ADM1 Parameter Calibration Method based on Partial Least Squares Regression Framework for Industrial-scale Anaerobic Digestion Modelling

by

Zhehua Xu

Thesis presented in partial fulfilment
of the requirements for the Degree

of

MASTER OF ENGINEERING
(CHEMICAL ENGINEERING)

in the Faculty of Engineering
at Stellenbosch University

Supervisor
Dr T.M. Louw

Co-Supervisor/s
Prof A.J. Burger

December 2019

DECLARATION

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: December 2019

PLAGIARISM DECLARATION

1. Plagiarism is the use of ideas, material and other intellectual property of another's work and to present is as my own.
2. I agree that plagiarism is a punishable offence because it constitutes theft.
3. I also understand that direct translations are plagiarism.
4. Accordingly all quotations and contributions from any source whatsoever (including the internet) have been cited fully. I understand that the reproduction of text without quotation marks (even when the source is cited) is plagiarism.
5. I declare that the work contained in this assignment, except where otherwise stated, is my original work and that I have not previously (in its entirety or in part) submitted it for grading in this module/assignment or another module/assignment.

Student number: ……..

Initials and surname: ……..

Signature: ……..

Date: ……..

ABSTRACT

Anaerobic Digestion Model 1 (ADM1) is the mainstay modelling tool for Anaerobic Digestion research and development. Its growing popularity is attributed to its sophisticated yet expandable structure. Not only does ADM1 encompass a broad range of biochemical, physicochemical and inhibition reactions, it provides the modeller a structured framework to add or remove reactions per application requirements. Two major challenges that ADM1 faces are the difficulty in translating common quality indicators into ADM1's 26 state variables, and the complication with calibrating a large number of model parameters – 58 by default. There is currently no consensus with regards to the parameter calibration approach. Researchers utilise various sensitivity analysis techniques to identify sensitive parameters, but the selection of parameters to be calibrated relies largely on the modeller's discretion. In some cases, decisions are simply made based on prior or expert knowledge.

Since the installation, operation and maintenance of advanced instrumentation are often expensive, most industrial digesters are inadequately monitored and thus intentionally over-designed. A model that can be used on-site with acceptable accuracy could serve as a soft sensor to forecast inhibition risks and automate preventive actions. Therefore, this study aimed to develop a standardised way to calibrate parameters when optimising ADM1 models built for industrial-scale digesters.

The proposed method, Partial Least Squares (PLS) Method, consists of four steps. In Step 1, a series of Monte Carlo simulations is carried out. For each Monte Carlo run, ADM1 is executed with all its model parameters sampled from independent probability distributions. These probability distributions were obtained by conducting a literature survey across 62 publications and all published parameters compiled into a domain which represents the uncertainty range of each parameter. In Step 2, a multivariate regression technique called PLS Regression (PLSR) is applied to the Monte Carlo results. The motives for employing PLSR are to reduce parameter dimensionality and to identify the underlying relationships between the model parameters and the model outputs. In Step 3, these relationships, which are mathematically described as PLS weights, loadings and latent variables, are utilised to guide parameter calibration. Lastly, the calibrated parameter set is validated against unseen data.

This method successfully improved, in the absence of any modeller's bias, the overall accuracy of a model based on data from an industrial-scale digester. The model is tasked to fit six typical plant measurements: Volatile Fatty Acids (VFA), ammonia, Volatile Suspended Solids (VSS), pH, methane gas flow & carbon dioxide gas flow. A configuration consisting of at least 500 Monte Carlo runs and two latent variables is required to produce a reasonably accurate fit. Although the use of more latent variables could enable PLSR to capture interactions of lesser weighted output variables, the model becomes increasingly prone to overfitting. However, it is envisaged that more latent variables would be necessary if more outputs are modelled. It is recommended to start the PLSR algorithm with one latent variable and only introduce more if necessary.

Different parameter calibration methods produce different model outcomes. The PLS Method was benchmarked against two other methods, namely the Group Method and the "Brute Force" Method. In the former method, kinetic parameters were grouped into the three groups of sensitivities (High, Medium, Low) as suggested in the ADM1 Scientific and Technical Report. The three groups are then calibrated sequentially in

order of decreasing sensitivity. The “Brute Force” Method involved calibrating all 58 parameters without any particular sequence, prioritisation or expert inputs. Lower and upper limits are, however, set as per the minimum and maximum values identified from the literature.

Besides proving to be a suitable method for industrial-scale digester modelling, the PLS Method was found to exhibit several unique traits:

- It is the only method that did not show signs of overfitting.
- It is the only method that concluded the model optimisation with all calibrated parameter values within the surveyed minimum and maximum range.
- It converges on the objective function 30-60% faster than the Group Method and 14 times quicker than the “Brute Force” Method

The success is attributed to the fundamentals of PLS regression. Unlike other regression methods where parameters are adjusted independently, PLS enables parameters to be manipulated collectively in a manner that ensures maximum impact on the outputs while considering collinearities among the parameters. This guided approach effectively mitigates the so-called “curse of dimensionality” and, potentially, overfitting and thereby speeds up the calibration process.

OPSOMMING

Anaerobiese Verteerder Model 1 (ADM1) is die hoof modelleringsinstrument vir Anaerobiese Verteerder navorsing en ontwikkeling. Sy groeiende populariteit word toegeskryf aan sy gesofistikeerde tog uitbreibare struktuur. ADM1 sluit nie net 'n wye bestek van biochemiese, fisikochemiese en inhibisie-reaksies in nie, dit verskaf ook die modelleerder met 'n gestruktureerde raamwerk om reaksies by te voeg of weg te neem in ooreenstemming met toepassingvereistes. Twee groot uitdagings wat ADM1 in die gesig staar is hoe moeilik dit is om gewone kwaliteit aanwysers in ADM1 se 26 toestandveranderlikes oor te dra, en die komplikasie met die kalibrering van 'n groot aantal model parameters – 58 by verstek. Daar is tans geen konsensus met betrekking tot die parameter-kalibrasie-benadering nie. Navorsers gebruik verskeie sensitiwiteit analisetegnieke om sensitiewe parameters te identifiseer, maar die keuse van parameters wat gekalibreer moet word steun grootliks op die modelleerder se diskresie. In sommige gevalle word besluite eenvoudig gemaak op voorafgaande of deskundige kennis.

Aangesien die installasie, bedryf en onderhoud van gevorderde instrumentasie dikwels duur is, is meeste industriële verteerders gebrekkig gemonitor en dus opsetlik oor-ontwerp. 'n Model wat op die perseel gebruik kan word met aanvaarbare akkuraatheid kan as 'n sagte sensor dien wat inhibisie risiko's kan voorspel en voorkomende aksies outomatiseer. Daarom is die doel van hierdie studie die ontwikkeling van 'n gestandaardiseerde manier om parameters te kalibreer wanneer ADM1-modelle geoptimeer word wat vir industriële verteerders gebou is.

Die voorgestelde metode, Parsiële Kleinste Kwadrate (PLS)-metode, bestaan uit vier stappe. In Stap 1, word 'n reeks Monte Carlo-simulasies uitgevoer. Vir elke Monte Carlo lopie, is ADM1 uitgevoer met al sy modelparameter monsters geneem uit onafhanklike waarskynlikheidsverdeling. Hierdie waarskynlikheidsverdeling is verkry deur 'n literatuuropname oor 62 publikasies en alle gepubliseerde parameters uit te voer en alle gepubliseerde parameters in 'n definisiegebied wat die onsekerheidsbestek van elke parameter voorstel, saam te stel. In Stap 2 word 'n meerveranderlike regressie-tegniek by name PLS Regressie (PLSR), toegepas op die Monte Carlo resultate. Die motivering om PLSR te gebruik is om parameter dimensionaliteit te verminder en om die onderliggende verhouding tussen modelparameters en die modeluitsette te identifiseer. In Stap 3 word hierdie verhoudings, wat wiskundig as PLS-gewigte, -ladings en latente veranderlikes beskryf word, gebruik om die kalibrasie van parameters te lei. Laastens word die gekalibreerde parameterstel gevalideer teen ongesiene data.

Hierdie metode het, in die afwesigheid van enige modelleerder se vooroordeel, die algehele akkuraatheid van 'n model gebaseer op data van 'n industriële-skaal verteerder, suksesvol verbeter. Die model is die taak opgelê om ses tipiese aanlegmetings te pas: VFA, ammoniak, VSS, pH, metaangasvloei en koolstofdioksiedgasvloei. 'n Konfigurasie wat uit ten minste 500 Monte Carlo-lopies en twee latente-veranderlikes bestaan, word benodig om 'n redelike akkurate passing te produseer. Al kan die gebruik van meer latente veranderlikes PLSR in staat stel om interaksies van minder gewigtige uitsetveranderlikes te vang, word die model meer geneig tot oorpassing. Dit word egter verwag dat meer latente-veranderlikes nodig sal wees as meer uitsette gemodelleer word. Dit word voorgestel om die PLSR-algoritme met een latente-veranderlike te begin en slegs meer in te voeg soos nodig.

Verskillende parameter kalibrasie metodes produseer verskillende model uitkomste. Die PLS-Metode is genormeer teen twee ander metodes, naamlik die Groep Metode en die “Brute Krag” Metode. In die eersgenoemde metode, is kinetiese parameters gegroepeer in drie groepe van sensitiwiteit (Hoog, Medium, Laag) soos voorgestel in die ADM1 Scientific and Technical Report. Die drie groepe word dan sekwensieel gekalibreer in orde van afnemende sensitiwiteit. Die “Brute Krag” Metode sluit kalibrasie van al 58 parameter in, sonder enige besondere orde, prioritisering of deskundige insette. Laer en hoër limiete is egter gestel soos per die minimum en maksimum waardes uit die literatuur geïdentifiseer.

Buiten die bewys dat dit ’n gepaste model is vir modellering van industriële-skaal verteerders, is die PLS-Metode gevind om verskeie unieke eienskappe te vertoon:

- Dit is die enigste metode wat nie tekens van oorpassing gewys het nie.
- Dit is die enigste metode wat die model optimering met al die gekalibreerde parameterwaardes binne die opname se minimum en maksimum bestek, gesluit het.
- Dit konvergeer 30–60% vinniger na die doelfunksie as die Groep Metode en 14 keer vinniger as die “Brute Krag” Metode.

Die sukses word toegeskryf aan die grondslag van PLS-regressie. Anders as ander regressiemetodes waar parameters onafhanklik aangepas word, stel PLS-konstruksies parameters in staat om gesamentlik gemanipuleer te word op ’n manier wat maksimum impak op die uitsette verseker terwyl kolineariteite onder parameters oorweeg word. Hierdie geleide benadering versag effektief die sogenaamde “vloek van dimensie” en, moontlik, oorpassing en daarby versnel dit die kalibrasieproses.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisors, Dr T.M Louw and Prof. A.J. Burger, for their continued faith and patience in me.

This journey would not have begun without those encouraging words from Prof. Burger during our first meeting. Apart from gaining theoretical knowledge, this research was also highly rewarding from a personal upliftment perspective. I have learnt to instil self-discipline and to maintain a healthy balance between work and life.

I am grateful to have Dr Louw as my supervisor. His enthusiasm and spontaneity in challenging uncharted research areas are truly inspiring. I have gained the courage to study new complex topics and think outside the box. Thank you for sharing your expertise in analytics and computational modelling. These skill sets will follow me throughout my career.

A special thanks to the dairy factory who generously provided the data for this research. Without it, the research would not be possible. Similarly, all researchers and contributors of ADM1 research are acknowledged for the data obtained in their study.

Appreciation is due to my company who sponsored this study.

Most importantly, to my family members, thank you all for your unwavering encouragement and understanding during this period. I will strive to use my learnings and give back to society, as well as to inspire the next generation of engineers.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Glossary	Description
AD	Anaerobic Digestion
ADM1	Anaerobic Digestion Model No. 1
BOD	Biochemical Oxygen Demand [mg/l]
CSTR	Continuously Stirred Tank Reactor
DS	Dry Solids content of the sludge [wt%]
GC	Gas Chromatography
HPLC	High Performance Liquid Chromatography
IQR	Interquartile Range
LCFA	Long-Chain Fatty Acids
LHS	Latin Hypercube Sampling
MAPE	Mean Absolute Percentage Error
MC	Monte Carlo
PCA	Principal Component Analysis
PCOD	Particulate Chemical Oxygen Demand
PCR	Principal Component Regression
PLSR	Partial Least Squares Regression
RMSE	Root Mean Squared Error
SCOD	Soluble Chemical Oxygen Demand
SCVFA	Short-chain Volatile Fatty Acids
SRT	Solids Retention Time
STR	Scientific and Technical Report
TCOD	Total Chemical Oxygen Demand
VFA	Volatile Fatty Acids [mg/l]
VSS	Volatile Suspended Solids [mg/l]
WAS	Waste-activated Sludge
WWTP	Wastewater Treatment Plant
VFA_{COD}	Total COD concentration of volatile fatty acids

NOMENCLATURE

ADM1 Dynamic State Variables

i	Variable	Unit	Description	*i*	Variable	Unit	Description
1	S_{su}	kgCOD/m³	Monosaccharides	16	X_{li}	kgCOD/m³	Lipids
2	S_{aa}	kgCOD/m³	Amino acids	17	X_{su}	kgCOD/m³	Monosaccharide degraders
3	S_{fa}	kgCOD/m³	Total LCFA	18	X_{aa}	kgCOD/m³	Amino acid degraders
4	S_{va}	kgCOD/m³	Total valerate	19	X_{fa}	kgCOD/m³	LCFA degraders
5	S_{bu}	kgCOD/m³	Total butyrate	20	X_{c4}	kgCOD/m³	C4-degraders
6	S_{pro}	kgCOD/m³	Total propionate	21	X_{pro}	kgCOD/m³	Propionate degraders
7	S_{ac}	kgCOD/m³	Total acetate	22	X_{ac}	kgCOD/m³	Acetate degraders
8	S_{h2}	kgCOD/m³	Hydrogen	23	X_{h2}	kgCOD/m³	Hydrogen degraders
9	S_{ch4}	kgCOD/m³	Methane	24	X_{I}	kgCOD/m³	Particulate inerts
10	S_{IC}	kmol C/m³	Inorganic carbon	25	S_{an}	kmol/m³	Anions
11	S_{IN}	kmol N/m³	Inorganic nitrogen	26	S_{cat}	kmol/m³	Cations
12	S_{I}	kgCOD/m³	Soluble inerts	27	$S_{h2,g}$	kgCOD/m³	Hydrogen (gas)
13	X_{c}	kgCOD/m³	Composites	28	$S_{ch4,g}$	kgCOD/m³	Methane (gas)
14	X_{ch}	kgCOD/m³	Carbohydrates	29	$S_{co2,g}$	kgCOD/m³	Carbon dioxide (gas)
15	X_{pr}	kgCOD/m³	Proteins				

ADM1 Stoichiometric Parameters

Parameter	Description	Parameter	Description
$f_{SI,XC}$	Soluble inerts fraction from composites	$f_{PRO,SU}$	Propionate fraction from monosaccharides
$f_{XI,XC}$	Particulate inerts fraction from composites	$f_{AC,SU}$	Acetate fraction from monosaccharides
$f_{CH,XC}$	Carbohydrates fraction from composites	$f_{H2,AA}$	Hydrogen fraction from amino acids
$f_{PR,XC}$	Proteins fraction from composites	$f_{VA,AA}$	Valerate fraction from amino acids
$f_{LI,XC}$	Lipids fraction from composites	$f_{BU,AA}$	Butyrate fraction from amino acids
$f_{FA,LI}$	Fatty acids fraction from lipids	$f_{PRO,AA}$	Propionate fraction from amino acids
$f_{H2,SU}$	Hydrogen fraction from monosaccharides	$f_{AC,AA}$	Acetate fraction from amino acids
$f_{BU,SU}$	Butyrate fraction from monosaccharides		

ADM1 Kinetic Parameters

Parameter	Unit	Description
k_{dis}	d^{-1}	Disintegration factor
k_{hyd_CH}	d^{-1}	Carbohydrates hydrolysis rate constant
k_{hyd_PR}	d^{-1}	Proteins hydrolysis rate constant
k_{hyd_LI}	d^{-1}	Lipids hydrolysis rate constant
K_{s_IN}	$kmol/m^3$	Inorganic nitrogen concentration threshold; growth ceases when exceeded
pH_{UL_acid}	-	pH threshold; no inhibition when pH is above this level
pH_{LL_acid}	-	pH threshold; full inhibition when pH is below this level
k_{m_su}	$COD.COD^{-1}.d^{-1}$	Monod maximum specific uptake rate for monosaccharide degraders
K_{s_su}	$kgCOD.m^{-3}$	Monod half saturation value for monosaccharide degradation
Y_{su}	$COD.COD^{-1}$	Biomass yield on uptake of monosaccharides
k_{dec_xsu}	d^{-1}	Decay rate constant of monosaccharide degraders
k_{m_aa}	$COD.COD^{-1}.d^{-1}$	Monod maximum specific uptake rate for amino acid degraders
K_{s_aa}	$kgCOD.m^{-3}$	Monod half saturation value for amino acid degradation
Y_{aa}	$COD.COD^{-1}$	Biomass yield on uptake of amino acids
k_{dec_xaa}	d^{-1}	Decay rate constant of amino acid degraders
k_{m_fa}	$COD.COD^{-1}.d^{-1}$	Monod maximum specific uptake rate for LCFA degraders
K_{s_fa}	$kgCOD.m^{-3}$	Monod half saturation value for LCFA degradation
Y_{fa}	$COD.COD^{-1}$	Biomass yield on uptake of LCFA
k_{dec_xfa}	d^{-1}	Decay rate constant of LCFA degraders
KI_{h2_fa}	$kgCOD.m^{-3}$	Hydrogen inhibitory concentration for LCFA degraders
k_{m_c4}	$COD.COD^{-1}.d^{-1}$	Monod maximum specific uptake rate for valerate & butyrate degraders
K_{s_c4}	$kgCOD.m^{-3}$	Monod half saturation value for valerate & butyrate degradation
Y_{c4}	$COD.COD^{-1}$	Biomass yield on uptake of valerate & butyrate
k_{dec_xc4}	d^{-1}	Decay rate constant of valerate & butyrate degraders
k_{m_pro}	$COD.COD^{-1}.d^{-1}$	Monod maximum specific uptake rate for propionate degraders
K_{s_pro}	$kgCOD.m^{-3}$	Monod half saturation value for propionate degradation
Y_{pro}	$COD.COD^{-1}$	Biomass yield on uptake of propionate
k_{dec_xpro}	d^{-1}	Decay rate constant of propionate degraders
KI_{h2_pro}	$kgCOD.m^{-3}$	Hydrogen inhibitory concentration for propionate degraders
k_{m_ac}	$COD.COD^{-1}.d^{-1}$	Monod maximum specific uptake rate for acetate degraders
K_{s_ac}	$kgCOD.m^{-3}$	Monod half saturation value for acetate degradation
Y_{ac}	$COD.COD^{-1}$	Biomass yield on uptake of acetate
k_{dec_xac}	d^{-1}	Decay rate constant of acetate degraders

Parameter	Unit	Description
KI_{nh3_ac}	$kgCOD.m^{-3}$	Free ammonia inhibitory concentration on acetate degraders
pH_{UL_ac}	-	pH threshold; no inhibition on acetate degradation when pH is above this level
pH_{LL_ac}	-	pH threshold; full inhibition on acetate degradation when pH is below this level
k_{m_h2}	$COD.COD^{-1}.d^{-1}$	Monod maximum specific uptake rate for hydrogen degraders
K_{s_h2}	$kgCOD.m^{-3}$	Monod half saturation value for hydrogen degradation
Y_{h2}	$COD.COD^{-1}$	Biomass yield on uptake of hydrogen
K_{dec_xh2}	d^{-1}	Decay rate constant of hydrogen degraders
pH_{UL_h2}	-	pH threshold; no inhibition on acetate degradation when pH is above this level
pH_{LL_h2}	-	pH threshold; full inhibition on acetate degradation when pH is below this level

CHAPTER 1

INTRODUCTION

1.1. Background

Anaerobic digestion (AD) has remained the mainstream approach for treating high strength organic waste since its invention as a waste/wastewater treatment technology. Its application is wide: ranging from municipal wastes such as municipal solid wastes and sewage sludge to industrial wastes such as livestock manure and food processing wastewater. Co-digestion of municipal wastes in combination with industrial wastes is also a well-accepted application (Angelidaki & Ellegaard, 2003).

Contrary to aerobic biological processes, anaerobic digestion can operate at significantly higher organic loading rates (i.e. more compact), produces lesser sludge as well as recovers energy from the waste as biogas (McCarty, 1964). These advantages are undeniably attractive to the industries because on-site space designated for waste treatment is often limited; plus the ever-increasing drive to reduce utility costs, and to create an environmentally sustainable image. Being a low energy-intensive process, the treatment plant is generally net energy positive, meaning that excess energy could be repurposed for other users in the form of either electricity or heat/steam. Another major cost-saving, which often left unaccounted for, is the disposal cost and penalties that would otherwise be incurred if no treatment was undertaken.

In spite of the benefits, modern anaerobic digestion plants still rely heavily on human monitoring and inputs due to lack of affordable advanced instrumentations. Anaerobic processes, in contrast to aerobic processes, require operators with higher technical abilities because the system is more susceptible to process upsets (Madsen, Holm-Nielsen & Esbensen, 2011). A severe process failure would require long periods to recover, and the financial impact of such a scenario remains the primary concern of adopting this technology (McCarty, 1964). To reduce risk, designers tend to undertake a conservative approach by selecting a lower organic loading rate intentionally (i.e. oversizing the digesters). This approach results in design redundancy and capital wastage.

Nonetheless, by 2007, there was already more than of 2250 anaerobic digestion plants implemented globally for treating industrial type wastewater (Van Lier, 2008). The field of application continues to broaden thanks to the immense research effort on the microbial, biochemical and physicochemical mechanisms within AD. It has led to the development of higher-rate reactors, wider digester operating temperature ranges and advancement in modelling and process control techniques. (Costello, Greenfield & Lee, 1991; Ge, Jensen & Batstone, 2011; Jimenez *et al.*, 2015).

1.2. Research Motivation and Rationale

Troubleshooting a full-scale anaerobic digestion plant is not straightforward; generally relies on operational experience and a trial and error approach. The difficulty is attributed to a lack of on-line process monitoring, automated diagnosis and management. Online instruments commonly employed on industrial scale are basic in functionality because advanced instruments are expensive, intricate and require a higher level of maintenance and calibration (Steyer *et al.*, 2002; Vanrolleghem & Lee, 2003).

Many useful data are acquired through manual sampling followed by offline analysis in a laboratory. Some constituents may be analysed on- site if simple and economical to perform, but others may require an external better-equipped laboratory. This process is cumbersome and does not allow the reactor to be managed via dynamic feedback; instead, it relies entirely on the plant operators' experience (Madsen *et al.*, 2011; Spanjers & Lier, 2006). Moreover, many sites do not have dedicated technical personnel to interpret the collected data correctly.

A process diagnosis & management tool, based on basic on-site obtainable data as inputs, would be highly valuable. The tool could be designed to foresee instability and to initiate corrective actions. Furthermore, since most existing digesters were designed rather conservatively, this tool would allow one to operate above the initial design set-points and exploit the true effective capacity of the digester (Liu, Olsson & Mattiasson, 2004). Developing such a tool before the advent of affordable advanced instruments would, otherwise, first require a reasonably accurate model.

Several mathematical models have been developed for AD process modelling. Despite differences in model structure and number of biochemical conversion processes incorporated, accuracy of the model outputs are fundamentally governed by the values that are assigned as the stoichiometric and kinetic parameters. *Parameter calibration* is thus an important step in model development.

Models with higher degrees of sophistication generally feature higher number of parameters. For instance, the most widely used model of recent years called Anaerobic Digestion Model No. 1 (ADM1) has 58 parameters by default. The high degrees of freedom mean that the calibration process can become extremely time-consuming if the parameters are adjusted one at a time. As such, sensitivity analysis techniques or expert knowledge are commonly employed to reduce the degrees of freedom. There is currently no common protocol regarding parameter calibration. Selection of parameters to be calibrated relies largely on the modeller's discretion. A method that is free from modeller's bias could offer a standardised way to calibrate ADM1 parameters.

Partial Least Squares Regression (PLSR) is a popular statistical tool used in chemometrics to identify and to regress the non-linear relationship between two matrices as a linear model. It is envisaged that this tool could present a methodical way to relate *ADM1 parameters* and the *model outputs*. An integral part of PLSR is dimensionality reduction. This means that a large number of model parameters could potentially be transformed into a smaller subset called latent variables, and simplify the calibration process due to the reduced degrees of freedom. In theory, a reduction in bias avoids overfitting.

Soft sensors have a prospective role in industrial AD plant operation. Using online measurements from conventional field sensors as inputs to a process model, a soft sensor allows other valuable process measurements that are expensive (e.g. microbial species) or time-consuming to analyse (e.g. VFA species) to be predicted instead. Access to this information would empower advanced monitoring and control strategies that further enhance the digester's stability.

As pointed out in a review paper by *Jimenez et al.* (2015), research involving soft sensoring generally utilise simplified models instead of ADM1. Simplified models can predict basic lumped measurements (e.g. total COD, VFA, etc.) which are appropriate for basic plant control, but inadequate for ultimate soft sensor development as it lacks the level of sophistication that ADM1 provides. Challenges that ADM1 faces are the high degrees of freedom and the model's requirement for detailed substrate composition. As ADM1 remains the forefront of AD modelling research, the ultimate goal should be aimed at addressing these challenges such that ADM1 is compatible for soft sensor applications.

1.3. Research Aim and Limitations

This research aims to develop a parameter calibration method that could be included in the ADM1 framework. The method shall adopt the concepts of PLSR and demonstrate whether model optimisation can be achieved without the need to manually select parameters for calibration. Performance of this method shall be benchmarked against other calibration methods.

The data available for the modelling demonstration in this thesis were sourced from the operational data log of an industrial-scale AD plant. Additional in-depth characterisation tests or change to sample frequency were not possible because access to the plant is restricted. Analytical constraints are normal for industrial applications due to affordability reasons (Arnell *et al.*, 2016). This limitation, however, did not impede the research objectives because the research focuses on framework development rather than the fit accuracy. In fact, for an industrial setting, high accuracy of every model outputs may not be unnecessary, as models are used for assessing changes in output trends (Batstone & Keller, 2003).

1.4. Research Questions and Objectives

1.4.1. Research Questions

1. How can the concept of Partial Least Squares Regression be exploited for parameter calibration?
2. What are the procedures when applying this method for industrial-scale anaerobic digestion modelling?
3. Is this method capable of simplifying the calibration process and avoid over-calibration?

1.4.2. Objectives

- Construct an ADM1 model on a numerical computational platform

- Translate experimental data obtained from an industrial-scale digester into the format required by ADM1
- Conduct a literature survey for stoichiometric and kinetic parameter values to establish the variance in each parameter
- Perform Monte Carlo simulation using the surveyed information to generate input and output data set
- Identify latent relationships between model parameters and outputs using PLSR algorithm and data set generated from Monte Carlo simulation
- Attempt parameter calibration by using the latent relationships as guidance
- Benchmark the new calibration method against other calibration methods

CHAPTER 2

LITERATURE REVIEW

2.1. Anaerobic Digestion Theory

2.1.1. Fundamental Biochemical Reactions

Hydrolysis

Hydrolysis (also referred to as solubilisation) involves the disintegration of complex, insoluble polymeric matter by extracellular enzymes into structurally smaller products. In this step, as illustrated in Figure 1, carbohydrates, proteins and lipids are broken down predominantly into monosaccharides, amino acids and LCFAs, respectively (Heukelekian, 1958). Being soluble, these products can enter the biomass and undergo further breakdown intracellularly. Carbohydrates were found to hydrolyse faster than proteins and lipids (Eastman & Ferguson, 1981).

Acidogenesis

The process of acidogensis (also referred to as fermentation) follows hydrolysis. In this step, acidogens convert monosaccharides into predominantly low molecular weight VFAs such as acetate, propionate and butyrate, lactic acid, H_2 and CO_2 (Gujer & Zehnder, 1983). Lactic acid is an intermediary compound that converts rapidly into VFAs; however, it could potentially accumulate within the digester when a high load of readily degradable substance such as glucose (Costello *et al.*, 1991) is received. The acidification products of amino acids are higher molecular weight VFAs such as i-butyrate, valerate and i-valerate, H_2, CO_2, ammonium and sulphides. Hydrogen production is mostly related to the acidification of monosaccharides rather than amino acids.

For proteinaceous substrates, hydrolysis is regarded as the rate-limiting step because the rate of fermentation is considerably faster (Pavlostathis & Gossett, 1988). Yu & Fang (2001) further observed, in a study on dairy wastewaters, that carbohydrates tend to suppress the degradation of proteins. This causes carbohydrates to acidify preferentially and more rapidly in comparison to proteins and lipids.

Fermentation pathways of LCFAs depend on the carbon structure of the acids. If the acid has an odd number of carbon atoms, both acetate and propionate will form. However, if the acid has even carbon counts, acetate is the only short-chained VFA that will be formed (McInerney & Bryant, 1981). Another fermentation product is molecular hydrogen. Hydrogen serves as a sink for the electrons liberated when LCFAs are oxidised (Gujer & Zehnder, 1983).

Acidification favours the VFA and hydrogen production pathway when the substrate COD concentration is low. However, at a higher COD concentration, the pathway would shift towards alcohols such as propanol and butanol when excessive amounts of low molecular weight VFAs accumulate. Gottschal & Morris (1981)

explained this metabolic shift as a mechanism for acidifiers to counter the VFA build-up and consequential pH inhibition. The pathway change is reported to trigger only when acetate or butyrate exceed a threshold concentration of 0.4 - 0.6 g/l (Jones & Woods, 1986).

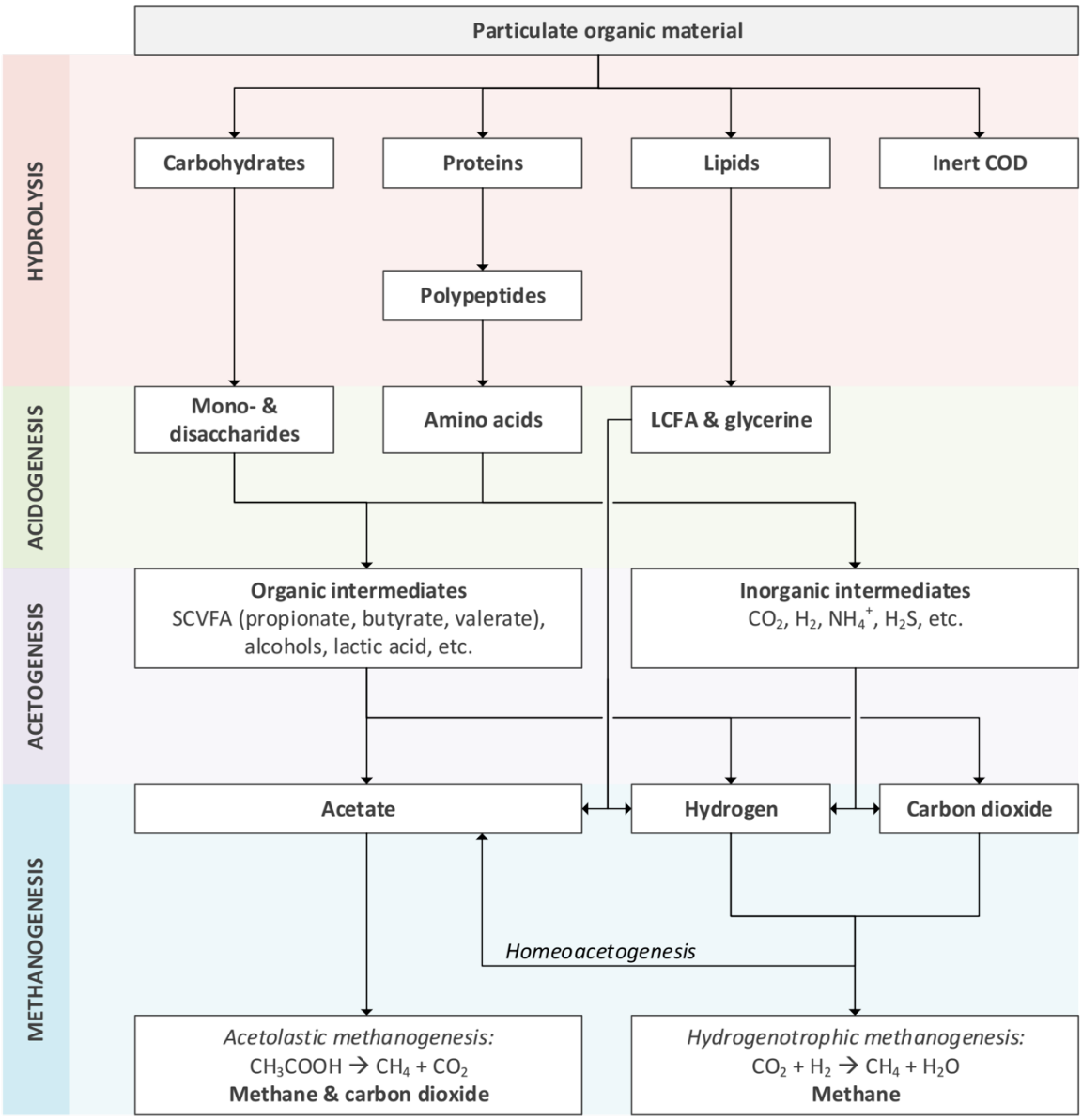


Figure 1: Breakdown of complex organic material to simpler components during anaerobic digestion - process scheme adapted from Gujer & Zehnder (1983); Siegrist et al. (2002); Madsen, Holm-Nielsen & Esbensen (2011)

Acetogenesis

According to McCarty & Smith (1986), the conversion from ethanol and propionate to acetate and hydrogen requires a Gibbs free energy $\Delta G^{o'}$ of 9.65 kJ/mol and 71.67 kJ/mol, respectively (Equation 1 & Equation 2). Conversion of other intermediate VFAs such as butyrate and valerate also holds positive free energy. This fact implies that these reactions will remain non-spontaneous until acetate and/or hydrogen concentrations are low enough to induce a negative free energy.

Hydrogen is consumed by H_2-utilising bacteria during methanogenesis to produce methane. Under sufficiently low hydrogen level, acidogens are observed to deviate from the ethanol (C_2H_6O) production pathway because ethanol acts as an electron sink. Instead, H_2 is produced from the oxidation of NADH, a process which leads to a preferential formation of acetate ($C_2H_3O_2^-$) (Wolin, 1982). The production ratio between acetate and ethanol is therefore dependent on the concentration of H_2-utilising methanogenic bacteria present during fermentation.

Oxidation of ethanol: $$CH_3CH_2OH + H_2O \rightarrow CH_3COO^- + H^+ + 2H_2$$ *Equation 1*

Oxidation of propionate: $$CH_3CH_2COO^- + 2H_2O \rightarrow CH_3COO^- + 3H_2 + CO_2$$ *Equation 2*

Methanogenisis

Methanogensis refers to the final carbon degradation step in which methane gas is produced. This process only occurs when all alternative forms of electron acceptors (e.g. O_2, NO_3^-, SO_4^{2-},) are depleted. Two major pathways are well known: (i) the uptake of acetic acid by acetoclastic methanogens; and (ii) the reduction of carbon dioxide by hydrogenotrophic methanogens.

The first pathway, termed acetolastic methanogenesis, follows the oxidation of acetate into carbon dioxide and methane (Equation 3). In the second pathway, termed hydrogenotrophic methanogenesis, methane is formed through the reduction of carbon dioxide by hydrogen (Equation 4). This pathway takes place only when acetate is depleted and carbon dioxide left as the sole electron acceptor. Theoretically, up to a third of the total methane could be produced via this route (Conrad, 1999).

$$CH_3COOH \rightarrow CH_4 + CO_2$$ *Equation 3*

$$\frac{1}{2}CO_2 + 2H_2 \rightarrow \frac{1}{2}CH_4 + H_2O$$ *Equation 4*

Hydrogenotrophic methanogenesis serves as an important sink for reducing hydrogen concentration in the digester. If hydrogen is allowed to accumulate, for example, due to suppressed methanogenic activity, fatty acid degrading organisms will become inhibited and initiate the reduction of low molecular VFAs (i.e. acetate, propionate) into alcohols and higher molecular VFAs (i.e. butyrate). This impedes methane production consequently because methanogens utilise products from acetogenesis as substrates (McInerney & Bryant, 1981).

On the contrary, if hydrogen is effectively consumed and kept below the inhibitory level, hydrogen production will regulate in conjunction with the hydrogen partial pressure (Pavlostathis & Giraldo-Gomez, 1991). Sustaining the syntrophy between acetogenic organisms (which produces hydrogen) and hydrogenotrophic methanogens (which consumes hydrogen) is therefore crucial for efficient AD operation.

Sulphate Reduction

Under anaerobic conditions, sulphate will be reduced to sulphide first before methane production occurs. The reason is that the biological reduction of sulphate is slightly more thermodynamically favoured than methanogenesis. The reaction (Equation 5) is mediated by sulphate reducing bacteria (SRB) which competes for the same electron donors as methanogens, and as a consequence, lesser acetate and hydrogen are available for methane production (Kalyuzhnyi & Fedorovich, 1998). Theoretically, 0.67 mg/l of COD is required to reduce 1 mg/l of sulphate (Liamleam & Annachhatre, 2007).

$$CH_3COO^- + SO_4^{2-} \rightarrow HS^- + HCO_3^- \qquad \textit{Equation 5}$$

Denitrification

Nitrate will undergo a series of reduction processes, termed denitrification, when subject to anaerobic conditions. Facultative anaerobic bacteria tend to use nitrate as an electron acceptor because the reduction reaction is highly favoured thermodynamically. During denitrification, nitrate is converted into ammonia and nitrogen gas, via nitrite as an intermediate (Tiedje, 1988).

The conversion path to ammonia, called ammonification, is a selective process that depends on the ammonia concentration. Under ammonia-limiting or nitrate-limiting conditions, ammonifiers compete well against denitrifiers which mediate the reduction of nitrate (Equation 6) and nitrite (Equation 7). As the reduction process utilises organic carbon, the theoretical COD required for (or loss due to) denitrification is 2.86 mg/l and 1.71 mg/l per mg/l of NO_3 and NO_2, respectively (Akunna, Bizeau & Moletta, 1992). This fact effectively implies that methane gas production will decrease when denitrification occurs.

$$NO_3^- + \frac{5}{24}\mathrm{C_6H_{12}O_6} + \mathrm{H_2O} \rightarrow \frac{1}{2}N_2 + \frac{5}{4}CO_2 + \frac{7}{4}H_2O + OH^- \qquad \textit{Equation 6}$$

$$NO_2^- + \frac{1}{8}\mathrm{C_6H_{12}O_6} + \mathrm{H_2O} \rightarrow \frac{1}{2}N_2 + \frac{3}{4}CO_2 + \frac{5}{4}H_2O + OH^- \qquad \textit{Equation 7}$$

2.1.2. Anaerobic Digestion Kinetics

Kinetics are key factors when developing dynamic models. They describe the rates of biomass growth, substrate uptake, product formation and microbial decay during biochemical reactions. Good understanding of microbial decay kinetics is particularly essential for anaerobic digestion because specific growth rates are much lower than aerobic processes (Pavlostathis & Giraldo-Gomez, 1991).

Several mathematical equations have been derived to describe biological growth kinetics. In general, they are based on the relationship between growth rate and substrate concentration. The kinetic model applied in ASM (Activated Sludge Model) is the Monod equation (Monod, 1942). It has further found success in describing anaerobic digestion processes other than the hydrolysis process which is best modelled by first-order kinetics.

The Monod equation describes microbial growth by relating it to the growth-limiting substrate's concentration (Equation 8). Parameters such as μ_{max} and K_s are empirical, which are obtained by fitting the observed substrate utilisation data using non-linear regression (Robinson & Tiedje, 1983). These parameters can differ between different studies even if the substrate is similar. Pavlostathis & Giraldo-Gomez (1991) explained this variability as a result of difference in mode of operation (batch or continuous) and/or operating conditions (e.g. temperature, pH) applied during each study.

$$\mu = \mu_{max} \frac{S}{K_s + S} \qquad \textit{Equation 8}$$

where:

μ is the specific microorganism growth rate [d^{-1}];

μ_{max} is the maximum specific microorganism growth rate [d^{-1}];

S is the limiting substrate concentration [$kgCOD.m^{-3}$];

K_s is the half-saturation constant;

From the Monod equation, the substrate uptake rate (k_m) can be calculated by relating the growth rate to the biomass yield ratio (Equation 9).

$$k_m = \mu_{max} Y \qquad \textit{Equation 9}$$

2.1.3. Toxicity & Inhibition

Toxicity related to pH is the most commonly encountered form of toxicity. It is caused by the presence of weak acids e.g. hydrogen sulphide (H_2S) and unionised VFAs, and weak bases e.g. unionised ammonium. Other forms of toxicity, such as biocides and heavy metals, are described further in the literature by van Haandel and van der Lubbe (2007).

Sulphides are formed when sulphate ions are reduced during methanogenesis or when sulphur-containing amino acids (cysteine and methionine) are degraded. H_2S, which is a toxin for methanogenic bacteria, exists in equilibrium with HS^- whereby their concentrations are functions of pH (Equation 10). The pK_a value of the H_2S/HS^- equilibrium is 6.99 at 30°C. This implies that at a pH of 6.99, 50% of the sulphides will be in the H_2S form and increasing concentrations at lower pH values.

$$H_2S \leftrightarrow HS^- + H^+ \qquad \textit{Equation 10}$$

Although each VFA has different pK_a values, VFA exists predominantly in the unionised form at lower pH values. Unionised VFAs are considered toxic because it can diffuse through the cell membrane and cause cell lysis due to a large internal pH drop as it dissociates within. The loss in microbial activity inhibits the degradation of hydrogen and organic acids (Siegrist, Renggli & Gujer, 1993).

An easily acidifying substrate with high organic content could lead to rapid accumulation of acetic acid. Unless sufficient alkalinity buffer is present or corrective action taken to reduce the influent loading rate, the decrease in pH will amplify the above-mentioned inhibition effects.

Ammonia is generated when the protein component of a waste is digested. Given the fact that NH_3/NH_4^+ equilibrium has a pKa of 9.3 at 25°C, ammonia toxicity is only a concern at high pH where the unionised form prevails. Microbial activity is reported to be inhibited when the free NH_3 level reaches 150 mg/l (Braun, Huber & Meyrath, 1981).

2.1.4. Commonly Monitored Process Indicators

This section discusses the relevance of measurements commonly monitored at an industrial-scale AD plant. Except for pH, temperature and volumetric flow rates, most of the measurements are off-line analyses from manually taken samples. Feedback control strategies are thus limited due to the long measurement delay. In spite of significant advancement made in recent years to develop sensors that allow in-situ measurements, affordability remains a major obstacle to the deployment of advanced instruments at industrial plants in developing countries (Jimenez *et al.*, 2015).

COD

Chemical Oxygen Demand (COD) indicates the generic pollution strength within a wastewater stream. It refers to the amount of oxygen required to completely oxidise the organic material present. Total COD (TCOD) is a term used to describe the COD content of a well-mixed wastewater sample inclusive of all entrained solids. Soluble COD (SCOD) refers to the COD of a filtered or centrifuged wastewater sample, whilst the contribution of the suspended solids to TCOD is referred to as particulate COD (PCOD). The relationship between the terms is described in Equation 11.

$$TCOD = SCOD + PCOD \qquad \textit{Equation 11}$$

TSS & VSS

Total Suspended Solids (TSS) indicates the total quantity of suspended, or non-filterable, solid particulates in a wastewater stream. Closely related to TSS is the Volatile Suspended Solids (VSS) component which describes the concentration of volatile solid particulates. When a sample is taken from the digester, VSS represents the organics content within both influent waste and biomass; however, given a highly soluble digester feed, the VSS concentration is an approximate quantitative measure of the biomass sludge inventory within the digester. This information allows one to control the food-to-microorganism (F/M) ratio and to detect biomass loss due to wash-outs.

In anaerobic digestion, little energy is allocated to cell growth and therefore the yield/growth of anaerobic biomass – particularly LCFAs, acetate and propionate degraders - is considerably slower compared to aerobic oxidation process (McCarty & Smith, 1986; Siegrist *et al.*, 1993). This fact further emphasizes the importance of tracking VSS in maintaining digester stability.

Inert, or non-biodegradable, particulates such as inorganic precipitates (calcium carbonate, struvite, etc.) may enter the digester as part of the substrate feed or form within the digester. Its concentration is calculated as the difference between the measured TSS and VSS concentrations.

pH & Alkalinity

pH indicates how acidic or basic an aqueous solution is, and is defined as the negative logarithm of hydrogen ion concentration (Equation 12). pH is regarded as an important parameter because some constituents (e.g. H_2S, NH_3) that are known to induce microbial toxicity or inhibition exists in equilibria relative to other harmless species. Fractional amounts of each species change as a function of hydrogen ion concentration. Key equilibria of relevance in AD research are the carbonate equilibrium ($CO_2/HCO_3^-/CO_3$), hydrogen sulphide equilibrium ($S_2^-/HS^-,H_2S$), ammonium equilibrium (NH_4^+, NH_3) and VFA equilibrium.

$$pH = -\log_{10}[H^+] \qquad \textit{Equation 12}$$

Alkalinity is a measure of the digester's pH buffering capacity and a key control parameter for steady digester functioning (McCarty & Smith, 1986). Without adequate buffering, a sudden pH drop will occur when VFA formation outpaces its metabolism rate, resulting in pH related toxicity. This occurrence is common, particularly during transient conditions. External alkalinity (e.g. lime or caustic soda) could be added as a countermeasure.

The carbon dioxide-bicarbonate system serves as the main pH buffering mechanism for anaerobic digestion (Equation 13 & Equation 14). Alkalinity is naturally formed through the production of CO_2 during the metabolism of VFAs, which then converts to bicarbonate. A portion of the produced CO_2 is stripped to the biogas as gaseous CO_2, whilst most of the CO_2 remain dissolved in the liquid as bicarbonate HCO_3^-. The stripping process is the primary contributor to alkalinity recovery during anaerobic digestion.

$$CO_2 + H_2O \leftrightarrow H_2CO_3 \qquad \textit{Equation 13}$$

$$H_2CO_3 \leftrightarrow HCO_3^- + H^+ \qquad \textit{Equation 14}$$

Another source of alkalinity is ammonia, which is produced when protein or organic nitrogen in the feed substrate is mineralised during the digestion process (van Haandel & van der Lubbe, 2007). The subsequent formation of ammonium carbonate serves as alkalinity (Equation 15). Products of sulphate reduction reaction also contribute to alkalinity (Equation 5).

$$NH_3 + H_2O + CO_2 \rightarrow NH_4HCO_3 \qquad \textit{Equation 15}$$

Excess amounts of VFAs are toxic to methanogenic bacteria. In a well-buffered system, pH measurements will fail to detect the accumulation of VFAs because pH change would occur only when alkalinity is close to depletion (Hawkes *et al.*, 1993). A popular monitoring strategy employed on industrial is to maintain a good ratio of VFA relative to alkalinity, which is known as the Ripley's ratio (Ripley *et al.*, 1986).

VFA

Total VFA level has been widely reported as a reliable process status indicator (Madsen *et al.*, 2011). According to Boe *et al.* (2010), a good understanding of individual VFA constituents (n-butyric, iso-butyric and propionate in particular) could provide even better insight on the digester's stress level and potentially act as a precautionary indicator towards process imbalance.

Measurement of individual VFA constituents is conventionally performed off-line using gas chromatography or high-performance liquid chromatography (HPLC) methods. Development of automated online devices for continuous measurement on industrial-scale plants are not yet mature in terms of affordability and robustness (Jimenez *et al.*, 2015). Nonetheless, the advantage of online VFA measurement is well acknowledged and much interest has been vested in its research and development.

Temperature

The operating temperature in anaerobic digesters has a prominent effect on the degradation efficiency and biogas production. In most cases, an increase in temperature up to 42 °C, will lead to improved performance (Donoso-Bravo *et al.*, 2009; Rebac *et al.*, 1995). The reason is that kinetic rates and equilibrium coefficients which governs biochemical reactions and physicochemical processes including gas-liquid transfer are all functions of temperature. Temperature influence on biogas production is approximately 3.4% per degree from 25 - 30°C and 1.6% per degree from 30 - 35°C (Bergland, Dinamarca & Bakke, 2015).

For digesters operating in the mesophilic temperature range, the influence of temperature on biochemical reactions and physicochemical processes can be corrected by the double Arrhenius equation (Equation 16) and van't Hoff equation (Equation 17) respectively, according to Batstone (2006).

$$\rho = A_1 \, exp\left(\frac{-E_{a1}}{RT}\right) - A_2 \, exp\left(\frac{-E_{a2}}{RT}\right) \qquad \textit{Equation 16}$$

where:

ρ is the microbial activity;

A is the pre-exponential constant (empirical);

E_a is the apparent activation energy [J.mol^{-1}];

R is the universal gas constant [J.K^{-1}.mol^{-1}];

T is the absolute temperature [K]

$$ln\frac{K_2}{K_1} = \frac{\Delta H°}{R}\left(\frac{1}{T_1} - \frac{1}{T_2}\right) \qquad \textit{Equation 17}$$

where:

K_1, K_2 are the equilibrium constants before and after temperature-correction respectively;

$\Delta H°$ is the enthalpy change [J.mol^{-1}];

T_1, T_2 are the initial and final temperatures respectively [K];

2.2. Overview of Anaerobic Digestion Modelling Development

Key milestones in anaerobic digestion modelling are listed in a review by Donoso-Bravo *et al.* (2011). This section reviews some of the milestones in greater detail, as these models contributed to the development of the Anaerobic Digestion Model No. 1 (ADM1). A comprehensive list of all AD models is provided in review papers by Appels *et al.* (2008) and Lauwers *et al.* (2013).

2.2.1. Two-microbial-culture model

In 1977, Hill & Barth (1977) developed a mathematical model to simulate the digestion process of animal waste. The development was motivated by the benefits that a dynamic model would allow one to study the digester's stability when it is subjected to different operating conditions.

The model considers only two microbial groups, namely acid-formers and methane-formers, to be responsible for the conversion of organics to VFAs and from VFAs to methane gas respectively. Other variables accounted in the model are commonly analysed parameters such as volatile matter/solids, VFAs, soluble organics, cations, carbon dioxide and ammonia. This model assumes the stoichiometry of soluble organics to be the same as glucose, and that VFAs is broadly representable as acetic acid.

Variables are represented by a set of multiple non-linear differential equations that ensure mass and charge balances are maintained continuously. All insoluble organics must be solubilised first before it is amenable to degradation. This conversion is simply based on a 1:1 stoichiometry and is not subjected to any kinetic rates. The only inhibitor taken into account by the model is the presence of unionised ammonia on the growth of methane-formers.

All kinetic and physicochemical constants applied for the model were sourced from previous investigation works related to similar waste substrate, and no parameter calibration was made. For model validation, the author selected four variables which were deemed most important for animal waste digestion, namely methane gas, volatile matter/solids, VFAs and alkalinity. For the steady-state period the model was found capable of fitting actual experiment data with reasonable accuracy but failed to predict well during transient periods.

2.2.2. Steady-state acid phase model

A study by Eastman & Ferguson (1981) pointed out the likelihood of the acid phase, which consists of the hydrolysis of particulates to soluble organics and the subsequent fermentation to VFAs, to be the rate-limiting step during anaerobic digestion. In retrospect, earlier models (Andres, 1969; Graef & Andrew, 1974) considered acetolastic methanogenesis as the rate-limiting step and thus did not include any acid phase mechanisms.

Hydrolysis, in particular, was reported to be a potential rate-limiting step in anaerobic digestion, especially when digesting particulate organic substances. The overall digestion rate could be constrained during periods of organic overloading as a result of hydrolysis being the slowest of all processes. Inclusion of hydrolysis kinetics into the model structure was thus acknowledged as a milestone development. For this reason,

understanding the composition make-up between particulate and soluble organics is regarded as crucial for establishing the rate-limiting factor in context. This fact was well supported by other researchers (Gujer & Zehnder, 1983; Pavlostathis & Gossett, 1988).

The acid phase model includes a first-order function to describe the hydrolysis step and Monod's equation to describe the growth of acid-formers (in association with the utilisation of soluble organics). Recognising that different particulates within a complex substrate could hydrolyse at dissimilar rates, a first-order function was proposed as the most appropriate method to describe the lumped/effective hydrolysis effect.

The research confirmed robustness of the acid phase across a wide range of pH and/or solids concentration. Following the "two-phase digestion" approach proposed by Pohland & Ghosh (1971) where two separate reactors are operated in series but under different conditions, Eastman & Ferguson (1981) proved that the digester's stability can be greatly improved when each reactor is operated at conditions most optimally suited for the acid-forming bacteria and methanogenic archaea.

2.2.3. Dynamic single-stage high-rate anaerobic reactor model

A major advancement in anaerobic modelling was presented in the work by Costello *et al.* (1991). Using the model framework developed by Mosey (1983), this model introduced the effects of hydrogen inhibition, product inhibition, and most notably, a larger ecosystem of anaerobic bacteria which are fundamentally involved in the degradation process. Unlike prior models which only considered a single acidogenic step to form acetic acids, this model accounts for intermediate volatile acids such as propionic and butyric acids as well.

A unique development made by the author was to incorporate a comprehensive set of inhibition and regulation functions into the model structure. It is implemented by multiplying the relevant inhibition functions to a substrate's utilisation rate formula as described by Monod's equation. The inclusion of inhibition and regulation effects based on hydrogen gas concentration in the biogas is an important feature of this model.

Inhibition functions influence substrate uptake rates of acidogens and acetogens, whilst regulation functions modify the production rates of the acidifiers' end-products (propionic acid, butyric acid, acetic acid and lactic acid). Other inhibitory phenomena built into the model were:

- Individualised pH inhibition function for each type of bacteria - glucose, lactic acid, propionic acid and butyric acid
- Product inhibition effect as a result of product accumulation and high product-to-substrate ratio:
    - A competitive inhibition function was proposed for propionic acid and butyric acid bacteria. Only acetic acid is considered as the cause of inhibition. VFAs, including lactic acid, were deemed non-influential
    - A non-competitive inhibition function was proposed for glucose and lactic acid bacteria. VFAs are the cause of inhibitory products.

2.2.4. Development of higher complexity models

Angelidaki *et al.* (1993) enhanced the model initially developed by Hill & Barth (1977), with the intention to simulate free ammonia inhibition on methanogens more accurately. This improvement is critical for digesters treating substrates containing high protein or ammonia in particular. As free ammonia concentration depends on pH and temperature the research focused on improving pH prediction and the temperature correction of dissociation constants. According to the authors, free ammonia inhibition could recover spontaneously and prevent system failure. It was reported that as VFAs accumulates, the reduction in pH would cause a shift in ammonia equilibrium, resulting in lower free ammonia concentration.

One of the first universal non-substrate specific models was created by Vavilin *et al. (*1994). The model is intentionally simplified to describe key anaerobic steps only, where carbohydrates, proteins and lipids are lumped as a single hydrolysed substrate term, while propionate served as the only fatty acid intermediate. Despite having a simplified structure the model included decay mechanisms for dead biomass to assimilate back to the ecosystem as degradable and non-degradable components. The inclusion of extra processes, such as sulphate reduction and syntrophic methanogenesis, further adds sophistication to this generic model. The sulphate reduction process involves the conversion of sulphates into sulphides, with acetate and propionate as substrates. Consequently, sulphate reducing bacteria are added to the model's ecosystem.

2.3. Anaerobic Digestion Model No. 1 (ADM1)

2.3.1. Introduction

In 1997, the IWA Task Group for Mathematical Modelling of Anaerobic Digestion Processes was formed with the task to create a generic anaerobic digestion model. Developed by international experts from different anaerobic digestion modelling backgrounds the model intends to provide a common basis – in terms of the model structure, process mechanisms and nomenclature – for future modelling research such that collaboration and validation of results are possible. The model was published in 2002, under volume 13 of IWA Scientific and Technical Report (STR), as "Anaerobic Digestion Model No.1 (ADM1)".

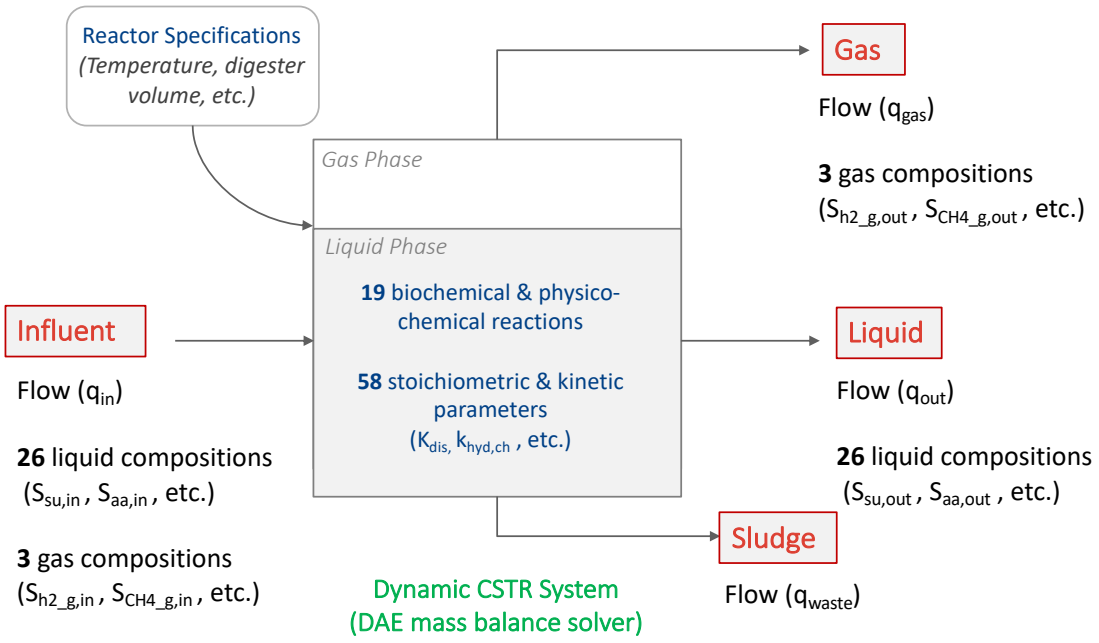


Figure 2: Block flow diagram depicting the concept of ADM1

The default/original version of ADM1 model features 29 dynamic state variables (Table 17). These variables represent 26 constituents present in the liquid phase, of which 14 are soluble and 12 are particulate, as well as 3 gas phase constituents within a CSTR system. State variables in the outlet streams are outputs calculated from the model's differential mass balance equations which describe the biochemical and physicochemical reactions taking place inside the digester. The conversion rate and kinetics of each reaction are governed by stoichiometric and kinetic parameters listed in Table 18.

2.3.2. Nomenclature and Units

Refer to Table 17 – Table 19 in the Appendix for definition of the nomenclatures used in ADM1.

2.3.3. Model Design Philosophy

In comparison with other AD models, the default version of ADM1 is already a highly sophisticated model. It has integrated physical and biochemical mechanisms widely accepted in AD research (Batstone *et al.*, 2002). Following the same format as a Petersen matrix to describe stoichiometric conversions and kinetic rates, ADM1 presents an organised structure that allows modellers to modify or add mechanisms in a manner which fellow researchers can interpret easily.

Another unique feature of ADM1 is the introduction of an additional breakdown step preceding hydrolysis called *disintegration*. This step is deemed necessary by the IWA Task Group, as there is a need to account for dead biomass and other particulates in the influent that have no distinct composition. ADM1 thus lumped these pool of particulates as a state variable called composite particulates (X_c). During disintegration, these composite particulates are broken down to carbohydrates (X_{ch}), proteins (X_{pr}), lipids (X_{li}), particulate inerts (X_I) and soluble inerts (S_I).

Effect of temperature is only partially incorporated in the default ADM1. By default, some physicochemical equilibrium constants are corrected by applying the Arrhenius equation (Equation 17). Kinetic parameters are not temperature-corrected but research has suggested that correcting kinetic parameters may improve model fitting (Bergland *et al.*, 2015). However, according to the IWA Scientific and Technical Report (STR), temperature variation less than 3 °C is known to have a negligible effect on ADM1's predictions.

Liquid Phase Mass Balance

Concentrations of soluble and particulate constituents in the liquid phase are governed by mass balance equations, Equation 18 and Equation 19, respectively. The term $\sum_{j=1}^{19} \rho_j \upsilon_{i,j}$ represents the net change in a constituent concentration (*i*) after undergoing all 19 biochemical processes (*j*). It is calculated by summing the specific kinetic rates (p_j) for each biochemical process multiplied by the corresponding stoichiometric coefficients ($v_{i,j}$). S_{IC} and S_{IN} are state variables representing inorganic carbon and inorganic nitrogen, which in the default ADM1 version refer to carbon dioxide and ammonia, respectively. These two terms serve as "sinks" for any excess carbon and nitrogen produced during the biochemical processes and ensures that the mass balance is closed.

Table 2 and Table 3 show the default Petersen matrices originally published in the STR but updated with the corrections (in blue) suggested by Rosen & Jeppsson (2006). The authors pointed out errors related to the law of conservancy of mass; therefore these corrections are considered fundamentally critical for the default ADM1. Firstly, the default parameter values for the carbon content (C_{xc}) and nitrogen content (N_{xc}) of composite particulates were found to cause carbon and nitrogen imbalances when composite particulates undergo the disintegration process as per the stoichiometric relationship proposed by the STR. To close the balance, the authors suggested that C_{xc} should be modified from 0.03 kmol C/kg COD to 0.02786 kmol C/kg COD, and N_{xc} modified from 0.002 kmol N/kg COD to 0.00269 kmol N/kg COD.

Secondly, only a selected few processes allocate their excess carbon and nitrogen to the S_{IC} and S_{IN} state variables. The authors, however, argued that the same philosophy should be applied for all 19 processes so

that carbon and nitrogen balances are truly fulfilled. These corrections must be incorporated because any discrepancy in S_{IC} and S_{IN} would influence carbon dioxide, ammonia and pH predictions.

$$\frac{dS_{liq,i}}{dt} = \frac{Q}{V_{liq}}\left(S_{in,i} - S_{liq,i}\right) + \sum_{j=1}^{19} \rho_j \upsilon_{i,j} \qquad \textit{Equation 18}$$

i = 1, 2, …, 12; i = 25, 26

$$\frac{dX_{liq,i}}{dt} = \frac{Q}{V_{liq}}\left(X_{in,i} - X_{liq,i}\right) + \sum_{j=1}^{19} \rho_j \upsilon_{i,j} \qquad \textit{Equation 19}$$

i = 13, 14, …, 24

where:

$S_{liq,i}$ represents the soluble state variables no. 1 – 12, 25 & 26 listed in Table 17;

$X_{liq,i}$ represents the particulate state variables no. 13 – 24 listed in Table 17;

Q is the influent volumetric flow, which is also equivalent to the flow exiting the digester;

V_{liq} is the digester volume;

$S_{in,i}$ and $X_{in,i}$ refer to the concentration of soluble and particulate constituents in the influent stream

Specific kinetic rates (ρ_j) are described in ADM1 in the form of Equation 20 which is effectively a combination of Monod kinetics (Equation 8) and Equation 9. The Task Group preferred to structure kinetic rates as *substrate uptake* rate instead of *biomass growth* rate because other kinetic models (e.g. inhibition, physicochemical) could be adapted easily into the model structure as add-on/extensions.

$$\rho_j = k_m \frac{S}{K_S + S} X \qquad \textit{Equation 20}$$

Gas Phase Mass Balance

By default, ADM1 considers only three gases: hydrogen, methane and carbon dioxide. These gases are represented as the last 3 state variables i.e. 27 – 29. Mass balance for the gas phase components (Equation 21) takes into account: (i) the gas outflow from the digester and (ii) the transfer of gases from the liquid phase to the gas phase, in which the rate of gas transfer is described by the general theory of two-film mass transfer (Equation 22).

$$\frac{dS_{gas,i}}{dt} = -\frac{q_{gas}}{V_{gas}} S_{gas,i} + \frac{V_{liq}}{V_{gas}} \rho_{T,i} \qquad \textit{Equation 21}$$

i = 27 - 29

$$\rho_{T,i} = k_{La}\left(K_H . P_{gas,i} - S_{liq,i}\right) \quad \textit{Equation 22}$$

where:

$S_{gas,i}$ is the gas state variables no. 27 – 29 listed in Table 17;

q_{gas} is the total gas flow exiting the digester;

V_{gas} is the gas headspace volume;

$p_{T,i}$ is the specific mass transfer rate of gas;

k_La is the volumetric gas-liquid mass transfer coefficient;

K_H is the Henry's law coefficient

ADM1 models all gases as ideal gases under the same temperature as the liquid contents (i.e. $T_{gas} = T$). The partial pressure of each gas component is thus determined per ideal gas law as described in Equation 23 - Equation 25. The denominators, 16 and 24, are COD equivalents for unit conversion purpose.

$$P_{gas,H_2} = S_{gas,H_2} . \frac{R . T_{gas}}{16} \quad \textit{Equation 23}$$

$$P_{gas,CH_4} = S_{gas,CH_4} . \frac{R . T_{gas}}{64} \quad \textit{Equation 24}$$

$$P_{gas,CO_2} = S_{gas,CO_2} . R . T_{gas} \quad \textit{Equation 25}$$

In addition to the three gases, water vapour contributes to the total headspace pressure (P_{gas}) as well because the headspace is considered to be saturated with water vapour. Its partial pressure ($P_{gas,H2O}$) is therefore discounted from P_{gas} when calculating the gas production rate q_{gas} (Equation 26). Water vapour pressure is highly dependent on temperature. Equation 27 is applied to correct the reference water vapour pressure at 298 K to the actual temperature (T).

$$q_{gas} = \frac{R . T}{P_{gas} - P_{gas,H_2O}} V_{liq} \left(\frac{\rho_{T,H_2}}{16} + \frac{\rho_{T,CH_4}}{64} + \rho_{T,CO_2}\right) \quad \textit{Equation 26}$$

where:

q_{gas} is the total gas volumetric flow leaving the digester

P_{gas} is the total headspace pressure. Under constant headspace pressure, P_{gas} is equivalent to the atmospheric pressure. Pgas = Patm = 1.013 bar

$$p_{gas,H2O} = 0.0313\ exp\left(5290\left(\frac{1}{298}\right) + \left(\frac{1}{T}\right)\right) \quad \textit{Equation 27}$$

Charge Balance

Besides mass balancing, ADM1 is set up to maintain a charge balance as well such that ionic neutrality is observed i.e. sum of cationic charges equals to the sum of anionic charges (Equation 28).

The cations group shown on the left-hand side of the equation includes hydrogen ion, ammonium ion and a cation component (S_{cat+}) which lumps any other metallic cations such as sodium and potassium that do not take part in the reaction. For the anions group, it includes bicarbonate ions, organic acids, hydroxide ions and an anionic component (S_{an-}). Similar to the cationic component, the anionic component lumps inert metallic anions such as sulphate and chloride together.

During each simulation time-step, this equation is solved for the hydrogen ion concentration (S_{H+}) and then Equation 12 is applied to determine the pH value. In other words, pH value is dependent on the concentrations of these ionic components in the digester influent as well as how these components evolve with time.

$$S_{H^+} + S_{NH_4^+} + S_{Cat^+} = S_{HCO_3^-} + \frac{S_{ac^-}}{64} + \frac{S_{pro^-}}{112} + \frac{S_{bu^-}}{160} + \frac{S_{va^-}}{208} + S_{OH^-} + S_{An^-} \qquad \textit{Equation 28}$$

where the ionic forms are calculated as follows:

$$S_{HCO_3^-} = \frac{K_{a,CO_2} S_{IC}}{K_{a,CO_2} + S_{H^+}}$$

$$S_{VFA^-} = \frac{K_{a,VFA} S_{VFA,total}}{K_{a,VFA} + S_{H^+}}$$

$$S_{NH4^+} = \frac{S_{H^+} S_{IN}}{K_{a,NH_4} + S_{H^+}}$$

$$S_{OH^-} = \frac{K_w}{S_{H^+}}$$

Inhibition Factors

As discussed in Section 2.1.3 the degradation process may be affected as a result of toxicity and inhibition. The derating effect on kinetic rates is described by applying the relevant inhibition factors (I_1, I_2…I_n) alongside the specific kinetic rates (Equation 20). The effective kinetic rate for a particular biochemical process (j) is thus as follows:

$$\rho_j = \frac{k_m S}{K_S + S} X . I_1 . I_2 \ldots I_n$$

where:

I is the inhibition function of inhibitor i on process j

An inhibition factor is a combination of one or more inhibition equations described in Table 1. Even though only four types of inhibition are featured in the default version of ADM1, it is regarded adequate to cover conditions encountered when treating ordinary substrates.

Table 1: Types of inhibition described in default ADM1

Description	Inhibited Process (j)	Equation
(1) pH Inhibition	All substrate uptake (j = 5 – 12)	If pH < pH_{UL}: $I_{pH} = \exp\left(-3\left(\frac{pH - pH_{UL}}{pH_{UL} - pH_{LL}}\right)^2\right)$ If pH > pH_{UL}: $I_{pH} = 1$
(2) Inhibition due to limited inorganic nitrogen	All substrate uptake (j = 5 – 12)	$I_{IN,lim} = \frac{1}{1 + K_{S,IN}/S_{I,IN}}$
(3) Hydrogen inhibition	LCFA, valerate, butyrate and propionate uptake only (j = 7 – 10)	$I_{h2} = \frac{1}{1 + S_{h2}/K_{I,h2}}$
(4) Free ammonia inhibition	Acetate uptake process only (j = 11)	$I_{NH3} = \frac{1}{1 + S_{NH3}/K_{I,NH3}}$

Table 2: Corrected default biochemical rate coefficients ($v_{i,j}$) and kinetic rate equations (ρ_j) for soluble organic compounds (Batstone *et al.*, 2002; Rosen & Jeppsson, 2006)

	Component → i	1	2	3	4	5	6	7	8	9	10	11	12	Kinetic Rate
j	Process ↓	S_{su}	S_{aa}	S_{fa}	S_{va}	S_{bu}	S_{pro}	S_{ac}	S_{h2}	S_{ch4}	S_{IC}	S_{IN}	S_I	(ρ_j, kg COD.m^{-3},d^{-1})
1	Disintegration										$-\sum_{i=1-9,11-24} C_i \nu_{1,i}$	$-\sum_{i=1-9,11-24} N_i \nu_{1,i}$	$f_{sI,xc}$	$k_{dis} X_c$
2	Hydrolysis of Carbohydrates	1									$-\sum_{i=1-9,11-24} C_i \nu_{2,i}$			$k_{hyd,ch} X_{ch}$
3	Hydrolysis of Proteins		1								$-\sum_{i=1-9,11-24} C_i \nu_{3,i}$			$k_{hyd,pr} X_{pr}$
4	Hydrolysis of Lipids	1 - $f_{fa,li}$		$f_{fa,li}$							$-\sum_{i=1-9,11-24} C_i \nu_{4,i}$			$k_{hyd,li} X_{li}$
5	Uptake of Sugars	-1				(1 - Y_{su}) $f_{bu,su}$	(1 - Y_{su}) $f_{pro,su}$	(1 - Y_{su}) $f_{ac,su}$	(1 - Y_{su}) $f_{h2,su}$		$-\sum_{i=1-9,11-24} C_i \nu_{5,i}$	-(Y_{su}) N_{bac}		$k_{m,su} \frac{S_{su}}{K_s + S_{su}} X_{su} I_1$
6	Uptake of Amino Acids		-1		(1 - Y_{aa}) $f_{va,aa}$	(1 - Y_{aa}) $f_{bu,aa}$	(1 - Y_{aa}) $f_{pro,aa}$	(1 - Y_{aa}) $f_{ac,aa}$	(1 - Y_{aa}) $f_{h2,aa}$		$-\sum_{i=1-9,11-24} C_i \nu_{6,i}$	N_{aa} -(Y_{aa}) N_{bac}		$k_{m,aa} \frac{S_{aa}}{K_s + S_{aa}} X_{aa} I_1$
7	Uptake of LCFA			-1			(1 - Y_{fa}) 0,7	(1 - Y_{fa}) 0,3			$-\sum_{i=1-9,11-24} C_i \nu_{7,i}$	-(Y_{fa}) N_{bac}		$k_{m,fa} \frac{S_{fa}}{K_s + S_{fa}} X_{fa} I_2$
8	Uptake of Valerate				-1		(1 - Y_{c4}) 0,54	(1 - Y_{c4}) 0,31	(1 - Y_{c4}) 0,15		$-\sum_{i=1-9,11-24} C_i \nu_{8,i}$	-(Y_{c4}) N_{bac}		$k_{m,c4} \frac{S_{va}}{K_s + S_{va}} X_{c4} \frac{1}{1 + S_{bu}/S_{va}} I_2$
9	Uptake of Butyrate					-1		(1 - Y_{c4}) 0,8	(1 - Y_{c4}) 0,2		$-\sum_{i=1-9,11-24} C_i \nu_{9,i}$	-(Y_{c4}) N_{bac}		$k_{m,c4} \frac{S_{bu}}{K_s + S_{bu}} X_{c4} \frac{1}{1 + S_{va}/S_{bu}} I_2$
10	Uptake of Propionate						-1	(1 - Y_{pro}) 0,57	(1 - Y_{pro}) 0,53		$-\sum_{i=1-9,11-24} C_i \nu_{10,i}$	-(Y_{pro}) N_{bac}		$k_{m,pro} \frac{S_{pro}}{K_s + S_{pro}} X_{pro} I_2$
11	Uptake of Acetate							-1		1 - Y_{ac}	$-\sum_{i=1-9,11-24} C_i \nu_{11,i}$	-(Y_{ac}) N_{bac}		$k_{m,ac} \frac{S_{ac}}{K_s + S_{ac}} X_{ac} I_3$
12	Uptake of Hydrogen								-1	1 - Y_{h2}	$-\sum_{i=1-9,11-24} C_i \nu_{12,i}$	-(Y_{h2}) N_{bac}		$k_{m,h2} \frac{S_{h2}}{K_s + S_{h2}} X_{h2} I_1$
13	Decay of X_{su}										$-\sum_{i=1-9,11-24} C_i \nu_{13,i}$	$-\sum_{i=1-9,11-24} N_i \nu_{13,i}$		$k_{dec,Xsu} X_{su}$
14	Decay of X_{aa}										$-\sum_{i=1-9,11-24} C_i \nu_{14,i}$	$-\sum_{i=1-9,11-24} N_i \nu_{14,i}$		$k_{dec,Xaa} X_{aa}$
15	Decay of X_{fa}										$-\sum_{i=1-9,11-24} C_i \nu_{15,i}$	$-\sum_{i=1-9,11-24} N_i \nu_{15,i}$		$k_{dec,Xfa} X_{fa}$
16	Decay of X_{C4}										$-\sum_{i=1-9,11-24} C_i \nu_{16,i}$	$-\sum_{i=1-9,11-24} N_i \nu_{16,i}$		$k_{dec,Xc4} X_{c4}$
17	Decay of X_{pro}										$-\sum_{i=1-9,11-24} C_i \nu_{17,i}$	$-\sum_{i=1-9,11-24} N_i \nu_{17,i}$		$k_{dec,Xpro} X_{pro}$
18	Decay of X_{ac}										$-\sum_{i=1-9,11-24} C_i \nu_{18,i}$	$-\sum_{i=1-9,11-24} N_i \nu_{18,i}$		$k_{dec,Xac} X_{ac}$
19	Decay of X_{h2}										$-\sum_{i=1-9,11-24} C_i \nu_{19,i}$	$-\sum_{i=1-9,11-24} N_i \nu_{19,i}$		$k_{dec,Xh2} X_{h2}$
		Monosaccharides (kg COD.m^{-3})	Amino Acids (kg COD.m^{-3})	Long Chain Fatty Acids (kg COD.m^{-3})	Total Valerate (kg COD.m^{-3})	Total Butyrate (kg COD.m^{-3})	Total Propionate (kg COD.m^{-3})	Total Acetate (kg COD.m^{-3})	Hydrogen Gas (kg COD.m^{-3})	Methane Gas (kg COD.m^{-3})	Inorganic Carbon (kmole C.m^{-3})	Inorganic Nitrogen (kmole N.m^{-3})	Soluble Inerts (kg COD.m^{-3})	Inhibition factors: $I_1 = I_{pH} I_{IN,lim}$ $I_2 = I_{pH} I_{IN,lim} I_{h2}$ $I_3 = I_{pH} I_{IN,lim} I_{NH3,Xac}$

Table 3: Corrected biochemical rate coefficients ($v_{i,j}$) and kinetic rate equations (ρ_j) for particulate components (Batstone *et al.*, 2002; Rosen & Jeppsson, 2006)

Component →	i	1	2	3	4	5	6	7	8	9	10	11	12	Kinetic Rate
j	Process ↓	X_c	X_{ch}	X_{pr}	X_{li}	X_{su}	X_{aa}	X_{fa}	X_{c4}	X_{pro}	X_{ac}	X_{h2}	X_I	(ρ_j, kg COD.m^{-3},d^{-1})
1	Disintegration	-1	$f_{ch,xc}$	$f_{pr,xc}$	$f_{li,xc}$								$f_{xI,xc}$	$k_{dis}X_c$
2	Hydrolysis of Carbohydrates		-1											$k_{hyd,ch}X_{ch}$
3	Hydrolysis of Proteins			-1										$k_{hyd,pr}X_{pr}$
4	Hydrolysis of Lipids				-1									$k_{hyd,li}X_{li}$
5	Uptake of Sugars					Y_{su}								$k_{m,su}\frac{S_{su}}{K_s+S_{su}}X_{su}I_1$
6	Uptake of Amino Acids						Y_{aa}							$k_{m,aa}\frac{S_{aa}}{K_s+S_{aa}}X_{aa}I_1$
7	Uptake of LCFA							Y_{fa}						$k_{m,fa}\frac{S_{fa}}{K_s+S_{fa}}X_{fa}I_2$
8	Uptake of Valerate								Y_{c4}					$k_{m,c4}\frac{S_{va}}{K_s+S_{va}}X_{c4}\frac{1}{1+S_{bu}/S_{va}}I_2$
9	Uptake of Butyrate								Y_{c4}					$k_{m,c4}\frac{S_{bu}}{K_s+S_{bu}}X_{c4}\frac{1}{1+S_{va}/S_{bu}}I_2$
10	Uptake of Propionate									Y_{pro}				$k_{m,pro}\frac{S_{pro}}{K_s+S_{pro}}X_{pro}I_2$
11	Uptake of Acetate										Y_{ac}			$k_{m,ac}\frac{S_{ac}}{K_s+S_{ac}}X_{ac}I_3$
12	Uptake of Hydrogen											Y_{h2}		$k_{m,h2}\frac{S_{h2}}{K_s+S_{h2}}X_{h2}I_1$
13	Decay of X_{su}	1				-1								$k_{dec,Xsu}X_{su}$
14	Decay of X_{aa}	1					-1							$k_{dec,Xaa}X_{aa}$
15	Decay of X_{fa}	1						-1						$k_{dec,Xfa}X_{fa}$
16	Decay of X_{C4}	1							-1					$k_{dec,Xc4}X_{c4}$
17	Decay of X_{pro}	1								-1				$k_{dec,Xpro}X_{pro}$
18	Decay of X_{ac}	1									-1			$k_{dec,Xac}X_{ac}$
19	Decay of X_{h2}	1										-1		$k_{dec,Xh2}X_{h2}$
		Composites (kg COD.m^{-3})	Carbohydrates (kg COD.m^{-3})	Proteins (kg COD.m^{-3})	Lipids (kg COD.m^{-3})	Sugar Degraders (kg COD.m^{-3})	Amino Acid Degraders (kg COD.m^{-3})	LCFA Degraders (kg COD.m^{-3})	Valerate and Butyrate Degraders (kg COD.m^{-3})	Propionate Degraders (kg COD.m^{-3})	Acetate Degraders (kmole C.m^{-3})	Hydrogen Degraders (kmole N.m^{-3})	Particulate Inerts (kg COD.m^{-3})	Inhibition factors: $I_1 = I_{pH}I_{IN,lim}$ $I_2 = I_{pH}I_{IN,lim}I_{h2}$ $I_3 = I_{pH}I_{IN,lim}I_{NH3,Xac}$

2.3.4. Model Limitations

Although ADM1 allows one to take into account a broad range of substrate composition and numerous reactions, the need to define a large number of influent state variables (Table 17) and model parameters (Table 18 & Table 19) is widely criticised as the model's main drawback (Bernard *et al.*, 2001; Donoso-Bravo *et al.*, 2011).

Experimental data obtainable in reality lack the level of description required by the ADM1 structure. It is acknowledged to be technically difficult to characterise the influent substrate experimentally in the term of input state variables, as well as costly to quantify different species of biomass within a consortium. Several authors have pointed out the impracticality to perform such a detailed analysis on a frequent basis (Donoso-Bravo *et al.*, 2011; Arnell *et al.*, 2016).

The ADM1 model, in its default version, does not consider the following reactions (Batstone *et al.*, 2002):

- Alternative products from acidogenesis of sugars
- Sulphate reduction and sulphide inhibition
- Nitrate reduction
- Weak acid and base inhibition
- LCFA inhibition
- Acetate oxidation and homoacetogenesis
- Solids precipitation

They are intentionally omitted to keep the model as simple as possible since these reactions are not relevant universally for all cases. The model instead relies on the user's discretion to implement model extensions where a certain biochemical reaction is considered to have significant influence on the model outputs. For instance, the process of sulphate reduction can be added by modifying the Petersen matrix as described by Batstone (2006); and LCFA inhibition can be accounted by incorporating the inhibition term proposed by Arnell *et al.* (2016) into the effective kinetic rate.

It is noted that other extensions in addition to the above-listed reactions exist. For example, Boubaker and Ridha (2008) proposed an extra inhibition factor to describe the effect of high total VFA concentration on methanogenesis and inorganic nitrogen profile more precisely. By including the factor into the acetate uptake equation, ADM1 would be able to detect reactor failure as a result of transient inhibition.

Furthermore, Bergland, Dinamarca & Bakke (2015) proved that temperature influences steady-state biogas production rates more remarkably than what the default ADM1 model predicts. The authors recommended temperature correction to be extended to kinetic parameters as well, such as the disintegration constants (k_{dis}), hydrolysis constants (k_{hyd}), Monod uptake constants (k_m) and mass transfer coefficient (k_La).

2.4. ADM1 Parameters Literature Survey

An ADM1 model requires an initial set of parameters before the simulation can commence. In the Scientific and Technical Report (STR) which the ADM1 framework is published, a table of suggested values for each parameter is provided. These values are reported to be applicable for generic cases and suitable for model initialisation. Alternatively, specific values may be referenced from previous studies of similar applications. The suggested values for mesophilic digesters are provided as two types, namely “high-rate” and “solids”. The former refers to digesters that receive a heterogeneous mixture of liquid and solid substrates, whilst the latter refers specifically to digesters treating homogenous solid substrates.

A literature survey across 42 articles (in addition to the 20 references reported in the STR) was conducted. The purpose of the survey is to understand the variability of each parameter. Refer to the Appendix, Section 8.3, for all the values obtained from the survey. There is a wide variation in the parameter values even for identical substrate types. As explained by Pavlostathis & Giraldo-Gomez (1991) and Donoso-Bravo *et al.*, (2011), parameter variations could be attributed to the difference in experiment conditions employed during each research study. The inoculum composition, operation mode (batch or continuous) and environmental conditions (such as pH, temperature, etc.) all influence the experimental data and hence the final calibrated parameters.

Although the survey covered a wide range of substrate types, there are arguably too few references to identify the parameter variability specific to each substrate category. Nonetheless, understanding the variability of parameter values across all categories is still valuable, as it potentially represents the “universal” uncertainty range of each parameter. Table 4 presents this “universal” range of values alongside the STR suggested values. It is noted that temperature ranges from 25 – 38°C in the survey, whereas the STR suggested values are based on a digester temperature of 35°C.

Table 4: Summary statistics of parameter values surveyed from literature for mesophilic digesters in comparison to the default values suggested in the ADM1 Scientific and Technical Report

Parameter	Suggested Values in ADM1 STR		Parameter Survey – All Categories				
	High-rate	Solids	Min	Max	Avg	LQ	UQ
$f_{SI,XC}$	0.1	0.1	0.013	0.422	0.185	0.087	0.321
$f_{XI,XC}$	0.25	0.25	0.02	0.55	0.25	0.13	0.38
$f_{CH,XC}$	0.2	0.2	0.0718	0.797	0.405	0.262	0.515
$f_{PR,XC}$	0.2	0.2	0.01	0.4	0.17	0.07	0.25
$f_{LI,XC}$	0.25	0.25	0.014	0.478	0.118	0.034	0.161
$f_{FA,LI}$	0.95	0.95	-	-	-	N/A	N/A
$f_{H2,SU}$	0.19	0.19	-	-	-	N/A	N/A
$f_{BU,SU}$	0.13	0.13	0.111	0.13	0.121	N/A	N/A
$f_{PRO,SU}$	0.27	0.27	0.27	0.54	0.405	N/A	N/A
$f_{AC,SU}$	0.41	0.41	0.202	0.41	0.306	N/A	N/A
$f_{H2,AA}$	0.06	0.06	-	-	-	N/A	N/A
$f_{VA,AA}$	0.23	0.23	0.23	0.309	0.284	0.273	0.299

Parameter	Suggested Values in ADM1 STR		Parameter Survey – All Categories				
	High-rate	Solids	Min	Max	Avg	LQ	UQ
$f_{BU,AA}$	0.26	0.26	0.186	0.29	0.253	0.236	0.271
$f_{PRO,AA}$	0.05	0.05	0.041	0.12	0.065	0.045	0.072
$f_{AC,AA}$	0.4	0.4	0.273	0.399	0.316	0.287	0.328
k_{dis}	0.4	0.5	0.001	1.743	0.450	0.115	0.782
k_{hyd_ch}	0.25	10	0.037	2.75	0.679	0.210	0.941
k_{hyd_pr}	0.2	10	0.0014	18.23	0.898	0.110	0.650
k_{hyd_li}	0.1	10	0.0086	2.1	0.45	0.10	0.73
K_{s_IN}	0.0001	0.0001	-	-	-	N/A	N/A
pH_{UL_acid}	5.5	5.5	5.5	8.5	6.5	5.5	8.0
pH_{LL_acid}	4	4	4	6	5	4	6
pH_{UL_acet}	5.5	5.5	5.5	6.7	6.0	5.8	6.4
pH_{LL_acet}	4	4	4	5.8	4.4	4.0	4.9
k_{m_su}	30	30	11.9	125	54.0	27.0	98.0
K_{s_su}	0.5	0.5	0.022	4.5	0.85	0.049	1.28
Y_{su}	0.1	0.1	0.01	0.17	0.09	0.01	0.15
k_{dec_xsu}	0.02	0.02	0.01	0.8	0.28	0.01	0.80
k_{m_aa}	50	50	19.8	53	35.2	27.3	41.8
K_{s_aa}	0.3	0.3	0.01	1.198	0.623	0.050	1.15
Y_{aa}	0.08	0.08	0.058	0.15	0.095	0.065	0.134
k_{dec_xaa}	0.02	0.02	0.02	0.8	0.19	0.02	0.45
k_{m_fa}	6	6	0.93	12	5.5	1.7	9.1
K_{s_fa}	0.4	0.4	0.024	9.21	1.79	0.10	2.22
Y_{fa}	0.06	0.06	0.004	0.055	0.043	0.027	0.055
k_{dec_xfa}	0.02	0.02	0.01	0.06	0.017	0.01	0.018
KI_{h2_fa}	5×10^{-6}	5×10^{-6}	3×10^{-6}	5×10^{-6}	4×10^{-6}	N/A	N/A
k_{m_c4}	20	20	5	60	18	7	22
K_{s_c4}	0.3	0.2	0.012	0.6	0.22	0.049	0.33
Y_{c4}	0.06	0.06	0.0193	0.066	0.055	0.043	0.066
k_{dec_xc4}	0.02	0.02	0.02	0.03	0.027	0.027	0.03
KI_{h2_c4}	1×10^{-5}	1×10^{-5}	1×10^{-8}	1×10^{-5}	5×10^{-6}	5×10^{-8}	8×10^{-6}
k_{m_pro}	13	13	0.16	100	15	5.5	16
K_{s_pro}	0.3	0.1	0.02	1.146	0.259	0.058	0.392
Y_{pro}	0.04	0.04	0.019	0.075	0.048	0.031	0.055
k_{dec_xpro}	0.02	0.02	0.001	0.06	0.022	0.008	0.046
KI_{h2_pro}	3.5×10^{-6}	3.5×10^{-6}	2.4×10^{-8}	8.0×10^{-6}	3.1×10^{-6}	4.7×10^{-8}	7.0×10^{-6}
k_{m_ac}	8	8	3.1	48	13	6.7	14
K_{s_ac}	0.15	0.15	0.011	0.93	0.26	0.045	0.49

Parameter	Suggested Values in ADM1 STR		Parameter Survey – All Categories				
	High-rate	Solids	Min	Max	Avg	LQ	UQ
Y_{ac}	0.05	0.05	0.014	0.1	0.048	0.027	0.073
k_{dec_xac}	0.02	0.02	0.001	0.05	0.022	0.008	0.036
KI_{nh3_ac}	0.0018	0.0018	0.00026	0.0223	0.0060	0.0009	0.0119
pH_{UL_ac}	7	7	6.7	7	6.9	6.7	7
pH_{LL_ac}	6	6	5.2	6	5.8	5.4	6
k_{m_h2}	35	35	1.68	209	50.6	25.3	59.0
K_{s_h2}	2.5×10^{-6}	7×10^{-5}	1×10^{-6}	0.0006	0.0001	8.3×10^{-6}	7.2×10^{-5}
Y_{h2}	0.06	0.06	0.0089	0.183	0.057	0.021	0.065
k_{dec_xh2}	0.02	0.02	0.001	0.3	0.080	0.003	0.227
pH_{UL_h2}	6	6	6	6.7	6.4	N/A	N/A
pH_{LL_h2}	5	5	5	5.8	5.4	N/A	N/A

LQ = Lower quartile; UQ = Upper quartile; N/A = Not applicable for parameters with less than 3 references

The STR suggested the model outputs' sensitivity relevant to each kinetic parameter. The level of sensitivity is categorised into 3 levels, namely (1) low or no sensitivity, (2) medium sensitivity under dynamic conditions and (3) high sensitivity under both steady-state and dynamic conditions. No suggestion was given for the stoichiometric parameters.

From the survey, it was observed that 31 articles have applied ADM1 modelling, wherein 305 counts of parameter calibrations were made. Researchers calibrate parameters to optimise the fit accuracy of their models. By grouping these counts according to their level of sensitivity suggested in the STR (Table 5), it is evident that parameters classified with high sensitivity indeed have a higher tendency (or bias/preference) to be calibrated. However, parameters with lower suggested sensitivities do not necessarily mean fewer adjustment frequency (cf. k_{m_su} versus k_{dec_xsu}). This observation supports the fact that parameters have an inter-correlated effect on the model outputs, and that STR-suggested sensitivities might not be a reliable indication of the sequence or priorities with regards to calibration.

Stoichiometric parameters are usually not calibrated because their conversions are theoretical. In contrast, composite particulate conversions ($f_{SI,XC}$, $f_{XI,XC}$, $f_{CH,XC}$, $f_{PR,XC}$, $f_{LI,XC}$) are modified frequently. Typically these conversions are determined through experimental observations (Lübken *et al.*, 2007; Mairet *et al.*, 2011). It is also pointed out in the STR that these conversions are highly variable and dependent on the substrate type and process. Adoption of the "composite particulate" (X_c) concept in ADM1 means that, depending on how these conversion parameters are defined, distribution of COD (as a result of disintegration process) to carbohydrates, proteins and lipids state variables has an indirect influence on the output sensitivities.

Table 5: Number of ADM1 parameter modifications observed during literature survey

Calibration Counts	Parameter Sensitivity as Suggested by STR			
	No Suggestion	**Low or No Sensitivity**	**Medium Sensitivity**	**High Sensitivity**
0 – 4	$f_{FA,LI}$; $f_{H2,SU}$; $f_{BU,SU}$; $f_{PRO,SU}$; $f_{AC,SU}$; $f_{H2,AA}$; $f_{VA,AA}$; $f_{BU,AA}$; $f_{PRO,AA}$; $f_{AC,AA}$	K_{s_IN} ; pH_{UL_acid} ; pH_{LL_acid} ; pH_{UL_acet} ; pH_{LL_acet} ; Y_{su} ; K_{s_aa} ; Y_{aa} ; k_{m_fa} ; K_{s_fa} ; Y_{fa} ; KI_{h2_fa} ; Y_{c4} ; Y_{pro} ; Y_{ac} ; Y_{h2} ; pH_{LL_h2}	k_{dec_xsu} ; k_{dec_xaa} ; k_{dec_xfa} ; k_{dec_xc4} ; k_{dec_xpro} ; k_{dec_xac} ; pH_{LL_ac} ; k_{dec_xh2} ; pH_{UL_h2}	pH_{UL_ac}
5 – 9	$f_{SI,XC}$	k_{m_su} ; K_{s_su} ; k_{m_aa} ; k_{m_c4} ; K_{s_c4} ; KI_{h2_c4} ; k_{m_h2}	KI_{h2_pro} ; KI_{nh3_ac} ; K_{s_h2}	
10 – 14			k_{hyd_li} ; K_{s_pro}	K_{s_ac}
15+	$f_{XI,XC}$; $f_{CH,XC}$; $f_{PR,XC}$; $f_{LI,XC}$		k_{m_pro}	k_{dis} ; k_{hyd_ch} ; k_{hyd_pr} ; k_{m_ac}

2.5. Current Practices of ADM1 Parameter Estimation

Parameter estimation is known to be a highly subjective procedure, as it relies on the modeller's discretion and experience (Koch *et al.*, 2010). The simplest approach would be to reference calibrated parameters from previous studies where similar substrate type is experimented; however such data is often neither available nor exactly applicable to the scenario at hand (Astals *et al.*, 2014).

Batch tests have been used to study model parameters. For instance, a specific bacteria type could be isolated in a batch system together with a particular soluble substrate. Given the substrate-limited condition, it would allow one to determine the specific substrate uptake rate which is a parameter required by ADM1. Some modellers prefer estimating kinetic parameters through the use of a batch test called BMP (Biomethane Potential) test first and then applying the parameters for continuous modelling (Antonopoulou *et al.*, 2012). BMP test involves introducing a calculated quantity of substrate into a known quantity of sludge based on a certain substrate to biomass ratio. Biogas production rate is continuously monitored whilst samples are extracted at fixed time intervals to monitor the evolution of specific constituents. Thereafter, kinetic parameters (e.g. hydrolysis rate, hydrogen and VFA uptake rates) can be approximated by fitting ADM1 against the batch experimental data using non-linear estimation methods.

Despite its popularity, Baltes *et al.* (1994) cautioned against estimating kinetic parameters of Monod growth kinetic model (which ADM1 uses) from batch systems, as the estimated parameters would fail under dynamic feed situations. Instead, the author suggested that the estimation should be applied to experiments with continuous, time-varying feed rates. Batstone, Tait and Starrenburg (2009) further pointed out that as a batch system the BMP testing is unable to conduct under the same conditions (except temperature) as the full-scale digester. Therefore, batch testing is deemed not entirely representative. Another drawback is that the test requires specialised experimental setup which is costly and time-consuming.

A major limitation of batch systems is the absence of inputs variance. The only model input is the initial conditions, which as a consequence, induces inadequate outputs sensitivity (Lokshina & Vavilin, 1999). Donoso-Bravo *et al.* (2011) concurred that experimental data from continuous systems are appropriate for kinetic parameter estimation, provided that the experiment ran at different dilution rates or where inputs are dynamic. Many publications have demonstrated estimation of ADM1 parameters using data obtained from continuous or semi-continuous systems on lab-scale (Blumensaat & Keller, 2005; Boubaker & Ridha, 2008); however only a few actually featured modelling of full-scale industrial plants (Batstone *et al.*, 2009; Girault & Steyer, 2010; López & Borzacconi, 2009).

In a study by Batstone, Tait & Starrenburg (2009), a full-scale plant operating at variable flow and organic loading was modelled. The authors compared the modelling performance when using biodegradability extent and hydrolysis parameter (k_{hyd}) estimated separately from BMP test data and 1.5 years full-scale plant data. Parameters estimated from the BMP test were found to result in poorer modelling performance, supposedly due to the estimated hydrolysis rates being too low. Hydrolysis values estimated from continuous data were an order of magnitude higher.

The typical procedure taken for estimating ADM1 parameters is summarised by Donoso-Bravo *et al.* (2011) in Figure 3. When minimising the objective function, it is common practice to only calibrate parameters that exhibit

high sensitivity towards the model outputs instead of the entire group of ADM1 parameters. By reducing the number of degrees of freedom and bias, it allows one to execute model fitting quicker and avoid parameter equifinality situations where more than one set of parameters have the same model fitting accuracy. For that reason, researchers often apply some sort of sensitivity analysis techniques to rank model parameters according to their sensitivities on model outputs.

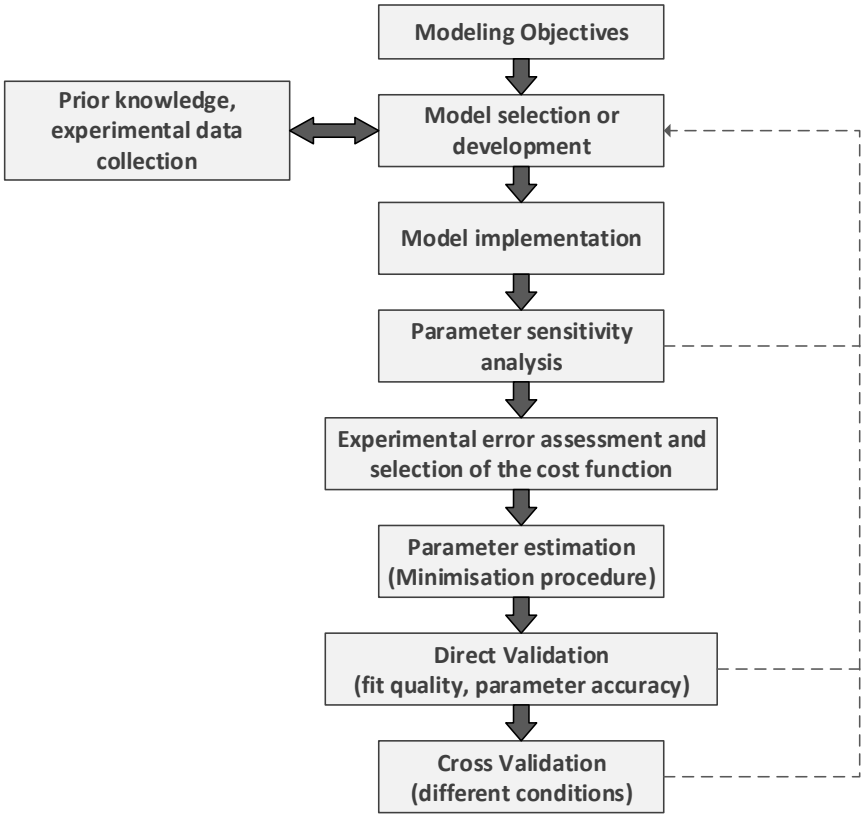


Figure 3: Parameter estimation procedure typically followed in anaerobic digestion modelling (Donoso-Bravo *et al.*, 2011)

A literature survey into the methods used in published ADM1 research was conducted (see Appendix, Table 20). The survey suggests that there is currently no protocol on parameter calibration, and in many cases, the method applied is not described explicitly. It is common practice that s*toichiometric* parameters are left uncalibrated and their default values (as suggested in the STR) retained unless sensitivity analysis suggests otherwise. For *kinetic* parameters, different approaches were noted:

- Only adjust kinetic parameters that are *suggested in the STR* as highly sensitive while parameters deemed less sensitive are left unchanged. e.g. Blumensaat & Keller (2005)
- Select kinetic parameters based on outcomes of a *sensitivity analysis* and only calibrate a few of the most sensitive ones. e.g. Jeong *et al.* (2005), Koch *et al.* (2010), Razaviarani & Buchanan (2015)
- Calibrate a selected group of parameters based on *expert knowledge* about the substrate or its degradation behaviour. e.g. Mairet *et al.* (2011)
- Calibrate parameters in groups in conjunction with digester design-related constants e.g. Bernard *et al.*, (2001) decoupled the calibration process by first classifying parameters into three subsets then calibrate them sequentially: (1) kinetic parameters, (2) transfer coefficient k_La and (3) the yield coefficients.

- Group parameters according to *different level of sensitivity*, then calibrate parameters in the high sensitivity subset first, followed by the next sensitivity subset if the objective function objective is not met e.g. Coelho et al. (2006) grouped the kinetic parameters into the 3 groups of sensitivities (High, Medium, Low) as suggested in the STR, and calibrate each group sequentially.

In an alternative approach, Girault & Steyer (2010) aimed at balancing total nitrogen, ammonium and COD whilst keeping the model parameters as default. To achieve nitrogen and ammonium balance, the nitrogen contents of composite particulate (X_c), protein (X_{Pr}) and soluble and particulate inerts (S_I, X_I) were calibrated. COD balance is achieved by adjusting the proportion of S_I and X_I relative to the total COD. These constants are normally fixed values by default.

In conclusion, it can be said that there is currently no consensus or common framework for sensitivity analysis and which parameters subset to be calibrated.

2.6. Sensitivity Analysis Techniques

The purpose of performing sensitivity analysis is to understand the change in model outputs as a result of a predefined change in model inputs (Sin *et al.*, 2011). Most commonly varied inputs in ADM1 are the kinetic and stoichiometric parameters but could also include influent conditions and system conditions.

2.6.1. Local Sensitivity Analysis

Local sensitivity analysis involves evaluating the differential changes in the model outputs with respect to the change in an input parameter value at a specific time (Equation 29). To determine the sensitivity of a complete parameter set, parameters are varied one at a time and the relative change in outputs are totalised as the weighted sum or as errors (Donoso-Bravo *et al.*, 2011).

While this technique is relevant for research in process control and problem identification, it is limited to describe the linear relationship between the input and output. It does not identify correlations between the inputs and does not consider the combined effect of two or more inputs on the outputs (Sin *et al.*, 2011).

$$\frac{dy_i}{d\theta_i} = \frac{y_j(\theta_i + \Delta\theta_i)}{\Delta\theta_i} \qquad \textit{Equation 29}$$

where:

y_i is the j^{th} output;

θ_i is the i^{th} input

2.6.2. Global Sensitivity Analysis

Global sensitivity analysis is often referred to in the literature as an analysis of variance (ANOVA) case because it is commonly performed in conjunction with uncertainty analysis (Saltelli *et al.*, 2008). A common use of these techniques is to rank each input according to their individual impact on the outputs' variance, by which the most influential parameter results in the most uncertainty in the outputs, and vice versa. Furthermore, it quantifies how much variance each input parameter is contributing to the outputs; therefore provides the plant designer valuable insights as to which uncertainty should be reduced in order to achieve a robust process (Sin *et al.*, 2011).

According to a review by Donoso-Bravo *et al.* (2011), various global sensitivity analysis techniques have been applied in WWTP modelling research. These include Morris Screening (Morris, 1991), Standard Regression Coefficient (SRC) method (Helton & Davis, 2003) and variance decomposition (Saltelli *et al.*, 2008).

The SRC method involves applying first-order linear regressions to each model output generated from the Monte Carlo (MC) simulation. Thereafter, the regression coefficients are normalised using the standard deviations of the MC simulation data. These coefficients represent the sensitivity of each parameter. As noted by Sin *et al.* (2011), this method may not be suited for non-linear models because it requires a coefficient of determination (R^2) above 0.7 to be valid.

For non-linear models, variance decomposition techniques (e.g. FAST, Sobol Indices) could be applied instead. These techniques decompose variance from results of Monte-Carlo into indices/components that represent the interaction between input variables and outputs (Sobol & Kucherenko, 2005).

Potentially, *multivariate regression* methods could present an alternative approach to global sensitivity analysis, given their ability to recognise underlying patterns between the model inputs and outputs, and to describe them linearly after mathematical transformation (Madsen *et al.*, 2011). Regardless of which method, the computational time is typically long due to the iterative MC runs, and thus a concern often pointed out by researchers.

2.7. Uncertainty Analysis Using Monte Carlo Simulation

Uncertainty analysis involves sampling values from uncertain inputs and then applying these samples through a mathematical model to generate output uncertainty – a process commonly referred to as propagation of uncertainty. Input uncertainty is a range of values presumed, based on limited knowledge, to represent the true value. Output uncertainty typically varies more when additional sources of input uncertainty are included (Sin *et al.*, 2009).

The Monte Carlo (MC) method is regarded as a simplistic yet effective way for propagating uncertainty. Apart from the advantage that the original process model requires no modification, MC results are intuitively easy to interpret and compatible with a wide variety of sensitivity analysis techniques (Diwekar & David, 2015).

Selecting an appropriate number of simulation/iteration runs and the type of sampling procedure are important aspects of the Monte Carlo method. This ensures that the MC results have adequate resolution for identification of each output's distribution. Latin Hypercube Sampling (LHS) and Random Sampling are two sampling procedures commonly applied in Monte Carlo simulations (Diwekar & David, 2015). While Random Sampling generates random values within the respective subset based on uniform probability, LHS incorporates probability as part of sampling by assigning weights to each sample element/parameter subset. Due to the stratified approach, LHS is generally considered to be more efficient and better suited for computationally expensive models. However, if stoichastic uncertainty exists, Random Sampling is the preferred option (Sin *et al.*, 2009).

Input uncertainty *framing* is a critical step of uncertainty analysis (Sin *et al.*, 2009). Framing is defined as the reasoning behind the input uncertainties i.e. range and statistical distribution of values selected. It influences the result produced by the subsequent sensitivity analysis and the interpretation thereof.

2.8. Multivariate Regression Methods

Modelling of dependent/output variable(s) from a large number of independent/input variables within a multivariate data set is known as “multivariate regression”. Linear regression methods are considered inadequate because multicollinearity often exists within the data set. Multicollinearity is a term that describes a scenario where some independent variables are highly correlated or collinear.

Kleinbaum *et al.*, (1998) noted that variance in the regression parameters increases as the degree of correlation increases; a phenomenon that leads to poor accuracy when predicting unseen data. For this reason, dimension reduction is the foremost step in multivariate regression methods. The objective of dimension reduction is to reduce the number of independent variables so that the influence of multicollinearity is lessened (Bair *et al.*, 2006).

Understanding the influence of a particular independent variable with respect to a specific dependent variable within a large multivariate data set is difficult but it is possible to identify the *underlying patterns/relationships* within the data set that describes their relationships (Madsen *et al.*, 2011). Two popular multivariate regression methods namely, PCR (Principal Component Regression) and PLSR (Partial Least Squares Regression), are known for their dimension reduction and pattern recognition functions.

PCR (Principal Component Regression)

The concept behind PCR lies in the statistical method called Principal Component Analysis (PCA). PCA brings about dimensionality reduction through a linear projection of high dimensionality data into a lower dimensional space while retaining the maximum variance in the data. A set of Principal Components constitutes the lower dimensional space (Maitra & Yan, 2008). The first principal component provides a direction in the data where the highest variation lies, meanwhile the subsequent principal components indicate the direction perpendicular (i.e. uncorrelated) to the previous component’s direction along which the next highest covariance exists. PCR is simply a regression of response variables using principal components as predictors.

The PCA algorithm is summarised as follows (Geladi & Kowalski, 1986):

- **Step 1:** Create a covariance matrix from the input variable matrix
- **Step 2**: Obtain the eigenvectors, together with a corresponding set of eigenvalues, by applying Singular Value Decomposition technique on the variance matrix. These eigenvectors are the so-called principal components. Collinearity is completely avoided due to orthogonality.
- **Step 3**: Rank the components in descending order of corresponding eigenvalues’ magnitude
- **Step 4**: The user may select the number of components to use for regression. In general, eigenvectors corresponding to small eigenvalues are omitted, since most of the information is held by the higher eigenvalues.

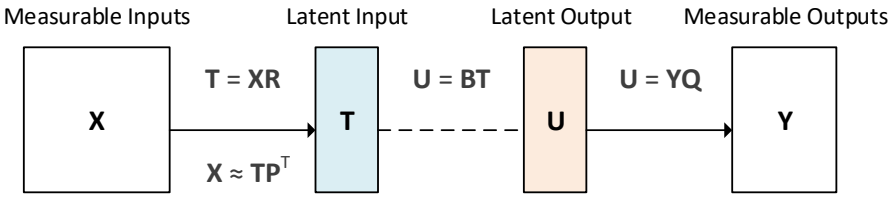
A notable drawback of PCA is the fact that the technique focuses solely on describing the variation in the predictive variables and does not factor any relations with the dependent variables into the directional searching. This implies that the directions of the components might not be the best for predicting the model outputs. Furthermore, some components with large weightings (i.e. eigenvalues) may retain information irrelevant for prediction. PCR is thus referred to as an *unsupervised* regression method (Maitra & Yan, 2008; Sarkar & Sobie, 2010).

PLSR (Partial Least Squares Regression)

Unlike PCA, the PLSR algorithm takes into account the correlation between the independent and dependent variables as part of its dimension reduction procedure. The new components created in PLSR, termed Latent Variables, aims to capture as much information in the original independent variables as well as the importance of each variable in relation to the dependent variables. As such, PLSR is considered a *supervised* regression method (Maitra & Yan, 2008).

An overview of the concept is illustrated in Figure 4 and the Nonlinear Iterative PLS (NIPALS) algorithm is outlined in Table 6. Consider a data set that consists of an input matrix X and an output matrix Y, whereby each matrix is composed of n number of samples for every X and Y variable. PLSR follows an iterative algorithm that identifies weight vectors (one for each data matrix), which upon linear combinations with the two data matrices, will give rise to two sets of latent variables (T and U) with maximum covariance.

The inner relation between the two latent variables is described by the regression coefficient vector **B**. At this point, the latent input and latent output variables correlate the greatest variation in the inputs to the outputs. According to Geladi & Kowalski (1986), a high number of latent variables selected for regression should be avoided because higher degree components may only describe data noise and induce collinearity problems. Moreover, there could be a risk of overfitting (Madsen *et al.*, 2011).



X: Input matrix; **Y**: Output matrix; **T**: Latent Input; **U**: Latent Output; **R**: X-weights; **P**: X-loading; **Q**: Y-loading; **F**: Y-residual

Figure 4: Illustration of the PLSR concept showing dimensions of the data matrices, weight and loading vectors

Table 6: Summary of the NIPALS algorithm for PLSR (Geladi & Kowalski, 1986; de Jong, 1993)

	Description	**Algorithm**
I. Data Preparation		
Step 1	Mean-centering	Calculate the average value of each variable within the data set, then subtract the averages from the corresponding variables.
Step 2	Variance scaling	Calculate the standard deviation of each variable within the data set, then divide each variable by the corresponding deviations. The resultant X and Y blocks are now termed $\mathbf{X}_0$ and $\mathbf{Y}_0$ respectively.
II. Determine 1st set of latent variables, weights and loading vectors		
Step 3	Initial guess for $\mathbf{u}_1$	Select a random column of $\mathbf{Y}_0$
Step 4	Find **t**: latent variable of **X**	$\mathbf{w}_1 = \mathbf{X}'_0\mathbf{u}_1$ $\mathbf{t}_1 = \mathbf{X}_0\mathbf{w}_1$
Step 5	Find **u**: latent variable of **Y**	$\mathbf{q}_1 = \mathbf{Y}'_0\mathbf{t}_1$ $\mathbf{u}_1 = \mathbf{Y}_0\mathbf{q}_1$
Step 6	Converging **t** and **u** values	Repeat steps 4 and 5 until **u** converges. These are the final latent variables, also known as scores.
Step 7	Find **b**: regression coefficient relating t and u	$\mathbf{u}_1 \approx \mathbf{b}_1\mathbf{t}_1$
Step 8	Find **p**: loading vector of **X**	Since $\mathbf{X}_0 \approx \mathbf{t}_1\mathbf{p}'_1$, the loading vector can be calculated as: $\mathbf{p}_1 = \mathbf{X}'_0\mathbf{t}_1 / (\mathbf{t}'_1\mathbf{t}_1)$
III. Determine 2nd set of latent variables, weights and loading vectors (Optional – Only if two or more latent variables are desired)		
Step 9	Deflate X and Y data matrices This step removes the variation attributed to the first input and output latent variables from their respective matrices.	Deflated data matrices are simply residuals from the previous regression: $\mathbf{X}_1 = \mathbf{X}_0 - \mathbf{t}_1\mathbf{p}'_1$ $\mathbf{Y}_1 = \mathbf{Y}_0 - \mathbf{b}_1\mathbf{t}_1\mathbf{q}'_1$
Step 10	Find new latent variables, weights and loading vectors	Repeat steps 4 – 7 using the deflated data matrices Indices of all newly determined vectors shall increase by 1
IV. Concluding steps		

	Description	Algorithm
Step 11	Extract all possible latent factors	Repeat step 10 until the number of PLS components meet the user's objective
Step 12	Collate all vectors calculated during each iteration	Compile the vectors into matrices: $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3, \ldots]$ $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots]$ $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \ldots]$ $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \ldots]$ $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \ldots]$
Step 13	Find **R**: alternative weighting matrix for **X** data matrix This step replaces **W** (which relates to the depleted **X** matrices) with **R** (which relates to the original **X** matrix), as it provides more relevant insights to the relationship between each X variables and the latent variable **T**.	Since $\mathbf{T} = \mathbf{XR}$, the alternate weighting matrix can be calculated as: $\mathbf{R} = (\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'\mathbf{X}$

Note: The subscripted indices beneath each vector indicates the latent variable number.

2.9. Model Objective Function & Validation

Objective function, also referred to as the cost function, is a criterion chosen by the modeller that measures the model's goodness of fit. Model optimisation aims to optimise this function by calibrating model parameters using numerical algorithms. Selection of an objective function is a critical factor as it influences how parameters are calibrated and hence the outcomes of model fitting (Batstone, Pind & Angelidaki, 2003).

According to Donoso-Bravo *et al.* (2011) the most popular objective function utilised for AD modelling is the sum of least squares. The concept lies in minimising the squared distance between measured and predicted data points (Equation 30). A slight variant is the Root Mean Square Error (RMSE), Equation 31, which is a commonly applied function in chemometrics to validate model predictions (Esbensen & Julius, 2010). Note that the output variables, y_{exp} and y_{sim}, are first scaled before RMSE is calculated.

Biogas flow is sometimes applied as the sole criterion of objective function evaluation (Donoso-Bravo *et al.*, 2011). However, based on the literature survey (Table 20), research works involving ADM1 tend to include other experiment measurements as well, such as gas constituents (commonly methane, carbon dioxide & hydrogen), VFA (total or individual species), VSS, NH_4, pH and alkalinity.

$$J(\theta) = min \sum_{t=1}^{N} \left(y_{exp}(t) - y_{sim}(t, \theta) \right)^2 \qquad \textit{Equation 30}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} \left(y_{exp}(t) - y_{sim}(t) \right)^2}{n}} \qquad \textit{Equation 31}$$

Evaluating a calibrated model against an unseen set of data is an important step of model development (Esbensen & Julius, 2010). A common model validation practice involves dividing experimental data into two sets intentionally: the first set is utilised for parameter calibration, while the second set is subjected to validation. Parameters calibrated from the first set are applied in the modelling of the second set and then the goodness-of-fit is evaluated against the first set's.

There are no definite rules on how the data set should be divided. For instance, Bernard *et al.* (2001) designated the steady-state period as the first set and the transient period as the second set. The author concluded that calibration using steady-state data is capable of producing valid modelling of transient behaviours. On the other hand, Barrera *et al.* (2015) partitioned the data into two subsets, both of which have different operating conditions (varying COD and sulphate loadings). Thamsiriroj & Murphy (2011) validated the first subset with another subset of which the reactor operated at distinctly different hydraulic retention time.

CHAPTER 3

RESEARCH METHODOLOGY

3.1. ADM1 Model Setup

3.1.1. Background of Full-scale Plant

The case study considered in this research work is a newly commissioned industrial-scale anaerobic digestion plant that is designed to treat wastewater generated from a dairy factory. Various types of milk products including cheese is produced at this factory. The treatment plant was constructed with the aim to reduce hydraulic and COD load to the local municipal sewage treatment plant; by doing so, valuable resources such as energy and water are recovered. Energy recovered in the form of biogas is used to produce steam whilst the treated wastewater is further upgraded for reuse as boiler feed water.

The design of this plant is categorised as an Anaerobic MBR (Membrane Bio-Reactor) process. Unlike conventional AD processes where biomass is separated by clarification or three-phase separators, this technology utilises externally pressurised ultrafiltration (UF) membranes to facilitate the solid-liquid separation. This unique process produces exceptionally clean effluent which could be fed to a reverse osmosis process without further treatment.

3.1.2. Plant Configuration

Wastewater produced by the factory consists of Cleaning-in-Place (CIP) water, flushing water and tanker rinse water. Also, whey is released intermittently into the wastewater stream to boost the COD concentration. The blended wastewater is sent over a self-cleaning filter with a filtration diameter of 1.5mm before entering the buffer tank. Figure 5 illustrates an overview of the Anaerobic MBR process.

Variations in flow, concentration and pH of the combined wastewater are equalised in the buffer tank. The tank is designed to operate at a minimum retention time of 8 hours to promote hydrolysis and pre-acidification. Nutrients such as ferric chloride and micro-nutrients are added into the buffer tank. Given the high pH CIP, no external alkalinity addition is required.

From the buffer tank, the pre-acidified wastewater is sent to the anaerobic digester (Figure 5 – Stream 1). The wastewater is pre-heated with treated wastewater in an heat exchanger, in order to minimise energy consumption. Organic matters such as VFAs in the wastewater are converted to biogas in the anaerobic digester. The produced biogas is fed to a steam boiler whilst any excess is flared (Figure 5 – Stream 2).

A portion of the anaerobic digester content is recycled and sprayed via spray nozzles onto the liquid surface to prevent foaming. The recirculation line passes through a heat exchanger to maintain the digester temperature as closely as possible to 35°C. Due to the nature of the wastewater (i.e. acidic whey and predominantly alkaline CIP), no external alkalinity is introduced into the digestion process. Hydrochloric acid dosing is available as a backup to ensure that pH is maintained below 7.5.

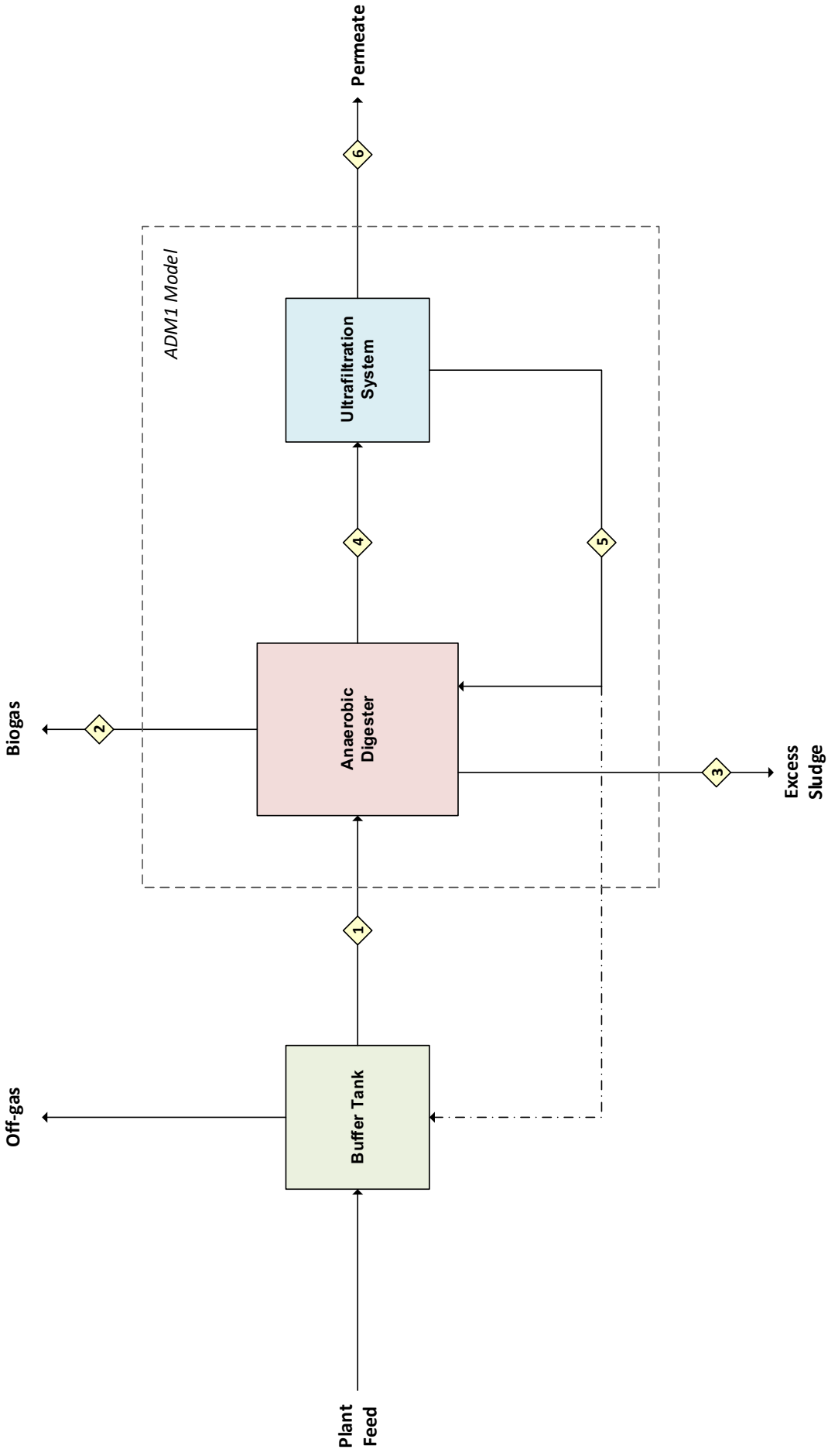


Figure 5: Schematic showing the main process flow of the full-scale plant. The enclosed section represents the anaerobic MBR system to which the ADM1 model is configured

In order to retain the anaerobic biomass in the digester and to produce a high-quality filtrate for the downstream reverse osmosis system, cross-flow type ultrafiltration (UF) is deployed. The contents of the anaerobic digester are continuously recirculated across these UF skids, where the pressure forces clean liquid through the membranes and leave particulate biomass behind in the return loop to the digester (Figure 5 – Streams 4 & 5). In order to maintain the flux rate at an acceptable level, the cake layer formation on the UF membrane surface is controlled by means of periodic backwash and chemical Cleaning-In-Place (CIP).

Waste Anaerobic Sludge (WAnS) is discharged from the anaerobic digester periodically to prevent excessive biomass concentration build-up which would jeopardise the membrane performance (Figure 5 – Stream 3). The excess WAnS is dewatered in a decanter centrifuge subsequently and the reject water is returned to the anaerobic digester.

3.1.3. Plant Data Analytical Methods

Data used in the research study are collected utilising either online measuring devices or offline (manual) sampling. The frequency of data recording, as summarised in Table 7, follows the operation philosophy set out by the plant designer. Note that some parameters are only analysed occasionally; therefore in order to construct a complete set of data that covers the entire period, analyses not measured daily were interpolated according to a "*4 days earlier and 3 days future"* basis. Refer to Appendix Section 8.4 for the raw data.

Table 7: List of measurements and sample frequency

Component	*Test Method*	Digester Feed (Stream 1)	Biogas (Stream 2)	WAnS (Stream 3)	Digester Contents	Permeate (Stream 6)
Flow [m³/day]	*Online flowmeter*	Daily total[1]	Daily total[2]	Daily total[3]	-	-
pH	*Online probe* [4]	Daily	-	-	Daily	-
Temperature [°C]	*Online probe* [4]	Daily	-	-	Daily	-
TCOD [mg/l]	*Hach TNTplus822* [5]	Daily	-	-	Daily	Daily
SCOD [mg/l]	*Hach TNTplus822* [5]	Daily	-	-	-	-
TSS [mg/l]	*APHA 2540* [6]	1 x / week	-	-	3 x / week	-
VSS [mg/l]	*APHA 2540* [6]	1 x / week	-	-	3 x / week	-
VFA [meq/l]	*Hach TNT872* [5]	1 x / week	-	-	-	3 x / week
Alkalinity [meq/l]	*APHA 2320* [6]	1 x / week	-	-	-	2 x / week
TN [mg N/l]	*Hach TNT823* [5]	1 x / week	-	-	-	1 x / week
NH_4-N [mg N/l]	*Hach LCK302* [5]	1 x / week	-	-	-	3 x / week
NO_2-N [mg N/l]	*Hach TNT839* [5]	3 x / week	-	-	-	3 x / week
NO_3-N [mg N/l]	*Hach TNT835* [5]	3 x / week	-	-	-	3 x / week
Total-P [mg P/l]	*Hach TNT844* [5]	1 x / week	-	-	-	1 x / week
PO_4-P [mg P/l]	*Hach TNT846* [5]	1 x / week	-	-	-	1 x / week
Ca [mg/l]	*Hach 0.8M EDTA* [5]	1 x / week	-	-	-	1 x / week
Mg [mg/l]	*Hach 0.8M EDTA* [5]	1 x / week	-	-	-	1 x / week
SO_4 [mg/l]	*Hach TNT864* [5]	2 x / week	-	-	-	2 x / week
CH_4 content [%]	*Online probe* [2]	-	Daily	-	-	-
CO_2 content [%]	*100% - CH_4%*	-	Daily	-	-	-

Notes

1 Endress & Hauser Proline Promag 50L Electromagnetic Flowmeter (Max error: ±0.5%)

2 Endress & Hauser Proline Prosonic Flow B 200 Ultrasonic Flowmeter (Max error: ±1.5%)

3 Endress & Hauser Proline Promag 50P Electromagnetic Flowmeter (Max error: ±0.5%)

4 Endress & Hauser Orbisint CPS11D Glass Electrode Sensor

5 Measured using Hach DR 6000TM spectrophotometer together with the Hach reagent as indicated

6 Standard Methods (APHA, 1992)

Table 8 presents a summary of the wastewater composition during the steady-state period. The average values exhibit similar characteristics (except for flow rate and ammonia) as expected by the plant designer. When comparing the composition against Cheese Whey Wastewater (CWW) compositions published by other

researchers, it is evident that the wastewater composition is highly variable. The reason for the variability, according to Carvalho, Prazeres & Rivas (2013), is that the composition is dependent on the fraction of non-valorised cheese whey and quantity of cleaning wastewater disposed of. Particularly for this case study, variability is expected because whey is introduced and blended into the other wastewater in batches.

Table 8: Comparison of measured data versus the digester's design basis and various published cheese whey wastewater characteristics

Component	Unit	Steady-state Data#			Design	Literature References		
		Min	Max	Avg	Basis*	1	2	3
Flow	m^3/d	440	1613	1158	1500	-	-	-
Temperature	°C	32.9	38.9	36.3	35 ± 3	-	-	-
COD	mg/l	3504	11939	7352	8260	5400 - 77300	71410	8800 - 25600
pH	-	5.3	8.8	6.3	4 - 7	4.3 - 8.7	5.92	4.0 - 4.6
Particulate COD (PCOD)	mg/l	542	6603	3448	4090	-	-	-
Soluble COD (SCOD)	mg/l	1495	5983	3904	4170	-	-	-
Total Suspended Solids (TSS)	mg/l	210	2580	1360	2600	-	-	1600 - 4800
Volatile Suspended Solids (VSS)	mg/l	200	2400	1270	-	-	-	-
Total Kjeldahl Nitrogen (TKN)	mg/l	14	355	193	360	-	1610	310 - 360
Ammonia Nitrogen (NH_4-N)	mg/l	5	91	40	720	-	161	52 - 71
Nitrate (NO_3-N)	mg/l	172	239	203	170	-	-	-
Nitrite (NO_2-N)	mg/l	81	111	94	70	-	-	-
Total Phosphorus (TP)	mg/l	5	58	44	62	-	-	6.6 - 7.2
Calcium	mg/l	24	116	56	60	-	-	-
Magnesium	mg/l	1	118	25	95	-	-	-
Sulphate	mg/l	5	58	44	45	-	-	-
Volatile Fatty Acids (VFA)	mg/l	301	2041	1279	-	-	-	-
Alkalinity	mg/l as $CaCO_3$	180	1248	673	-	-	-	-
Lactose	g/l	-	-	-	-	-	44.37	0.178 - 0.182
Proteins	g/l	-	-	-	-	2.3 - 33.5	9.06	-
Fats & oils	g/l	-	-	-	-	0.4 - 5.7	-	1.83 - 3.76

\# Refers to Day 80 – 230 of actual plant operation. Data corresponds to the Digester Feed stream (Stream 1).
* Refers to the reference design basis followed by the plant designer.
Lit. Ref. 1 Kalyuzhnyi, Perez Martinez and Rodriguez Martinez (1997)
Lit. Ref. 2 Yang, Yu and Hwang (2003)
Lit. Ref. 3 Rivas *et al.* (2010)

3.1.4. Translating full-scale plant data to ADM1

ADM1 requires the influent substrate to be described as **26 state variables**, of which 20 variables are COD-based variables. These variables are far more sophisticated than the rather basic type of measurements taken in full-scale operation (cf. Table 7). The use of some reasonable assumptions is therefore necessary in order to decompose the measurement data.

SCOD and PCOD measurements are considered as key starting points for COD decomposition. The theoretical relationship between each state variables with these two measurements is illustrated in Figure 6. The first step involves estimating the biodegradability of the COD because it allows one to classify the proportion of inerts and the biodegradable components with respect to SCOD and PCOD respectively.

The second step involves differentiating the biodegradable components into the more sophisticated components as required by ADM1 structure. Published literature that have reported composition of similar substrate type had to be referenced since the plant measurements do not have such analysis.

Lastly, all components are converted to kgCOD/m^3 based on their theoretical specific COD. The philosophy used for non-COD based state variables is discussed further in subsequent sections.

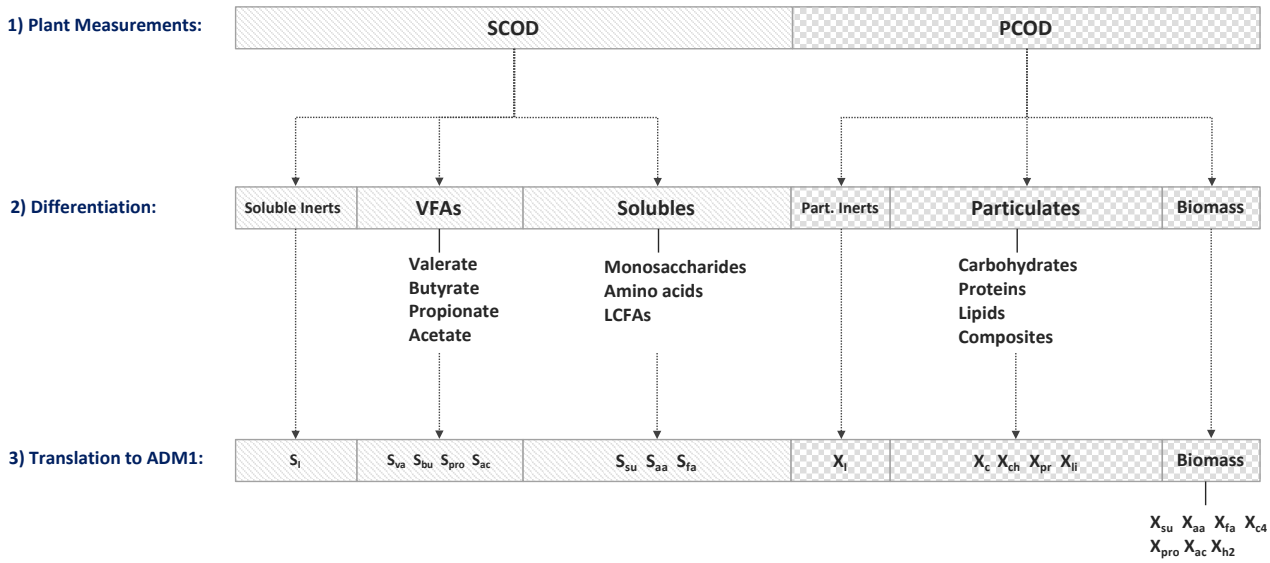


Figure 6: Schematic describing how COD measurements are differentiated and translated into ADM1 state variables

3.1.5. Substrate Biodegradability

Establishing the biodegradability of the substrate is one of the most critical steps when translating the substrate characteristics to ADM1 input components. It allows one to differentiate the degradable organic components of the COD apart from the inert fraction, which as a result, defines the amount of COD available for reactions. Biodegradability factor therefore has a direct influence on all modelling outputs.

To evaluate a substrate's biodegradability, a BMP (Bio-Methane Potential) test could be performed (Angelidaki *et al.*, 2009). The BMP test aims to quantify the volume of methane produced from a known quantity of

substrate per given quantity of biomass, thereafter Equation 32 is applied to quantify the biodegradable portion of the substrate.

$$f_d = \frac{B_0}{350\ COD_t} VS \qquad \textit{Equation 32}$$

where:

f_d is the biodegradable part of COD_t;

B_0 is the ultimate methane potential [$Nm^3.CH_4$ ton VS-1];

COD_t is the total COD of substrate added at the start of the test [kg $COD.m^{-3}$];

VS is the concentration of volatile solids [$kg.m^{-3}$];

However, as discussed in Section 2.4, using parameters (including biodegradability factor) determined from a batch test may not be appropriate for modelling a full-scale continuous plant. It is proposed that the substrate's biodegradability is approximated from the continuous plant data by evaluating the amount of COD consumed for: (i) methane gas production, (ii) biomass production and (iii) denitrification against the total amount of COD that has entered the digester during the steady-state period.

For biodegradability evaluation, "steady-state" period is defined as Day 80 to Day 230 of plant operation, as it is evident from Figure 7 that plant ramp-up took place predominantly from Day 1 to Day 79.

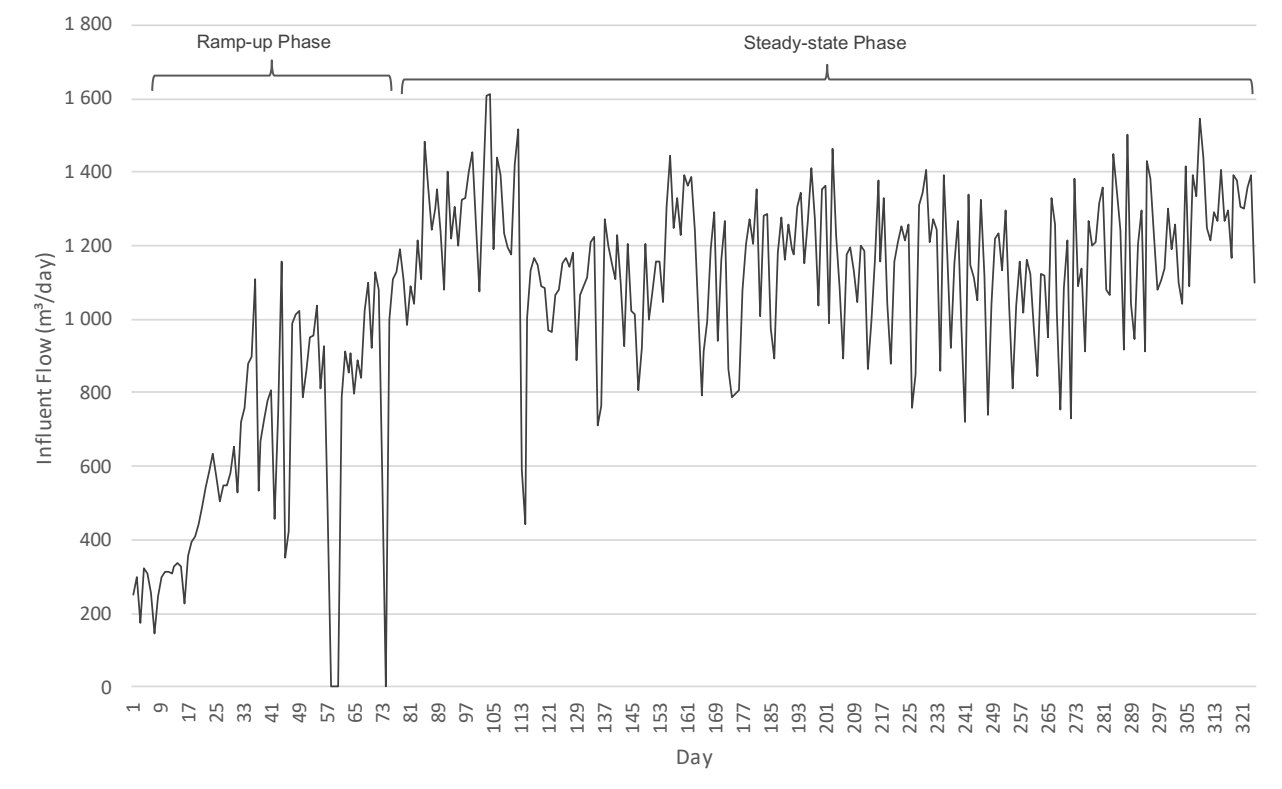


Figure 7: Daily influent volumetric flow into digester, showing the ramp-up period (Day 1 - Day 79) and the steady-state period (Day 80 - Day 230)

Equation 33 describes how the biodegradability factor (f_d) is estimated from the continuous plant data. It should be noted that this factor refers to the overall degree of biodegradability describing both soluble and particulate components in the substrate. Although specific biodegradability of soluble components may differ from that of the particulate components, provided that the digestion process is not hydrolysis limited, a universal factor is considered to be reasonable.

$$f_d \approx \frac{Total\ COD\ converted}{Total\ COD\ entered\ digester}$$

$$\approx \frac{COD_{CH4} + COD_{BM}}{\left(COD_{reactor,i} - COD_{reactor,f}\right) + \left(COD_{inf} - COD_{eff}\right)} \qquad \textit{Equation 33}$$

where:

f_d is the biodegradable part of total COD within the digester during the steady-state period;

COD_{CH4} is the theoretical amount of COD attributed to methane gas produced during the steady-state period [kg];

COD_{BM} is the amount of COD consumed for microbial growth (evaluated as the excess sludge wasted) during the steady-state period [kg];

$COD_{reactor,i}$ is the amount of COD present in the digester on the first evaluated day i.e. Day 80 [kg];

$COD_{reactor,f}$ is the amount of COD present in the digester on the last evaluated day i.e. Day 230 [kg];

COD_{inf} is the amount of COD that entered the digester as influent during the steady-state period [kg];

COD_{eff} is the amount of COD that exits the digester as during the steady-state period [kg]

Denitrification

It is important to take into account the potential loss to denitrification, particularly when treating substrates with high concentrations of nitrates and nitrites, such as in this case.

The denitrification term (COD_{DN}) is defined to quantify the portion of total COD lost due to denitrification. Since the ADM1 model employed in this study is the default model, which does not have the mechanism to simulate the denitrification process, the amount of biodegradable COD would be an overestimation unless the loss to denitrification is discounted from the influent COD. The biodegradability factor without denitrification is hence corrected from Equation 34 as:

$$f_d' \approx \frac{COD_{CH4} + COD_{BM}}{\left(COD_{reactor,i} - COD_{reactor,f}\right) + \left(COD_{inf} - COD_{eff}\right) - COD_{DN}} \approx 0.88 \qquad \textit{Equation 34}$$

where:

COD_{DN} is the amount of COD consumed due to denitrification of nitrates and nitrites during the steady-state period [kg];

Refer to the Appendix, Section 8.6.1, for an example of how the above factor is calculated.

Sulphate Reduction

COD loss to sulphate reduction is deliberately omitted. Based on the theoretical loss as described in Section 2.1.1, the low sulphate concentration in the influent implies that the impact will be insignificant.

3.1.6. Influent Soluble State Variables

The biodegradable portion of SCOD, which represents the hydrolysed products (monosaccharides, amino acids and LCFA) and VFAs, can be determined once the biodegradability factor is known. Next, to differentiate the COD contribution of hydrolysed products and VFAs, VFA constituents are first determined because VFA measurements are available.

VFA is measured as a lumped concentration in mg/l. Compositional make-up of individual VFA constituents (valerate, butyrate, propionate & acetate) is approximated from a study investigating acidification of lactose wastewater (Yu & Fang, 2001). Thereafter, the COD contribution of each VFA constituents is calculated based on their specific COD content (Grau et al., 2007).

Once the COD content of VFA is established the remainder of SCOD refers to the hydrolysed products. Compositional make-up of monosaccharides, amino acids and LCFA is assumed to have identical proportion as the carbohydrates, protein and lipids fractions respectively. Composition of carbohydrates, protein and lipids for cheese whey wastewater is widely reported in literature. Similar to the method followed for VFA, literature compositional guidelines are referenced and their specific COD values applied accordingly.

Some dissolved hydrogen is expected in the influent since partial acidification will occur in the buffer tank. However, its concentration is negligible in comparison to the concentration within the methanogenic digester; hence the state variable for dissolved hydrogen is set to zero. Methanogenic activity in the buffer tank is expected to be non-existent, given the fact that no sludge was recirculated to the buffer tank. The state variable for dissolved methane is thus set to zero as well.

Cations are calculated by summing the concentration of dissolved Ca^{2+}, Mg^{2+}, Na^{+} and K^{+} ions. The bivalent ions are plant measurements, whereas the monovalent ions are referenced from a dairy wastewater characterisation study (Danalewich et al., 1998). Remaining anions is calculated by ionic balancing:

$$Anions = Cations - [OH^-] - [HCO_3^-] - [VFAs] \qquad \textit{Equation 35}$$

An overview of the methodology applied to estimate soluble state variables is provided in Table 9 and Table 10. Refer to the Appendix, Section 8.6.2, for calculation examples.

Table 9: Translating full-scale plant data to influent soluble state variables in ADM1

State variable	Description	Method of estimation
$S_{su,in}$	Monosaccharides	(SCOD x f_d' – VFA_{COD}) x η_{su}
$S_{aa,in}$	Amino acids	(SCOD x f_d' – VFA_{COD}) x η_{aa}
$S_{fa,in}$	Long-chain fatty acids	(SCOD x f_d' – VFA_{COD}) x η_{fa}
$S_{va,in}$	Valerate	VFA x $\eta_{va,VFA}$ x Y_{va}
$S_{bu,in}$	Butyrate	VFA x $\eta_{bu,VFA}$ x Y_{bu}
$S_{pro,in}$	Propionate	VFA x $\eta_{pro,VFA}$ x Y_{pro}
$S_{ac,in}$	Acetate	VFA x $\eta_{ac,VFA}$ x Y_{ac}
$S_{h2,in}$	Dissolved hydrogen	Set to zero
$S_{ch4,in}$	Dissolved methane	Set to zero
$S_{IC,in}$	Inorganic carbon	Converted from bicarbonate alkalinity measurements
$S_{IN,in}$	Inorganic nitrogen	Converted from ammonium-N measurements
$S_{I,in}$	Soluble inerts	SCOD x (1 – f_d')
$S_{an,in}$	Anions	Ionic balancing (see Equation 35)
$S_{cat,in}$	Cations	Sum of Ca^{2+}, Mg^{2+}, Na^{+} & K^{+} ions as kmol/m³

Table 10: Definition of terms denoted in Table 9

Term	Description	Unit	Definition/Reference
SCOD	Soluble COD concentration	kg COD/m³	Obtained from actual plant sampling and analysis
f_d'	Biodegradability w/o denitrification	-	Equation 34 from Section 3.3.4
VFA_{COD}	VFA COD	kg COD/m³	Sum of COD associated with valeric acid, butyric acid, propionic acid and acetic acid
η_{su}	Monosaccharides fraction	-	Assume similar proportion as carbohydrates fraction (η_{ch} = **0.44**). See definition in Table 12 for η_{ch}
η_{aa}	Amino acids fraction	-	Assume similar proportion as protein fraction (η_{pr} = **0.40**). See definition in Table 12 for η_{pr}
η_{fa}	Long-chain fatty acids fraction	-	Assume similar proportion as lipids fraction (η_{li} = **0.16**). See definition in Table 12 for η_{li}
VFA	Volatile fatty acids concentration	mg/l	Obtained from actual plant sampling and analysis
$\eta_{va,VFA}$	Valerate fraction	-	Approximated from acidification of dairy wastewater (Yu & Fang, 2001) as **0.09**
$\eta_{bu,VFA}$	Butyrate fraction	-	Approximated from acidification of dairy wastewater (Yu & Fang, 2001) as **0.37**
$\eta_{pro,VFA}$	Propionate acid fraction	-	Approximated from acidification of dairy wastewater (Yu & Fang, 2001) as **0.22**

Term	Description	Unit	Definition/Reference
$\eta_{ac,VFA}$	Acetate fraction	-	Approximated from acidification of dairy wastewater (Yu & Fang, 2001) as **0.31**
Y_{va}	COD equivalent of valerate	kg COD. $(kg\ va)^{-1}$	Y_{va} = **2.039216** (Grau *et al.*, 2007)
Y_{bu}	COD equivalent of butyrate	kg COD. $(kg\ va)^{-1}$	Y_{bu} = **1.818182** (Grau *et al.*, 2007)
Y_{pro}	COD equivalent of propionate	kg COD. $(kg\ va)^{-1}$	Y_{pro} = **1.513514** (Grau *et al.*, 2007)
Y_{ac}	COD equivalent of acetate	kg COD. $(kg\ va)^{-1}$	Y_{ac} = **1.066667** (Grau *et al.*, 2007)
Ca^{2+}	Calcium ions	$kmol.m^3$	Obtained from actual plant sampling and analysis as mg/l
Mg^{2+}	Magnesium ions	$kmol.m^3$	Obtained from actual plant sampling and analysis as mg/l
Na^{+}	Sodium ions	$kmol.m^3$	Interpolated from dairy wastewater characterisation survey (Danalewich *et al.*, 1998) as **0.0536** $kmol/m^3$
K^{+}	Potassium	$kmol.m^3$	Interpolated from dairy wastewater characterisation survey (Danalewich *et al.*, 1998) as **0.00387** $kmol/m^3$

3.1.7. Influent Particulate State Variables

The composite particulate variable (X_c) is defined in ADM1 to represent a broad consortium of organic matter that includes dead biomass and miscellaneous organic matter with complex composition. It is not possible to estimate this variable given the available plant data. However, assuming disintegration is complete in the buffer tank, all particulate COD in the influent wastewater would have already disintegrated and X_c may be set to zero. This philosophy is in line with the recommendation made by Batstone *et al.* (2015) and Arnell *et al.*, (2016) which suggests that X_c shall be disregarded when describing feed substrate due to its broad definition as well as the difficulty to quantify it by means of analytical methods.

Acidifiers (X_{su}, X_{aa}, X_{fa}) are expected to be present in the influent, considering the fact that some degree of pre-acidification will occur in the buffer tank. The amount of acidifiers is estimated based on its yield. According to a study by Yu & Fang (2002) which investigated acidogenesis of dairy wastewater, biomass yield is observed to be 0.26 g-VSS per g-COD removed. The amount of COD removed due to acidogenesis in this case study is however unknown because no data regarding the raw untreated wastewater entering the buffer tank is available. For this reason, COD of this stream is approximated by basing the ratio of COD converted to VFAs similar to that of the referenced study.

Acetogens and methanogens (X_{c4}, X_{pro}, X_{ac}, X_{h2}) are considered negligible. Growth of these species are not favoured since pH of the buffer tank remains at most times below 6.5 (Yu & Fang, 2001). An overview of the methodology applied to estimate soluble state variables is provided in Table 11 and Table 12. Refer to the Appendix, Section 8.6.3, for calculation examples.

Table 11: Translating full-scale plant data to particulate soluble state variables in ADM1

State variable	Description	Method of estimation
$X_{c,in}$	Composite particulate	Set to zero
$X_{ch,in}$	Carbohydrates	$(PCOD \times f_d' - X_{degr}) \times \eta_{ch}$
$X_{pr,in}$	Proteins	$(PCOD \times f_d' - X_{degr}) \times \eta_{pr}$
$X_{li,in}$	Lipids	$(PCOD \times f_d' - X_{degr}) \times \eta_{li}$
$X_{su,in}$	Monosaccharide degraders/biomass	$[VFA_{COD} \div f_{acid} - SCOD \times f_d'] \times Y_{acid} \times Y_{bm} \times \eta_{ch}$
$X_{aa,in}$	Amino acids degraders/biomass	$[VFA_{COD} \div f_{acid} - SCOD \times f_d'] \times Y_{acid} \times Y_{bm} \times \eta_{pr}$
$X_{fa,in}$	Long-chain fatty acids degraders/biomass	$[VFA_{COD} \div f_{acid} - SCOD \times f_d'] \times Y_{acid} \times Y_{bm} \times \eta_{li}$
$X_{c4,in}$	Valerate & Butyrate degraders/biomass	Set to zero
$X_{pro,in}$	Propionate degraders/biomass	Set to zero
$X_{ac,in}$	Acetate degraders/biomass	Set to zero
$X_{h2,in}$	Hydrogen degraders/biomass	Set to zero
$X_{I,in}$	Particulate inerts	$PCOD \times (1 - f_d')$

Table 12: Definition of terms denoted in Table 11

Term	Description	Unit	Definition/Reference
PCOD	Particulate COD	kg COD/m³	Obtained from actual plant sampling and analysis. PCOD = TCOD – SCOD
X_{degr}	Degraders COD	kg COD/m³	Sum of all degraders (i.e. X_{su}, X_{aa}, X_{fa}, X_{c4}, X_{pro}, X_{ac}, X_{h2})
η_{ch}	Carbohydrates fraction	-	Composition of typical cheese whey wastewater (Kalyuzhnyi *et al.*, 1997; Yang *et al.*, 2003): η_{ch} = **0.44**
η_{pr}	Protein fraction	-	Composition of typical cheese whey wastewater (Kalyuzhnyi *et al.*, 1997; Yang *et al.*, 2003): η_{pr} = **0.40**
η_{li}	Lipids fraction	-	Composition of typical cheese whey wastewater (Kalyuzhnyi *et al.*, 1997; Yang *et al.*, 2003): η_{li} = **0.16**
f_{acid}	Acidification	-	Fraction of COD in the buffer tank converted to VFAs and alcohols as a result of acidification of dairy wastewater (Yu & Fang, 2002): f_{acid} = **0.484**
Y_{acid}	Acidogen Yield	g VSS/g COD	Production rate of acidogens during acidification of dairy wastewater at pH of 6.5 (Yu & Fang, 2002): Y_{acid} = **0.26**
Y_{bm}	COD-equivalent of biomass	g COD/g VSS	Commonly accepted COD value for biomass (Eastman & Ferguson, 1981): Y_{bm} = **1.42**

3.1.8. Sludge Extraction

Sludge is periodically wasted from the digester such that a F/M (Food-to-Mass) ratio of no less than 0.3 is maintained. The sludge wasting process is initiated manually under the discretion of the plant operator, and the amount of sludge to be wasted is determined according to the daily COD loading relative to the digester's VSS content. An average ratio of 0.4 was maintained during the experiment.

Having sufficient, but not excessive, biomass sludge with respect to the COD loading is critical for good digestion performance, as it promotes optimum microbial activity (Chen & Hashimoto, 1996). Optimal F/M ratio varies with different types of substrates. For example, F/M ratio for cellulosic substrates were found be around 0.5 while substrates with high-fat content tend to degrade well at F/M ratios between 0.33 – 1.25 *(*Chynoweth *et al.*, 1993; Raposo *et al.*, 2006).

ADM1's mass balance equation calculates the net change in concentration of each particulate component (cf. Equation 19) during each ODE time-step. Although it takes into account the influent concentration, growth and decay reactions, the default equation does not consider losses due to sludge wastage. In that regard, an additional "sludge wastage" term ($X_{liq,sw}$) is introduced into Equation 19, which modifies to Equation 36.

$$\frac{dX_{liq,i}}{dt} = \frac{Q}{V_{liq}}\left(X_{in,i} - X_{liq,i}\right) + \sum_{j=1-19} \rho_j \upsilon_{i,j} - X_{liq,sw} \qquad \textit{Equation 36}$$

where:

$$X_{liq,sw} = \frac{Q_{sludge} \times X_{liq,i}}{V_{liq}}$$

where:

$X_{liq,sw}$ is the concentration of a particulate variable relative to the digester volume that is lost due to sludge wastage [kgCOD/m^3];

Q_{sludge} is the volume of sludge wasted daily [m^3/day];

$X_{liq,i}$ is the concentration of a particulate variable within the digester [kgCOD/m^3]

3.1.9. Mass Balance Modification

As described in Section 3.3.2 the type of AD process employed in this study is an Anaerobic MBR process whereby a CSTR digester is coupled with an externally pressurised ultrafiltration (UF) process. Even though ADM1 is adequate to model the CSTR process, by default it does not take into account the biomass separation and recirculation process. Because practically all particulates are held back by the UF membrane, the mass balance equation for particulates (Equation 36) further simplifies to Equation 37.

$$\frac{dX_{liq,i}}{dt} = \frac{Q}{V_{liq}} X_{in,i} + \sum_{j=1-19} \rho_j \upsilon_{i,j} - X_{liq,sw} \qquad \textit{Equation 37}$$

3.1.10. Computational Software Setup

The ADM1 model was coded and implemented using Scilab version 5.5.2. Refer to the Appendix, Section 8.7, for the Scilab codes. The model is designed to retrieve data from a Microsoft Excel spreadsheet which contains information required to initialise the simulation, such as:

- Influent volume for each time interval
- Influent state variables (calculated from actual plant water analysis) for each time interval
- Actual plant measurands (e.g. VFA, N-NH_4, VSS, pH, CH_4 and CO_2) for each time interval
- Sludge wastage volume for each time interval
- Digester initial state variables
- Lower and upper bounds of parameters for which Monte Carlo is applied

Besides stoichiometric and kinetic parameters, all other model parameters (k_La, pKa, acidity, carbon & nitrogen contents, etc.) are assigned their default values throughout the study, as they are mostly theoretical values and unlikely to change. Model inputs related to the digester were defined according to the digester's specifications (Table 13).

Table 13: Values specific to the digester modelled in this study

Model Input	Description	Unit	Value
$V_{digester}$	Volume of CSTR	m^3	Actual liquid volume in digester: $V_{digester}$ = **2875**
$V_{headspace}$	Volume of headspace	m^3	Actual volume above liquid in digester: $V_{headspace}$ = **165**
T	Digester temperature	K	Temperature is controlled by heat exchanging digester contents with steam. Set-point is 35°C and fluctuation is maintained well within 3°C despite slight fluctuations in the influent temperature (cf. Table 8). Constant temperature applied for model: K = **308**
P_{atm}	Atmospheric pressure	bar	Plant is situated near sea level. P_{atm} = **1.013**

3.1.11. Limitations & Assumptions

- Concentrations of toxic or inhibiting substances present in the raw wastewater are assumed to remain constant throughout the study period
- A single biodegradability value is assumed to be generalisable across the entire study period
- Temperature correction is only applied for those parameters suggested for the default ADM1 model because temperature fluctuation was well controlled within 3°C
- It is assumed that homogenous mixing is achieved in the CSTR, such that the extracted sludge has an identical composition as the digester's content.
- The formation and accumulation of inorganic precipitates such as calcium carbonate and struvite are not considered in the modelling

3.2. PLS Method

3.2.1. Definition of Terms

Key terminologies applicable to the PLS Method are defined below:

- **"STR"** – Refers to the IWA Scientific and Technical Report, which elaborates the ADM1 model framework and the default value for each parameter.
- **"Parameters"** – Refers to the 58 stoichiometric and kinetic parameters related to the biochemical reactions in ADM1.
- **"Model inputs"** – Refers to the information required by ADM1 in order to commence simulation. These are analyses recorded relating to the digester feed stream and digester conditions during actual plant operation, which are subsequently translated into ADM1 format.
- **"Measured outputs"** – Refers to the 6 plant measurements recorded during actual plant operation, which are subjected to modelling i.e. VFA, Ammonium, VSS, pH, Methane flow and Carbon dioxide flow.
- **"Model outputs"** – Refers to the results produced from ADM1 simulation, of which correspond to the 6 plant measurements.
- **"Input Matrix"** – Refers to the first of the two data matrices required by the PLSR algorithm. In this study, it contains all the randomised parameter sets produced during Monte Carlo.
- **"Output Matrix"** – Refers to the second of the two data matrices required by the PLSR algorithm. In this study, it contains all the model outputs produced during Monte Carlo.

3.2.2. Concept Introduction

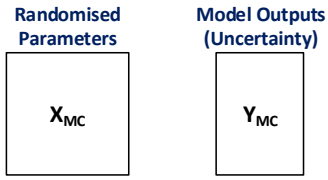
Calibrating large numbers of parameters in ADM1 is a challenge. Besides being a lengthy process, overfitting is known to occur. The PLS method developed in this thesis aims to exploit the merits of PLS regression. Firstly, through its dimension reduction function, calibration is expected to simplify since the scope reduces from 58 parameters to a few latent variables. Secondly, PLS regression is capable of mapping a latent relationship between two multivariate matrices (input matrix and output matrix) while taking into account the collinearity between the inputs. Therefore, by specifying the parameters and model outputs as the two matrices, one could establish how each parameter is transformed in order to effect the highest variance in the outputs. This understanding could allow parameters to calibrate in a guided/supervised manner and potentially shorten the time required to optimise or recalibrate a model.

Figure 8 illustrates how the PLSR framework is employed to extract the latent relationship and to generate a newly calibrated parameter set. An overview is provided in this section. Detailed description of each step is provided in the subsequent sections.

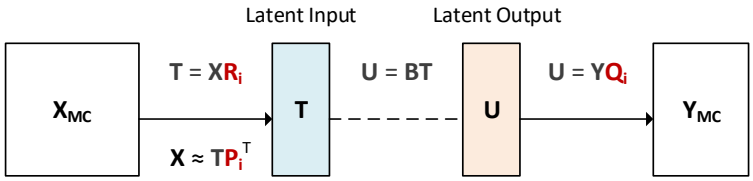
The first step involves the use of Monte Carlo to generate the two matrices. The input matrix $\mathbf{X_{MC}}$ includes randomised parameter values, whilst the output matrix $\mathbf{Y_{MC}}$ collects the model outputs corresponding to each randomised set of parameters.

How PLSR is Utilised in the PLS Method

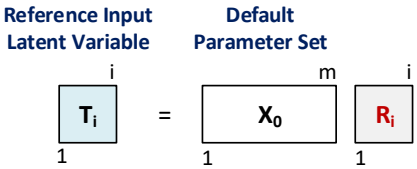
***Step 1**: Produce X_{MC} and Y_{MC} matrices using Monte Carlo*



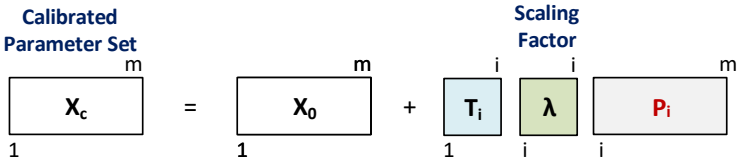
***Step 2**: Apply PLSR algorithm (*Figure 4*) to extract relationship as reference weights and loadings (R_i, P_i, Q_i)*



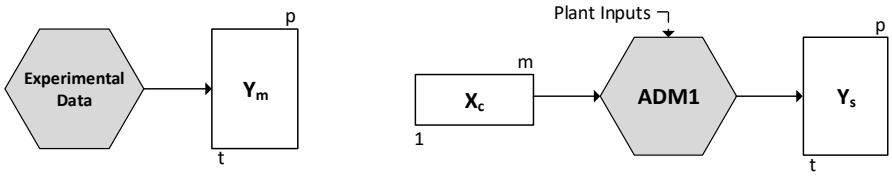
***Step 3.1**: Use the extracted input weight vector (R_i) to produce a reference input latent variable (T_i)*



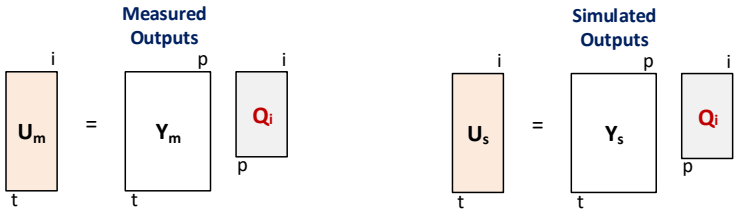
***Step 3.2**: Generate a calibrated parameter set by adjusting the scaling factor (λ). λ is initially set as zero.*



Consolidate experimental data and ADM1 outputs (based on calibrated parameter set) into respective output matrices.



***Step 3.3**: Apply the extracted output weight vector (Q_i) to transform simulated and measured outputs to latent variables*



Optimise the model by repeating Steps 3.2 &3.3 until the objective function (U_{diff}) is minimised

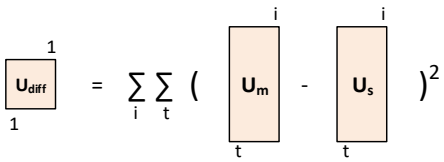


Figure 8: Illustration showing how the PLSR framework is incorporated into the PLS Method. First, the relationship between the parameters and the model outputs are mapped as weights and loadings. Thereafter, these PLS constructs are applied to guide parameter calibration. i - no. of latent variables; m - no. of parameters; p - no. of outputs; t - no. of time intervals simulated.

In the subsequent step, PLSR algorithm uncovers the latent relationship in the form of parameter weighting vector (**R**), input loading vector (**P**) and output weighting vector (**Q**). These PLS constructs describe how the parameters and outputs need to be transformed such that the covariance between the input & output latent variables are maximised.

Parameter calibration takes place in Step 3. The PLS Method exploits the fact that **P** represents the vector direction between the latent input variable(s) **T** and **X**, which corresponds to maximum change in the output latent variable(s). Hence, by making use of a scaling factor (**λ**) to adjust **T**, effectively all parameters within **X** would alter relative to the relationship captured in **P**. This newly calibrated parameter set is then applied in ADM1 to produce new simulated outputs $\mathbf{Y_s}$. Note that **λ** is a diagonal matrix with each element corresponding to the scaling of a particular latent input variable.

The final part of Step 3 evaluates the model's objective function. **Q** represents the weighting that each output contributes towards the output latent variable **U**. With a smaller dimension than **Y**, **U** serves a useful score that represents a consolidation of the model outputs. Each output's sensitivity, as learnt from the Monte Carlo simulation (Step 1), is taken into account during the calculation of **U**.

The objective function is defined as the difference between the output latent variables of the simulated data and measured data. A model is said to be optimised when this function is minimised. At each iteration, **λ** is adjusted depending on the previous objective function.

Lastly, once the optimised parameter set is identified, a validation check is performed. This final step is deemed necessary to ensure that the parameter set does not only fit the training data but fits new unseen data as well.

3.2.3. Step 1 – Uncertainty Analysis Using Monte Carlo Method

In this step, Monte Carlo (MC) simulation is applied by running ADM1 repeatedly using randomised stoichiometric and kinetic parameters. The objective is to propagate parameter uncertainties into the outputs such that sensitivity analysis can be applied in Step 3.

The lower and upper bounds of the range, in which the parameters are randomised, were based on the minimum and maximum values observed in the literature survey (Table 4). For fractional stoichiometric parameters, the algorithm is set up in such a way that the sum of fractional parameters adds up to 1.

Random Sampling is selected for this study because firstly the literature survey took reference from articles that covered many types of substrates; and secondly, these referenced experiments were carried out under different conditions and with different inoculum compositions. As such, there is no valid population distribution available to apply stratified sampling methods. Parameter uncertainties are considered stochastic and randomisation follows a uniform probability distribution.

At the end of each MC run, the randomised set of parameters and simulation results are collected in an "input" matrix and an "output" matrix, herewith referred to as $\mathbf{X_{MC}}$ and $\mathbf{Y_{MC}}$ respectively in accordance to the PLSR nomenclature described in Section 2.8. $\mathbf{X_{MC}}$ possesses a dimension that has rows corresponding to Monte Carlo samples and columns corresponding to randomised parameter sets.

Specific to this study, the "output" matrix includes 6 components constituted from ADM1 simulated outputs (Table 14). Although a total of 29 dynamic state variables are available in ADM1, only 6 plant measurands are applicable for model fitting. Evolution of these 6 components across all time intervals are stored as $\mathbf{Y_{MC}}$.

Table 14: Description of the 6 outputs/measurands included in the output matrix

Output Matrix Component	ADM1 Constituents
VFA	Sum of S_{fa}, S_{va}, S_{bu}, S_{pro}, S_{ac}
Ammonium [S_{IN}]	S_{IN}
VSS	Sum of X_c, X_{ch}, X_{pr}, X_{li}, X_{su}, X_{aa}, X_{fa}, X_{c4}, X_{pro}, X_{ac}, X_{h2}, X_I
pH	Calculated from charge balance
Methane Gas Flow Rate [q_{CH4}]	Calculated from gas phase mass balance
Carbon Dioxide Gas Flow Rate [q_{CO2}]	Calculated from gas phase mass balance

3.2.4. Step 2 – Determining PLSR Weights and Loadings

The intention of applying PLSR in this study is not for its main function i.e. regression. Instead, the technique is utilised for its sub-function which identifies weight vectors that, in linear combination with input variables, would induce greatest output sensitivity. In some literature this sub-function is referred to as the *underlying interaction* (Chin, Marcolin & Newsted, 2003; Maitra & Yan, 2008).

Using the $\mathbf{X_{MC}}$ and $\mathbf{Y_{MC}}$ matrices obtained in Step 1, the PLSR technique as described in Section 2.8 is applied to map out the underlying interactions which are represented as the weight and loading vectors (i.e. **r**, **p**, **q**). The key outcome from performing PLSR is to generate these vectors corresponding to the number of latent variables selected.

An alternative normalisation method, instead of mean-centering and variance scaling, was applied on the X and Y matrices. Equation 38 normalises the data generated from the Monte Carlo simulations into a range between -1 and +1. In other words, minimum values at each time interval are normalised to a value of -1 and the maximum values are normalised to +1.

$$\boldsymbol{X_{Norm}} = 2\big((\boldsymbol{X} - \boldsymbol{X}_{min})/(\boldsymbol{X}_{max} - \boldsymbol{X}_{min})\big) - 1 \qquad \textit{Equation 38}$$

$$\boldsymbol{Y_{Norm}} = 2\big((\boldsymbol{Y} - \boldsymbol{Y}_{min})/(\boldsymbol{Y}_{max} - \boldsymbol{Y}_{min})\big) - 1$$

where:

$\mathbf{X}_{Norm}$ is the normalised **X** matrix;

X is the "input" matrix that contains ADM1 parameters used for each Monte Carlo runs;

$\mathbf{X}_{max}$ is a vector that contains the maximum values in X at each time interval

$\mathbf{X}_{min}$ is a vector that contains the minimum values in X at each time interval

3.2.5. Step 3 – Model Fitting

Contrary to conventional model fitting procedures whereby only parameters selected by the modellers are calibrated, this study proposes to adjust the *input latent variable(s)* during each iteration. In essence, as the latent variable(s) change, *all* parameters will adjust proportionally according to the interaction construct with the outputs as identified in Step 2. Parameter calibration thus does not depend on parameter ranking or the modeller's discretion.

First, the weight vector ($\mathbf{r}_i$) is applied on the default ADM1 parameters ($\mathbf{x}_0$) to generate its latent input variable $\mathbf{t}_i$ (Equation 39). Thereafter, a scaling vector, $\boldsymbol{\lambda}$, is introduced to scale the input latent variable(s) - this is the *only* term that adjusts during each iteration. As $\boldsymbol{\lambda}$ changes, a newly calibrated set of ADM1 parameters is generated (Equation 40). It should be noted that $\mathbf{x_c}$ is a normalised vector; therefore to obtain the new parameter set in its original scale, a reversal of the normalisation procedure was carried out.

$$\boldsymbol{t}_i = \boldsymbol{x}_0 \boldsymbol{r}_i \qquad \textit{Equation 39}$$

$$\boldsymbol{x}_c = \boldsymbol{x}_0 + \sum \boldsymbol{\lambda}_i \boldsymbol{t}_i \boldsymbol{p}_i{}^T \qquad \textit{Equation 40}$$

where:

$\mathbf{x}_c$ is the newly calibrated parameters;

$\mathbf{x}_0$ is the default ADM1 parameters;

r is the input weight vector;

$\mathbf{t}_i$ is the i-th input latent variable;

p is the input loading vector;

λ is the scaling vector

Definition of the objective function is described in Equation 41. For each iteration step, simulated outputs are compared to the measured outputs in the form of output latent variable(s) $\mathbf{u}_j$. The use of latent variable(s) as an objective function eliminates the need to assign aggregates to each output, as it is already weighted intrinsically to the outputs' sensitivities. The vector **λ** is initially set as [0,0,..] and is calibrated until $\mathbf{u}_{diff}$ is minimised. Assigning a value of 0 to **λ** means that calibration process always initiates with the default ADM1 parameters i.e. $\mathbf{x_c} = \mathbf{x_0}$. Objective function minimisation is performed using the default *LeastSq* function in Scilab. This function is based on the quasi-Newton method.

$$\boldsymbol{u}_{diff} = min \sum_{j=1}^{i} \sum_{t=1}^{n} \left(\boldsymbol{u}_{m,j}(t) - \boldsymbol{u}_{s,j}(t) \right)^2 \qquad \textit{Equation 41}$$

where:

$\mathbf{u}_{diff}$ is the model objective function

n is the total number of time intervals;

i is the total number of latent variables selected;

$\mathbf{u}_{m,j}$ is the j-th output latent variable for the *measured* outputs at time interval t;

$\mathbf{u}_{s,j}$ is the j-th output latent variable for the *simulated* outputs at time interval t;

An example calculation demonstrating how u_{diff} is calculated is shown in Section 8.6.4. To evaluate the fitting accuracy of individual outputs, conventional RMSE (Equation 31) is applied.

3.2.6. Step 4 – Validation

The process of generating PLSR vectors (Step 3) and parameter calibration (Step 4) is based on experimental data (plant measurands) gathered from Day 1 to Day 230. There is a possibility, as pointed out in Section 2.9, that the calibrated parameter set only fits this specific set of data and not a true representation of the digester's mechanics.

Validation is applied as the method to validate the calibrated parameter set. Using unseen data collected from the subsequent 90 days of plant operation (Day 231 to Day 320), ADM1 ran with its simulation timeframe extended to 320 days. This extended period is referred to as the Validation Period. The simulated outputs are thereafter evaluated in terms of U_{diff} and RMSE .

3.3. Research Limitations

The methodology proposed in this study uses the default parameter set as a baseline starting point for model calibration. It is thus noted that the calibrated parameters could risk being a local search/minima solution.

Monte Carlo simulations are structured according to the parameters gathered from literature survey. The extent of uncertainty propagation, and hence the min/max values used for normalising the dataset, is dependent on the quantity of published data available for each parameter.

Quality of experimental data may be influenced by measurement error and low sampling frequency of off-line (i.e. manual) measurements (Guisasola *et al.*, 2006). Since it is not the scope of this research to interfere with the sampling and analysis schedule of the full-scale plant, there is limited control over the integrity of the experimental data, and hence necessitates assumptions to be made when fractionating the influent substrate into the ADM1 format. In that respect, this desktop study focuses primarily on the procedure and effectiveness of the parameter calibration framework.

3.4. Methodology Map

A methodology map summarising the four steps of PLS Method is presented in Figure 9 and Figure 10. It illustrates how the proposed method integrates into ADM1.

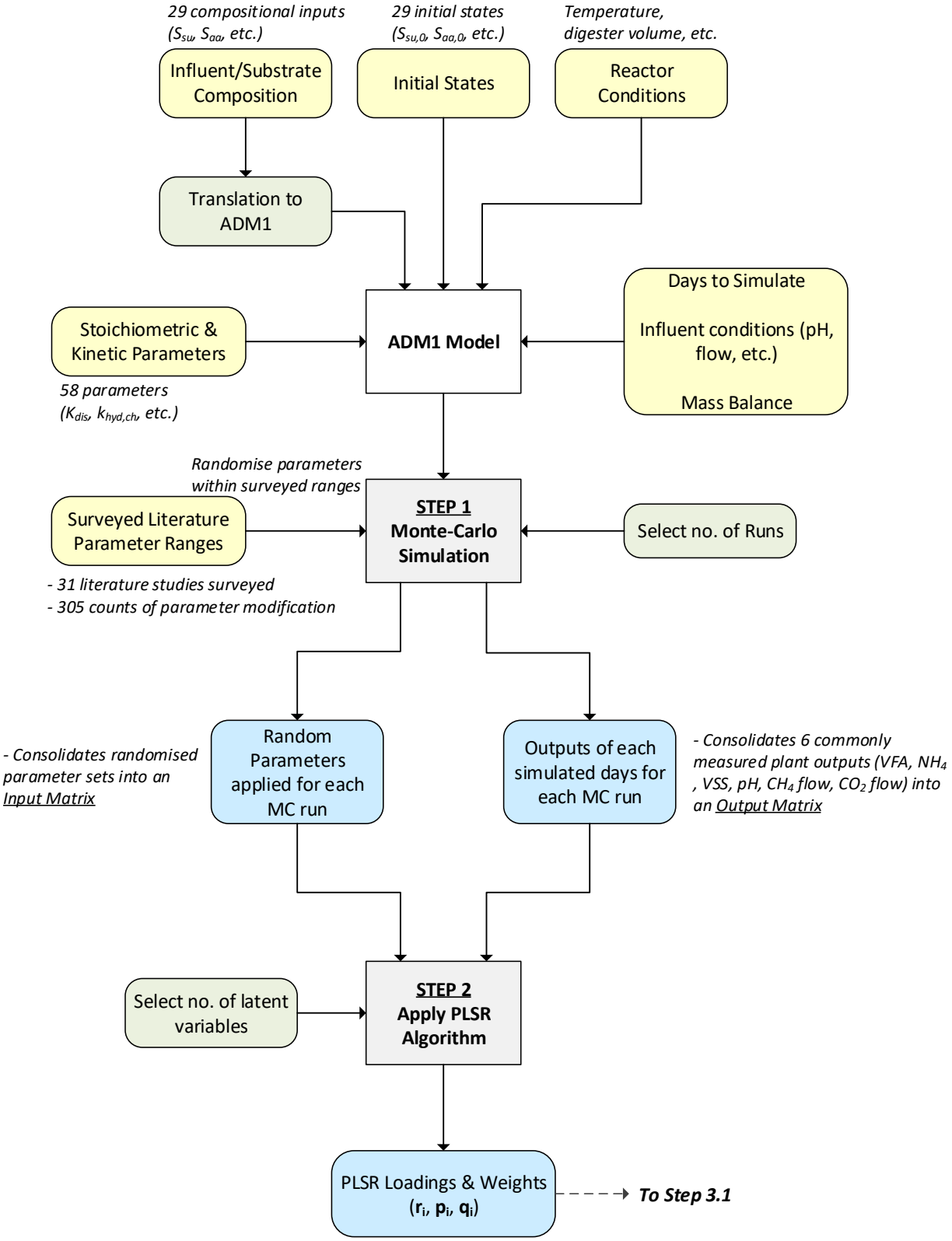


Figure 9: Overview of the PLS Method for ADM1 parameter calibration – Part 1 of 2

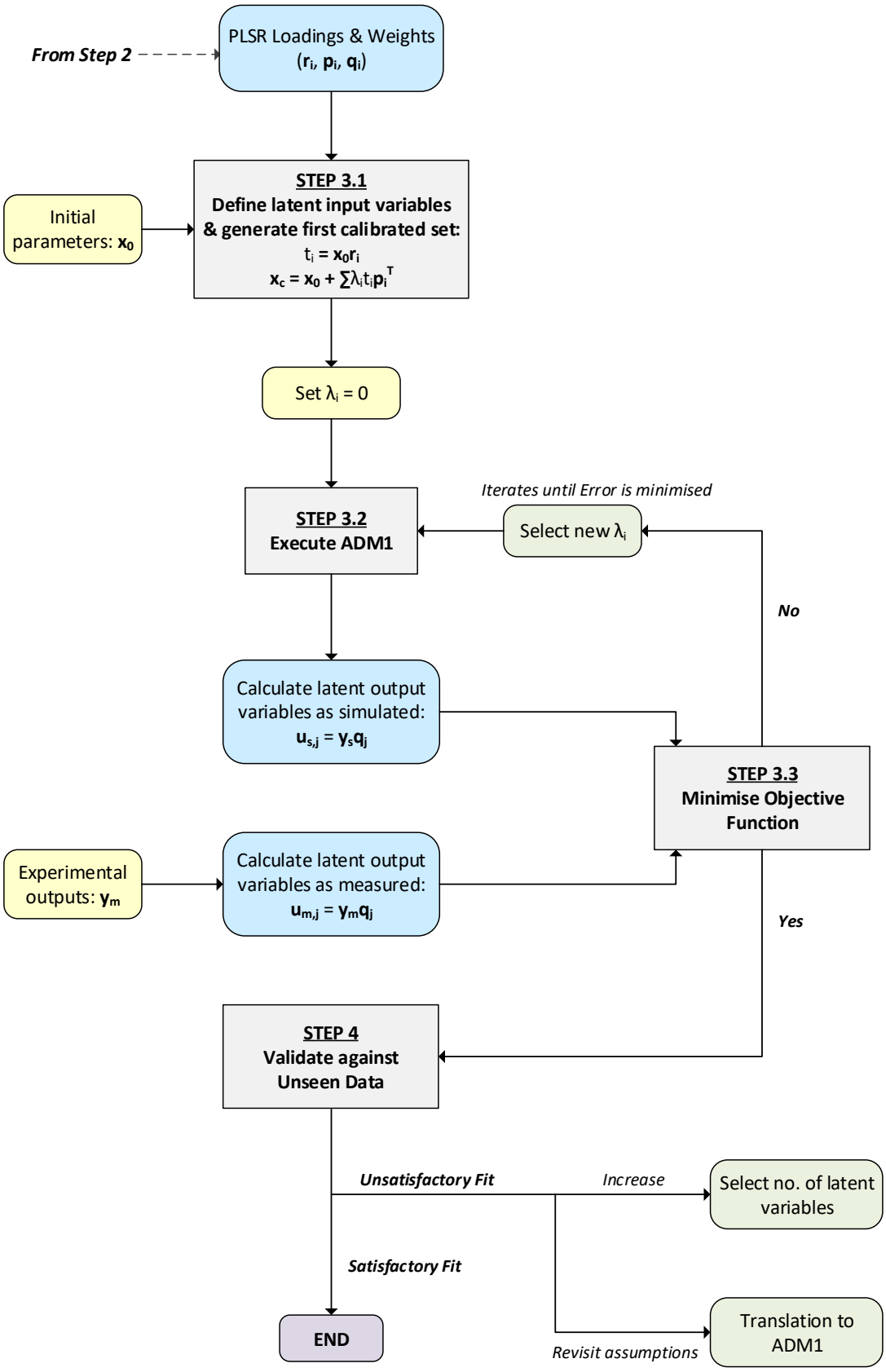


Figure 10: Overview of the PLS Method for ADM1 parameter calibration – Part 2 of 2

3.5. Model Benchmarking

As discussed in the literature review (Section 2.5), there is currently no single mainstream method for parameter calibration. For that reason, it is crucial to benchmark the PLS Method against other methods in order to verify the applicability of this method in terms of (i) model accuracy and (ii) optimisation speed.

The first method selected for benchmarking the PLS Method shall be referred to as the **"Group Method"**. This method was used by Coelho et al. (2006) to model digestion of dairy type wastewater. In this method, kinetic parameters were grouped into the 3 groups of sensitivities (High, Medium, Low) as suggested in the STR. The 3 groups are then calibrated sequentially in order of decreasing sensitivity. Stoichiometric parameters are not calibrated. Selection of this method is also supported by the observations made from the literature survey (Table 5) where parameters labelled with higher sensitivities are more exposed to calibration. Two variants of the Group Method were investigated, namely:

- **"Group Method (Unbounded)"** – This term refers to the application of the Group Method without any lower and upper limits imposed on the parameters during calibration. In other words, parameters are allowed to take on any low or high values.
- **"Group Method (Bounded)"** – This term refers to the application of the Group Method but with lower and upper limits assigned to each parameter according to the minimum and maximum values identified from the literature survey, respectively.

The second method selected for benchmarking against the PLS Method shall be referred to as the **"Brute Force Method"**. In this method, all 58 parameters are calibrated without any particular sequence, prioritisation or expert inputs. Lower and upper limits are, however, set as per the minimum and maximum values identified from the literature survey.

In order to benchmark the model accuracies between different methods, a single score representing the overall model accuracy was required. **MAPE** (Mean Absolute Percentage Error) is selected for this purpose because it expresses model accuracy in the form of percentage which is scale-independent. As all 6 outputs have different scales, RMSE (as applied for PLS Method) would affect bias towards outputs with larger scales (i.e. CH_4, CO_2) when an average RMSE score across all model outputs is taken. The objective function for both Group Method and Brute Force Method is expressed as the "Average MAPE of the six outputs".

CHAPTER 4

DEVELOPMENT OF THE PLS METHOD

4.1. ADM1 Simulation using Default Parameters

Figure 11 shows the projections of the 6 outputs when default parameters (as recommended by the STR) are applied. This outcome represents the *uncalibrated* model. It is evident, by visual inspection, that all simulated outputs except Ammonia (S_{IN}), followed the *trend* of the industrial plant measurements reasonably well. The projected trend of Ammonia is considered irregular because, even though there were close resemblances at certain time intervals (e.g. Day 100 – 120, Day 165 – 240 and Day 290 – 310), contradictory responses were observed across other time intervals.

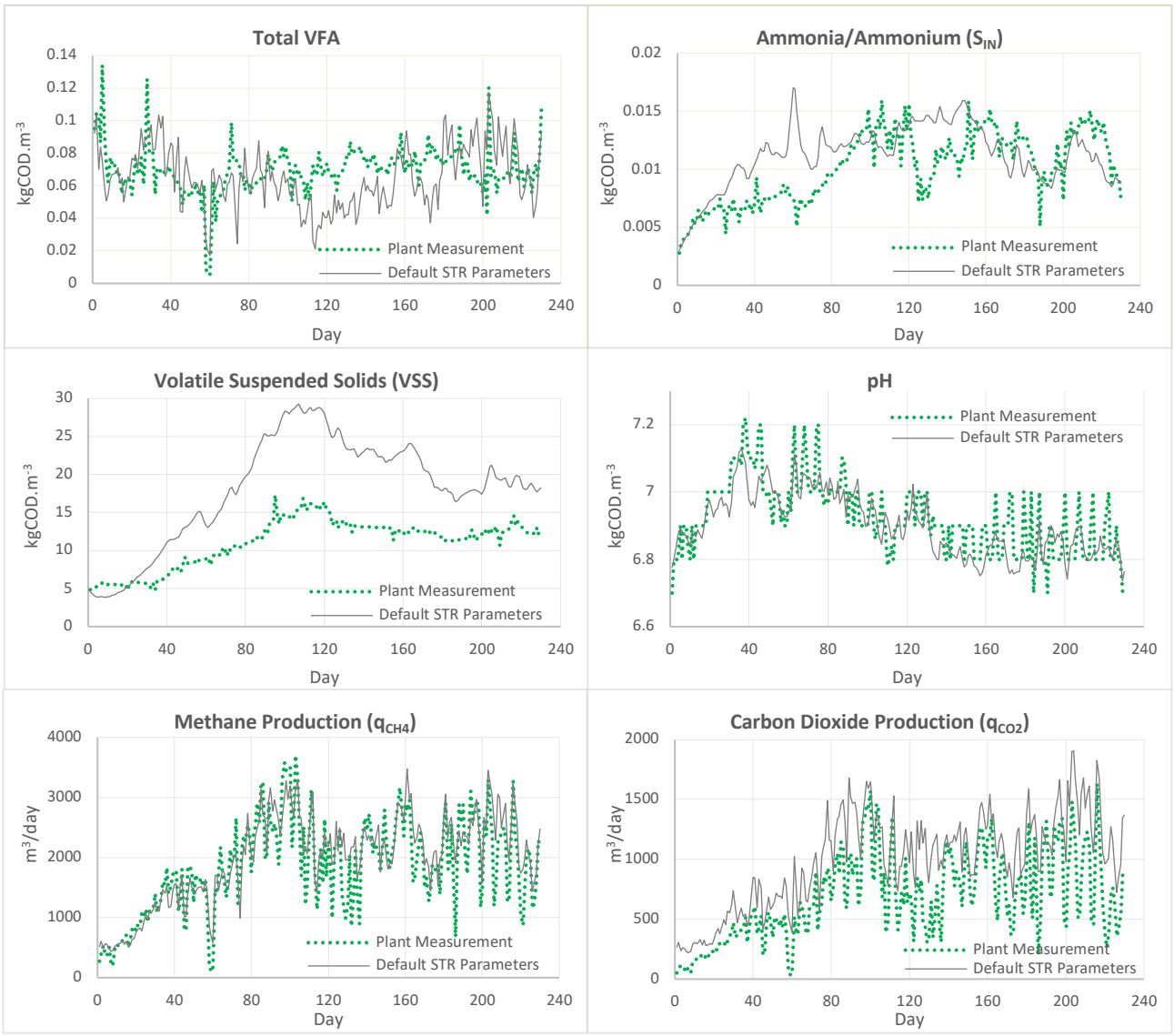


Figure 11: ADM1 simulation using default parameters. Projected values are represented in grey lines and actual plant measurements are plotted as green dots.

This observation is possibly a consequence of applying ADM1 for an influent with highly variable substrate composition. The fact that a *fixed* compositional ratio is used for translating particulate COD concentration into the ADM1 state variables representing carbohydrates, proteins and fats, means that discrepancy is inevitable when the substrate's composition ratio changes. In reality, such as in this scenario, the substrate composition is dynamically *variable.* How the wastewater is constituted depends on the factory's activities (types of CIP's, section cleaned, etc.) and the timing of whey addition. Since ammonia production is closely linked to the degradation of protein content in the substrate, a wastewater stream with low whey addition could result in an overestimation of protein, and consequently an overestimated ammonia/ammonium response (e.g. Day 30 – 100).

Volatile suspended solids (VSS) concentration was grossly overestimated. However, considering the fact that the projection has a close resemblance in its trend, the poor fit is likely a result of incorrect kinetic parameters related to organic particulate degradation such as hydrolysis, biomass growth or biomass decay. Calibration is therefore expected to target these parameters.

pH, VFA and methane gas flow (q_{CH4}) exhibited relatively good fits even though parameters were uncalibrated. Good pH fit is expected because, as pointed out by Donoso-Bravo *et al.* (2011), pH will remain stable in a well-buffered digester. Since the digester possesses a long hydraulic retention time (+/- 2 days), most of the alkalinity produced during methanisation (i.e. in the form of bicarbonates and ammonium) is retained. Furthermore, the predominantly alkaline substrate guarantees a consistent alkalinity buffer.

Conversion to methane from the three components (acetate, carbon dioxide and hydrogen) is stoichiometric based. The production rate is thus relative to the concentrations of these components, which are dependent on the influent's composition and degradation kinetics. In other words, how COD is fractionated into ADM1 state variables as well as the calibrated model parameters are key influencing factors. Nevertheless, it is still possible to obtain fairly accurate prediction of methane gas production using uncalibrated parameters, because regardless of how COD is fractionated, all degradable COD will participate unrestrictedly in the methanogenesis reactions when no inhibition is in effect.

Simulated carbon dioxide flow (q_{CO2}) was consistently higher than the plant measurement despite having a similar trend. This outcome could be a result of a lower than actual pH prediction and/or over-prediction of inorganic carbon (S_{IC}).

4.2. Sensitivity Analysis using Monte Carlo and PLSR

4.2.1. Monte Carlo Simulation

Model uncertainties were propagated through a series of Monte Carlo (MC) runs. During each MC run, ADM1 was executed using a randomised set of parameters. In order to understand the relationship between uncertainty with respect to the number of MC runs, simulations were carried out in batches of 250, 500, 1000 and 1500 MC runs. A plot for 500 MC is composed of 500 lines, each of which represents the output that corresponds to a particular random parameter set. The extent of uncertainty in each output is expressed by its variance, or graphically speaking, higher uncertainty is portrayed in the form of a wider *uncertainty band* (i.e.

spread across the vertical axis). For ease of viewing, only the mean and percentiles are plotted in Figure 12 and Figure 13.

VFA has the widest uncertainty band among all outputs where the difference between the upper bounds and lower bounds could be a factor of as high as 1000 (cf. Figure 12 and Figure 13). This outcome is expected given the fact that the "VFA" term is a sum of all volatile fatty acid constituents (acetate, propionate, etc.) which collectively are influenced, directly or indirectly, by a majority of biochemical reactions. Evolution of these constituents' concentrations is thus related to most ADM1 parameters. Kinetic rates of preceding degradation processes (disintegration, hydrolysis, acidogenesis and biomass decay) influence VFA accumulation, whilst parameters associated with the uptake kinetics affect VFA consumption.

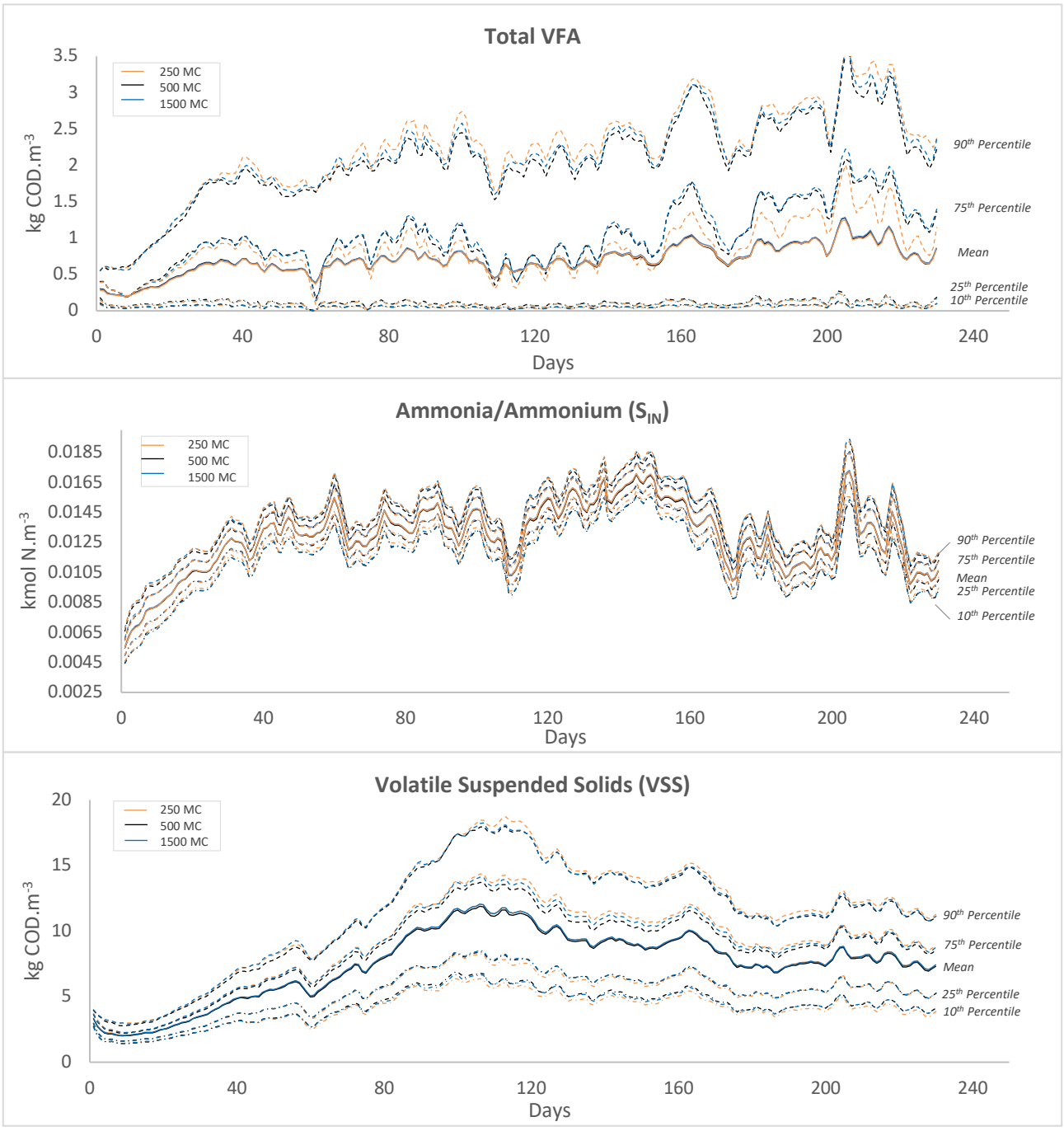


Figure 12: Monte Carlo results for VFA, S_{IN} and VSS at 250, 500 and 1500 Monte Carlo runs. The uncertainty band is represented using mean, 10th, 25th, 75th and 90th percentile values.

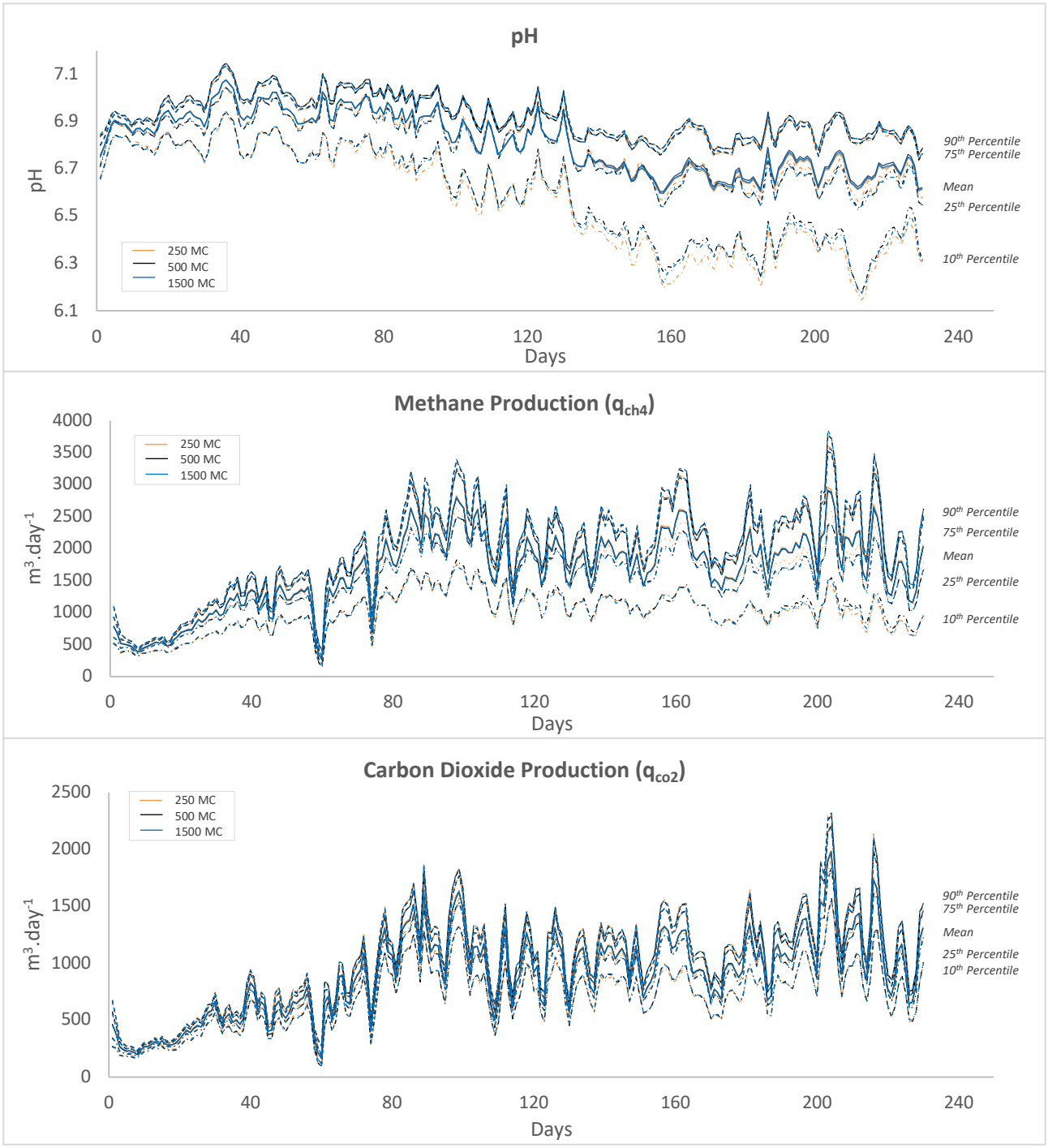


Figure 13: Monte Carlo results for pH, CH_4 and CO_2 at 250, 500 and 1500 Monte Carlo runs. The uncertainty band is represented using mean, 10th, 25th, 75th and 90th percentile values.

It is interesting to note that the VFA line representing the mean of the MC simulations is a magnitude higher than the actual VFA measurements (~0.5 vs ~0.05). Additionally, the large gap between the 90th percentile and 75th percentile suggests that the parameter ranges tested during MC simulations can result in extreme uncertainties with a tendency towards VFA accumulation. This observation implies that the digester is likely not inhibited.

There were several cases of complete system failures during MC simulation. Methane gas production was seen to cease after around Day 100, during which pH drops below the threshold at which hydrogen inhibition

on methanogeneis becomes effective. Elevated hydrogen concentration (S_{H2}) were also detected (Appendix, Figure 31). Inhibition signifies that VFA uptake rate is retarded and explains the build-up of VFA.

It is not practical to run MC simulations for all possible combinations due to a large number of parameters. Therefore, it is important to understand how many MC runs are necessary in order to capture the trends required for PLSR. From the above graphs, 250 MC runs were noted to have percentiles different from those of higher MC runs, whereas 500 MC runs showed a closer statistical resemblance to 1500 MC runs. Convergence continues as the number of MC runs increase. This observation suggests that it may be redundant to execute a high number of MC runs unless calibration time is not a critical consideration for the modeller.

Another point to consider, however, is the fact that parameters are randomised stochastically. It is possible to simulate a batch of 500 MC runs without generating an exact randomised parameter set as in a 1000 MC batch, even though both scenarios produce similar-looking uncertainty bands. From a statistical perspective, redundancy is good because too few MC runs may risk poor data quality and result in inaccurate PLSR evaluation. Modellers are encouraged to strike a balance between data quality and redundancy.

The minimum and maximum values at each time interval are important for the purpose of data normalisation. Although it is evident from the figures that fairly similar minimum and maximum values could be propagated given a different number of MC runs, it may not always be true because each parameter set is unique. Hence, it is advisable to run an adequate number of MC runs so that the *absolute* minimum and maximum values can be identified.

Refer to the Appendix, Section 8.5.2, for the Monte Carlo plots of all 26 state variables.

4.2.2. Outlier Removal

Outputs from the Monte Carlo simulation are stored in an “output” matrix. First step of the PLSR algorithm involves normalising this matrix; however, considering the presence of abnormally high and low values (as a result of system failures), it was presumed that normalisation without outlier removal could produce a skewed data set. A simple outlier classification technique based on the IQR (Interquartile Range) was applied in this study. The conventional rule states that any data point with a value below Q1 minus 1.5 x IQR, or above Q3 plus 1.5 x IQR, shall be classified as an outlier. Q1 and Q3 refer to the 1st quartiles and 3rd quartiles of a data set respectively.

Although it is acknowledged that all data points may contain valuable information, the influence of outliers on PLSR and hence the model fitting performance is unknown. Different degrees of outlier removal was therefore investigated in this thesis by testing various outlier classification thresholds: 1.5 x IQR, 1.0 x IQR and 3.0 x IQR. The objective is to verify whether outlier removal is necessary; if so, what is the outlier removal extent required.

Figure 14 and Figure 15 demonstrate how the uncertainty bands change when outliers are removed. As tighter outlier classification thresholds are applied, more data points are removed, resulting in smaller uncertainty bands. The substantial shrinking of the uncertainty bands from their original bounds confirmed that MC

simulations do yield extremities. These extreme values are unidirectional – either upper bound (e.g. VFA and VSS) or lower bound (e.g. S_{IN}, pH, q_{CO2} and q_{CH4}). This trend corresponds to the signs when a digester fails. Biogas production ceases due to inhibited methanogens, which results in VFA accumulation and consequent pH drop.

Increasing the number of MC runs has limited impact on the Monte Carlo data after outliers are removed. As shown in Figure 32 – Figure 35 in the Appendix, similar results were produced for MC above 500 runs while the graph representing 250 runs displayed some deviations. Normalisation of the "output" matrix is thus independent of the number of MC runs provided that a sufficiently high number of MC runs is selected.

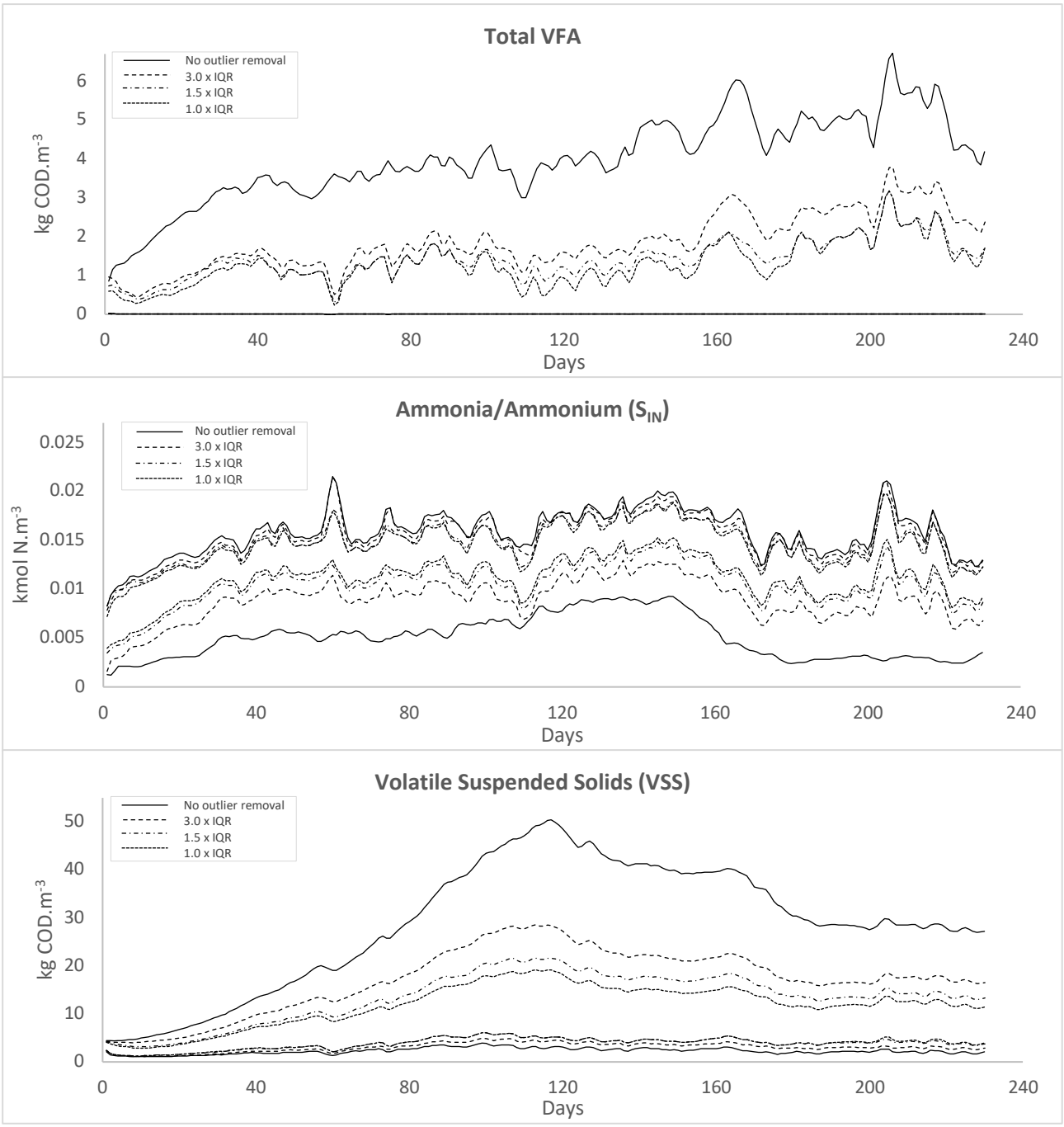


Figure 14: Maximum and minimum bounds of 1500 Monte Carlo runs for VFA, S_{IN} & VSS before and after outlier removal

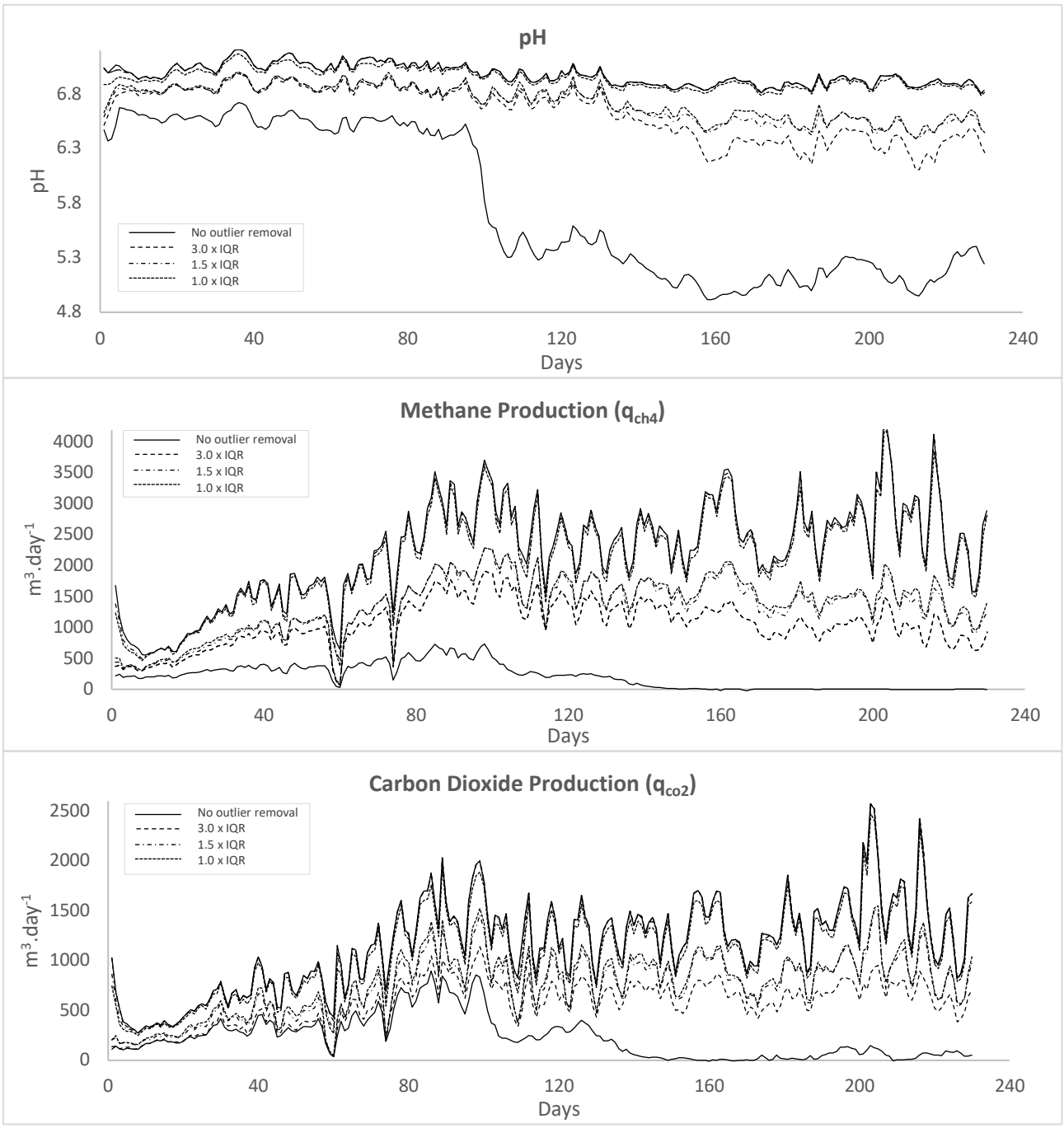


Figure 15: Maximum and minimum bounds of 1500 Monte Carlo runs for pH, q_{ch4} & q_{co2} before and after outlier removal

4.2.3. PLSR Evaluation

Figure 16 shows the evolution of the output loading vector (**q**) for the *first* latent variable. Note that in this elaboration the output matrix is pre-treated with outlier removal based on 1.5 x IQR. Loading vectors for the 6 outputs are presented in percentages relative to their weight contribution towards the output latent variable (**U**). While VFA, S_{IN}, and VSS exhibited increasing weighting, the remaining outputs pH, q_{CH4} and q_{CO2} have decreasing influence over time. An interesting observation to note is that as the number of MC runs increase, the lines tend to be closer. This suggests that a threshold number of runs exist, beyond which the change in loading vectors become insignificant.

According to the loading vector distribution, VFA displayed the highest weighting towards the latent output variable. However, as previously noted, VFA's uncertainty band possessed the most extreme variances. In that respect, normalisation is likely to produce very small normalised values unless the values are close to the upper/lower bounds. Since the latent output variable is calculated as a linear combination of normalised outputs and **q** loading vector, the *effective* weighting, or contribution, of VFA towards **U** is actually insignificant. In contrast, VSS and S_{IN} have a tighter sensitivity band and relatively high weighting. As such, these two components are expected to have more influence than VFA on **U**.

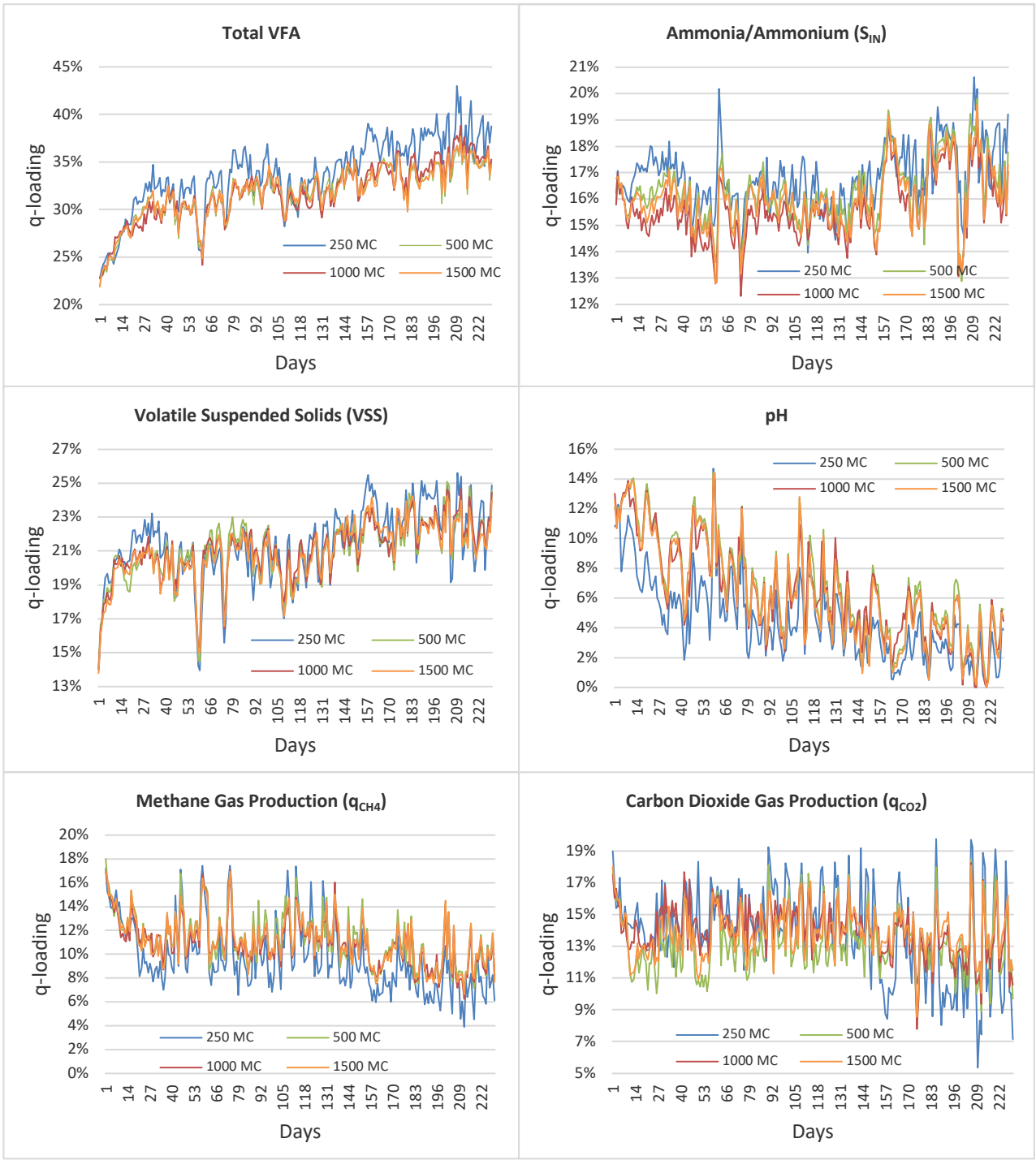


*Figure 16: Evolution of output loading vector (**q**) for the 1st latent variable*

4.3. Model Optimisation

The model objective function (U_{diff}) is defined as the combined differences in the output latent variables (**U**) before and after parameter calibration. A reduced U_{diff} value resembles an improvement in the overall weighted fit accuracy of the 6 outputs. Calibration simply involves adjusting the scaling vector (**λ**) iteratively until the objective function can no longer be minimised further. At this point, the model is deemed optimised. Goodness-of-fit for each output variable is evaluated in terms of RMSE. Note that in this thesis RMSE serves only as a comparative measure and does not form part of the objective function criterion.

Figure 14 shows two calibrated models. The first example, labelled as "PLS Method (1)" on the graph, utilised PLSR vectors that were constructed using 1500 Monte Carlo runs without outlier removal and two latent variables. The calibration process started with **λ** = [1,1] and ended as [0.035, 1.657] after 105 calibration runs. In relation to the default simulation (grey line), the objective function reduced from a value of 182 to 116 while a substantial improvement in the RMSE of VSS was achieved.

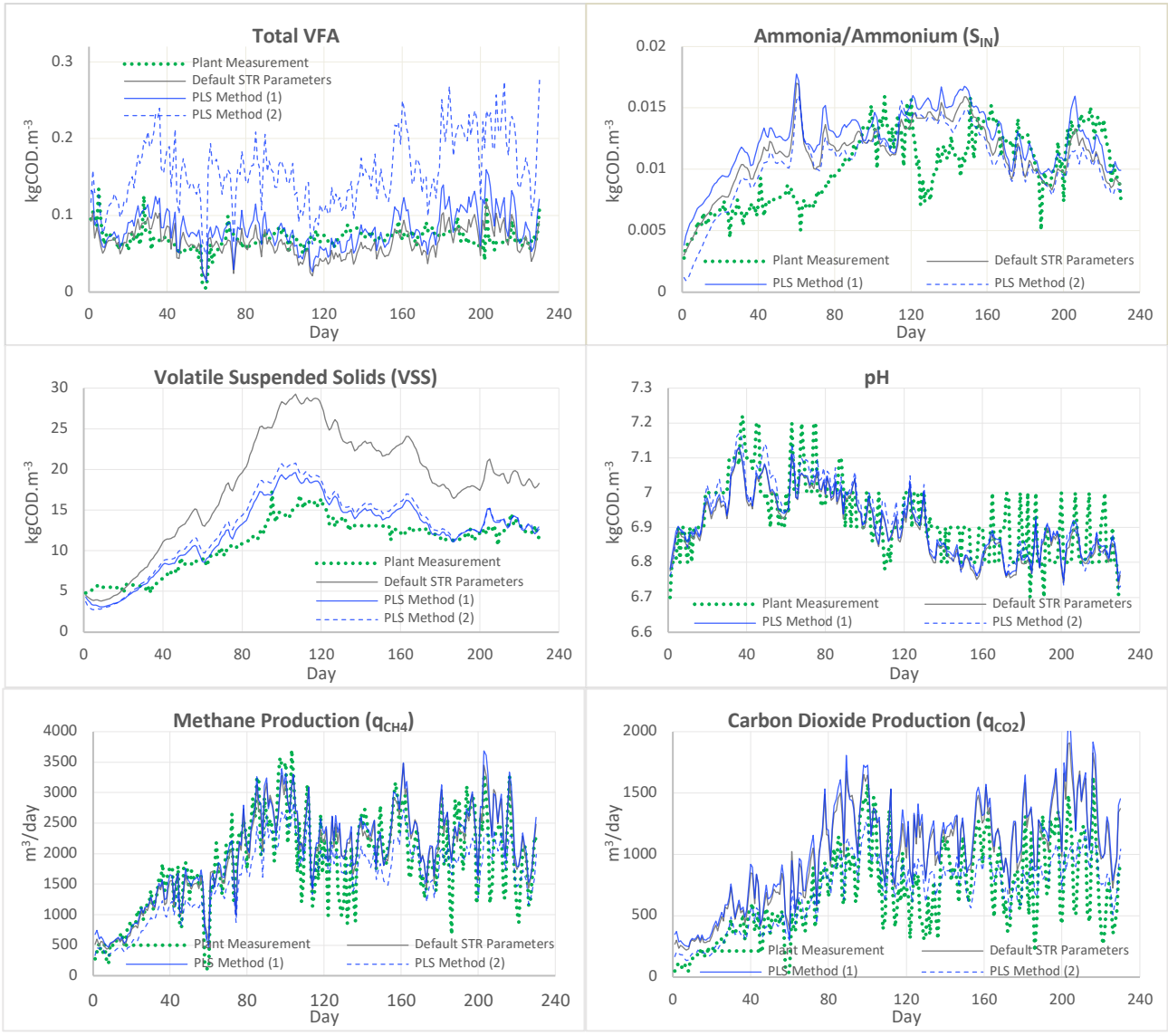


Figure 17: Two ADM1 simulations with similar objective function after calibrated using the PLS Method. Green lines represent actual plant measurements; grey lines represent simulation before calibration; solid blue lines represent Method (1)is based on 1500 Monte Carlo runs, no outlier removal and two latent variables; dotted blue lines represent Method (2) which is based on 1500 Monte Carlo runs with outlier removal and 4 latent variables

Yet, besides VSS and pH, other output variables ended with higher RMSE values after calibration. This outcome correlates with the fact that VSS has the highest effective weighting towards the value of U. In other words, as long as the VSS model fit improves, U_{diff} could still achieve a significant reduction, even if fitting accuracies of other components are compromised.

To exemplify the impact of normalisation, "PLS Method (2)" presents a simulation which saw a similar magnitude reduction on the objective function (~39%) after calibration. This example utilised PLSR vectors that were based on 1500 Monte Carlo runs with outlier removal (1.5 x IQR) and four latent variables. It can be seen that VFA has a much poorer fit than the previous example. VFA tends to compromise its fit accuracy in favour for other outputs with higher effective weighting because its influence on the objective function is relatively insignificant. Extreme upper and lower bounds in combination with higher mean uncertainty (0.5-1.0; Figure 12) than the range of interest (0.05-0.1; Figure 17), explains why the normalised VFA values are inevitably small. This is especially true since the starting calibration parameter set is the default STR parameters which already has a reasonably acceptable fit.

In addition to a very poor VFA fitment, VSS, pH and q_{CH4} also fitted worse in Method (2). Despite so, similar objective function reduction to Method (1) was still achieved due to superior accuracy in the S_{IN} and q_{CO2} models. These two outputs boast small uncertainty bands (Figure 12; Figure 13) and high q-loading vector distributions (Figure 16), as such they are the second and third most effective weighting outputs on the objective function following VSS. A substantial improvement in q_{CO2} fitment hence directly translate to a significant objective function reduction.

4.4. Effect of Outlier Removal & Number of Latent Variables on Model Fitting

According to Geladi & Kowalski (1986), the number of latent variables selected is critical because nonlinearities in non-linear models can only be described by assigning multiple latent variables. This section aims to establish the relationship between outliers and the number of latent variables with respect to model fit accuracy.

Results plotted in Figure 18 showed that fitting performance fare worse as more outliers are removed. This outcome suggests that data generated from the Monte Carlo simulation that would normally qualify as outliers could in fact possess information critical to the development of the PLS interaction constructs. Lesser outlier removal further enables the use of fewer latent variables without expending more fitting time. For instance, outlier removal thresholds based on IQR multipliers of 1.0 and 1.5 required four latent variables in order to attain similar objective function improvement. However, these scenarios are disregarded due to signs of overfitting (Figure 19) where VSS and q_{CO2} fits exceptionally well but at the expense of VFA. Outlier removal is therefore concluded to be an unnecessary procedure prior to applying the PLSR algorithm.

The tendency for lower weighted outputs to become influenced during calibration is higher when more latent variables are introduced. Latent variables can thus be viewed as "sensitivity controls". Modeller's discretion must be exercised though, because more latent variables tend to increase overfitting risk and also increase the number of iterations required to complete the calibration.

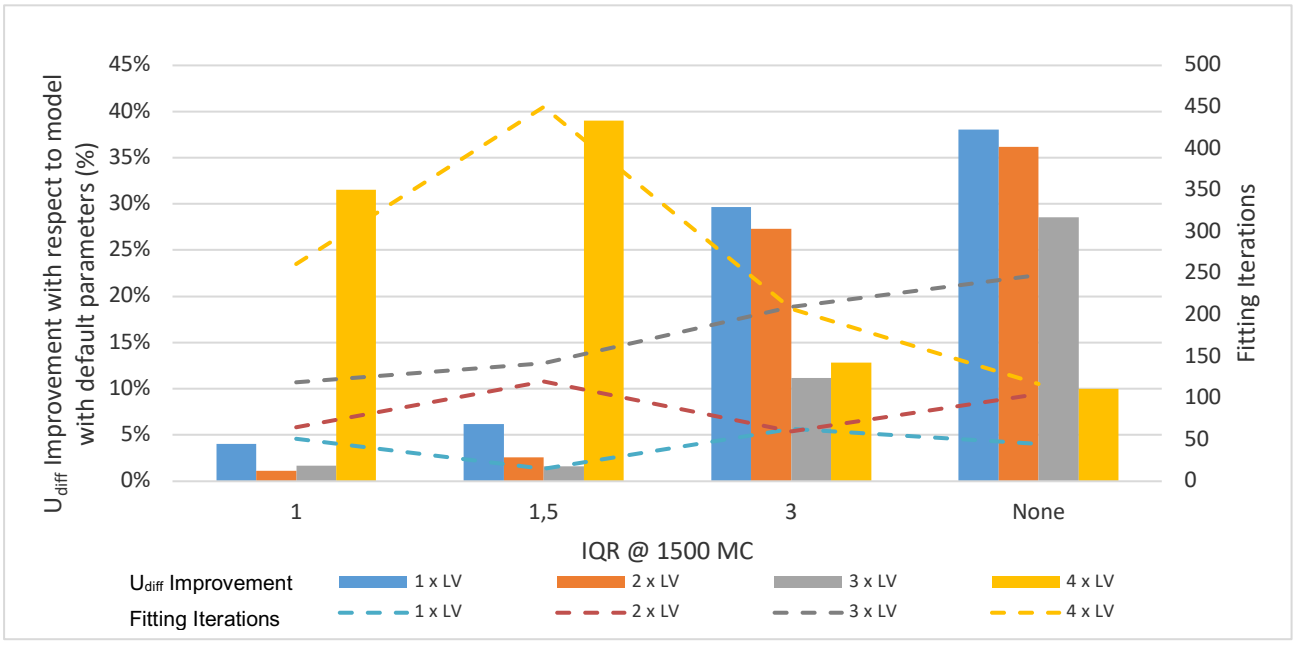


Figure 18: Model fitting performance at various IQR (extent of outlier removal) and number of latent variables

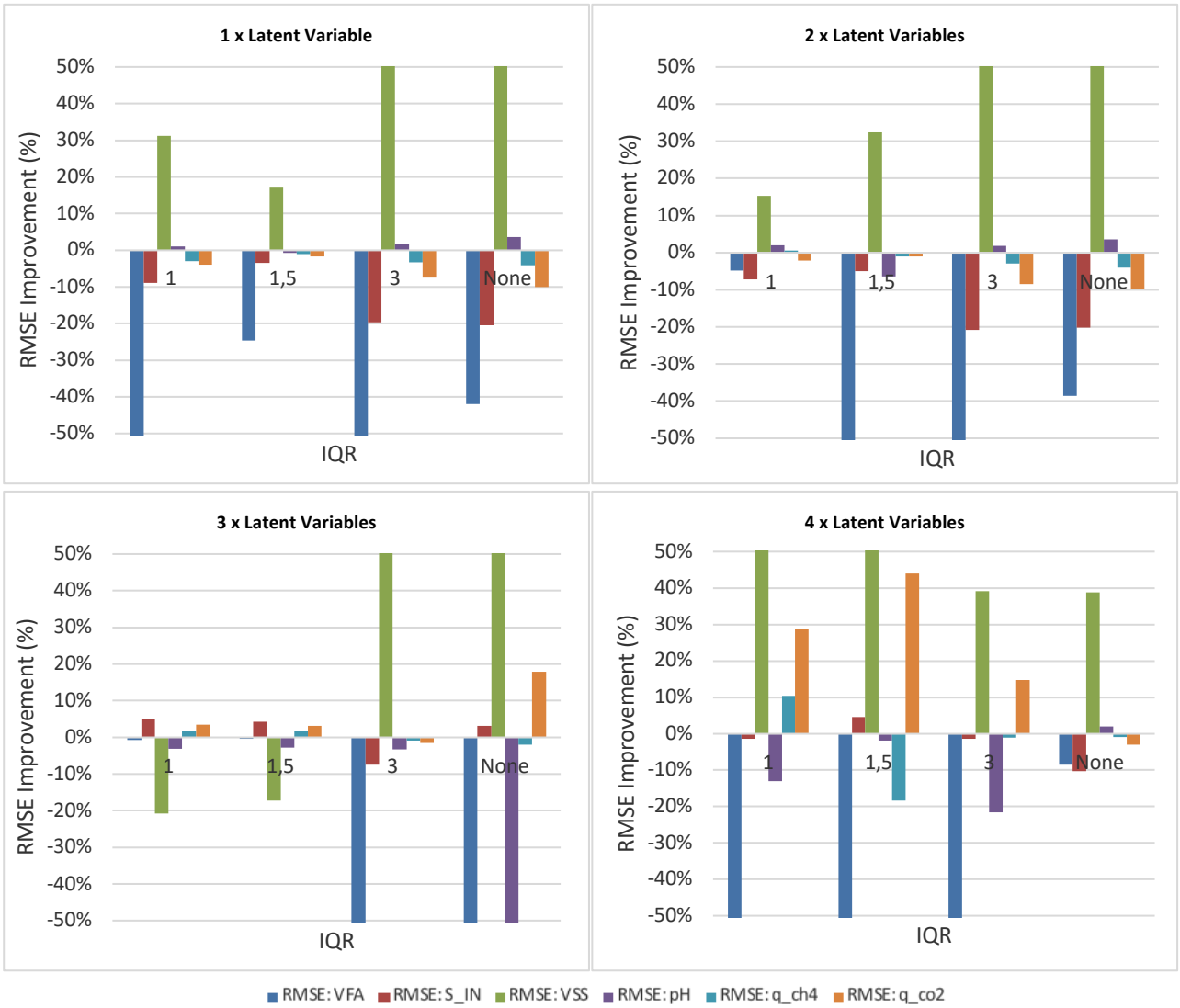


Figure 19: Output RMSE at various outlier removal and number of latent variables

4.5. Conclusion

The PLS Method has demonstrated to be a viable parameter calibration method for ADM1 optimisation. The number of Monte Carlo runs and latent variables chosen is important factors that could influence the calibration outcome. A high number of Monte Carlo runs is welcomed as it produces data redundancy which allows PLSR to identify underlying trends more accurately. In addition, data normalisation relies on the upper and lower bounds and mean values from the Monte Carlo data. 500 or more Monte Carlo runs were found to be statistically adequate for normalisation purpose, but since parameter randomisation is done stochastically, it is advisable to simulate as many as practically possible such that all major uncertainty trends are captured.

The outlier removal procedure is unnecessary even though Monte Carlo produces extreme values. In fact, all data serve as valuable information for PLSR. Utilising too many latent variables may induce overfitting, while too few latent variables could leave lower weighted outputs unaffected. It is advisable to start with one latent variable and increase only if necessary. In this thesis, two latent variables were found to produce the best result.

CHAPTER 5

BENCHMARKING AGAINST OTHER PARAMETER CALIBRATION METHODS

5.1. Results: Model Fit Accuracy

5.1.1. Total VFA

Figure 20 presents the VFA simulations produced by the 4 methods alongside an uncalibrated model which utilises default parameters. A residual plot is plotted adjacent to allow easy interpretation of the model accuracy at each simulation interval. High volatility is evident across all methods, with the PLS method consistently over-predicts and both Group methods under-predict. Even though the Group Methods exhibited superior MAPE scores (i.e. lower residuals) during calibration, they failed to continue the trend than during the validation period. This attributes to the fact that the Group Methods are more prone to overfitting.

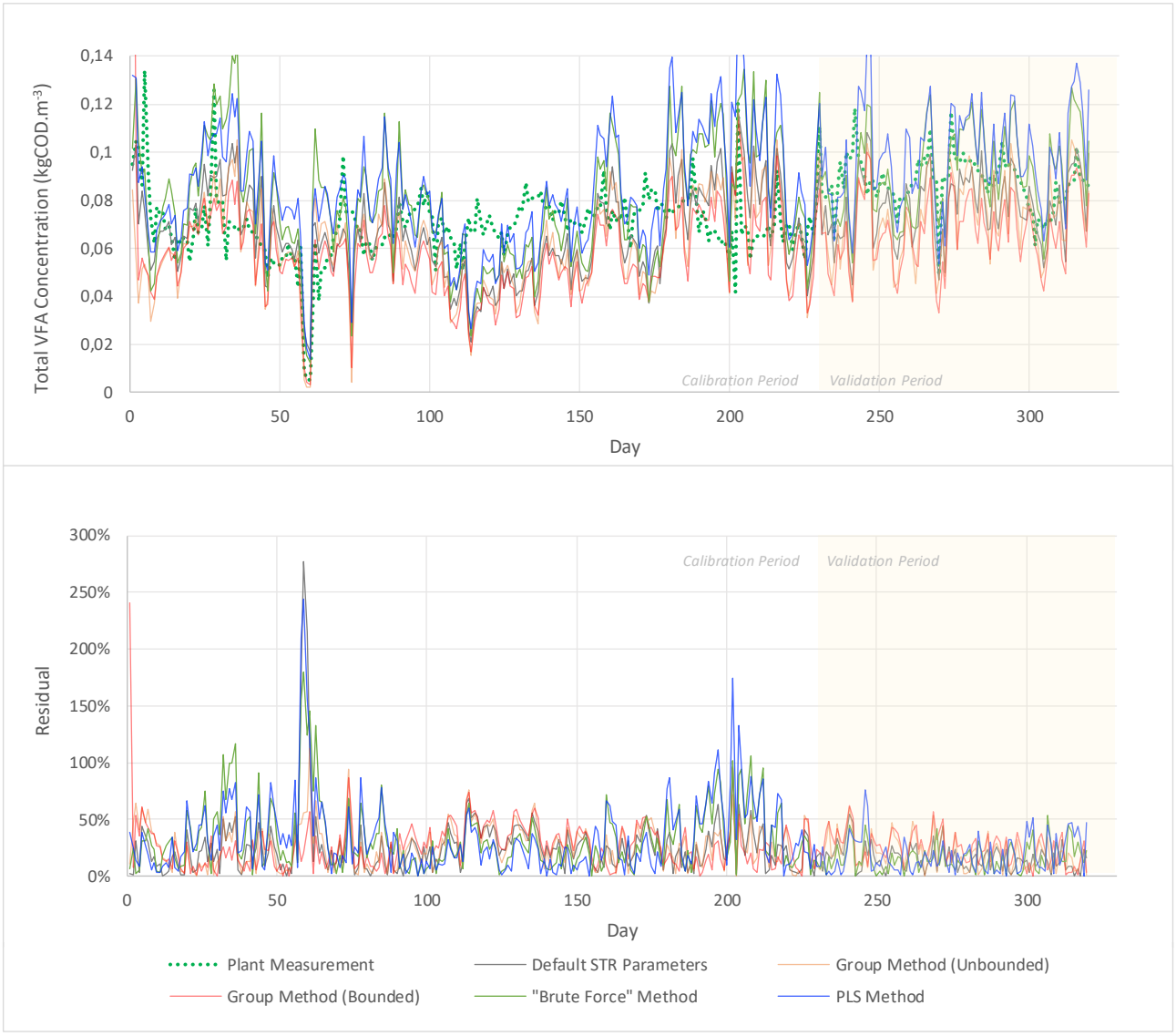


Figure 20: Graphical comparison between various parameter calibration methods and residual error plot - Total VFA

During the calibration process, all methods were seen to prioritise on improving the fit accuracy of VSS instead of VFA (Figure 22). The reason is that the uncalibrated model for VSS has the worst fit out of all 6 outputs and any marginal improvement brings about the largest influence on the objective function. In exchange for better VSS fit accuracy, all methods concluded with poorer MAPE scores for VFA fitment after calibration.

Since the Group Method (Unbounded) is unrestricted during calibration, the disintegration and hydrolysis rates became over-exaggerated in attempt to reduce the VSS concentration. The increase in particulate breakdown rates corresponds to an increase in the production of VFA. To counteract against the rise in VFA concentration, Monod uptake rate ($k_{m,ac}$) and the half-saturation value of acetate ($K_{s,ac}$) were the only two parameters that could be calibrated in the first group of parameters which relate to VFA; hence these parameters were calibrated to very high values. Then, during the second group calibration, the decay rate of monosaccharide degraders ($k_{dec,su}$) was modified to an exceptionally high value which accelerates the death of this specific degraders. The breakdown of monosaccharides into VFA is consequently throttled due to a lack of degraders. A build-up of monosaccharides (S_{su}) is expected and could reach levels that do not correlate with reality. Verification is not possible because no experimental data is available.

A similar scenario for Group Method (Bounded) was noted. However, in this case, the maximum values for disintegration and hydrolysis rates are capped. As such, VFA production in response to particulate breakdown is not as excessive in comparison. This explains why kinetic parameters for acetate ($k_{m,ac}$ and $K_{s,ac}$) were left uncalibrated. $k_{dec,su}$ is again the main manipulated parameter in the second sensitivity group and maxed out against its upper cap.

The “Brute Force” Method attempted to improve VSS model fitment through several mechanisms. Firstly, hydrolysis rates for carbohydrates and lipids are maxed out; and secondly, the yield constants of most degraders were decreased. Both factors attribute to the reduction in VSS concentration. Elevated hydrolysis rates also implied quicker formation of monosaccharides (S_{su}), amino acids (S_{aa}) and LCFA (S_{fa}) than the uncalibrated model. The subsequent conversion to VFA should be rapid, but maxed out decay rates of all acidogens, acidogensis is throttled and these three components are expected to accumulate. Decay rate of acetate is also maxed out. This explains why VFA concentration is relative high while methane production is the lowest out of all methods (Figure 24).

Similar to other methods, the PLS Method elevated the values of hydrolysis and biomass decay parameters (Table 15) to lower VSS concentration. An increase in the stoichiometric parameter $f_{CH,XC}$ was noted. This adjustment has allowed composite particulates (x_c), which are formed from biomass lysis, to fractionate predominantly into carbohydrates instead of other forms of organic particulates. Combined with an increase in both hydrolysis rate (k_{hyd_ch}) and Monod uptake rate of sugar ($k_{m,su}$), the rate at which simpler short-chain VFA (such as acetate and propionate) and hydrogen are produced is expected to be higher. Thus, as seen in Figure 14, VFA concentration is higher than that of the uncalibrated model.

5.1.2. Ammonia/Ammonium (S_{IN})

From Figure 21, it can be seen that all 4 methods produce identical ammonium trends as the uncalibrated model, despite each method having significantly different parameters. Model fitting is grossly overestimated during plant start-up (transient conditions) but improved when transitioning past Day 150 (steady-state conditions) and into the validation period. As explained in Section 4.1, this particular wastewater does not have a fixed protein content. In fact, to improve the model's accuracy, substrate characterisation accuracy is regarded as more critical than parameter calibration. Nonetheless, a distinction between different methods exists. Group Methods performed worse than other methods because of its over-exaggerated protein hydrolysis rate. Although the "Brute Force" Method also has high protein hydrolysis rate, the breakdown of amino acids (and hence ammonium production) is restricted by the high decay rate of amino acids biomass.

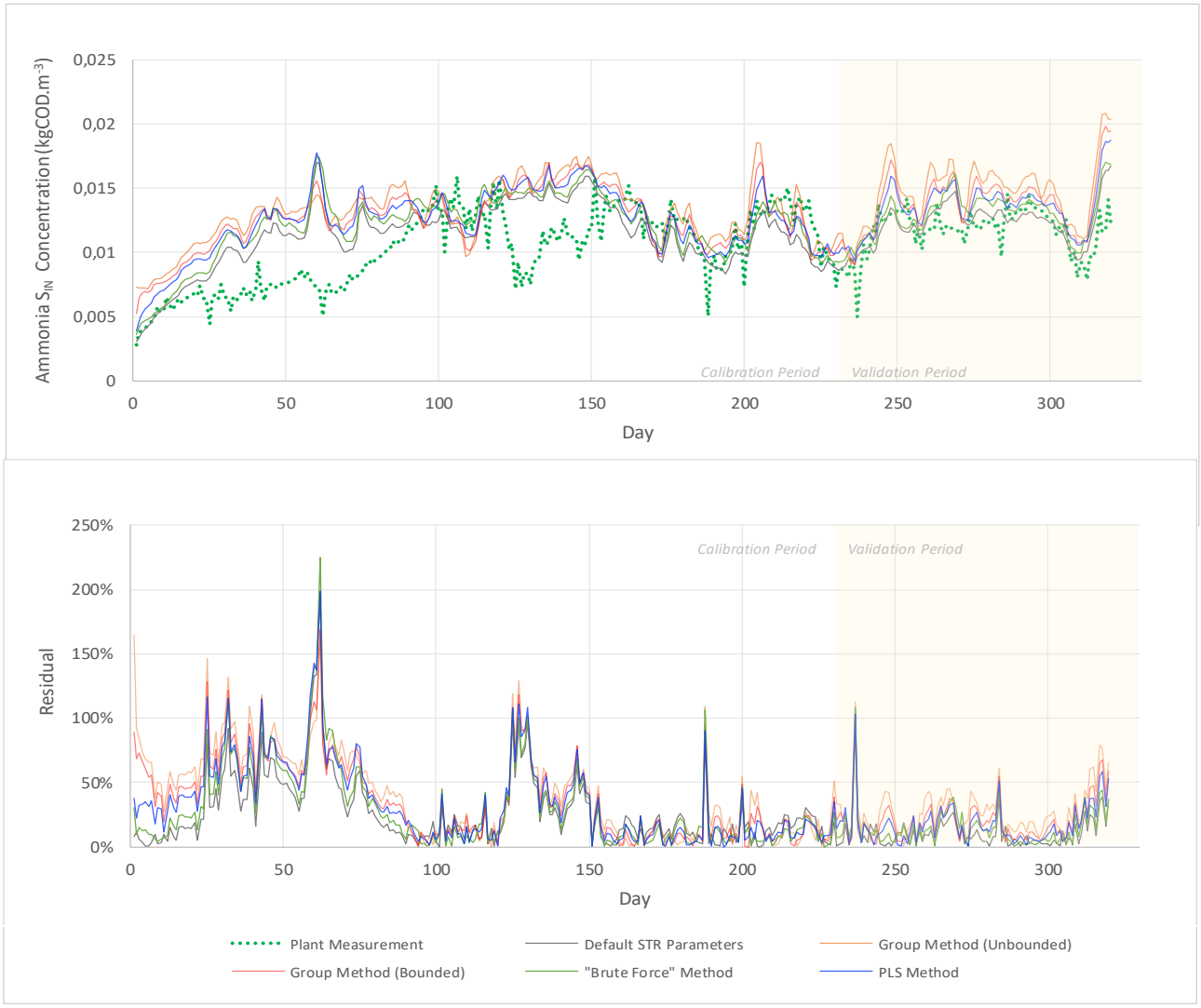


Figure 21: Graphical comparison between various model optimisation methods and residual error plot – Ammonia

5.1.3. Volatile Suspended Solids (VSS)

All 4 methods succeeded in improving the model fitting accuracy of VSS (Figure 22). Apart from the PLS method, other methods saw an increase in MAPE score (i.e. higher residuals) when transitioning into the validation period despite having superior MAPE scores during calibration (Figure 28). This occurrence suggests that these methods may have overfitted.

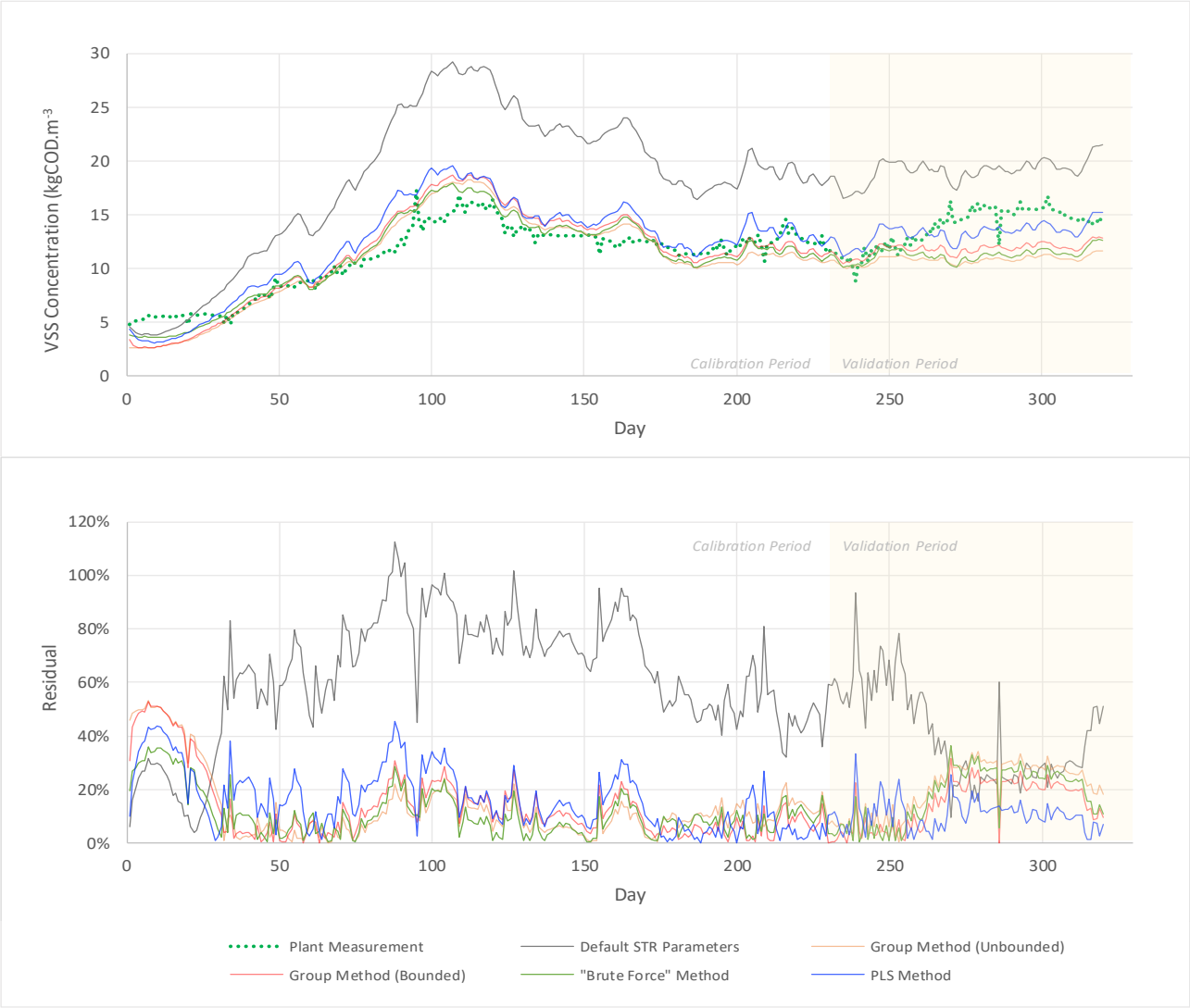


Figure 22: Graphical comparison between various model optimisation methods and residual error plot - VSS

5.1.4. pH

pH is the most accurately modelled output. All methods, including the uncalibrated model, showed a low level of residuals consistently across both periods. It is evident from Figure 23 that the Group Method (Unbounded) and the “Brute Force” Method have overestimated pH projections. Since pH is indirectly correlated to the soluble carbon dioxide concentration, lower carbon dioxide production in the models produced by these two methods explains why pH is higher. Carbon dioxide is formed when monosaccharides, amino acids and propionate are degraded. However, as described in Section 5.1.1, parameters were calibrated in such a way that the metabolism of monosaccharides and amino acids are slowed down tremendously. Consequently, the formation of carbon dioxide and propionate is lowered.

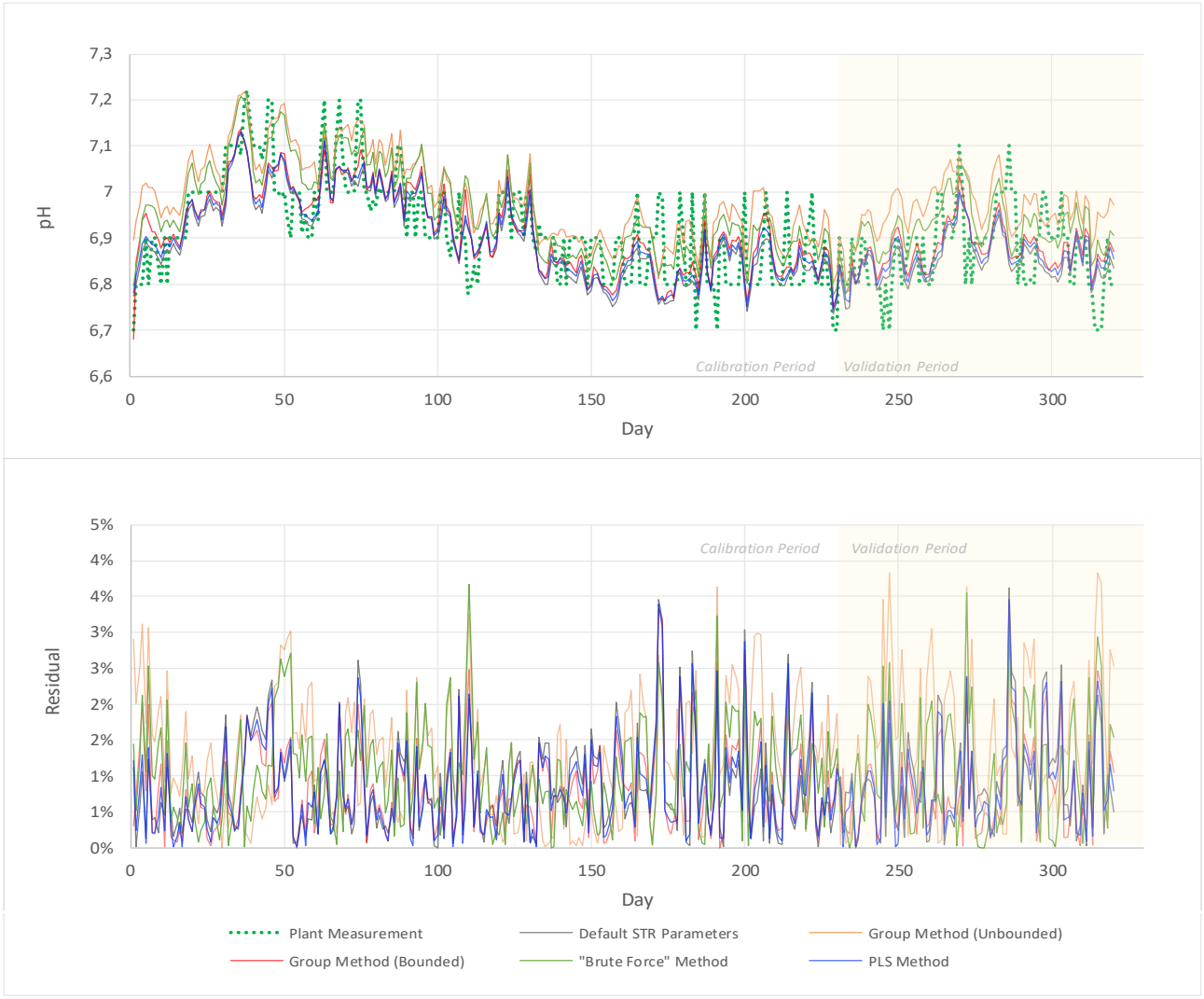


Figure 23: Graphical comparison between various model optimisation methods and residual error plot - pH

5.1.5. Methane (q_{CH4}) & Carbon Dioxide Production (q_{CO2})

Calibrated models produced by the Group Method (Bounded) and PLS Method were found to project higher biogas volumes than the uncalibrated model. For the PLS Method, higher gas volumes is a result of increased Monod uptake rates for acetate ($k_{m,ac}$) and hydrogen ($k_{m,h2}$), which translates to faster acetoclastic methanogensis and hydrogenotrophic methanogenesis respectively. Another contributing factor to higher gas volumes is the reduced inhibition effect imposed on propionate and acetate uptake. The increase of two inhibition parameters, namely KI_{nh3_ac} and KI_{h2_pro}, means that methanogenesis and propionate degradation are less inhibited by free ammonia and hydrogen respectively.

The other two methods yielded opposite results. For the “Brute Force” Method, high methanogens decay rate ($k_{dec,xac}$) in combination with reduced acetate degrader yield (Y_{ac}) after calibration implies that acetoclastic methanogenesis (and hence methane gas production) is partly impeded by biomass deficiency. The root cause for low biogas production, however, is a result of the bottleneck as explained in Section 5.1.1. Without adequate monosaccharides and amino acids degradation, the substrates required for biogas production is limited. Nonetheless, by visual inspection or MAPE scores, both methods displayed superior biogas fitting accuracy.

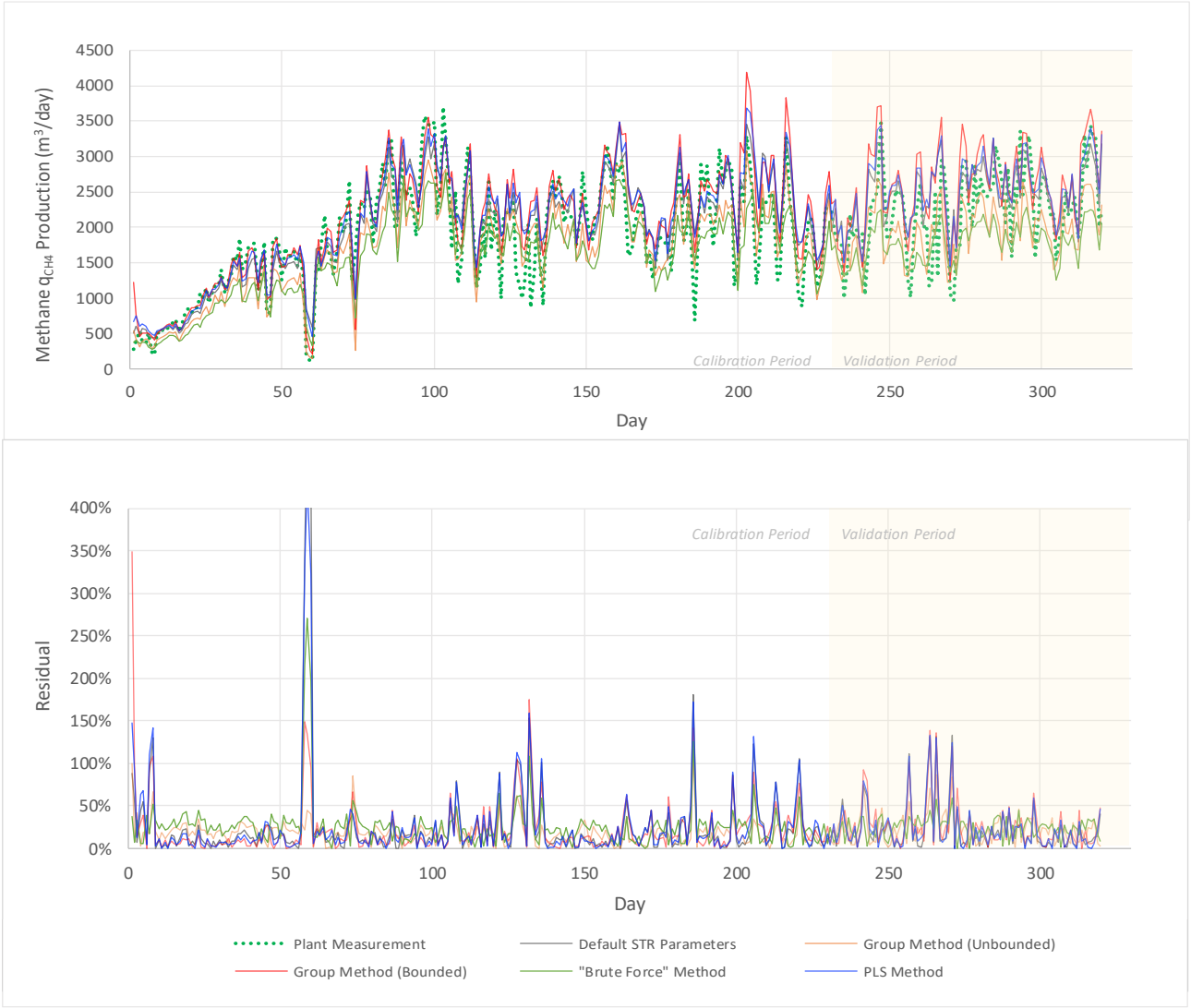


Figure 24: Graphical comparison between various model optimisation methods and residual error plot – CH_4 production

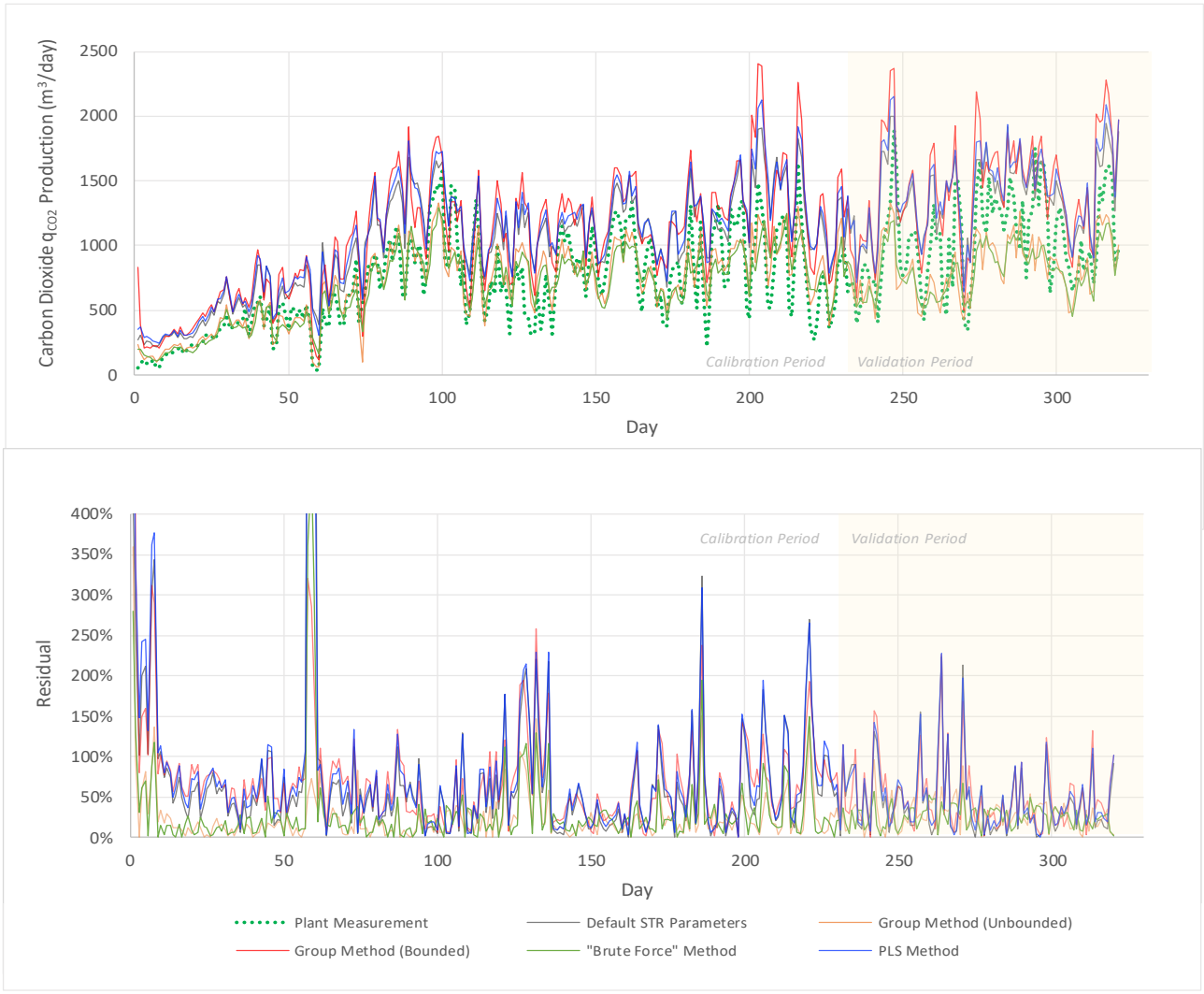


Figure 25: Graphical comparison between various model optimisation methods and residual error plot – CO_2 production

5.2. Results: Parameter Calibration Speed

The number of iterations, or time, taken to complete the calibration differs much between the 4 methods. Figure 26 illustrates how the Grand Average MAPE score evolves with each iteration step of calibration. Grand Average MAPE is defined as the average of all 6 outputs' MAPE values.

Model optimisation using the PLS Method is notably faster, even though it did not achieve a grand average MAPE as low as the other methods. This method arrived at its minimal MAPE of 32.5% within 40 iterations whereas the Group Methods took at least 100 iterations to attain similar MAPE reduction. Meanwhile, the "Brute Force" method required more than 4500 iterations. The shortened duration is attributed to the reduced degrees of freedom where only two latent variables instead of 58 parameters need to be manipulated. It is learnt from previous sections that a lower MAPE score during the calibration period does not necessarily guarantee a better fit when tested against unseen data. As such, the MAPE scores shown on Figure 26 serve only as a quantitative measure of the model accuracy *during calibration*. It is necessary to scrutinise how the scores change from the calibration period to the validation period holistically in order to identify the best method. Another possible reason for PLS Method to exhibit higher grand average MAPE is that its objective function is defined to minimum RMSE instead of MAPE.

Both Group Methods demonstrated that MAPE improves as calibration progresses from a higher sensitivity group to a lower sensitivity group. This stepwise reduction in MAPE has decreasing effectiveness since the room for objective function improvement becomes increasingly limited. When parameters are unbounded, lower MAPE score could be attained but at the expense of more iterations and extreme calibrated values.

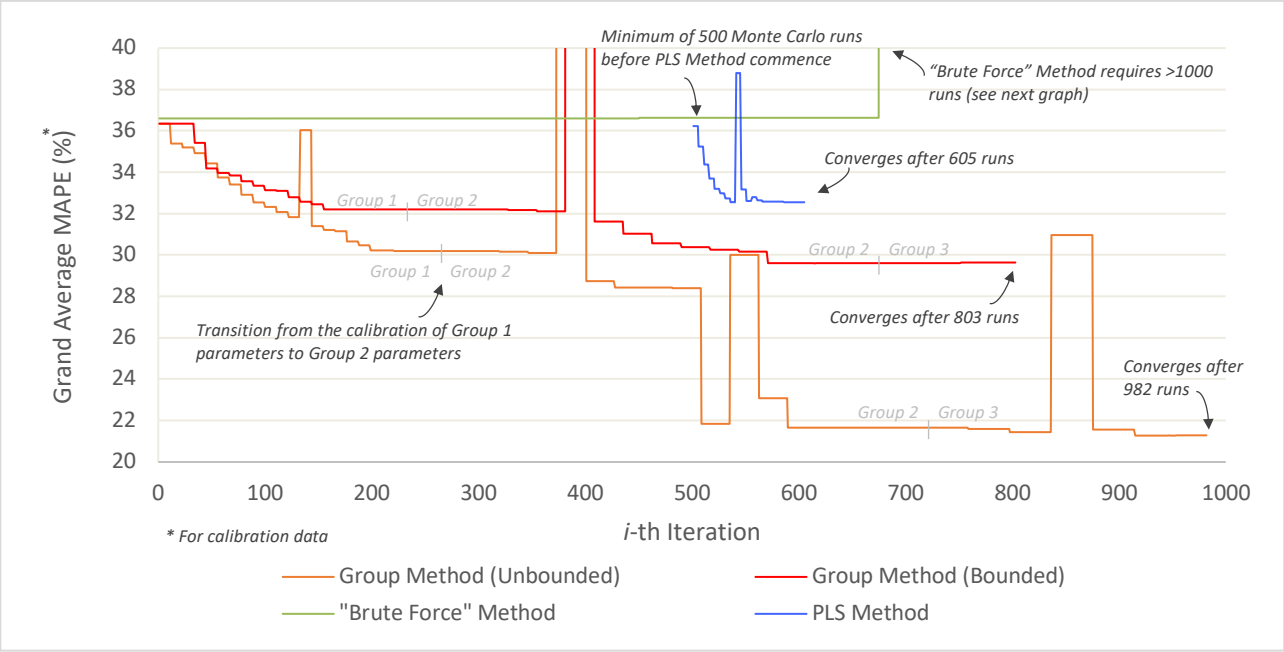


Figure 26: Iterations/time taken by various methods to optimise model i.e. minimise MAPE

The "Brute Force" Method took over 8000 iterations before completing the optimisation (Figure 27). This is due to the large number of parameters, or degrees of freedom, that the Newton algorithm needs to manipulate every time in order to map/identify the next solution step. At approximately 3.5 minutes to compute an iteration,

the “Brute Force” Method took more than 20 days to complete, whereas the PLS Method only took 6 hours. A PC running with Windows 10 64-bit, Intel® i5 processor and 8 gigabytes of RAM was used for the computation.

If the PLS Method were to compare to other methods that involve sensitivity analysis, then the duration taken to generate the Monte Carlo data (i.e. 500 iterations) could be regarded as part of sensitivity analysis, which effectively means the calibration duration is merely 105 runs. The same would also apply if the PLS loadings are reusable for regular calibration.

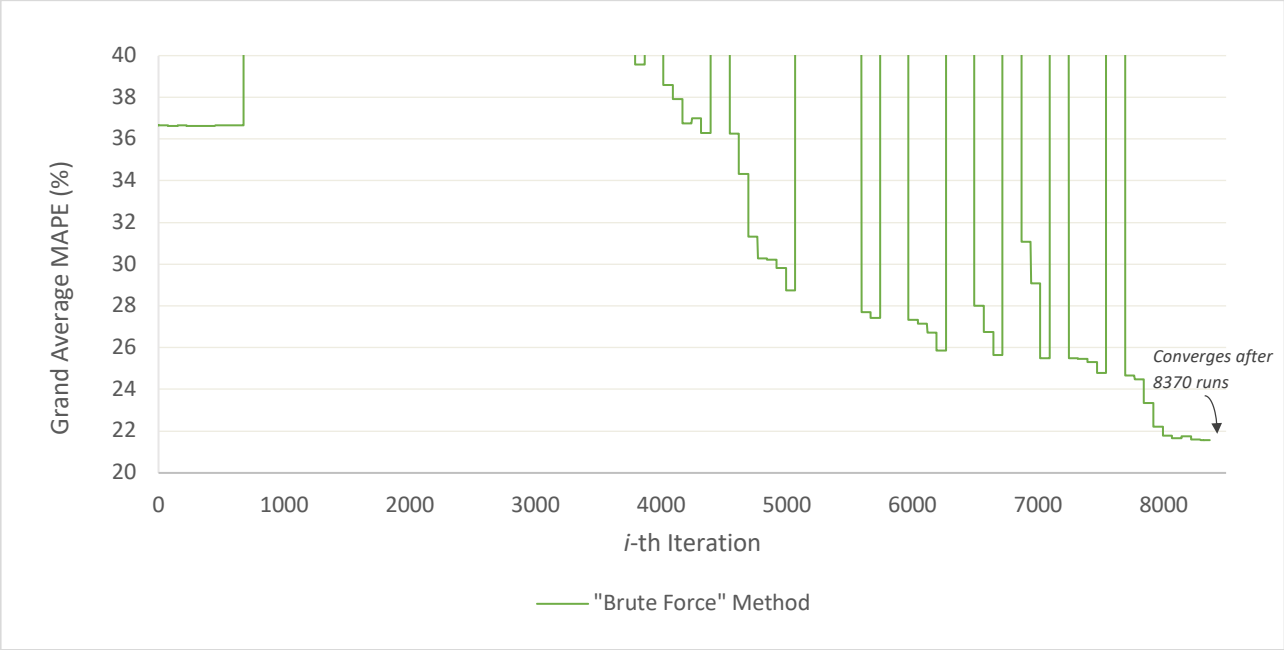


Figure 27: Iterations/time taken by “Brute Force” method to optimise model

5.3. Calibrated Parameters

Table 15 presents a summary of the parameters before and after model calibration. In accordance with the concept behind the PLS Method, *all* parameters underwent some degree of calibration but only a few adjusted more than 30% with respect to the default STR value. Although PLS Method is unbounded during calibration, all of its parameters have remained within the minimum and maximum surveyed values. It is noted that the parameters calibrated by this method does not correlate to the sensitivity groups reported by the STR. In contrast, the Group Methods only calibrate the most sensitive parameters and seems to always over-exaggerate the calibration. The "Brute Force" Method saw most of the parameters undergo significant calibration, whereby some have shown the tendency to exceed the minimum or maximum limits.

Table 15: Calibrated ADM1 parameters produced by the various model optimisation methods. Parameters are colour-coded according to the level of sensitivity reported by the STR: Red = high, Blue = medium, Green = low. Values in parenthesis indicate the percentage change from the default STR parameters.

Parameter	Literature Survey		Default STR	Group Method (Unbounded)	Group Method (Bounded)	"Brute Force" Method	PLS Method
	Min	Max					
$f_{SI,XC}$	0.013	0.422	0.1	0.1	0.1	0.1	0.113 (↑13%)
$f_{XI,XC}$	0.02	0.55	0.25	0.25	0.25	0.25	0.22 (↓11%)
$f_{CH,XC}$	0.0718	0.797	0.2	0.2	0.2	0.2	0.290 **(↑45%)**
$f_{PR,XC}$	0.01	0.4	0.2	0.2	0.2	0.2	0.17 (↓14%)
$f_{LI,XC}$	0.014	0.478	0.25	0.25	0.25	0.25	0.203 (↓19%)
$f_{FA,LI}$	0.665	1	0.95	0.95	0.95	0.95	0.965 (↑2%)
$f_{H2,SU}$	0	0.19	0.19	0.19	0.19	0.19	0.17 (↓9%)
$f_{BU,SU}$	0.111	0.13	0.13	0.13	0.13	0.13	0.153 (↑17%)
$f_{PRO,SU}$	0.27	0.54	0.27	0.27	0.27	0.27	0.24 (↓13%)
$f_{AC,SU}$	0.202	0.41	0.41	0.41	0.41	0.41	0.438 (↑7%)
$f_{H2,AA}$	0.042	0.078	0.06	0.06	0.06	0.06	0.063 (↑5%)
$f_{VA,AA}$	0.23	0.309	0.23	0.23	0.23	0.23	0.237 (↑3%)
$f_{BU,AA}$	0.186	0.29	0.26	0.26	0.26	0.26	0.26 (0%)
$f_{PRO,AA}$	0.041	0.12	0.05	0.05	0.05	0.05	0.048 (↓4%)
$f_{AC,AA}$	0.273	0.399	0.4	0.4	0.4	0.4	0.392 (↓2%)
k_{dis}	0.001	1.743	0.4	23.1 **(↑5680%)**	1.565 **(↑291%)**	0.068 **(↓83%)**	0.444 (↑11%)
k_{hyd_ch}	0.037	2.75	0.25	151 **(↑60000%)**	2.75 **(↑1000%)**	2.75 **(↑1000%)**	0.391 **(↑56%)**
k_{hyd_pr}	0.0014	18.23	0.2	61 **(↑30000%)**	1.803 **(↑800%)**	0.112 **(↓44%)**	0.638 **(↑219%)**
k_{hyd_li}	0.0086	2.1	0.1	0.058 **(↓42%)**	0.1 (0%)	2.1 **(↑2000%)**	0.17 **(↑67%)**
K_{s_IN}	0.00007	0.0013	0.0001	0.0001 (0%)	0.0001 (0%)	0.0013 **(↑1200%)**	0.00011 (↑13%)
pH_{UL_acid}	5.5	8.5	5.5	5.5 (0%)	5.5 (0%)	5.5 (0%)	5.6 (↑2%)
pH_{LL_acid}	4	6	4	4 (0%)	4 (0%)	4 (0%)	4.2 (↑5%)
k_{m_su}	11.9	125	30	30 (0%)	30 (0%)	11.9 **(↓60%)**	44.0 **(↑47%)**
K_{s_su}	0.022	4.5	0.5	0.5 (0%)	0.5 (0%)	2.469 **(↑394%)**	0.815 **(↑63%)**
Y_{su}	0.01	0.17	0.1	0.1 (0%)	0.1 (0%)	0.01 **(↓90%)**	0.12 (↑17%)

Parameter	Literature Survey		Default STR	Group Method (Unbounded)	Group Method (Bounded)	"Brute Force" Method	PLS Method
	Min	Max					
k_{dec_xsu}	0.01	0.8	0.02	7.7 (↑38500%)	0.8 (↑3900%)	0.8 (↑3900%)	0.06 (↑195%)
k_{m_aa}	19.8	53	50	50 (0%)	50 (0%)	52.97 (↑6%)	52.5 (↑5%)
K_{s_aa}	0.01	1.198	0.3	0.3 (0%)	0.3 (0%)	1.198 (↑300%)	0.314 (↑4%)
Y_{aa}	0.058	0.15	0.08	0.08 (0%)	0.08 (0%)	0.058 (↓28%)	0.087 (↑8%)
k_{dec_xaa}	0.02	0.8	0.02	0.02 (0%)	0.02 (0%)	0.8 (↑3900%)	0.10 (↑389%)
k_{m_fa}	0.93	12	6	6 (0%)	6 (0%)	11.85 (↑98%)	7.05 (↑18%)
K_{s_fa}	0.024	9.21	0.4	0.4 (0%)	0.4 (0%)	9.21 (↑2200%)	0.758 (↑89%)
Y_{fa}	0.004	0.055	0.06	0.06 (0%)	0.055 (↓8%)	0.0184 (↓69%)	0.067 (↑12%)
k_{dec_xfa}	0.01	0.06	0.02	0.02 (0%)	0.02 (0%)	0.06 (↑200%)	0.03 (↑34%)
KI_{h2_fa}	3×10^{-6}	5×10^{-6}	5×10^{-6}	6×10^{-6} (↑20%)	5×10^{-6} (0%)	3×10^{-6} (↓40%)	5×10^{-6} (0%)
k_{m_c4}	5	60	20	20 (0%)	20 (0%)	56 (↑180%)	24 (↑20%)
K_{s_c4}	0.012	0.6	0.3	0.3 (0%)	0.3 (0%)	0.6 (↑50%)	0.349 (↑16%)
Y_{c4}	0.0193	0.066	0.06	0.06 (0%)	0.06 (0%)	0.066 (↑10%)	0.0625 (↑4%)
k_{dec_xc4}	0.02	0.03	0.02	0.02 (0%)	0.02 (0%)	0.02 (0%)	0.02 (↑2%)
KI_{h2_c4}	1×10^{-8}	1×10^{-5}	1×10^{-5}	1.3×10^{-5} (↑30%)	1×10^{-5} (0%)	1×10^{-7} (↓99%)	1.1×10^{-5} (↑10%)
k_{m_pro}	0.16	100	13	13 (0%)	13 (0%)	100 (↑670%)	14.46 (↑11%)
K_{s_pro}	0.02	1.146	0.3	0.3 (0%)	0.3 (0%)	1.146 (↑74%)	0.482 (↑60%)
Y_{pro}	0.019	0.075	0.04	0.04 (0%)	0.04 (0%)	0.019 (↑53%)	0.043 (↑7%)
k_{dec_xpro}	0.001	0.06	0.02	0.02 (0%)	0.02 (0%)	0.023 (↑15%)	0.025 (↑26%)
KI_{h2_pro}	2.4×10^{-8}	8×10^{-6}	3.5×10^{-6}	2×10^{-4} (↑6470%)	8×10^{-6} (↑129%)	8×10^{-6} (↑129%)	5×10^{-6} (↑43%)
k_{m_ac}	3.1	48	8	205 (↑2460%)	8.07 (↑1%)	48 (↑500%)	12.6 (↑57%)
K_{s_ac}	0.011	0.93	0.15	3.48 (↑2220%)	0.119 (↓21%)	0.097 (↓35%)	0.246 (↑64%)
Y_{ac}	0.014	0.1	0.05	0.05 (0%)	0.05 (0%)	0.014 (↓72%)	0.055 (↑10%)
k_{dec_xac}	0.001	0.05	0.02	0.02 (0%)	0.02 (0%)	0.05 (↑150%)	0.024 (↑17%)
KI_{nh3_ac}	0.00026	0.0223	0.0018	0.0018 (0%)	0.0018 (0%)	0.00026 (↓86%)	0.0047 (↑160%)
pH_{UL_ac}	6.7	7	7	7 (0%)	7 (0%)	7 (0%)	7 (0%)
pH_{LL_ac}	5.2	6	6	6 (0%)	6 (0%)	6 (0%)	6.1 (↑1%)
k_{m_h2}	1.68	209	35	35 (0%)	35 (0%)	1.68 (↓95%)	53.5 (↑53%)
K_{s_h2}	1×10^{-6}	0.0006	2.5×10^{-5}	1.8×10^{-4} (↑620%)	1×10^{-6} (↓96%)	1×10^{-6} (↓96%)	7.7×10^{-5} (↑220%)
Y_{h2}	0.0089	0.183	0.06	0.06 (0%)	0.06 (0%)	0.183 (↑205%)	0.072 (↑19%)
k_{dec_xh2}	0.001	0.3	0.02	0.02 (0%)	0.02 (0%)	0.012 (↓40%)	0.047 (↑134%)
pH_{UL_h2}	6	6.7	6	6 (0%)	6 (0%)	6 (0%)	6 (0%)
pH_{LL_h2}	5	5.8	5	5 (0%)	5 (0%)	5 (0%)	5.2 (↑3%)

5.4. Benchmark Summary

MAPE results from the 4 parameter calibration methods are summarised in Figure 28. The dotted and solid bar graphs represent the calibration period and validation period respectively. Generally speaking, validation error should be higher than the calibration error because calibration is carried out using known data, which means in this context the solid bars should be higher than the dotted bars. However, in this study, MAPE values have appeared otherwise lower during the validation period. This phenomenon is possible because the data used for model training includes both transient (plant ramp-up) and steady-state conditions, whereas validation only considered data under steady-state conditions. Transient data has high variance and is more likely to include "noise"; therefore, the training data is considered to be more complex to model than the validation set.

The Group Methods showed repeated signs of overfitting with respect to VSS, pH and CH_4. The term "overfitting" refers to a scenario where the model achieved superior MAPE during calibration period but failed to fit well during the validation period. A possible cause for this discrepancy is the fact that the first parameter group tends to calibrate too far beyond the surveyed range or past meaningful values. By calibrating parameters in segregated groups, the synergistic effect between two parameters of different groups could be left unaccounted. This resulted in some parameters in the second and third groups to over-exaggerate while attempting to counter the overfitting caused by the prior group's calibration.

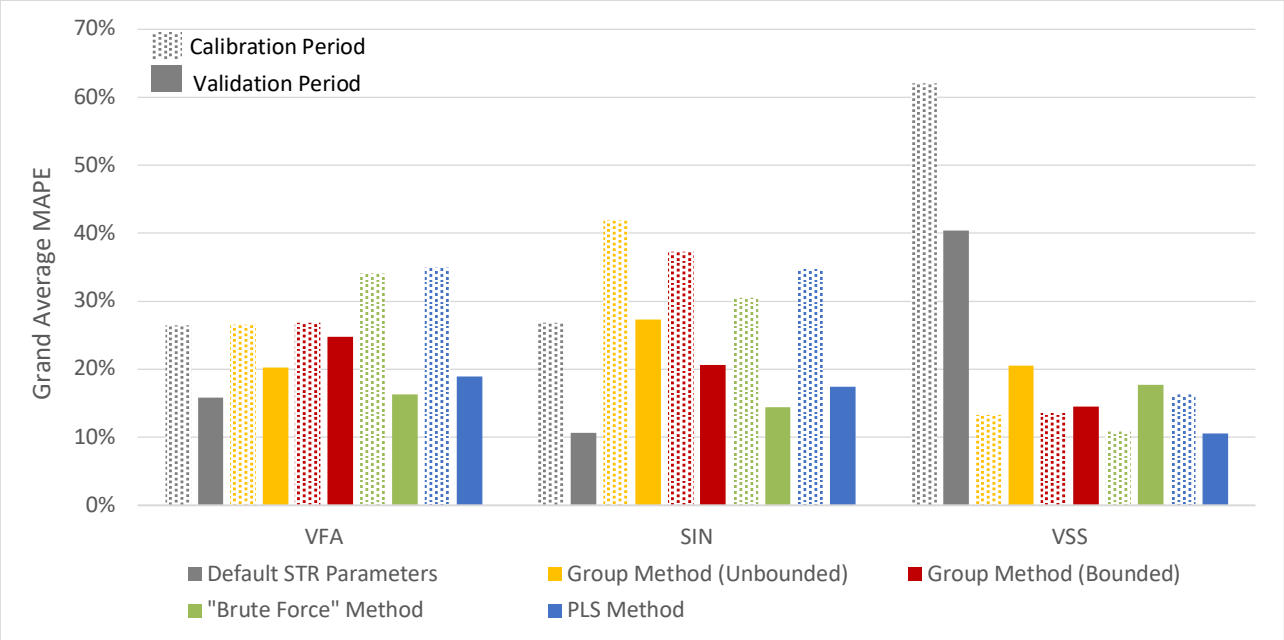


Figure 28: Comparing the accuracy of models produced by various parameter calibration methods during calibration period and validation period. Higher MAPE indicates poorer model accuracy.

The "Brute Force" Method was found to provide the best overall accuracy. It is one of the two investigated methods that could model CO_2 production with high accuracy. The other outputs were also modelled satisfactorily. VSS is the only output that overfitted. The long duration required to optimise a model using this method is a practicality problem. Moreover, it is suspected that this method has forced the model to fit through the accumulation of monosaccharides and amino acids. If true, this calibrated model shall be deemed invalid.

The PLS method has shown to be a more reliable method. There were no signs of overfitting; and overall, it was able to predict unseen data better than the Group Methods. Even though this method was unable to improve the model for CH_4 and CO_2, it was capable of projecting correct movement of the trends including sudden low and high spikes across transient and steady-state conditions. Another advantage that this method offers is the relatively short calibration duration.

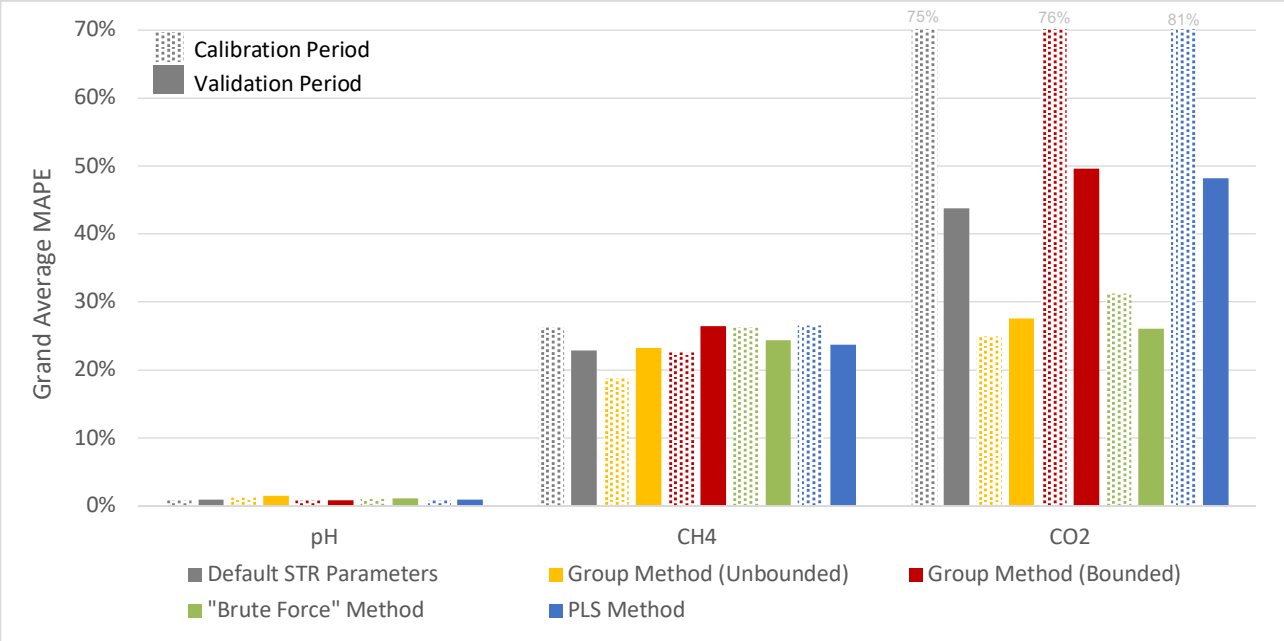


Figure 29: Comparing the accuracy of models produced by various parameter calibration methods during calibration period and validation period. Higher MAPE indicates poorer model accuracy

Table 16 ranks the various methods according to the accuracy of each output model and their optimisation speed. None of the methods was able to improve the fitting for VFA, S_{IN} and CH_4 than a model simply based on default STR parameters; however, it was understood that the fitting accuracy for these outputs had to be compromised in order to improve the fitting of VSS. The behaviour and outcome of model optimisation are dependent on which output(s) are included in the objective function. For example, if the objective function considers CH_4 only and VSS excluded, then the Group Methods will perform much better since it will not over-calibrate hydrolysis and decay parameters to force-fit VSS. This explains why selective calibration of only high sensitivity parameters, such as the Group Method, worked well for most literature (Table 20) where the objective function featured only VFA and gases.

Table 16: Ranking of the various parameter calibration methods according to the model's accuracy during validation and the duration taken to complete the calibration

Method	Ranking of Model Accuracy (Validation Period)						Calibration Speed*
	VFA	S_{IN}	VSS	pH	CH_4	CO_2	
Default STR Parameters	1	1	5	3	1	3	N/A
Group Method (Unbounded)	4	5	4	5	2	2	1.6
Group Method (Bounded)	5	4	2	1	5	5	1.3
"Brute Force" Method	2	2	3	4	4	1	14
PLS Method	3	3	1	2	3	4	1.0

* Order of magnitude relative to the PLS Method

CHAPTER 6

CONCLUSION AND RECOMMENDATIONS

6.1. Summary of Findings

Industrial-scale AD plants currently do not perform compositional analysis with the same level of detail as required by ADM1 for model input. For practical reasons, industrial operations focus only on key performance indicators that can be analysed conveniently and economically. For instance, knowing the "total VFA concentration" is adequate for plant operation, but it is mandatory for ADM1 that VFA is described as four specific short-chain VFAs. The methodology and assumptions applied to translate these indicators into ADM1 state variables are therefore major sources of model uncertainties. It is important that they are scrutinised or verified through further investigations, if possible.

The parameter calibration method proposed in this thesis successfully improved the overall accuracy of an ADM1 model based on data from an industrial-scale anaerobic digester. It makes use of the PLSR algorithm to capture the underlying interaction effects between ADM1 model parameters and the model outputs as PLS weights and loadings. The data matrices, onto which PLSR are applied, are generated from a Monte Carlo simulation that repeatedly ran ADM1 with randomised parameter sets. Each parameter is constrained to randomise within the range of values surveyed from other ADM1 research works.

Model optimisation is accomplished by minimising the objective function using a least squares method. Instead of assigning the aggregate of RMSE from various output graphs as the objective function, this study demonstrated that optimisation can be achieved by defining the objective function as the *difference in PLS output latent variables between the simulated data and actual data*. The advantage of this alternative approach is that it does not require the modeller to stipulate how the aggregate weightings are distributed. The latent variables intrinsically take into account the sensitivities of each output variable.

Increasing the number of Monte Carlo iterations enhances data redundancy which is essential for the PLSR algorithm to capture underlying trends more accurately. Furthermore, it promotes the development of the absolute upper bounds, lower bounds and mean values in the uncertainty data. These components influence how data is normalised and ultimately the influence each output has on the objective function. For the industrial-scale plant studied in this thesis, 500 or more Monte Carlo runs were found to be statistically adequate for normalisation purpose. However, since parameter randomisation is done stochastically, it is advisable to execute as many Monte Carlo runs as practically feasible such that all major uncertainty trends are exposed. Given adequate iterations, the only input required by the modeller is the choice of number of latent variables.

Utilising more latent variables enables the PLSR algorithm to capture the interactions of lesser weighted outputs. This enhances the influence of calibration on these variables during model optimisation. Utilising too many latent variables may induce overfitting, while too few latent variables could leave lower weighted outputs

unaffected. Specific to this study where six outputs were modelled, as few as two latent variables were found to be adequate in producing improved models. Additional latent variables may need to be introduced if more outputs are to be modelled.

The uncertainty data gathered from the Monte Carlo simulation *do not* require any outlier removal before it is subjected to the PLSR algorithm. Removing outliers was found to cause poor model fitting. All data points, including those that would normally be classified as outliers, were found to contain valuable information for PLS constructs development.

The PLS Method can become computationally demanding. In general, increasing the number of Monte Carlo runs or the number of latent variables tends to increase the number of iterations taken to calibrate the parameter set. The number of simulated time intervals (e.g. days) also influences the computation requirements; however, this factor is not within the modeller's control. It is thus critical to remain conservative when selecting the number of Monte Carlo runs and to start the PLSR algorithm with one latent variable; introducing more if necessary.

Different parameter calibration methods approach the process differently and thus yield different model outcomes. When benchmarked against other methods such as the Group Method and "Brute Force" Method, the PLS Method displayed promising results:

- It is the only method that did not show signs of overfitting.
- It is the only method that concluded the model optimisation with all calibrated parameter values within the surveyed minimum and maximum range.
- Overall model accuracy is quite acceptable, although less superior than the "Brute Force" Method.
- It converges on the objective function was 30-60% faster than the Group Method and 14 times quicker than the "Brute Force" Method

The success is attributed to the fundamentals of PLS regression, in which the interactions between the parameters and the model sensitivity of each outputs are mapped into PLS constructs and latent variables. Unlike other methods where parameters are adjusted one by one, PLS constructs enable parameters to be manipulated collectively in a manner that ensures maximum impact on the outputs while considering collinearities among the parameters. This guided approach effectively prevents overfitting during calibration and speeds up the process.

Although the "Brute force" method enables a better model accuracy, its extreme long execution duration renders it a rather impractical option for industrial (on-site) use. Moreover, many of its calibrated parameters are capped by either the maximum or minimum values of the surveyed range. This fact could suggest that the substrate bears no resemblance to the substrates studied in the survey, and thereby reiterates the importance of how experimental data (COD, VFA, etc.) are translated into ADM1 state variables.

Lastly, it is noted that the outcome of model optimisation could differ depending on the outputs included in the objective function evaluation. For instance, the poor performance of the Group Methods is attributed to the inclusion of VSS as part of the objective function. Over-calibration of the hydrolysis and decay parameters in

the first parameter group subsequently resulted in a corrective reaction (over-calibration) in the second parameter group.

6.2. Final Conclusions

An ADM1 parameter calibration method based on the PLS regression framework, called “PLS Method”, has been developed. Benchmarking the PLS Method against two other parameter calibration methods confirmed the PLS Method as a suitable alternative method for industrial-scale digester modelling. Not only does it eliminates the risk of overfitting, the total calibration duration is also attractively shorter.

By using this method, it is no longer necessary for modellers to take any decision on the sensitivity analysis method to apply, or the model parameters to calibrate. Instead, all parameters will undergo calibration simultaneously in accordance to the PLS constructs extracted from simple Monte Carlo simulations. A notable drawback of this method is the fact that a Monte Carlo simulation is a mandatory procedure regardless of the number of outputs predicted, whereas methods that do not require prerequisite sensitivity analysis (e.g. Group Method) could commence immediately.

The methodology consists of four steps which modellers can apply to optimise ADM1:

1. Perform a Monte Carlo simulation using values within the surveyed parameter ranges.
2. Apply a PLSR algorithm on the Monte Carlo results to produce PLS constructs.
3. Optimise the model by minimising the objective function, which is defined as the difference in PLS output latent variables between the simulated outputs and experimental outputs.
4. Validate the calibrated parameter set against unseen data.

6.3. Recommendations

Prior to the availability of affordable advanced instruments, regular calibration of a plant model serving as a soft sensor may be the short-term answer to address the dynamic nature of industrial wastewater compositions. The fact that the PLS latent structures do not change unless the model structure (i.e. rate equations, stoichiometry, etc.) changes, it is not necessary to run Monte Carlo simulations when the model re-calibrates. This means that the soft sensor model using the PLS Method will be able to update parameters within a short time on a regular, or even online, basis using three latest plant measurements.

It is inevitable that assumptions or references to other literature have to be made when translating experimental data into the ADM1 structure. A major uncertainty noted pertains to how soluble COD and particulate COD values are translated into the various ADM1 state variables. As a prerequisite to ADM1 modelling, it is recommended that a once-off substrate characterisation test is carried out. The purpose of this test is to establish the COD fractions between carbohydrates, proteins and lipids, as well as to identify the compositional make-up by the various VFAs. Understanding these factors will ensure that the substrate is correctly described within ADM1.

6.4. Future Research

In spite of the extensive research in ADM1 there is currently no consolidated database of calibrated parameters. It is encouraged that the research community continue to build on the parameter database established in this study. A growing database does not only benefit the framework presented in this study, it would serve well as a directory for similar substrate type or a statistical resource for the development of new techniques. For the convenience of future research, resources developed in this study such as the parameter database and Scilab codes can be reused.

The applicability of the proposed framework should be trialled under different scenarios in which different substrate types and loading rates are experimented. Further investigation could look at using a different parameter set (besides the default set given in the STR) as the starting set for optimisation.

Hydrogen is widely acknowledged as a key methanogenic activity inhibitor and a precursor to process failure. Future research should, whenever possible, include hydrogen concentration measurements into the experimental setup and to include it as one of the modelled outputs.

REFERENCES

Akunna, J.C., Bizeau, C. & Moletta, R. 1992. Denitrification in anaerobic digesters: Possibilities and influence of wastewater COD/N-NOX ratio. *Environmental Technology (United Kingdom)*. 13(9):825–836.

Angelidaki, I. & Ellegaard, L. 2003. Codigestion of manure and organic wastes in centralized biogas plants: Status and future trends. *Applied Biochemistry and Biotechnology - Part A Enzyme Engineering and Biotechnology*. 109(1–3):95–105.

Angelidaki, I., Ellegaard, L. & Ahring, B.K. 1993. A mathematical model for dynamic simulation of anaerobic digestion of complex substrates: Focusing on ammonia inhibition. *Biotechnology and Bioengineering*. 42(2):159–166.

Angelidaki, I., Alves, M., Bolzonella, D., Borzacconi, L., Campos, J.L., Guwy, A.J., Kalyuzhnyi, S., Jenicek, P., et al. 2009. Defining the biomethane potential (BMP) of solid organic wastes and energy crops: A proposed protocol for batch assays. *Water Science and Technology*. 59(5):927–934.

Antonopoulou, G., Gavala, H.N., Skiadas, I. V. & Lyberatos, G. 2012. ADM1-based modeling of methane production from acidified sweet sorghum extract in a two stage process. *Bioresource Technology*. 106(1):10–19.

APHA. 1992. *Standard Methods for the Examination of Water and Wastewater*. 18th ed. ed. Washington: American Public Health Association.

Appels, L., Baeyens, J., Degrève, J. & Dewil, R. 2008. Principles and potential of the anaerobic digestion of waste-activated sludge. *Progress in Energy and Combustion Science*. 34(6):755–781.

Arnell, M., Astals, S., Åmand, L., Batstone, D.J. & Jensen, P.D. 2016. Modelling anaerobic co-digestion in Benchmark Simulation Model No . 2 : Parameter estimation , substrate characterisation and plant-wide integration. *Water Research*. 98:138–146.

Arnell, M., Astals, S., Åmand, L., Batstone, D.J., Jensen, P.D. & Jeppsson, U. 2016. Modelling anaerobic co-digestion in Benchmark Simulation Model No. 2: Parameter estimation, substrate characterisation and plant-wide integration. *Water Research*. 98(2):138–146.

Astals, S., Batstone, D.J., Mata-alvarez, J. & Jensen, P.D. 2014. Bioresource Technology Identification of synergistic impacts during anaerobic co-digestion of organic wastes. *Bioresource Technology*. 169:421–427.

Bair, E., Hastie, T., Paul, D. & Tibshirani, R. 2006. Prediction by supervised principal components. *Journal of the American Statistical Association*. 101(473):119–137.

Baltes, M., Schneider, R., Sturm, C. & Reuss, M. 1994. Unstructured Growth Models. (1989):480–488.

Barrera, E.L., Spanjers, H., Solon, K., Amerlinck, Y., Nopens, I. & Dewulf, J. 2015. Modeling the anaerobic digestion of cane-molasses vinasse: Extension of the Anaerobic Digestion Model No. 1 (ADM1) with sulfate reduction for a very high strength and sulfate rich wastewater. *Water Research*. 71:42–54.

Batstone, D.J. 2006. Mathematical modelling of anaerobic reactors treating domestic wastewater: Rational criteria for model use. *Reviews in Environmental Science and Biotechnology*. 5(1):57–71.

Batstone, D.J. & Keller, J. 2003. Industrial applications of the IWA anaerobic digestion model No . 1. *Water Science and Technology*. 47(1):199–206.

Batstone, D.J., Keller, J., Angelidaki, I., Kalyuzhnyi, S.V., Pavlostathis, S.G., Rozzi, A., Sanders, W.T.M., Siegrist, H., et al. 2002. *Anaerobic Digesiton Model No. 1*. London, UK: International Water Association (IWA) Publishing.

Batstone, D.J., Pind, P.F. & Angelidaki, I. 2003. Kinetics of thermophilic, anaerobic oxidation of straight and branched chain butyrate and valerate. *Biotechnology and Bioengineering*. 84(2):195–204.

Batstone, D.J., Tait, S. & Starrenburg, D. 2009. Estimation of hydrolysis parameters in full-scale anerobic digesters. *Biotechnology and Bioengineering*. 102(5):1513–1520.

Batstone, D.J., Puyol, D., Flores-Alsina, X. & Rodríguez, J. 2015. Mathematical modelling of anaerobic digestion processes: applications and future needs. *Reviews in Environmental Science and Biotechnology*. 14(4):595–613.

Bergland, W.H., Dinamarca, C. & Bakke, R. 2015. Temperature Effects in Anaerobic Digestion Modeling. *Proceedings of the 56th SIMS*. (1):261–269.

Bernard, O., Hadj-Sadok, Z., Dochain, D., Genovesi, A. & Steyer, J.P. 2001. Dynamical model development and parameter identification for an anaerobic wastewater treatment process. *Biotechnology and bioengineering*. 75(4):424–438.

Blumensaat, F. & Keller, J. 2005. Modelling of two-stage anaerobic digestion using the IWA Anaerobic Digestion Model No . 1 (ADM1). 39(1):171–183.

Boe, K., Batstone, D.J., Steyer, J.P. & Angelidaki, I. 2010. State indicators for monitoring the anaerobic digestion process. *Water Research*. 44(20):5973–5980.

Boubaker, F. & Ridha, B.C. 2008. Modelling of the mesophilic anaerobic co-digestion of olive mill wastewater with olive mill solid waste using anaerobic digestion model No . 1 (ADM1). 99(1):6565–6577.

Braun, R., Huber, P. & Meyrath, J. 1981. Ammonia toxicity in liquid piggery manure digestion. *Biotechnology Letters*. 3(4):159–164.

Carvalho, F., Prazeres, A.R. & Rivas, J. 2013. Cheese whey wastewater: Characterization and treatment. *Science of the Total Environment*. 445–446:385–396.

Chen, T.-H. & Hashimoto, A.G. 1996. Effects of pH and Substrate: Inoculum Ratio on Batch Methane Fermentation. *Bioresource Technology*. 56:179–186.

Chin, W.W., Marcolin, B.L. & Newsted, P.R. 2003. A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic-Mail Motion/Adoption Study. *Information Systems Research*. 14(2):189–217.

Chynoweth, D., Turick, C., Owens, J., Jerger, D. & Peck, M. 1993. Biochemical methane potential of biomass and waste feedstocks. *Biomass and Bioenergy*. 5:95–111.

Conrad, R. 1999. Contribution of hydrogen to methane production and control of hydrogen concentrations in methanogenic soils and sediments. *FEMS Microbiology Ecology*. 28:193–202.

Costello, D.J., Greenfield, P.F. & Lee, P.L. 1991. Dynamic modelling of a single-stage high-rate anaerobic reactor-I. Model derivation. *Water Research*. 25(7):847–858.

Danalewich, J.R., Papagiannis, T.G., Belyea, R.L., Tumbleson, M.E. & Raskin, L. 1998. Characterization of dairy waste streams, current treatment practices, and potential for biological nutrient removal. *Water Research*. 32(12):3555–3568.

Diwekar, U. & David, A. 2015. Uncertainty Analysis and Sampling Techniques. In New York: Springer *BONUS Algorithm for Large Scale Stochastic Nonlinear Programming Problems*.

Donoso-Bravo, A., Retamal, C., Carballa, M., Ruiz-Filippi, G. & Chamy, R. 2009. Influence of temperature on the hydrolysis, acidogenesis and methanogenesis in mesophilic anaerobic digestion: Parameter identification and modeling application. *Water Science and Technology*. 60(1):9–17.

Donoso-Bravo, A., Mailier, J., Martin, C., Rodríguez, J., Aceves-Lara, C.A. & Wouwer, A. Vande. 2011. Model selection, identification and validation in anaerobic digestion: A review. *Water Research*. 45(17):5347–5364.

Eastman, J. a & Ferguson, J.F. 1981. Solubilization of particulate organic carbon during the acid phase of anaerobic digestion. *Journal of the Water Pollution Control Federation*. 53(3):352–366.

Esbensen, K.H. & Julius, L.P. 2010. Representative Sampling, Data Quality, Validation - A Necessary Trinity in Chemometrics. *Comprehensive Chemometrics*. 4:1–20.

Ge, H., Jensen, P.D. & Batstone, D.J. 2011. Increased temperature in the thermophilic stage in temperature

phased anaerobic digestion (TPAD) improves degradability of waste activated sludge. *Journal of Hazardous Materials*. 187(1–3):355–361.

Geladi, P. & Kowalski, B.R. 1986. Partial least-squares regression: a tutorial. *Analytica Chimica Acta*. 185(C):1–17.

Girault, R. & Steyer, J. 2010. Influent fractionation and parameter calibration for ADM1: Lab-scale and full-scale experiments. *Wastewater*. (ii):1–12. [Online], Available: https://wwwappli.nantes.inra.fr/psdr/Biodecol2/R32_Girault_Steyer_2010.pdf.

Gottschal, J.C. & Morris, J.G. 1981. The induction of acetone and butanol production in cultures of Clostridium acetobutylicum by elevated concentrations of acetate and butyrate. *FEMS Microbiology Letters*. 12(4):385–389.

Grau, P.Ã., Gracia, M. De, Vanrolleghem, P.A. & Ayesa, E. 2007. A new plant-wide modelling methodology for WWTPs. *Water Research*. 41:4357–4372.

Guisasola, a, Baeza, J.A., Carrera, J., Sin, G., Vanrolleghem, P.A. & Lafuente, J. 2006. The Influence of Experimental Data Quality and Quantity on Parameter Estimation Accuracy. *Education for Chemical Engineers*. 1(1):139–145.

Gujer, W. & Zehnder, A.J.B. 1983. Conversion Processes in Anaerobic Digestion. *Water Sci Technol*. 15:127–167. [Online], Available: http://www.iwaponline.com/wst/01508/wst015080127.htm.

van Haandel, A. & van der Lubbe, J. 2007. *Handbook Biological Waste Water Treatment*. First ed. Leidschendam: Quist Publishing.

Hawkes, F.R., Guwy, A.J., Rozzi, A.G. & Hawkes, D.L. 1993. A new instrument for on-line measurement of bicarbonate alkalinity. *Water Research*. 27(1):167–170.

Heukelekian, H. 1958. Basic Principles of Sludge Digestion. In New York: Reinhold Publishing Corp. *Biological Treatment of Sewage and Industrial Wastes, vol. II*.

Hill, D.T. & Barth, C.L. 1977. A dynamic model for simulation of animal waste digestion. *Journal of the Water Pollution Control Association*. 10(10):2129–2143.

Jeong, H.-S., Suh, C.-W., Lim, J.-L., Lee, S.-H. & Shin, H.-S. 2005. Analysis and application of ADM1 for anaerobic methane production. *Bioprocess and biosystems engineering*. 27(2):81–89.

Jimenez, J., Latrille, E., Harmand, J., Robles, A., Ferrer, J., Gaida, D., Wolf, C., Mairet, F., et al. 2015. Instrumentation and control of anaerobic digestion processes: a review and some research challenges. *Reviews in Environmental Science and Biotechnology*. 14(4):615–648.

Jones, D.T. & Woods, D.R. 1986. Acetone-Butanol Fermentation Revisited. 50(4):484–524.

de Jong, S. 1993. SIMPLS: an alternative approach squares regression to partial least. *Elsevier Science Publishers B.V*. 18:2–263.

Kalyuzhnyi, S. V. & Fedorovich, V. V. 1998. Mathematical modelling of competition between sulphate reduction and methanogenesis in anaerobic reactors. *Bioresource Technology*. 65(3):227–242.

Kalyuzhnyi, S. V., Perez Martinez, E. & Rodriguez Martinez, J. 1997. Anaerobic treatment of high-strength cheese-whey wastewaters in laboratory and pilot UASB-reactors. *Bioresource Technology*. 60(1):59–65.

Kleinbaum, D.G., Kupper, L.L., Muller, K.E. & Nizam, A. 1998. *Applied regression analysis and other multivariable methods*. 3rd ed. Belmont, CA, US: Thomson Brooks/Cole Publishing Co.

Koch, K., Lübken, M., Gehring, T., Wichern, M. & Horn, H. 2010. Biogas from grass silage – Measurements and modeling with ADM1. *Bioresource Technology*. 101(21):8158–8165.

Lauwers, J., Appels, L., Thompson, I.P., Degrève, J., Van Impe, J.F. & Dewil, R. 2013. Mathematical modelling of anaerobic digestion of biomass and waste: Power and limitations. *Progress in Energy and Combustion Science*. 39(4):383–402.

Liamleam, W. & Annachhatre, A.P. 2007. Electron donors for biological sulfate reduction. *Biotechnology Advances*. 25(5):452–463.

Van Lier, J.B. 2008. High-rate anaerobic wastewater treatment: Diversifying from end-of-the-pipe treatment to resource-oriented conversion techniques. *Water Science and Technology*. 57(8):1137–1148.

Liu, J., Olsson, G. & Mattiasson, B. 2004. Monitoring and control of an anaerobic upflow fixed-bed reactor for high-loading-rate operation and rejection of disturbances. *Biotechnology and Bioengineering*. 87(1):43–53.

Lokshina, L.Y. & Vavilin, V.A. 1999. Kinetic analysis of the key stages of low temperature methanogenesis. 117:285–303.

López, I. & Borzacconi, L. 2009. Modelling a full scale UASB reactor using a COD global balance approach and state observers. *Chemical Engineering Journal*. 146(1):1–5.

Lübken, M., Wichern, M., Schlattmann, M., Gronauer, A. & Horn, H. 2007. Modelling the energy balance of an anaerobic digester fed with cattle manure and renewable energy crops. *Water Research*. 41(18):4085–4096.

Madsen, M., Holm-Nielsen, J.B. & Esbensen, K.H. 2011. Monitoring of anaerobic digestion processes: A review perspective. *Renewable and Sustainable Energy Reviews*. 15(6):3141–3155.

Mairet, F., Bernard, O., Ras, M., Lardon, L. & Steyer, J. 2011. Testing the ability of ADM1 to represent the anaerobic digestion of microalgae. *Integrated Assessment*. 151–158.

Maitra, S. & Yan, J. 2008. *Principle Component Analysis and Partial Least Squares : Two Dimension Reduction Techniques for Regression*.

McCarty, P.L. 1964. Anaerobic Waste Treatment Fundamentals. *Chemistry and microbiology*. 95(9):107–112.

McCarty, P.L. & Smith, D.P. 1986. Anaerobic wastewater treatment. *Environmental Science & Technology*. 20(12):1200–1206.

McInerney, M.J. & Bryant, M.P. 1981. *Basic Principles of Bioconversions in Anaerobic Digestion and Methanogenesis*. New York: Plenum Press.

Monod, J. 1942. The Growth of Bacterial Cultures. (Paris: Hermann et Cie).

Mosey, F.E. 1983. Mathematical modelling of the anaerobic digestion process: Regulatory mechanisms for the formation of short-chain volatile acids from glucose. *Water Science and Technology*. 15(8–9):209–232.

Pavlostathis, S.G. & Giraldo-Gomez, E. 1991. Kinetics of anaerobic treatment. *Water Science and Technology*. 24(8):35–59.

Pavlostathis, S.G. & Gossett, J.M. 1988. Preliminary Conversion Mechanisms in Anaerobic Digestion of Biological Sludges. *Journal of Environmental Engineering*. 114:575–592.

Pohland, F.G. & Ghosh, S. 1971. Developments in Anaerobic Stabilization of Organic Wastes - The Two-Phase Concept. *Environmental Letters*. 1(4):255–266.

Raposo, F., Banks, C.J., Siegert, I., Heaven, S. & Borja, R. 2006. Influence of inoculum to substrate ratio on the biochemical methane potential of maize in batch tests. *Process Biochemistry*. 41(6):1444–1450.

Razaviarani, V. & Buchanan, I.D. 2015. Calibration of the Anaerobic Digestion Model No. 1 (ADM1) for steady-state anaerobic co-digestion of municipal wastewater sludge with restaurant grease trap waste. *Chemical Engineering Journal*. 266(1):91–99.

Rebac, S., Ruskova, J., Gerbens, S., van Lier, J.B., Stams, A.J.M. & Lettinga, G. 1995. High-rate anaerobic treatment of wastewater under psychrophilic conditions. *Journal of Fermentation and Bioengineering*. 80(5):499–506.

Ripley, A.L.E., Boyle, W.C., Converse, J.C., Ripley, L.E., Boyle, W.C. & Converse, J.C. 1986. Improved Alkalimetric Monitoring for Anaerobic Digestion of High-strength Wastes. 58(5):406–411.

Rivas, J., Prazeres, A.R., Carvalho, F. & Beltrán, F. 2010. Treatment of cheese whey wastewater: Combined Coagulation - Flocculation and aerobic biodegradation. *Journal of Agricultural and Food Chemistry*. 58(13):7871–7877.

Robinson, J.A. & Tiedje, J.M. 1983. Nonlinear estimation of Monod growth kinetic parameters from a single substrate depletion curve . Nonlinear Estimation of Monod Growth Kinetic Parameters from a Single Substrate Depletion Curvet. 45(5):1453–1458.

Rosen, C. & Jeppsson, U. 2006. Aspects on ADM1 Implementation within the BSM2 Framework. *Technical report*. 1–37.

Saltelli, S., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. & Tarantola, S. 2008. *Global Sensitivity Analysis: The Primer*. Chichester: John Wiley & Sons Ltd.

Sarkar, A.X. & Sobie, E.A. 2010. Regression analysis for constraining free parameters in electrophysiological models of cardiac cells. *PLoS Computational Biology*. 6(9).

Siegrist, H., Renggli, D. & Gujer, W. 1993. Mathematical modelling of anaerobic mesophilic sewage sludge treatment. *Water Science and Technology*. 27(2):25–36.

Siegrist, H., Vogt, D., Garcia-Heras, J.L. & Gujer, W. 2002. Mathematical model for meso- and thermophilic anaerobic sewage sludge digestion. *Environmental Science and Technology*. 36(5):1113–1123.

Sin, G., Gernaey, K. V., Neumann, M.B., van Loosdrecht, M.C.M. & Gujer, W. 2009. Uncertainty analysis in WWTP model applications: A critical discussion using an example from design. *Water Research*. 43(11):2894–2906.

Sin, G., Gernaey, K. V., Neumann, M.B., van Loosdrecht, M.C.M. & Gujer, W. 2011. Global sensitivity analysis in wastewater treatment plant model applications: Prioritizing sources of uncertainty. *Water Research*. 45(2):639–651.

Sobol, I.M. & Kucherenko, S.S. 2005. Global sensitivity indices for nonlinear mathematical models. Review. *Wilmott*. 2005(1):56–61.

Spanjers, H. & Lier, J.B. va. 2006. Instrumentation in anaerobic treatment – research and practice. *Water Science & Technology*. 53(4–5):63.

Steyer, J.P., Bouvier, J.C., Conte, T., Gras, P. & Sousbie, P. 2002. Evaluation of a four year experience with a fully instrumented anaerobic digestion process. *Water Science and Technology*. 45(4–5):495–502.

Thamsiriroj, T. & Murphy, J.D. 2011. Modelling mono-digestion of grass silage in a 2-stage CSTR anaerobic digester using ADM1. *Bioresource Technology*. 102(2):948–959.

Tiedje, J.M. 1988. Ecology of denitrification and dissimilatory nitrate reduction to ammonium. *Environmental Microbiology of Anaerobes*.

Vanrolleghem, P.A. & Lee, D.S. 2003. On-line monitoring equipment for wastewater treatment processes: State of the art. *Water Science and Technology*. 47(2):1–34.

Vavilin, V.A., Vasiliev, V.B., Ponomarev, A. V. & Rytow, S. V. 1994. Simulation model “methane” as a tool for effective biogas production during anaerobic conversion of complex organic matter. *Bioresource Technology*. 48(1):1–8.

Wolin, M.J. 1982. Hydrogen transfer in microbial communities. In A.T. Bull & J.H. Slater (eds.). London: Academic Press, Inc. *Microbial Interactions in Communities*.

Yang, K., Yu, Y. & Hwang, S. 2003. Selective optimization in thermophilic acidogenesis of cheese-whey wastewater to acetic and butyric acids: Partial acidification and methanation. *Water Research*. 37(10):2467–2477.

Yu, H.G. & Fang, H.H. 2002. Acidogenesis of dairy wastewater at various pH levels. *Water Science and*

Technology. 45(10):201–206.

Yu, H.Q. & Fang, H.H.P. 2001. Acidification of mid- and high-strength dairy wastewaters. *Water Research*. 35(15):3697–3705.

APPENDICES

8.1. Appendix A – ADM1 Nomenclature

Dynamic State Variables

Table 17: Description of the state variables used in ADM1 models

i	Variable	Unit	Description	*i*	Variable	Unit	Description
1	S_{su}	kgCOD/m³	Monosaccharides	16	X_{li}	kgCOD/m³	Lipids
2	S_{aa}	kgCOD/m³	Amino acids	17	X_{su}	kgCOD/m³	Monosaccharide degraders
3	S_{fa}	kgCOD/m³	Total LCFA	18	X_{aa}	kgCOD/m³	Amino acid degraders
4	S_{va}	kgCOD/m³	Total valerate	19	X_{fa}	kgCOD/m³	LCFA degraders
5	S_{bu}	kgCOD/m³	Total butyrate	20	X_{c4}	kgCOD/m³	C4-degraders
6	S_{pro}	kgCOD/m³	Total propionate	21	X_{pro}	kgCOD/m³	Propionate degraders
7	S_{ac}	kgCOD/m³	Total acetate	22	X_{ac}	kgCOD/m³	Acetate degraders
8	S_{h2}	kgCOD/m³	Hydrogen	23	X_{h2}	kgCOD/m³	Hydrogen degraders
9	S_{ch4}	kgCOD/m³	Methane	24	X_{I}	kgCOD/m³	Particulate inerts
10	S_{IC}	kmol C/m³	Inorganic carbon	25	S_{an}	kmol/m³	Anions
11	S_{IN}	kmol N/m³	Inorganic nitrogen	26	S_{cat}	kmol/m³	Cations
12	S_{I}	kgCOD/m³	Soluble inerts	27	$S_{h2,g}$	kgCOD/m³	Hydrogen (gas)
13	X_{c}	kgCOD/m³	Composites	28	$S_{ch4,g}$	kgCOD/m³	Methane (gas)
14	X_{ch}	kgCOD/m³	Carbohydrates	29	$S_{co2,g}$	kgCOD/m³	Carbon dioxide (gas)
15	X_{pr}	kgCOD/m³	Proteins				

Model Parameters

(i) Stoichiometric Parameters

Table 18: Description of the stoichiometric parameters used in ADM1 models

Parameter	Unit	Description	Parameter	Unit	Description
$f_{SI,XC}$	-	Soluble inerts fraction from composites	$f_{PRO,SU}$	-	Propionate fraction from monosaccharides
$f_{XI,XC}$	-	Particulate inerts fraction from composites	$f_{AC,SU}$	-	Acetate fraction from monosaccharides

Parameter	Unit	Description	Parameter	Unit	Description
$f_{CH,XC}$	-	Carbohydrates fraction from composites	$f_{H2,AA}$	-	Hydrogen fraction from amino acids
$f_{PR,XC}$	-	Proteins fraction from composites	$f_{VA,AA}$	-	Valerate fraction from amino acids
$f_{LI,XC}$	-	Lipids fraction from composites	$f_{BU,AA}$	-	Butyrate fraction from amino acids
$f_{FA,LI}$	-	Fatty acids fraction from lipids	$f_{PRO,AA}$	-	Propionate fraction from amino acids
$f_{H2,SU}$	-	Hydrogen fraction from monosaccharides	$f_{AC,AA}$	-	Acetate fraction from amino acids
$f_{BU,SU}$	-	Butyrate fraction from monosaccharides			

(ii) Kinetic Parameters

Table 19: Description of the kinetic parameters used in ADM1 models

Parameter	Unit	Description
k_{dis}	d^{-1}	Disintegration factor
k_{hyd_CH}	d^{-1}	Carbohydrates hydrolysis rate constant
k_{hyd_PR}	d^{-1}	Proteins hydrolysis rate constant
k_{hyd_LI}	d^{-1}	Lipids hydrolysis rate constant
K_{s_IN}	$kmol/m^3$	Inorganic nitrogen concentration threshold; growth ceases when exceeded
pH_{UL_acid}	-	pH threshold; no inhibition when pH is above this level
pH_{LL_acid}	-	pH threshold; full inhibition when pH is below this level
k_{m_su}	$COD.COD^{-1}.d^{-1}$	Monod maximum specific uptake rate for monosaccharide degraders
K_{s_su}	$kgCOD.m^{-3}$	Monod half saturation value for monosaccharide degradation
Y_{su}	$COD.COD^{-1}$	Biomass yield on uptake of monosaccharides
k_{dec_xsu}	d^{-1}	Decay rate constant of monosaccharide degraders
k_{m_aa}	$COD.COD^{-1}.d^{-1}$	Monod maximum specific uptake rate for amino acid degraders
K_{s_aa}	$kgCOD.m^{-3}$	Monod half saturation value for amino acid degradation
Y_{aa}	$COD.COD^{-1}$	Biomass yield on uptake of amino acids
k_{dec_xaa}	d^{-1}	Decay rate constant of amino acid degraders
k_{m_fa}	$COD.COD^{-1}.d^{-1}$	Monod maximum specific uptake rate for LCFA degraders
K_{s_fa}	$kgCOD.m^{-3}$	Monod half saturation value for LCFA degradation
Y_{fa}	$COD.COD^{-1}$	Biomass yield on uptake of LCFA
k_{dec_xfa}	d^{-1}	Decay rate constant of LCFA degraders
KI_{h2_fa}	$kgCOD.m^{-3}$	Hydrogen inhibitory concentration for LCFA degraders

Parameter	Unit	Description
k_{m_c4}	$COD.COD^{-1}.d^{-1}$	Monod maximum specific uptake rate for valerate & butyrate degraders
K_{s_c4}	$kgCOD.m^{-3}$	Monod half saturation value for valerate & butyrate degradation
Y_{c4}	$COD.COD^{-1}$	Biomass yield on uptake of valerate & butyrate
k_{dec_xc4}	d^{-1}	Decay rate constant of valerate & butyrate degraders
k_{m_pro}	$COD.COD^{-1}.d^{-1}$	Monod maximum specific uptake rate for propionate degraders
K_{s_pro}	$kgCOD.m^{-3}$	Monod half saturation value for propionate degradation
Y_{pro}	$COD.COD^{-1}$	Biomass yield on uptake of propionate
k_{dec_xpro}	d^{-1}	Decay rate constant of propionate degraders
KI_{h2_pro}	$kgCOD.m^{-3}$	Hydrogen inhibitory concentration for propionate degraders
k_{m_ac}	$COD.COD^{-1}.d^{-1}$	Monod maximum specific uptake rate for acetate degraders
K_{s_ac}	$kgCOD.m^{-3}$	Monod half saturation value for acetate degradation
Y_{ac}	$COD.COD^{-1}$	Biomass yield on uptake of acetate
k_{dec_xac}	d^{-1}	Decay rate constant of acetate degraders
KI_{nh3_ac}	$kgCOD.m^{-3}$	Free ammonia inhibitory concentration on acetate degraders
pH_{UL_ac}	-	pH threshold; no inhibition on acetate degradation when pH is above this level
pH_{LL_ac}	-	pH threshold; full inhibition on acetate degradation when pH is below this level
k_{m_h2}	$COD.COD^{-1}.d^{-1}$	Monod maximum specific uptake rate for hydrogen degraders
K_{s_h2}	$kgCOD.m^{-3}$	Monod half saturation value for hydrogen degradation
Y_{h2}	$COD.COD^{-1}$	Biomass yield on uptake of hydrogen
K_{dec_xh2}	d^{-1}	Decay rate constant of hydrogen degraders
pH_{UL_h2}	-	pH threshold; no inhibition on acetate degradation when pH is above this level
pH_{LL_h2}	-	pH threshold; full inhibition on acetate degradation when pH is below this level

8.2. Appendix B – Model Optimisation Method Survey

Table 20: Survey of current ADM1 model optimisation methods

Reference	Experiment Type	Starting Parameter Set	No. of parameters calibrated	Measured Outputs	Sensitivity Analysis	Simulation Platform	Model Optimisation Method
Coelho *et al.* (2006)	Continuous Lab-scale	Default STR set	12	CH_4 gas flow, TCOD	Adopted sensitivities as reported in the STR.	AQUASIM	Minimising Chi-square function by first calibrating parameter group with highest sensitivity, then other two groups classified with decending sensitivities.
Boubaker and Ridha (2008)	Semi-continuous Lab-scale	(i) Disintegration and hydrolysis rate constants set according to past literature values. (ii) Stoichiometric parameters determined from substrate chemical composition. (iii) Other parameters follow default STR set	3	Biogas flow, CH_4 & CO_2 gas composition, VFA, NH4, pH	Applied but did not specify which method.	MATLAB/Simulink	Method was not specified. Only the most sensitive parameters were calibrated.
Blumensaat and Keller (2005)	Continuous Pilot-scale	(i) Stoichiometric parameters follow default STR set (ii) Kinetic parameters with low sensitivities set according to past literature values. (iii) Kinetic parameters with low sensitivities follow default STR set	6	Biogas flow, CH_4 & CO_2 gas composition, VFAs including Propionate & Acetate	Adopted sensitivities as reported in the STR.	MATLAB/Simulink	Method was not specified. Only eight of the most sensitive parameters reported in STR were calibrated.
Razaviarani and Buchanan (2015)	Batch Lab-scale & Semi-continuous Lab-scale	(i) Disintegration and kinetic rates obtained from BMP fitting are used (ii) Hydrolysis rates set according to past literature values. (iii) Other parameters follow default STR set	14	Biogas flow, CH_4 & CO_2 gas composition, VSS, Alkalinity, pH, NH_4, COD, VFAs including Acetate, Propionate, Butyrate & Valerate	None	GPS-X	Method was not specified. BMP test results used for calibration. Fitting criteria for certain parameters are specific: (i) Disintegration rate is calibrated by fitting methane gas production while hydrolysis rates are held constant. (ii) Kinetic rates for acidogenesis of monosaccharides & LCFA and acetogenesis of all VFA constituents are calibrated by fitting VFA constituents.

Reference	Experiment Type	Starting Parameter Set	No. of parameters calibrated	Measured Outputs	Sensitivity Analysis	Simulation Platform	Model Optimisation Method
							(iii) Only stoichiometric parameters with high variability as reported in the STR are calibrated.
Koch *et al.* (2010)	Continuous Lab-scale	(i) Fractionation parameters related to disintegration are estimated using detailed substrate composition. (ii) Hydrolysis rate for protein set according to past literature values. (iii) Other parameters follow default STR set	11	Biogas flow, CH_4, CO_2 & H_2 gas composition	None	MATLAB/Simulink	Modified Nash-Sutcliffe coefficient applied to produce goodness-of-fit maps, which assists in identifying the most sensitive parameters to calibrate.
Wichern *et al.* (2009)	Semi-continuous Lab-scale	Default STR set	8	Biogas flow, CH_4, CO_2 & H_2 gas composition, pH, TS, VFAs including Acetate, Propionate & Butyrate	Single Step Variation Method (SVM)	MATLAB/Simulink	(i) Manual calibration based on expert knowledge (ii) Use of Genetic Algorithm in conjuction with Square Error Sum function
Mairet *et al.* (2011)	Semi-continuous Lab-scale	(i) Fractionation parameters related to disintegration are estimated using detailed substrate composition. (ii) Other parameters follow default STR set.	12	Biogas flow, CH_4 gas composition, pH, TCOD, SCOD, NH_4, VFAs including Acetate, Propionate, Butyrate & Valerate	None	MATLAB/Simulink	(i) pH inhibition factor of acetate calibrated manually based on expert knowledge (ii) Hydrolysis rates and half saturation constants of carbohydrates, proteins and lipids calibrated by trial and error
Jurado *et al.* (2016)	Semi-continuous Lab-scale	Default STR set	7	Biogas flow, CH_4 gas composition, pH, VFAs including Acetate, Propionate & Butyrate	None	AQUASIM	Secant method applied in conjunction with Square Error Sum. Calibrated parameters are selected based on expert knowledge
Antonopoulou *et al.* (2012)	Batch Lab-scale & Semi-continuous Lab-scale	(i) Maximum specific uptake rates for hydrogen and VFAs obtained from BMP fitting are used. (ii) Other parameters follow default STR set.	4	Biogas flow, CH_4 gas composition, pH	None	AQUASIM	Secant method applied for BMP fitting. Maximum specific uptake rates for hydrogen and VFAs calibrated against BMP test results
Jeong *et al.* (2005)	Batch Lab-scale	Default STR set	10	Biogas flow, CH_4 gas composition, VFAs including Acetate, Propionate & Butyrate	Dynamic sensitivity analysis	MATLAB	Genetic Algorithms applied in conjuction with Square Error Sum on sensitivity analysis results

Reference	Experiment Type	Starting Parameter Set	No. of parameters calibrated	Measured Outputs	Sensitivity Analysis	Simulation Platform	Model Optimisation Method
Thamsiriroj and Murphy (2011)	Continuous Lab-scale	(i) Stoichiometric parameters estimated from VS content of substrate. (ii) 6 kinetic parameters referenced from another study on similar substrate. (iii) Other kinetic parameters follow default STR set.	1	Biogas flow, CH_4 gas composition, Total VFA, pH, DS & VS content	None	MATLAB/Simulink	Method was not specified.
Barrera *et al.* (2015)	Continuous Lab-scale	Default STR set	4	Biogas flow, CH_4 & CO_2 & H_2S gas composition, pH, COD, VFAs including Acetate & Propionate, SO_4 & H_2S	Local relative sensitivity analysis	MATLAB/Simulink	(i) Select parameters to calibrate based on expert knowledge and sensitivity analysis (ii) Minimise mean absolute relative errors
Derbal *et al.* (2009)	Continuous Full-scale	(i) Disintegration and hydrolysis parameters are estimated from BMP fitting. (ii) Other parameters follow default STR set.	4	Biogas flow, CH_4 & CO_2 gas composition, pH, TCOD & SCOD, Total VFA, NH_4 and Alkalinity	None	Not specified	Method was not specified.
Lübken *et al.* (2007)	Continuous Pilot-scale	Default STR set	11	Biogas flow, CH_4 & CO_2 & H_2 gas composition, pH, VFAs including Acetate & Propionate	None	MATLAB/Simulink	(i) Select parameters to calibrate based on expert knowledge (ii) Referenced some kinetic parameters from another study (iii) Calibrate parameters iteratively. Method was not specified.
Normak *et al.* (2015)	Batch Lab-scale	(i) 6 kinetic parameters referenced from other studies on similar substrate. (ii) Other kinetic parameters follow default STR set.	11	Biogas flow, CH_4 gas composition, pH, Total VFA	None	MATLAB/Simulink	Minimise residual sum of squares of errors based on biogas flow
Ozkan-Yucel & Gökçay (2010)	Continuous Full-scale	Default STR set	10	Biogas flow, pH, Total VFA, TCOD	Applied but did not specify which method.	AQUASIM	Minimise residual sum of squares of errors based on total VFA, pH and biogas flow
Lee *et al.* (2009)	Semi-continuous Lab-scale	Default STR set	4	CH4, TCOD, SCOD, Acetate	Local relative sensitivity analysis	Not specified	Method was not specified.

Reference	Experiment Type	Starting Parameter Set	No. of parameters calibrated	Measured Outputs	Sensitivity Analysis	Simulation Platform	Model Optimisation Method
Chen *et al.* (2016)	Batch Lab-scale	Default STR set	7	Biogas flow, CH_4 & H_2 gas composition, VFAs including Acetate & Propionate	Local relative sensitivity analysis	AQUASIM	(i) Select parameters to calibrate based on sensitivity analysis (ii) Minimise residual sum of squares of errors based on biogas flow
Fezzani & Cheikh (2009)	Semi-continuous Lab-scale	Default STR set	2	Biogas flow, pH, Phenol	Applied but did not specify which method.	MATLAB/Simulink	Method was not specified.
Silva *et al.* (2009)	Semi-continuous Lab-scale	Default STR set	2	Methane flow, pH, Acetic acid	Local relative sensitivity analysis	AQUASIM	Secant method applied only on two most sensitive parameters
Biernacki *et al.* (2013)	Continuous Full-scale	Default STR set	4	Biogas flow	None	MATLAB/Simulink	Simplex method algorithm
Bulkowska *et al.* (2015)	Semi-continuous Lab-scale	Default STR set	10	Biogas flow, CH4 gas composition, pH, VFAs including Acetate, Propionate & Butyrate	None	MATLAB/Simulink	Genetic Algorithms
Ngo *et al.* (2016)	Continuous Lab-scale	Default STR set	17	Methane flow, Soluble COD, Sugar, Protein, Lipids, pH, VFAs including Acetate & Propionate	None	GPS-X	Method was not specified.
Mendes *et al.* (2015)	Continuous Lab-scale	(i) Stoichiometric parameters estimated from substrate composition. (ii) Kinetic parameters follow default STR set.	4	Biogas flow, CH_4 & CO_2 & H_2 gas composition, pH, VFAs including Acetate & Propionate	Local relative sensitivity analysis	MATLAB/Simulink	(i) Select parameters to calibrate based on sensitivity analysis (ii) Minimise residual sum of squares of errors based on methane concentration only
Yu *et al.* (2012)	Continuous Pilot-scale	Default STR set	7	CH_4 gasflow, VFAs including Acetate, Propionate, Butyrate & Valerate	Local relative sensitivity analysis	AQUASIM	(i) Select parameters to calibrate based on sensitivity analysis and expert knowledge (ii) Minimise residual sum of squares of errors
Biernacki *et al.* (2013)	Batch Lab-scale	Default STR set	4	Biogas flow	3-D Graph	MATLAB/Simulink	Simplex method algorithm

Reference	Experiment Type	Starting Parameter Set	No. of parameters calibrated	Measured Outputs	Sensitivity Analysis	Simulation Platform	Model Optimisation Method
Colussi *et al.* (2016)	Batch Lab-scale	Default STR set	4	CH_4 & CO_2 gas flows	None	AQUASIM	Secant method
Galí *et al.* (2009)	Semi-continuous Lab-scale	Default STR set	1	CH_4 gas flow	None	MATLAB/Simulink	Minimise residual sum of squares of errors based on methane flow
Shi *et al.* (2014)	Semi-continuous Lab-scale	(i) 5 kinetic parameters referenced from other studies on similar substrate. (ii) Other kinetic parameters follow default STR set.	3	CH_4 gas flow, pH	Local relative sensitivity analysis	AQUASIM	(i) Select parameters to calibrate based on sensitivity analysis (ii) Minimise residual sum of squares of errors
This study	Continuous Full-scale	Default STR set	All	Total VFA, NH_4, VSS, pH, CH_4 gas flow & CO_2 gas flow	Monte Carlo & Partial Least Squares Regression	SCILAB	(i) Apply PLS regression on Monte Carlo data to identify underlying relationships between parameters and model outputs (ii) Calibrate by manipulating PLS input latent variables instead of parameters (iii) Minimise objective function which is expressed in the form of PLS output latent variables

8.3. Appendix C – Parameter Survey Data

Notes on Parameter Survey Tables

General: For literature references that have applied ADM1 simulation, only parameters that differ from the STR suggested parameters are indicated. In other words, only modified parameters (as a result of model fitting) are noted.

Note 1: Literature references highlighted in yellow are references provided in STR (Batstone et al., 2002).

Note 2: Substrate is tested with excess carbohydrates, protein and lipid compositions in individual experiments. It is assumed that the hydrolysis constants obtained can be represented as k_{hyd_ch}, k_{hyd_pr} and k_{hyd_li} respectively.

Note 3: Single first-order hydrolysis constant was determined to describe lumped effect of disintegration and hydrolysis. The constant is classified under disintegration constant (k_{dis}) based on the assumption that disintegration is the rate limiting step.

References

Reference numbers listed below correspond to the "lit. ref." numbers indicated on the parameter survey tables.

1 Batstone, D.J. (2000). *High-rate Anaerobic Treatment of Complex Wastewater*. Thesis, (PhD). University of Queensland, Brisbane.

2 Boon, F. (1994). *Influence of pH, High Volatile Fatty Acid Concentrations and Partial Hydrogen Pressure on Hydrolysis*. Thesis, (MSc). Wageningen University, Wageningen.

3 Eastman, J.A., Ferguson, J.F. (1981). Solubilization of Particulate Organic Carbon during the Acid Phase of Anaerobic Digestion. W*ater Pollution Control Federation*, **53**, 352-366.

4 Gavala, H.N., Skiadas, I.V. and Lyberatos, G. (1999). On the performance of a centralized digestion facility receiving seasonal agroindustrial wastewaters. *Water Science and Technology*, **40**, 339-346.

5 Gavala, H.N. and Lyberatos, G. (2001). Influence of anaerobic culture acclimation on the degration kinetics of various substrates. *Biotechnology and Bioengineering*, **74**, 181-195.

6 Gujer, W. and Zehnder, A.J.B. (1983). Conversion processes in anaerobic digestion. *Water Science and Technology*, **15**, 127-167.

7 Lokshina, L.Y. and Vavilin, V.A. (1999). Kinetic analysis of the key stages of low temperature methanogenesis. *Ecology Modeling*, **117**, 285-303.

8 Novak, J.T. and Carlson, D.A. (1970). The kinetics of anaerobic long chain fatty acid degradation. *Water Pollution Control Federation*, **42**, 1932-1943.

9 O'Rourke. (1968). *Kinetics of Anaerobic Treatment at Reduced Temperatures*. Thesis, (PhD). Standford University, Standford.

10 Palenzuella Rollon, A. (1999). Anaerobic Digestion of Fish Processing Wastewater with Special Emphasis on Hydrolysis of Suspended Solids. Thesis, (PhD). Wageningen University, Wageningen.

11 Pavlostathis, S.G. and Giraldo-Gomez, E. (1991). Kinetics of anaerobic treatment: A critical review. *Critical Reviews in Environmental Control*, **21**, 411-490.

12 Ramsay, I.R. (1997). Modeling and Control of High-Rate Anaerobic Wastewater Treatment Systems. Thesis, (PhD). University of Queensland, Brisbane.

13 Romli, M., Keller, J., Lee, P.J. and Greenfield, P.F. (1995). Model prediction and verification of a two-stage high-rate anaerobic wastewater treatment system subjected to shock loads. *Process Safety Progress*, **73**, 151-154.

14 Salminen, E., Rintala, J., Lokshina, L. and Vavilin, V.A. (2000). Anaerobic batch degradation of solid poultry slaughterhouse waste. *Water Science and Technology*, **41**, 33-41.

15 Siegrist, H., Vogt, D., Garcia-Heras, J. and Gujer, W. (2002). Mathematical model for meso and thermophilic anaerobic sewage sludge digestion. *Environmental Science and Technology*, **36**, 1113-1123.

16 Skiadas, I.V., Gavala, H.N. and Lyberatos, G. (2000). Modeling of the periodic anaerobic baffled reactor (PABR) based on the retaining factor concept. *Water Research*, **34**, 3691-3905.

17 Stamatelatou, K. (1999). *Optimization of Anaerobic Digestion Systems*. Thesis, (PhD). University of Patras, Patras.

18 Vavilin, V.A. and Lokshina, L.Y. (1996). Modeling of volatile fatty acids degradation kinetics and evaluation of microorganism activity. *Bioresource Technology*, **57**, 69-80.

19 Vavilin, V.A., Lokshina, L.Y., Rytov, S.V., Kotsyurbenko, O.R., Nozhevnikova, A.N. and Parshina, S.N. (1997). Modeling methanogenesis during anaerobic conversion of complex organic matter at low temperatures. *Water Science and Technology*, **36**, 531-538.

20 Vavilin, V.A., Rytov, S.V., Lokshina, L.Ya. and Rintala, J.A. (1999). Description of hydrolysis and acetoclastic methanogenensis as the rate-limiting steps during anaerobic conversion of solid waste into methane. Proceedings of the Second International Symposium on Anaerobic Digestion of Solid Wastes, Barcelona, p.1-4.

21 Coelho, N., Capela, I. and Droste, R.L. (2006). Application of ADM1 to a UASB treating complex wastewater in different feeding regimes. Proceedings of the WEFTEC, Dallas. TX: Water Environment Foundation, p.7123-7135.

22 Boubaker, F. and Ridha, B. (2008). Modeling of the mesophilic anaerobic co-digestion of olive mill wastewater with olive mill solid waste using anaerobic digestion model No. 1 (ADM1). *Bioresource Technology*, **99**, 6565-6577.

23 Blumensaat, F. and Keller, J. (2005). Modeling of two-stage anaerobic digestion using the Anaerobic Digestion Model No. 1 (ADM1). *Water Research*, **39**, 171-183.

24 Razaviarani, V. and Buchanan, I. (2015). Calibration of the Anaerobic Digestion Model No. 1 (ADM1) for steady-state anaerobic co-digestion of municipal wastewater sludge with restaurant grease trap waste. *Chemical Engineering Journal*, **266**, 91-99.

25 Lübken, M., Kosse, P., Koch, K., Gehring, T. and Wichern, M. (2015). Influent fractionation for modeling continuous anaerobic digestion processes. *Advances in Biochemical Engineering/Biotechnology*, **151**, 137-169.

26 Koch, K., Lübken, M., Gehring, T., Wichern, M. and Horn, H. (2010). Biogas from grass silage - Measurements and modeling with ADM1. *Bioresource Technology*, **101**, 8158-8165.

27 Wichern, M., Gehring, T., Fischer, K., Andrade, D., Lübken, M., Koch, K., Gronauer, A. and Horn, H. (2009). Monofermentation of grass silage under mesophilic conditions: Measurements and mathematical modeling with ADM1. *Bioresource Technology*, **100**, 1675-1681.

28 Mairet, F., Bernard, O., Ras, M., Lardon, L. and Steyer, J. (2011). Testing the ability of ADM1 to represent the anaerobic digestion of microalgae. Proceedings of the 8th IWA Symposium on Systems Analysis and Integrated Assessment, p.151-158

29 Jurado, E., Antonopoulou, G., Lyberatos, G., Gavala, H.N. and Skiadas, I.V. (2016). Continuous anaerobic digestion of swine manure: ADM1-based modeling and effect of addition of swine manure fibers pretreated with aqueous ammonia soaking. *Applied Energy*, **172**, 190-198.

30 Antonopoulou, G., Gavala, H.N., Skiadas, I.V. and Lyberatos, G. (2012). ADM1-based modeling of methane production from acidified sweet sorghum extract in a two stage process. *Bioresource Technology*, **106**, 10-19.

31 Jeong, H., Suh, C., Lim, J., Lee, S. and Shin, H. (2005). Analysis and application of ADM1 for anaerobic methane production. *Bioprocess Biosyst Eng*, **27**, 81-89.

32 Thamsiriroj, T. and Murphy, J.D. (2011). Modeling mono-digestion of grass silage in a 2-stage CSTR anaerobic digester using ADM1. *Bioresource Technology*, **102**, 948-959.

33 Barrera, E.L., Spanjers, H., Solon, K., Amerlinck, Y., Nopens, I. and Dewulf, J. (2015). Modeling the anaerobic digestion of cane-molasses vinasse: Extension of the Anaerobic Digestion Model No. 1 (ADM1) with sulfate reduction for a very high strength and sulfate rich wastewater. *Water Research*, **71**, 42-54.

34 Derbal, K., Bencheikh-lehocine, M., Cecchi, F., Meniai, A.-H. and Pavan, P. (2009). Application of the IWA ADM1 model to simulate anaerobic co-digestion of organic waste with waste activated sludge in mesophilic condition. *Bioresource Technology*, **100**, 1539-1543.

35 Lübken, M., Wichern, M., Schlattmann, M., Gronauer, A. and Hom, H. (2007). Modeling the energy balance of an anaerobic digester fed with cattle manure and renewable energy crops. *Water research*, **41**, 4085-4096.

36 Normak, A., Suurpere, J., Suitso, I., Jogi, E., Kokin, E. and Pitk, P. (2015). Improving ADM1 model to simualte anaerobic digestion start-up with inhibition phase based on cattle slurry. *Biomass and Bioenergy*, **80**, 260-266.

37 Ozkan-Yucel, U.G. and Gökcay, C.F. (2010). Application of ADM1 model to a full scale anaerobic digester under dynamic organic loading conditions. *Environmental Technology*, **31**, 633-640.

38 Lee, M.Y., Suh, C.W., Ahn, Y.T. and Shin, H.S. (2009) Variation of ADM1 by using temperature-phased anaerobic digestion (TPAD) operation. *Bioresource Technology*, **100**, 2816-2822.

39 Chen, X., Chen, Z., Wang, X., Huo, C., Hu, Z., Xiao, B. and Hu, M. (2016). Application of ADM1 for modeling of biogas production from anaerobic digestion of Hydrilla verticillata. *Bioresource Technology*, **211**, 101-107.

40 Fezzani, B. and Cheikh, R. (2009). Extension of the anaerobic digestion model No. 1 (ADM1) to include phenol compounds biodegradation processes for simulating the anaerobic co-digestion of olive mill wastes at mesophilic temperature. *Journal of Hazardous Material*s, **172**, 1430-1438.

41 Silva, F., Nadais, H., Prates, A., Arroja, L. and Capela, I. (2009). Modeling of anaerobic treatment of evaporator condensate (EC) from a sulphite pulp mill using the IWA anaerobic digestion model no. 1 (ADM1). *Chemical Engineering Journal*, **148**, 319-326.

42 Koch, K., Wichern, M., Lübken, M. and Horn, H. (2009). Mono fermentation of grass silage by means of loop reactors. *Bioresource Technology*, **100**, 5934-5940.

43 Hu, Z. and Yu, H. (2005). Application of rumen microorganisms for enhanced anaerobic fermentation of corn stover. *Process Biochemistry*, **40**, 2371-2377.

44 Biernacki, P., Steinigeweg, S., Borchert, A. and Uhlenhut, F. (2013). Application of Anaerobic Digestion Model No. 1 for describing anaerobic digestion of grass, maize, green weed silage, and industrial glycerine. *Bioresource Technology*, **127**, 188-194.

45 Bulkowska, K., Bialobrzewski, I., Gusiatin, Z.M., Klimiuk, E. and Pokój, T. (2015). ADM1-based modeling of anaerobic codigestion of maize silage and cattle manure - calibration of parameters and model verification (part II). *Archives of Environmental Protection*, **41**, 20-27.

46 Ngo, A., Nguyen, H., Le, C., Goel, R., Terashima, M. and Yasui, H. (2016). A dynamic simulation of methane fermentation process receiving heterogeneous food wastes and modeling acidic failure. *Journal of Material Cycles and Waste Management*, **18**, 239-247.

47 Rivas-Garćia, P., Álvarez, J.B., Estrada-Baltazar, A. and Navarrete-Bolaños, J.L. (2013). Numerical study of microbial population dynamics in anaerobic digestion through the Anaerobic Digestion Model No. 1 (ADM1). *Chemical Engineering Journal*, **228**, 87-92.

48 Mendes, C., Esquerre, K. and Queiroz, L.M. (2015). Application of Anaerobic Digestion Model No. 1 for simulating anaerobic mesophilic sludge digestion. *Waste Management*, **35**, 89-95.

49 Yu, L., Zhao, Q., Ma, J., Frear, Craig and Chen S. (2012). Experimental and modeling study of a two-stage pilot scale high solid anaerobic digester system. *Bioresource Technology*, **124**, 8-17.

50 Biernacki, P., Steinigeweg, S., Borchert, A. Uhlenhut, F. and Brehm, A. (2013). Application of Anaerobic Digestion Model No. 1 for describing an existing biogas power plant. *Biomass and Bioenergy*, **59**, 441-447.

51 Colussi, I., Cortesi, A., Gallo, V. and Vitanza, R. (2016). Biomethanization of Brewer's Spent Grain evaluated by application of the anaerobic digestion model No. 1. *Environmental Progress & Sustainable Energy*, **35**, 1055-1060.

52 Galí, A., Benabdallah, T., Astals, S. and Mata-Alvarez, J. (2009). Modified version of ADM1 model for agro-waste application. *Bioresource Technology*, **100**, 2783-2790.

53 Shi, X., Yuan, X., Wang, Y., Zeng, S., Qiu, Y., Guo, R. and Wang, L. (2014). Modeling of the methane production and pH value during the anaerobic co-digestion of dairy manure and spent mushroom substrate. *Chemical Engineering Journal*, **244**, 258-263.

54 Rosén, C. and Jeppsson, U. (2006). Aspects on ADM1 implementation within the BSM2 framework. Tech. Report no. LUTEDX/(TEIE-7224)/1-35/(2006). Department of Industrial Electrical Engineering and Automation, Lund University, Lund, Sweden.

55 Neves, L., Goncalo, E., Oliveira, R. and Alves, M.M. (2008). Influence of composition on the biomethanation potential of restaurant waste at mesophilic temperatures. *Waste Management*, **28**, 965-972.

56 Arnell, M., Astals, S., Amand, L., Batstone, D.J., Jensen, P. and Jeppsson, U. (2016). Modeling anaerobic co-digestion in Benchmark Simulation Model No. 2: Parameter estimation, substrate characterisation and plant-wide integration. *Water Research*, **98**, 138-146.

57 Garćia-Heras, J.L. (2002). Reactor sizing process kinetics and modeling of anaerobic digestion of complex wastes. *Biomethanisation of the organic fraction for municipal solid wastes*. International Water Association. London.

58 Shimizu, T., Kudo, K., Nasu, Y. (1993). Anaerobic waste activated sludge digestion - a bioconversion mechanism and kinetic model. *Biotechnology and Bioengineering*, **41**, 1082-1091.

59 Masse, L., Massé, D.I., Kennedy, K.J. and Chou, S.P. (2002). Neutral Fat Hydrolysis and Long-Chain Fatty Acid Oxidation During Anaerobic Digestion of Slaugtherhouse Wastewater. *Biotechnology and Bioengineering*, **79**, 43-52.

60 Liebetrau, J., Kraft, E., Bidlingmaier, W. (2004). The influence of the hydrolysis rate of co-substrates on process behaviour. Proceedings of the Tenth World Congress on Anaerobic, Canadian Association on Water Quality, Montreal, p.1296-1300,

61 Lokshina, L.Y., Vavilin, V.A., Salminen, E. and Rintala, J. (2003). Modeling of Anaerobic Degradation of Solid Slaughterhouse Waste. *Applied Biochemistry and Biotechnology*, **109**, 15-32.

62 Song, H., Clarke, W. (2009). Cellulose hydrolysis by a methanogenic culture enriched from landfill waste in a semi-continuous reactor. *Bioresource Technology*, **100**, 1268-1273.

Stoichiometric Parameters Survey – Parameter 1 to 15

Description					Stoichiometric Parameters														
Lit. Ref.	Substrate Category	Substrate Type	Applied ADM1	Temp	$f_{SI,XC}$	$f_{XI,XC}$	$f_{CH,XC}$	$f_{PR,XC}$	$f_{LI,XC}$	$f_{FA,LI}$	$f_{H2,SU}$	$f_{BU,SU}$	$f_{PRO,SU}$	$f_{AC,SU}$	$f_{H2,AA}$	$f_{VA,AA}$	$f_{BU,AA}$	$f_{PRO,AA}$	$f_{AC,AA}$
				°C															
6	Acetate	Acetate (Max)	No	25-35															
6	Acetate	Acetate (Min)	No	25-35															
11	Acetate	Acetate (Max)	No	25-35															
11	Acetate	Acetate (Min)	No	25-35															
18	Acetate	Acetate	No	37															
7	Agricultural solid wastes	Forest soil	No	30															
7	Agricultural solid wastes	Forest soil (Max)	No	20															
7	Agricultural solid wastes	Forest soil (Min)	No	20															
7	Agricultural solid wastes	Pond silt	No	28															
25	Aquatic culture	Blue algae	No													0.293	0.242	0.041	0.325
25	Aquatic culture	*Chlorella vulgaris*	No													0.278	0.271	0.064	0.301
25	Aquatic culture	*Scenedesmus obliquus*	No													0.275	0.265	0.063	0.315
28	Aquatic culture	Microalgae	Yes			0.3	0.08	0.4	0.22										
39	Aquatic culture	*Hydrilla verticillata*	Yes	35		0.55	0.072	0.313	0.017										
6	Butyrate	Butyrate	No	35															
11	Butyrate	Butyrate (Max)	No	35-60															
11	Butyrate	Butyrate (Min)	No	35-60															
57	Carbohydrates	Carbohydrates (Max)	No	35															
57	Carbohydrates	Carbohydrates (Min)	No	35															
35	Cattle manure + REC	Cattle manure + REC	Yes																
17	Cellulose	Cellulose	No	28															
60	Cellulose	Cellulose[3]	No	35															
62	Cellulose	Cellulose	No																
4	Dairy	Dairy	No	35															
21	Dairy	Dairy	Yes	35															
25	Energy crops	Rye	No													0.272	0.264	0.078	0.328
25	Energy crops	Soybean	No													0.298	0.29	0.063	0.287
25	Energy crops	Sweet potato	No													0.262	0.257	0.055	0.319
30	Energy crops	Sorghum extract	Yes																
43	Energy crops	Corn stover	No	20-40															
52	Energy crops	Rape	Yes		0.122	0.166	0.556	0.126	0.122										
52	Energy crops	Sunflower	Yes		0.184	0.078	0.506	0.198	0.034										
8	Fatty acids	Linoleate	No	37															
8	Fatty acids	Myristate	No	37															
8	Fatty acids	Oleate	No	37															
8	Fatty acids	Palmitate	No	37															
8	Fatty acids	Stearate	No	37															
14	Fatty acids	Slaughterhouse (palmitate)	No	35															
14	Fatty acids	Slaughterhouse (stearate)	No	35															
20	Food wastes	Food waste	No	37															
38	Food wastes	Dog food + Flour	Yes	35															
46	Food wastes	Heterogeneous food wastes	Yes	35															
50	Food wastes	Food waste	Yes			0.21	0.183	0.268	0.338										
55	Food wastes	Restaurant waste[2]	No																
60	Food wastes	Kitchen waste[3]	No	35															
52	Fruit wastes	Apple pulp	Yes		0.422	0.255	0.256	0.011	0.055										
52	Fruit wastes	Orange pulp	Yes		0.153	0.337	0.477	0.02	0.014										
52	Fruit wastes	Pear pulp	Yes		0.367	0.134	0.399	0.016	0.084										
6	General	General/Non-specific	No	33															
6	General	General/Non-specific (Max)	No	34-40															
6	General	General/Non-specific (Min)	No	34-40															
54	General	General/Non-specific	No																

Stoichiometric Parameters Survey – Parameter 1 to 15 (Cont'd)

Description					Stoichiometric Parameters													
Lit. Ref.	Substrate Category	Substrate Type	Applied ADM1	Temp	$f_{SI,XC}$	$f_{XI,XC}$	$f_{CH,XC}$	$f_{PR,XC}$	$f_{LI,XC}$	$f_{FA,LI}$	$f_{H2,SU}$	$f_{BU,SU}$	$f_{PRO,SU}$	$f_{AC,SU}$	$f_{H2,AA}$	$f_{VA,AA}$	$f_{BU,AA}$	$f_{PRO,AA}$	$f_{AC,AA}$
				°C															
6	Glucose	Glucose	No	30															
11	Glucose	Glucose (Max)	No	35-37															
11	Glucose	Glucose (Min)	No	35-37															
16	Glucose	Glucose	No	35															
17	Glucose	Glucose	No																
31	Glucose	Glucose	Yes	35								0.111	0.54	0.202					
11	H_2/butyrate	H_2/butyrate (Max)	No	33-60															
11	H_2/butyrate	H_2/butyrate (Min)	No	33-60															
44	Industrial glycerine	Industrial glycerine	Yes			0.02	0.492	0.01	0.478										
51	Industrial solid wastes	Brewer's spent grain	Yes			0.133	0.403	0.244	0.22										
60	Industrial solid wastes	Biowaste[3]	No	35															
61	Industrial solid wastes	Solid household waste	No	35															
5	Lactose	Lactose	No	35															
57	Lipids	Lipids (Max)	No	35															
57	Lipids	Lipids (Min)	No	35															
58	Lipids	Lipids	No																
59	Lipids	Lipids	No	25															
4	Livestock manure	Piggery	No	35															
19	Livestock manure	Pig manure	No	28															
29	Livestock manure	Swine manure	Yes																
36	Livestock manure	Cattle manure	Yes																
47	Livestock manure	Cattle manure	Yes	35		0.41047	0.2093	0.36047	0.01977										
50	Livestock manure	Cattle manure	Yes			0.509	0.264	0.185	0.042										
50	Livestock manure	Chicken manure	Yes			0.356	0.306	0.219	0.119										
52	Livestock manure	Pig manure	Yes		0.143	0.033	0.461	0.202	0.161										
53	Livestock manure	Dairy manure + SMS	Yes																
13	Molasses	Molasses	No	35															
33	Molasses	Cane-molasses vinasse	Yes																
41	Molasses	EC from SPM with molasses	Yes	35															
2	MWS	Primary sludge	No	35															
3	MWS	Primary sludge	No	35															
3	MWS	Primary sludge	No	35															
3	MWS	Primary sludge	No	35															
9	MWS	Primary sludge (Max)	No	35															
9	MWS	Primary sludge (Min)	No	35															
15	MWS	Primary sludge	No	35															
23	MWS	MWS	Yes	35															
34	MWS	MWS	Yes																
37	MWS	MWS	Yes	35															
48	MWS	MWS	Yes	35															
24	MWS + GTW	MWS + GTW	Yes																
49	OFMSW	OFMSW	Yes	35	0.075	0.075	0.5	0.28	0.068										
4	Olive	Olive-mill	No	35															
22	Olive	Olive	Yes	35		0.45	0.35	0.074	0.1										
40	Olive	OMSW	Yes	37	0.013	0.45	0.35	0.074	0.1										
6	Propionate	Propionate (Max)	No	25-35															
6	Propionate	Propionate (Min)	No	25-35															
18	Propionate	Propionate	No	37															

Stoichiometric Parameters Survey – Parameter 1 to 15 (Cont'd)

Description					Stoichiometric Parameters														
Lit. Ref.	Substrate Category	Substrate Type	Applied ADM1	Temp	$f_{SI,XC}$	$f_{XI,XC}$	$f_{CH,XC}$	$f_{PR,XC}$	$f_{LI,XC}$	$f_{FA,LI}$	$f_{H2,SU}$	$f_{BU,SU}$	$f_{PRO,SU}$	$f_{AC,SU}$	$f_{H2,AA}$	$f_{VA,AA}$	$f_{BU,AA}$	$f_{PRO,AA}$	$f_{AC,AA}$
				°C															
5	Proteins	Gelatin	No	35															
7	Proteins	Proteins	No	28															
12	Proteins	Casein	No	35															
12	Proteins	Casein	No	35															
12	Proteins	Casein	No	35															
12	Proteins	Casein	No	35															
12	Proteins	Casein	No	35															
12	Proteins	Casein	No	35															
25	Proteins	Casein	No													0.299	0.28	0.072	0.279
25	Proteins	Egg	No													0.309	0.26	0.068	0.273
25	Proteins	Galantine	No													0.293	0.186	0.12	0.399
57	Proteins	Proteins (Max)	No	35															
57	Proteins	Proteins (Min)	No	35															
25	Silage	Grass silage	No													0.292	0.236	0.045	0.334
25	Silage	Maize silage	No													0.304	0.236	0.044	0.321
26	Silage	Grass silage	Yes	38		0.379	0.401	0.187	0.033										
27	Silage	Grass silage	Yes																
32	Silage	Grass silage	Yes	37		0.075	0.797	0.095	0.033										
42	Silage	Grass silage	No	38															
44	Silage	Grass silage	Yes			0.21	0.54	0.21	0.04										
44	Silage	Green weed silage	Yes			0.21	0.604	0.135	0.051										
44	Silage	Maize silage	Yes			0.14	0.695	0.11	0.055										
45	Silage	Maize silage + cattle manure	Yes	29															
1	Slaughterhouse	Slaughterhouse	No	33															
10	Slaughterhouse	Fish waste (Max)	No	33															
10	Slaughterhouse	Fish waste (Min)	No	33															
14	Slaughterhouse	Slaughterhouse	No	35															
56	Slaughterhouse	Paunch + Blood + DAF sludge	Yes	35		0.2			0.3										
61	Slaughterhouse	Solid slaughterhouse waste	No	35															

Kinetic Parameters Survey – Parameter 16 to 30

Description					Kinetic Parameters												
Lit. Ref.	Substrate Category	Substrate Type	Applied ADM1	Temp	k_{dis}	k_{hyd_ch}	k_{hyd_pr}	k_{hyd_li}	K_{S_IN}	pH_{UL_acid}	pH_{LL_acid}	k_{m_su}	K_{S_su}	Y_{su}	k_{dec_xsu}	k_{m_aa}	K_{S_aa}
				°C	d^{-1}	d^{-1}	d^{-1}	d^{-1}	$kgCOD.m^{-3}$	-	-	$COD.COD^{-1}d^{-1}$	$kgCOD.m^{-3}$	$COD.COD^{-1}$	d^{-1}	$COD.COD^{-1}d^{-1}$	$kgCOD.m^{-3}$
6	Acetate	Acetate (Max)	No	25-35													
6	Acetate	Acetate (Min)	No	25-35													
11	Acetate	Acetate (Max)	No	25-35													
11	Acetate	Acetate (Min)	No	25-35													
18	Acetate	Acetate	No	37													
7	Agricultural solid wastes	Forest soil	No	30	0.54												
7	Agricultural solid wastes	Forest soil (Max)	No	20	0.09												
7	Agricultural solid wastes	Forest soil (Min)	No	20	0.031												
7	Agricultural solid wastes	Pond silt	No	28	0.013												
25	Aquatic culture	Blue algae	No														
25	Aquatic culture	*Chlorella vulgaris* (green alga)	No														
25	Aquatic culture	*Scenedesmus obliquus* (green alga)	No														
28	Aquatic culture	Microalgae	Yes														
39	Aquatic culture	*Hydrilla verticillata*	Yes	35	0.18		0.62									35	0.58
6	Butyrate	Butyrate	No	35													
11	Butyrate	Butyrate (Max)	No	35-60													
11	Butyrate	Butyrate (Min)	No	35-60													
57	Carbohydrates	Carbohydrates (Max)	No	35		2											
57	Carbohydrates	Carbohydrates (Min)	No	35		0.5											
35	Cattle manure + REC	Cattle manure + REC	Yes			0.31	0.31	0.31		8	6						
17	Cellulose	Cellulose	No	28		0.15											
60	Cellulose	Cellulose[3]	No	35	0.066												
62	Cellulose	Cellulose	No			0.45											
4	Dairy	Dairy	No	35		0.13	0.24										
21	Dairy	Dairy	Yes	35		2.75	0.15					97.961	1.745	0.01		42	
25	Energy crops	Rye	No														
25	Energy crops	Soybean	No														
25	Energy crops	Sweet potato	No														
30	Energy crops	Sorghum extract	Yes														
43	Energy crops	Corn stover	No	20-40		0.94	0.94	0.94									
52	Energy crops	Rape	Yes		0.24												
52	Energy crops	Sunflower	Yes		0.23												
8	Fatty acids	Linoleate	No	37													
8	Fatty acids	Myristate	No	37													
8	Fatty acids	Oleate	No	37													
8	Fatty acids	Palmitate	No	37													
8	Fatty acids	Stearate	No	37													
14	Fatty acids	Slaughterhouse (palmitate)	No	35													
14	Fatty acids	Slaughterhouse (stearate)	No	35													
20	Food wastes	Food waste	No	37	0.41												
38	Food wastes	Dog food + Flour	Yes	35	1												
46	Food wastes	Heterogeneous food wastes	Yes	35								40	0.05		0.01		0.01
50	Food wastes	Food waste	Yes		1.043	1.044	0.233	0.98									
55	Food wastes	Restaurant waste[2]	No			0.32	0.24	0.12									
60	Food wastes	Kitchen waste[3]	No	35	0.34												
52	Fruit wastes	Apple pulp	Yes		0.15												
52	Fruit wastes	Orange pulp	Yes		0.29												
52	Fruit wastes	Pear pulp	Yes		0.18												
6	General	General/Non-specific	No	33													
6	General	General/Non-specific (Max)	No	34-40		0.13	0.03	0.08									
6	General	General/Non-specific (Min)	No	34-40		0.041	0.02	0.4									
54	General	General/Non-specific	No														

Kinetic Parameters Survey – Parameter 16 to 30 (Cont'd)

Description					Kinetic Parameters											
Lit. Ref.	Substrate Category	Substrate Type	Applied ADM1	Temp	k_{dis}	k_{hyd_ch}	k_{hyd_pr}	k_{hyd_li}	K_{s_IN}	pH_{UL_acid}	pH_{LL_acid}	k_{m_su}	K_{s_su}	Y_{su}	k_{dec_xsu}	k_{m_aa}	K_{s_aa}
				°C	d^{-1}	d^{-1}	d^{-1}	d^{-1}	$kgCOD.m^{-3}$	-	-	$COD.COD^{-1}d^{-1}$	$kgCOD.m^{-3}$	$COD.COD^{-1}$	d^{-1}	$COD.COD^{-1}d^{-1}$	$kgCOD.m^{-3}$
6	Glucose	Glucose	No	30								51	0.022	0.14			
11	Glucose	Glucose (Max)	No	35-37								125	0.63	0.17			
11	Glucose	Glucose (Min)	No	35-37								29	0.023	0.01			
16	Glucose	Glucose	No	35								5067	0.049				
17	Glucose	Glucose	No														
31	Glucose	Glucose	Yes	35										0.05			
11	H_2/butyrate	H_2/butyrate (Max)	No	33-60													
11	H_2/butyrate	H_2/butyrate (Min)	No	33-60													
44	Industrial glycerine	Industrial glycerine	Yes		1.3236	1.2516	0.0018	0.0086									
51	Industrial solid wastes	Brewer's spent grain	Yes		0.823	0.941	1.056	0.124									
60	Industrial solid wastes	Biowaste[3]	No	35	0.12												
61	Industrial solid wastes	Solid household waste	No	35	0.95												
5	Lactose	Lactose	No	35		106											
57	Lipids	Lipids (Max)	No	35				0.7									
57	Lipids	Lipids (Min)	No	35				0.1									
58	Lipids	Lipids	No					0.76									
59	Lipids	Lipids	No	25				0.63									
4	Livestock manure	Piggery	No	35		0.28	0.68										
19	Livestock manure	Pig manure	No	28	0.096												
29	Livestock manure	Swine manure	Yes														
36	Livestock manure	Cattle manure	Yes									11.9	4.5			19.8	0.3
47	Livestock manure	Cattle manure	Yes	35													
50	Livestock manure	Cattle manure	Yes		1.54	0.037	0.099	0.225									
50	Livestock manure	Chicken manure	Yes														
52	Livestock manure	Pig manure	Yes		0.17												
53	Livestock manure	Dairy manure + SMS	Yes		0.365		18.23										
13	Molasses	Molasses	No	35						5.5	4	120	1.28	0.07	0.02		
33	Molasses	Cane-molasses vinasse	Yes														
41	Molasses	EC from SPM with molasses	Yes	35													
2	MWS	Primary sludge	No	35			0.2										
3	MWS	Primary sludge	No	35		0.3	0.28										
3	MWS	Primary sludge	No	35		0.41	0.39										
3	MWS	Primary sludge	No	35		0.58	0.58										
9	MWS	Primary sludge (Max)	No	35		1.94	0.1	0.17									
9	MWS	Primary sludge (Min)	No	35		0.21	0.0096	0.0096									
15	MWS	Primary sludge	No	35	0.25					5.5	4.5	27	0.05	0.15	0.8	27	0.05
23	MWS	MWS	Yes	35	1												
34	MWS	MWS	Yes		0.5	1.017	0.3842	0.999									
37	MWS	MWS	Yes	35		1	1	1				35	0.5				
48	MWS	MWS	Yes	35								20.22				41.12	
24	MWS + GTW	MWS + GTW	Yes		0.2	0.75	0.7	2.1				37.4	0.496				
49	OFMSW	OFMSW	Yes	35	0.15												
4	Olive	Olive-mill	No	35		0.19	0.35										
22	Olive	Olive	Yes	35	0.006												
40	Olive	OMSW	Yes	37	0.001												
6	Propionate	Propionate (Max)	No	25-35													
6	Propionate	Propionate (Min)	No	25-35													
18	Propionate	Propionate	No	37													

Kinetic Parameters Survey – Parameter 16 to 30 (Cont'd)

Description					Kinetic Parameters												
Lit. Ref.	Substrate Category	Substrate Type	Applied ADM1	Temp	k_{dis}	k_{hyd_ch}	k_{hyd_pr}	k_{hyd_li}	K_{s_IN}	pH_{UL_acid}	pH_{LL_acid}	k_{m_su}	K_{s_su}	Y_{su}	k_{dec_xsu}	k_{m_aa}	K_{s_aa}
				°C	d^{-1}	d^{-1}	d^{-1}	d^{-1}	$kgCOD.m^{-3}$	-	-	$COD.COD^{-1}d^{-1}$	$kgCOD.m^{-3}$	$COD.COD^{-1}$	d^{-1}	$COD.COD^{-1}d^{-1}$	$kgCOD.m^{-3}$
5	Proteins	Gelatin	No	35			2.7										
7	Proteins	Proteins	No	28			0.12										
12	Proteins	Casein	No	35													
12	Proteins	Casein	No	35						5.5	4					36	1.198
12	Proteins	Casein	No	35						5.5	4					28	1.027
12	Proteins	Casein	No	35						7.2	5					53	1.198
12	Proteins	Casein	No	35													
12	Proteins	Casein	No	35													
25	Proteins	Casein	No														
25	Proteins	Egg	No														
25	Proteins	Galantine	No														
57	Proteins	Proteins (Max)	No	35			0.8										
57	Proteins	Proteins (Min)	No	35			0.25										
25	Silage	Grass silage	No														
25	Silage	Maize silage	No														
26	Silage	Grass silage	Yes	38		0.5	0.8	0.5									
27	Silage	Grass silage	Yes		0.26					8.5	6						
32	Silage	Grass silage	Yes	37	0.05												
42	Silage	Grass silage	No	38		0.6	0.6	0.6									
44	Silage	Grass silage	Yes		1.7433	0.7366	0.0104	0.0149									
44	Silage	Green weed silage	Yes		0.8168	0.6659	0.0014	0.0513									
44	Silage	Maize silage	Yes		0.7705	0.6865	0.2446	0.1216									
45	Silage	Maize silage + cattle manure	Yes	29	0.1												
1	Slaughterhouse	Slaughterhouse	No	33			0.29	0.12									
10	Slaughterhouse	Fish waste (Max)	No	33			0.15										
10	Slaughterhouse	Fish waste (Min)	No	33			0.1										
14	Slaughterhouse	Slaughterhouse	No	35	0.7												
56	Slaughterhouse	Paunch + Blood + DAF sludge	Yes	35	10	0.2	0.3	0.1									
61	Slaughterhouse	Solid slaughterhouse waste	No	35	0.35												

Kinetic Parameters Survey – Parameter 31 to 45

Description					Kinetic Parameters														
Lit. Ref.	Substrate Category	Substrate Type	Applied ADM1	Temp	Y_{aa}	k_{dec_xaa}	k_{m_fa}	K_{s_fa}	Y_{fa}	k_{dec_xfa}	KI_{h2_fa}	k_{m_c4}	K_{s_c4}	Y_{c4}	k_{dec_xc4}	KI_{h2_c4}	k_{m_pro}	K_{s_pro}	Y_{pro}
				°C	$COD.COD^{-1}$	d^{-1}	$COD.COD^{-1}d^{-1}$	$kgCOD.m^{-3}$	$COD.COD^{-1}$	d^{-1}	$kgCOD.m^{-3}$	$COD.COD^{-1}d^{-1}$	$kgCOD.m^{-3}$	$COD.COD^{-1}$	d^{-1}	$kgCOD.m^{-3}$	$COD.COD^{-1}d^{-1}$	$kgCOD.m^{-3}$	$COD.COD^{-1}$
6	Acetate	Acetate (Max)	No	25-35															
6	Acetate	Acetate (Min)	No	25-35															
11	Acetate	Acetate (Max)	No	25-35															
11	Acetate	Acetate (Min)	No	25-35															
18	Acetate	Acetate	No	37															
7	Agricultural solid wastes	Forest soil	No	30															
7	Agricultural solid wastes	Forest soil (Max)	No	20															
7	Agricultural solid wastes	Forest soil (Min)	No	20															
7	Agricultural solid wastes	Pond silt	No	28															
25	Aquatic culture	Blue algae	No																
25	Aquatic culture	*Chlorella vulgaris*	No																
25	Aquatic culture	*Scenedesmus obliquus*	No																
28	Aquatic culture	Microalgae	Yes																
39	Aquatic culture	*Hydrilla verticillata*	Yes	35															
6	Butyrate	Butyrate	No	35								5.6	0.013	0.066	0.027				
11	Butyrate	Butyrate (Max)	No	35-60								14	0.298	0.066	0.027				
11	Butyrate	Butyrate (Min)	No	35-60								5.3	0.012	0.066	0.027				
57	Carbohydrates	Carbohydrates (Max)	No	35															
57	Carbohydrates	Carbohydrates (Min)	No	35															
35	Cattle manure + REC	Cattle manure + REC	Yes									13.7	0.357				5.5	0.392	
17	Cellulose	Cellulose	No	28															
60	Cellulose	Cellulose[3]	No	35															
62	Cellulose	Cellulose	No																
4	Dairy	Dairy	No	35															
21	Dairy	Dairy	Yes	35			7					60					100		
25	Energy crops	Rye	No																
25	Energy crops	Soybean	No																
25	Energy crops	Sweet potato	No																
30	Energy crops	Sorghum extract	Yes									9.1					13		
43	Energy crops	Corn stover	No	20-40															
52	Energy crops	Rape	Yes																
52	Energy crops	Sunflower	Yes																
8	Fatty acids	Linoleate	No	37			10	5.19	0.055	0.01									
8	Fatty acids	Myristate	No	37			1.6	1.23	0.053	0.01									
8	Fatty acids	Oleate	No	37			8.174	9.21	0.054	0.01									
8	Fatty acids	Palmitate	No	37			2.03	0.41	0.054	0.01									
8	Fatty acids	Stearate	No	37			1.88	0.295	0.055	0.01									
14	Fatty acids	Slaughterhouse (palmitate)	No	35			201	0.1	0.004										
14	Fatty acids	Slaughterhouse (stearate)	No	35			363	0.1	0.021										
20	Food wastes	Food waste	No	37															
38	Food wastes	Dog food + Flour	Yes	35													12.5	0.3	
46	Food wastes	Heterogeneous food wastes	Yes	35		0.1		0.024		0.01							16.25	0.02	
50	Food wastes	Food waste	Yes																
55	Food wastes	Restaurant waste[2]	No																
60	Food wastes	Kitchen waste[3]	No	35															
52	Fruit wastes	Apple pulp	Yes																
52	Fruit wastes	Orange pulp	Yes																
52	Fruit wastes	Pear pulp	Yes																
6	General	General/Non-specific	No	33															
6	General	General/Non-specific (Max)	No	34-40															
6	General	General/Non-specific (Min)	No	34-40															
54	General	General/Non-specific	No																

Kinetic Parameters Survey – Parameter 31 to 45 (Cont'd)

Description					Kinetic Parameters													
Lit. Ref.	Substrate Category	Substrate Type	Applied ADM1	Temp	Y_{aa}	k_{dec_xaa}	k_{m_fa}	K_{s_fa}	Y_{fa}	k_{dec_xfa}	KI_{h2_fa}	k_{m_c4}	K_{s_c4}	Y_{c4}	k_{dec_xc4}	KI_{h2_c4}	k_{m_pro}	K_{s_pro}	Y_{pro}
				°C	$COD.COD^{-1}$	d^{-1}	$COD.COD^{-1}d^{-1}$	$kgCOD.m^{-3}$	$COD.COD^{-1}$	d^{-1}	$kgCOD.m^{-3}$	$COD.COD^{-1}d^{-1}$	$kgCOD.m^{-3}$	$COD.COD^{-1}$	d^{-1}	$kgCOD.m^{-3}$	$COD.COD^{-1}d^{-1}$	$kgCOD.m^{-3}$	$COD.COD^{-1}$
6	Glucose	Glucose	No	30															
11	Glucose	Glucose (Max)	No	35-37															
11	Glucose	Glucose (Min)	No	35-37															
16	Glucose	Glucose	No	35													19	0.021	
17	Glucose	Glucose	No																
31	Glucose	Glucose	Yes	35										0.0193				0.582	0.052
11	H_2/butyrate	H_2/butyrate (Max)	No	33-60															
11	H_2/butyrate	H_2/butyrate (Min)	No	33-60															
44	Industrial glycerine	Industrial glycerine	Yes																
51	Industrial solid wastes	Brewer's spent grain	Yes																
60	Industrial solid wastes	Biowaste[3]	No	35															
61	Industrial solid wastes	Solid household waste	No	35															
5	Lactose	Lactose	No	35															
57	Lipids	Lipids (Max)	No	35															
57	Lipids	Lipids (Min)	No	35															
58	Lipids	Lipids	No																
59	Lipids	Lipids	No	25															
4	Livestock manure	Piggery	No	35															
19	Livestock manure	Pig manure	No	28															
29	Livestock manure	Swine manure	Yes				0.93					13.1					6.56		
36	Livestock manure	Cattle manure	Yes									12.2	0.6				3.5	0.4	
47	Livestock manure	Cattle manure	Yes	35															
50	Livestock manure	Cattle manure	Yes																
50	Livestock manure	Chicken manure	Yes																
52	Livestock manure	Pig manure	Yes																
53	Livestock manure	Dairy manure + SMS	Yes																
13	Molasses	Molasses	No	35								41	0.28	0.066	0.03	0.000008	15	0.373	0.055
33	Molasses	Cane-molasses vinasse	Yes														16		
41	Molasses	EC from SPM with molasses	Yes	35															
2	MWS	Primary sludge	No	35															
3	MWS	Primary sludge	No	35															
3	MWS	Primary sludge	No	35															
3	MWS	Primary sludge	No	35															
9	MWS	Primary sludge (Max)	No	35															
9	MWS	Primary sludge (Min)	No	35															
15	MWS	Primary sludge	No	35	0.15	0.8	12	1	0.045	0.06	0.000003						11	0.02	0.05
23	MWS	MWS	Yes	35													9	0.2	
34	MWS	MWS	Yes																
37	MWS	MWS	Yes	35								5					2.2		
48	MWS	MWS	Yes	35															
24	MWS + GTW	MWS + GTW	Yes				5.9	0.3815				14.1	0.193				17.1	0.0635	
49	OFMSW	OFMSW	Yes	35															
4	Olive	Olive-mill	No	35															
22	Olive	Olive	Yes	35															
40	Olive	OMSW	Yes	37															
6	Propionate	Propionate (Max)	No	25-35													0.31	1.146	0.05
6	Propionate	Propionate (Min)	No	25-35													0.16	0.06	0.025
18	Propionate	Propionate	No	37													23	0.151	0.019

Kinetic Parameters Survey – Parameter 31 to 45 (Cont'd)

Description					Kinetic Parameters													
Lit. Ref.	Substrate Category	Substrate Type	Applied ADM1	Temp	Y_{aa}	k_{dec_xaa}	k_{m_fa}	K_{s_fa}	Y_{fa}	k_{dec_xfa}	KI_{h2_fa}	k_{m_c4}	K_{s_c4}	Y_{c4}	k_{dec_xc4}	KI_{h2_c4}	k_{m_pro}	K_{s_pro}	Y_{pro}
				°C	$COD.COD^{-1}$	d^{-1}	$COD.COD^{-1}d^{-1}$	$kgCOD.m^{-3}$	$COD.COD^{-1}$	d^{-1}	$kgCOD.m^{-3}$	$COD.COD^{-1}d^{-1}$	$kgCOD.m^{-3}$	$COD.COD^{-1}$	d^{-1}	$kgCOD.m^{-3}$	$COD.COD^{-1}d^{-1}$	$kgCOD.m^{-3}$	$COD.COD^{-1}$
5	Proteins	Gelatin	No	35															
7	Proteins	Proteins	No	28															
12	Proteins	Casein	No	35													20	0.056	0.055
12	Proteins	Casein	No	35	0.085	0.02													
12	Proteins	Casein	No	35	0.085	0.02													
12	Proteins	Casein	No	35	0.058	0.02													
12	Proteins	Casein	No	35								22	0.062	0.055	0.03	0.000008			
12	Proteins	Casein	No	35								32	0.08	0.066	0.03	0.000008			
25	Proteins	Casein	No																
25	Proteins	Egg	No																
25	Proteins	Galantine	No																
57	Proteins	Proteins (Max)	No	35															
57	Proteins	Proteins (Min)	No	35															
25	Silage	Grass silage	No																
25	Silage	Maize silage	No																
26	Silage	Grass silage	Yes	38												5.0E-08			
27	Silage	Grass silage	Yes													5.4E-08	13		
32	Silage	Grass silage	Yes	37								13.7	0.357				5.5	0.392	
42	Silage	Grass silage	No	38															
44	Silage	Grass silage	Yes																
44	Silage	Green weed silage	Yes																
44	Silage	Maize silage	Yes																
45	Silage	Maize silage + cattle manure	Yes	29									0.23			1.0E-08	8.5	0.15	
1	Slaughterhouse	Slaughterhouse	No	33															
10	Slaughterhouse	Fish waste (Max)	No	33															
10	Slaughterhouse	Fish waste (Min)	No	33															
14	Slaughterhouse	Slaughterhouse	No	35															
56	Slaughterhouse	Paunch + Blood + DAF sludge	Yes	35															
61	Slaughterhouse	Solid slaughterhouse waste	No	35								7.04	0.036	0.039			5.76	0.08	0.075

Surveyed Kinetic Parameters – Parameter 46 to 58

Description					Kinetic Parameters															
Lit. Ref.	Substrate Category	Substrate Type	Applied ADM1	Temp	k_{dec_xpro}	KI_{h2_pro}	k_{m_ac}	K_{s_ac}	Y_{ac}	k_{dec_xac}	KI_{nh3_ac}	pH_{UL_ac}	pH_{LL_ac}	k_{m_h2}	K_{s_h2}	Y_{h2}	k_{dec_xh2}	pH_{UL_h2}	pH_{LL_h2}
				°C	d^{-1}	$kgCOD.m^{-3}$	$COD.COD^{-1}d^{-1}$	$kgCOD.m^{-3}$	$COD.COD^{-1}$	d^{-1}	$kgCOD.m^{-3}$			$COD.COD^{-1}d^{-1}$	$kgCOD.m^{-3}$	$COD.COD^{-1}$	d^{-1}		
6	Acetate	Acetate (Max)	No	25-35			6.2	0.93	0.076	0.036									
6	Acetate	Acetate (Min)	No	25-35			3.1	0.028	0.032	0.012									
11	Acetate	Acetate (Max)	No	25-35			19	0.93	0.076	0.004									
11	Acetate	Acetate (Min)	No	25-35			3.4	0.011	0.014	0.036									
18	Acetate	Acetate	No	37			7.9	0.213	0.038										
7	Agricultural solid wastes	Forest soil	No	30															
7	Agricultural solid wastes	Forest soil (Max)	No	20															
7	Agricultural solid wastes	Forest soil (Min)	No	20															
7	Agricultural solid wastes	Pond silt	No	28															
25	Aquatic culture	Blue algae	No																
25	Aquatic culture	*Chlorella vulgaris*	No																
25	Aquatic culture	*Scenedesmus obliquus*	No																
28	Aquatic culture	Microalgae	Yes										5.2						
39	Aquatic culture	*Hydrilla verticillata*	Yes	35			5	0.26						28					
6	Butyrate	Butyrate	No	35															
11	Butyrate	Butyrate (Max)	No	35-60															
11	Butyrate	Butyrate (Min)	No	35-60															
57	Carbohydrates	Carbohydrates (Max)	No	35															
57	Carbohydrates	Carbohydrates (Min)	No	35															
35	Cattle manure + REC	Cattle manure + REC	Yes				7.1								3.0E-05				
17	Cellulose	Cellulose	No	28															
60	Cellulose	Cellulose[3]	No	35															
62	Cellulose	Cellulose	No																
4	Dairy	Dairy	No	35															
21	Dairy	Dairy	Yes	35			42.788	0.457						88.89					
25	Energy crops	Rye	No																
25	Energy crops	Soybean	No																
25	Energy crops	Sweet potato	No																
30	Energy crops	Sorghum extract	Yes				5							2					
43	Energy crops	Corn stover	No	20-40															
52	Energy crops	Rape	Yes																
52	Energy crops	Sunflower	Yes																
8	Fatty acids	Linoleate	No	37															
8	Fatty acids	Myristate	No	37															
8	Fatty acids	Oleate	No	37															
8	Fatty acids	Palmitate	No	37															
8	Fatty acids	Stearate	No	37															
14	Fatty acids	Slaughterhouse (palmitate)	No	35															
14	Fatty acids	Slaughterhouse (stearate)	No	35															
20	Food wastes	Food waste	No	37															
38	Food wastes	Dog food + Flour	Yes	35			6.5												
46	Food wastes	Heterogeneous food wastes	Yes	35	0.001	4.0E-06	19	0.1		0.001				33.3	1.00E-06		0.001		
50	Food wastes	Food waste	Yes																
55	Food wastes	Restaurant waste[2]	No																
60	Food wastes	Kitchen waste[3]	No	35															
52	Fruit wastes	Apple pulp	Yes																
52	Fruit wastes	Orange pulp	Yes																
52	Fruit wastes	Pear pulp	Yes																
6	General	General/Non-specific	No	33										25	0.0006	0.056			
6	General	General/Non-specific (Max)	No	34-40															
6	General	General/Non-specific (Min)	No	34-40															
54	General	General/Non-specific	No																

Surveyed Kinetic Parameters – Parameter 46 to 58 (Cont'd)

Description					Kinetic Parameters														
Lit. Ref.	Substrate Category	Substrate Type	Applied ADM1	Temp	k_{dec_xpro}	KI_{h2_pro}	k_{m_ac}	K_{s_ac}	Y_{ac}	k_{dec_xac}	KI_{nh3_ac}	pH_{UL_ac}	pH_{LL_ac}	k_{m_h2}	K_{s_h2}	Y_{h2}	k_{dec_xh2}	pH_{UL_h2}	pH_{LL_h2}
				°C	d^{-1}	$kgCOD.m^{-3}$	$COD.COD^{-1}.d^{-1}$	$kgCOD.m^{-3}$	$COD.COD^{-1}$	d^{-1}	$kgCOD.m^{-3}$			$COD.COD^{-1}.d^{-1}$	$kgCOD.m^{-3}$	$COD.COD^{-1}$	d^{-1}		
6	Glucose	Glucose	No	30															
11	Glucose	Glucose (Max)	No	35-37															
11	Glucose	Glucose (Min)	No	35-37															
16	Glucose	Glucose	No	35			48	0.034											
17	Glucose	Glucose	No				6.4	0.035	0.063	0.02									
31	Glucose	Glucose	Yes	35				0.259	0.1							0.0282			
11	H_2/butyrate	H_2/butyrate (Max)	No	33-60										64	0.0006	0.183	0.009		
11	H_2/butyrate	H_2/butyrate (Min)	No	33-60										1.68	0.000018	0.014			
44	Industrial glycerine	Industrial glycerine	Yes																
51	Industrial solid wastes	Brewer's spent grain	Yes																
60	Industrial solid wastes	Biowaste[3]	No	35															
61	Industrial solid wastes	Solid household waste	No	35															
5	Lactose	Lactose	No	35															
57	Lipids	Lipids (Max)	No	35															
57	Lipids	Lipids (Min)	No	35															
58	Lipids	Lipids	No																
59	Lipids	Lipids	No	25															
4	Livestock manure	Piggery	No	35															
19	Livestock manure	Pig manure	No	28															
29	Livestock manure	Swine manure	Yes				45.02												
36	Livestock manure	Cattle manure	Yes				11.1	0.5			0.0223								
47	Livestock manure	Cattle manure	Yes	35															
50	Livestock manure	Cattle manure	Yes																
50	Livestock manure	Chicken manure	Yes																
52	Livestock manure	Pig manure	Yes																
53	Livestock manure	Dairy manure + SMS	Yes				16.34												
13	Molasses	Molasses	No	35	0.01	0.000008	9.4	0.384	0.048	0.02		7	6	43	0.000088	0.06	0.009	6	5
33	Molasses	Cane-molasses vinasse	Yes				12							43		0.07			
41	Molasses	EC from SPM with molasses	Yes	35			13.2	0.06											
2	MWS	Primary sludge	No	35															
3	MWS	Primary sludge	No	35															
3	MWS	Primary sludge	No	35															
3	MWS	Primary sludge	No	35															
9	MWS	Primary sludge (Max)	No	35															
9	MWS	Primary sludge (Min)	No	35															
15	MWS	Primary sludge	No	35	0.06	1.0E-06	13	0.04	0.025	0.05	0.0012	6.7	5.8	44	0.000001	0.045	0.3	6.7	5.8
23	MWS	MWS	Yes	35		3.5E-06	9	0.15			0.0011				7.00E-06				
34	MWS	MWS	Yes																
37	MWS	MWS	Yes	35			10	0.18								0.05			
48	MWS	MWS	Yes	35			13.8							26.01					
24	MWS + GTW	MWS + GTW	Yes				10.9	0.0961											
49	OFMSW	OFMSW	Yes	35			8.16	0.026											
4	Olive	Olive-mill	No	35															
22	Olive	Olive	Yes	35			9	0.65			0.0028								
40	Olive	OMSW	Yes	37															
6	Propionate	Propionate (Max)	No	25-35	0.041														
6	Propionate	Propionate (Min)	No	25-35	0.01														
18	Propionate	Propionate	No	37			19	0.107	0.027										
5	Proteins	Gelatin	No	35															
7	Proteins	Proteins	No	28															
12	Proteins	Casein	No	35	0.01	0.000008	8.4	0.096	0.048	0.02		7	6						
12	Proteins	Casein	No	35															
12	Proteins	Casein	No	35															
12	Proteins	Casein	No	35															
12	Proteins	Casein	No	35															
12	Proteins	Casein	No	35															
25	Proteins	Casein	No																
25	Proteins	Egg	No																
25	Proteins	Galantine	No																
57	Proteins	Proteins (Max)	No	35															
57	Proteins	Proteins (Min)	No	35															

Surveyed Kinetic Parameters – Parameter 46 to 58 (Cont'd)

Description					Kinetic Parameters															
Lit. Ref.	Substrate Category	Substrate Type	Applied ADM1	Temp	k_{dec_xpro}	KI_{h2_pro}	k_{m_ac}	K_{s_ac}	Y_{ac}	k_{dec_xac}	KI_{nh3_ac}	pH_{UL_ac}	pH_{LL_ac}	k_{m_h2}	K_{s_h2}	Y_{h2}	k_{dec_xh2}	pH_{UL_h2}	pH_{LL_h2}
				°C	d^{-1}	$kgCOD.m^{-3}$	$COD.COD^{-1}.d^{-1}$	$kgCOD.m^{-3}$	$COD.COD^{-1}$	d^{-1}	$kgCOD.m^{-3}$			$COD.COD^{-1}.d^{-1}$	$kgCOD.m^{-3}$	$COD.COD^{-1}$	d^{-1}		
5	Proteins	Gelatin	No	35															
7	Proteins	Proteins	No	28															
12	Proteins	Casein	No	35	0.01	0.000008	8.4	0.096	0.048	0.02		7	6						
12	Proteins	Casein	No	35															
12	Proteins	Casein	No	35															
12	Proteins	Casein	No	35															
12	Proteins	Casein	No	35															
12	Proteins	Casein	No	35															
25	Proteins	Casein	No																
25	Proteins	Egg	No																
25	Proteins	Galantine	No																
57	Proteins	Proteins (Max)	No	35															
57	Proteins	Proteins (Min)	No	35															
25	Silage	Grass silage	No																
25	Silage	Maize silage	No																
26	Silage	Grass silage	Yes	38		4.6E-08	4.4								5.60E-05				
27	Silage	Grass silage	Yes			4.8E-08					0.0084				4.20E-05				
32	Silage	Grass silage	Yes	37			7.1								3.0E-05				
42	Silage	Grass silage	No	38															
44	Silage	Grass silage	Yes																
44	Silage	Green weed silage	Yes																
44	Silage	Maize silage	Yes																
45	Silage	Maize silage + cattle manure	Yes	29		2.4E-08	7.64	0.6			0.00026								
1	Slaughterhouse	Slaughterhouse	No	33															
10	Slaughterhouse	Fish waste (Max)	No	33															
10	Slaughterhouse	Fish waste (Min)	No	33															
14	Slaughterhouse	Slaughterhouse	No	35															
56	Slaughterhouse	Paunch + Blood + DAF sludge	Yes	35															
61	Slaughterhouse	Solid slaughterhouse waste	No	35			14.5	0.085	0.027					208.5	9.60E-06	0.009			

8.4. Appendix D – Plant Data

This section collates all raw data gathered from the industrial-scale AD plant. Original measurements are represented in *black* font, and interpolated/calculated data are represented in *blue* font.

All missing measurements, except for TSS and VSS, are subjected to interpolation. Interpolations are estimated as averages according to a "*4 days earlier and 3 days future"* basis.

Missing TSS and VSS values are calculated according to a pre-determined VSS/TSS and PCOD/VSS ratios, respectively. These ratios are estimated from specific days where data sets are more comprehensive i.e. consist of TCOD, SCOD, TSS and VSS.

Digester Feed (Stream 1) – Day 1 - 60

(Removed due to confidentiality reasons)

Digester Feed (Stream 1) – Day 61 - 150

(Removed due to confidentiality reasons)

Digester Feed (Stream 1) – Day 151 - 240

(Removed due to confidentiality reasons)

Digester Feed (Stream 1) – Day 241 - 325

(Removed due to confidentiality reasons)

Reactor Content (Stream 4) and Product (Stream 6) – Day 1 - 90

(Removed due to confidentiality reasons)

Reactor Content (Stream 4) and Product (Stream 6) – Day 91 - 180
(Removed due to confidentiality reasons)

Reactor Content (Stream 4) and Product (Stream 6) – Day 181 - 270

(Removed due to confidentiality reasons)

Reactor Content (Stream 4) and Product (Stream 6) – Day 271 - 325

(Removed due to confidentiality reasons)

Biogas (Stream 2) and Excess Sludge (Stream 3) – Day 1 - 85

(Removed due to confidentiality reasons)

Biogas (Stream 2) and Excess Sludge (Stream 3) – Day 86 - 170

(Removed due to confidentiality reasons)

Biogas (Stream 2) and Excess Sludge (Stream 3) – Day 171 - 255

(Removed due to confidentiality reasons)

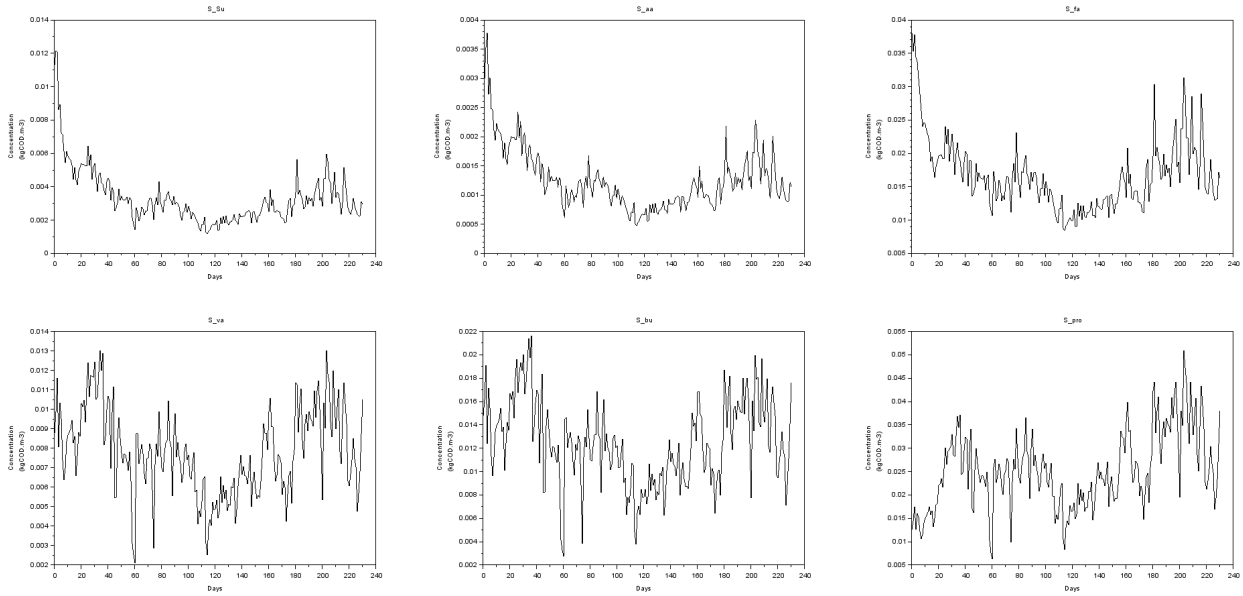
Biogas (Stream 2) and Excess Sludge (Stream 3) – Day 256 - 325

8.5. Appendix E – Supplementary Graphs

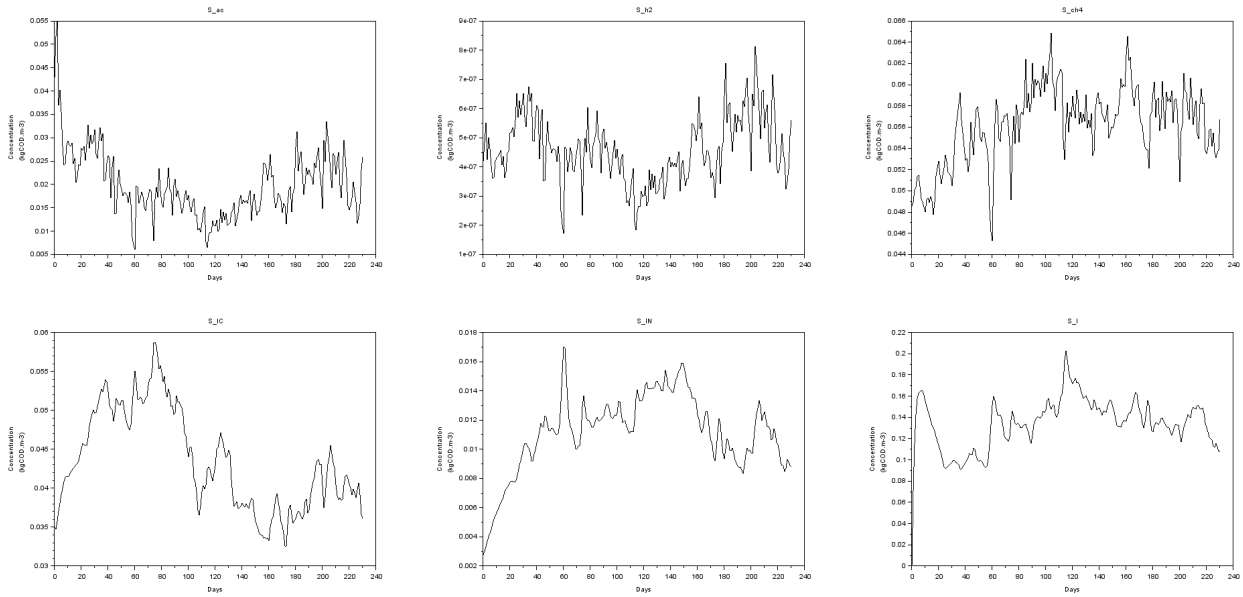
8.5.1. Simulation using Default Parameters

This section presents evolution plots of 27 state variables when ADM1 is executed using default parameters. Anions and cations are not plotted.

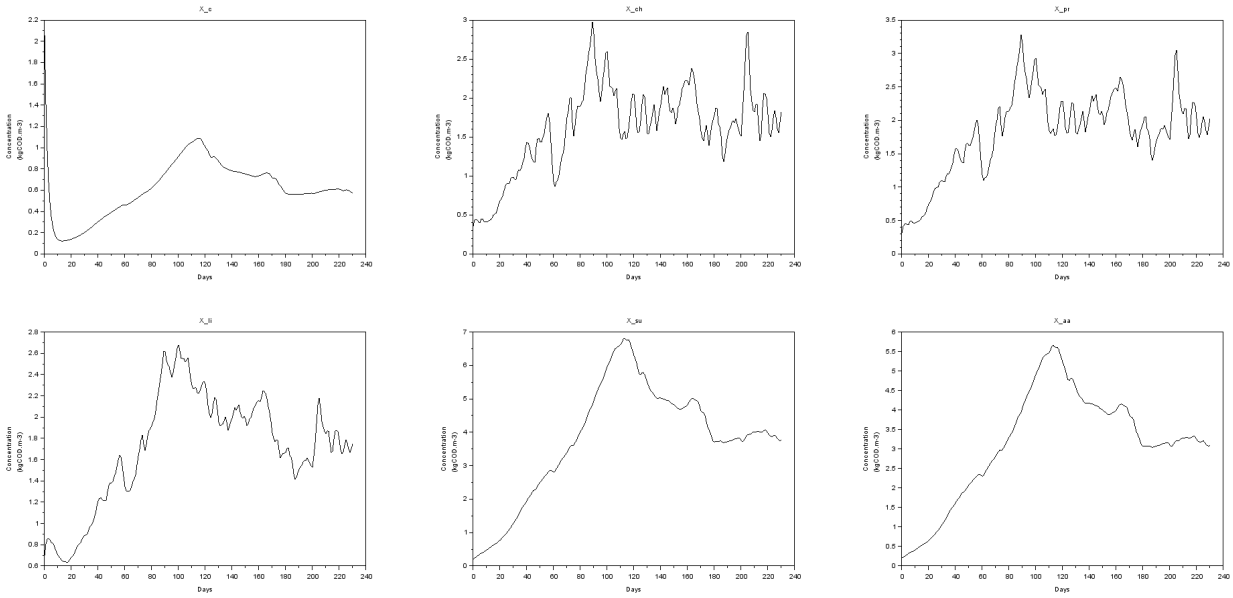
State variables 1 – 6: S_{su}, S_{aa}, S_{fa}, S_{va}, S_{bu}, S_{pro}:



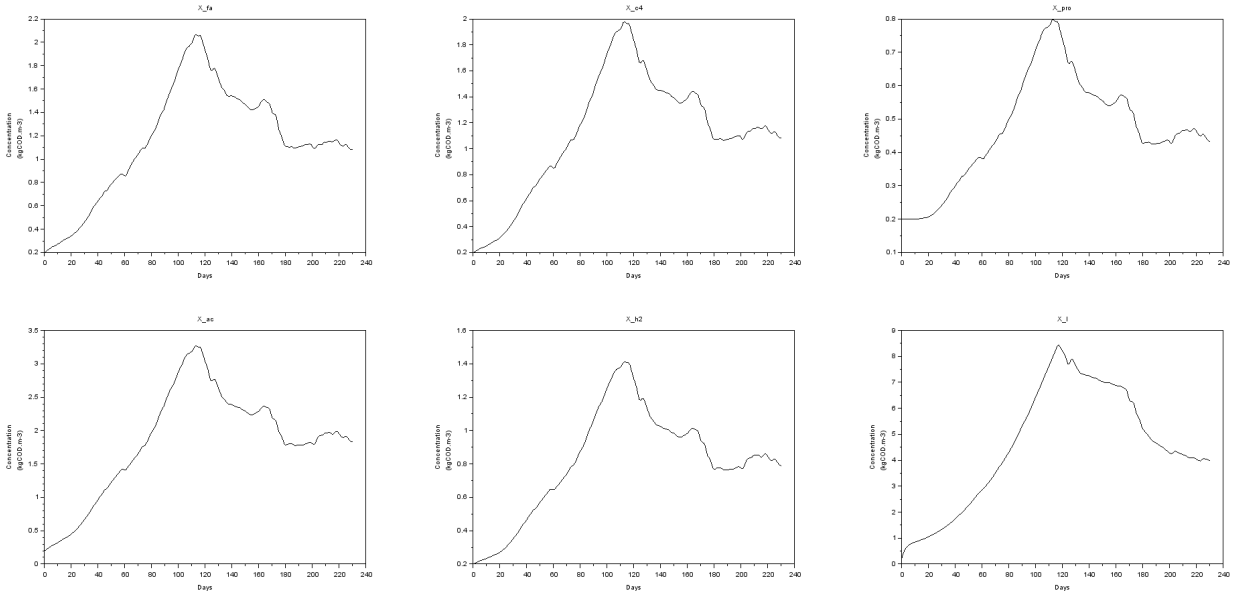
State variables 7 – 12: S_{ac}, S_{h2}, S_{ch4}, S_{IC}, S_{IN}, S_I:



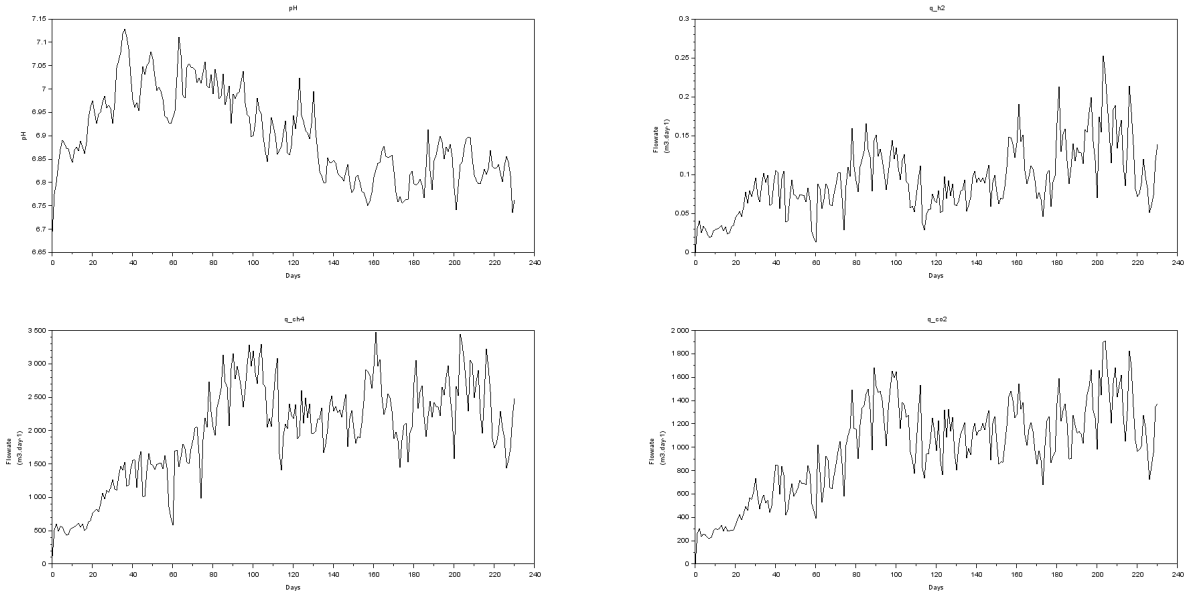
State variables 13 – 18: X_c, X_{ch}, X_{pr}, X_{li}, X_{su}, X_{aa}:



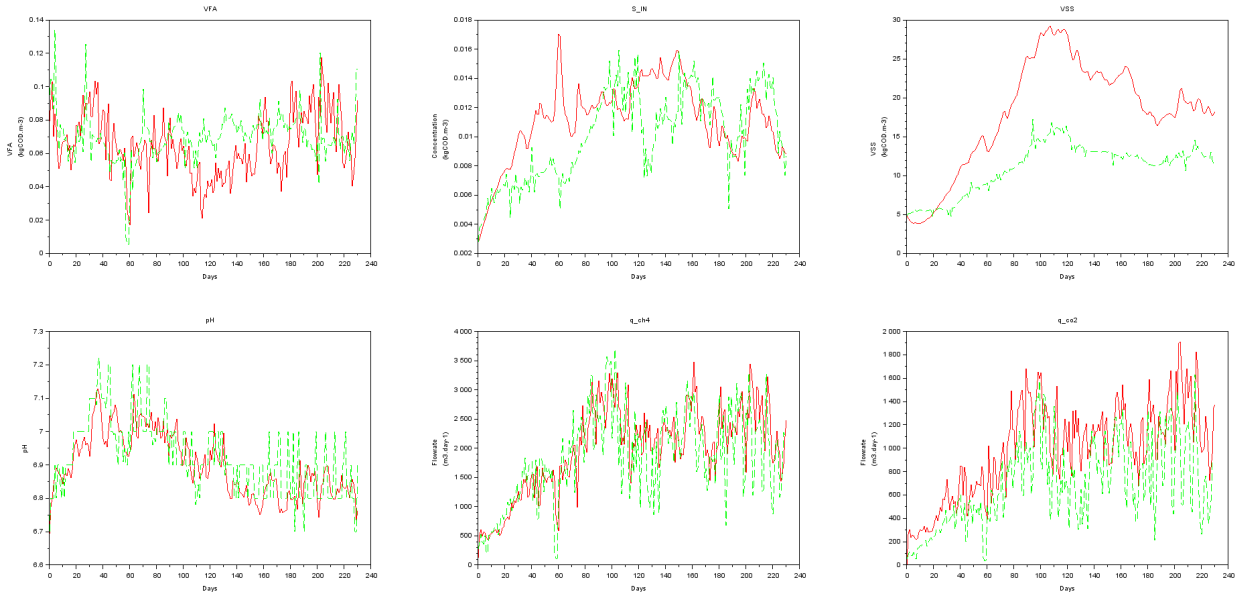
State variables 19 – 24: X_{fa}, X_{c4}, X_{pro}, X_{ac}, X_{h2}, X_I:



State variables 27 – 29: q_{h2}, q_{ch4}, q_{co2}:

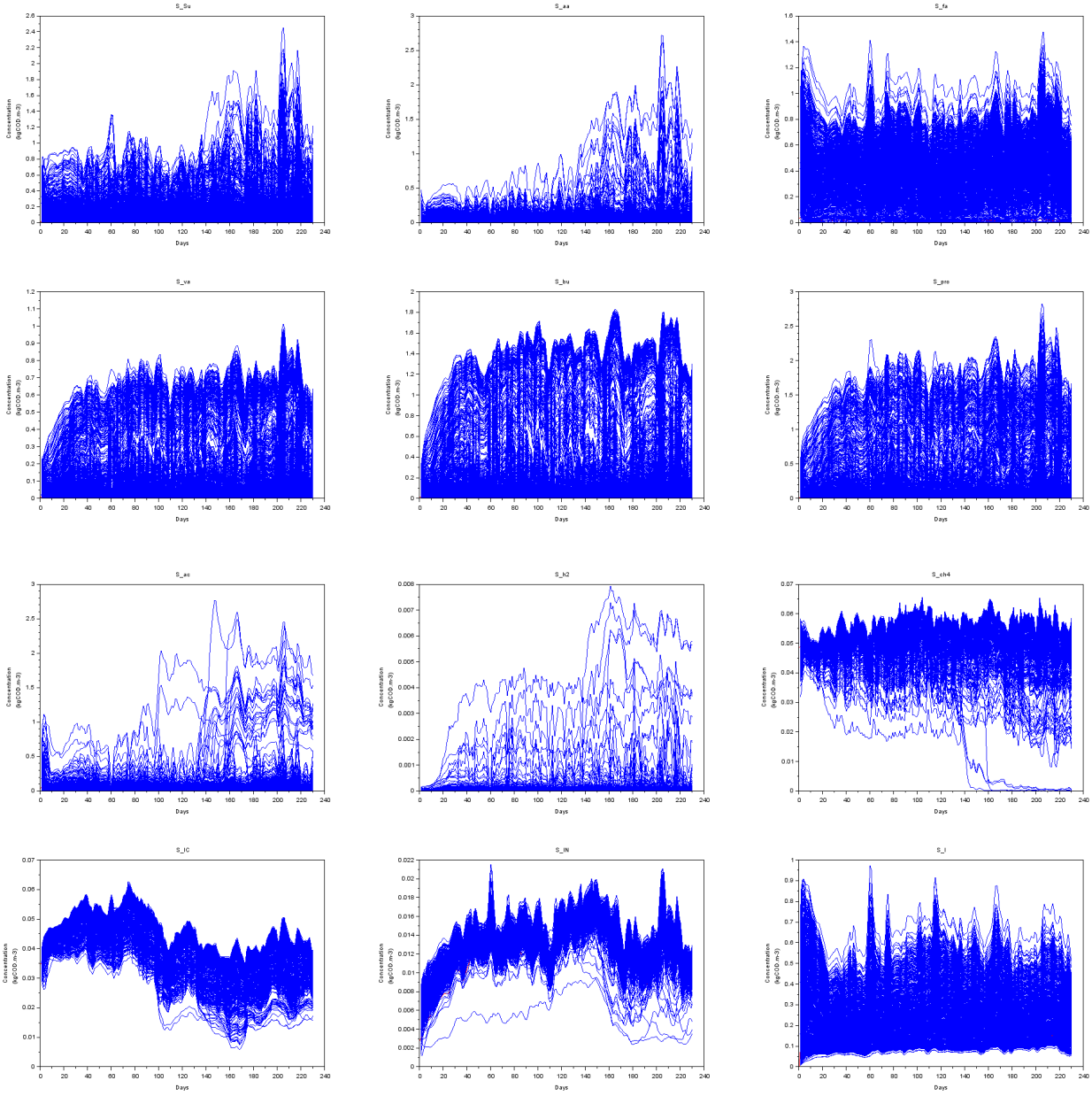


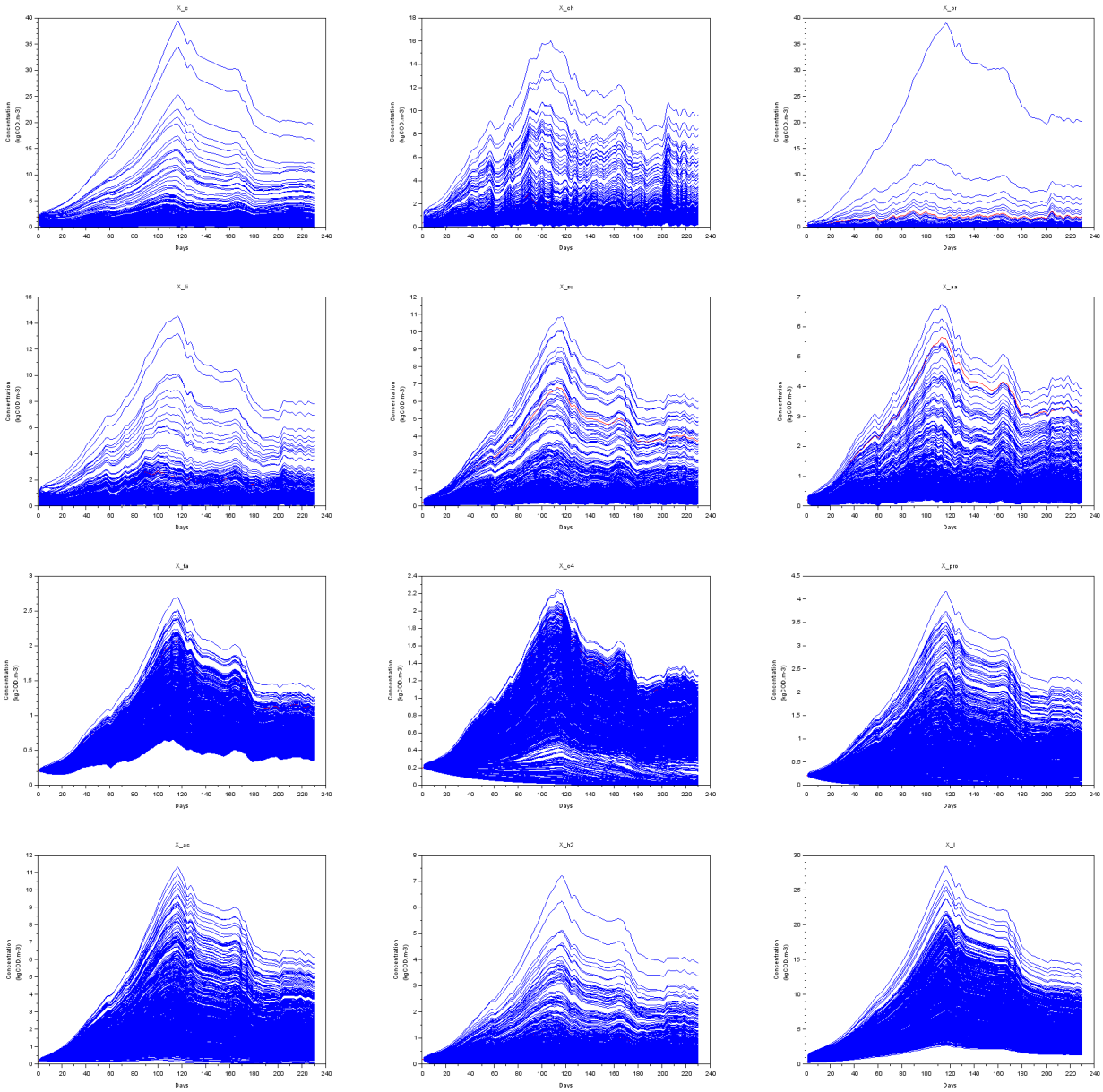
Plots of 6 plant measurements (green line) versus simulated results (red line):

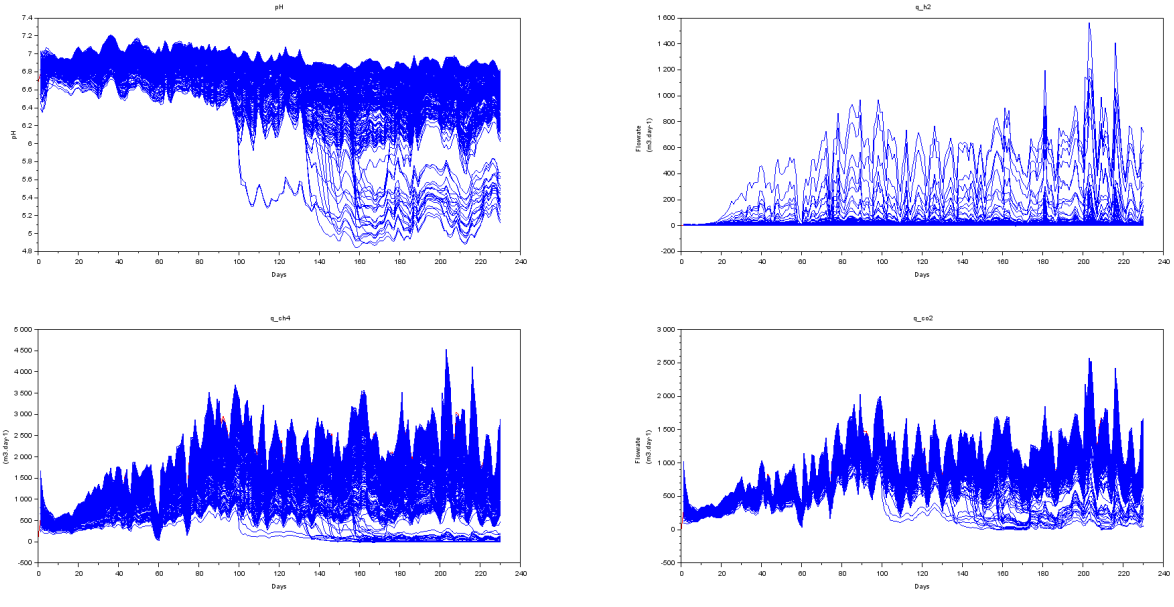


8.5.2. Monte Carlo Graphs

This section presents the 1000 runs Monte Carlo plots of all 26 state variables.







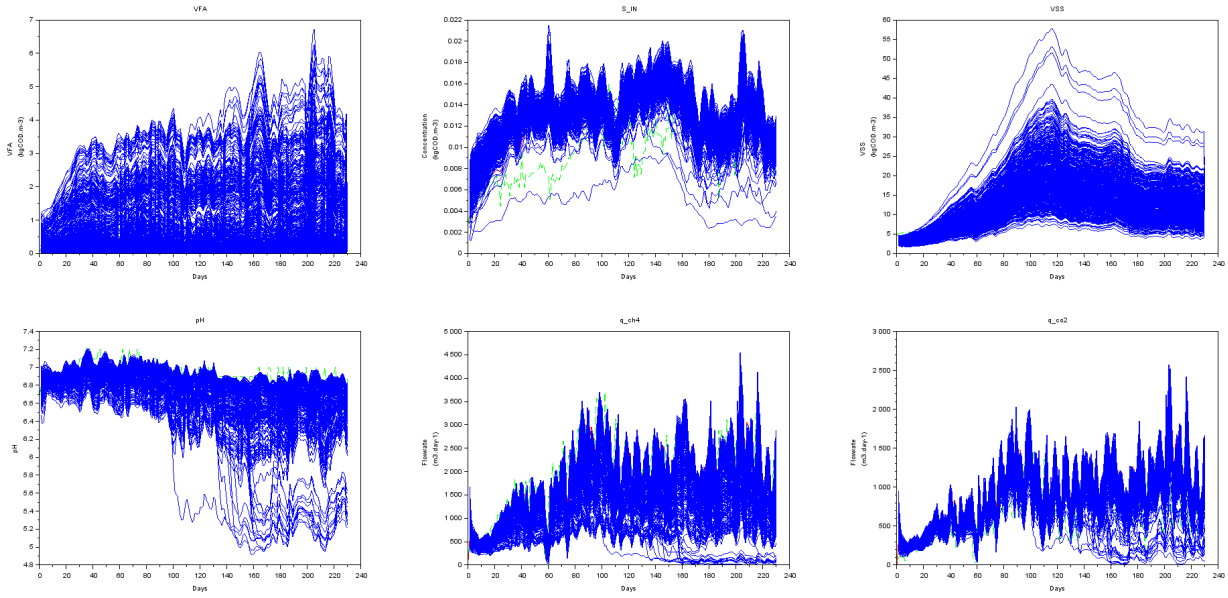


Figure 30: Plot showing 500 Monte Carlo simulation runs

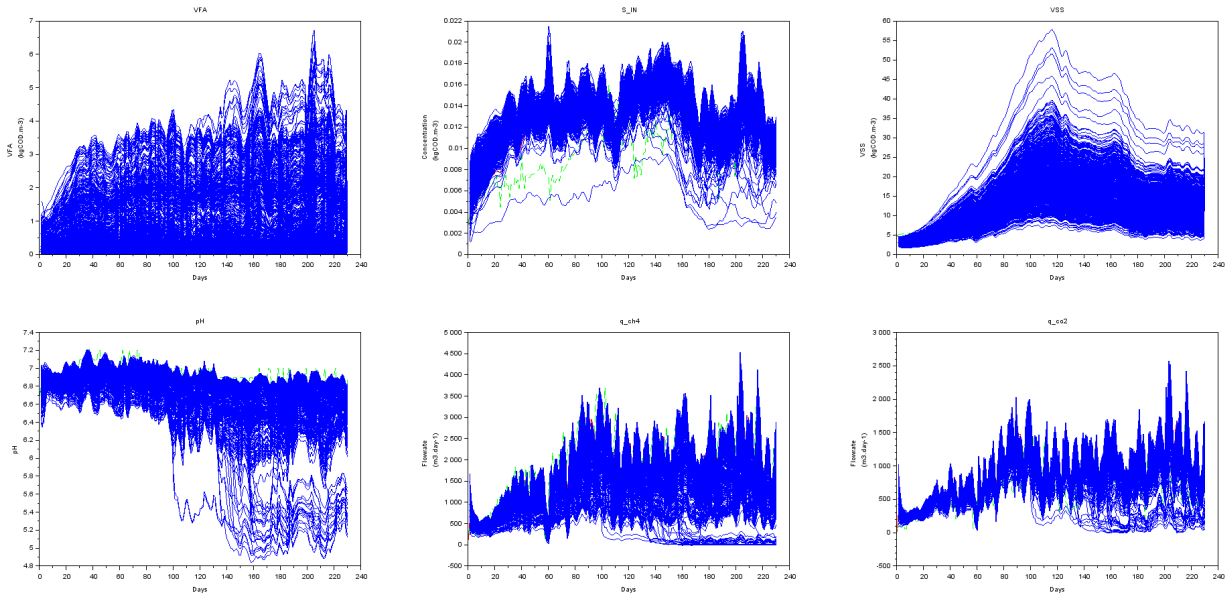


Figure 31: Plot showing 1000 Monte Carlo simulation runs

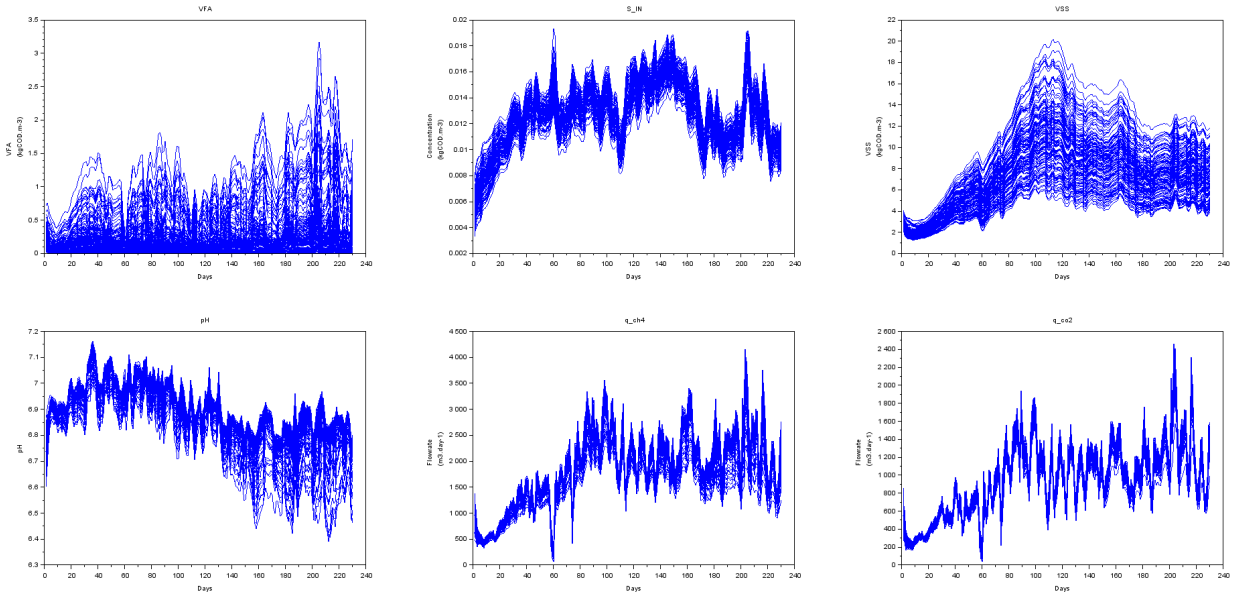


Figure 32: 250 Monte Carlo simulation runs with outliers beyond ±1.5 x IQR removed

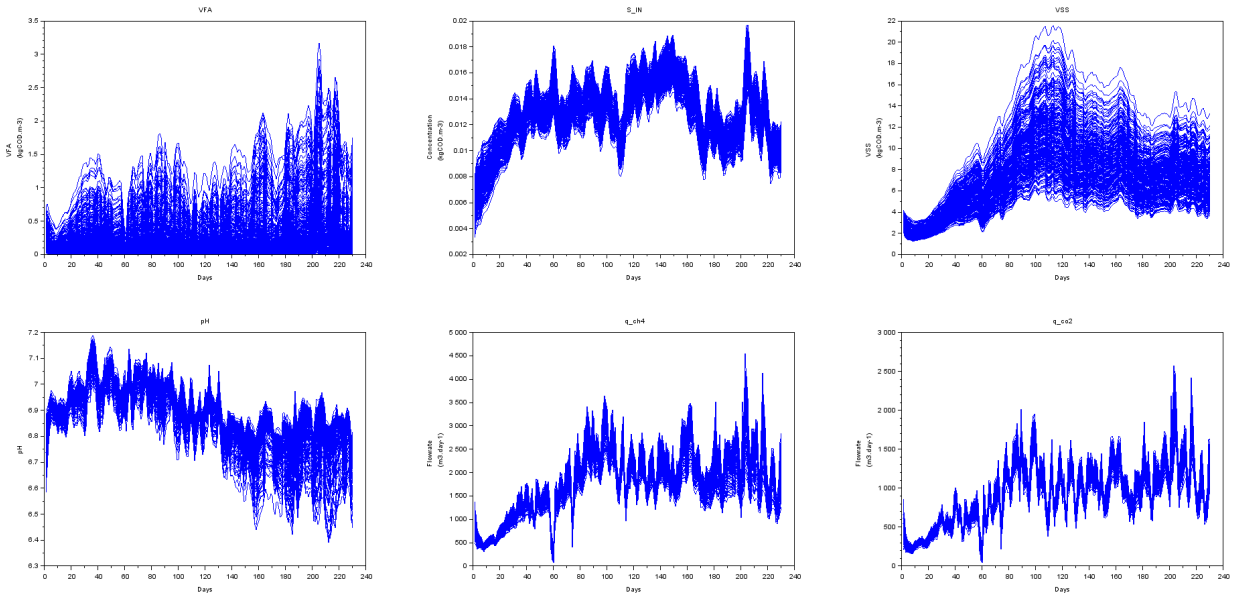


Figure 33: 500 Monte Carlo simulation runs with outliers beyond ±1.5 x IQR removed

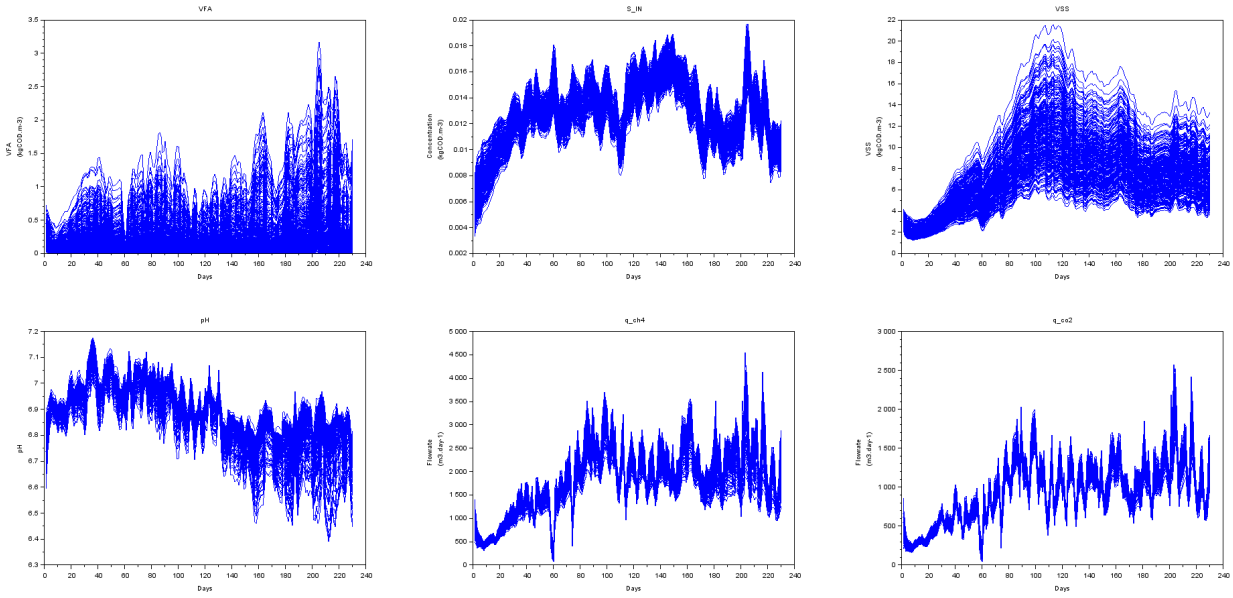


Figure 34: 1000 Monte Carlo simulation runs with outliers beyond ±1.5 x IQR removed

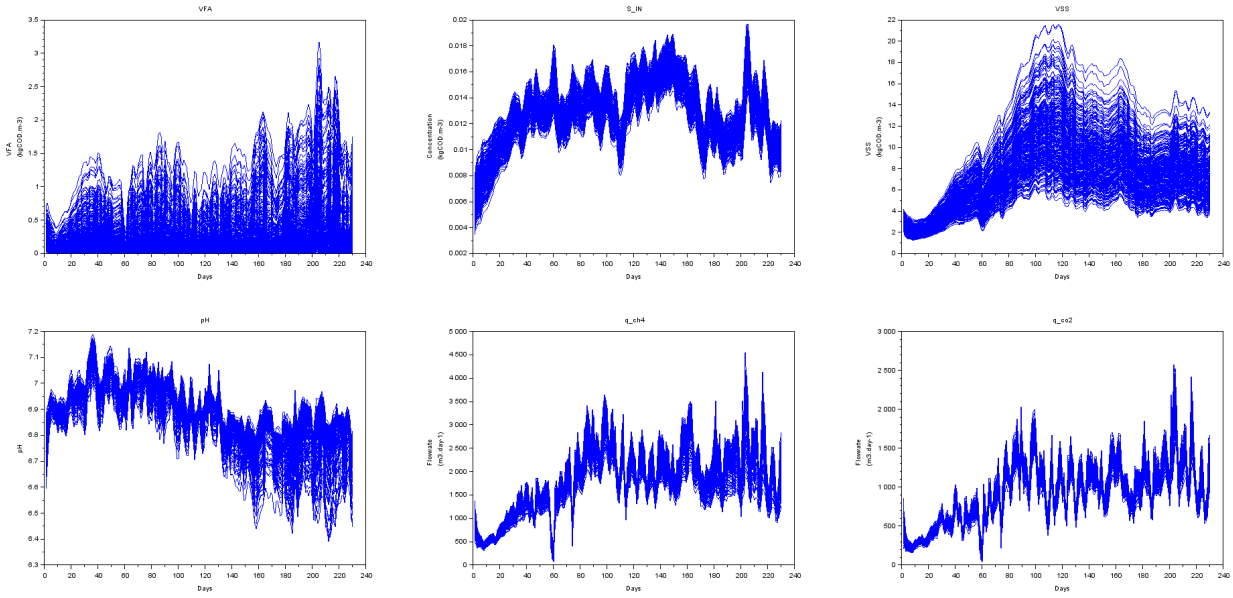


Figure 35: 1500 Monte Carlo simulation runs with outliers beyond ±1.5 x IQR removed

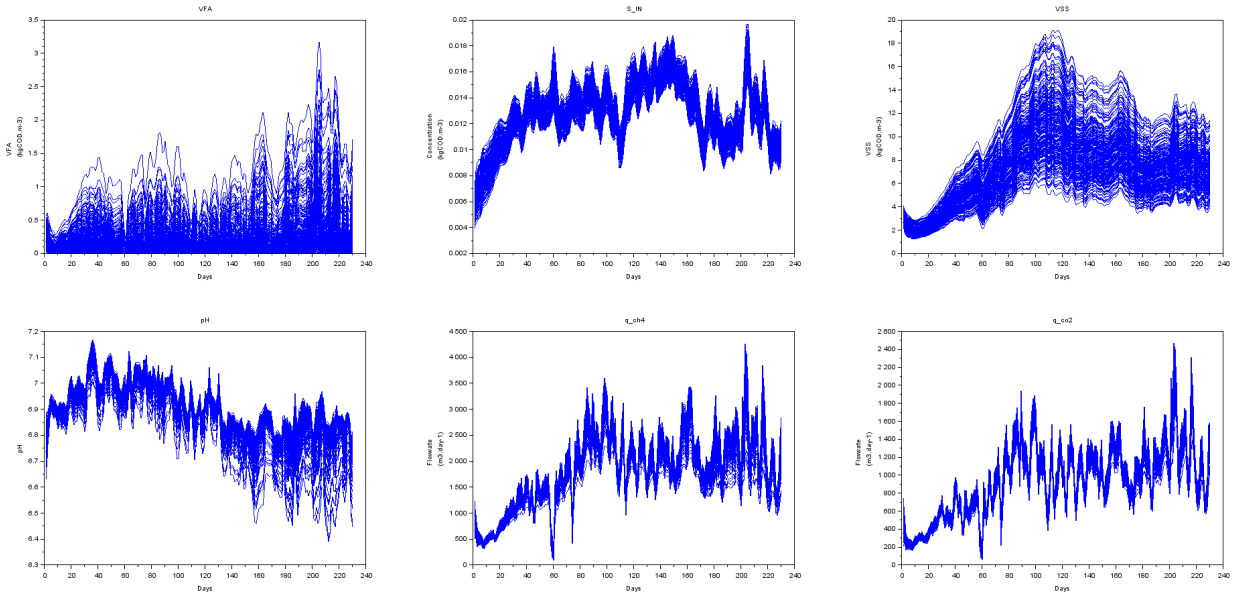


Figure 36: 500 Monte Carlo simulation runs with outliers beyond ±1.0 x IQR removed

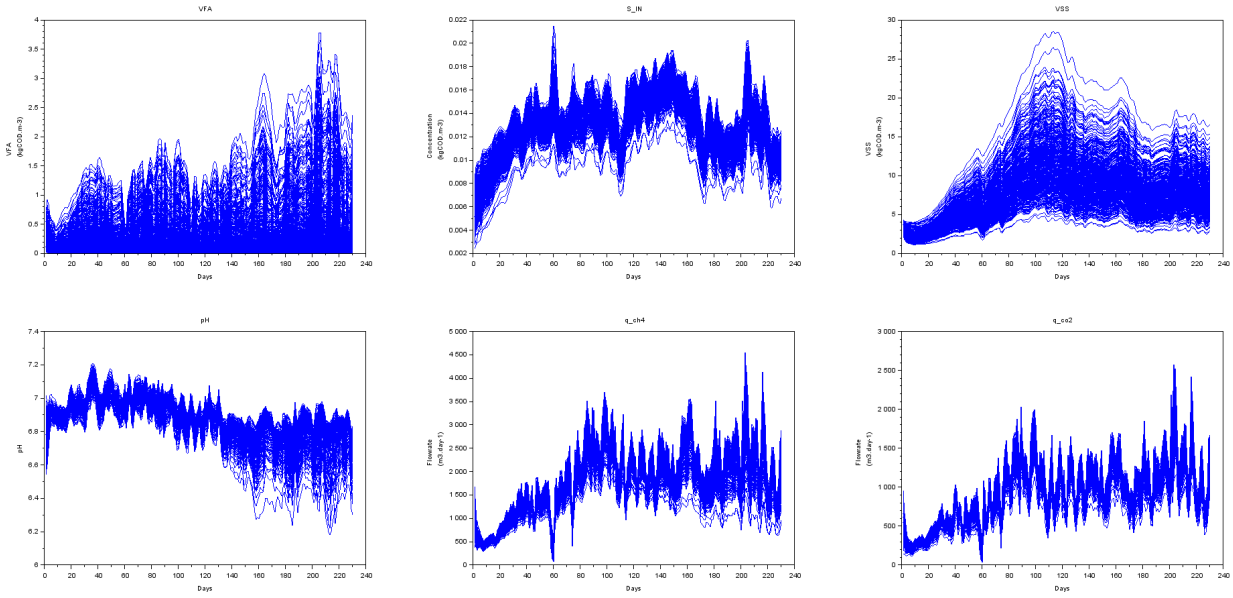


Figure 37: 500 Monte Carlo simulation runs with outliers beyond ±3.0 x IQR removed

8.6. Appendix F – Example Calculations

8.6.1. Biodegradability Factor

In this study, the biodegradability factor is estimated using Equation 35. Components in this equation are calculated from actual plant data where only the steady-state period is considered.

Step 1: COD load attributed to methane gas (COD_{CH4})

The volume of methane gas produced is summed and then converted to COD using the theoretical methane yield (0.35 m3/kg COD):

$$COD_{CH4} = \sum_{Day\ 80}^{Day\ 230} V_{CH4} \div 0.35\ \frac{m^3}{kgCOD}$$

$$= 328508\ m^3 \div 0.35\ \frac{m^3}{kgCOD}$$

$$= 938594\ kgCOD$$

Step 2: COD load attributed to biomass production (COD_{BM})

The volume of sludge wasted is summed and then converted to COD using the theoretical biomass COD content (1.42 kg COD/kg VSS). Given that the inconsistent solid concentration in digester, density of the sludge extracted is expected to fluctuate significantly. It is hence assumed that the average density (8.6 kg TSS/m^3) across the steady-state period is applicable. The ratio of organic solids to total solids, VSS/TSS, is taken as the average ratio observed from the plant data.

$$COD_{BM} = \sum_{Day\ 80}^{Day\ 230} V_{BM} \times 16\ \frac{kg\ TSS}{m^3} \times 0.89\ \frac{kg\ VSS}{kg\ TSS} \times 1.42\ \frac{kg\ COD}{kg\ Biomass}$$

$$= 6765\ m^3 \times 8.6\ \frac{kg}{m^3} \times 0.89\ \frac{kg\ VSS}{kg\ TSS} \times 1.42\ \frac{kg\ COD}{kg\ Biomass}$$

$$= 73531\ kgCOD$$

Step 3: Net COD load change in the digester content

The net COD load change is calculated as the difference between the first and last day of steady-state period:

$$COD_{reactor,i} - COD_{reactor,f} = \left(TCOD_{reactor,Day\ 80} - TCOD_{reactor,Day\ 230}\right) \times V_{reactor}$$

$$= \left(10378\frac{mg}{l} - 11028\frac{mg}{l}\right) \times 2875\ m^3$$

$$= 1868\ kgCOD$$

Step 4: Net COD load entered and exited the reactor

The net COD load processed by the reactor is calculated as the accumulated difference in COD load between the influent and treated evaluated from the first day to the last day of steady-state period:

$$COD_{inf} - COD_{eff} = \sum_{Day\ 80}^{Day\ 230} \left(TCOD_{influent} - TCOD_{effluent}\right) \times V_{reactor}$$

$$= 1292741\ kg - 15450\ kg$$

$$= 1277291\ kgCOD$$

Step 5: COD loss to denitrification

The amount of COD lost to denitrification is estimated using the theoretical COD demand for nitrate and nitrite denitrification, which are 2.86 mg/l and 1.71 mg/l per mg/l NO_3 and NO_2, respectively.

$$COD_{DN} = \sum_{Day\ 80}^{Day\ 230} kg\ NO_3 \times 2.86\ + kg\ NO_2\ \times\ 1.71$$

$$= 35583\ kg\ NO_3 \times 2.86 - 16465\ kg\ NO_2\ \times\ 1.71$$

$$= 129922\ kgCOD$$

Step 6: Calculate biodegradability factor (f_d')

Substituting all the components into Equation 35:

$$f_d' \approx \frac{COD_{CH4} + COD_{BM}}{\left(COD_{reactor,i} - COD_{reactor,f}\right) + \left(COD_{inf} - COD_{eff}\right) - COD_{DN}} \qquad \textit{Equation 42}$$

$$\approx \frac{938594 + 73531}{1868 + 1277291 - 129922} \approx \mathbf{0.88}$$

8.6.2. Translating Soluble Plant Measurements to ADM1 Format

The influent data is translated into ADM1 format by following the methods described in Table 9 and Table 10. Note that for demonstration purpose, the example calculation only consider one day of data (Day 230). Similar philosophy is applied for all time intervals.

Step 1: Translate SCOD to VFA constituents

Total VFA measured on Day 230 = 1528 mg/l

(i) Valerate: $S_{va,in} = VFA \times \eta_{va,VFA} \times Y_{va}$

$= (1528 \text{ mg/l} / 1000) \times 0.09 \times 2.039216 \text{ kg COD.(kg va)}^{-1}$

$= 0.288511 \text{ kg COD/m}^3\text{.day}$

(ii) Butyrate: $S_{bu,in} = VFA \times \eta_{bu,VFA} \times Y_{bu}$

$= (1528 \text{ mg/l} / 1000) \times 0.37 \times 1.818182 \text{ kg COD.(kg bu)}^{-1}$

$= 1.028956 \text{ kg COD/m}^3\text{.day}$

(iii) Propionate: $S_{pro,in} = VFA \times \eta_{pro,VFA} \times Y_{pro}$

$= (1528 \text{ mg/l} / 1000) \times 0.22 \times 1.513514 \text{ kg COD.(kg pro)}^{-1}$

$= 0.513922 \text{ kg COD/m}^3\text{.day}$

(iv) Acetate: $S_{ac,in} = VFA \times \eta_{ac,VFA} \times Y_{ac}$

$= (1528 \text{ mg/l} / 1000) \times 0.31 \times 1.066667 \text{ kg COD.(kg ac)}^{-1}$

$= 0.513106 \text{ kg COD/m}^3\text{.day}$

Step 2: Translate SCOD to Monosaccharides, Amino acids & LCFA

SCOD measured on Day 230 = 3655 mg/l

$VFA_{COD} = S_{va,in} + S_{bu,in} + S_{pro,in} + S_{ac,in} = 0.289 + 1.029 + 0.514 + 0.513 = 2.345 \text{ kg COD/m}^3\text{.day}$

(i) Monosaccharides: $S_{su,in} = (SCOD \times fd' - VFA_{COD}) \times \eta_{su}$

$= [(3655 \text{ mg/l} / 1000) \times 0.88 - 2.345] \times 0.44$

$= 0.386 \text{ kg COD/m}^3\text{.day}$

(ii) Amino acids: $S_{aa,in} = (SCOD \times fd' - VFA_{COD}) \times \eta_{aa}$

$= [(3655 \text{ mg/l} / 1000) \times 0.88 - 2.345] \times 0.40$

$= 0.351 \text{ kg COD/m}^3\text{.day}$

(iii) LCFA: $S_{fa,in} = (SCOD \times fd' - VFA_{COD}) \times \eta_{fa}$

$= [(3655 \text{ mg/l} / 1000) \times 0.88 - 2.345] \times 0.16$

$= 0.137 \text{ kg COD/m}^3\text{.day}$

Step 3: Translate SCOD to Soluble Inerts

(i) Soluble Inerts: $S_{I,in} = SCOD \times (1 - fd') = (3655 / 1000) \times (1 - 0.88) = 0.439 \text{ kg COD/m}^3\text{.day}$

Step 4: Translate Alkalinity & Ammonium to Inorganic Carbon & Inorganic Nitrogen

Alkalinity measured on Day 230 = 950 mg/l as $CaCO_3$

Ammonium concentration measured on Day 230 = 32 mg/l as N

(i) Inorganic Carbon: $S_{IC} = (950 \text{ mg/l} / 1000) / 44 = 0.022 \text{ kmol/m}^3$

(ii) Inorganic Nitrogen: $S_{IN} = (32 \text{ mg/l} / 1000) / 14 = 0.0023 \text{ kmol/m}^3$

Step 5: Calculate Cations & Anions

Calcium concentration measured on Day 230 = 42 mg/l

Magnesium concentration measured on Day 230 = 1 mg/l

Sodium concentration referenced = 0.0546 kmol/m^3

Potassium concentration referenced = 0.00387 kmol/m^3

(i) Cations: $S_{cat} = [Ca^{2+}] + [Mg^{2+}] + [Na^{+}] + [K^{+}]$

$= (42 \text{ mg/l} / 1000 / 40 \text{ kmol/kg}) + (1 \text{ mg/l} / 1000 / 24 \text{ kmol/kg}) + 0.0546 \text{ kmol/m}^3 + 0.00387 \text{ kmol/m}^3$

$= 0.0586 \text{ kmol/m}^3$

Hydrogen concentration $[H^{+}]$ calculated from Equation 2 = 3.16×10^{-7} kmol/m^3

Ammonium concentration $[NH_4^{+}]$ calculated from Equation 32 = 0.0021 kmol/m^3

Hydroxide concentration $[OH^{-}]$ calculated from Equation 33 = 6.58×10^{-8} kmol/m^3

Bicarbonate concentration $[HCO3^{-}]$ calculated from Equation 30 = 0.115 kmol/m^3

VFA concentrations $[Ac^{-}]$, $[Pro^{-}]$, $[Bu^{-}]$, $[Va^{-}]$ calculated from Equation 31

(ii) Anions: $S_{an} = [H^+] + [NH_4^+] + S_{cat} – [OH^-] – [HCO3^-] – [Ac^-] – [Pro^-] – [Bu^-] – [Va^-]$

$= 3.16 \times 10^{-7}$ kmol/m^3 + 0.0021 kmol/m^3 + 0.0586 kmol/m^3 - 6.58×10^{-8} kmol/m^3 - 0.115 kmol/m^3 – 0.0058 kmol/m^3 – 0.00334 kmol/m^3 – 0.00469 kmol/m^3 – 0.001 kmol/m^3

= 0.0332 kmol/m^3

Step 6: Set all other soluble components to zero

(i) Dissolved hydrogen: $S_{h2,in} = 0$

(ii) Dissolved methane: $S_{ch4,in} = 0$

8.6.3. Translating Particulate Plant Measurements to ADM1 Format

The influent data is translated into ADM1 format by following the methods described in Table 11 and Table 12. Note that for demonstration purpose, the example calculation only consider one day of data (Day 230). Similar philosophy is applied for all time intervals.

Step 1: Estimate Total Acidogens

(Note that this step is necessary for this specific case as insufficient data about the raw is available)

SCOD measured on Day 230 = 3655 mg/l

VFA_{COD} calculated in previous section = 2.345 kg COD/m^3.day

f_{acid} = Fraction of COD in buffer tank converted to VFAs and alcohols as a result of acidification of dairy wastewater is reported to be 0.484.

Y_{acid} = Yield of acidogens during acidification of dairy wastewater at pH 6.5 is reported to be 0.26 g VSS/ g COD

Y_{bm} = COD content of biomass = 1.42 g COD/ g VSS

Firstly, the total COD entering the buffer tank is assumed using f_{acid}. The difference between this estimated COD and the measured TCOD is the COD load converted during acidification. Using the typical yield Y_{acid} an estimate for the total acidogens is obtained.

(i) Total Acidogens: $X_{degr} = [VFA_{COD} \div f_{acid} – SCOD \times fd'] \times Y_{acid} \times Y_{bm}$

= (2.345 kg COD/m^3.day ÷ 0.484 – 3655 mg/l / 1000 x 0.88) x 0.26 g VSS/ g COD x 1.42 g COD / g VSS

= 0.60 kg COD/m^3.day

Step 2: Calculate composition of various acidogens

Carbohydrates fraction η_{ch} = 0.44

Protein fraction η_{pr} = 0.40

Lipids fraction η_{li} = 0.16

Assuming acidogen composition is relative to the fractional composition of carbohydrates, proteins and lipids:

Monosaccharide degraders: $X_{su,in}$ = Acidogens x η_{ch} = 0.60 x 0.44 = 0.264 kg COD/m^3.day

Amino acids degraders: $X_{aa,in}$ = Acidogens x η_{ch} = 0.60 x 0.40 = 0.240 kg COD/m^3.day

LCFA degraders: $X_{fa,in}$ = Acidogens x η_{ch} = 0.60 x 0.16 = 0.096 kg COD/m^3.day

Step 3: Translate PCOD to Carbohydrates, Proteins & Lipids

SCOD measured on Day 230 = 3655 mg/l

TCOD measured on Day 230 = 7060 mg/l

PCOD = TCOD – SCOD = 7060 mg/l – 3655 mg/l = 3405 mg/l

Carbohydrates: $X_{ch,in}$ = (PCOD x fd' – X_{degr}) x η_{ch}

= (3405 / 1000 x 0.88 – 0.60) x 0.44

= 1.05 kg COD/m^3.day

Proteins: $X_{pr,in}$ = (PCOD x fd' – X_{degr}) x η_{ch}

= (3405 / 1000 x 0.88 – 0.60) x 0.40

= 0.96 kg COD/m^3.day

Lipids: $X_{li,in}$ = (PCOD x fd' – X_{degr}) x η_{ch}

= (3405 / 1000 x 0.88 – 0.60) x 0.16

= 0.38 kg COD/m^3.day

Step 4: Translate PCOD to Particulate Inerts

(i) Soluble Inerts: $S_{I,in}$ = PCOD x (1 – fd') = (3405 / 1000) x (1 – 0.88) = 0.409 kg COD/m^3.day

Step 5: Set all other particulate components to zero

(i) Composite particulate: $X_{c,in}$ = 0

(ii) C4 degraders: $X_{c4,in}$ = 0

(iii) Propionate degraders: $X_{pro,in}$ = 0

(iv) Acetate degraders: $X_{ac,in} = 0$

(v) Hydrogen degraders: $X_{h2,in} = 0$

8.6.4. Calculating Objective function (U_{diff})

The objective during model optimisation is to minimise the objective function ($\mathbf{u}_{diff}$), which is defined in this study as the difference between the latent variables of measured variables ($\mathbf{u}_m$) and the latent variables of simulated results ($\mathbf{u}_s$). Objective function evaluation is performed using Equation 43 every time after parameters are calibrated.

$$\boldsymbol{u}_{diff} = min \sum_{i=1}^{p} \sum_{t=1}^{n} \left(\boldsymbol{u}_{m,i}(t) - \boldsymbol{u}_{s,i}(t) \right)^2$$

where:

$\mathbf{u}_{diff}$ is the model objective function

n is the total number of time intervals;

p is the total number of latent variables selected;

$\mathbf{u}_{m,i}$ is the output latent variable for the *measured* outputs at time interval t;

$\mathbf{u}_{s,i}$ is the output latent variable for the *simulated* outputs at time interval t;

The following data set is referenced for example calculation. For demonstration purpose the simulation scope is simplified to two time intervals and a PLSR consisting of two latent variables.

Time Interval (Day)	Simulated output vector, $\mathbf{Y}_s$	Measured output vector, $\mathbf{Y}_m$
1	VFA: 0.18299	VFA: 0.134103
	S_IN: 0.006705	S_IN: 0.004251
	VSS: 2.595955	VSS: 4.932968
	pH: 6.909906	pH: 6.9
	q_CH4: 566.698	q_CH4: 358.3214
	q_CO2: 251.5566	q_CO2: 80.10716
2	VFA: 0.163243	VFA: 0.091351

	S_IN: 0.007085	S_IN: 0.004566
	VSS: 2.556666	VSS: 5.038868
	pH: 6.90786	pH: 6.8
	q_CH4: 541.7001	q_CH4: 507.744
	q_CO2: 248.7446	q_CO2: 111.456

From the PLSR results (i.e. Step 3), an output loading vector (**q**) corresponding to two latent variables is obtained:

Time Interval (Day)	1st Latent Variable Output Loading Vector, $\mathbf{q_1}$	2nd Latent Variable Output Loading Vector, $\mathbf{q_2}$
1	0.875497	0.160572
	0.321773	0.77736
	0.665149	0.251696
	0.50236	0.623451
	0.309176	0.681115
	0.449083	0.509797
2	0.910941	0.126062
	0.332756	0.752034
	0.68398	0.244362
	0.498085	0.636422
	0.261499	0.721064
	0.347944	0.601484

Before applying the loading vector, the output vectors first need to be normalised. From the Monte Carlo results (i.e. Step 2), maximum and minimum values at each time interval are noted as follows:

Time Interval (Day)	Maximum values, $\mathbf{Y}_{max}$	Minimum values, $\mathbf{Y}_{min}$
1	VFA: 0.703089	VFA: 0.010929
	S_IN: 0.008758	S_IN: 0.00301
	VSS: 3.725297	VSS: 1.178361
	pH: 7.037791	pH: 6.829438
	q_CH4: 625.6168	q_CH4: 273.5294
	q_CO2: 292.2564	q_CO2: 119.7491
2	VFA: 0.735666	VFA: 0.00726
	S_IN: 0.009058	S_IN: 0.003373
	VSS: 3.781064	VSS: 1.196652
	pH: 7.032901	pH: 6.827943
	q_CH4: 573.372	q_CH4: 226.8117
	q_CO2: 275.7924	q_CO2: 90.36569

Normalising using the algorithm given in Equation 39 thus results in:

Time Interval (Day)	Normalised simulated output vector, $\mathbf{Y}_{s,norm}$	Normalised measured output vector, $\mathbf{Y}_{m,norm}$
1	VFA: -0.50283	VFA: -0.64409
	S_IN: 0.285665	S_IN: -0.5682
	VSS: -0.37646	VSS: 0.651489
	pH: -0.22758	pH: -0.32267
	q_CH4: 0.665317	q_CH4: -0.51835
	q_CO2: 0.528139	q_CO2: -0.99663
2	VFA: -0.57171	VFA: -0.76911

	S_IN: 0.305893	S_IN: -0.5803
	VSS: -0.40668	VSS: 0.676209
	pH: 0.689323	pH: 0.421334
	q_CH4: 0.817221	q_CH4: 0.621261
	q_CO2: 0.708264	q_CO2: -0.77252

Finally, the difference in latent variables for both simulated and measured outputs is determined. Latent variables are calculated by applying the vector multiplication function: **u** = **Yq**. It should be noted that, instead of taking the difference between full latent variables, the difference is calculated for each sub-latent variable components and then summed as grand total. This approach is preferred as it prevents high and low value components from cancelling each other out.

Time Interval (Day)	$U_{s,1}$	$U_{s,2}$	$U_{m,1}$	$U_{m,2}$	$(U_{s,1} - U_{m,1})^2$	$(U_{s,2} - U_{m,2})^2$
1	-0.4402	-0.080	-0.5638	-0.1034	0.0152	0.0005
	0.0919	0.2220	-0.1828	-0.4416	0.0754	0.440
	-0.2504	-0.0947	0.4333	0.1639	0.4675	0.066
	-0.1143	-0.1418	-0.1620	-0.2011	0.0022	0.0035
	0.2057	0.4531	-0.1602	-0.3530	0.1339	0.6499
	0.2770	0.3145	-0.4475	-0.5080	0.5250	0.6766
2	-0.5207	-0.0720	-0.7006	-0.0969	0.0323	0.0006
	0.1017	0.2300	-0.1930	-0.4364	0.0869	0.4441
	-0.2781	-0.0993	0.4625	0.1652	0.5485	0.0700
	0.3433	0.4387	0.2098	0.2681	0.0178	0.0290
	0.2137	0.5892	0.1624	0.4479	0.0026	0.0199
	0.2464	0.4260	-0.2687	-0.4646	0.2654	0.7932
				Grand Total - U_{diff}	Sum(0.0152,…,0.7932) = **5.369**	

8.7. Appendix G – SCILAB Codes

```
//=========================================================================================================
//ADM1 MODEL
//IWA TASK GROUP, SCIENTIFIC AND TECHNICAL REPORT NO.13
//File: Main.sce
//=========================================================================================================

//1. INITIALISATION
//==================
clear;                                                                          //clears all variables
clearglobal;
stacksize('max');

exec('Inputs.sce',-1);                                                          //Load all model inputs
exec('Parameters.sce',-1);                                                      //Initiate parameters & constants
exec('Functions.sce',-1);                                                       //Initiate ADM1 functions

MC_data = zeros(1,33);
VFA_data = zeros(1,1);
VSS_data = zeros(1,1);


//2. SENSITIVITY ANALYSIS
//========================
if run_mc == "Yes" then                                                 //If true, a new set of MC data will be generated
    sim_mode = "single";
    exec('ADM1.sce',-1);                                                        //Run ADM1 model
    exec('Plot_Sens.sce',-1);                                                   //plot default results in red lines

    sim_mode = "monte-carlo";
    for mont_car = 1:MC_nr
        disp(mont_car)
        exec('Rand_Parameters.sce',-1);
        exec('Inputs.sce',-1);
        exec('ADM1.sce',-1);
    end

    MC_data_test = MC_data(2:size(MC_data,1),:);
    VFA_data_test = VFA_data(2:size(VFA_data,1),:);
    VSS_data_test = VSS_data(2:size(VSS_data,1),:);
    t = t(2:length(t));
    if run_mc_plot == "Yes" then
        for mont_car = 1:MC_nr
            count = 1;
            DSV = zeros(sim_dur,33);
            for index = (mont_car*sim_dur + 1) : ((mont_car + 1) * sim_dur)
                DSV(count,:) = MC_data_test(index,:);                   //Record Dynamic State Variables of each MC run
                VFA_s(count,:) = VFA_data_test(index,:);                //Record total VFA of each MC run
                VSS_s(count,:) = VSS_data_test(index,:);                //Record VSS of each MC run
                count = count + 1;
            end
            DSV = DSV';
            exec('Plot_Sens.sce',-1);                                           //plot MC results in blue lines
        end
    end
end

if run_mc == "No" then                                  //If true, MC data will be extracted from external source (.txt or .xls)
    MC_data_test = zeros(sim_dur*(MC_nr+1),33);
    sim_mode = "single";
    exec('ADM1.sce',-1);                                                        //Run ADM1 model
    exec('Plot_Sens.sce',-1);                                                   //plot default results in red lines

    MC_data_test(1:sim_dur,1:33) = MC_data(2:size(MC_data,1),:);
    clear('MC_data');
    co = size(MC_lib,2);
    MC_data_test((sim_dur+1):size(MC_data_test,1),1:33) = MC_lib(1:sim_dur*MC_nr,1:co);
    VFA_data_test = [MC_data_test(:,4)+MC_data_test(:,5)+MC_data_test(:,6)+MC_data_test(:,7)];
```

```
    VSS_data_test =
[MC_data_test(:,13)+MC_data_test(:,14)+MC_data_test(:,15)+MC_data_test(:,16)+MC_data_test(:,17)+MC_data_test(:,18)+MC
_data_test(:,19)+MC_data_test(:,20)+MC_data_test(:,21)+MC_data_test(:,22)+MC_data_test(:,23)+MC_data_test(:,24)];

    t = linspace(t_0+1,sim_dur,sim_dur);
    if run_mc_plot == "Yes" then
        sim_mode = "monte-carlo";
        for mont_car = 1:MC_nr
            count = 1;
            DSV = zeros(sim_dur,33);
            for index = (mont_car*sim_dur + 1) : ((mont_car + 1) * sim_dur)
                DSV(count,:) = MC_data_test(index,:);
                VFA_s(count,:) = VFA_data_test(index,:);
                VSS_s(count,:) = VSS_data_test(index,:);
                count = count + 1;
            end
            DSV = DSV';
            exec('Plot_Sens.sce',-1);                                    //plot MC results in blue lines
        end
    end
end


//3. PLSR
//========
if run_plsr == "Yes" then
    MC_data_pls =
[(MC_data_test(:,4)+MC_data_test(:,5)+MC_data_test(:,6)+MC_data_test(:,7)),MC_data_test(:,11),(MC_data_test(:,13)+MC_da
ta_test(:,14)+MC_data_test(:,15)+MC_data_test(:,16)+MC_data_test(:,17)+MC_data_test(:,18)+MC_data_test(:,19)+MC_data_t
est(:,20)+MC_data_test(:,21)+MC_data_test(:,22)+MC_data_test(:,23)),MC_data_test(:,30),MC_data_test(:,32),MC_data_test(:,
33)];                                                    //arrange matrix to [VFA;S_IN;VSS;pH;q_ch4;q_co2]
    pls_Ran_Par = Ran_Par;

    clear('MC_data_test');
    clear('VFA_data_test');
    clear('VSS_data_test');

    exec('PLSR.sce',-1);

    fprintfMat('W_pls.xls',W_pls);
    fprintfMat('Q_pls.xls',Q_pls);
    fprintfMat('P_pls.xls',P_pls);
    fprintfMat('R_pls.xls',R_pls);
end


//4. MODEL FITTING
//=================
if run_modfit == "Yes" then
    exec('Mod_Fit.sce',-1);
end
```

```
//=============================================================================================
//ADM1 MODEL
//IWA TASK GROUP, SCIENTIFIC AND TECHNICAL REPORT NO.13
//File: Inputs.sce
//=============================================================================================

//1. DEFINITION OF SYSTEM
//=========================
V_digester = 2875;                                          //[m3] volume of digester tank
V_headspace = 165;                                          //[m3] volume of digester headspace
Q_in = 600;                                                 //[m3/day] flow rate into digester
Q_was = 0;                                                  //[m3/day] waste flow rate discharged from digester
Q_eff = Q_in - Q_was;                                       //[m3/day]    flow rate out of AD system
T = 308;                                                    //[K] Temperature
P_atm = 1.013;                                              //[bar] Pressure of atmosphere


//2. OTHER SIMULATION INPUTS
//====================
t_0 = 0;
S_h_ion = 0.00000001;                                       //initial S_h_ion guess
sim_dur = 230;                                              //days to simulate
MC_nr = 500;                                                //select no. of Monte Carlo runs
no_LV = 2;                                                  //select no. of latent variables
lamda0 = [1;1];                                             //initial guess of Lamda value for parameter fitting:
                                                            e.g. select [1] if no_LV = 1; [1;1] if no_LV = 2; etc.
sim_dur_CV = 320;

run_mc = "No";                                              //select yes to run Monte-Carlo algorithm
run_mc_plot = "No";                                         //select yes to plot Monte-Carlo results
run_plsr = "No";                                            //select yes to run PLSR algorithm
run_modfit = "No";                                          //select yes to run Model Fitting algorithm
//run_pca = "No";                                           //select yes to run PCA algorithm (INACTIVE)

counter = 1; //plot counter


//3. INFLUENT COMPOSITION
//=========================
sheets = readxls("Inputs.xls");                 //influent composition with respect to time stored in Excel sheet "Inputs"
s1 = sheets(1);
s2 = sheets(2);
s3 = sheets(3);
s4 = sheets(4);
s5 = sheets(5);
s6 = sheets(6);
s7 = sheets(7);

if run_mc == "No" then
MC_lib = read('MC_lib.txt',sim_dur*MC_nr,33)
//sheets = readxls("Ran_Par_lib.xls");
//Ran_Par_lib = sheets(1);
Ran_Par = read('Ran_Par_lib.txt',MC_nr,58)
end

ro = size(s1,1);
co = size(s1,2);
t_data = s1.value(2:ro,1);
inf_data = s1.value(2:ro,2:co-3);
Q_in = s1.value(2:ro,co-2);
Q_was = s1.value(2:ro,co-1);
T_data = s1.value(2:ro,co);

ro = size(s2,1);
co = size(s2,2);
ini_data = s2.value(2:ro,2:co);

ro = size(s3,1);
co = size(s3,2);
q_ch4_m = s3.value(2:ro,co-6);
q_co2_m = s3.value(2:ro,co-5);
pH_m = s3.value(2:ro,co-4);
S_IN_m = s3.value(2:ro,co-3);
```

```
VFA_m = s3.value(2:ro,co-2);
VSS_m = s3.value(2:ro,co-1);
C_was = s3.value(2:ro,co);

ro = size(s4,1);
co = size(s4,2);
surv_min = (s4.value(2,2:co))';
surv_max = (s4.value(3,2:co))';

ro = size(s5,1);                                          //Load data for cross-validation
co = size(s5,2);
t_data_cv = s5.value(2:ro,1);
inf_data_cv = s5.value(2:ro,2:co-3);
Q_in_cv = s5.value(2:ro,co-2);
Q_was_cv = s5.value(2:ro,co-1);
T_data_cv = s5.value(2:ro,co);

ro = size(s6,1);                                          //Load data for cross-validation
co = size(s6,2);
q_ch4_m_cv = s6.value(2:ro,co-6);
q_co2_m_cv = s6.value(2:ro,co-5);
pH_m_cv = s6.value(2:ro,co-4);
S_IN_m_cv = s6.value(2:ro,co-3);
VFA_m_cv = s6.value(2:ro,co-2);
VSS_m_cv = s6.value(2:ro,co-1);
C_was_cv = s6.value(2:ro,co);

ro = size(s7,1);                                          //Load data for cross-validation
co = size(s7,2);
CV_Par = (s7.value(2,2:co))';


//4. REACTOR INITIAL COMPOSITION
//==============================

DSV_0 = ini_data';
```

```
//=========================================================================================================
//ADM1 MODEL
//IWA TASK GROUP, SCIENTIFIC AND TECHNICAL REPORT NO.13
//File: Functions.sce
//
//<Modifications>
//Note 1: Corrected Carbon and Nitrogen balances for all 19 processes [Rosen & Jeppson,2006]
//=========================================================================================================

//Function 1 - Charge Balance
//============================
function balance=charge_bal(S_h_ion, S_cat, S_IN, S_IC, S_ac, S_pro, S_bu, S_va, S_an, Ka, cpc)
    balance = S_cat+...
              ((S_IN.*S_h_ion)./(Ka.nh4+S_h_ion)) + ...
              S_h_ion-((Ka.co2.*S_IC)./(Ka.co2+S_h_ion)) - ...
              ((Ka.ac.*S_ac)./(Ka.ac+S_h_ion))./cpc.ac - ...
              ((Ka.pro*S_pro)./(Ka.pro+S_h_ion))./cpc.pro - ...
              ((Ka.bu.*S_bu)./(Ka.bu+S_h_ion))./cpc.bu - ...
              ((Ka.va.*S_va)./(Ka.va+S_h_ion))./cpc.va - ...
              (Ka.h2o./S_h_ion)-S_an;
endfunction


//Function 2 - pH Inhibition: Acetogens & Acidogens
//==================================================
function pH_inhibit=I_pH(pH)
    if pH < I_pH_ul then
        pH_inhibit = exp(-3*((pH-I_pH_ul)/(I_pH_ul-I_pH_ll))^2)
    else
        pH_inhibit = 1
    end
endfunction


//Function 3 - pH Inhibition: Acetate Degraders
//==============================================
function pH_ac_inhibit=I_pH_ac(pH)
    if pH < I_pH_ac_ul then
        pH_ac_inhibit = exp(-3*((pH-I_pH_ac_ul)/(I_pH_ac_ul-I_pH_ac_ll))^2)
    else
        pH_ac_inhibit = 1
    end
endfunction


//Function 4 - pH Inhibition: H2 Degraders
//=========================================
function pH_h2_inhibit=I_pH_h2(pH)
    if pH < I_pH_h2_ul then
        pH_h2_inhibit = exp(-3*((pH-I_pH_h2_ul)/(I_pH_h2_ul-I_pH_h2_ll))^2)
    else
        pH_h2_inhibit = 1
    end
endfunction


//Function 5 - Inorganic Nitrogen Inhibition
//===========================================
//Limit to growth due to lack of IN
function IN_inhibit=I_IN_lim(S_IN, Ks_IN)
    if S_IN < 0 then
        IN_inhibit = 0
    else
        IN_inhibit = 1/(1+Ks_IN/S_IN)
    end
endfunction


//Function 6 - Hydrogen Inhibition
//================================
function h2_inhibit=I_h2(S_h2, KI_h2)
    h2_inhibit = 1/(1+S_h2/KI_h2)
endfunction
```

```
//Function 7 - Ammonia Inhibition
//==============================
function nh3_inhibit=I_nh3_ac(S_nh3, KI_nh3_ac)
      nh3_inhibit = 1/(1+S_nh3/KI_nh3_ac)
endfunction


//Function 8 - ADM1 Model
//=======================
function dDSVdt=f_dDSVdt(t, DSV)
//       global('DSV_data','DSV_t','t_count','sim_dur');
S_su = DSV(1);S_aa = DSV(2);S_fa = DSV(3);S_va = DSV(4);S_bu = DSV(5);S_pro = DSV(6);
S_ac = DSV(7);S_h2 = DSV(8);S_ch4 = DSV(9);S_IC = DSV(10);S_IN = DSV(11);S_I = DSV(12);
X_c = DSV(13);X_ch = DSV(14);X_pr = DSV(15);X_li = DSV(16);X_su = DSV(17);X_aa = DSV(18);
X_fa = DSV(19);X_c4 = DSV(20);X_pro = DSV(21);X_ac = DSV(22);X_h2 = DSV(23);X_I = DSV(24);
S_an = DSV(25);S_cat = DSV(26);S_h2_g = DSV(27);S_ch4_g = DSV(28);S_co2_g = DSV(29);

m = zeros(19,26);


//Equilibrium & Charge Balances
//-----------------------------
h_solver_list = list(charge_bal,S_cat,S_IN,S_IC,S_ac,S_pro,S_bu,S_va,S_an,Ka,cpc);
S_h_ion = fsolve(S_h_ion,h_solver_list);

pH = -log10(S_h_ion);                                                          //pH of reactor content

S_nh4_ion = (S_IN*S_h_ion)/(Ka.nh4+S_h_ion);                                   //IN equilibrium
S_nh3 = S_IN-S_nh4_ion;
S_hco3 = (Ka.co2*S_IC)/(Ka.co2+S_h_ion);                                       //IC equilibrium
S_co2 = S_IC-S_hco3;


//Liquid-Gas Transfer
//-------------------
p_h2 = S_h2_g/M.h2*R*T;                                  //[bar] partial pressure of h2
p_ch4 = S_ch4_g/M.ch4*R*T;                               //[bar] partial pressure of CH4
p_co2 = S_co2_g*R*T;                                     //[bar] partial pressure of CO2
p_h2o = p_h2o_0*exp(deltaHvap0_h2o/(R*100)*(1/T_0-1/T)); //[bar] partial pressure of h2o, van't Hoff equation
P_headspace = p_co2+p_h2+p_ch4+p_h2o;                    //[bar] total gas phase pressure

kr_h2 = kLa*(S_h2-M.h2*KH.h2*p_h2);
kr_ch4 = kLa*(S_ch4-M.ch4*KH.ch4*p_ch4);
kr_co2 = kLa*(S_co2-KH.co2*p_co2);

q_gas = R*T/(P_atm-p_h2o)*V_digester*(kr_h2/M.h2+kr_ch4/M.ch4+kr_co2);          //[m3.d-1] total gas flow


//Process 1: Disintegration
//-------------------------
m(1,10) = C.Xc - ...
                 f.xi_xc*C.XI - ...
                 f.si_xc*C.SI - ...
                 f.pr_xc*C.pr - ...
                 f.ch_xc*C.ch - ...
                 f.li_xc*C.li;                   //[kmol.kgCOD-1] determines amount/portion of C in composite
                                                  particulates that disintegrates to inorganic carbon, Note 1
m(1,11) = N.Xc - ...
                 f.xi_xc*N.XI - ...
                 f.si_xc*N.SI - ...
                 f.pr_xc*N.aa;                   //[kmol.kgCOD-1] determines amount/portion of N in composite
                                                   particulates that disintegrates to inorganic nitrogen
m(1,12) = f.si_xc;                               //[kgCOD.kgCOD-1] catabolic yield of soluble inerts from
                                                   composite particulates

m(1,13) = -1;
m(1,14) = f.ch_xc;
m(1,15) = f.pr_xc;
m(1,16) = f.li_xc;
m(1,24) = f.xi_xc;
kr(1) = kdis*X_c;                                                              //kinetic rate

//Process 2: Hydrolysis carbohydrates
```

```
//---------------------------------
m(2,1) = 1;
m(2,10) = C.ch-C.su;                                                    //Note 1
m(2,14) = -1;
kr(2) = khyd.ch*X_ch;


//Process 3: Hydrolysis proteins
//-----------------------------
m(3,2) = 1;
m(3,10) = C.pr-C.aa;                                                    //Note 1
m(3,15) = -1;
kr(3) = khyd.pr*X_pr;


//Process 4: Hydrolysis lipids
//---------------------------
m(4,1) = 1-f.fa_li;
m(4,3) = f.fa_li;
m(4,10) = C.li - ...
          (1-f.fa_li)*C.su - ...
          f.fa_li*C.fa;                                                 //Note 1
m(4,16) = -1;
kr(4) = khyd.li*X_li;


//Process 5: Uptake of Sugars
//---------------------------
m(5,1) = -1;
m(5,5) = (1-Y.su)*f.bu_su;                          //(Yield of other substances)*(fractional yield of butyrate)
m(5,6) = (1-Y.su)*f.pro_su;
m(5,7) = (1-Y.su)*f.ac_su;
m(5,8) = (1-Y.su)*f.h2_su;
m(5,10) = C.su - ...
          (1-Y.su)*f.bu_su*C.bu - ...
          (1-Y.su)*f.pro_su*C.pro - ...
          (1-Y.su)*f.ac_su*C.ac - ...
          Y.su*C.bm;                                    //determines amount/portion of carbon in monosaccharides
                                                          that is utilised for CO2 production
m(5,11) = -Y.su*N.bm;                                   //consumption of nitrogen into growth of monosaccharide
                                                          degraders
m(5,17) = Y.su;
kr(5) = km.su*X_su*S_su/(KS.su+S_su)* ...
        I_pH(pH)* ...
        I_IN_lim(S_IN,KS.IN);


//Process 6: Uptake of Amino Acids
//--------------------------------
m(6,2) = -1;
m(6,4) = (1-Y.aa)*f.va_aa;
m(6,5) = (1-Y.aa)*f.bu_aa;
m(6,6) = (1-Y.aa)*f.pro_aa;
m(6,7) = (1-Y.aa)*f.ac_aa;
m(6,8) = (1-Y.aa)*f.h2_aa;
m(6,10) = C.aa - ...
          (1-Y.aa)*f.va_aa*C.va - ...
          (1-Y.aa)*f.bu_aa*C.bu - ...
          (1-Y.aa)*f.pro_aa*C.pro - ...
          (1-Y.aa)*f.ac_aa*C.ac - ...
          Y.aa*C.bm;
m(6,11) = N.aa - Y.aa*N.bm;
m(6,18) = Y.aa;
kr(6) = km.aa*X_aa*S_aa/(KS.aa+S_aa)* ...
        I_pH(pH)* ...
        I_IN_lim(S_IN,KS.IN);




//Process 7: Uptake of LCFA
```

```
//-------------------------
m(7,3) = -1;
m(7,7) = 0.7*(1-Y.fa);
m(7,8) = 0.3*(1-Y.fa);
m(7,10) = C.fa - ...
          0.7*(1-Y.fa)*C.ac - ...
          Y.fa*C.bm;                                                    //Note 1
m(7,11) = -Y.fa*N.bm;
m(7,19) = Y.fa;
kr(7) = km.fa*X_fa*S_fa/(KS.fa+S_fa)* ...
        I_pH(pH)* ...
        I_IN_lim(S_IN,KS.IN)*I_h2(S_h2,KI.h2_fa);


//Process 8: Uptake of Valerate
//-----------------------------
m(8,4) = -1;
m(8,6) = 0.54*(1-Y.c4);
m(8,7) = 0.31*(1-Y.c4);
m(8,8) = 0.15*(1-Y.c4);
m(8,10) = C.va - ...
          0.54*(1-Y.c4)*C.pro - ...
          0.31*(1-Y.c4)*C.ac - ...
          Y.c4*C.bm;                                                    //Note 1
m(8,11) = -Y.c4*N.bm;
m(8,20) = Y.c4;
kr(8) = km.c4*X_c4*S_va/(KS.c4+S_va)*1/(1+S_bu/S_va)* ...
        I_pH(pH)* ...
        I_IN_lim(S_IN,KS.IN)* ...
        I_h2(S_h2,KI.h2_c4);


//Process 9: Uptake of Butyrate
//-----------------------------
m(9,5) = -1;
m(9,7) = 0.8*(1-Y.c4);
m(9,8) = 0.2*(1-Y.c4);
m(9,10) = C.bu - ...
          0.8*(1-Y.c4)*C.ac - ...
          Y.c4*C.bm;                                                    //Note 1
m(9,11) = -Y.c4*N.bm;
m(9,20) = Y.c4;
kr(9) = km.c4*X_c4*S_bu/(KS.c4+S_bu)*1/(1+S_va/S_bu)* ...
        I_pH(pH)* ...
        I_IN_lim(S_IN,KS.IN)* ...
        I_h2(S_h2,KI.h2_c4);


//Process 10: Uptake of Propionate
//--------------------------------
m(10,6) = -1;
m(10,7) = 0.57*(1-Y.pro);
m(10,8) = 0.43*(1-Y.pro);
m(10,10) = C.pro - ...
           0.57*(1-Y.pro)*C.ac - ...
           Y.pro*C.bm;

m(10,11) = -Y.pro*N.bm;
m(10,21) = Y.pro;
kr(10) = km.pro*X_pro*S_pro/(KS.pro+S_pro)* ...
         I_pH(pH)* ...
         I_IN_lim(S_IN,KS.IN)* ...
         I_h2(S_h2,KI.h2_pro);


//Process 11: Uptake of Acetate
//-----------------------------
m(11,7) = -1;
m(11,9) = (1-Y.ac);
m(11,10) = C.ac - ...
           (1-Y.ac)*C.ch4 - ...
           Y.ac*C.bm;
m(11,11) = -Y.ac*N.bm;
```

```
m(11,22) = Y.ac;
kr(11) = km.ac*X_ac*S_ac/(KS.ac+S_ac)* ...
            I_pH_ac(pH)* ...
            I_IN_lim(S_IN,KS.IN)* ...
            I_nh3_ac(S_nh3,KI.nh3_ac);


//Process 12: Uptake of Hydrogen
//------------------------------
m(12,8) = -1;
m(12,9) = (1-Y.h2);
m(12,10) = -(1-Y.h2)*C.ch4 - Y.h2*C.bm;
m(12,11) = -Y.h2*N.bm;
m(12,23) = Y.h2;
kr(12) = km.h2*X_h2*S_h2/(KS.h2+S_h2)* ...
            I_pH_h2(pH)*...
            I_IN_lim(S_IN,KS.IN);


//Process 13: Decay of Monosaccharide Degraders
//---------------------------------------------
m(13,10) = C.bm - C.Xc;                                          //Note 1
m(13,11) = N.bm - N.Xc;                                          //Note 1
m(13,13) = 1;
m(13,17) = -1;
kr(13) = kdec.xsu*X_su;


//Process 14: Decay of Amino Acid Degraders
//-----------------------------------------
m(14,10) = C.bm - C.Xc;                                          //Note 1
m(14,11) = N.bm - N.Xc;                                          //Note 1
m(14,13) = 1;
m(14,18) = -1;
kr(14) = kdec.xaa*X_aa;


//Process 15: Decay of LCFA Degraders
//-----------------------------------
m(15,10) = C.bm - C.Xc;                                          //Note 1
m(15,11) = N.bm - N.Xc;                                          //Note 1
m(15,13) = 1;
m(15,19) = -1;
kr(15) = kdec.xfa*X_fa;


//Process 16: Decay of C4 Degraders
//---------------------------------
m(16,10) = C.bm - C.Xc;                                          //Note 1
m(16,11) = N.bm - N.Xc;                                          //Note 1
m(16,13) = 1;
m(16,20) = -1;
kr(16) = kdec.xc4*X_c4;


//Process 17: Decay of Propionate Degraders
//-----------------------------------------
m(17,10) = C.bm - C.Xc;                                          //Note 1
m(17,11) = N.bm - N.Xc;                                          //Note 1
m(17,13) = 1;
m(17,21) = -1;
kr(17) = kdec.xpro*X_pro;


//Process 18: Decay of Acetate Degraders
//--------------------------------------
m(18,10) = C.bm - C.Xc;                                          //Note 1
m(18,11) = N.bm - N.Xc;                                          //Note 1
m(18,13) = 1;
m(18,22) = -1;
kr(18) = kdec.xac*X_ac;

//Process 19: Decay of Hydrogen Degraders
```

```
//---------------------------------------
m(19,10) = C.bm - C.Xc;                                                    //Note 1
m(19,11) = N.bm - N.Xc;                                                    //Note 1
m(19,13) = 1;
m(19,23) = -1;
kr(19) = kdec.xh2*X_h2;


//State variables balance - Digester
//-----------------------------------
DSV_change = m'*kr;                                          //totaling concentration changes (summation of all
                                                              rows in matrix) for each dynamic state variable


DSV_change(8) = DSV_change(8) - kr_h2;                                     //gas-liquid transfer kinetic rates
DSV_change(9) = DSV_change(9) - kr_ch4;
DSV_change(10) = DSV_change(10) - kr_co2;
DSV_change = DSV_change';

t0 = floor(t)+1;
I = inf_data(t0,:)+...
      (t-t_data(t0))/(t_data(t0+1)-t_data(t0))...
      .*(inf_data(t0+1,:)-inf_data(t0,:));                   //interpolate influent composition relative to influent
                                                              data and t in ode function

Q_I = Q_in(t0)+...
      (t-t_data(t0))/(t_data(t0+1)-t_data(t0))...
      .*(Q_in(t0+1)-Q_in(t0));                              //interpolate influent flow relative to influent data and
                                                              t in ode function

Q_was_I = Q_was(t0)+...
      (t-t_data(t0))/(t_data(t0+1)-t_data(t0))...
      .*(Q_was(t0+1)-Q_was(t0));                            //interpolate WAns flow relative to influent data and t
                                                              in ode function

//C_was_I = C_was(t0)+...
//      (t-t_data(t0))/(t_data(t0+1)-t_data(t0))...
//      .*(C_was(t0+1)-C_was(t0));                          //interpolate WAns concentration relative to influent
                                                              data and t in ode function


Q_eff = Q_I;

for i = 1 : 12
      dDSVdt(i) = Q_I*I(i)/V_digester - ...
                  Q_eff*DSV(i)/V_digester + ...
                  DSV_change(i);                                           //mass balance around digester
end

for j = 13 : 24
      f_X = DSV(j)/sum(DSV(13:24));
      dDSVdt(j) = Q_I*I(j)/V_digester - ...
                  Q_was_I*DSV(j)/V_digester + ...
                  DSV_change(j);                                           //mass balance around digester
end

for k = 25 : 26
      dDSVdt(k) = Q_I*I(k)/V_digester - ...
                  Q_eff*DSV(k)/V_digester + ...
                  DSV_change(k);                                           //mass balance around digester
end

//State variables balance - Headspace
//-----------------------------------
dDSVdt(27) = -S_h2_g*q_gas/V_headspace + kr_h2*V_digester/V_headspace;
dDSVdt(28) = -S_ch4_g*q_gas/V_headspace + kr_ch4*V_digester/V_headspace;   //gas phase mass balances
dDSVdt(29) = -S_co2_g*q_gas/V_headspace + kr_co2*V_digester/V_headspace;

endfunction
```

```
//=============================================================================================
//ADM1 MODEL
//IWA TASK GROUP, SCIENTIFIC AND TECHNICAL REPORT NO.13
//File: Parameters.sce
//=============================================================================================

//1. Physico-chemical
//====================
R = 0.08314;                                                    //[bar.K-1.M-1] ideal gas law constant
kLa = 200;                                                      //[d-1] oxygen liquid-gas transfer coefficient

pKa.ac = 4.76;                                                  //pKa of acetate at T=298K
pKa.bu = 4.84;                                                  //pKa of butyrate at T=298K
pKa.co2 = 6.35;                                                 //pKa of CO2 at T=298K
pKa.h2o = 14;                                                   //pKa of water at T=298K
pKa.nh4 = 9.25;                                                 //pKa of ammonium at T=298K
pKa.pro = 4.88;                                                 //pKa of propionate at T=298K
pKa.va = 4.8;                                                   //pKa of valerate at T=298K
deltaH0_Ka.co2 = 7646;                                          //[J] enthalpy of reaction: CO2 -> HCO3
deltaH0_Ka.h2o = 55900;                                         //[J] enthalpy of reaction: h2o -> OH + H+
deltaH0_Ka.nh4 = 51965;                                         //[J] enthalpy of reaction: nh4 -> nh3
deltaH0_KH.ch4 = -14240;                                        //[J] enthalpy of reaction of CH4(g) -> CH4(l)
deltaH0_KH.co2 = -19410;                                        //[J] enthalpy of reaction of CO2(g) -> CO2(l)
deltaH0_KH.h2 = -4180;                                          //[J] enthalpy of reaction of h2(g) -> h2(l)

k_AB_co2 = 10^14;                                               //kinetic constant for Acid-Base CO2-HCO3 reaction
Ka.ac = 10^(-pKa.ac);                                           //acetate acidity constant (without temperature correction)
Ka.bu = 10^(-pKa.bu);                                           //butyrate acidity constant (without tempeature correction)
Ka.co2 = 10^(-pKa.co2)*exp(deltaH0_Ka.co2/(R*100)*(1/298-1/T));  //CO2 acidity constant (temperature corrected)
Ka.h2o = 10^(-pKa.h2o)*exp(deltaH0_Ka.h2o/(R*100)*(1/298-1/T));  //water acidity constant (temperature corrected)
Ka.nh4 = 10^(-pKa.nh4)*exp(deltaH0_Ka.nh4/(R*100)*(1/298-1/T));  //ammonium acidity constant (temperature corrected)
Ka.pro = 10^(-pKa.pro);                                         //propionate acidity constant (without temperature correction)
Ka.va = 10^(-pKa.va);                                           //valerate acidity constant (without temperature correction)

KH.ch4 = 0.0014*exp(deltaH0_KH.ch4/(R*100)*(1/298-1/T));        //[Mliq.Mgas-1] non-dimensional Henry's law constant for
                                                                  CH4 (temperature corrected)
KH.co2 = 0.035*exp(deltaH0_KH.co2/(R*100)*(1/298-1/T));         //[Mliq.Mgas-1] non-dimensional Henry's law constant for
                                                                  CO2 (temperature corrected)
KH.h2 = 0.00078*exp(deltaH0_KH.h2/(R*100)*(1/298-1/T));         //[Mliq.Mgas-1] non-dimensional Henry's law constant for
                                                                  H2 (temperature corrected)

p_h2o_0 = 0.0313;                                               //[bar] vapour pressure of water at STP
T_0 = 298;                                                      //[K] temp at STP
deltaHvap0_h2o = 43980;                                         //[J.mol-1] heat of vaporisation at STP

//2. Biochemical Stoichiometric
//==============================
C.ch = 0.0313;                                                  //[mol.gCOD-1] carbon content in carbohydrates
C.pr = 0.03;                                                    //[mol.gCOD-1] carbon content in proteins
C.li = 0.022;                                                   //[mol.gCOD-1] carbon content in lipids
C.su = 0.0313;                                                  //[mol.gCOD-1] carbon content in monosaccharides
C.aa = 0.03;                                                    //[mol.gCOD-1] carbon content in amino acids
C.fa = 0.0217;                                                  //[mol.gCOD-1] carbon content in LCFA
C.va = 0.024;                                                   //[mol.gCOD-1] carbon content in valerate
C.bu = 0.025;                                                   //[mol.gCOD-1] carbon content in butyrate
C.pro = 0.0268;                                                 //[mol.gCOD-1] carbon content in propionate
C.ac = 0.0313;                                                  //[mol.gCOD-1] carbon content in acetate
C.ch4 = 0.0156;                                                 //[mol.gCOD-1] carbon content in methane
C.bm = 0.0313;                                                  //[mol.gCOD-1] carbon content in biomass
C.SI = 0.03;                                                    //[mol.gCOD-1] carbon content in soluble inert COD
C.Xc = 0.02                                                     //[mol.gCOD-1] carbon content in composite particulates
C.XI = 0.03;                                                    //[mol.gCOD-1] carbon content in particulate inert COD

N.aa = 0.007;                                                   //[mol.gCOD-1] nitrogen content in amino acids
N.bm = 0.08/14;                                                 //[mol.gCOD-1] nitrogen content in biomass, Note 1
N.SI = 0.06/14;                                                 //[mol.gCOD-1] nitrogen content of soluble inert COD, Note 1
N.Xc = 0.0376/14;                                               //[mol.gCOD-1] nitrogen content of composite particulates, Note 1
N.XI = 0.06/14;                                                 //[mol.gCOD-1] nitrogen content of particulate inert COD, Note 1

nu.su_1 = 0.495;                                                //fraction of monosaccharides that degrades to acetate only
nu.su_2 = 0.345;                                                //fraction of monosaccharides that degrades to acetate and propionate
nu.su_3 = 1-nu.su_1-nu.su_2;                                    //fraction of monosaccharides that degrades to butyrate only
```

```
f.pro_su = 0.78*nu.su_2;                    //[kgCOD.kgCOD-1] catabolic yield of propionate from monosaccharides
f.bu_su = 0.83*nu.su_3;                     //[kgCOD.kgCOD-1] catabolic yield of butyrate from monosaccharides
f.ac_su = 0.67*nu.su_1+0.22*nu.su_2;        //[kgCOD.kgCOD-1] catabolic yield of acetate from monosaccharides
f.h2_su = 0.33*nu.su_1+0.17*nu.su_3;        //[kgCOD.kgCOD-1] catabolic yield of hydrogen from monosaccharides

f.va_aa = 0.23;                             //[kgCOD.kgCOD-1] catabolic yield of valerate from amino acids
f.bu_aa = 0.26;                             //[kgCOD.kgCOD-1] catabolic yield of butyrate from amino acids
f.pro_aa = 0.05;                            //[kgCOD.kgCOD-1] catabolic yield of propionate from amino acids
f.ac_aa = 0.4;                              //[kgCOD.kgCOD-1] catabolic yield of acetate from amino acids
f.h2_aa = 0.06;                             //[kgCOD.kgCOD-1] catabolic yield of hydrogen from amino acids

f.fa_li = 0.95;                             //[kgCOD.kgCOD-1] catabolic yield of LCFAs from lipids

f.ch_xc = 0.2;                              //[kgCOD.kgCOD-1] catabolic yield of carbohydrates from composite particulates
f.pr_xc = 0.2;                              //[kgCOD.kgCOD-1] catabolic yield of protein from composite particulates
f.si_xc = 0.1;                              //[kgCOD.kgCOD-1] catabolic yield of soluble inerts from composite particulates
f.xi_xc = 0.25;                             //[kgCOD.kgCOD-1] catabolic yield of particulate inerts from composite particulates
f.li_xc = 1-f.ch_xc-f.pr_xc-f.si_xc-f.xi_xc; //[kgCOD.kgCOD-1] catabolic yield of lipids from composites particulates

M.h2 = 16;                                  //[kgCOD.kmol-1] molecular mass of hydrogen
M.ch4 = 64;                                 //[kgCOD.kmol-1] molecular mass of methane
cpc.ac = 64;                                //[gCOD.charge-1] COD content per charge of acetate
cpc.pro = 112;                              //[gCOD.charge-1] COD content per charge of propionate
cpc.bu = 160;                               //[gCOD.charge-1] COD content per charge of butyrate
cpc.va = 208;                               //[gCOD.charge-1] COD content per charge of valerate

//3. Biochemical Kinetic
//=======================
kdis = 0.4;                                 //[d-1] disintegration rate constant of composite particulates
khyd.ch = 0.25;                             //[d-1] carbohydrates hydrolysis rate constant
khyd.li = 0.1;                              //[d-1] lipids hydrolysis rate constant
khyd.pr = 0.2;                              //[d-1] proteins hydrolysis rate constant

km.su = 30;                                 //[kgCODS.kgCODX.d-1] specific Monod maximum uptake rate for monosaccharide degraders
km.aa = 50;                                 //[kgCODS.kgCODX.d-1] specific Monod maximum uptake rate for amino acid degraders
km.fa = 6;                                  //[kgCODS.kgCODX.d-1] specific Monod maximum uptake rate for LCFA degraders
km.c4 = 20;                                 //[kgCODS.kgCODX.d-1] specific Monod maximum uptake rate for butyrate and valerate degraders
km.pro = 13;                                //[kgCODS.kgCODX.d-1] specific Monod maximum uptake rate for propionate degraders
km.ac = 8;                                  //[kgCODS.kgCODX.d-1] specific Monod maximum uptake rate for acetate degraders
km.h2 = 35;                                 //[kgCODS.kgCODX.d-1] specific Monod maximum uptake rate for hydrogen degraders

kdec.xsu = 0.02;                            //[d-1] decay rate constant of monosaccharide degraders
kdec.xaa = 0.02;                            //[d-1] decay rate constant of amino acid degraders
kdec.xfa = 0.02;                            //[d-1] decay rate constant of LCFA degraders
kdec.xpro = 0.02;                           //[d-1] decay rate constant of propionate degraders
kdec.xc4 = 0.02;                            //[d-1] decay rate constant of butyrate and valerate degraders
kdec.xac = 0.02;                            //[d-1] decay rate constant of acetate degraders
kdec.xh2 = 0.02;                            //[d-1] decay rate constant of hydrogen degraders

KS.su = 0.5;                                //[kgCOD.m-3] Monod half saturation constant for monosaccharide degradation
KS.aa = 0.3;                                //[kgCOD.m-3] Monod half saturation constant for amino acid degradation
KS.fa = 0.4;                                //[kgCOD.m-3] Monod half saturation constant for LCFA degradation
KS.c4 = 0.3;                                //[kgCOD.m-3] Monod half saturation constant for butyrate and valerate degradation
KS.pro = 0.3;                               //[kgCOD.m-3] Monod half saturation constant for propionate degradation
KS.ac = 0.15;                               //[kgCOD.m-3] Monod half saturation constant for acetate degradation
KS.h2 = 2.5*10^-5;                          //[kgCOD.m-3] Monod half saturation constant for hydrogen uptake
KS.IN = 0.0001;                             //[M] inorganic nitrogen concentration at which growth ceases

KI.h2_c4 = 10^-5;                           //[kgCOD.m-3] hydrogen inhibitory concentration for butyrate and valerate degraders
KI.h2_fa = 5 * 10^-6;                       //[kgCOD.m-3] hydrogen inhibitory concentration for LCFA degraders
KI.h2_pro = 3.5*10^-6;                      //[kgCOD.m-3] hydrogen inhibitory concentration for propionate degraders
KI.nh3_ac = 0.0018;                         //[kgCOD.m-3] free ammonia inhibitory concentration for acetate degraders

Y.su = 0.1;                                 //[kgCODX.kgCODS-1] yield of biomass on uptake of monosaccharides
Y.aa = 0.08;                                //[kgCODX.kgCODS-1] yield of biomass on uptake of amino acids
Y.fa = 0.06;                                //[kgCODX.kgCODS-1] yield of biomass on uptake of LCFA
Y.c4 = 0.06;                                //[kgCODX.kgCODS-1] yield of biomass on uptake of valerate or butyrate
Y.pro = 0.04;                               //[kgCODX.kgCODS-1] yield of biomass on uptake of propionate
Y.ac = 0.05;                                //[kgCODX.kgCODS-1] yield of biomass on uptake of acetate
Y.h2 = 0.06;                                //[kgCODX.kgCODS-1] yield of biomass on uptake of hydrogen

I_pH_ll = 4;                                //pH level at which full inhibition applies
I_pH_ul = 5.5;                              //pH level at which no inhibition applies
```

```
l_pH_ac_ll = 6;                    //pH level at which full inhibition of acetate degradation applies
l_pH_ac_ul = 7;                    //pH level at which no inhibition of acetate degradation applies
l_pH_h2_ll = 5;                    //pH level at which full inhibition of hydrogen degradation applies
l_pH_h2_ul = 6;                    //pH level at which no inhibition of hydrogen degradation applies
```

```
//=============================================================================================================
//ADM1 MODEL
//IWA TASK GROUP, SCIENTIFIC AND TECHNICAL REPORT NO.13
//File: Rand_Parameters.sce
//=============================================================================================================


//Biochemical Stoichiometric
//==============================
f.li_xc = -1;
while (f.li_xc < surv_min(5))
        f.si_xc = surv_min(1)+(surv_max(1)-surv_min(1))*rand();
        f.xi_xc = surv_min(2)+(surv_max(2)-surv_min(2))*rand();
        f.ch_xc = surv_min(3)+(surv_max(3)-surv_min(3))*rand();
        f.pr_xc = surv_min(4)+(surv_max(4)-surv_min(4))*rand();
        f.li_xc = 1-f.ch_xc-f.pr_xc-f.si_xc-f.xi_xc;
        while (f.li_xc > surv_max(5))
                f.si_xc = surv_min(1)+(surv_max(1)-surv_min(1))*rand();
                f.xi_xc = surv_min(2)+(surv_max(2)-surv_min(2))*rand();
                f.ch_xc = surv_min(3)+(surv_max(3)-surv_min(3))*rand();
                f.pr_xc = surv_min(4)+(surv_max(4)-surv_min(4))*rand();
                f.li_xc = 1-f.ch_xc-f.pr_xc-f.si_xc-f.xi_xc;
        end
 end

f.fa_li = surv_min(6)+(surv_max(6)-surv_min(6))*rand();

f.ac_su = -1;
while (f.ac_su < surv_min(10))
      f.h2_su = surv_min(7)+(surv_max(7)-surv_min(7))*rand();
      f.bu_su = surv_min(8)+(surv_max(8)-surv_min(8))*rand();
      f.pro_su = surv_min(9)+(surv_max(9)-surv_min(9))*rand();
      f.ac_su = 1-f.h2_su-f.bu_su-f.pro_su;
      while (f.ac_su > surv_max(10))
            f.h2_su = surv_min(7)+(surv_max(7)-surv_min(7))*rand();
            f.bu_su = surv_min(8)+(surv_max(8)-surv_min(8))*rand();
            f.pro_su = surv_min(9)+(surv_max(9)-surv_min(9))*rand();
            f.ac_su = 1-f.h2_su-f.bu_su-f.pro_su;
      end
end

f.ac_aa = -1;
while (f.ac_aa < surv_min(15))
      f.h2_aa = surv_min(11)+(surv_max(11)-surv_min(11))*rand();
      f.va_aa = surv_min(12)+(surv_max(12)-surv_min(12))*rand();
      f.bu_aa = surv_min(13)+(surv_max(13)-surv_min(13))*rand();
      f.pro_aa = surv_min(14)+(surv_max(14)-surv_min(14))*rand();
      f.ac_aa = 1-f.h2_aa-f.va_aa-f.bu_aa-f.pro_aa;
      while (f.ac_aa > surv_max(15))
            f.h2_aa = surv_min(11)+(surv_max(11)-surv_min(11))*rand();
            f.va_aa = surv_min(12)+(surv_max(12)-surv_min(12))*rand();
            f.bu_aa = surv_min(13)+(surv_max(13)-surv_min(13))*rand();
            f.pro_aa = surv_min(14)+(surv_max(14)-surv_min(14))*rand();
            f.ac_aa = 1-f.h2_aa-f.va_aa-f.bu_aa-f.pro_aa;
      end
end

//Biochemical Kinetic
//=======================
kdis = surv_min(16)+(surv_max(16)-surv_min(16))*rand();
khyd.ch = surv_min(17)+(surv_max(17)-surv_min(17))*rand();
khyd.pr = surv_min(18)+(surv_max(18)-surv_min(18))*rand();
khyd.li = surv_min(19)+(surv_max(19)-surv_min(19))*rand();

KS.IN = surv_min(20)+(surv_max(20)-surv_min(20))*rand();

l_pH_ll = 10;
while l_pH_ll > l_pH_ul
      l_pH_ul = surv_min(21)+(surv_max(21)-surv_min(21))*rand();
      l_pH_ll = surv_min(22)+(surv_max(22)-surv_min(22))*rand();
end

km.su = surv_min(23)+(surv_max(23)-surv_min(23))*rand();
```

```
KS.su = surv_min(24)+(surv_max(24)-surv_min(24))*rand();
Y.su = surv_min(25)+(surv_max(25)-surv_min(25))*rand();
kdec.xsu = surv_min(26)+(surv_max(26)-surv_min(26))*rand();

km.aa = surv_min(27)+(surv_max(27)-surv_min(27))*rand();
KS.aa = surv_min(28)+(surv_max(28)-surv_min(28))*rand();
Y.aa = surv_min(29)+(surv_max(29)-surv_min(29))*rand();
kdec.xaa = surv_min(30)+(surv_max(30)-surv_min(30))*rand();

km.fa = surv_min(31)+(surv_max(31)-surv_min(31))*rand();
KS.fa = surv_min(32)+(surv_max(32)-surv_min(32))*rand();
Y.fa = surv_min(33)+(surv_max(33)-surv_min(33))*rand();
kdec.xfa = surv_min(34)+(surv_max(34)-surv_min(34))*rand();
Kl.h2_fa = surv_min(35)+(surv_max(35)-surv_min(35))*rand();

km.c4 = surv_min(36)+(surv_max(36)-surv_min(36))*rand();
KS.c4 = surv_min(37)+(surv_max(37)-surv_min(37))*rand();
Y.c4 = surv_min(38)+(surv_max(38)-surv_min(38))*rand();
kdec.xc4 = surv_min(39)+(surv_max(39)-surv_min(39))*rand();
Kl.h2_c4 = surv_min(40)+(surv_max(40)-surv_min(40))*rand();

km.pro = surv_min(41)+(surv_max(41)-surv_min(41))*rand();
KS.pro = surv_min(42)+(surv_max(42)-surv_min(42))*rand();
Y.pro = surv_min(43)+(surv_max(43)-surv_min(43))*rand();
kdec.xpro = surv_min(44)+(surv_max(44)-surv_min(44))*rand();
Kl.h2_pro = surv_min(45)+(surv_max(45)-surv_min(45))*rand();

km.ac = surv_min(46)+(surv_max(46)-surv_min(46))*rand();
KS.ac = surv_min(47)+(surv_max(47)-surv_min(47))*rand();
Y.ac = surv_min(48)+(surv_max(48)-surv_min(48))*rand();
kdec.xac = surv_min(49)+(surv_max(49)-surv_min(49))*rand();
Kl.nh3_ac = surv_min(50)+(surv_max(50)-surv_min(50))*rand();
l_pH_ac_ul = surv_min(51)+(surv_max(51)-surv_min(51))*rand();
l_pH_ac_ll = surv_min(52)+(surv_max(52)-surv_min(52))*rand();

km.h2 = surv_min(53)+(surv_max(53)-surv_min(53))*rand();
KS.h2 = surv_min(54)+(surv_max(54)-surv_min(54))*rand();
Y.h2 = surv_min(55)+(surv_max(55)-surv_min(55))*rand();
kdec.xh2 = surv_min(56)+(surv_max(56)-surv_min(56))*rand();
l_pH_h2_ul = surv_min(57)+(surv_max(57)-surv_min(57))*rand();
l_pH_h2_ll = surv_min(58)+(surv_max(58)-surv_min(58))*rand();
```

```
//=========================================================================================================
//IWA TASK GROUP, SCIENTIFIC AND TECHNICAL REPORT NO.13
//File: ADM1.sce
//=========================================================================================================

t = linspace(t_0,sim_dur,sim_dur+1);                                                    //simulation range
%ODEOPTIONS=[1,0,0,%inf,0,2,1000,12,5,0,-1,-1];                                         //ODE configurations
DSV = ode(DSV_0,t_0,t,f_dDSVdt);                                                        //calling ODE function: f_dDSVdt

//pH & Gas Flows
//================
pH = 0;
q_h2 = 0;
q_ch4 = 0;
q_co2 = 0;

for i = 1:size(DSV,'c')
    h_solver_list = list(charge_bal,DSV(26,i),DSV(11,i),DSV(10,i),DSV(7,i),DSV(6,i),DSV(5,i),DSV(4,i),DSV(25,i),Ka,cpc);
    S_h_ion = fsolve(S_h_ion,h_solver_list);
    pH(length(pH)+1) = -log10(S_h_ion);                                                 //pH of reactor content

    S_h2 = DSV(8,i);
    S_ch4 = DSV(9,i);
    S_co2 = DSV(10,i)-(Ka.co2*DSV(10,i)/(Ka.co2+S_h_ion));

    p_h2 = DSV(27,i)/M.h2*R*T;                                                          //partial pressure of h2 [bar]
    p_ch4 = DSV(28,i)/M.ch4*R*T;                                                        //partial pressure of CH4 [bar]
    p_co2 = DSV(29,i)*R*T;                                                              //partial pressure of CO2 [bar]
    p_h2o = p_h2o_0*exp(deltaHvap0_h2o/(R*100)*(1/T_0-1/T));                            //[bar] partial pressure of h2o, van't Hoff equation
    P_headspace = p_co2+p_h2+p_ch4+p_h2o;                                               //total gas phase pressure [bar]

    kr_h2 = kLa*(S_h2-M.h2*KH.h2*p_h2);
    kr_ch4 = kLa*(S_ch4-M.ch4*KH.ch4*p_ch4);
    kr_co2 = kLa*(S_co2-KH.co2*p_co2);

    q_gas = R*T/(P_atm-p_h2o)*V_digester*(kr_h2/M.h2+kr_ch4/M.ch4+kr_co2);               //total gas flow [m3.d-1]
    q_h2(length(q_h2)+1) = p_h2/P_headspace*q_gas;
    q_ch4(length(q_ch4)+1) = p_ch4/P_headspace*q_gas;
    q_co2(length(q_co2)+1) = p_co2/P_headspace*q_gas;
end

DSV($+1,:) = pH(2:length(pH))';
DSV($+1,:) = q_h2(2:length(q_h2))';
DSV($+1,:) = q_ch4(2:length(q_ch4))';
DSV($+1,:) = q_co2(2:length(q_co2))';
for j = 2 : size(DSV,2)
    MC_data($+1,:) = DSV(:,j)';
end

DSV = DSV';

VFA_s = [DSV(:,4)+DSV(:,5)+DSV(:,6)+DSV(:,7)];
VSS_s =
[DSV(:,13)+DSV(:,14)+DSV(:,15)+DSV(:,16)+DSV(:,17)+DSV(:,18)+DSV(:,19)+DSV(:,20)+DSV(:,21)+DSV(:,22)+DSV(:,23)];

for i = 1 : sim_dur
    VFA_data($+1) = VFA_s(i);
    VSS_data($+1) = VSS_s(i);
end

DSV = DSV';

if sim_mode == "single" then
Ran_Par =
[f.si_xc,f.xi_xc,f.ch_xc,f.pr_xc,f.li_xc,f.fa_li,f.h2_su,f.bu_su,f.pro_su,f.ac_su,f.h2_aa,f.va_aa,f.bu_aa,f.pro_aa,f.ac_aa,kdis,khyd.c
h,khyd.pr,khyd.li,KS.IN,I_pH_ul,I_pH_ll,km.su,KS.su,Y.su,kdec.xsu,km.aa,KS.aa,Y.aa,kdec.xaa,km.fa,KS.fa,Y.fa,kdec.xfa,KI.h2
_fa,km.c4,KS.c4,Y.c4,kdec.xc4,KI.h2_c4,km.pro,KS.pro,Y.pro,kdec.xpro,KI.h2_pro,km.ac,KS.ac,Y.ac,kdec.xac,KI.nh3_ac,I_pH_
ac_ul,I_pH_ac_ll,km.h2,KS.h2,Y.h2,kdec.xh2,I_pH_h2_ul,I_pH_h2_ll];
                                                                                        //keep track of the random parameters used for each Monte-Carlo run
end

if sim_mode == "monte-carlo" then
```

```
Ran_Par(mont_car+1,:) =
[f.si_xc,f.xi_xc,f.ch_xc,f.pr_xc,f.li_xc,f.fa_li,f.h2_su,f.bu_su,f.pro_su,f.ac_su,f.h2_aa,f.va_aa,f.bu_aa,f.pro_aa,f.ac_aa,kdis,khyd.c
h,khyd.pr,khyd.li,KS.IN,I_pH_ul,I_pH_ll,km.su,KS.su,Y.su,kdec.xsu,km.aa,KS.aa,Y.aa,kdec.xaa,km.fa,KS.fa,Y.fa,kdec.xfa,KI.h2
_fa,km.c4,KS.c4,Y.c4,kdec.xc4,KI.h2_c4,km.pro,KS.pro,Y.pro,kdec.xpro,KI.h2_pro,km.ac,KS.ac,Y.ac,kdec.xac,KI.nh3_ac,I_pH_
ac_ul,I_pH_ac_ll,km.h2,KS.h2,Y.h2,kdec.xh2,I_pH_h2_ul,I_pH_h2_ll];
                                                    //keep track of the random parameters used for each Monte-Carlo run
end
```

```
//=============================================================================================
//ADM1 MODEL
//IWA TASK GROUP, SCIENTIFIC AND TECHNICAL REPORT NO.13
//File: PLSR.sce
//=============================================================================================

//PLSR Initialisation
//====================
output_no = size(MC_data_pls,2);

pls_data_sel = zeros(MC_nr,output_no * sim_dur);
row_count = 1;

for i = 0 : MC_nr
    col_count = 1;
    for j = 1 : sim_dur
        row = i * sim_dur + j;
        pls_data_sel(row_count,col_count:j*output_no) = MC_data_pls(row,:);
        col_count = col_count + output_no;
    end
    row_count = row_count + 1;
end

pls_data_sel = pls_data_sel(2:size(pls_data_sel,1),:);

//PLSR Algorithm
//==============

Y_pls = pls_data_sel;                                                    //Raw output data - Y_pls
X_pls = pls_Ran_Par;                                                     //Raw input data - X_pls


Y_sort = gsort(Y_pls,'r','d');
Y_IQR = iqr(Y_sort,1);
qt_index = 3/4*(size(Y_sort,1)+1);
lower_qt = (Y_sort(ceil(qt_index),:)+Y_sort(floor(qt_index),:))/2;
upper_qt = lower_qt + Y_IQR;

low_bound = lower_qt - 99 * Y_IQR;
up_bound = upper_qt + 99 * Y_IQR;
row_remv = 0;

for row_Y = 1 : size(Y_pls,1)
    outlr = "No"
    for col_Y = 1 : size(Y_pls,2)
        if Y_pls(row_Y,col_Y) < low_bound(col_Y) then
            outlr = "Yes";
            disp(row_Y)
            disp(col_Y)
            row_remv($+1) = row_Y;
        end
        if Y_pls(row_Y,col_Y) > up_bound(col_Y) then
            outlr = "Yes";
            disp(row_Y)
            disp(col_Y)
            row_remv($+1) = row_Y
        end
    end
end

Y_new = zeros(1,size(Y_pls,2));
X_new = zeros(1,size(X_pls,2));
for row_Y = 1 : size(Y_pls,1)
    remove = "No"
    for k = 1 : length(row_remv)
        if row_Y == row_remv(k) then
            remove = "Yes"
        end
    end
    if remove == "No" then
        Y_new($+1,:) = Y_pls(row_Y,:)
        X_new($+1,:) = X_pls(row_Y,:)
    end
```

```
end

Y_pls = Y_new(2:size(Y_new,1),:);
X_pls = X_new(2:size(X_new,1),:);

Y_0 = zeros(size(Y_pls,1),size(Y_pls,2));
X_0 = zeros(size(X_pls,1),size(X_pls,2));
W_1 = zeros(size(X_pls,2),1);
W_pls = zeros(size(X_pls,2),no_LV);
P_pls = zeros(size(X_pls,2),no_LV);
T_pls = zeros(size(X_pls,1),no_LV);
Q_pls = zeros(size(Y_pls,2),no_LV);
U_pls = zeros(size(Y_pls,1),no_LV);
b_coeff = zeros(no_LV,1);
Y_pred = zeros(size(Y_pls,1),size(Y_pls,2));
R_pls_norm = zeros(size(X_pls,2),no_LV);
LV_Y = zeros(size(Y_pls,1),no_LV * size(Y_pls,2));
mean_Y = zeros(size(Y_pls,1),size(Y_pls,2));

Y_min = min(Y_pls,"r");
Y_max = max(Y_pls,"r");

numrows = size(Y_pls,1);
for i = 1:numrows                                                    //Normalise & standardise Y_pls dataset
    Y_0(i,:) = 2*((Y_pls(i,:) - Y_min) ./ (Y_max - Y_min)) - 1;
//      Y_0(i,:) = ((Y_pls(i,:) - Y_min) ./ (Y_max - Y_min));
end

Y_0i = Y_0;

numrows = size(X_pls,1);
for i = 1:numrows                                                    //Normalise & standardise X_pls dataset
    X_0(i,:) = 2*((X_pls(i,:) - surv_min') ./ (surv_max' - surv_min')) - 1;
//      X_0(i,:) = ((X_pls(i,:) - surv_min') ./ (surv_max' - surv_min'));
end

X_0i = X_0;

first_iter = "Yes"
for j = 1 : no_LV
    if first_iter == "Yes" then
        U_0 = Y_0(:,1);                                              //Initialise U by choosing first column of Y_pls
    else
        U_0 = U_1;
        X_0 = X_1;
        Y_0 = Y_1;
        end
converge = "No";                                                     //This section repeats until U converges
tolerance = 0.00000001;
while (converge == "No")
    Err = 0;
    W_1x = X_0' * U_0;

    numrows = size(W_1x,1);
    max_data = max(W_1x,"r");
    min_data = min(W_1x,"r");
    for i = 1:numrows
        W_1(i,:) = (W_1x(i,:) - min_data)/(max_data - min_data);
    end

    T_1 = X_0 * W_1
    Q_1x = Y_0' * T_1;

    numrows = size(Q_1x,1);
    max_data = max(Q_1x,"r");
    min_data = min(Q_1x,"r");
    for i = 1:numrows
        Q_1(i,:) = (Q_1x(i,:) - min_data)/(max_data - min_data);
    end

    U_1 = Y_0 * Q_1;

    for i = 1 : size(U_1,1)
```

```
            Err = Err + (U_1(i) - U_0(i))^2;
        end
disp(Err)
        if (Err < tolerance) then                                   //Stops loop when Error is smaller than tolerance
            converge = "Yes"
            first_iter = "No"
        else Err_0 = Err
            U_0 = U_1
end
end
b_1 = (T_1' * T_1) \ (U_1' * T_1);
P_1 = (T_1' * T_1) \ (X_0' * T_1);
X_1 = X_0 - (T_1 * P_1');
Y_1 = Y_0 - (b_1 * T_1 * Q_1');
b_coeff(j,1) = b_1;
W_pls(:,j) = W_1;
P_pls(:,j) = P_1;
T_pls(:,j) = T_1;
Q_pls(:,j) = Q_1;
U_pls(:,j) = U_1;
low_index = j*size(Y_pls,2)-(size(Y_pls,2)-1);
up_index = j*size(Y_pls,2);
LV_Y(:,low_index : up_index) = b_1* T_1 * Q_1';                     //Calculates Y_pls-prediction for each LV
end

R_pls = (X_0i' * X_0i) \ (X_0i' * T_pls);
b_pls = R_pls * diag(b_coeff) * Q_pls';
Y0_pred = X_0i * b_pls;

Pred_Err = 0;
SS_tot = 0;
mean_data_Y = mean(Y_0i,"r");

for num_col = 1 : size(Y0_pred,2)
        for num_row = 1 : size(Y0_pred,1)
            Pred_Err = Pred_Err + (Y_0i(num_row,num_col) - Y0_pred(num_row,num_col))^2;          //Residual sum of squares
            SS_tot = SS_tot + (Y_0i(num_row,num_col) - mean_data_Y(num_col))^2;                  //Total sum of sqaures
            mean_Y(num_row,:) = mean_data_Y;
        end
end

R_sqr = 1 - (Pred_Err/SS_tot);                                                                  //Coefficient of determination
disp(R_sqr);

Y_pred = mean_Y + LV_Y(:,1:size(Y_pls,2));

for k = 2 : no_LV
        Y_pred = Y_pred + LV_Y(:,(k * size(Y_pls,2)-(size(Y_pls,2)-1)): k * size(Y_pls,2));
end

max_data = max(R_pls,"r");
min_data = min(R_pls,"r");
for num_col = 1 : size(R_pls,2)
        for num_row = 1:size(R_pls,1)
            R_pls_norm(num_row,num_col) = (R_pls(num_row,num_col) - min_data(num_col))/(max_data(num_col) -
min_data(num_col));
        end
end

LV_Y_perc = zeros(size(LV_Y,1),size(LV_Y,2));
sum_elem = zeros(size(Y_pls,1),size(Y_pls,2));

for j = 1 : size(LV_Y,1)
        for k = 1 : size(Y_pls,2)
            for l = 1 : no_LV
                sum_elem(j,k) = sum_elem(j,k) + abs(LV_Y(j,l * size(Y_pls,2) - (size(Y_pls,2) - 1) + k - 1));
            end
        end
end

for i = 1 : no_LV
        for j = 1 : size(LV_Y,1)
            count = 1;
```

```
            for k = (i * size(Y_pls,2) - (size(Y_pls,2) - 1)) : i * size(Y_pls,2)
                LV_Y_perc(j,k) = abs(LV_Y(j,k)) / sum_elem(j,count);
                count = count + 1;
            end
        end
    end
```

```
//=========================================================================================================
//ADM1 MODEL
//IWA TASK GROUP, SCIENTIFIC AND TECHNICAL REPORT NO.13
//File: Optimisation.sce
//=========================================================================================================

global counter
global mf_X_data
global U_diff_data
global lamda_data

function U_diff=Mod_Opt(lamda)

lamda = lamda';

global ('counter')
global ('mf_X_data')
global ('U_diff_data')
global ('lamda_data')

exec('Parameters.sce',-1)
mf_X_0i =
[f.si_xc,f.xi_xc,f.ch_xc,f.pr_xc,f.li_xc,f.fa_li,f.h2_su,f.bu_su,f.pro_su,f.ac_su,f.h2_aa,f.va_aa,f.bu_aa,f.pro_aa,f.ac_aa,kdis,khyd.c
h,khyd.pr,khyd.li,KS.IN,I_pH_ul,I_pH_ll,km.su,KS.su,Y.su,kdec.xsu,km.aa,KS.aa,Y.aa,kdec.xaa,km.fa,KS.fa,Y.fa,kdec.xfa,KI.h2
_fa,km.c4,KS.c4,Y.c4,kdec.xc4,KI.h2_c4,km.pro,KS.pro,Y.pro,kdec.xpro,KI.h2_pro,km.ac,KS.ac,Y.ac,kdec.xac,KI.nh3_ac,I_pH_
ac_ul,I_pH_ac_ll,km.h2,KS.h2,Y.h2,kdec.xh2,I_pH_h2_ul,I_pH_h2_ll];

mf_X = 0;

mf_X_0 = 2*((mf_X_0i    - surv_min') ./ (surv_max' - surv_min')) - 1;

mf_t = mf_X_0 * R_pls;
mf_X_e = mf_t * P_pls';
mf_X_res = mf_X_0 - mf_X_e;

mf_X_delta = (lamda .* mf_t) * P_pls';
mf_X_c = mf_X_res + mf_X_delta;
mf_X = (mf_X_c + 1)/2 .* (surv_max' - surv_min') + surv_min';

neg = 0;
for k = 1 : size(mf_X,2)
      if mf_X(k) < 0 then
            neg = neg + 1;
            mf_X(k) = surv_min(k);
            end
end

if mf_X(22) > mf_X(21) then
      mf_X(22) = mf_X(21) - 1;
end

disp(strcat(["No. of neg. parameters = ",string(neg)]))

      f_xc = mf_X(1) + mf_X(2) + mf_X(3) + mf_X(4) + mf_X(5);
      f_su = mf_X(7) + mf_X(8) + mf_X(9) + mf_X(10);
      f_aa = mf_X(11) + mf_X(12) + mf_X(13) + mf_X(14) + mf_X(15);

      mf_X(1) = mf_X(1)/f_xc;
      mf_X(2) = mf_X(2)/f_xc;
      mf_X(3) = mf_X(3)/f_xc;
      mf_X(4) = mf_X(4)/f_xc;
      mf_X(5) = mf_X(5)/f_xc;

      mf_X(7) = mf_X(7)/f_su;
      mf_X(8) = mf_X(8)/f_su;
      mf_X(9) = mf_X(9)/f_su;
      mf_X(10) = mf_X(10)/f_su;

      mf_X(11) = mf_X(11)/f_aa;
      mf_X(12) = mf_X(12)/f_aa;
      mf_X(13) = mf_X(13)/f_aa;
      mf_X(14) = mf_X(14)/f_aa;
      mf_X(15) = mf_X(15)/f_aa;
```

```
f.si_xc = mf_X(1); f.xi_xc = mf_X(2); f.ch_xc = mf_X(3); f.pr_xc = mf_X(4); f.li_xc = mf_X(5); f.fa_li = mf_X(6); f.h2_su = mf_X(7);
f.bu_su = mf_X(8); f.pro_su = mf_X(9); f.ac_su = mf_X(10);
f.h2_aa = mf_X(11); f.va_aa = mf_X(12); f.bu_aa = mf_X(13); f.pro_aa = mf_X(14); f.ac_aa = mf_X(15); kdis = mf_X(16);
khyd.ch = mf_X(17); khyd.pr = mf_X(18); khyd.li = mf_X(19); KS.IN = mf_X(20);
l_pH_ul = mf_X(21); l_pH_ll = mf_X(22); km.su = mf_X(23); KS.su = mf_X(24); Y.su = mf_X(25); kdec.xsu = mf_X(26); km.aa =
mf_X(27); KS.aa = mf_X(28); Y.aa = mf_X(29); kdec.xaa = mf_X(30);
km.fa = mf_X(31); KS.fa = mf_X(32); Y.fa = mf_X(33); kdec.xfa = mf_X(34); KI.h2_fa = mf_X(35); km.c4 = mf_X(36); KS.c4 =
mf_X(37); Y.c4 = mf_X(38); kdec.xc4 = mf_X(39); KI.h2_c4 = mf_X(40);
km.pro = mf_X(41); KS.pro = mf_X(42); Y.pro = mf_X(43); kdec.xpro = mf_X(44); KI.h2_pro = mf_X(45); km.ac = mf_X(46);
KS.ac = mf_X(47); Y.ac = mf_X(48); kdec.xac = mf_X(49); KI.nh3_ac = mf_X(50);
l_pH_ac_ul = mf_X(51); l_pH_ac_ll = mf_X(52); km.h2 = mf_X(53); KS.h2 = mf_X(54); Y.h2 = mf_X(55); kdec.xh2 = mf_X(56);
l_pH_h2_ul = mf_X(57); l_pH_h2_ll = mf_X(58);

MC_data = zeros(1,33);
disp(strcat(["Running Mod Fit ",string(lamda)]))
S_h_ion = 0.00000001;                                                    //initial S_h_ion guess
exec('ADM1.sce',-1);                                                     //Run ADM1 model
disp("OK")
sim_data = MC_data(2:size(MC_data,1),:);
sim_data_sel0 =
[(sim_data(:,4)+sim_data(:,5)+sim_data(:,6)+sim_data(:,7)),sim_data(:,11),(sim_data(:,13)+sim_data(:,14)+sim_data(:,15)+sim_
data(:,16)+sim_data(:,17)+sim_data(:,18)+sim_data(:,19)+sim_data(:,20)+sim_data(:,21)+sim_data(:,22)+sim_data(:,23)),sim_d
ata(:,30),sim_data(:,32),sim_data(:,33)];                                //arrange matrix to [VFA;S_IN;VSS;pH;q_ch4;q_co2]
m_data_sel0 = [VFA_m,S_IN_m,VSS_m,pH_m,q_ch4_m,q_co2_m];
m_data_sel0 = m_data_sel0(2:size(m_data_sel0,1)-1,:);

                                                                         //Convert to Latent Var.

sim_data_sel = zeros(1,output_no * sim_dur);
m_data_sel = zeros(1,output_no * sim_dur);
col_count = 1;

for j = 1 : sim_dur
    sim_data_sel(1,col_count:j*output_no) = sim_data_sel0(j,:);
    m_data_sel(1,col_count:j*output_no) = m_data_sel0(j,:);
    col_count = col_count + output_no;
end

Y_sim = 2*((sim_data_sel - Y_min) ./ (Y_max - Y_min)) - 1;
Y_m = 2*((m_data_sel - Y_min) ./ (Y_max - Y_min)) - 1;

U_sim_i = zeros(sim_dur,no_LV);
U_m_i = zeros(sim_dur,no_LV);
U_diff_i = zeros(sim_dur,no_LV);
for k = 1 : no_LV
    for p = 1 : size(Y_sim,2)
        U_sim_i(p,k) = Y_sim(1,p) * Q_pls(p,k);
        U_m_i(p,k) = Y_m(1,p) * Q_pls(p,k);
        U_diff_i(p,k) = (U_sim_i(p,k) - U_m_i(p,k))^2;
    end
end

U_diff = sum(U_diff_i);

output_s = sim_data_sel0;
output_m = m_data_sel0;

opt_lsqr0 = zeros(size(output_s,1),size(output_s,2));

for j = 1 : size(output_s,2)
    for i = 1 : size(output_s,1)
        opt_lsqr0(i,j) = (output_s(i,j) - output_m(i,j))^2;
    end
end

opt_lsqr = sqrt(sum(opt_lsqr0,'r')/sim_dur);

disp(strcat(["U_diff = ",string(U_diff)]))

mf_X_data($+1,:) = mf_X;
U_diff_data($+1,:) = U_diff;
lamda_data($+1,:) = lamda;
```

```
exec('Plot_Mod_Fit.sce',-1);

endfunction

exec('Parameters.sce',-1)
mf_X_0i =
[f.si_xc,f.xi_xc,f.ch_xc,f.pr_xc,f.li_xc,f.fa_li,f.h2_su,f.bu_su,f.pro_su,f.ac_su,f.h2_aa,f.va_aa,f.bu_aa,f.pro_aa,f.ac_aa,kdis,khyd.c
h,khyd.pr,khyd.li,KS.IN,I_pH_ul,I_pH_ll,km.su,KS.su,Y.su,kdec.xsu,km.aa,KS.aa,Y.aa,kdec.xaa,km.fa,KS.fa,Y.fa,kdec.xfa,KI.h2
_fa,km.c4,KS.c4,Y.c4,kdec.xc4,KI.h2_c4,km.pro,KS.pro,Y.pro,kdec.xpro,KI.h2_pro,km.ac,KS.ac,Y.ac,kdec.xac,KI.nh3_ac,I_pH_
ac_ul,I_pH_ac_ll,km.h2,KS.h2,Y.h2,kdec.xh2,I_pH_h2_ul,I_pH_h2_ll];

mf_X_data = zeros(1,size(mf_X_0i,2));
U_diff_data = zeros(1,no_LV);
lamda_data = zeros(1,no_LV);

[fopt,xopt] = leastsq(list(Mod_Opt),lamda0,"ar",100,25,0.1,0.01,lamda0);

fprintfMat('Mod_Fit_X.xls',mf_X_data);
fprintfMat('Mod_Fit_U_diff.xls',U_diff_data);
fprintfMat('Mod_Fit_lamda.xls',lamda_data);
```

```
//=========================================================================================================
//ADM1 MODEL
//IWA TASK GROUP, SCIENTIFIC AND TECHNICAL REPORT NO.13
//File: Cross_Validate.sce
//=========================================================================================================

f.si_xc = CV_Par(1); f.xi_xc = CV_Par(2); f.ch_xc = CV_Par(3); f.pr_xc = CV_Par(4); f.li_xc = CV_Par(5); f.fa_li = CV_Par(6);
f.h2_su = CV_Par(7); f.bu_su = CV_Par(8); f.pro_su = CV_Par(9); f.ac_su = CV_Par(10);
f.h2_aa = CV_Par(11); f.va_aa = CV_Par(12); f.bu_aa = CV_Par(13); f.pro_aa = CV_Par(14); f.ac_aa = CV_Par(15); kdis =
CV_Par(16); khyd.ch = CV_Par(17); khyd.pr = CV_Par(18); khyd.li = CV_Par(19); KS.IN = CV_Par(20);
I_pH_ul = CV_Par(21); I_pH_ll = CV_Par(22); km.su = CV_Par(23); KS.su = CV_Par(24); Y.su = CV_Par(25); kdec.xsu =
CV_Par(26); km.aa = CV_Par(27); KS.aa = CV_Par(28); Y.aa = CV_Par(29); kdec.xaa = CV_Par(30);
km.fa = CV_Par(31); KS.fa = CV_Par(32); Y.fa = CV_Par(33); kdec.xfa = CV_Par(34); KI.h2_fa = CV_Par(35); km.c4 =
CV_Par(36); KS.c4 = CV_Par(37); Y.c4 = CV_Par(38); kdec.xc4 = CV_Par(39); KI.h2_c4 = CV_Par(40);
km.pro = CV_Par(41); KS.pro = CV_Par(42); Y.pro = CV_Par(43); kdec.xpro = CV_Par(44); KI.h2_pro = CV_Par(45); km.ac =
CV_Par(46); KS.ac = CV_Par(47); Y.ac = CV_Par(48); kdec.xac = CV_Par(49); KI.nh3_ac = CV_Par(50);
I_pH_ac_ul = CV_Par(51); I_pH_ac_ll = CV_Par(52); km.h2 = CV_Par(53); KS.h2 = CV_Par(54); Y.h2 = CV_Par(55); kdec.xh2
= CV_Par(56); I_pH_h2_ul = CV_Par(57); I_pH_h2_ll = CV_Par(58);

sim_dur0 = sim_dur;
sim_dur = sim_dur_CV;
t_data = t_data_cv;
inf_data = inf_data_cv;
Q_in = Q_in_cv;
Q_was = Q_was_cv;
T_data = T_data_cv;
q_ch4_m = q_ch4_m_cv;
q_co2_m = q_co2_m_cv;
pH_m = pH_m_cv;
S_IN_m = S_IN_m_cv;
VFA_m = VFA_m_cv;
VSS_m = VSS_m_cv;
C_was = C_was_cv;

sim_mode = "cross validate";
MC_data = zeros(1,33);
VFA_data = zeros(1,1);
VSS_data = zeros(1,1);

exec('ADM1.sce',-1);                                                    //Run ADM1 model

output_s = [VFA_s DSV(11,:)' VSS_s pH(2:length(pH)) q_ch4(2:length(q_ch4)) q_co2(2:length(q_co2))];
output_m = [VFA_m S_IN_m VSS_m pH_m q_ch4_m q_co2_m];

CV_lsqr0 = zeros(size(output_s,1),size(output_s,2));

for j = 1 : size(output_s,2)
    for i = 1 : size(output_s,1)
        CV_lsqr0(i,j) = (output_s(i,j) - output_m(i,j))^2;
    end
end

CV_lsqr = sqrt(sum(CV_lsqr0,'r')/sim_dur_CV);
//CV_lsqr = sum(CV_lsqr0,'r');

sim_dur = sim_dur0;                                                     //days to simulate

exec('Plot_Sens.sce',-1);                                               //plot results in red lines
```

```
//=========================================================================================================
//ADM1 MODEL
//IWA TASK GROUP, SCIENTIFIC AND TECHNICAL REPORT NO.13
//File: Plot_Sens.sce
//=========================================================================================================

//Graphs Plotting
//================

if sim_mode == "single" then

scf(1);
//clf(1);
fig=get("current_figure")
fig.figure_position
fig.figure_size=[1936,1056]
subplot(231)
plot(t,DSV(1,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_Su')
subplot(232)
plot(t,DSV(2,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_aa')
subplot(233)
plot(t,DSV(3,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_fa')
subplot(234)
plot(t,DSV(4,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_va')
subplot(235)
plot(t,DSV(5,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_bu')
subplot(236)
plot(t,DSV(6,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_pro')

scf(2);
//clf(2);
fig=get("current_figure")
fig.figure_position
fig.figure_size=[1936,1056]
subplot(231)
plot(t,DSV(7,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_ac')
subplot(232)
plot(t,DSV(8,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_h2')
subplot(233)
plot(t,DSV(9,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_ch4')
subplot(234)
plot(t,DSV(10,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kmoleC.m-3)"]);
title('S_IC')
subplot(235)
```

```
plot(t,DSV(11,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kmoleN.m-3)"]);
title('S_IN')
subplot(236)
plot(t,DSV(12,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_I')

scf(3);
//clf(3);
fig=get("current_figure")
fig.figure_position
fig.figure_size=[1936,1056]
subplot(231)
plot(t,DSV(13,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_c')
subplot(232)
plot(t,DSV(14,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_ch')
subplot(233)
plot(t,DSV(15,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_pr')
subplot(234)
plot(t,DSV(16,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_li')
subplot(235)
plot(t,DSV(17,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_su')
subplot(236)
plot(t,DSV(18,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_aa')

scf(4);
//clf(4);
fig=get("current_figure")
fig.figure_position
fig.figure_size=[1936,1056]
subplot(231)
plot(t,DSV(19,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_fa')
subplot(232)
plot(t,DSV(20,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_c4')
subplot(233)
plot(t,DSV(21,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_pro')
subplot(234)
plot(t,DSV(22,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_ac')
subplot(235)
plot(t,DSV(23,:),'r-')
```

```
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_h2')
subplot(236)
plot(t,DSV(24,:),'r-')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_I')

scf(5);
//clf(5);
fig=get("current_figure")
fig.figure_position
fig.figure_size=[1936,1056]
subplot(221)
plot(t,pH(2:length(pH))','r-')
xlabel("Days");
ylabel(["pH"]);
title('pH')
subplot(222)
plot(t,q_h2(2:length(q_h2))','r-')
xlabel("Days");
ylabel(["Flowrate";"(m3.day-1)"]);
title('q_h2')
subplot(223)
plot(t,q_ch4(2:length(q_ch4))','r-')
xlabel("Days");
ylabel(["Flowrate";"(m3.day-1)"]);
title('q_ch4')
subplot(224)
plot(t,q_co2(2:length(q_co2))','r-')
xlabel("Days");
ylabel(["Flowrate";"(m3.day-1)"]);
title('q_co2')

scf(6);
//clf(6);
fig=get("current_figure")
fig.figure_position
fig.figure_size=[1936,1056]
subplot(231)
plot(t,VFA_s(1:length(VFA_s))','r-')
plot(t,VFA_m(2:length(VFA_m))','g--')
xlabel("Days");
ylabel(["VFA";"(kgCOD.m-3)"]);
title('VFA')
subplot(232)
plot(t,DSV(11,:),'r-')
plot(t,S_IN_m(2:length(S_IN_m))','g--')
xlabel("Days");
ylabel(["Concentration";"(kgmoleN.m-3)"]);
title('S_IN')
subplot(233)
plot(t,VSS_s(1:length(VSS_s))','r-')
plot(t,VSS_m(2:length(VSS_m))','g--')
xlabel("Days");
ylabel(["VSS";"(kgCOD.m-3)"]);
title('VSS')
subplot(234)
plot(t,pH(2:length(pH))','r-')
plot(t,pH_m(2:length(pH_m))','g--')
xlabel("Days");
ylabel(["pH"]);
title('pH')
subplot(235)
plot(t,q_ch4(2:length(q_ch4))','r-')
plot(t,q_ch4_m(2:length(q_ch4_m))','g--')
xlabel("Days");
ylabel(["Flowrate";"(m3.day-1)"]);
title('q_ch4')
subplot(236)
plot(t,q_co2(2:length(q_co2))','r-')
plot(t,q_co2_m(2:length(q_co2_m))','g--')
```

```
xlabel("Days");
ylabel(["Flowrate";"(m3.day-1)"]);
title('q_co2')

end

if sim_mode == "monte-carlo" then

scf(1);
//clf(1);
fig=get("current_figure")
fig.figure_position
fig.figure_size=[1936,1056]
subplot(231)
plot(t,DSV(1,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_Su')
subplot(232)
plot(t,DSV(2,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_aa')
subplot(233)
plot(t,DSV(3,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_fa')
subplot(234)
plot(t,DSV(4,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_va')
subplot(235)
plot(t,DSV(5,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_bu')
subplot(236)
plot(t,DSV(6,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_pro')

scf(2);
//clf(2);
fig=get("current_figure")
fig.figure_position
fig.figure_size=[1936,1056]
subplot(231)
plot(t,DSV(7,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_ac')
subplot(232)
plot(t,DSV(8,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_h2')
subplot(233)
plot(t,DSV(9,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_ch4')
subplot(234)
plot(t,DSV(10,:))
xlabel("Days");
ylabel(["Concentration";"(kmoleC.m-3)"]);
title('S_IC')
subplot(235)
plot(t,DSV(11,:))
xlabel("Days");
ylabel(["Concentration";"(kmoleN.m-3)"]);
```

```
title('S_IN')
subplot(236)
plot(t,DSV(12,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_I')

scf(3);
//clf(3);
fig=get("current_figure")
fig.figure_position
fig.figure_size=[1936,1056]
subplot(231)
plot(t,DSV(13,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_c')
subplot(232)
plot(t,DSV(14,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_ch')
subplot(233)
plot(t,DSV(15,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_pr')
subplot(234)
plot(t,DSV(16,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_li')
subplot(235)
plot(t,DSV(17,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_su')
subplot(236)
plot(t,DSV(18,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_aa')

scf(4);
//clf(4);
fig=get("current_figure")
fig.figure_position
fig.figure_size=[1936,1056]
subplot(231)
plot(t,DSV(19,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_fa')
subplot(232)
plot(t,DSV(20,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_c4')
subplot(233)
plot(t,DSV(21,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_pro')
subplot(234)
plot(t,DSV(22,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_ac')
subplot(235)
plot(t,DSV(23,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_h2')
```

```
subplot(236)
plot(t,DSV(24,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('X_I')

scf(5);
//clf(5);
fig=get("current_figure")
fig.figure_position
fig.figure_size=[1936,1056]
subplot(221)
plot(t,DSV(30,:))
xlabel("Days");
ylabel(["pH"]);
title('pH')
subplot(222)
plot(t,DSV(31,:))
xlabel("Days");
ylabel(["Flowrate";"(m3.day-1)"]);
title('q_h2')
subplot(223)
plot(t,DSV(32,:))
xlabel("Days");
ylabel(["Flowrate";"(m3.day-1)"]);
title('q_ch4')
subplot(224)
plot(t,DSV(33,:))
xlabel("Days");
ylabel(["Flowrate";"(m3.day-1)"]);
title('q_co2')

scf(6);
//clf(6);
fig=get("current_figure")
fig.figure_position
fig.figure_size=[1936,1056]
subplot(231)
plot(t,VFA_s(2:length(VFA_s))')
xlabel("Days");
ylabel(["VFA";"(kgCOD.m-3)"]);
title('VFA')
subplot(232)
plot(t,DSV(11,:))
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_IN')
subplot(233)
plot(t,VSS_s(2:length(VSS_s))')
xlabel("Days");
ylabel(["VSS";"(kgCOD.m-3)"]);
title('VSS')
subplot(234)
plot(t,DSV(30,:))
xlabel("Days");
ylabel(["pH"]);
title('pH')
subplot(235)
plot(t,DSV(32,:))
xlabel("Days");
ylabel(["Flowrate";"(m3.day-1)"]);
title('q_ch4')
subplot(236)
plot(t,DSV(33,:))
xlabel("Days");
ylabel(["Flowrate";"(m3.day-1)"]);
title('q_co2')

end

if sim_mode == "cross validate" then

scf(6);
```

```
//clf(6);
fig=get("current_figure")
fig.figure_position
fig.figure_size=[1936,1056]

subplot(231)
plot(t(1:sim_dur),VFA_s(1:sim_dur)','b-')
plot(t(sim_dur+1:sim_dur_CV+1),VFA_s(sim_dur+1:length(VFA_s))','r-')
plot(t,VFA_m(2:sim_dur_CV+2)','g--')
xlabel("Days");
ylabel(["VFA";"(kgCOD.m-3)"]);
title(["VFA";strcat(["RMSE = ",string(CV_lsqr(1))])]);
subplot(232)
plot(t(1:sim_dur),DSV(11,1:sim_dur),'b-')
plot(t(sim_dur+1:sim_dur_CV+1),DSV(11,sim_dur+1:size(DSV,2)),'r-')
plot(t,S_IN_m(2:sim_dur_CV+2)','g--')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title(['S_IN';strcat(["RMSE = ",string(CV_lsqr(2))])])
subplot(233)
plot(t(1:sim_dur),VSS_s(1:sim_dur)','b-')
plot(t(sim_dur+1:sim_dur_CV+1),VSS_s(sim_dur+1:length(VSS_s))','r-')
plot(t,VSS_m(2:sim_dur_CV+2)','g--')
xlabel("Days");
ylabel(["VSS";"(kmoleN.m-3)"]);
title(['VSS';strcat(["RMSE = ",string(CV_lsqr(3))])])
subplot(234)
plot(t(1:sim_dur),pH(2:sim_dur+1)','b-')
plot(t(sim_dur+1:sim_dur_CV+1),pH(sim_dur+2:length(pH))','r-')
plot(t,pH_m(2:sim_dur_CV+2)','g--')
xlabel("Days");
ylabel(["pH"]);
title(['pH';strcat(["RMSE = ",string(CV_lsqr(4))])])
subplot(235)
plot(t(1:sim_dur),q_ch4(2:sim_dur+1)','b-')
plot(t(sim_dur+1:sim_dur_CV+1),q_ch4(sim_dur+2:length(q_ch4))','r-')
plot(t,q_ch4_m(2:sim_dur_CV+2)','g--')
xlabel("Days");
ylabel(["Flowrate";"(m3.day-1)"]);
title(['q_ch4';strcat(["RMSE = ",string(CV_lsqr(5))])])
subplot(236)
plot(t(1:sim_dur),q_co2(2:sim_dur+1)','b-')
plot(t(sim_dur+1:sim_dur_CV+1),q_co2(sim_dur+2:length(q_co2))','r-')
plot(t,q_co2_m(2:sim_dur_CV+2)','g--')
xlabel("Days");
ylabel(["Flowrate";"(m3.day-1)"]);
title(['q_co2';strcat(["RMSE = ",string(CV_lsqr(6))])])

end

if sim_mode == "pls" then

scf(6);
//clf(6);
fig=get("current_figure")
fig.figure_position
fig.figure_size=[1936,1056]
subplot(231)
plot(t,DSV_temp(:,1)')
xlabel("Days");
ylabel(["VFA";"(kgCOD.m-3)"]);
title('VFA')
subplot(232)
plot(t,DSV_temp(:,2)')
xlabel("Days");
ylabel(["Concentration";"(kgCOD.m-3)"]);
title('S_IN')
subplot(233)
plot(t,DSV_temp(:,3)')
xlabel("Days");
ylabel(["VSS";"(kgCOD.m-3)"]);
title('VSS')
subplot(234)
```

```
plot(t,DSV_temp(:,4)')
xlabel("Days");
ylabel(["pH"]);
title('pH')
subplot(235)
plot(t,DSV_temp(:,5)')
xlabel("Days");
ylabel(["Flowrate";"(m3.day-1)"]);
title('q_ch4')
subplot(236)
plot(t,DSV_temp(:,6)')
xlabel("Days");
ylabel(["Flowrate";"(m3.day-1)"]);
title('q_co2')

end
```