	$k_a S_{ND} X_{B,H}$	
7 'Hydrolysis' of entrapped organics			1		$-1$										$k_b \frac{X_S/X_{B,H}}{K_X + (X_S/X_{B,H})} \left[ \left( \frac{S_O}{K_{O,H} + S_O} \right) + \eta_b \left( \frac{K_{O,H}}{K_{O,H} + S_O} \right) \left( \frac{S_{NO}}{K_{NO} + S_{NO}} \right) \right] X_{B,H}$	
8 'Hydrolysis' of entrapped organic nitrogen												1	$-1$		$\rho_7 (X_{ND}/X_S)$	
Observed Conversion Rates [ML <sup>-3</sup> T <sup>-1</sup> ]		$r_i = \sum_j v_{ij} \rho_j$									$r_i = \sum_j v_{ij} \rho_j$					
Stoichiometric Parameters: Heterotrophic yield: $Y_H$ Autotrophic yield: $Y_A$ Fraction of biomass yielding particulate products: $f_P$ Mass N/Mass COD in biomass: $i_{XB}$ Mass N/Mass COD in products from biomass: $i_{XP}$	Soluble inert organic matter [M(COD)L <sup>-3</sup> ] Readily biodegradable substrate [M(COD)L <sup>-3</sup> ] Particulate inert organic matter [M(COD)L <sup>-3</sup> ] Slowly biodegradable substrate [M(COD)L <sup>-3</sup> ] Active heterotrophic biomass [M(COD)L <sup>-3</sup> ] Active autotrophic biomass [M(COD)L <sup>-3</sup> ] Particulate products arising from biomass decay [M(COD)L <sup>-3</sup> ] Oxygen (negative COD) [M(-COD)L <sup>-3</sup> ] Nitrate and nitrite nitrogen [M(N)L <sup>-3</sup> ]															<b>Kinetic Parameters:</b> Heterotrophic growth and decay: $\mu_H, K_S, K_{O,H}, K_{NO}, b_H$ Autotrophic growth and decay: $\mu_A, K_{NH}, K_{O,A}, b_A$ Correction factor for anoxic growth of heterotrophs: $\eta_F$ Ammonification: $k_a$ Hydrolysis: $k_b, K_X$ Correction factor for anoxic hydrolysis: $\eta_b$
											NH <sub>4</sub> <sup>+</sup> and NH <sub>3</sub> nitrogen [M(N)L <sup>-3</sup> ] Soluble biodegradable organic nitrogen [M(N)L <sup>-3</sup> ] Particulate biodegradable organic nitrogen [M(N)L <sup>-3</sup> ] Alkalinity – Molar units					