Fast characterization of the organic matter, instrumentation and modelling for the AD process performances prediction

Cyrille Charnier

To cite this version:

Cyrille Charnier. Fast characterization of the organic matter, instrumentation and modelling for the AD process performances prediction. Agricultural sciences. Montpellier SupAgro, 2016. English. NNT : 2016NSAM0016 . tel-02801514v2

HAL Id: tel-02801514
https://hal.inrae.fr/tel-02801514v2

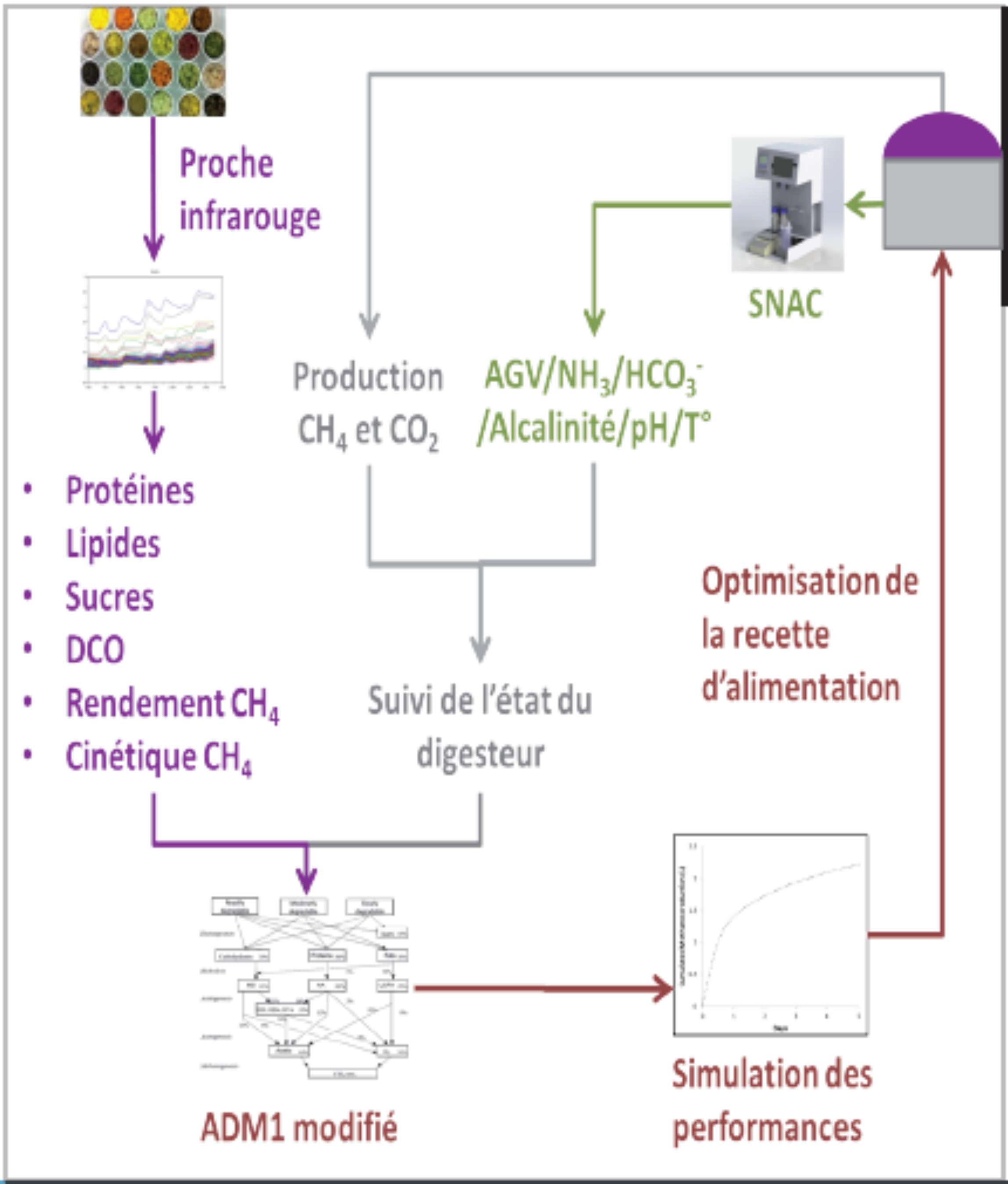
Submitted on 5 Jun 2020

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

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

Cyrille CHARNIER

PhD thesis 2016


Proche
infrarouge
Protéines
Lipides
Sucres
DCO
Rendement CH_4
Cinétique CH_4
Production
CH_4 et CO_2
SNAC
$AGV/NH_3/HCO_3^-$
/Alcalinité/pH/T°
Suivi de l'état du
digesteur
Optimisation de
la recette
d'alimentation
ADM1 modifié
Simulation des
performances

CARACTÉRISATION RAPIDE DE LA MATIÈRE ORGANIQUE, INSTRUMENTATION ET MODÉLISATION POUR LA PRÉDICTION DES PERFORMANCES DES PROCÉDÉS DE DIGESTION ANAÉROBIE

FAST CHARACTERIZATION OF THE ORGANIC MATTER, INSTRUMENTATION AND MODELLING FOR THE ANAEROBIC DIGESTION PROCESS PERFORMANCES PREDICTION

INRA
SCIENCE & IMPACT

Membre fondateur de
agreenium

THÈSE

Pour obtenir le grade de
Docteur

Délivré par **Montpellier Supagro**
Préparée au sein de l'école doctorale **GAIA**
et de l'unité de recherche **INRA UR 050**
Spécialité : **Agroressource, Procédés, Aliments, Bioproduits (APAB)**
Présentée par **Cyrille Charnier**

Fast Characterization of the Organic Matter, Instrumentation and Modelling for the AD Process Performances Prediction

Caractérisation rapide de la matière organique, instrumentation et modélisation pour la prédiction des performances des procédés de digestion anaérobie

Thèse soutenue le 25 novembre 2016 devant le jury composé de

Mr Fabrice Béline, Directeur de recherche, UR OPAALE, IRSTEA.	Rapporteur
Mme. Véronique Bellon-Maurel, Directrice de recherche, UMR ITAP, IRSTEA	Examinatrice
Mme. Patricia Camacho, Project manager biogas consulting, SUEZ environnement	Examinatrice
Mr. Marc Heran, Professeur, IEM-GPM, UM2	Examinateur
Mr Eric Latrille, Ingénieur de recherche, UR LBE, INRA.	Encadrant
Mr Jérémie Miroux, CEO, BioEnTech	Encadrant
Mme. Marie-Noëlle Pons, Directrice de recherche, LRGP, CNRS.	Rapportrice
Mr Jean-Philippe Steyer, Directeur de recherche, UR LBE, INRA.	Directeur de thèse

Abstract

Anaerobic digestion is an important pillar of the European circular economy, producing methane and organic fertilizers from waste. The development of the anaerobic digestion sector goes through co-digestion and feeding strategy optimization. Its development requires the biological state estimation of the plant, substrate characterization and predictive models simulating the plant performances, for which current solutions are not suitable. In this thesis, a titration sensor coupling pH and electrical conductivity for the estimation of volatile fatty acids, inorganic carbon and ammonia has been designed, improving the accuracy on volatile fatty acids estimation by 14.5 compared to current sensors. Along with biogas analyses, it allows estimating the biological state of the unit. Besides, fast near infrared spectroscopic analysis, estimating in a matter of minute carbohydrate, protein and lipid contents, chemical oxygen demand, methane production yield and kinetics, has been developed. Fast substrate characterization is then used to implement a modified Anaerobic Digestion Model n°1 which predicts the performances of a plant under optimal condition. Coupling biological state estimation to this approach enables to correct the prediction with the current state of the plant. This approach provides a powerful predictive tool for advanced control of anaerobic digestion plants and feeding strategy optimization.

Keywords: Anaerobic digestion, Titration, Near-infrared, PLS regression, Modelling, Characterization.

Résumé

La digestion anaérobie est un des piliers de l'économie circulaire européenne, produisant du méthane et des engrais organiques à partir de déchets. Le développement de ce secteur passe par la co-digestion et l’optimisation de l'alimentation des procédés. Cela nécessite l'estimation de l'état biologique du digesteur, la caractérisation du substrat ainsi que l’utilisation de modèles prédictifs simulant les performances du digesteur, pour lesquels les solutions actuelles ne sont pas adaptées. Dans cette thèse, un capteur titrimétrique couplant pH et conductivité électrique pour l'estimation des concentrations en acides gras volatils, carbone inorganique et azote ammoniacale a été conçu, améliorant la précision d'estimation des acides gras volatils de 14,5 par rapport aux capteurs actuels. Couplé à l’analyse du biogaz, il permet d'estimer l'état biologique du procédé. En parallèle, une analyse spectroscopique proche-infrarouge, estimant les teneurs en glucides, protéines, lipides, demande chimique en oxygène, rendement et cinétique de production de méthane a été développée réduisant le temps de caractérisation des substrats à quelques minutes. Cette caractérisation rapide des substrats est ensuite utilisée pour implémenter le modèle de digestion anaérobie ADM1 de l’IWA et prédire les performances d'un digesteur dans des conditions optimales. Le couplage de l’estimation de l'état biologique à cette approche permet de corriger la prédiction en prenant en compte l'état actuel du digesteur. Cette approche fournit un outil prédictif puissant pour le contrôle avancé des unités de digestion anaérobie ainsi que l'optimisation de la recette d'alimentation.

Mots-clés: Digestion anaérobie, Titration, Proche-infrarouge, Régression PLS, Modélisation, Caractérisation.

Résumé étendu

Le développement de la digestion anaérobie passe aujourd’hui par la diversification des substrats utilisés et leur utilisation en co-digestion. La quantité de déchets organiques disponibles donne de belles perspectives au secteur de la digestion anaérobie, mais sa nouveauté et l'absence de solutions industrielles adaptées à la diversité des déchets organiques pénalisent la rentabilité et le développement de ce secteur. La conception des unités de co-digestion doit être rationnalisée pour améliorer la rentabilité des unités en termes d'investissements et de frais de fonctionnement. Cela permettra de rassurer les acteurs et investisseurs de la filière, facilitant l’émergence de nouveaux projets. Les digesteurs actuels souffrent en effet de la sous-optimisation de leur production de méthane, souvent liée à une mauvaise gestion de la recette d'alimentation, qui pénalise la rentabilité des unités. En parallèle, des problèmes de conception et d’importants soucis de supervision desservent le secteur de la digestion anaérobie. L’apport de solutions de supervision en termes de capteurs, méthodes et modèles d’optimisation de la recette d'alimentation doit se focaliser sur trois objectifs :

- estimer l’état biologique réel du réacteur
- définir la meilleure stratégie d'alimentation et ses performances associées dans des conditions de production optimales
- définir, en fonction de l'état biologique et de la recette d’alimentation, les conditions opératoires permettant la meilleure production.

L'état de l'art sur la supervision de la digestion anaérobie révèle que de nombreux capteurs, méthodes et modèles ont été mis au point ces dernières années afin d’optimiser les performances des digesteurs. En 2002, le développement du modèle de digestion anaérobie n°1 (ADM1) de l’IWA a été une avancée majeure pour l’optimisation du procédé de digestion anaérobie. ADM1 modélise en effet avec précision les performances de digestion anaérobie de l’unité, il prend en compte à la fois l’état biologique du digesteur et les caractéristiques des substrats. Cependant, il souffre d’un nombre élevé de paramètres dont les valeurs numériques sont difficiles à estimer. De plus, l’étape d’hydrolyse, développée originellement pour représenter la digestion anaérobie des boues de station d’épuration, est trop simplement modélisée pour représenter la complexité de l’hydrolyse des substrats solides. Ce type de modèle ne peut donc être utilisé qu’à condition de définir simplement les paramètres du modèle et de modifier l’étape d’hydrolyse.

La première étape est donc de définir des capteurs et méthodes permettant l’estimation de l’état biologique du digesteur ainsi que la caractérisation des substrats adaptés à la co-digestion et à la diversité des recettes d’alimentation qu’elle implique. Les analyses de la production de dioxyde de carbone et de méthane sont aujourd'hui matures et permettent d'apprécier la qualité et quantité de biogaz produit. Cependant, il est difficile de corréler la production de biogaz aux performances de l'unité sans connaitre ni les caractéristiques des substrats, ni les concentrations en intermédiaires réactionnels tels que les acides gras volatils ou l'azote ammoniacal. Or, les solutions disponibles pour estimer les intermédiaires réactionnels dans la phase liquide, restent peu adaptées à la co-digestion de substrats divers ainsi qu'aux changements fréquents de recette. Ce manque de capteurs disponibles pour estimer les intermédiaires réactionnels solubles restreint l'appréciation de l'état biologique du digesteur et constitue un premier frein à l'optimisation du procédé.

Un second frein réside dans la difficulté de caractérisation des substrats. Les analyses permettant d'estimer les performances de digestion anaérobie des déchets organiques sont longues et fastidieuses, ce qui limite leurs utilisations. Le manque de caractérisation induit des risques d'inhibition importants en cas de changement de recette encourageant l'opérateur à conserver une recette d'alimentation fixe. Ce manque de flexibilité dans la recette d'alimentation provoque des difficultés pour son optimisation. Ces difficultés d'optimisation de l'alimentation du digesteur s'aggraveront avec l'augmentation du stress sur les ressources/déchets due au nombre croissant d'installations de biogaz.

Compte tenu du contexte actuel de la digestion anaérobie, cette thèse se concentre sur :

- Une estimation précise des concentrations en intermédiaires réactionnels (à savoir les acides gras volatils (AGV), le carbone inorganique (IC) et l'azote ammoniacal (TAN)) dans le digesteur afin d'apprécier son état biologique
- La caractérisation des déchets organiques solides par des analyses rapides, reproductibles et précises permettant d'estimer la teneur en protéines, glucides et lipides, la demande chimique en oxygène (DCO), le rendement en méthane et la cinétique de production de méthane d'une large gamme de substrats solides tels que les fruits, les légumes, les fumiers, les céréales, les huiles, les graisses, la viande et les poissons.
- La modélisation dynamique du processus de digestion anaérobie avec un accent particulier sur la modélisation de l'étape d'hydrolyse et une implémentation facilitée du modèle ADM1.

La première partie de cette thèse, basée sur l'estimation des intermédiaires réactionnels en milieu liquide révèle que les capteurs actuels basés sur une méthode de titration avec évaluation du volume titrant en point clés fournissent, dans certains cas, une estimation erronée des AGV, IC et TAN. En cas de changements limités dans la recette d'alimentation, un biais constant entre valeur estimée et réelle est présent. Dans ces conditions, malgré une estimation biaisée, ces capteurs permettent de suivre la dynamique de production des AGV, IC et TAN. Cependant, un recalibrage des capteurs est nécessaire à tout changement de la recette d'alimentation. Ce type de capteurs n'est donc pas adapté au suivi des procédés fonctionnant en co-digestion, flexibles dans leur recette d'alimentation. Une solution plus prometteuse est l'estimation des AGV, IC et TAN sur la base de l'analyse non linéaire de la capacité du tampon sur l'ensemble du processus de titration.

Nous montrons dans cette thèse que les analyses non-linéaires de la capacité tampon sont plus précises que les analyses en points clés du volume titrant et sont surtout moins sensibles à la présence d'autres constituants dans l'échantillon. Elles sont donc plus appropriées pour le suivi de la co-digestion. Néanmoins, la capacité-tampon, définie à partir du pH et du volume de titrant, est rapidement limitée pour distinguer les nombreux composés solubles présents dans les digesteurs, ce qui peut conduire à des estimations erronées. Cette étude montre que les méthodes existantes estimant les concentrations en AGV, IC et TAN ne sont pas assez précises et robustes pour la supervision des unités. Pour pallier à ce problème, un nouveau capteur nommé SNAC (Système de titration pour l'azote ammoNiacal, les Acides gras volatils et le Carbone inorganique) a été développé dans cette thèse. Il utilise une méthode de résolution non linéaire combinant conductivité électrique et pH pour identifier plus précisément les différents composés et les concentrations en AGV, IC et TAN.

SNAC a été testé sur 24 échantillons de différents digesteurs dans une gamme de 0-0,16 mol.L^{-1} TAN, 0,01-0,21 mol.L^{-1} IC and 0-0,04 mol.L^{-1} AGV. L'écart type de la gamme d'échantillons testés est de 0,012 mol.L^{-1} TAN, 0,015 mol.L^{-1} IC and 0,003 mol.L^{-1} AGV. Le coefficient de détermination R² entre valeurs estimées par SNAC et valeurs de référence est de 0,95, 0,94 and 0,95 pour respectivement TAN, IC et AGV. Sur les mêmes échantillons, les méthodes de titration couramment utilisées et basées sur une analyse du volume titrant en points clés fournissent des erreurs types 14,5 et 1,2 fois plus élevées pour l'estimation des concentrations en AGV et IC respectivement. Ces résultats valident la précision de SNAC. Cette thèse montre que l'estimation précise des intermédiaires réactionnels tels que les AGV, IC et TAN est possible par titration en combinant la conductivité électrique et le pH avec des méthodes de résolution non-linéaires. Elle a abouti à la conception, la réalisation et la mise au point d'un prototype du capteur SNAC en cours d'utilisation sur site industriel. Combiné avec les débits de gaz et l'analyse de leur composition, SNAC fournit une solution adaptée pour l'estimation de l'état biologique des digesteurs. Une description de l'invention ainsi que des résultats obtenus sont publiés dans Water Research (Charnier et al., 2016). De plus, SNAC a fait l'objet d'une demande de brevet d'invention (n°15/59767).

Résumé étendu

La seconde partie de cette thèse traite de la caractérisation des déchets organiques solides pour la prédiction des performances des procédés. Les substrats utilisés en digestion anaérobie sont composés de matière organique complexe et hétérogène. Pour faciliter leur analyse, leur composition est décrite par des paramètres généraux tels que les teneurs en glucides, protéines et lipides, la demande chimique en oxygène (DCO), le rendement du méthane et la cinétique de production de méthane. Ces paramètres sont nécessaires pour la supervision des digesteurs mais leur analyse est longue et coûteuse. De plus, les techniques de laboratoire associées ne sont pas encore adaptées à des analyses de déchets solides. Ainsi, ces paramètres sont rarement évalués conjointement. Dans un premier temps, les méthodes actuelles d'analyse biochimique des déchets solides ont été testées. Notre étude montre en particulier des erreurs sur l'estimation de la DCO et du taux de glucide des déchets solides donnant des résultats peu répétables. Nous avons montré qu'une hydrolyse acide de la matière organique préalable à l'analyse de la DCO et du taux de sucre améliore la précision et la répétabilité des résultats. Les erreurs standards de laboratoire (SEL), estimées sur 153 échantillons analysés en duplicat, sont de 60 $mgO_2.gMS^{-1}$ et 63 $mgO_2.gMS^{-1}$ pour respectivement la DCO et le taux de glucides avec pré-hydrolyse acide.

Les relations entre composition biochimique et performances de digestion anaérobie ont été étudiées. Les corrélations existantes entre la DCO, le rendement en méthane, la biodégradabilité, la cinétique de la production de méthane et la teneur en glucides, protéines et lipides ont été évaluées sur 222 échantillons en utilisant une analyse en composantes principales et les coefficients de corrélation de Pearson. Ces travaux révèlent une corrélation positive entre le rendement méthane et la cinétique de production de méthane. Pour sa simplicité, l'analyse de DCO remplace souvent le rendement méthane, mais la corrélation négative révélée entre la DCO et la biodégradabilité indique que la DCO ne peut, en réalité, pas remplacer l'analyse de rendement en méthane. Les teneurs en glucides, protéines et lipides modulent différemment les performances de digestion anaérobie. Dans les conditions expérimentales étudiées, les glucides sont corrélés avec la matière organique facilement biodégradable tandis que les protéines sont légèrement anti-corrélées avec la biodégradabilité. Les lipides sont corrélés avec un rendement en méthane élevé, mais une production lente. Les corrélations obtenues révèlent que des relations entre la composition de la matière organique et les performances de digestion anaérobie existent.

Ces résultats conduisent à l'idée que la cinétique de production de méthane, fastidieuse à estimer, pourrait être prédite à partir de la composition biochimique. Cependant, une régression des moindres carrés partiels (PLSr) utilisant le rendement méthane, la DCO et les teneurs en glucides, lipides et protéines ne permet pas de prédire avec précision la cinétique de production de méthane.

Ceci démontre que ces indicateurs fournissent une description insuffisante de la matière organique. Néanmoins, ils révèlent une relation entre composition biochimique et performance de digestion anaérobie. Ces travaux ouvrent la voie vers la caractérisation de la matière organique avec des méthodes plus descriptives qui pourraient être utilisées pour la prédiction des performances du processus de digestion anaérobie.

La spectroscopie proche infrarouge (SPIR) et la spectroscopie par fluorescence 3D couplée à l'extraction chimique séquentielle ont été rapportées comme deux méthodes rapides fournissant une description suffisante de la matière organique pour prédire sa biodégradabilité via une PLSr. Ainsi, les avantages et inconvénients de SPIR et de la fluorescence 3D couplée à l'extraction chimique séquentielle pour la prédiction du rendement méthane et de la DCO ont été évalués sur 28 échantillons divers, divisés en 24 échantillons de calibration et 4 échantillons de validation. Les échantillons balayent une gamme de 0,141-0,717 $LCH_4.gMS^{-1}$ et 935-2334 $mgO_2.gMS^{-1}$ pour le rendement méthane et la DCO respectivement. En utilisant la spectroscopie par fluorescence 3D couplée à l'extraction chimique séquentielle, les erreurs standards de validation croisée et de prédiction sont respectivement de 0,314-0,155 $LCH_4.gMS^{-1}$ pour le rendement méthane et 324-459 $mgO_2.gMS^{-1}$ pour la DCO. Avec la SPIR, les erreurs standards de validation croisée et de prédiction sont respectivement de 0,057-0,092 $LCH_4.gMS^{-1}$ pour le rendement méthane et 120-193 $mgO_2.gMS^{-1}$ pour la DCO. Les résultats montrent que la fluorescence 3D couplée à l'extraction chimique séquentielle a l'avantage de prédire la bio-accessibilité de la matière organique, mais que la SPIR est plus rapide et précise pour la prédiction de la DCO et du rendement méthane. Ainsi, pour la suite de l'étude, la SPIR a été préférée pour caractériser les substrats.

Nous avons construit des modèles PLSr pour prédire la composition biochimique, le rendement méthane et le temps de production de méthane, à partir duquel est calculée la cinétique de production de méthane. Les modèles ont été calibrés avec près de 300 échantillons et validés avec un ensemble indépendant de 43 échantillons dans un large éventail de déchets solides utilisés par les unités de digestion anaérobie. Les erreurs standards de validation croisée sont de 90 $mgO_2.gMS^{-1}$ pour les glucides, $2,5.10^{-2}g.gMS^{-1}$ pour les lipides, $7,2.10^{-3}g.gMS^{-1}$ pour l'azote, 99 $mgO_2.gMS^{-1}$ pour la demande d'oxygène chimique, 0,026 $LCH_4.gMS^{-1}$ pour le rendement méthane et 0,34 jours en moyenne pour le temps de production de méthane. Les erreurs standards de prédiction sont de 35 $mgO_2.gMS^{-1}$ pour les glucides, $9,4.10^{-3}g.gMS^{-1}$ pour les lipides, $5,2.10^{-3}g.gMS^{-1}$ pour l'azote, 39 $mgO_2.gMS^{-1}$ pour la DCO, 0,025 $LCH_4.gMS^{-1}$ pour le rendement méthane et 0,24 jours en moyenne pour le temps de production de méthane. Cette étude montre que la SPIR permet une estimation rapide et efficace de la composition biochimique et des performances méthanogènes des déchets solides. Ces travaux ont fait l’objet d’un article scientifique : Charnier, C., Latrille, E.,Jimenez, J., Lemoine, M., Boulet, J.C., Miroux, J., Steyer, J.P. Fast characterization of solid waste content with near infrared spectroscopy in anaerobic digestion (*Soumis en avril 2016 à Waste Management)*.

Afin d'assurer l'application industrielle de l'analyse par SPIR, des indicateurs de l'intervalle de confiance ont été développés pour la prédiction des teneurs en glucides, azote et lipides, la DCO, le rendement du méthane et le temps de production de méthane. Cette estimation de l’intervalle de confiance de la prédiction est basée sur la distance euclidienne minimale entre l'échantillon et l'ensemble d'étalonnage. L’estimation de l’intervalle de confiance améliore la robustesse de l’analyse SPIR facilitant l’optimisation de la recette d’alimentation des unités de digestion anaérobie.

Si ces différents paramètres sont des indicateurs clés pour optimiser la recette d’alimentation, ils sont toutefois peu informatifs sur le lien entre la composition moléculaire de la matière organique et les performances de la digestion anaérobie. Les modèles PLSr calibrés pour la prédiction de ces 7 indicateurs sont une occasion unique d'étudier en profondeur la relation entre la composition de la matière organique et ces indicateurs. La contribution des différentes longueurs d'onde est obtenue par l'analyse des *b* coefficients. Ces longueurs d’ondes sont ensuite comparées avec l'absorbance des molécules en SPIR de composés purs rapportés dans la littérature. L’étude parallèle des *b* coefficients et des longueurs d'onde donne un aperçu des contributions de chaque molécule à ces 7 indicateurs. Elle met en évidence le rôle du phénol dans la fraction non biodégradable de la matière organique et l'importance des alcanes et alcènes sur le rendement en méthane d'un substrat. La prédiction de la teneur en glucides est liée avec la teneur en molécules oxygénées telles que les alcools. Dans une deuxième étape, les *b* coefficients des modèles PLSr ont été comparés les uns aux autres. Les relations analysées confirment le lien entre les molécules lipidiques et le rendement en méthane ou la demande chimique en oxygène. Ces relations indiquent aussi un impact négatif de la teneur en alcools sur la DCO, mais positif sur le rendement méthane et la biodégradabilité. Les molécules biodégradables sont liées avec une cinétique de production de méthane rapide. En contradiction avec la pensée dominante, les *b* coefficients impliqués dans les prédictions de la DCO et du rendement du méthane sont anti-colinéaires et presque orthogonaux avec ceux liés à la prédiction de la cinétique de production de méthane. Les relations obtenues par l’analyse des *b* coefficients sont compatibles avec les corrélations évaluées entre la composition biochimique et les performances du procédé de digestion anaérobie analysées par corrélation de Pearson. Seule la teneur en protéines est liée avec des molécules rapidement et hautement biodégradables par analyse de *b* coefficients alors qu'elle n’est pas corrélée avec le rendement ou la cinétique de production de méthane par corrélation de Pearson. Ceci pourrait être expliqué par une légère différence sur le jeu de données utilisé. Ces travaux ont fait l'objet d'un article scientifique : Charnier, C., Latrille, E.,Roger, J.M. , Miroux, J., Steyer, J.P. Parallel study of PLS *b* coefficients and near infrared wavelengths to assess molecules contribution to chemical oxygen demand, methane production yield and kinetics (*soumis prochainement)*.

Comme nous le montre l'étude du contexte actuel du secteur de la digestion anaérobie, l'optimisation de la stratégie d'alimentation est un enjeu essentiel de la co-digestion et nécessite une caractérisation rapide du substrat, la connaissance de l'état biologique du digesteur mais aussi l'utilisation de modèles prédictifs. Les deux premières étapes ayant été traitées, la suite de cette thèse se concentre sur l'utilisation de modèles prédictifs. Le modèle ADM1 a été choisi pour ses qualités prédictives mais l'étape d'hydrolyse/désintégration a dû être modifiée. Le mécanisme d'hydrolyse est mécanistiquement mal compris ce qui rend les modèles parfois non appropriés pour décrire les résultats expérimentaux. Des travaux précédents ont montré que plusieurs cinétiques de bio-dégradation sont observées correspondant à des fractions de substrat de bioaccessibilités différentes. La question étudiée ici est, en présence de plusieurs fractions biodégradables, celles-ci se dégradent-elles de façon simultanée ou séquentielle ? Nous avons montré qu'un modèle qui considère la dégradation simultanée des fractions peut ne pas avoir la résolution nécessaire pour représenter correctement les différentes cinétiques. Cette étude propose d'abord une discussion sur la structure du modèle d'hydrolyse de la matière organique (en bioprocédés aérobies et anaérobies). La modélisation de l'hydrolyse séquentielle des multiples fractions de matière organique a été évaluée comme une alternative à l'hydrolyse simultanée. Le modèle de l'hydrolyse séquentielle convient à tous les substrats biodégradables tandis que le modèle de l'hydrolyse simultanée n'est parfois pas applicable. L'hydrolyse séquentielle est donc préférée pour la modélisation de l'hydrolyse des substrats solides bien que le mécanisme biologique d'hydrolyse ne soit pas pleinement compris. Ces travaux ont fait l'objet d'un article scientifique : Jimenez, J., Charnier, C., Latrille, E., Torrijos, M., Harmand, J., Patureau, D., Sperandio, M., Morgenroth, E., Beline, F., Ekama, G., Vanrolleghem, P.A., Robles, A., Seco, A., Batstone, D.J., Steyer, J.P., Modelling hydrolysis: simultaneous versus sequential biodegradation of the hydrolysable fractions *(soumission en cours)*.

Un des inconvénients majeurs d'ADM1 réside dans l'estimation des paramètres du modèle. Les travaux sur la SPIR développés dans cette thèse permettent une estimation rapide des paramètres décrivant le substrat nécessaire à l'implémentation de l'ADM1. Ces données sont ensuite implémentées dans l'ADM1 modifié afin de simuler les performances de digestion anaérobie. Cette approche a été validée par comparaison des données expérimentales par rapport aux données prédites par simulation sur différentes expériences. Elle facilite l'utilisation d'ADM1 pour prédire les performances optimales de dégradation. L'utilisation de cette approche permet de prédire l'évolution de chaque variable estimée dans le modèle ADM1 telle que la concentration d'AGV, d'ammoniaque ou encore la composition du biogaz. Elle fournit ainsi un outil de contrôle avancé des unités de digestion anaérobie et d'optimisation de leur recette d'alimentation. Cette approche couplant SPIR et ADM1 pour la prédiction des performances d'un digesteur a fait l'objet d'une demande de brevet d'invention (n°16/51050), d'une demande divisionnaire du brevet d'invention n°16 51050 ainsi que d'un article scientifique (Charnier, C., Latrille, E.,Jimenez,J., Torrijos,M., Sousbie,P., Lemoine, M., Miroux, J., Steyer, J.P. Fast prediction of methane production rate for anaerobic digestion control using near infrared spectroscopy *(soumission en cours)*).

Cette thèse montre que la SPIR couplée à la PLSr permet une estimation rapide des paramètres décrivant le substrat dans ADM1 alors que le capteur SNAC permet l'estimation de l'état biologique du réacteur nécessaire pour ajuster les paramètres relatifs à l'état du digesteur d'ADM1. L'utilisation combinée de la SPIR et SNAC permet donc une implémentation facilitée de l'ADM1 modifié afin de prédire les performances de digestion anaérobie liées à une recette d'alimentation. Ce travail vise à aider la supervision des unités de méthanisation par l'opérateur humain. L'utilisation combinée des outils « SPIR/SNAC/ADM1 modifié » pourrait être un outil puissant pour l'optimisation du fonctionnement des unités de méthanisation, en particulier pour l'optimisation de la recette d'alimentation et la prédiction des performances associées.

Cependant, l'application de ses solutions à échelle industrielle nécessite de futurs développements. La méthode SNAC, développée en laboratoire doit être validée pour le suivi d'un digesteur industriel. A cette fin, un capteur utilisant la méthode SNAC a été prototypé et installé sur site industriel. Le prototype estime la température, le pH, la conductivité électrique, le ratio d'alcalinité et les concentrations en AGV, IC et TAN. L'utilisation conjointe de SNAC, du débit et de la composition du biogaz produit devrait permettre l'estimation et le suivi de l'état biologique du digesteur. Par exemple, les concentrations en composés inhibiteurs et l'activité des micro-organismes pourront être estimées au fil du temps. L'étape suivante consistera à estimer les paramètres ADM1 relatifs au procédé en utilisant SNAC ainsi que le débit et la composition du biogaz produit. L'estimation de ces paramètres couplés à la SPIR permettra l'implémentation de l'ADM1 modifié afin de prédire les performances de l'unité. Ce test à échelle industrielle permettra de comparer les performances prédites par ADM1 avec les performances réelles de l'unité. Des difficultés pourraient résider sur la prédiction du pH, l'adaptation de la flore microbienne ou encore la présence de composés non mesurés tels que le souffre ou les antibiotiques. Si cette approche est validée à échelle industrielle, elle pourrait être étendue afin de mieux contrôler et anticiper la qualité et quantité de digestat produit. Un projet a été soumis à l'ADEME permettant de tester et améliorer SNAC ainsi que l'approche SPIR plus ADM1 à l'échelle d'un territoire afin d'y optimiser les flux de déchets organiques.

Enfin, l'analyse par SPIR doit actuellement être effectuée sur des échantillons secs. Trois inconvénients sont liés au séchage : les composés volatils ne peuvent pas être analysés, le processus de séchage ralentit considérablement l'analyse et l'utilisation de la SPIR en ligne n'est pas envisageable. Pour une application à grande échelle, la haute fréquence de changement et l'hétérogénéité des substrats est une menace pour la caractérisation du substrat et, de fait, la prédiction des performances des unités associées à une recette d'alimentation. L'approche développée sur échantillons secs utilisant la SPIR permet de caractériser plus d'échantillons à partir d'un substrat hétérogène dans un court laps de temps, ce qui correspond d'avantage aux besoins industriels. Néanmoins, un futur développement intéressant serait d'adapter cette technologie à des échantillons humides pour la réalisation de capteurs SPIR capables de caractériser en temps réel tout substrat de l'unité. L'obstacle est l'absorption de l'eau en SPIR qui perturbe l'analyse. Des travaux sont actuellement en cours pour déterminer la faisabilité de cette approche sur substrat humide.

Résumé étendu

List of Communications

Papers in Peer Reviewed International Journals

Charnier, C., Latrille, E., Lardon, L., Miroux, J., Steyer, J.P., 2016. Combining pH and electrical conductivity measurements to improve titrimetric methods to determine ammonia nitrogen, volatile fatty acids and inorganic carbon concentrations. Water Res. 95, 268–279. doi:10.1016/j.watres.2016.03.017

Charnier, C., Latrille, E.,Jimenez, J., Lemoine, M., Boulet, J.C., Miroux, J., Steyer, J.P. Fast characterization of solid waste content with near infrared spectroscopy in anaerobic digestion (*Submitted in April 2016 to Waste Management)*.

Charnier, C., Latrille, E.,Jimenez,J., Torrijos,M., Sousbie,P., Lemoine, M., Miroux, J., Steyer, J.P. Fast prediction of methane production rate for anaerobic digestion control using near infrared spectroscopy *(Being submitted)*.

Jimenez, J., Charnier, C., Latrille, E., Torrijos, M., Harmand, J., Patureau, D., Sperandio, M., Morgenroth, E., Beline, F., Ekama, G., Vanrolleghem, P.A., Robles, A., Seco, A., Batstone, D.J., Steyer, J.P., Modelling hydrolysis: simultaneous versus sequential biodegradation of the hydrolysable fractions *(Being submitted)*.

Charnier, C., Latrille, E.,Roger, J.M. , Miroux, J., Steyer, J.P. Parallel study of PLS *b* coefficients and near infrared wavelengths to assess molecules contribution to chemical oxygen demand, methane production yield and kinetics (*Waiting for the acceptance of the other papers about near infrared)*.

Patents

Demande de brevet d'invention n°16/51050 related the use of near infrared coupled with partial least square regression model for the prediction of AD performances using ADM1

Demande divisionnaire du brevet d'invetion n°16 51050 related the use of near infrared for the prediction of AD performances

Demande de brevet d'invention n°15/59767 related to the System of titration for total ammonia nitrogen, volatile fatty Acids and inorganic Carbon

List of Communication

Communications in International Conferences

Kouas, M., Charnier, C., Harmand, J., Sousbie, S.S.P., Steyer, J.P., Torrijos, M., 2002. Kinetic modelling of anaerobic digestion with new fractionation using simple model and ADM1. AD14, IWA Word Congress on Anaerobic Digestion, Viña del Mar, Chile, November 2015 *(oral presentation)*.

Jimenez, J., Charnier, C., Latrille, E., Torrijos, M., Harmand, J., Patureau, D.,Sperandio, M., Morgenroth, E.,Beline, F., Ekama, G., Vanrolleghem, P.A., Robles, A.,Seco, A.,Batstone, D.J., Steyer, J.P., Modelling hydrolysis: simultaneous versus sequential biodegradation of the hydrolysable fractions, WWTmod2016, April 2016, Annecy, France *(oral presentation)*.

Charnier, C., Latrille, E, Moscoviz R., Miroux, J., Steyer, J.P., Impact of the biochemical composition on the methane production yield and kinetic, XII DAAL, IWA Symposium and Workshop Latin American Anaerobic Digestion, October 2016, Cusco, Peru *(accepted for oral presentation)*

Charnier, C., Latrille, E., Juznic-Zonta, Z., Miroux, J., Steyer, J.P., Feeding strategy optimization assisted by near infrared and real time process monitoring to improve methane production, International Conference Progress in Biogas IV, March 2017,Stuttgart, Germany (*submitted in June 2016*).

Communications in National Conferences

Charnier, C., Latrille, E., Jimenez, J., Torrijos, M., Sousbie, P., Lemoine, M., Miroux, J., Steyer, J.P. , Optimisation du fonctionnement des unités de digestion anaérobie par une meilleure connaissance des substrats et la prédiction de leur biodégradation, Journée Recherche Innovation, Février 2016, Limoges, France *(présentation orale)*.

Charnier, C., Latrille, E, Miroux, J., Steyer, J.P., Combiner la mesure du pH et de la conductivité afin d'augmenter la précision et la robustesse de l'estimation en azote ammoniacal, acides gras volatils et carbone inorganique par titration, Journée Recherche Innovation, Février 2016, Limoges, France *(présentation orale)*.

Charnier, C., Latrille, E., Pacaud, S., Miroux, J., Fick, M., Steyer, J.P., Prédiction des performances de digestion des substrats : apport pour le pilotage du digesteur. 5èmes JOURNEES de la METHANISATION Applications Agricoles et Industrielles ,Décembre 2016, Chambéry, France *(acceptée pour une présentation orale)*.

Charnier, C., Rossard, V., Latrille, E., Boulet, J.C., Gogé, F., Steyer, J.P., Science reproductible appliquée à la qualité des modèles de calibration et des prédictions à partir de mesures en spectrométrie proche infra-rouge : développement de l'application ChemFlow , 14ème Journées de la Mesure et de la Métrologie, Octobre 2016, La Chaussée-Saint-Victor, France *(accepté pour une présentation orale)*.

Lardon, L., Charnier, C., Latrille, E., Juznic-Zonta, Z., Miroux, J., Steyer, J.P., Apport de la modélisation en continu pour les procèdes de méthanisation : exemples d'applications industrielle et agricole, Journée Recherche Industrie 2015, Rennes, France (*Poster*).

Table of Contents

Chapter 1. Anaerobic Digestion in Europe and in France. 1

1.1 Chapter Guidelines 3
1.2 Technical and Graphical Abstracts 5
1.3 World Energy Challenge 7
1.4 Biogas over European Union 7
1.5 France, a Potential Vector of Growth for the European Anaerobic Digestion Sector 10
1.6 Contents of the Thesis 14

Chapter 2. Materials and Methods 19

2.1 Materials and Methods Guidelines 21
2.2 Titrimetric Methods to Determine Ammonia Nitrogen, Volatile Fatty Acids and Inorganic Carbon Concentrations 23
  2.2.1 Experimental Titration Device 23
  2.2.2 Titration Protocol 25
  2.2.3 Analytical Methods for Reference Data 25
  2.2.4 Titration Model 25
  2.2.5 Signal Processing Coupling pH and Electrical Conductivity Measurements 30
2.3 Analytical Methods for Solid Waste Characterization 32
  2.3.1 Biochemical Composition of Solid Waste 32
  2.3.2 Bio-accessibility Assessment by Chemical Sequential Extraction 33
  2.3.3 Anaerobic digestion experiments 34
2.4 Spectroscopic Analyses 35
  2.4.1 Near Infrared 35
  2.4.2 3D Fluorescence 36
2.5 Solid Waste Analyzed 36
  2.5.1 Samples Analyzed for Fast Spectroscopic Characterization 36
  2.5.2 Samples Analyzed for Correlation Analyses 39
2.6 Correlation Analysis and Regression Models 41
  2.6.1 Pearson's Correlation and Cosine 41
  2.6.2 Principal Component Analysis 41
  2.6.3 Partial Least Square Regression Model and Performance 41
2.7 Dynamic Anaerobic Digestion Models 44
  2.7.1 Dynamic Models: Modified ADM1 44
  2.7.2 ADM1 Modelling Approach Using NIRs 45

Chapter 3. Combining pH and Electrical Conductivity Measurements to Improve Titrimetric Methods to Determine Ammonia Nitrogen, Volatile Fatty Acids and Inorganic Carbon Concentrations 49

3.1 Chapter Guidelines 51
3.2 Technical and Graphical Abstracts 53
3.3 Introduction of the Chapter 55
  3.3.1 Key Parameters to Estimate the Biological State of Anaerobic Digestion Processes 55
  3.3.2 Available Sensors for VFA, IC and TAN Concentrations Measurement 57
3.4 Materials and Methods Related to this Chapter 61

3.5 Results of the Existing Methods 61
  3.5.1 Estimation of VFA and IC 61
  3.5.2 Estimation of TAN Using Non-Linear Resolution 66
3.6 Accurate Estimation of VFA, IC and TAN Concentrations by Coupling pH and Electrical Conductivity Measurement. 68
  3.6.1 Building of a Decision Tree 69
  3.6.2 Example of a Titration Using SNAC 71
  3.6.3 Discussion of the Advantages between the Existing Methods and SNAC 73
3.7 SNAC Prototype 76
3.8 Conclusion 78

Chapter 4. Relations between Biochemical Composition and Anaerobic Digestion Performances of Solid Waste 81

4.1 Chapter Guidelines 83
4.2 Technical and Graphical Abstracts 85
4.3 Introduction of the Chapter 87
  4.3.1 Analytical Methods for Biochemical Characterization of Substrates 87
  4.3.2 Biodegradability and Methane Production Rate Estimation 88
4.4 Materials and Methods Related to this Chapter 91
4.5 Validation of Biochemical Analyses on Solid Waste 91
4.6 Dataset Exploration Using a Principal Component Analysis 94
4.7 Variables Correlation Exploration Using a Principal Component Analysis 95
4.8 Variables Correlation Analysis using Pearson Correlation Coefficients 98
4.9 Discussion about Variables Correlation Analysis 99
4.10 Predictability of the Kinetics of Methane Production from Biochemical Composition 102
4.11 Conclusion on the Existing Relation between the Biochemical Composition and the Methane Production Performances 103

Chapter 5. Characterization of the Solid Waste Biochemical Composition and Anaerobic Digestion Performances Using Spectroscopic Analyses 105

5.1 Chapter Guidelines 107
5.2 Technical and Graphical Abstracts 109
5.3 Introduction of the Chapter 111
  5.3.1 BMP Estimation on Early Biogas Production 111
  5.3.2 Spectroscopic Analyses 114
5.4 Materials and Methods Related to this Chapter 116
5.5 Comparison of 3D Fluorescence and Near Infrared Spectroscopy for Biodegradability Prediction 117
5.6 Fast Characterization of Solid Waste Biochemical Composition with Near Infrared Spectroscopy 121
  5.6.1 Combinations of the Calibration Model Using PLS Regression 121
  5.6.2 Model Validation on an Independent Dataset 124
  5.6.3 Discussion Regarding the Selected Model 124
5.7 Prediction of the Methane Production Rate in Batch Conditions Using NIR Spectroscopy 127
  5.7.1 Prediction of the Methane Yield 127
  5.7.2 Prediction of the Methane Production Time 127
  5.7.3 Methane Production Rate Prediction 132

5.7.4 Simple Indicator of Methane Production Performances 132
5.8 Confidence Interval Determination Using Spectrum Distance from the Calibration Set 133
5.9 Conclusion on NIRs Characterization 139

Chapter 6. Parallel Study of PLS *b* Coefficients and Near Infrared Wavelengths to Assess Molecules Contribution to Chemical Oxygen Demand, Methane Production Yield and Kinetics. 141

6.1 Chapter Guidelines 143
6.2 Technical and Graphical Abstracts 145
6.3 Introduction of the Chapter 147
6.4 Materials and Methods Related to this Chapter 151
6.5 Homogenization of the Chemometric Treatment 151
6.6 b Coefficients Analysis and Molecules Contribution to the Predicted Parameters 152
6.7 Relations between the *b* Coefficients of the Models 157
6.8 Conclusion 160

Chapter 7. Challenging the Simultaneous Hydrolysis Concept and Fast Implementation of ADM1 for the Anaerobic Co-digestion Performances Simulation 169

7.1 Chapter Guidelines 171
7.2 Technical and Graphical Abstracts 173
7.3 Introduction of the Chapter 175
7.4 Materials and Methods Related to this Chapter 181
7.5 Modelling Hydrolysis: Simultaneous versus Sequential Biodegradation of the Hydrolysable Fractions 182
7.5.1 Definition of Simultaneous and Sequential Concepts 182
7.5.2 Results Obtained on Several Substrates 183
7.5.3 Fractions Biodegradability: Case Studies on Apple and Wheat Straw 186
7.5.4 Case of Continuous Reactor 188
7.5.5 Conclusion on Hydrolysis Modelling 190
7.6 ADM1 Implementation Assisted by Near Infrared Spectroscopy for the Anaerobic Digestion Performances Simulation 191
7.6.1 ADM1 Calibration from NIR Prediction 191
7.6.2 Discussion on the MPR Prediction Using NIR 194
7.7 Example of Industrial Application 194
7.8 Conclusion 195

Chapter 8. General Conclusion and Outlooks 197

8.1 Conclusion 199
8.2 Outlooks 204
8.3 Projected development 205

Chapter 9. References 209

List of Figures

Figure 1: World global nutrients ($N+P_2O_5+K_2O$) consumption and forecast *(FAO, 2015).* ------------------------ 8
Figure 2: Percentage of biogas plants in European countries in 2013 by type *(EBA, 2014).* -------------------- 9
Figure 3: Number of biogas plant evolution in Europe in 2012 and 2013 *(EBA, 2014).* ------------------------ 10
Figure 4: Biogas rate of return depending on the type of substrate they are digesting *(G.Bastide, 2014).* -- 13
Figure 5: Feeding strategy optimization graphical abstract -- 17
Figure 6: Calibration and validation set features of the experimental data. Description of the methane production time required to reach a certain ratio of the methane yield from 5 to 95%.---------------------- 39
Figure 7: Boxplot of data set characteristics. Figure 7A represents the lipid, protein, carbohydrate (carbo) contents, COD, methane yield (CH4Yield) and biodegradability (BD) on centered and scaled data. Figure 7B represents the methane production time expressed in days needed to reach a percent of the methane yield from 5 to 95%. --- 40
Figure 8: Modified ADM1 with four different particulate inputs X_{RC}, X_{MC}, X_{SC} and X_{NE}. ------------------------ 45
Figure 9: Modeling approach diagram-- 47
Figure 10: Effect of pH on ammonia and volatile fatty acids inhibition. Red arrows indicate the inhibitory effect of non-ionized form of TAN and VFA on the bacteria.-- 57
Figure 11: Estimation of total VFA (tVFA) concentrations from the recalibrated equation of Ripley et al., (1986) (equation 4) and the recalibrated equation of Nordmman (1977) (equation 5) plotted against the tVFA measured by gas chromatography for 52 energy crop and slurry digestate samples *(Purser et al., (2014)).*--- 59
Figure 12: Buffer capacity curve for a complex aqueous system containing 5mg $CO_2.L^{-1}$, 7 mg o-PO_4-$P.L^{-1}$, 15 mg NH_4^+-$N.L^{-1}$ and 0.6 $meq.L^{-1}$ of an unspecified soap *(Van Vooren 2000).*---------------------------- 60
Figure 13: Estimation of IC concentrations by the method of Nordmann (1977), Ripley et al (1986), Purser et al (2014) and Zaher (2005) -- 62
Figure 14: Estimation of VFA concentrations by the method of Nordmann (1977), Ripley et al (1986), Purser et al (2014) and Zaher (2005). The bisector is represented by a straight line on the graphics.----- 64
Figure 15: Buffer intensity of lactic acid, acetate and carbonate and their influences on the estimation of VFA and IC by Nordmann (1977). The VFA analysis thresholds indicate the range of pH in which only IC and VFA are supposed to influence the titration using Nordmann (1977) equations.---------------- 65
Figure 16: Simulation of Nordmann (1977) estimation for IC and VFA on 100 days supposing a constant concentration of lactic acid (0.03 mol.L-1)-- 66
Figure 17: TAN concentrations estimation based on buffer capacity using the method developped by Zaher (2005) on the left part and SNAC on the right part. --- 67
Figure 18: Comparison of the estimation on buffer conductivity and buffer capacity for VFA and TAN concentration on up titration--- 68
Figure 19: SNAC decision tree --- 70
Figure 20: Titration of an industrial effluent, comparison of buffer capacity and buffer conductivity ----- 72
Figure 21: Comparison of VFA estimation using buffer capacity, Zaher (2005), or buffer capacity coupled to buffer conductivity (SNAC) --- 73
Figure 60: SNAC prototype, front picture.--- 76
Figure 61: SNAC web interface designed in Rshiny-- 77
Figure 62: Part of the SNAC titration report in French --- 77
Figure 22: Confidence regions for k_{hyd} (day^{-1}), and f_d (frac) using different fitted data sets, including whole full-scale gasflow data sets (480 measurements for each digester), effluent solids (450 measurements for each digester), one example of a 30-day gasfllow data set (30 measurements), and the BMP tests (10 gas measurements in triplicate batches). The arrows around estimates on solids data indicate that the confidence regions are unbounded in k_{hyd}. The inset (top left) indicates a magnified view of the batch test confidence region *(Batstone et al, 2009).*-------------------------------------- 90
Figure 23: Effect of pre-sulfuric acid hydrolysis on glucose estimation --- 92

Figure 24: 3D representation of the dataset on the three first principal components of the PCA. Spheres are colored depending on their biochemical composition, turquoise spheres: lipid content up to 50%, yellow spheres: protein content up to 50%, red spheres: carbohydrate content up to 50%, grey spheres represent the remaining data. ---- 95

Figure 25: PCA variables representation on PC1/PC2 (up part) and PC1/PC3 (down part). BD and CH4Yield stand respectively for biodegradability and methane yield. Time 05 to 95 represent the times needed to reach a percent of the methane yield from 5 to 95%. ---- 97

Figure 26: Matrix of Pearson's correlation coefficients between lipid, protein and carbohydrate contents, COD, methane yield (CH4YIELD) and biodegradability (BD), methane production time expressed in days needed to reach a percent of the methane yield from 5 to 95%. ---- 99

Figure 27: Correlation network between lipids, proteins and carbohydrates content, COD, methane yield (CH4YIELD) and biodegradability (BD), methane production time expressed in days needed to reach 70 percent of the methane yield. The level correlation is represented by the intensity and thickness of the link. ---- 101

Figure 28: Prediction of methane production time needed to reach 70% of the methane yield using carbohydrate, protein, lipid contents, COD, biodegradability and methane yield as variables of PLS regression model. ---- 103

Figure 29: Correlation between respiration index (RI4) and biogas production (GB21) measured in samples of excavated waste from landfills *(Cossu and Raga, 2008)*. ---- 111

Figure 30: Accumulated specific methane production for three different samples (i.e. mixed food waste as a household waste, banana stem as an agricultural waste, and butter residues as a lipid rich waste). The experimental data is represented by the solid line, waste, and butter residues as a lipid rich waste). The experimental data is represented by the solid lines and the symbols, while the predicted gas profile is shown by the dashed line *(Strömberg et al., 2015)*. ---- 112

Figure 31: Scheme of an example of Envital® assay in microplate based on fluorescence measurement *(Bellaton et al., (2016))*. ---- 113

Figure 32: Molecule families-like in 3D spectra of manure based on (Muller et al, 2014). I: Tyrosine-like II: Tryptophan-like III: Microbial product protein-like IV: Fulvic acid-like V: Glycolated protein-like VI: Lignocellulose-like VII: Humic acid-like ---- 114

Figure 33: Observed versus predicted biodegradability using chemical sequential extraction and 3D fluorescence *(Jimenez et al., 2014)*. ---- 115

Figure 34: Predicted using NIR spectra vs measured BMP values. The solid line indicates the 1:1 relationship. For each point, horizontal lines indicate the Standard Deviation of the BMP test of the duplicates samples and the vertical lines indicate the Standard Deviation of the NIR measurement of the triplicates samples. The dotted line represents the Dunne effect. Calibration data set are represented by blue circles and validation data set by purple squares *(Lesteur et al., 2011)*. ---- 115

Figure 35: Comparison of cross-validation and validation PLS performances of NIR (left) and 3D fluorescence coupled with chemical sequential extraction (right) for the prediction of methane yield (up) and COD (bottom). The blue squares (♦) represent the prediction of the independent validation set whereas the red crosses (+) represent the estimation of the calibration set. The line represents the bisector between reference and predicted or estimated values. ---- 120

Figure 36: Prediction performances of carbohydrate, nitrogen and lipid contents and COD, respectively on the top left, top right, bottom left and bottom right. The blue squares (♦) represent the prediction of the independent validation set whereas the red crosses (+) represent the estimation of the calibration set. The line represents the bisector between reference and predicted values. ---- 126

Figure 37: PLS regression performances on cross-validation with 3 different chemometric treatments and considering the methane production time as correlated or independent values. Treatment 1 corresponds to absorbance and SNV, treatment 2 to absorbance, SNV and Savitzky-Golay filter and treatment 3 to absorbance, SNV and detrending. ---- 130

Figure 38: PLS performances on the validation set (up) and representation of the predicted value of both cross-validation and validation (bottom). ---- 131

Figure 39: Cumulated Methane production curve for anaerobic of salad (left) and cauliflower (right) in batch. Comparison of predicted with reference data -- 132
Figure 40: Estimation of the validation set substrates performances expressed as a function of the methane yield and the production time of 70% of the methane yield. Abscissa arrow indicates the methane production kinetics and ordinate arrow the methane yield. -- 133
Figure 41: Relations between the distance of a spectrum from the calibration set and its error of prediction. Example of the methane yield prediction -- 135
Figure 42: Predictions confidence interval depending on the minimal Euclidian distance from the spectrum to the calibration dataset for carbohydrate, nitrogen and lipid contents, COD, methane yield and methane production time at 70% of the methane yield. The purple line models the confidence interval. --- 137
Figure 43: Prediction errors on the validation set depending on the minimal Euclidian distance from the spectrum to the calibration dataset for carbohydrate, nitrogen and lipid contents, COD, methane yield and methane production time at 70% of the methane yield. The purple line corresponds to the equations of the confidence interval established on the calibration set (Table 10). ------------------------ 138
Figure 44: *b* coefficients (Bcoef) of carbohydrate, protein and lipid contents, biodegradability, COD, methane yield and methane production time prediction from spectra pretreated by absorbance/SNV/detrending -- 153
Figure 45: *b* coefficients (Bcoef) cosine matrix of carbohydrate, protein and lipid contents, methane yield (CH4Yield), COD, biodegradability and methane production time (ProductionTime). --------------- 159
Figure 46: The ADM1 model with the following biochemical processes: (1) acidogenesis from monosaccharides (MS), (2) acidogenesis from amino acids (AA), (3) acetogenesis from long chain fatty acids (LCFA), (4) acetogenesis from propionate, (5) acetogenesis from butyrate (HBu) and valerate (HVa), (6) aceticlastic methanogenesis, and (7) hydrogenotrophic methanogenesis *(Batstone et al., 2002)*.--- 176
Figure 47: Disintegration COD flux in the modified model including two different fractions of composite, Xcr and Xcs *(Mottet et al., 2013)*. --- 178
Figure 48: Conceptual comparison of surface-based kinetics with and without particle breakup *(Dimock and Morgenroth, 2006)*.-- 178
Figure 49: Experimental batch assay of mango. (a) Cumulative biogas production per reactor volume. (b) Gas production rate (computed from cumulative derivative) *(García-Gen et al., 2015b)*.-------------- 179
Figure 50: Structure of simplified particle break-up model. Illustration of one hypothesis leading to sequential degradation based on the particle size *(Yasui et al., 2008)*. -- 179
Figure 51: Illustration of the chemical sequential extraction (Jimenez et al., 2015a) to estimate the bioaccessibility. X_{RC}, X_{MC} and X_{SC} stand respectively for readily, moderately and slowly accessible organic matter. -- 181
Figure 52: Schematic definition of simultaneous and sequential concepts. Xrc, Xmc and Xsc stands respectively for the readily biodegradable, the moderately biodegradable and the slowly biodegradable fractions. To simplify the figure, Xne is not represented.--------------------------------------- 183
Figure 53: Methane Production Rate curves obtained experimentally and by the simulations with simultaneous model and with the sequential model (a: carrot, b: cauliflower, d: lettuce, e: wheat straw, e: potatoes).-- 185
Figure 54: Anaerobic biodegradation of the apple (a) and the wheat straw (b) fractions------------------ 186
Figure 55: Conceptual scenarios proposed to explain accessible fractions hydrolysis kinetics and fed-batch digestion of wheat -- 187
Figure 56: Illustration of the sequential approach in a continuous reactor ------------------------------------ 188
Figure 57: Continuous reactor simulation using simultaneous model or sequential model based either on bio-accessibility limitation (access) or inhibition.--- 190
Figure 58: Cumulated methane production curve for anaerobic of salad (left) and cauliflower (right) in batch. A and B, comparison of predicted and modelled data using ADM1 considering the particulate organic matter as a single input or fractioning the particulate organic matter in four inputs with their own kinetics of disintegration for salad and cauliflower respectively. B. C and D,

modelling of the methane cumulated production for salad and cauliflower AD taking into account the error of prediction for salad and cauliflower respectively. Methane production times and methane yield were chosen at ± 1 SEP of the predicted value. The grey aera represents the range of confidence. -- 193
Figure 63: Plant performances prediction using ADM1 plus NIRs --- 195
Figure 59: Feeding strategy optimization, current and future development ----------------------------------- 203

List of Tables

Table 1: Feed in Tariffs and premiums for biogas injection and electricity production in 2016 in France - 12
Table 2: Reference analyses data of the samples used for the experiments. Volatile fatty acids (VFA) concentration was analyzed by GC, Inorganic Carbon (IC) by TOC and Total Ammoniacal Nitrogen (TAN) by TKN. Partial Alkalinity (PA) and Intermediate Alkalinity (IA) were estimated using the formula developed in Ripley et al (1986) -- 24
Table 3: Dataset features for NIR calibration and validation-- 38
Table 4: comparison of accuracy of the estimation between the existing methods and SNAC on volatile fatty acids, inorganic carbon and total ammonia nitrogen -- 75
Table 5: Comparison of analysis methods for the estimation of proteins and carbohydrates. Selected methods are in bold. -- 93
Table 6: Dataset features used for methane yield and COD prediction --- 117
Table 7: PLS model performance on cross validation and validation set for COD and methane yield prediction using 3D fluorescence spectroscopy coupled with chemical sequential extraction (FLUO) or near infrared spectroscopy (NIRs). -- 119
Table 8: Influence of spectra treatment on PLSr performances, selected model values are in bold. For calibration dataset details, see Materials and Methods section 2.5.1.-- 123
Table 9: Validation of the model on an independent dataset for nitrogen, carbohydrate and lipid contents and COD prediction. For validation dataset details, see Materials and Methods section 2.5.1.124
Table 10: Relation between the Minimal Euclidian Distance (MED) and the Confidence Interval (CI) for carbohydrate, nitrogen and lipid contents, COD, methane yield and methane production time prediction at 70% of the methane yield. In the table, the variables are ranked by chemometric treatment. --- 136
Table 11: Influence of alkane and alkene on NIR spectra. The black marks represent the range of absorbance --- 148
Table 12: Influence of alcohol, ketone, ester, epoxide, aldehyde and carboxylic acid on NIR spectra. The black marks represent the range of absorbance -- 149
Table 13: Influence of organic molecules containing nitrogen on NIR spectra. The black marks represent the range of absorbance -- 150
Table 14: Influence of organic molecules containing sulfur on NIR spectra. The black marks represent the range of absorbance --- 150
Table 15: Sparse PLSr performances. Methane production time is expressed as a mean quadratic error for the 19 time predicted along the methane production curve from 5 to 95% of the methane yield. -- 152
Table 16: Parameters of simulation of the batches using simultaneous model or sequential model ---- 184

Nomenclature

ADM1	Anaerobic digestion model n°1
BMP	Biochemical methane potential
BSA	Bovine serum albumin
BSM2	Benchmark Simulation Model n°2
CI	Confidence interval
COD	Chemical oxygen demand
DOM	Dissolved organic matter
Ed	Euclidian distance
FOS	Volatile fatty acids estimation from Nordmann, 1977
HEM	Hexane extractible matter
IA	Intermediate alkalinity
IC	Inorganic carbon
KeepX	Number of variables to keep in X
max	Maximum
Md	Mahalanobis distance
MED	Minimal Euclidian distance
MEOM	Moderately extractible organic matter
min	Minimal
MPR	Methane production rate
NAS	Net Analyte Signal
NEOM	Non extractible organic matter
NIR	Near infrared
NIRs	Near infrared spectroscopy
PA	Partial alkalinity
PCA	Principal components analysis
PCC	Pearson correlation coefficients
PLS	Partial least square
PLSr	Partial least square regression
REOM	Readily extractible organic matter
RSD	Relative standard deviation
SD	Standard deviation
SDr	Standard deviation of repeatability
SEC	Standard error of calibration
SECV	Standard error of cross-validation
SEL	Standard error of laboratory
SEOM	Slowly extractible organic matter
SEP	Standard error of prediction
SNAC	System of titration for total ammonia Nitrogen, volatile fatty Acids and inorganic Carbon

sPLS	Sparse PLS
TA	Total alkalinity
TAC	Inorganic carbon estimation from Nordmann, (1977)
TAN	Total Ammoniacal Nitrogen
TKN	Total Kjeldahl Nitrogen
TS	Total solid
VFA	Volatile fatty acids
VS	Volatile solid
X_{MC}	Moderately accessible fraction
X_{NE}	Not extractible fraction
X_{RC}	Readily accessible fraction
X_{SC}	Slowly accessible fraction
Z	Non-reactive ions
β	Buffer capacity
λ	Molar conductivity
б	Conductivity

Remerciements

C'est avec émotion que j'écris cette partie des plus importantes. Plaisir d'avoir fini la rédaction et de faire le point sur trois années de recherche, Impatience de voir évoluer les projets développés, Appréhension qu'ils ne voient jamais le jour, Espoir qu'ils participent à faire avancer la recherche et facilite le développement du secteur, Envie de les mener à maturité, Désir de se lancer dans de nouveaux projets. Cela marque le début d'un nouvel inconnu et la fin d'une étape riche en échanges et collaborations, Je m'en excuse donc par avance ; une thèse étant un pour moi un projet collaboratif ; les remerciements sont un peu long. Je souhaite tout d'abord remercier mes encadrants et directeurs de thèse, Eric Latrille, Jérémie Miroux et Jean-Philippe Steyer. Ensuite viennent les membres du jury que je veux remercier d'avoir répondu rapidement présent pour la soutenance.

Eric merci pour ton aide et tes enseignements précieux sur tous les aspects techniques de cette thèse. Tu m'as appris à coder des modèles dynamiques, faire des PLS, des interfaces de prototype, du proche infra-rouge…etc. Malgré ton emploi du temps chargé, j'ai toujours pu faire un saut dans ton bureau quand un problème se présentait pour en discuter avec toi et aboutir à une solution. Merci aussi d'avoir partagé avec moi ton amour de la nature et de la région. Je me souviens des excursions en kayak, des canyoning ou encore de tes justes conseils œnologiques !

Jérémie, merci de m'avoir fait confiance pour cette thèse, d'avoir cru aux projets développés et compris la subtilité d'une thèse CIFRE entre confidentialité et partage. Merci aussi de favoriser et financer l'aboutissement des recherches en solutions industrielles, preuves pour moi de ta confiance dans ces solutions. Tu m'as permis de rencontrer de nombreux acteurs de la méthanisation pour comprendre les différentes visions du secteur. Merci aussi à ta famille et toi pour votre chaleureux accueil dans votre maison lyonnaise.

Jean-Philippe, je ne peux m'empêcher de commencer par te remercier pour tes encouragements permanents au long de cette thèse. Tu sais trouver les mots pour complimenter tes doctorants. Plus sérieusement, merci pour ton expertise et tes conseils, ton aide sur la rédaction des manuscrits et brevets et ton soutien sur cette thèse. Comme toute personne au LBE, j'ai pu bénéficier de la renommée du laboratoire que tu as su maintenir à haut niveau pendant ton mandat ainsi que de l'important tissu universitaire et industriel qui entoure le LBE. Cette thèse, riche en collaborations, en est un des fruits.

Merci à Julie Jimenez pour ses précieux conseils et sa formation sur les problématiques de modélisation ADM1, caractérisations biochimiques de la matière, fractionnement biochimique et analyses par fluorescence 3D. Dans un contexte moins professionnel, merci à toi, Pipine pour ces excellents souvenirs tant en randonnées, voyages qu'innombrables soirées.

Merci à Margaux Lemoine, que j'ai la chance d'avoir eu en stage pour m'accompagner pendant cette thèse, pour la qualité et le sérieux de son travail. Tes travaux ont vraiment été majeurs pour moi. Merci aussi pour les riches échanges et discussions que nous avons eu, dans un cadre professionnel ou non, ainsi que pour ta gentillesse. Je te souhaite beaucoup de bonheur dans ta vie et future carrière d'institutrice.

Merci aussi à Laurent Lardon pour ses conseils au début de ma thèse et cette doctrine qui m'a accompagnée « c'est en forgeant qu'on devient forgeron ». J'aurais aimé avoir eu l'occasion de travailler d'avantage ensemble mais comprends tout à fait les priorités de la vie de famille. J'espère que tout se passe pour le mieux au Danemark.

Thanks also to the BioEnTech Team, Jérémie Miroux, Zivko Juznic-Zonta, Francesco Novelis and Malcolm Djurick for the good atmosphere during every meeting and afterwards, Francesco this one is just for you: "SNAAAAC!". I have plenty of reasons to thank you but I'll keep it short. Zivko thanks for your advice and the rich exchange we had about modeling, Francesco thanks for your support and time on the SNAC prototype, Malcolm thanks for your help and kindness dealing with customers and redacting reports.

Merci à Jean-Michel Roger et Victor Alcaraz Gonzales pour leurs précieux conseils lors des comités de thèses. Merci particulièrement à Jean-Michel de m'avoir reçu à l'IRSTEA et d'avoir pris du temps pour travailler ensemble sur le proche Infra-rouge. Merci enfin pour ton expertise et aide sur les *b* coefficients. Merci aussi à Jean-Claude Boulet pour son aide sur la boite à outils FACT et ses conseils sur l'analyse des distances.

Merci à l'équipe de l'ITE pour votre aide et professionnalisme. Plus particulièrement à Romain Cresson, Olivier Azam, Bruno Marty et Stéphanie Deville. J'espère avoir d'autres occasions de collaborer avec vous.

Merci à Michel Torrijos, Philippe Sousbie et Mokhles Kouas pour leur expertise sur la digestion des substrats en batch. Merci de m'avoir aidé et donné l'occasion de collaborer avec vous sur la partie modélisation, ce qui a été essentiel dans ma thèse. Merci à Guillaume Guizard, sans qui le prototype SNAC n'aurait jamais vu le jour. Merci à Arnaud Helias de m'avoir donné l'opportunité d'enseigner. Merci à Jérôme Harmand pour nos échanges sur la modélisation. Merci à Marjolaine Hamelin pour son aide et expertise sur la rédaction des articles.

Merci à mes chers colocataires du Taudis joyeux. Pour ceux qui l'ignorent, le Taudis joyeux est une auberge (le mot est maintenant validé) située au cœur d'un quartier plein de charme où la guitare berce vos journées, vous trouvez en cette demeure, outre un corbeau d'adoption et de nombreuses créations originales 100% recyclées, une ambiance chaleureuse et de nombreuses opportunités de festoyer. Bref, merci à l'Espagnolo , Mr. Porc et le Roux pour cette année et demie passée ensemble dans la bonne humeur et le partage.

Merci à Thibaut pour son accueil au LBE, un agro ça rassure toujours, et ces excellentes soirées passées ensemble. Evite les toits mais on remonte la fontaine quand tu veux ! Merci à Marie-Lou et Wendy pour leur sympathie et gentillesse. Merci aussi à Mamie-Lou et Clapoche pour ces bons moments passés ensemble. Merci à Lucille pour tes recommandations sur la thèse et les riches discussions que nous avons eu, entre autres, sur les trajets Narbonne-Toulouse. Merci à Silvio de ta bonne humeur, promis, on trouvera un moment pour pêcher ! Thank you Afifi/Selfifi for your permanent good mood, I will remember your motto « It's France It's Freedom », take care! Merci à tous ceux avec qui j'ai pu passer de bons moments autant au laboratoire qu'en dehors, Alice, Angel, Clément, Felipe, Myriam, Benjamin, Florence, Virginie, Quentin, Helias, Violette, Elsa, Diane, Hélène, Isabelle, Anaïs.

Merci à toutes les personnes que j'ai eu la chance de côtoyer au LBE, Annie, Sylvie et Alexandra à l'accueil qui m'ont toujours aidé avec gentillesse. Merci aussi aux personnes que j'ai moins eu l'occasion de connaitre mais avec qui les moments passés étaient excellents : Kim, Nicolas, Eric, Renaud, Nori, Audrey, Fréderic, Hélène, Jérôme, Dominique, Valérie, Denis, Nadine, Véronique.

Merci à Fabrice Béline et Marie-Noëlle Pons pour leur travail de rapporteur sur cette thèse. Merci à eux aussi, ainsi qu'à Patricia Camacho, Véronique Bellon-Maurel et Marc Heran pour avoir accepté d'être membres du jury et leurs précieuses remarques durant la soutenance de thèse.

Merci à ma famille et plus particulièrement mes parents pour m'avoir soutenu et encouragé au long de mes études tout en me laissant libre de mes choix. Et enfin, un grand merci à ma compagne avec qui je partage joies et peines depuis maintenant quelques années, et je l'espère pour beaucoup d'autres. Merci de m'avoir accompagné, soutenu et même rejoint (presque ^^). Merci pour tes précieux conseils et corrections, toi à qui j'ai fait subir tous mes premiers brouillons de diverses communications. Je te souhaite toute la réussite possible, tu le mérites et en es capable. Maintenant que c'est écrit tu ne peux plus le nier.

« Combien de belles épopées bactériennes, combien de drames, combien de bonheurs bactériens demeureront à jamais ignorés de nous, plus récents occupants de la croûte terrestre...

Dans le cœur de tout homme, il y a une bactérie qui sommeille.»

Encyclopédie du savoir relatif et absolu, Bernard Werber.

Chapter 1. Anaerobic Digestion in Europe and in France.

Chapter 1: Anaerobic Digestion in Europe and in France.

Chapter 1. Anaerobic Digestion in Europe and in France.

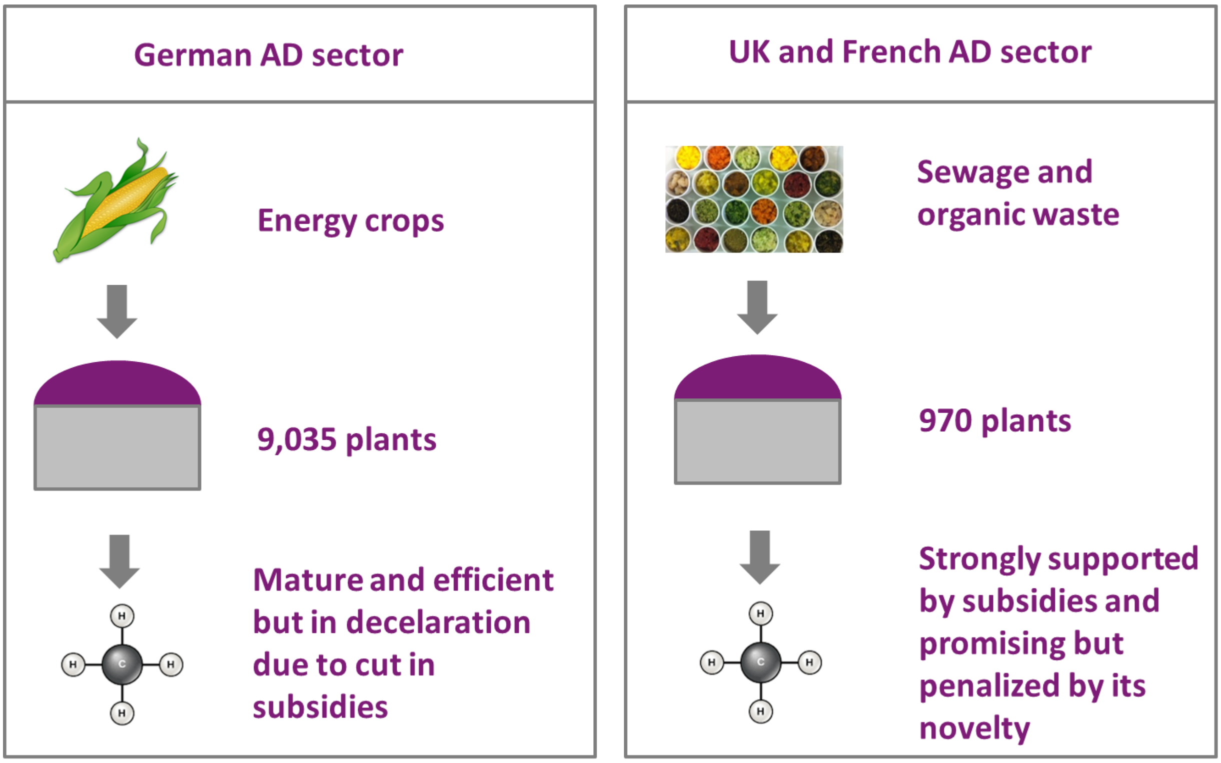
Chapter 1. Anaerobic Digestion in Europe and in France. 1
1.1 Chapter Guidelines 3
1.2 Technical and Graphical Abstracts 5
1.3 World Energy Challenge 7
1.4 Biogas over European Union 7
1.5 France, a Potential Vector of Growth for the European Anaerobic Digestion Sector 10

1.1 Chapter Guidelines

This chapter provides the context to the Anaerobic digestion (AD) sector in Europe. Based on world energy report, AD European report and French current AD sector status assessment, it explains why this PhD focuses on developing solutions to monitor the co-digestion of solid waste.

1.2 Technical and Graphical Abstracts

Positive change about global warming occurred in 2014 when worldwide economy and CO_2 emissions increase appeared to be decoupled for the first time in 40 years. It is a positive signal on the development of low carbon energy supply, even if nonrenewable energy such as nuclear energy represents more than 63% of them. Among the renewable energy in operation, anaerobic digestion is of particular value in Europe for its ability to produce both fertilizer and energy from waste. However, the sector, led by Germany far ahead with 9,035 of the 14,572 plants installed in Europe, is penalized by the deceleration of the German biogas industry. Indeed, the German model has been criticized by the European commission which reported environmental issues from the use of energy crops for biogas production. Thus, forerunner countries such as France or United Kingdom producing biogas from sewage, organic waste and on landfills are without any doubt the most promising alternatives for AD development. The quantity of available resources promises a great future to this alternative model. Nonetheless, its novelty and diversity penalize the first plants profitability which suffers from immature technologies, unadapted plants monitoring and under-exploitation.


German AD sector
Energy crops
9,035 plants
Mature and efficient but in decelaration due to cut in subsidies
H
H
C
H
H
UK and French AD sector
Sewage and organic waste
970 plants
Strongly supported by subsidies and promising but penalized by its novelty
H
H
C
H
H

Chapter 1: Anaerobic Digestion in Europe and in France.

1.3 *World Energy Challenge*

The global energy system is one of the main 21st century challenges. Until 2014, a continued increase in global greenhouse-gas emissions has been observed, correlated with the global economy development. A very positive change occurred in 2014 when global economy and CO_2 emissions increase appeared to be decoupled for the first time in 40 years. Indeed, that year, global energy-related CO_2 emissions stayed flat at 32.2 Gt while global economy increased by 3% (International Energy Agency, 2015).

Low-carbon energy supply plays an important role in decoupling economic growth to CO_2 emission. In 2014, new equipment capable of producing 128 GW were installed which represented an increase of 50% of the total low carbon energy supply (International Energy Agency, 2014). Deploying low-carbon technologies aims to save 37 Gt CO_2 over the next years until 2040. Nonetheless, 63% of the low carbon power came from nuclear power in 2014 and aims to be reduced to 50% in the 450 Scenario relative to the Bridge Scenario, even if nuclear power is facing an uncertain future (International Energy Agency, 2015). Renewable energy technologies must play a critical and rising role in the low-carbon energy supply. In 2014, wind power accounted for the largest share of growth in renewables-based generation (34%), followed by hydropower (30%) and solar technologies (18%). The agreement reached at the 21st Conference of the Parties (COP21) in Paris in 2015 could reverse current unsustainable global energy system if the levers available to accelerate research, development, demonstration and deployment in renewable energies are mobilized to give the right signals to innovators and financiers.

Since cities drive economic, energetic and population growth, sustainable actions in urban areas, such as urban waste valorization or rooftop solar panels, are the instruments of a transition to a sustainable global energy system. Among the promising developments, this thesis will focus on the production of energy from waste using Anaerobic Digestion (AD). By its own, wastewater treatment represents 2-4% percent of the energy consumption in developed countries. The ability of AD to turn waste treatment from an energy-consuming process into an energy-producing process could thus have a major impact on CO_2 emission reductions.

1.4 *Biogas over European Union*

The use of fertilizer is coupled with economic growth and is foreseen to increase in the next years (Figure 1). If the forecast for global supply increases faster than the global demand, West Europe will continue to depend on nitrogen and phosphate imports (FAO, 2015). Thanks to the current reliance on both natural gas and fertilizer imports of the European Union countries as well as the 20%

renewable energies target by 2020, anaerobic digestion is gaining in popularity. AD is indeed an important pillar of the European circular economy: it produces renewable methane and recycles nutrients in the form of organic fertilizers. In 2013, the European Union (EU) produced 13.4 million tons oil equivalent of biogas primary energy. The total installed electrical capacity from biogas reached 7,852 MW with 47.5 TWh of electricity delivered to the network and 48.7 TWh of thermal energy produced. The electricity produced from biogas in Europe is equivalent to seven nuclear reactors with the addition of thermal energy production and it covers the energy consumption of 5.4 million average European households (EBA, 2014). Additionally, if all the landfilled biodegradable waste were anaerobically digested, it would produce 400,000 tons of nitrogen, 120 tons of phosphorus and 450 tons of potassium (EBA, 2015) which represent respectively 12%, 8% and 32% of the fertilizer imports in West Europe in 2014.

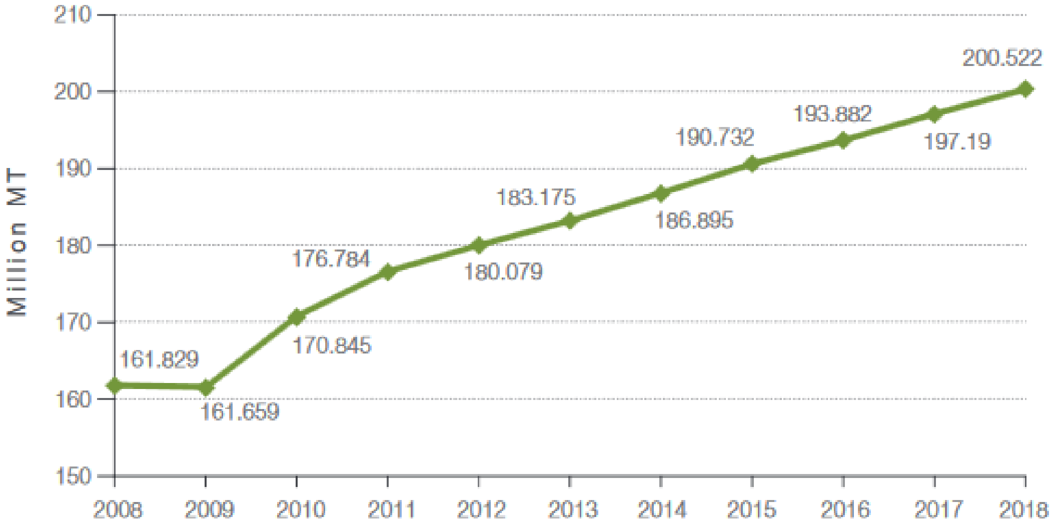


Figure 1: World global nutrients ($N+P_2O_5+K_2O$) consumption and forecast *(FAO, 2015).*

The European AD sector, composed of 14,572 plants, is led by Germany and Italy far ahead with 9,035 and 1,390 biogas units respectively, and then comes Switzerland at the third place with 620 biogas plants (Figure 3). Two third of the European plants are agriculture-based biogas plants that run on energy crops and agricultural residues. Agricultural based biogas plants are driven by Germany and Italy in which they represent more than 80% of the biogas plants. On July, 28th 2013, the European Commission published a working document on the sustainability of solid and gaseous biomass used for electricity, heating and cooling. In the section about biogas, the European commission reported environmental issues from the use of energy crops for biogas production. The report fostered the use of organic waste to reduce greenhouse gases emissions and the competition with food crops. These recommendations impacted the actual mainstream anaerobic digestion model and slowed down the sector. According to EurObserv'ER, growth was about 10.2% in biogas production and 5.5% in the plants installation in 2013 while in 2012 biogas production increased by 16.9% in biogas production and by 12% in biogas plant installation (EurObserv'ER, 2014). The deceleration is expected to continue. Indeed since August 2014, the German biogas industry is suffering from cut in the Feed-In-Tariffs (FIT) because of the reform of the Renewable Energy Act (EEG). The deceleration is foreseen to pursue in 2016 with cutting of the FIT for biogas plants above 100 kW in Germany (EBA, 2014).

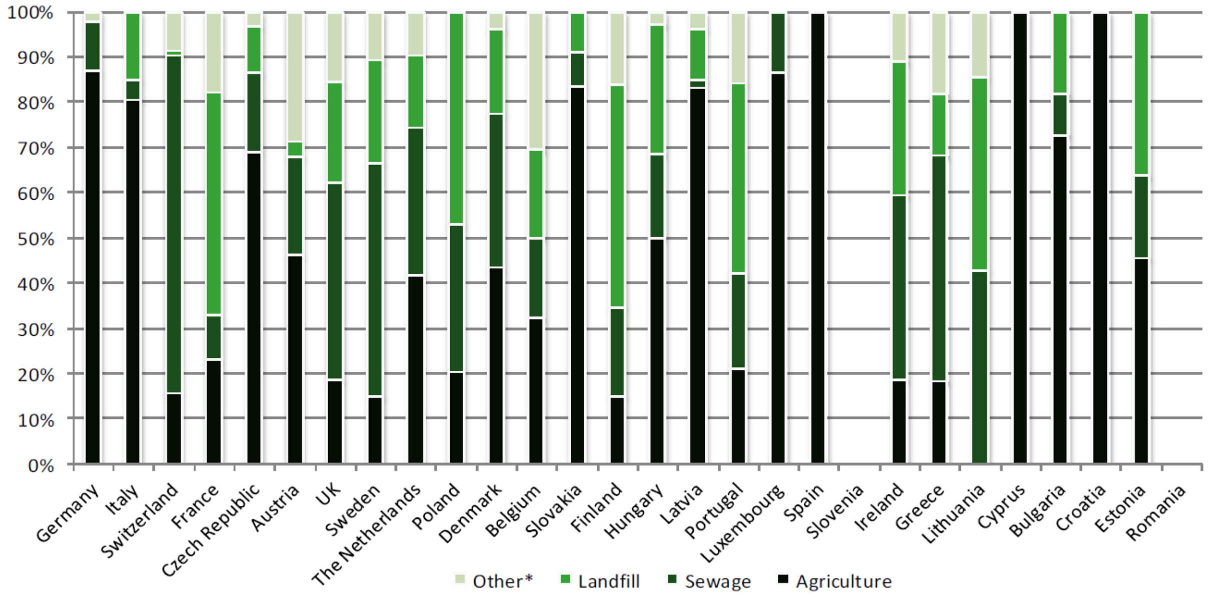


Figure 2: Percentage of biogas plants in European countries in 2013 by type *(EBA, 2014).*

While the global anaerobic digestion sector is penalized by the deceleration of the German biogas industry, other economic models gain in interest. In accordance with the European Commission recommendations, AD producing biogas from sewage, organic waste and on landfills grow in forerunners countries such as France, United Kingdom (UK), Poland and Finland (Figure 3). On October 14th 2014, the amendment 2102 of article 27 of the energy transition reinforced the will of France to develop a new AD model and prohibited the use of dedicated energy crops. The potentialities offered by AD from bio-waste are indeed enormous. In 2015, only a third of the bio-waste generated over the year across EU is valorized through anaerobic digestion and composting (BSAG, EBA, ECN, ESPP, 2015). The development of anaerobic digestion in circular economy could bring benefits for the EU balance of payments with an estimated household cost savings of 1.7–2.9 billion €/year for the EU (Ellen MacArthur Foundation / McKinsey), synergies with bio-energies and bio-fertilizers both reducing CO_2 emission. From this statement, indicators show that, after Germany, France could be the Europe's new vector for growth of AD.

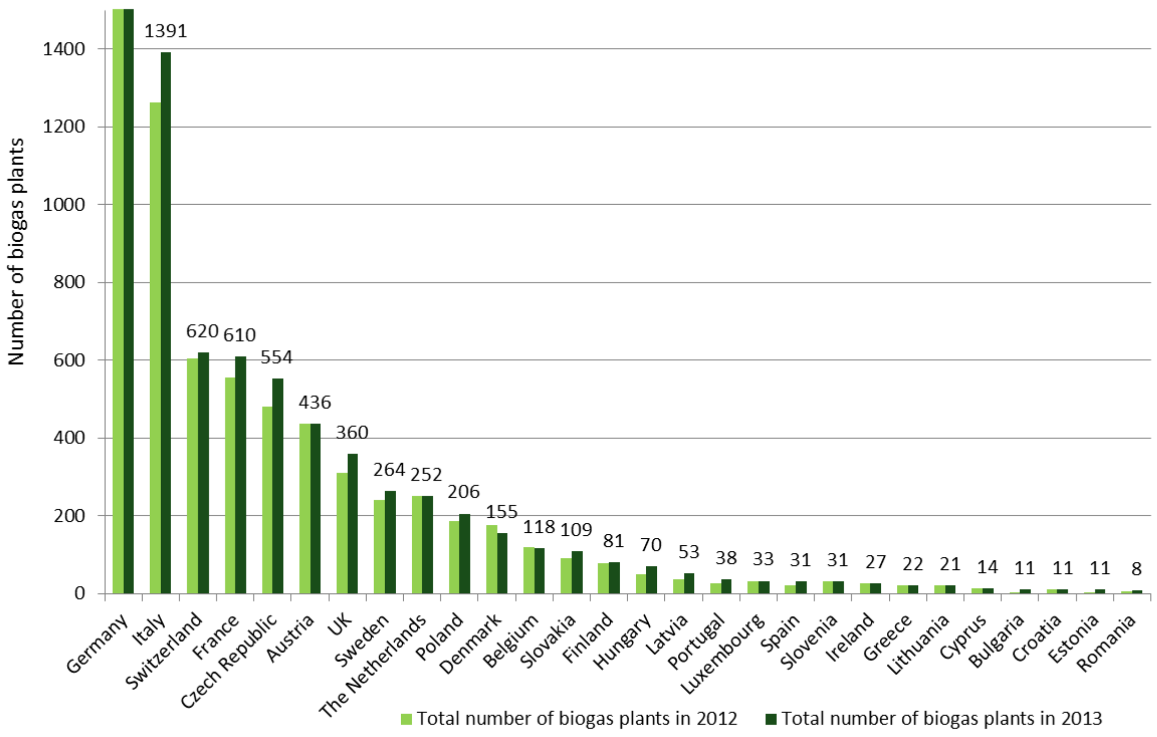


Figure 3: Number of biogas plant evolution in Europe in 2012 and 2013 *(EBA, 2014).*

1.5 France, a Potential Vector of Growth for the European Anaerobic Digestion Sector

According to EurObserv'ER, (2014), along with UK, France is without any doubt the most promising country for AD development. The French Observation and Statistics Office reported an increase in the biogas production by 17% in 2013 and by 13% in 2012. For 10 years, regulation and public aids have been in favor of AD development in France. Ecology Minister, Ségolène Royal, reaffirmed her support for AD with the project to launch 1 500 biogas plants between 2014 and 2017. Public supports focus especially on biogas plants treating farm waste with the governmental project "Plan EMAA" ("Energie Méthanisation Autonomie Azote") which aims to increase the number of agricultural biogas plants from 90 in 2012 to 1000 in 2020, with an average rate of 130 plants per year between 2013 and 2020. Biogas plants treating farm waste are strongly supported thanks to their ability to produce both biogas and fertilizers. Digestate, the byproduct of biogas, is indeed supposed to replace conventional fertilizers and reduce the environmental impact of agriculture.

France is the leader in Europe of the young organic waste AD sector. Since the deceleration of the market in their own country, German companies are eager to valorize their strong know-how on AD of energy crops and to install biogas plants in the emerging French market. Besides, new companies commercialize very new processes thought to be adapted to the sector. The actual statement reveals a lack of knowledge when dealing with the large diversity of substrates used in France. The technologies developed on energy crops indeed turns out to be unadapted to agricultural waste and appropriate solutions are slow in coming.

Although only a third of waste resources are today exploited, 40% of the biogas plants reported difficulties in substrates supply to feed the plants in 2015 (E-cube, 2015). These supply difficulties are mainly due to the lack of available resources inventory, many resources localization remaining unknown to the plant operators. Furthermore, the biogas plant operators are suspicious of accepting new substrate and changing the feeding strategy. Only few data about the resources are available and feeding poorly characterized substrates can lead to process inhibition or failure.

The treatment of agricultural waste such as manure, slurry and straw with technologies adapted from energy crops remains a problem. 94% of the 56 installations surveyed by E-Cube (E-cube, 2015) reported operational incidents and 70% of the plants reported a profitability below their expectation. Moreover, 44% of the units reported a dysfunction of their AD plant and 31% a dysfunction in the co-generation unit. The main reason for dysfunction of the AD plant is a poor adaptation of the material to the substrates. The diversity of substrates used induces quick changes in the operating conditions of the plant that can induce inhibitions or failures. The difficulties to anticipate the biological changes are attributed to a lack of formation for the operators as well as the instrumentation not adapted to the plant. A frequent solution found by constructors is the oversizing of biogas plants, which increases the cost of the installations. The hydraulic retention time ranges from a few hours up to 68 days in industrial effluent treatment plants, from 6 to 40 days on wastewater treatment plants and from 20 to 200 days on farm waste treatment plants (ATEE-Club biogas, 2011). This under-optimization of the plants affects the profits of this sector.

Due to reactor-oversizing and the difficulty to duplicate the plant designs, initial investment in biogas units in France remains high and depends on the digested substrate. The mean values are about 4,100 €/kWe (kiloWatt electric) for biogas plants treating industrial effluent, 6,300 €/kWe for biogas plants treating farm waste and 7,000 €/kWe for collective biogas plants treating mixed organic waste or wastewater treatment plants. Moreover, the cost of installation is not proportional to the capacity of biogas production: for biogas plants treating farm waste, investments vary from 15,000 €/kWe for a capacity of 35 kWe to 5,600 €/kWe for a capacity of 500 kWe (Bastide, 2015). The type of substrate treated and the electrical capacity of the plant are difficult to decouple, mean electrical power is 200 kWe, 441 kWe, 870 Kwe and 2,050 KWe for biogas plants treating agricultural waste (pulled up by collective biogas plants), industrial effluent, wastewater and household waste respectively. The investment needed to produce 1 kWe for a 200 kWe biogas plant treating farm waste is around 50% more expensive than for a 440 Kwe biogas plant treating an industrial effluent. Operational expenditures (OPEX) vary with operational choices from 500 €/kWe to 1,500€/kWe and do not depend on the biogas production (E-cube, 2015). Since 75% of the biogas plant incomes are from electricity sales, biogas plants treating farm waste are penalized by the initial investment and compensate the loss of earning treating waste with treatment fees, FIT and premium for agricultural waste treatment which decrease with the increase of electrical power.

Government subsidies bring a financial support to biogas plant installations, targeted on the FIT and premiums for good practice, as reported in Table 1. Premiums are added to the FIT. In the case of "mixed" feedstock and biogas injection, the premium is weighted and calculated in proportion to the feedstock quantities used by the plant.

Table 1: Feed in Tariffs and premiums for biogas injection and electricity production in 2016 in France

		Feed in Tariffs (FIT)	Energy efficiency	Livestock processing	Hazardous waste treatment	Exclusively agricultural or agri-business waste	Exclusively household waste
Electricity production	min €/kWh	0.119	0	0	▨	▨	▨
	max €/kWh	0.1337	0.04	0.026	▨	▨	▨
Biomethane injection	min €/kWh	0.45	▨	▨	0.24	0.02	0
	max €/kWh	0.95	▨	▨		0.03	0.05

FIT and premiums are based on the kWh production cost and decrease with the biogas produced, encouraging the development of small biogas plants treating farm waste despite the high initial investment needed. Thanks to public aids, as shown in Figure 4, the rate of return after subvention is equivalent for each type of unit, but farm biogas plants profitability is today subventions dependent.

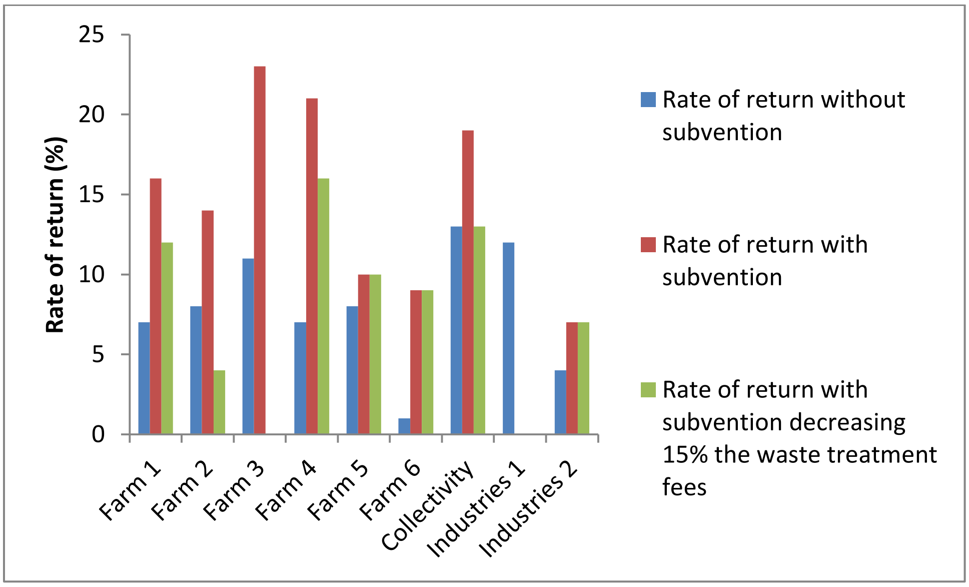


Figure 4: Biogas rate of return depending on the type of substrate they are digesting *(G.Bastide, 2014).*

To conclude, AD is probably the major biological process involved in waste valorization. AD is a microbial process that converts organic matter into biogas mainly composed of CO_2 and CH_4. It has a low energy requirement, low sludge production and the potentiality to produce both valuable intermediates and energy (Moletta, 2011).

French biogas sector is interesting for its will to digest organic waste and could represent a strong innovation place thanks to government subsidies for R&D and demonstration of industrial system. The quantity of available resources promises a great future for the sector but its novelty and diversity penalize its development. The first plants profitability suffers from immature technologies, unadapted plants monitoring and the under-exploitation of the plants. From the current study, four points of French AD model development are highlighted:

- Improvements of the plants design are required to be more cost-efficient and adapted to the AD of solid organic waste, reducing both investment and operational expenditure.
- Sensors must be adapted to estimate the biological state of AD plants digesting a great diversity of organic waste.
- Process monitoring must be adapted to the development of new supervision systems, taking into account the complexity of the AD biological reactions and a fine control of processes with feeding changes. It aims to secure the plant without plant oversizing and to optimize global performances and profitability
- Characterization and inventory of the available resources must be done. Biogas plants suffer already of feeding optimization difficulties while stress on available resources will increase with the number of biogas plants. Standard and fast methods of

characterization have to be developed to better anticipate the AD performances of the substrates.

1.6 Contents of the Thesis

This PhD will specifically focus on the three last points, mainly concerning biogas plant monitoring. AD technology has been successfully monitored for many years, see for example (Jimenez et al., 2015b). AD processes dealing with liquid effluent have been monitored using simple measurements such as biogas output flowrate and pH (Steyer et al., 1999) and AD technology applied on organic solid waste has been reported as mature (Mata-Alvarez et al., 2000).

In the current AD context of energy production detailed in this chapter, diversification of substrates and development of Anaerobic co-Digestion (AcoD) are needed in order to keep a highly performant and profitable process. The development of AcoD for the production of energy from organic waste, the growing number of installations, their diversity, their instability, the accumulation of inhibitors and the non-adapted process monitoring are real challenges for the long-term viability of the installations (Steyer et al. (2006), Chen et al., (2008)). Maintaining a highly performant process goes through an increase in the robustness, the performance and the monitoring of these installations.

Since the feeding strategy is the major factor governing the overall process, the development of AcoD requires defining optimal feeding strategies. Challenges are related to the variability and diversity of the available inputs and to the difficulty to forecast process performances (Karthikeyan and Visvanathan, 2013). Thus, for an appropriate process monitoring, it is fundamental to deeply characterize the inputs, to estimate the biological state of the process, and eventually to develop advanced dynamic models that enable to predict the performances of the process depending on its feeding strategy and its current biological state.

The state of art on anaerobic digestion monitoring, which will be develop in Chapter 3, Chapter 4 and Chapter 7, reveals that many sensors, methods and models have been developed other the years to optimize plant performances. However, the developed solutions are not fully adapted to frequent changes in the inputs. To answer the need of flexibility in the feeding strategy, new sensors, methods and models must be developed based on a fast characterization of the substrate, an estimation of the plant biological state and modelling co-digestion of the solid organic waste. Plants feeding strategy optimization will be studied through 3 axes:

- First axis is based on the estimation of the actual biological state of the reactor. While carbon dioxide and methane production analyses are matured, the sensors developed for liquid phase analysis are not adapted to solid organic waste co-digestion. Most of the cost efficient sensors estimating soluble compounds concentrations are based on titration methods but have a high risk of failure in case of frequent change in the inputs. The development of a sensor which would enable the estimation of main indicators of the biological state without calibration depending on the inputs is strongly needed.

- Second axis is based on the prediction of the substrate AD performances to optimize feeding strategy. Current laboratory methods are not standard and sometimes poorly adapted to solid organic waste characterization. Their accuracy and time of analysis penalize the substrate characterization and therefore feeding strategy optimization. Fast characterization of the solid organic waste based on spectroscopic analyses has been developed. While promising, the methods developed focused on BMP estimation. BMP is essential but not sufficient for feeding strategy optimization. Other parameters such as biochemical composition, COD and methane production kinetics are needed. Thus, studies on fast spectroscopic characterization must be pursued to provide a detailed characterization of the substrate.
- Third axis is based on the modelling of AD process. The complexity of AD makes the plant monitoring tricky for the human operator. First, the relation between the inputs characterization and the AD plant performances must be deeply understood. Second, since plant performances depend on both biological state and feeding strategy, the development of a dynamic model taking into account both actual biological state and substrate performances, is extremely relevant to help the operator reaching its optimal level of production. Actual models such as the anaerobic digestion model n°1 (ADM1) provide a detailed overview of the process and enable to simulate accurately the process performances. However, hydrolysis step of ADM1 model which has been developed to represent the AD of wastewater sludge is unsuitable for the modelling of solid waste AD. Since hydrolysis is often the limiting step of solid organic waste AD, it must be studied to provide an accurate representation of the co-digestion process.

After a section describing the materials and methods (Chapter 2) used during this PhD, these three axes are developed into 5 chapters as shown in Figure 5. First, Chapter 3 deals with the estimation of volatile fatty acids, inorganic carbon and ammoniacal nitrogen concentration in the plant to assess its biological state. The aim of this work is first to compare the accuracy of the estimations from Nordmann (1977), Ripley et al (1986), Purser et al (2014) and Zaher (2005) methods on 24 samples and second to develop an innovative sensor in order to improve the monitoring of AD plants. We decided to go further with the buffer capacity method since it is the most promising one to deal with additional component interferences as recommended by Zaher et al. (2003). To do so, we added a variable that could allow differentiating the component of a plant sample in an easier way. Thus, an electrical conductivity sensor was implemented in the titration vessel. The method was applied to 24 plant samples chosen to represent a wide range of AD process to assess the accuracy of prediction using both buffer capacity and buffer conductivity.

Chapter 4 described the relations between the biochemical composition of the organic matter and the AD performances. The aim of this chapter is to understand the existing correlations between carbohydrate, protein or lipid contents, COD, methane yield and the kinetics of methane production to facilitate plant monitoring with partial characterization of the substrates. For that purpose, the relations are highlighted using Principal Component Analysis (PCA). Then, Pearson's correlation test was carried out on these parameters measured for 222 fully characterized samples to deeply investigate the existing correlations. Finally, since the analysis of the biochemical composition is faster than methane production performances analysis, PLS regression was carried out to evaluate the feasibility of methane production kinetics prediction from the biochemical composition.

Chapter 5 is about solid organic waste characterization. This study is first focused on the development of repeatable and accurate analyses for the estimation of carbohydrate, protein, lipid contents and COD in a wide range of solid samples such as fruits, vegetables, manure, cereals, oils, fats, meat and fishes. Secondly, NIR analyses are performed on the same samples with the aim to correlate protein, carbohydrate, lipid contents and COD with NIR measurements. Eventually, the possibility to predict the Methane Production Rate (MPR) of solid substrate using NIR spectroscopy is discussed.

Chapter 6 focuses on the interpretation of the partial least square *b* coefficients associated to the biochemical and AD performances parameters predictions. Parallel study of *b* coefficients and near infrared wavelengths allows sampling relevant wavelengths involved in the estimation of these parameters. The relation between these wavelengths and the potential molecules related to these wavelengths is established and discussed. Eventually, cosine between the *b* coefficients of the different PLSr calibrations developed in Chapter 5 is estimated to highlights the relations between the models.

Chapter 7 is about dynamic modeling of the AD process by a modified ADM1. The first section will challenge the classical parallel concept of multi-substrates hydrolysis to provide a methodology applicable to different substrates. The second section of this study aims to reduce the difficulties of ADM1 parameters estimation. The composition of the substrate predicted by spectroscopic methods is used to implement ADM1. This approach reduces the ADM1 parameters estimation such as methane yield or hydrolysis kinetics from 60 days to a matter of minute, making it suitable for industrial application. Eventually, the accuracy of this approach has been estimated on two case studies.

Eventually, conclusion and outlooks of this thesis are discussed in Chapter 8.

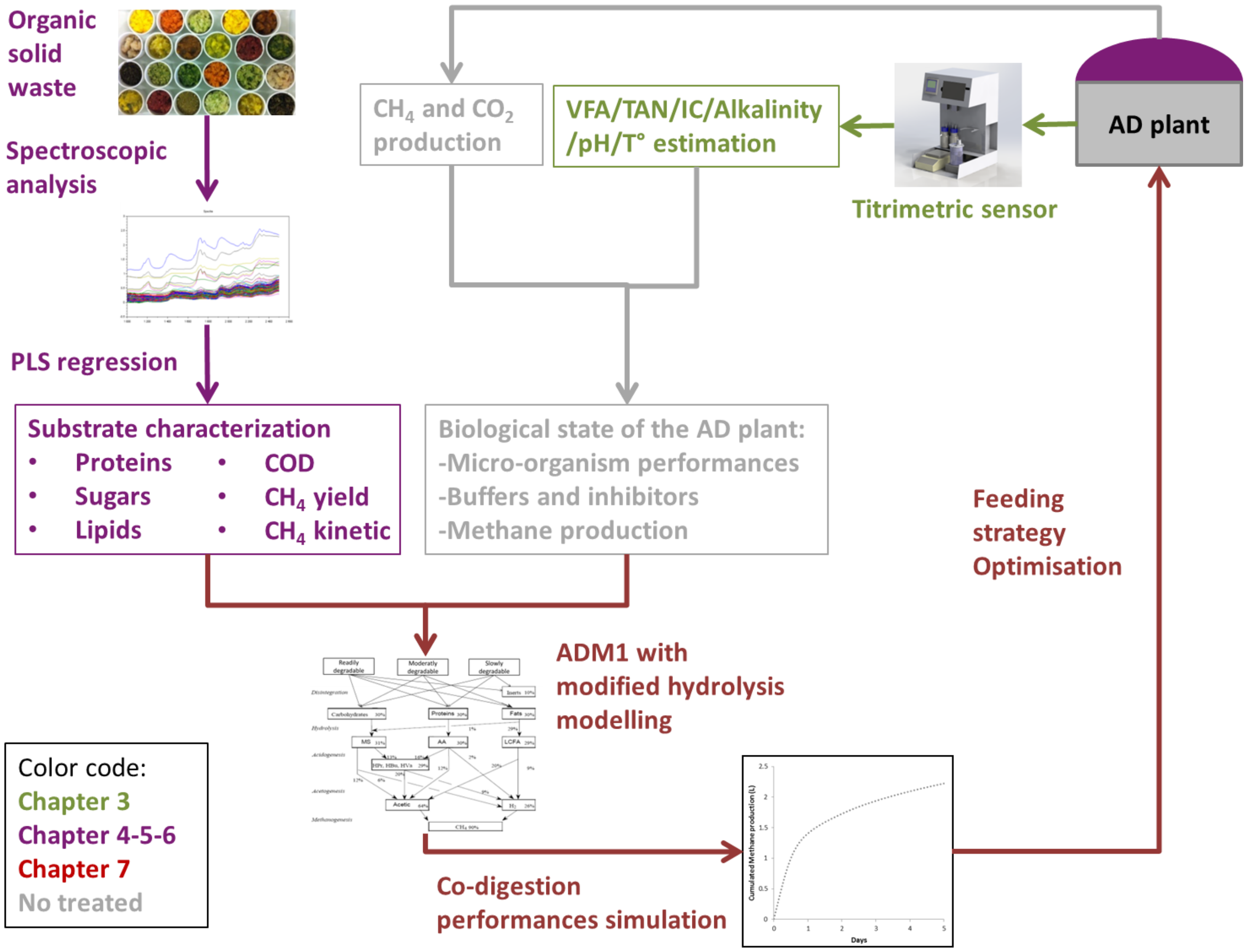

Organic solid waste
Spectroscopic analysis
PLS regression
Substrate characterization
• Proteins • COD
• Sugars • CH_4 yield
• Lipids • CH_4 kinetic
CH_4 and CO_2 production
VFA/TAN/IC/Alkalinity /pH/T° estimation
Titrimetric sensor
AD plant
Biological state of the AD plant:
-Micro-organism performances
-Buffers and inhibitors
-Methane production
Feeding strategy Optimisation
ADM1 with modified hydrolysis modelling
Co-digestion performances simulation
Color code:
Chapter 3
Chapter 4-5-6
Chapter 7
No treated

Figure 5: Feeding strategy optimization graphical abstract

Chapter 2. Materials and Methods

Chapter 2. Materials and Methods 19

2.1 Materials and Methods Guidelines 21
2.2 Titrimetric Methods to Determine Ammonia Nitrogen, Volatile Fatty Acids and Inorganic Carbon Concentrations 23
2.2.1 Experimental Titration Device 23
2.2.2 Titration Protocol 25
2.2.3 Analytical Methods for Reference Data 25
2.2.4 Titration Model 25
2.2.5 Signal Processing Coupling pH and Electrical Conductivity Measurements 30
2.3 Analytical Methods for Solid Waste Characterization 32
2.3.1 Biochemical Composition of Solid Waste 32
2.3.2 Bio-accessibility Assessment by Chemical Sequential Extraction 33
2.3.3 Anaerobic digestion experiments 34
2.4 Spectroscopic Analyses 35
2.4.1 Near Infrared 35
2.4.2 3D Fluorescence 36
2.5 Solid Waste Analyzed 36
2.5.1 Samples Analyzed for Fast Spectroscopic Characterization 36
2.5.2 Samples Analyzed for Correlation Analyses 39
2.6 Correlation Analysis and Regression Models 41
2.6.1 Pearson's Correlation and Cosine 41
2.6.2 Principal Component Analysis 41
2.6.3 Partial Least Square Regression Model and Performance 41
2.7 Dynamic Anaerobic Digestion Models 44
2.7.1 Dynamic Models: Modified ADM1 44
2.7.2 ADM1 Modelling Approach Using NIRs 45

2.1 Materials and Methods Guidelines

This chapter does not present results but details every materials and methods used during this PhD. The choice of a single "Materials and Methods" section avoids repetitions. For an easier readability, at the beginning of each following chapter, a reference to the specific Materials and Methods sections used is made. The first section, 2.2, focuses on titrimetric methods and liquid sample analyses, then analytical methods for solid waste are developed in section 2.3. Section 2.4 focuses on spectroscopic methods, section 2.5 details the different datasets used for correlation analyses and the calibration of spectroscopic methods and section 2.6 explains the mathematical tools used for correlation analyses and development of regression models. Eventually section 2.7 details the dynamic models used to predict AD performances.

2.2 Titrimetric Methods to Determine Ammonia Nitrogen, Volatile Fatty Acids and Inorganic Carbon Concentrations

2.2.1 Experimental Titration Device

Twenty four plant samples coming from diverse anaerobic digestion units were titrated, 9 from plants treating industrial effluents like vinasses or paper mill wastewater, 7 treating manure, sewage sludge and agri-food waste, 4 treating manure and cereal waste, 2 treating goat farm waste and 2 treating sewage sludge. Details are available in Table 2. All the plant listed were CSTR working in standard mesophilic conditions except the plant treating manure, sewage sludge and agri-food waste n°4 which was a dry batch process. The energy production of the plants varies from 0 to 1500 electrical kilowatts (kWe). Seven plants produce between 1000 and 1500 kWe, 4 plants produce between 500 and 1000 kWe and the other 13 plants produce less than 250 kWe. Dry matter content was in a range of 0.43-16.33 %. Hence, a protocol was designed depending on the dry matter content of the samples according to the protocol used for the titration of compost (Costello and Sullivan (2013)). If the dry matter content was below 10%, the titration was carried out with 2 L of digestate without pre-treatment. If the dry matter content was above 10%, the sample was diluted with distilled water to reach a maximum content of 10% and solubilized overnight; filtration (4 mm) was then used to remove the biggest particles with a colander. Titration was carried out in a 3 L stirred tank reactor (Applikon®, Delft, The Netherlands) with a working volume of 2 L. The reactor was equipped with a stirring system made of two Rushton impellers, stirring speed set to 530 rpm. Two sensors were connected to the reactor for measuring the pH and temperature (InPro 3253i/SG, Mettler Toledo®, Colombus, USA) and the conductivity and temperature (InPro 7100i/12/120, Mettler Toledo®, Colombus, USA). Both probes were connected to a computer for on-line data acquisition with one acquisition every 15 s for sensor and 30 s for actuator by a transmitter M300 (Mettler Toledo®, Colombus, USA). The analog to digital converter (ADC) was a TES module (Leroy Automation®, Toulouse, France). NaOH (1 M) and HCl (1 M) were added using two peristaltic pumps with a fix flow-rate of 4 $mL.min^{-1}$. A peristaltic air-pump was connected to the sparger with an equivalent water flow-rate of 0.7 $L.min^{-1}$. The pumps were controlled by the computer through on/off outputs. The gas output was connected to a liquid trap and a bottle of 500 mL containing 400 mL of 3 M NaOH with 0.002 % Thymolphtalein in order to ensure a constant pressure inside the reactor and the safety of the gas output in case of H_2S exhaustion.

Table 2: Reference analyses data of the samples used for the experiments. Volatile fatty acids (VFA) concentration was analyzed by GC, Inorganic Carbon (IC) by TOC and Total Ammoniacal Nitrogen (TAN) by TKN. Partial Alkalinity (PA) and Intermediate Alkalinity (IA) were estimated using the formula developed in Ripley et al (1986)

Substrate characteristics	Plant power (kWe)	VFA (mol.L^{-1})	IC (mol.L^{-1})	TAN (mol.L^{-1})	PA (mol.L^{-1})	IA (mol.L^{-1})
Manure, sewage sludge and agri-food waste 1	[1000-1500]	0.003	0.098	0.090	0.058	0.038
Manure, sewage sludge and agri-food waste 2	[1000-1500]	0.020	0.098	0.090	0.055	0.052
Manure, sewage sludge and agri-food wastes 3	[1000-1500]	0.031	0.105	0.052	0.040	0.039
Manure, sewage sludge and agri-food wastes 4	[1000-1500]	0.011	0.211	0.104	0.089	0.030
Manure, sewage sludge and agri-food wastes 5	<250	0.004	0.115	0.060	0.044	0.031
Manure, sewage sludge and agri-food wastes 6	[500-1000]	0.002	0.215	0.135	0.102	0.071
Manure, sewage sludge and agri-food wastes 7	<250	0.029	0.115	0.060	0.044	0.046
Manure and cereal waste 1	[500-1000]	0.001	0.066	0.033	0.031	0.081
Manure and cereal waste 2	<250	0.005	0.111	0.031	0.038	0.026
Manure and cereal waste 3	<250	0.039	0.111	0.059	0.038	0.045
Manure and cereal waste 4	<250	0.025	0.054	0.036	0.028	0.029
Goat farm waste 1	<250	0.002	0.007	0.004	0.003	0.0022
Goat farm waste 2	<250	0.025	0.020	0.012	0.009	0.018
Paper mill wasterwater 1	[500-1000]	0.000	0.038	0.076	0.016	0.009
Paper mill wasterwater 2	[500-1000]	0.000	0.038	0.000	0.018	0.006
Sewage sludge 1	<250	0.000	0.116	0.061	0.043	0.019
Sewage sludge 2	<250	0.000	0.116	0.061	0.037	0.028
Industrial effluent 1	[1000-1500]	0.001	0.132	0.034	0.064	0.011
Industrial effluent 2	[1000-1500]	0.001	0.132	0.034	0.066	0.001
Vinasses 1	<250	0.013	0.029	0.000	0.008	0.015
Vinasses 2	<250	0.013	0.029	0.000	0.007	0.014
Vinasses 3	<250	0.013	0.027	0.164	0.008	0.015
Vinasses 4	<250	0.007	0.013	0.078	0.005	0.007
Vinasses 5	<250	0.013	0.027	0.162	0.008	0.016

2.2.2 Titration Protocol

Dealing with samples from plants treating complex substrate, a working volume of 2 L was chosen to have representative samples of the plants in non-homogenous conditions. Once 2 L of plant sample is added to the reactor, a stirring is applied for 20 minutes to reach the thermodynamic equilibrium in the closed reactor. Then, the down-titration is carried out by adding continuously HCl (1 M) until the pump is automatically switched off when pH is below 2. Then, stripping of CO_2 lasts for 20 minutes switching the air pump on. Eventually, up-titration is carried out by adding continuously NaOH (1 M) until the pump is automatically switched off when pH is higher than 11.5. The flasks containing NaOH and HCl are weighted to check and correct the flowrate of acid and base pumps.

2.2.3 Analytical Methods for Reference Data

VFA, IC and TAN concentrations were measured in the liquid fraction and after dilution in case of dry matter content up to 10 %. TAN was measured using TKN Buchi AutoKjeldahl Unit K-370 (Buchi®, Flawil, Switzerland), Kjeldhal method without acid hydrolysis. To analyze VFA and IC, the liquid fraction is filtrated to 0.45 µm with cellulose-acetate filter (Sartorius®, Goettingen, Germany). VFA were analyzed with a gas chromatograph Perkin Clarus 580 (PerkinElmer®, Waltham, USA) with an Elite-FFAP crossbond® carbowax® 15 m column connected to a FID detector at 280 °C and nitrogen at 6 $mL.min^{-1}$ as gas carrier. IC was measured using TOC-Vscn, total organic carbon analyzer, (Shimadzu®, Kyoto, Japan), liquid Inorganic Carbon method.

2.2.4 Titration Model

A first check was applied to the experimental data collected from the computer. The synchronization of the sensors and actuators was verified. Then, a smooth spline function was done using R® 3.03 to remove noise and to compute first and second derivatives. The pre-treated data were then exported to Scilab® 5.5.0. Four models have been implemented to check the hypothesis and the accuracy based on Nordmann, (1977); Ripley et al., (1986); Purser et al., (2014); Zaher, (2005). The five points method or the eight points method developed in Moosbrugger et al., (1993) and Lahav and Loewenthal (2000) were not implemented because they require both the concentrations in sulphide and phosphorous to be accurate, which is not compatible with an easy off-line application. The *leastsq* function was used to solve nonlinear problem with Scilab® 5.5.0. It minimizes the square value of a function which is usually the difference between experimental and modelled data.

2.2.4.1 Existing Model based on linear resolution

Nordmann (1977) proposed to carry out a titration down to pH=5 and then to pH=4.4 and hypothesized that the whole IC is transformed into H_2CO_3 at pH=5 and that 30% of VFA are titrated between pH=5 and pH=4.4. The formula of Nordmann (1977) was first implemented as:

$$IC = \frac{[HCl]*V_{HCl-pH5}}{V_S} \qquad VFA = \frac{10}{3} * \frac{[HCl]*V_{HCl-pH5-4.4}}{V_S}$$

(1)

IC and VFA concentration are expressed in mol.L^{-1}, $V_{HCl-pH5}$ is the needed volume of HCl to reach pH 5, $V_{HCl-pH5-4.4}$ is the added volume of HCl between pH 5 and 4.4 and $[HCl]$ is the concentration of HCl (1 M).

Ripley et al (1986) worked with the notion of alkalinity. A first titration down to pH=5.75 provides the partial alkalinity (PA) while a second titration from pH=5.75 to pH=4.3 provides the Intermediate Alkalinity (IA). These authors assumed that PA relates to IC and IA approximates VFA and that the rate PA/IA provides a good indicator to monitor a plant. Jenkins et Morgan (1983) refined this method and assumed that IC contributes to 80% of PA and VFA about 65% of IA. Ripley et al. (1986) formula was implemented using the factor converting PA in IC and IA in VFA proposed by Jenkins et Morgan., (1983), respectively $\frac{10}{8}$ and $\frac{10}{6.5}$ as:

$$IC = \frac{10}{8} * \frac{[HCl]*V_{HCl-pH5.75}}{V_S} \qquad VFA = \frac{10}{6.5} * \frac{[HCl]*V_{HCl-pH5.75-4.3}}{V_S} \qquad (2)$$

Purser et al (2014) recalibrated the equations of Nordmann (1977) and Ripley et al (1986) with a corrective factor for the VFA concentration estimation. The calibrations were done on animal slurry and energetic crops, and a second on food waste for Nordmann (1977) formula. These three recalibrations were implemented (equation 3.1-3.2-3.3). The obtained estimations of VFA were expressed in mg.L^{-1}, thus, it was divided by 1000 and by the molar mass of acetic acid in order to have it in acetic acid equivalent (mol.L^{-1}). This modification does not change the results since acetate was the main component of VFA in our samples.

The formulas proposed by Purser et al (2014) with the estimation expressed in mol.L^{-1} are:

- Nordmann (1977) calibration for animal slurry and energetic crops:

$$VFA = \left[-340 - 7160 * \frac{[HCl]*V_{HCl-pH5}}{V_S} + 124540 * \frac{[HCl]*V_{HCl-pH5-4.4}}{V_S}\right]/60000 \tag{3.1}$$

- Nordmann (1977) calibration for food waste:

$$VFA = \left[-523 - 5920 * \frac{[HCl]*V_{HCl-pH5}}{V_S} + 131110 * \frac{[HCl]*V_{HCl-pH5-4.4}}{V_S}\right]/60000 \tag{3.2}$$

- Ripley et al (1986) calibration for animal slurry and energetic crops:

$$VFA = \left[-767 - 1197 * \frac{[HCl]*V_{HCl-pH5.75}}{V_S} + 54460 * \frac{[HCl]*V_{HCl-pH5.75-4.3}}{V_S}\right]/60000 \tag{3.3}$$

2.2.4.2 Existing model based on nonlinear resolution

The model developed by Van Vooren (2000) and improved by Zaher (2005) was implemented. This method is based on the buffer capacity defined as $\beta = \frac{dCt}{dpH}$ where *Ct* is the concentration of the associated counter-ions in solution used for the titration. A model simulates the buffer capacity during the titration based on this equation, assuming that the volume of titration is constant. In order to simplify the model, the compounds were supposed to be diluted enough to consider that the activity of the compounds can be estimated by their concentration:

$$\frac{d[H_3O^+]}{dpH} = \frac{d[H_3O^+]}{dlog([H_3O^+])} = \frac{d[H_3O^+]}{-\frac{d[H_3O^+]}{\ln(10)[H_3O^+]}} = -\ln(10)[H_3O^+] \tag{4.1}$$

$$\frac{d[\mathrm{OH^-}]}{dpH} = \frac{\mathrm{K_e}*\ln(10)}{[\mathrm{H_3O^+}]} \tag{4.2}$$

$$\frac{d[NH_4^+]}{dpH} = -\frac{N_t*[\mathrm{H_3O^+}]*\ln(10)*\mathrm{K_N}}{(\mathrm{K_N} + [\mathrm{H_3O^+}])^2} \tag{4.3}$$

$$\frac{d[\mathrm{VFA^-}]}{dpH} = \frac{\mathrm{VFA_t}*\mathrm{K_{AGV}}*\ln(10)*[H_3O^+]}{(\mathrm{K_{AGV}} + \mathrm{H_3O^+})^2} \tag{4.4}$$

$$\frac{d[\mathrm{HCO_3^-}]}{dpH} = \frac{[\mathrm{H_3O^+}]*\ln(10)*C_t*\mathrm{K_C}*([\mathrm{H_3O^+}]^2-\mathrm{K_{C2}K_C})}{(\mathrm{K_C}*[\mathrm{H_3O^+}]+\mathrm{K_{C2}}*\mathrm{K_C} + [\mathrm{H_3O^+}]^2)^2} \tag{4.5}$$

where VFAt, Ct and Nt represent respectively the total amount of VFA, C and N in solution. K_N is the equilibrium constant of NH_4^+/NH_3 ($10^{-9.24}$ at 25°C), K_{VFA} the one of acetate/acetic acid ($10^{-4.76}$ at 25°C), K_C is for HCO_3^-/H_2CO_3 and K_{C2} for HCO_3^-/CO_3^{2-} (respectively $10^{-6.35}$ and $10^{-10.33}$ at 25°C). Zaher (2005) used this equation for phosphorous and lactate as well. In order to fit every experimental data, blind buffers were also introduced based on the equation of their acid form (*i.e.*, VFA if mono-acid and IC if di-acid). Eventually, using the electro-neutrality concept, the buffer capacity was modeled knowing the concentration of components:

$$\beta = \sum \gamma * \frac{dX}{dpH} \tag{5}$$

where X represents the components in solution and γ their valence. Thus, estimations of VFA, IC, TAN, lactate and phosphorous concentrations were obtained by minimizing the sum squared error between the model and the experimental data by nonlinear solver. In the case of a simplified equation taking into account only VFA, IC and TAN, equation 5 could be written for the up titration as:

$$\beta = \frac{d\mathrm{HCO_3^-}}{dpH} + 2 * \frac{d\mathrm{CO_3^{2-}}}{dpH} + \frac{d\mathrm{VFA^-}}{dpH} - \frac{dNH_4^+}{dpH} + \frac{d\mathrm{OH^-}}{dpH} - \frac{dH_3O^+}{dpH}$$

2.2.4.3 SNAC model based on nonlinear resolution

Some modifications were done to the model developed by Van Vooren (2000) and Zaher (2005). First of all, we assumed a thermodynamic equilibrium between gaseous and liquid phase for TAN and IC. Supposing a species A with its basic form A^- and acid form AH in equilibrium with the gaseous phase according to Henry's law, the following equation is obtained:

$$\mathrm{A_t} = \begin{cases} \mathrm{A_{gaz}} = \frac{\mathrm{AH}}{\mathrm{K_H}} \\ \mathrm{A_{soluble}} = \mathrm{A^-} + \mathrm{AH} \end{cases} \Rightarrow \quad \mathrm{A_t} = \frac{\mathrm{AH}}{\mathrm{K_H}} + \mathrm{A^-} + \mathrm{AH} \tag{6}$$

Using a liquid/gas phase equation in a dynamic model allows us to take into account the stripping of CO_2 all along the titration, even during the equilibrium phase if the experimental conditions are designed to reduce the transfer in order to reach quickly the liquid/gas equilibrium. The solution retained in this study was to use a closed reactor. Models not taking the liquid/gas equilibrium into account tend to under-estimate the concentration of volatile species and usually solve it by a quick titration (Lahav and Morgan, 2004). Then, the possible variation of volume due to the addition of a titrant volume ($\mathrm{V}_{titrant}$) during the titration was taken into account. n_{ion} is the total amount of counter ion during titration expressed in mole, $[ion]_{solution}$ is its concentration in the solution composed of the initial solution plus the added titrant volume, $[ion]_{titrant}$ is the molarity of the titrant solution.

$$\begin{cases} n_{\text{ion}} = (\mathrm{V}_{solution-initial} + \mathrm{V}_{titrant}) * [ion]_{solution} \\ n_{\text{ion}} = [ion]_{titrant} * \mathrm{V}_{titrant} \end{cases}$$

$$\Rightarrow \mathrm{V}_{titrant} = \frac{\mathrm{V}_{solution-initial} * [ion]_{solution}}{[ion]_{titrant} - [ion]_{solution}} \tag{7}$$

The concentration of the titrating counter ion can be defined by the equation 5 with the electro-neutrality concept. The modification of the volume during the titration induces the dilution of the concentration of non-reactive ions during the titration. Thus, we introduced Z^+ and Z^- respectively the inert ions with a positive and a negative charge, for instance respectively K^+ and Cl^-. Z^+ and Z^- are estimated at the initial condition as the residual conductivity once the known component are subtracted assuming that Z^+ balance the charge of negative ions and Z^- the one of positive ions. An estimation of the molar conductivity of Z^+ and Z^- had to be made. Z^+ was supposed to be mainly composed of Na^+/ K^+ and Z^- of Cl^-. This estimation was done because it allows us to obtain an estimated initial titrant volume of 0 estimating Z^+ and Z^- on the initial conductivity. It indicates that the estimation was correct. The conductivity of the solution (σ) was modeled depending on the concentration of the component listed below and their molar conductivity (λ_x) as:

$$\sigma = \lambda_{\mathrm{OH-}} * [\mathrm{OH}^-] + \lambda_{\mathrm{H3O+}} * [\mathrm{H_3O}^+] + \Big[\lambda_{\mathrm{HCO3-}} * [\mathrm{HCO}_3^-] + \lambda_{\mathrm{CO32-}} * [\mathrm{CO}_3^{2-}] + \lambda_{\mathrm{Acetate}} * [\mathrm{VFA}^-] + \lambda_{\mathrm{NH4+}} * [\mathrm{NH}_4^+] + \lambda_{\mathrm{Cl-}} * [Z^-] + \frac{(\lambda_{\mathrm{Na+}} + \lambda_{\mathrm{K+}})}{2} * [Z^+]\Big] * \frac{\mathrm{V}_{solution-initial}}{\mathrm{V}_{solution-initial} + \mathrm{V}_{titrant}} + \lambda_{cournterions} * [ion]_{titrant} * \frac{\mathrm{V}_{titrant}}{\mathrm{V}_{solution-initial} + \mathrm{V}_{titrant}} \tag{8}$$

The molar conductivity was expressed in mS.m^{-2}.mol^{-1} and set to $\lambda_{\mathrm{H3O+}}$= 34.96; $\lambda_{\mathrm{OH-}}$ = 19.8; $\lambda_{\mathrm{Na+}}$= 5.01; $\lambda_{\mathrm{Cl-}}$ = 7.63; $\lambda_{\mathrm{NH4+}}$ = 7.35; $\lambda_{\mathrm{K+}}$ = 7.35; $\lambda_{\mathrm{HCO3-}}$ = 4.45; $\lambda_{\mathrm{CO32-}}$ = 13.86; $\lambda_{\mathrm{Acetate}}$ = 4.09 (Lide, (2003)). The molar conductivity of the compounds is well known at infinite dilution. At concentration above 10^{-3} mol.L^{-1} the molar conductivity will start to be affected by the concentration of the compounds.

The buffer capacity was calculated using the same principle as Van Vooren (2000) but taking into account that the volume of the solution and the concentration of TAN, IC and VFA change during the titration. The following formula is for instance obtained during the down titration using HCl as titrant solution.

$$\beta = \left(\frac{d[NH_4^+]}{dpH} - \frac{d[\mathrm{HCO_3^-}]}{dpH} - 2 * \frac{d[\mathrm{CO_3^{2-}}]}{dpH} - \frac{d[\mathrm{AGV^-}]}{dpH}\right) * \frac{\mathrm{V}_{solution}}{\mathrm{V}_{solution} + \mathrm{V}_{titrant}} + \frac{d\frac{\mathrm{V}_{solution}}{\mathrm{V}_{solution} + \mathrm{V}_{titrant}}}{dpH} * \\ ([\mathrm{NH_4^+}] + [Z^+] - [Z^-] - [\mathrm{HCO_3^-}] - 2 * [\mathrm{CO_3^{2-}}] - [\mathrm{AGV^-}]) - \frac{d[\mathrm{OH^-}]}{dpH} + \frac{d[H_3O^+]}{dpH} \tag{9}$$

The concept of "conductivity buffering" was introduced and defined as the derivative of the conductivity by the pH and defined theoretically as:

$$\frac{\mathrm{d}\sigma}{dpH} \\ = \frac{d\left(\frac{\mathrm{V}_{solution-initial}}{\mathrm{V}_{solution-initial} + \mathrm{V}_{titrant}} (\lambda_{\mathrm{HCO3-}}[\mathrm{HCO_3^-}] + \lambda_{\mathrm{CO32-}}[\mathrm{CO_3^{2-}}] + \lambda_{\mathrm{Acetate}}[\mathrm{AGV^-}] + \lambda_{\mathrm{NH4+}}[\mathrm{NH_4^+}] + \lambda_{\mathrm{Cl-}}[Z^-] + \lambda_{\mathrm{Na+/K+}}[Z^+])\right)}{dpH} \\ + \lambda_{\mathrm{OH-}} * \frac{d[\mathrm{OH^-}]}{dpH} + \lambda_{\mathrm{H3O+}} * \frac{[\mathrm{H_3O^+}]}{dpH} + \lambda_{cournterions} * \beta$$

$$= \frac{d\frac{\mathrm{V}_{solution-initial}}{\mathrm{V}_{solution-initial} + \mathrm{V}_{titrant}}}{dpH} * (\lambda_{\mathrm{HCO3-}}[\mathrm{HCO_3^-}] + \lambda_{\mathrm{CO32-}}[\mathrm{CO_3^{2-}}] + \lambda_{\mathrm{Acetate}}[\mathrm{AGV^-}] + \lambda_{\mathrm{NH4+}}[\mathrm{NH_4^+}] + \\ \lambda_{\mathrm{Cl-}}[Z^-] + \lambda_{\mathrm{Na+/K+}}[Z^+]) + \frac{\mathrm{V}_{solution-initial}}{\mathrm{V}_{solution-initial} + \mathrm{V}_{titrant}} * \left[\lambda_{\mathrm{HCO3-}} * \frac{d[\mathrm{HCO_3^-}]}{dpH} + \lambda_{\mathrm{CO32-}} * \frac{d[\mathrm{CO_3^{2-}}]}{dpH} + \lambda_{\mathrm{Acetate}} * \right. \\ \left. \frac{d[\mathrm{AGV^-}]}{dpH} + \lambda_{\mathrm{NH4+}} * \frac{d[NH_4^+]}{dpH}\right] + \lambda_{\mathrm{OH-}} * \frac{d[\mathrm{OH^-}]}{dpH} + \lambda_{\mathrm{H3O+}} * \frac{d[H_3O^+]}{dpH} + \lambda_{cournterions} * \beta \tag{10}$$

2.2.5 *Signal Processing Coupling pH and Electrical Conductivity Measurements*

According to the experimental data, we defined four functions $\mathrm{V}_{titrant} = f(pH), \sigma = f(pH), \beta = f(pH), \frac{d\sigma}{dpH} = f(pH)$. Working with $\frac{d\sigma}{dpH}$, a possibility to correct the pH shift between experimental data and the model has been found. This shift is due to a loss of accuracy of the pH sensor at extreme pH values.

2.2.5.1 *Experimental Error Correction Improving the Accuracy of the Estimation*

The first step of the analysis was to correct pH shifts between the model and the experimental data. The pH discrepancy is caused by measurement errors at extreme values. This error is difficult to predict and varies in each titration. The shift between experimental and theoretical data is calculated around pH=2, where only water impacts the curve. This shift is calculated on $\frac{d\sigma}{dpH}$ because it keeps the same value between pH 2 and 2.5 which means that experimental and theoretical data are represented by two parallel lines in this range. Then, the calculated shift is corrected by a translation of the experimental data. The pKa of VFA, IC and TAN are found by analyzing the second derivative of $\mathrm{V}_{titrant} = f(pH)$ or $\sigma = f(pH)$ depending if the estimation is done either on the buffer capacity or the conductivity buffering. For more accuracy, the pKa are fixed in a range of ±0.4 pH unit of their theoretical values as recommend by Van Vooren (2000) for down titration and 0.3 for up titration.

2.2.5.2 Coupling the Measurement of pH and Electrical Conductivity

Six samples from vinasses digester, which have the advantage to be already well characterized in their composition, were first used to decide which variables could be used to obtain the most accurate estimation of VFA, TAN, and IC. As expected from Van Vooren (2000), estimation using the derivative functions gave more accurate results than using non-derivate functions. Nonetheless, these signals are also more sensitive to noise; thus, a cubic interpolation of the pH and electrical conductivity signals was performed in order to smooth them (see 2.2.4). Then, we focused on $\beta = f(pH)$ and $\frac{d\sigma}{dpH} = f(pH)$. These two functions have many interesting features; one of them is that they have local extremums at each component pKa. Thus, using their second derivative, the pKa of the components present in the solution were found. The conductivity gives another advantage: VFA and TAN have opposite responses. Indeed, a loss of charge during the titration decreases the electrical conductivity when an accumulation of charge increases it. So the up titration of NH_4^+ to NH_3 induces a decrease of the electrical conductivity when the up titration of the acid form of VFA (VFAH) to the basic form of VFA (VFA^-) increases it. Supposing a salt composed of the compound E^+ and D^-, the dissociation of E-D during down titration increases the electrical conductivity while the titration of acids decreases it. Thus, it becomes easier to distinguish the components present in the sample. Despite its interest, conductivity buffering presents many signals during the down titration which are not taken into account by the model. These signals could be due to hydrolysis and cell destruction which disturb the response of electrical conductivity by releasing ions in solution.

Thus, buffer capacity was chosen as the best variable to estimate VFA and IC concentration during down titration as assumed by Van Vooren (2000) but taking into account the variation of volume during the titration. Since IC represent the main component of down titration, it was not necessary to implement any other components than IC and VFA to fit the experimental data. Zaher (2005) demonstrated that the estimation of VFA concentration is more accurate on up titration than on down titration. In accordance with the results obtained by Zaher (2005), our first analysis validates that up titration is more accurate to estimate VFA concentration since IC is removed at the end of the down titration. Indeed, the pKa of IC and VFA are near each other hence, IC disturbs the estimation of VFA.

2.3 Analytical Methods for Solid Waste Characterization

2.3.1 Biochemical Composition of Solid Waste

2.3.1.1 Carbohydrate Measurements

The protocol described by Dubois et al., (1956) was used because it has been reported as the most accurate and repeatable protocol for carbohydrate content determination (Jimenez, 2012). Two different methods were carried out on solid waste. Using the first methods, freeze-dried and grinded samples were suspended in distilled water at 1 $g.L^{-1}$ as recommended by Buffiere et al., (2008). Using the second methods, 0.050 g of freeze-dried and grinded samples were suspended in a shaker for 6 h in 5 mL sulfuric acid (98% mass) instead of water. Subsequently, for both of them, 1 mL of the solution was added to 1 mL of phenol at 5% in a hemolysis tube. After vortexing, 5 mL of sulphuric acid (98% mass) were added. After 10 minutes, the tubes were closed and vortexed. Absorbance of the samples was measured at 490 nm with a spectrophotometer at least 30 minutes after closing the tubes. Results are expressed in glucose equivalents, and converted to COD equivalents by multiplying the glucose equivalent by 1065. The COD of glucose is expressed in $mgO_2.gTS^{-1}$.

2.3.1.2 Protein Measurements

Two different methods have been tested: total Kjeldahl nitrogen (TKN) and Folin reagent. The TKN method analyzed the nitrogen content of the samples which was then converted into protein using an average ratio of 6.25 (Izhaki, 1993)(Girault et al., 2012) while the Folin reagent method expressed a protein concentration in BSA equivalents.

Folin reagent method: the applied protocol was described by Lowry et al., (1951). Bovine serum albumin (BSA) made from 0 to 100 $mgBSA.L^{-1}$ was used for the calibration. The method was adapted to deal with solid waste; freeze-dried and grinded samples were suspended in molecular water at 1 $g.L^{-1}$. 0.5 mL of this solution was then added to a hemolysis tube containing 2.5 mL of a solution comprising 50 mL of sodium bicarbonate $NaCO_3$ at 2% with NaOH (0.1N), 1 mL of copper sulfate $CuSO_4$ solution at 1% and sodium and potassium tartrate $C_4H_4KNaO_6$. After 10 minutes, 0.25 mL of Folin solution (commercial solution) was added. After vortexing, the tubes were stored in the dark and at room temperature during 30 minutes. The reaction provides a blue coloration which is proportional to the protein concentration. The absorbance of the samples was then measured at 750 nm on a spectrophotometer. Results were expressed in BSA equivalent ($mgBSA.L^{-1}$).

TKN method: 0.050 g of dried and grinded samples were added to a matrix and then suspended in 5 mL of distilled water. This was followed by the addition of 10 mL of sulfuric acid solution (98% in mass) and 0.366 g.L^{-1} copper selenite. The sample was then heated up to 420°C until the hydrolysis was completed, generally for 3 hours. Total ammonia nitrogen was finally measured using TKN Buchi AutoKjeldahl Unit K-370 (Buchi®, Flawil, Switzerland), Kjeldhal method. Results were expressed in grams of nitrogen per gram of dry matter.

2.3.1.3 Lipid measurements

Lipids were analyzed as heptane extractable material (HEM). 1 g of freeze-dried and grinded sample was extracted with 25 mL of hot and pressurized heptane using an extra-Accelerated Solvent Extractor ASE 200 (Thermo Fisher Scientific®, Sunnyvale, California 94085 USA). The extracted solution was collected in a 60 mL glass vial. The heptane was evaporated under a N_2 flow. The quantity of extracted fatty matter was measured once the remaining sample was dried at 105°C during 2 hours. Results were expressed in grams of HEM per gram of dry matter (g.gTS^{-1}).

2.3.1.4 COD measurements

The COD of samples and extracts were measured according to the ISO15705:2002 international standard using a COD Vario Tube Test MR (Tintometer GmbH, Division Aqualytic, Dortmund, Germany). The validity range of the assay varies between 22 and 1 500 mg0$_2$.L^{-1}. A 2 mL liquid sample is required. For the negative control, 2 mL of distilled water were added to the test tube. Vials were heated to 150°C for 2 h and the resulting oxygen consumption was determined by spectrophotometry (Muller and Jimenez, 2014). Two different methods were carried out on solid waste. The first method implied the suspension of freeze-dried and grinded samples in distilled water at 1 g.L^{-1}, as recommended by Buffiere et al., (2008). With the second method, 0.050 g of freeze-dried and grinded sample was suspended in a shaker for 6 h in 5 mL of sulfuric acid (98% in mass) instead of water. The results were expressed in milligrams of O_2 per gram of dry matter (mgO_2.gTS^{-1}).

2.3.2 Bio-accessibility Assessment by Chemical Sequential Extraction

The chemical sequential extractions were applied on fresh substrates based on Jimenez et al., (2015a):

- Dissolved Organic Matter (DOM) was obtained by liquid/solid phase separation performed by sample centrifugation (18600 g, 20 minutes, 4°C)
- Readily Extractible Organic Matter (REOM) obtained by a saline basic extraction (8 mL.gTS^{-1} NaCl and NaOH 10 mM) during 15min, at 30°C and 300 rpm of the remaining pellet from DOM extraction

- Moderately Extractible Organic Matter (MEOM) obtained by 4 sequential strong basic extractions (8 mL.gTS^{-1} NaOH 0.1 M) of the remaining pellet during 4 h, at 30°C and 300 rpm of the remaining pellet from REOM extraction
- Slowly Extractible Organic Matter (SEOM) was obtained by 2 sequential strong acid extractions (8 mL.gTS^{-1} sulfuric acid 72%) during 3 h, at 30°C and 300 rpm of the remaining pellet from MEOM extraction
- The remaining pellet from SEOM extraction is considered as Non Extractible Organic Matter (NEOM).

The percentage of extraction of each fraction was measured by a COD analysis of the supernatant.

2.3.3 Anaerobic digestion experiments

2.3.3.1 Batch Experiments to Determine Methane Yield and Methane Production Kinetics

MPR was estimated from the biological degradation of the substrate. Experiments were carried out in 8.5 L batch reactors with 6 L of sludge under mesophilic conditions as described in García-Gen et al., (2015). The substrates were grinded to approximately 1 cm. The reactor was fed at 1 gVS.L^{-1} in eight successive batches with the same substrate to adapt the micro-organism to the substrate. The biogas produced passed through a moisture trap and through a milligas counter fitted with a digital output (MGC-1 gas flow meters, DR.-ING.RITTER APPARTEBAU GMBH & CO, Bochum, Germany); the data were recorded online. Biogas composition was determined using a gas chromatograph (Perkin Elmer GC Clarus 480) equipped with two columns: the first one (RtUBond) was used at 110°C to separate CO_2 and H_2S from other gases, the second (RtMolsieve) was used at 70°C to separate O_2, N_2 and CH_4. The injector and detector temperature was 200°C. The detection of gaseous compounds was done using a thermal conductivity detector. The calibration was carried out with a standard gas composed of 0.1% H_2S, 0.5% O_2, 10% N_2, 25% CO_2 and 64.4% CH_4 (special gas from Linde gas S.A.). The methane production of the last batch was used to obtain the MPR of the substrate. Methane yield was estimated as the maximal methane produced from 1 gram of dry matter on the eighth batch. Methane production times were selected on the MPR as values corresponding to the times required to reach a certain ratio of the methane yield from 5 to 95% with a regular increment of 5%.

2.3.3.2 BMP Test

0.8 gVS of the substrate were digested in batch anaerobic flasks. The volume of each flask was 573 mL, with a working volume of 200 mL. The flask contained a macroelements solution (NH_4Cl, 26.6 g.L^{-1}; KH_2PO_4, 10 g.L^{-1}; $MgCl_2$, 6 g.L^{-1}; $CaCl_2$, 3 g.L^{-1}), an oligoelement solution ($FeCl_2$, 2 g.L^{-1}; $CoCl_2$, 0.5 g.L^{-}

^{1}; $MnCl_2$, 0.1 g.L^{-1}; $NiCl_2$, 0.1 g.L^{-1}; $ZnCl_2$, 0.05 g.L^{-1}; H_3BO_3, 0.05 g.L^{-1}; Na_2SeO_3, 0.05 g.L^{-1}; $CuCl_2$, 0.04 g.L^{-1}; Na_2MoO_4, 0.01 g.L^{-1}), bicarbonate buffer ($NaHCO_3$, 50 g.L^{-1}), an anaerobic sludge at 5 gVS.L^{-1} and substrate at 5 gTS.L^{-1}. Once the flasks were prepared, degasification with nitrogen was carried out to obtain anaerobic conditions and the bottles were closed with red butyl rubber septum-type stopper which were air tight. Triplicate bottles were incubated at 35°C. Biogas volume was monitored by using a manometric device (LEO 2, KELLER). Biogas composition was determined using GC as for batch experiments (2.3.3.1). The endogenous biogas production was determined following the biogas production of the sludge in the same condition without addition of substrates. Then, endogenous biogas production was withdrawn from the biogas production of the BMP test.

2.3.3.3 Batch Experiments for Hydrolysis Modelling

Two additional batch experiments were carried out for hydrolysis modelling:

- wheat straw: after each chemical sequential extraction step, the remaining pellet was used as substrate for a BMP digestion test in order to evaluate each fraction's biodegradation
- Apple: after each chemical sequential extraction step, the remaining pellet was used as substrate for a BMP digestion test in order to evaluate each fraction's biodegradation

2.4 Spectroscopic Analyses

2.4.1 Near Infrared

Freeze-dried and grinded samples were scanned in reflectance over 1000-2500 nm, with a resolution of 2.5*10^{6} nm, using a BUCHI NIR-Flex N-500 solids spectrophotometer fitted with a vial accessory (Buchi®, Flawil, Switzerland). Each sample was measured three times and shaken between each replicate in order to evaluate sample homogeneity. The resulting spectrum was the average of the three replicates which was used to construct the Partial Least Squares regression (PLSr) model. Once the PLSr model calibrated, the Standard Deviation of Repeatability (SDr) of the PLSr was assessed comparing the variance in the estimation of the triplicate with equation 11.

$$SDr = \sqrt{\frac{1}{N}\sum Variance(triplicate)} \tag{11}$$

Three different chemometric treatments of the spectrum were tested. Absorbance (A) was calculated from reflectance (R) as A=log(1/R). For treatment one, absorbance was calculated then SNV was applied. For treatment two, absorbance was calculated then SNV was applied and eventually Savitzky–Golay filter with a second derivative. For treatment three, absorbance was calculated then SNV, and ultimately, detrending were applied.

2.4.2 3D Fluorescence

Each soluble fraction extracted by chemical sequential extractions (section 2.3.2) was analyzed by 3D fluorescence spectroscopy. The fluorescence spectrometer used was a PerkinElmer LS55 in liquid phase fluorescence mode (PerkinElmer®, Waltham, Massachusetts, USA). Excitation wavelengths varied from 200 to 600 nm with increments of 10 nm. The slit width of emission and excitation monochromators was fixed at 10 nm. Scanning monochromator speed was about 1200 $nm.s^{-1}$. Fluorescence values were recorded every 0.5 nm between 200 and 600 nm. Based on Jimenez et al., (2014a), the spectrum were decomposed in seven zones, the intensity of each zone was recorded. For PLSr use, the sample was described by 21 indicators which are the intensity of fluorescence of one zone times the percentage of COD represented by the extracted fraction associated to the spectrum.

2.5 Solid Waste Analyzed

2.5.1 Samples Analyzed for Fast Spectroscopic Characterization

This study was conducted with 90 samples chosen to represent a wide range of potential substrates for AD. More precisely, the analyses were conducted on 15 FOG, 25 fruits, 41 vegetables, 27 farm waste such as straw, manure or grass, 25 types of cereal, 6 kinds of meat or fish extracts, 4 types of micro-algae. All the samples were freeze-dried and grinded to less than 1 mm in size before any analysis. These 90 samples were mixed together to artificially increase the dataset with 195 mixed organic waste and improve the robustness of the regression models. The characteristics of these mixed organic waste were set as the weighted average of the initial substrates. The MPR of the mixed organic waste were created in silico using the modified ADM1 (2.7.1) with parameters of kinetics defined on the MPR of the initial substrates. This approach was validated by García-Gen et al., (2015a). It leads to a database of 338 samples for COD, carbohydrate and nitrogen contents. The analysis of the lipid content was only performed on 231 of these samples, as the required amount of organic matter for lipid analyses was sometimes too large for certain samples with limited availability. The analysis of the methane production time was performed on 275 of these samples. The analysis of the methane production yield was performed on 232 of these samples as the endogenous respiration disturbed the methane yield assessment on substrates with a slow methane production.

Before any correlation with NIRs could be established, the dataset was split into two in order to produce an independent validation set as recommended by (Dardenne, 2010). Thus, an independent validation set comprising 43 samples (38 for lipids, 42 for methane production times) was created for the model validation, including 7 vegetables and 36 mixed organic waste. The remaining dataset of 295 samples (193 for lipids, 194 for methane production times) was used to calibrate the model. The mean content of the calibration set was about 723mgO_2.gTS^{-1}, 0.03 g.gTS^{-1}, 0.06 g.gTS^{-1}, 1229 mgO_2.gTS^{-1} and 0.316 L.gTS^{-1} for carbohydrates, nitrogen, lipids, COD and methane yield respectively. The mean contents of the validation set were about 672 mgO_2.gTS^{-1}, 0.032 g.gTS^{-1}, 0.021 g.gTS^{-1}, 1284 mgO_2.gTS^{-1} and 0.270 L.gTS^{-1} for carbohydrates, nitrogen, lipids, COD and methane yield respectively. Table 3 summarizes the data set characteristics, with Standard Deviance (SD), Relative Standard Deviance (RSD), minimal (min) and maximal (max) values.

Table 3: Dataset features for NIR calibration and validation

		Calibration set	Validation set
Carbohydrates	Max ($mg0_2$. gTS^{-1})	1220	913
	Mean ($mg0_2.gTS^{-1}$)	723	672
	Min ($mg0_2.gTS^{-1}$)	0	299
	SD ($mg0_2.gTS^{-1}$)	286	152
	RSD	0.40	0.23
	N-Sample	295	43
Nitrogen	Max ($g.gTS^{-1}$)	0.14	0.09
	Mean ($g.gTS^{-1}$)	0.03	0.03
	Min ($g.gTS^{-1}$)	0.00	0.01
	SD ($g.gTS^{-1}$)	$2.9*10^{-2}$	$1.8*10^{-2}$
	RSD	1.06	0.55
	N-Sample	295	43
Lipids	Max ($g.gTS^{-1}$)	1.03	0.13
	Mean ($g.gTS^{-1}$)	0.06	0.02
	Min ($g.gTS^{-1}$)	0.00	0.01
	SD ($g.gTS^{-1}$)	0.18	0.02
	RSD	2.88	0.94
	N-Sample	193	38
COD	Max ($mg0_2.gTS^{-1}$)	2884	1605
	Mean ($mg0_2.gTS^{-1}$)	1229	1284
	Min ($mg0_2.gTS^{-1}$)	876	1103
	SD ($mg0_2.gTS^{-1}$)	294	121
	RSD	0.24	0.09
	N-Sample	296	42
Methane yield	Max ($g.gTS^{-1}$)	0.717	0.350
	Mean ($g.gTS^{-1}$)	0.316	0.270
	Min ($g.gTS^{-1}$)	0.129	0.167
	SD ($g.gTS^{-1}$)	0.093	0.041
	RSD	0.294	0.152
	N-Sample	194	38

Figure 6 represents the times required to reach a certain ratio of the methane yield, from 5 to 95%, of both calibration and validation sets.

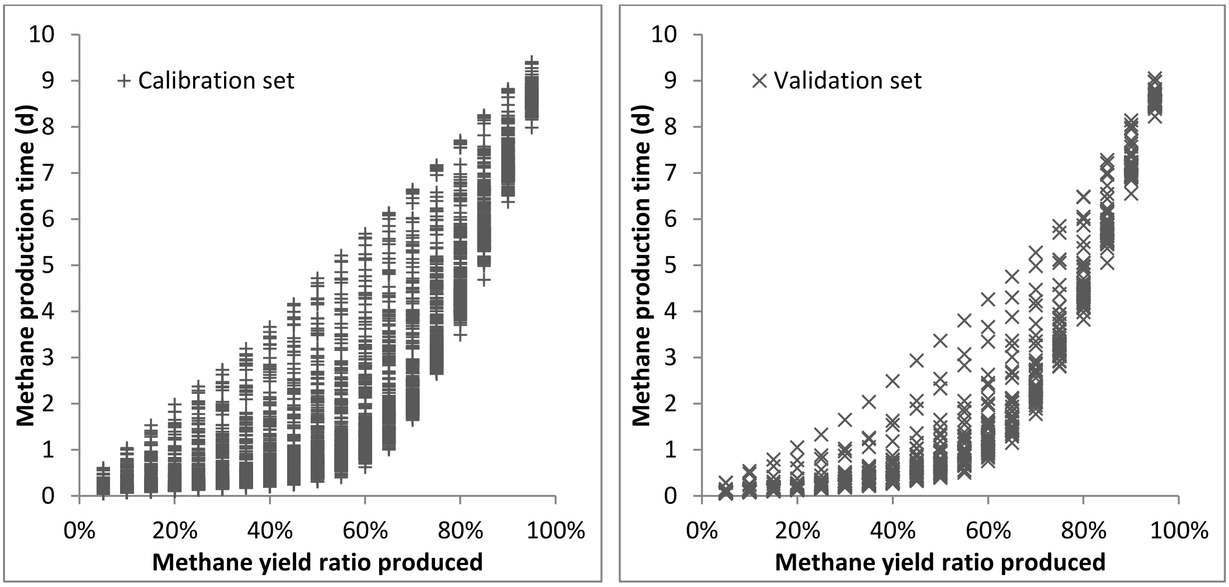


Figure 6: Calibration and validation set features of the experimental data. Description of the methane production time required to reach a certain ratio of the methane yield from 5 to 95%.

2.5.2 Samples Analyzed for Correlation Analyses

In order to estimate the relation between carbohydrate, protein and lipid contents, COD, methane yield and methane production times, only fully characterized samples were considered, reducing the dataset detailed above down to 222 samples. In details, analyses were carried out on Fat, Oil and Grease (FOG), fruits, vegetables, farm waste, cereal, meat or fish extracts, micro-algae and mixed organic waste. The mean content was about 710 $mgO_2.gTS^{-1}$, 0.032 $g.gTS^{-1}$, 0.034 $g.gTS^{-1}$, 1205 $mgO_2.gTS^{-1}$, 0.70 and 0.297 $L.gTS^{-1}$ for carbohydrates, proteins, lipids, COD, biodegradability and methane yield respectively. From 5 to 95% of the methane yield, the methane production mean times were respectively 0.07 d^{-1}, 0.12 d^{-1}, 0.16 d^{-1}, 0.22 d^{-1}, 0.28 d^{-1}, 0.35 d^{-1}, 0.44 d^{-1}, 0.54 d^{-1}, 0.68 d^{-1}, 0.86 d^{-1}, 1.10 d^{-1}, 1.43 d^{-1} 1.88 d^{-1}, 2.53 d^{-1}, 3.43 d^{-1}, 4.55 d^{-1}, 5.80 d^{-1}, 7.13 d^{-1}, 8.54 d^{-1}. The standard deviations were about 246 $mgO_2.gTS^{-1}$, 0.029 $g.gTS^{-1}$, 0.094 $g.gTS^{-1}$, 148 $mgO_2.gTS^{-1}$, 0.15 and 0.063 $L.gTS^{-1}$ for carbohydrates, proteins, lipids, COD, biodegradability and methane yield respectively. From 5 to 95% of the methane yield, the standard deviation of the methane production times were respectively 0.05 d^{-1}, 0.10 d^{-1}, 0.14 d^{-1}, 0.19 d^{-1}, 0.25 d^{-1}, 0.31 d^{-1}, 0.38 d^{-1}, 0.45 d^{-1}, 0.52 d^{-1}, 0.59 d^{-1}, 0.67 d^{-1}, 0.75 d^{-1} 0.81 d^{-1}, 0.82 d^{-1}, 0.77 d^{-1}, 0.65 d^{-1}, 0.51 d^{-1}, 0.35 d^{-1}, 0.18 d^{-1}. Figure 7A summarizes the data set characteristics for carbohydrates, proteins, lipids, COD, biodegradability and methane yield using boxplot. Data were scaled and centered in order to represent the different data on the same figure. Figure 7B summarizes the data set characteristics of methane production time expressed in days needed to reach a percent of the methane yield from 5 to 95 %.

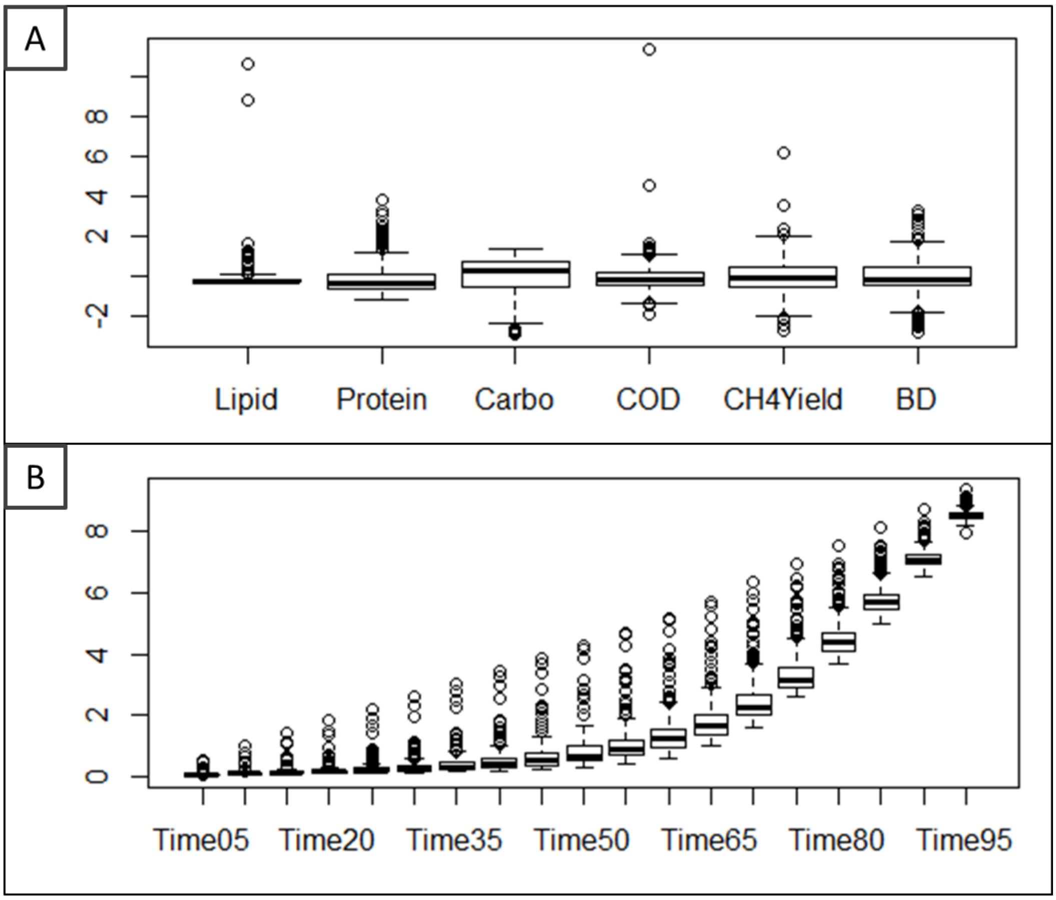


Figure 7: Boxplot of data set characteristics. Figure 7A represents the lipid, protein, carbohydrate (carbo) contents, COD, methane yield (CH4Yield) and biodegradability (BD) on centered and scaled data. Figure 7B represents the methane production time expressed in days needed to reach a percent of the methane yield from 5 to 95%.

2.6 Correlation Analysis and Regression Models

2.6.1 Pearson's Correlation and Cosine

Pearson's correlations calculate the correlations between the variables two by two by dividing their covariance with the product of their standard deviation. It provides a coefficient which varies in a range from -1 to 1, indicating respectively an anti-correlation or a correlation. Permutation tests with 9999 permutations were performed on the Pearson Correlation Coefficient (PCC) indicating if the coefficients were significantly different (alpha=0.05) from 0.

The correlations between the *b* coefficients chosen for each PLSr model were assessed using cosine, package *lsa* from R.3.2.2, which measures the cosine between two vectors. Cosine varies in a range from -1 to 1, indicating respectively a negative co-linearity or a positive co-linearity. A value of 0 indicates the orthogonality of the two vectors. Heat maps and correlation networks were drawn using the package *corrplot* from R.3.2.2 which takes in consideration either PCC and p-values. The non-significant correlations were represented as non-existing correlations.

2.6.2 Principal Component Analysis

Principal Component Analysis (PCA) reduces the dimensionality of the data while retaining as much variance as possible (Al-Kandari and Jolliffe, 2005). PCA projects the data into the space spanned by orthogonal axis named principal components, which are a linear combination of the original variables. In this study, the original variables are carbohydrate, protein and lipid contents, COD, methane yield, biodegradability and methane production time expressed in days needed to reach a percent of the methane yield from 5 to 95%. For the PCA analysis, data were centered and scaled. The number of principal components has been chosen to represent 90% of the total variance. PCA was used as an exploratory tool to identify the variables that explain most of the variance within the data set and to assess the existing correlations between variables. In order to verify the accuracy of the PCA, projections of the samples on the principal components were also observed.

2.6.3 Partial Least Square Regression Model and Performance

2.6.3.1 PLS

The PLS model was performed using the *FACT* tool box with Scilab® 5.5.0. The models were created individually for the prediction of carbohydrate, nitrogen and lipid contents, COD and methane yield. Methane production times were either considered as independent values when running the PLS regression with the tool box *FACT*, Scilab®5.5.0, or as a vector using the package *MixOmics*, R®3.2.2.

First, the models were calibrated using the calibration dataset. Cross-validations were performed on the latter to optimize the number of latent variables. The Jackknife method was used, splitting the dataset into 15 blocks of equivalent sizes. Each block was alternatively removed from the calibration dataset used for the cross-validation in order to estimate the accuracy of the prediction. Thanks to these results, the number of latent variables and the treatment applied to the spectra were chosen to maximize the accuracy and robustness of prediction. Finally the chosen treatment was applied to the spectra of the independent validation dataset. The PLS model was then built on the calibration dataset applied to the treated spectrum of the validation set. The final accuracy and robustness of the model was assessed, based on the recommendation of Dardenne (2010). For the calibration of the model Standard Error of Calibration (SEC), Standard Error of Cross Validation (SECV) and determination coefficients between predicted and reference values (R^2) were followed. The SEC was unfortunately not accessible using MixOmics but the Q^2 and SECV were used to assess the performance of the PLS. Parameters such as the Root Mean Square Error of Prediction (RMSEP), Standard error of Prediction (SEP) and bias were used to assess the accuracy of the PLS on the validation data set.

2.6.3.2 Sparse-PLS

In order to reduce the number of *b* coefficients, sparse PLS was used. Sparse PLS model was performed using the package *MixOmics* with R®3.2.2 (Fourati et al., 2016). The models were calibrated on a calibration dataset. Cross-validations were performed on it to optimize the number of latent variables, the number of variables to keep in *X*-loadings (*KeepX*) and assess a first estimation of the model performances. As for PLS (2.6.3.1), the Jackknife method was used for cross-validation. For the calibration of the model Standard Error of Calibration (SEC) and Standard Error of Cross Validation (SECV) were assessed, for the validation of the model, Standard error of Prediction (SEP) was assessed. Once the model validated, *b* coefficients were analyzed. The vectors of *b* coefficients returned by the sparse PLS had a dimension equivalent to the one of a spectrum. Positive and negative peaks of the vectors of *b* coefficients were visually compared to the data reported on the literature.

2.6.3.3 Distance from the calibration set

To define the relation between the Confidence Interval (CI) of prediction and the distance between a sample and the calibration dataset on the final PLSr model, groups of samples with similarity were subsequently removed from the calibration set and used as validation sets to simulate new samples that would be analyzed. The prediction errors were assessed as well as Euclidian and Mahalanobis distances as described in equations 12 and 13.

Mahalanobis distance:

$$Md = diag\left(\sqrt{U * X' * X * U'}\right) \quad (12)$$

Euclidian distance:

$$Ed_{min} = \min\left[\sqrt{diag[(U - X) * (U - X)]}\right] \quad (13)$$

$$Ed_{mean} = \text{mean}\left[\sqrt{diag[(U - X) * (U - X)]}\right]$$

where U represents the validation spectrum whose distance to the calibration set is analyzed. *X* represents the calibration set of the spectrum and *diag* the function that calculates the diagonal of the matrix. The samples of the calibration set with errors of calibration above 5% were removed from X for distance analysis. It was assumed that samples with high errors of calibration are not correctly taken into account by the PLS regression. These distances were calculated on the processed spectrum but also on the processed spectrum projected on the space described by the PLS loading vectors replacing *U* by *Uproj* and *X* by *Xproj* in equations 12 and 13 with *Uproj* and *Xproj* described as:

$$Uproj = U * (X' * X) * P * [(P' * X' * X * P)^{-1}] * P' \quad (14)$$

$$Xproj = X * (X' * X) * P * [(P' * X' * X * P)^{-1}] * P' \quad (15)$$

where *P* represents the loading vectors of the PLS model.

2.6.3.4 Modelling of the confidence interval based on the distance from the calibration set

Based on the distance and prediction errors analyses (section 2.6.3.3), the relation between the CI and the distance between a sample and the calibration dataset were modelled. To do so, the samples were ranked from the lowest to the highest distances to the calibration set and split in 10 groups with equal number of samples. The first group contains the samples with the lowest distances to the calibration set and the last one, the samples with the highest distances from the calibration set. In each group, the 5% of samples with the highest errors of prediction were selected to represent the CI. The relation between the CI and the distance from the calibration set was modelled using equation 16:

$$Confidence\ interval = a * Distance^{n} + \|b\| \quad (16)$$

The parameters (*a,b,n*) were optimized in order to minimize the square value of the difference between the experimental confidence interval and the confidence interval modelled through equation 16 using the function *leastsq* from Scilab®.5.5.0.

2.7 Dynamic Anaerobic Digestion Models

2.7.1 Dynamic Models: Modified ADM1

ADM1 benchmarked on the BSM2 report (Rosén and Jeppsson, 2008) was implemented in Scilab®. Using the same input as in the BSM2 report, a simulation of 200 days was run to reach steady state. Both relative and absolute errors between the 45 ADM1 variables output from the simulation and from the BSM2 report were investigated. Only minor errors were encountered, the largest relative and absolute errors were in the range of 10^{-4} and 10^{-5} respectively. ADM1 model was then modified based on the work of Ramirez et al., (2009). The degradation of the particulate organic matter has been implemented with Contois kineticss to describe both disintegration and hydrolysis. Hydrolysis was supposed to be non-limiting. Three micro-organism groups were considered for the hydrolysis of carbohydrates, proteins and lipids. A consortium regrouping all the hydrolytic micro-organism was considered for the disintegration. As recommended by (Jimenez et al., 2016), a modification of this model was tested and four different particulate inputs were implemented: X_{RC}, X_{MC}, X_{SC} and X_{NE} (Figure 8) corresponding to the fractionation simulating accessibility, *i.e.* DOM + REOM as readily hydrolysable fraction X_{RC}, MEOM as moderately hydrolysable X_{MC} and SEOM as slowly hydrolysable X_{SC} and NEOM as poorly hydrolysable $X_{\boldsymbol{NE}}$ (REOM, MEOM, SEOM and NEOM are detailed in section 2.3.2). These four particulate inputs were degraded by the same biomass but had different μ_{max} and composition in the modified ADM1. A "switching" function was also introduced to simulate sequential hydrolysis (Equation 17).

$$\rho(i) = k_m \times \frac{S(i)/X}{K_{S(i)} + S(i)/X} \times X \times \frac{1}{1 + \sum_{1\ to\ (i-1)} S / K_I} \tag{17}$$

where:

- k_m is the growth rate of hydrolytic bacteria (d^{-1})
- K_S is the half-saturation coefficient of hydrolytic bacteria
- S is the concentration of organic matter contained in the fraction considered (kg COD/m^3)
- X is the hydrolytic biomass (kg COD/m^3)

- $\sum_{1\ to\ (i-1)} S$ represents the sum of the particulate inputs corresponding to more accessible fraction than *S*.
- K_I is the switching concentration for a fraction to another fraction in the switching function (kg COD/m^3)

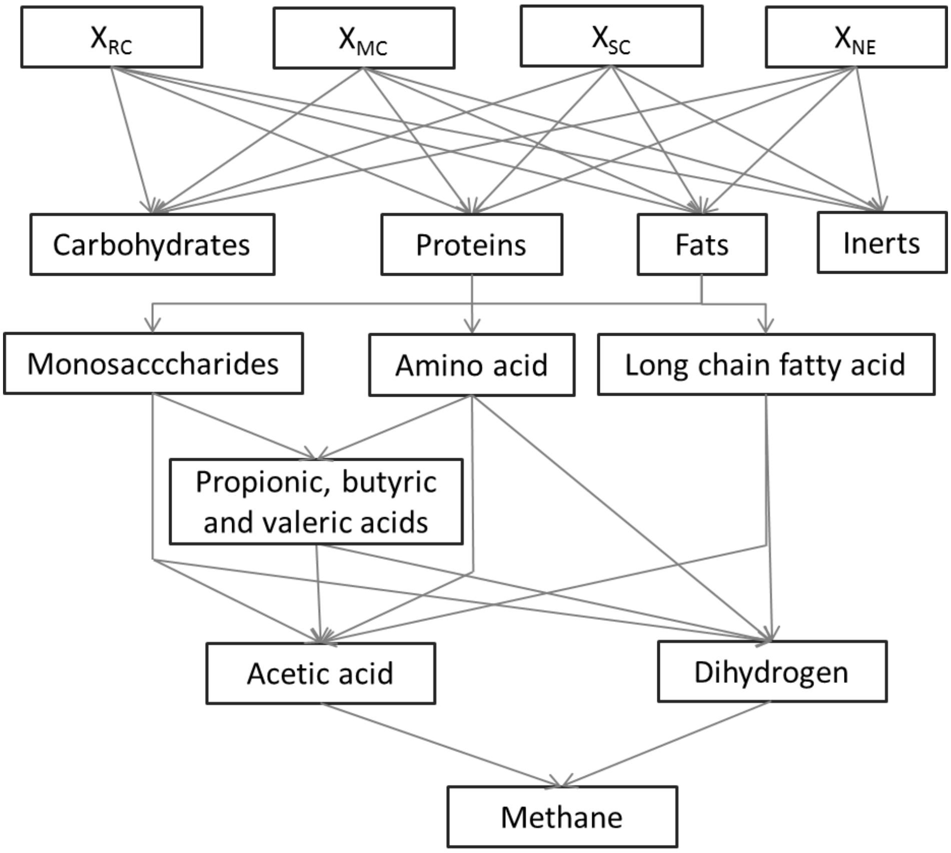

X_{RC}
X_{MC}
X_{SC}
X_{NE}
Carbohydrates
Proteins
Fats
Inerts
Monosacccharides
Amino acid
Long chain fatty acid
Propionic, butyric and valeric acids
Acetic acid
Dihydrogen
Methane

Figure 8: Modified ADM1 with four different particulate inputs X_{RC}, X_{MC}, X_{SC} and X_{NE}.

2.7.2 ADM1 Modelling Approach Using NIRs

To shorten the time needed to implement ADM1, estimations based on NIRs have been used. Conversion of the NIRs predicted analyses values into ADM1 parameters are represented in Figure 9 and detailed below.

2.7.2.1 Carbohydrate, Protein, Lipid and Inert Contents

Nitrogen is converted in protein COD equivalent multiplying TKN in gram by 6.25*1.42. Carbohydrates prediction were based on DuBois et al., (1956) protocol in COD equivalent. Lipids prediction is expressed in gram under heptane extraction per gram of total solid, thus a conversion rate of 2.8 between gram and COD was applied (Batstone et al., 2002). It represents the averaged COD of 1 gram of oil. COD and methane yield prediction were used to assess the non-biodegradable content of the substrate.

2.7.2.2 Methane Production Rate

The prediction of the MPR is made of the predicted methane yield at each of the 19 methane production times (5-95 %). The initial point was added as zero methane produced at time 0. The MPR was then interpolated. A smooth spline with the function *smooth.spline* was done using R® 3.2.2 to remove noise.

2.7.2.3 Particulate Inputs with Different Bio-accessibility

If X_{RC}, X_{MC}, X_{SC} and X_{NE} inputs were implemented, the repartition of the initial COD in X_{RC}, X_{MC}, X_{SC} and X_{NE} was determined by analyzing the changes in the methane production slope. To automatize this procedure, the methane production curve was estimated using linear regression on the first 10% of the methane produced. The first fraction end point is defined as the first point where the difference between the linear regression and the experimental data is above 10% of the methane produced. The percentage of X_{RC} is then defined as the methane produced until the first fraction end point divided by the total amount of methane produced. From this point until an additional production of 10%, a second linear regression is used to define the X_{MC} percentage and so until the whole curve is estimated using linear regression. If the number of fractions is up to 4, the step of 10% is increased by 5% and so on until the number of fractions is inferior or equal to 4.

2.7.2.4 Estimation of the Kinetics

The kinetics was assessed based on the predicted MPR. For the initialization, a maximal specific growth rate (μ_{max}) of 3 d^{-1} was set, representing the mean μ_{max} found in the 275 samples. The initialization of the ADM1 was made in accordance with the experimental protocol. Seven batches were successively simulated with an input of 1 $gTS.L^{-1}$ in a 6 liter reactor under mesophilic conditions. The estimation of the kinetics was made on the eighth batch. The estimation of the kinetics was considered as a non-linear problem. Using ADM1 from (Rosén and Jeppsson, 2008), the μ_{max} of the first order disintegration kinetics was optimized in order to minimize the squared value of the difference between the predicted methane production and the modeled methane production using the function *leastsq* from Scilab®.5.5.0. However, using the modified ADM1 (see Material and Methods 2.7.1), too many μ_{max} terms are present (4) and their estimation had to be made by trial and error to minimize the difference between the predicted methane production and the modeled methane production.

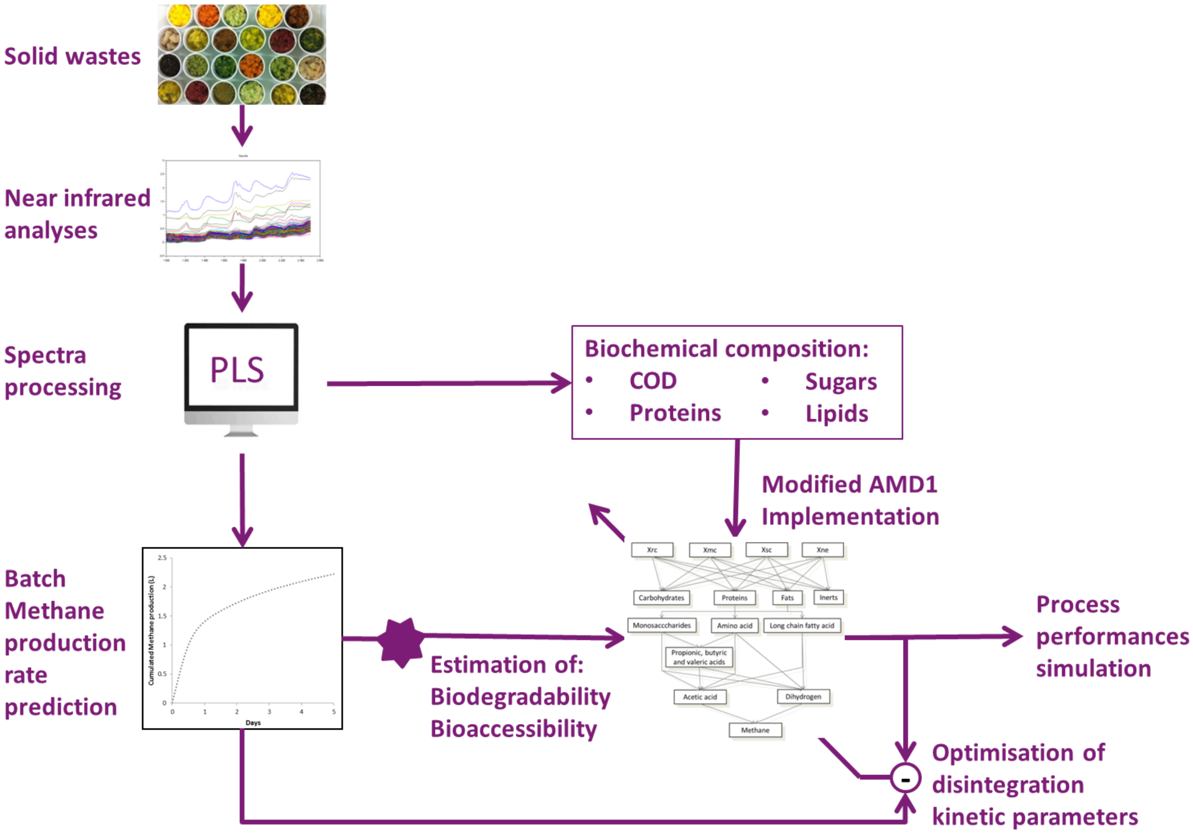

Solid wastes
Near infrared analyses
Spectra processing
PLS
Biochemical composition:
COD
Sugars
Proteins
Lipids
Modified AMD1 Implementation
Xrc
Xmc
Xsc
Xne
Carbohydrates
Proteins
Fats
Inerts
Monosacccharides
Amino acid
Long chain fatty acid
Propionic, butyric and valeric acids
Acetic acid
Dihydrogen
Methane
Batch Methane production rate prediction
Days
Cumulated Methane production (L)
Estimation of: Biodegradability Bioaccessibility
Process performances simulation
-
Optimisation of disintegration kinetic parameters

Figure 9: **Modeling approach diagram**

2.7.2.5 Confidence Range of Simulation

For the prediction of the confidence range of simulation, the same approach as detailed above was used, but ± 1 SEP were added to the 19 methane production times and the methane production yield before to produce the MPR. ± 1 SEP was also added independently to carbohydrate, protein and lipid contents and COD. Every scenario was then simulated. The confidence interval of the simulation was defined has the minimal interval containing all the scenarios.

Chapter 3. Combining pH and Electrical Conductivity Measurements to Improve Titrimetric Methods to Determine Ammonia Nitrogen, Volatile Fatty Acids and Inorganic Carbon Concentrations

Chapter 3. Combining pH and Electrical Conductivity Measurements to Improve Titrimetric Methods to Determine Ammonia Nitrogen, Volatile Fatty Acids and Inorganic Carbon Concentrations 49

3.1 Chapter Guidelines 51
3.2 Technical and Graphical Abstracts 53
3.3 Introduction of the Chapter 55
  3.3.1 Key Parameters to Estimate the Biological State of Anaerobic Digestion Processes 55
  3.3.2 Available Sensors for VFA, IC and TAN Concentrations Measurement 57
3.4 Materials and Methods Related to this Chapter 61
3.5 Results of the Existing Methods 61
  3.5.1 Estimation of VFA and IC 61
  3.5.2 Estimation of TAN Using Non-Linear Resolution 66
3.6 Accurate Estimation of VFA, IC and TAN Concentrations by Coupling pH and Electrical Conductivity Measurement. 68
  3.6.1 Building of a Decision Tree 69
  3.6.2 Example of a Titration Using SNAC 71
  3.6.3 Discussion of the Advantages between the Existing Methods and SNAC 73
3.7 SNAC Prototype 76
3.8 Conclusion 78

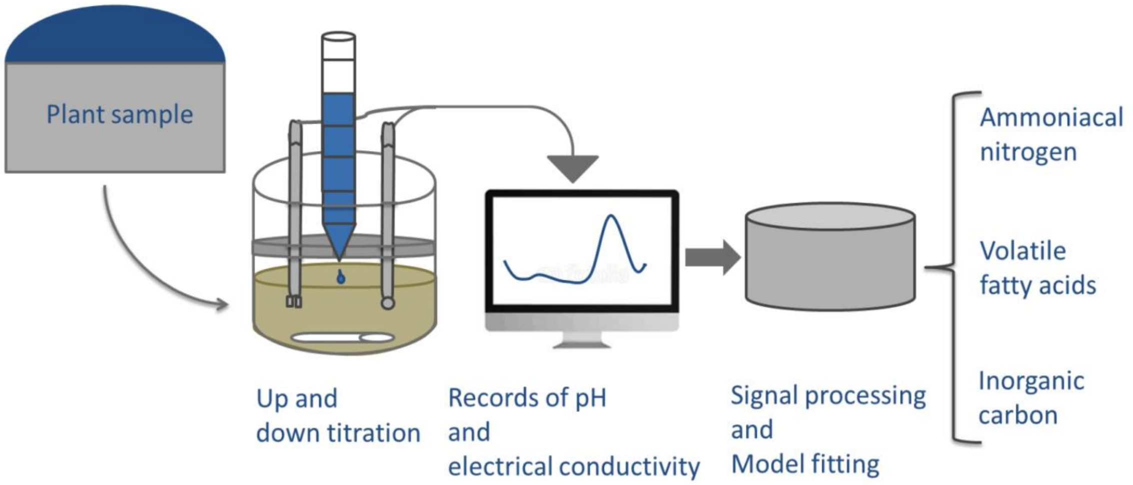
3.1 Chapter Guidelines

This chapter is based on the publication (Charnier et al., 2016)

This PhD aims to develop a method that allows predicting the AD process performances based on feeding strategy. To do so, a required step is the estimation of the biological state. It can be seen as the starting point of the simulation. Thus, this chapter presents in section 3.3, the state of art on biological state estimation and focuses especially on analyses of soluble compounds using titrimetric methods. Then a short reference to the “Materials and Methods” section is made in 3.4. The second section, 3.5, presents the results obtained using the existing titrimetric methods for the estimation of volatile fatty acids (VFA), inorganic carbon (IC) and total ammonia Nitrogen (TAN). The section 3.6 presents the results obtained by the new titrimetric method developed in this PhD and finally, section 3.7, presents the prototype developed with SNAC method.

3.2 Technical and Graphical Abstracts

Volatile fatty acids (VFA), inorganic carbon (IC) and total ammonia nitrogen (TAN) are key variables in the current context of anaerobic digestion (AD). Accurate measurements like gas chromatography and infrared spectrometry have been developed to follow the concentration of these compounds but none of these methods are affordable for small AD units. Only titration methods answer the need for small plant monitoring. The existing methods accuracy was assessed in this study and reveals a lack of accuracy and robustness to control AD plants. To solve these issues, a new titrimetric device to estimate the VFA, IC and TAN concentrations with an improved accuracy was developed. This device named SNAC (System of titration for total ammonia Nitrogen, volatile fatty Acids and inorganic Carbon) has been developed combining the measurement of electrical conductivity and pH. SNAC were tested on 24 different plant samples in a range of 0-0.16 mol.L^{-1} TAN, 0.01-0.21 mol.L^{-1} IC and 0-0.04 mol.L^{-1} VFA. The standard error was about 0.012 mol.L^{-1} TAN, 0.015 mol.L^{-1} IC and 0.003 mol.L^{-1} VFA. The coefficient of determination R² between the estimated and reference data was 0.95, 0.94 and 0.95 for TAN, IC and VFA respectively. Using the same data, current methods based on key pH points lead to standard error more than 14.5 times higher on VFA and more than 1.2 times higher on IC. These results show that SNAC is an accurate tool to improve the management of AD plant.


Plant sample
Up and down titration
Records of pH and electrical conductivity
Signal processing and Model fitting
Ammoniacal nitrogen
Volatile fatty acids
Inorganic carbon

3.3 Introduction of the Chapter

3.3.1 Key Parameters to Estimate the Biological State of Anaerobic Digestion Processes

Several variables are found to strongly influence the process such as micro-organisms activity, pH, temperature, alkalinity, concentrations of free ammonia, hydrogen, sodium, potassium, heavy metals, volatile fatty acids (VFA), ions concentration and other compounds (Appels et al., 2008). However, AD process performances are mainly evaluated through the methane production. Methane production is usually estimated from both biogas production and biogas composition. Biogas production is monitored through flow sensors that are frequently part of the standard digester instrumentation. Gas composition measurements are less common. The basic gas analyzers are based on infrared absorption measurements to determine carbon dioxide and methane concentrations and can be either on-line or off-line. Several of the sensors today available on the market are mainly found in units with a minimal production of 150-200 kWe (Jimenez et al., 2015b). More sophisticated gas analyzers can estimate specific compounds of interest for the process performances estimation or the biogas valorization such as hydrogen sulfide or volatile organic compounds.

Temperature is another key variable of the process because it drives the microbial population and the kinetics of reactions according to the Arrhenius law and thermodynamic equilibrium such as gas-liquid equilibrium according to Henry law or acido-bacic equilibrium affecting the pKa. Three commonly used temperature sensors are available: resistance thermometer, thermo-element, and thermistor (Jimenez et al., 2015b).

pH is a critical variable, its value indicates the risk of acidification and process failure. Immersion of the pH probe in liquid phase for on-line measurements led to the development of hydraulic, mechanical and chemical cleaning strategies, to limit the development of biofilm and probes clogging. On-line liquid sensors such as temperature or pH sensors are usually accurate but depending on the mixing in the reactor, the sensors can be misleading, estimating local values not representative of the plant. For more details, Harremoes et al. (1993) provided an overview on the good practice of sensors installation in wastewater treatment plants for a guaranteed accuracy.

In terms of monitoring, pH measurement cannot be the sole indication of acidification because in well-buffered medium, high acids concentration would have to form in order to cause a detectable drop in pH, by which process failure would already occur (Lahav and Morgan, 2004). Thus, acid production, mainly represented by VFA production, has to be monitored. There has been a significant amount of research directed towards on-line VFA determination techniques in particular Gas Chromatography (GC) headspace techniques (Boe et al., 2007 and Boe et al., 2008) but these methods remain too expensive for most industrial plants.

Mainly produced during acidogenesis, acid forms of VFA compounds inhibit the following steps of the process: methanogenesis and acetogenesis. Up to a threshold, VFAs accumulate in the reactor, slow down and eventually stop the whole process. The inhibition is represented by the acid form of VFAs and is therefore correlated with pH. Thus, a second important compound concentration to monitor is the concentration of inorganic carbon (IC). Indeed IC can buffer the plant acidification thanks to the pKa of H_2CO_3/HCO_3^- which is at 6.3 at 25°C (Figure 10).

Total Ammonia Nitrogen (TAN), the end-product of anaerobic digestion of proteins, urea and nucleic acids is another important variable to monitor. Unlike the importance of ammonia for bacterial growth at lower concentrations, high concentrations of ammonia may cause a severe disturbance in the anaerobic process performance (Rajagopal et al., 2013). Hejnfelt and Angelidaki, (2009) reported that 3 $g.L^{-1}$ of TAN could inhibit 50% of the methanogenesis. TAN inhibitory effect depends on the pH, the inhibition is represented by free ammonia, the basic form of TAN (pKa = 9.2, at 25°C), because of its permeability through the cell membrane (Koch et al. (2010)). Due to their inhibition or pH buffering effects, VFA, IC and TAN concentrations are three of the main parameters to monitor AD plants (Molina et al. (2009)).

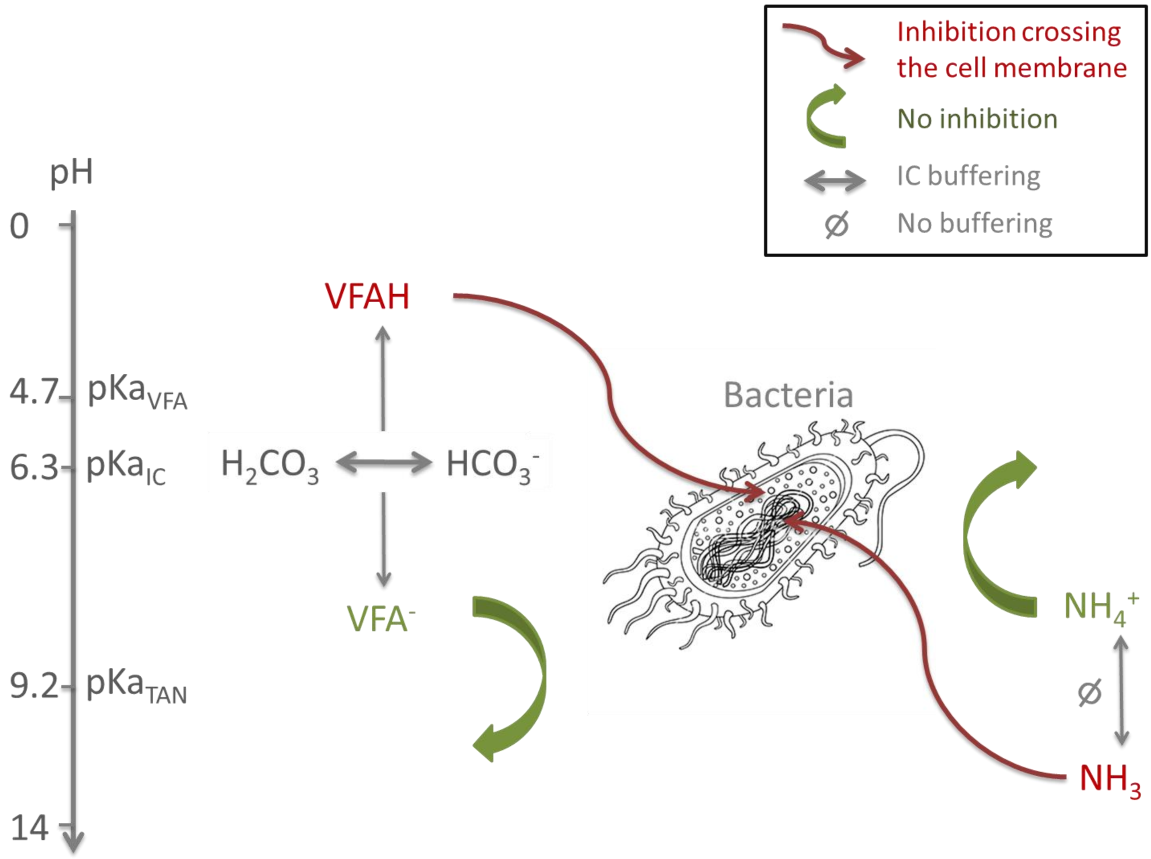


Figure 10: Effect of pH on ammonia and volatile fatty acids inhibition. Red arrows indicate the inhibitory effect of non-ionized form of TAN and VFA on the bacteria.

3.3.2 Available Sensors for VFA, IC and TAN Concentrations Measurement

Over the years, research has been focused on developing a methodology and a sensor that allow estimating the concentrations of VFA, IC and TAN in AD samples. An often encountered problem is the volatility of IC, TAN and to a smaller extent VFAs, which can induce a loss of IC, TAN and VFA from samples to be analyzed. Nonetheless, a simple titration method is deemed to be adapted as demonstrated by Spanjers and Van Lier (2006). Nowadays, most of the titration tools used in anaerobic digestion are based on titrant's volume analysis on key points. Nordmann (1977) and Ripley et al. (1986) developed two similar methods to estimate the ratio VFA/IC. In both methods, a first step of the titration provides an estimation of IC followed by a second step of the titration to estimate VFA.

Several studies have been conducted to adapt these formulas to on-line titrators for liquid industrial effluents. A first study on wastewater from textile finishing industries designed an on-line device using commercially available component to estimate VFA concentrations (Feitkenhauer et al. (2002)). It was followed by Bouvier et al. (2002) who designed a on-line sensor estimating both VFA and IC concentrations using a recalibration of the Ripley et al. (1986) equations. This device found commercial applications under the name Anasense® (Applitek Belgium) and has proven to be suitable to monitor industrial scale plants (Ruiz et al., 2005). Even if these methods are suitable for industrial effluents, their common drawback is the underlying assumption that VFA and IC represent the whole alkalinity of the samples.

Thus, additional studies have been done to estimate the true value of VFA and IC concentrations taking into account already known concentrations of phosphorous and sulphide compounds to calculate alkalinity. Moosbrugger et al. (1993) recommended a first accurate method to estimate IC and VFA concentration based on 5 pH points. Two of these pH points are at 0.5 unit to each side of the pKa of acetate, two to the pKa of HCO_3^-/H_2CO_3 and the last point is usually between pH=3 and pH=2 in order to correct the pH values measured. Hence, knowing the concentration of sulfur and phosphate compounds in solutions, these authors advocated that the equations and the software developed estimated an accurate value of IC and VFA concentrations. A method using eight pH points was also proposed using almost the same methodology (Lahav and Loewenthal, (2000)). Even if more accurate than the equation developed by Nordmann (1977) and Ripley et al (1986), this method is difficult to apply because its accuracy depends on the exact knowledge of the concentrations of sulphide and phosphorous. Moreover, they are not the only components that account for total alkalinity.

While there was no major study published on this topic over the last years, Purser et al. (2014) recently advocated for a recalibration of the method developed by Nordmann (1977) depending on the substrate used. These new equations reduced the absolute mean error by 600 % to 800 % depending on the substrate (Figure 11). The same year, Vannecke et al., (2014) evaluated the use of the methods developed by Moosbrugger et al., (1993) and Lahav and Loewenthal, (2000) and concluded that the method developed by Moosbrugger et al., (1993) is more accurate to monitor anaerobic digesters treating solid waste. Eventually, Lützhøft et al., (2014) confirmed the interest of titration methods to monitor AD plants. They compared the different methods and concluded that the simpler the method (eg., using only two pH points as did Nordmann (1977) and Ripley et al (1986)), the better the results. However linear correlation between titration results and GC VFA measurements was found in this study, it highlighted that the use of titration methods on complex matrix such as digested manure with high concentrations of interfering components induces an overestimation of VFA concentration.

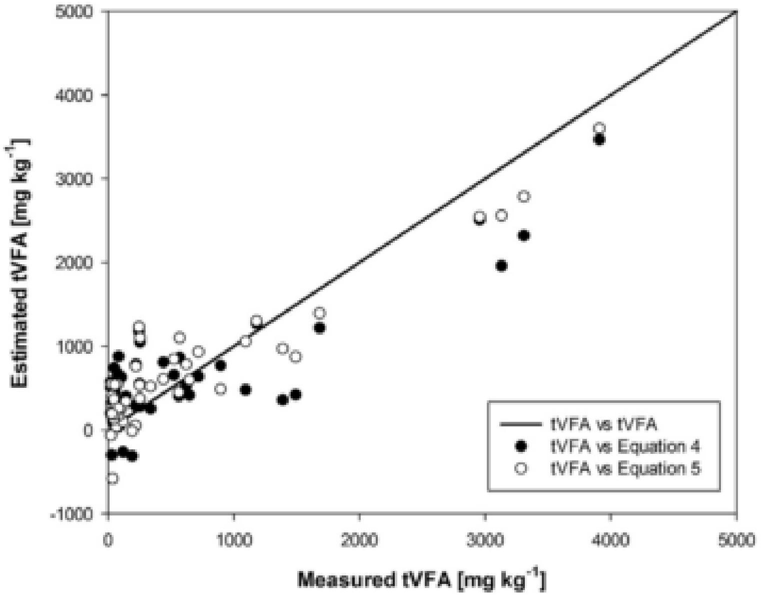


Figure 11: Estimation of total VFA (tVFA) concentrations from the recalibrated equation of Ripley et al., (1986) (equation 4) and the recalibrated equation of Nordmman (1977) (equation 5) plotted against the tVFA measured by gas chromatography for 52 energy crop and slurry digestate samples *(Purser et al., (2014)).*

All these methods have a major drawback as they only take a limited number of components into account. Thus, the only way to obtain accurate results is to recalibrate each method for each new substrate. Hence, some researches focused on a method able to provide accurate results with any kind of component present in the solution (Van Vooren, (2000), Van De Steene et al., (2002)). This method is based on the whole titration curve and the buffer capacity is calculated as $\beta = \frac{dCt}{dpH}$ where Ct is the concentration of the associated counter-ions used for the titration of the solution (Figure 12). A model simulates the buffer capacity during the titration depending on the concentrations of VFA and IC. By minimizing the error between the model and the experiment, an estimation of VFA and IC concentrations is obtained. However, even if promising, this method can make errors in distinguishing different components.

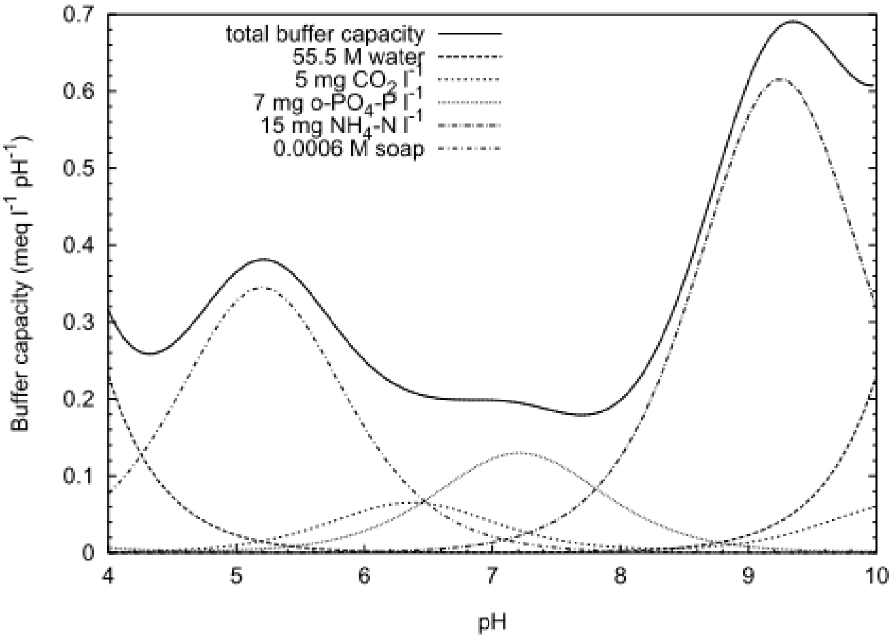


Figure 12: Buffer capacity curve for a complex aqueous system containing 5mg CO_2.L^{-1}, 7 mg o-PO_4-P.L^{-1}, 15 mg NH_4^+-N.L^{-1} and 0.6 meq.L^{-1} of an unspecified soap *(Van Vooren 2000)*.

This work was followed by Zaher et al (2003) and Zaher (2005) who proposed to develop this method adding an up-titration to remove IC from the equation and to estimate TAN concentration. These authors also recommended implementing other components than VFA, IC and TAN in the model, like lactic acid, phosphorus or even blind buffer in order to better fit experimental data. This method meets the problem of the very large number of different components that can be present in a plant sample and loses its accuracy in case of a solution with too many components. Another problem is the presence of complexes which are impossible to distinguish from acids with only a pH probe.

The aim of this thesis is first to compare the accuracy of the estimations from Nordmann (1977), Ripley et al (1986), Purser et al (2014) and Zaher (2005) methods on 24 samples and secondly to develop an innovative sensor in order to improve the monitoring of AD plants. We decided to go further with the buffer capacity method since it is the most promising one to deal with additional component interferences as recommended by Zaher et al. (2003). To do so, we added a variable that could allow differentiating the component of a plant sample in an easier way. Thus, an electrical conductivity sensor was implemented in the titration vessel. It has been shown that combining pH and conductivity can be a powerful tool for the estimation of VFA and IC concentration (Aceves-Lara et al. (2012)). Because the conductivity is influenced by the ionic charge of the component in the media, it becomes easier to distinguish the components. Using buffer capacity, the effect of compounds on the titration curve depends only of their concentration and pKa, while the effect on the electrical conductivity depends on the concentration and the pKa but also on the molar conductivity which is specific to the compounds. The protocol of titration was adapted with the addition of up-titration after down-titration in a closed reactor in order to obtain an estimation of IC, VFA and TAN concentrations. Then a way to treat and combine these variables was developed to obtain a robust and accurate method. This was applied to 24 plant samples chosen to represent a wide range of AD process.

3.4 Materials and Methods Related to this Chapter

Detailed information about the materials and methods related to this chapter are provided in the “Materials and Methods“ section 2.2.

3.5 Results of the Existing Methods

3.5.1 Estimation of VFA and IC

As described in the “Materials and Methods” sections 2.2.4.1 and 2.2.4.2, the four methods (Nordmann, (1977); Ripley et al., (1986); Purser et al., (2014) and Zaher (2005)) were tested on the data collected during the titration of the 24 AD plant samples. Nordmann (1977), Ripley et al. (1986) and Zaher (2005) provided acceptable estimation of IC concentration with a strong correlation, respectively $R^2=0.91$, $R^2=0.87$ and $R^2=0.87$ and standard errors of 0.019, 0.044 and 0.033 mol.L^{-1} since IC is the main component influencing the down titration (Figure 13). Nonetheless, Zaher (2005) and in a stronger manner, Ripley et al (1986) tend to underestimate IC with a coefficient of the regression of 0.65 and 0.60 respectively while using Nordmann (1977) the regression coefficient is about 1.03. Thus, the assumption that 80% of the IC was titrated at pH=5.75 made by Jenkins et Morgan, (1983) was not accurate for the samples analyzed and a recalibration of this assumption could be made as it was done for VFA. But we agreed on the choice of Purser et al. (2014) not to work on the recalibration of IC estimation since good correlation was observed using Nordmann (1977) equations.

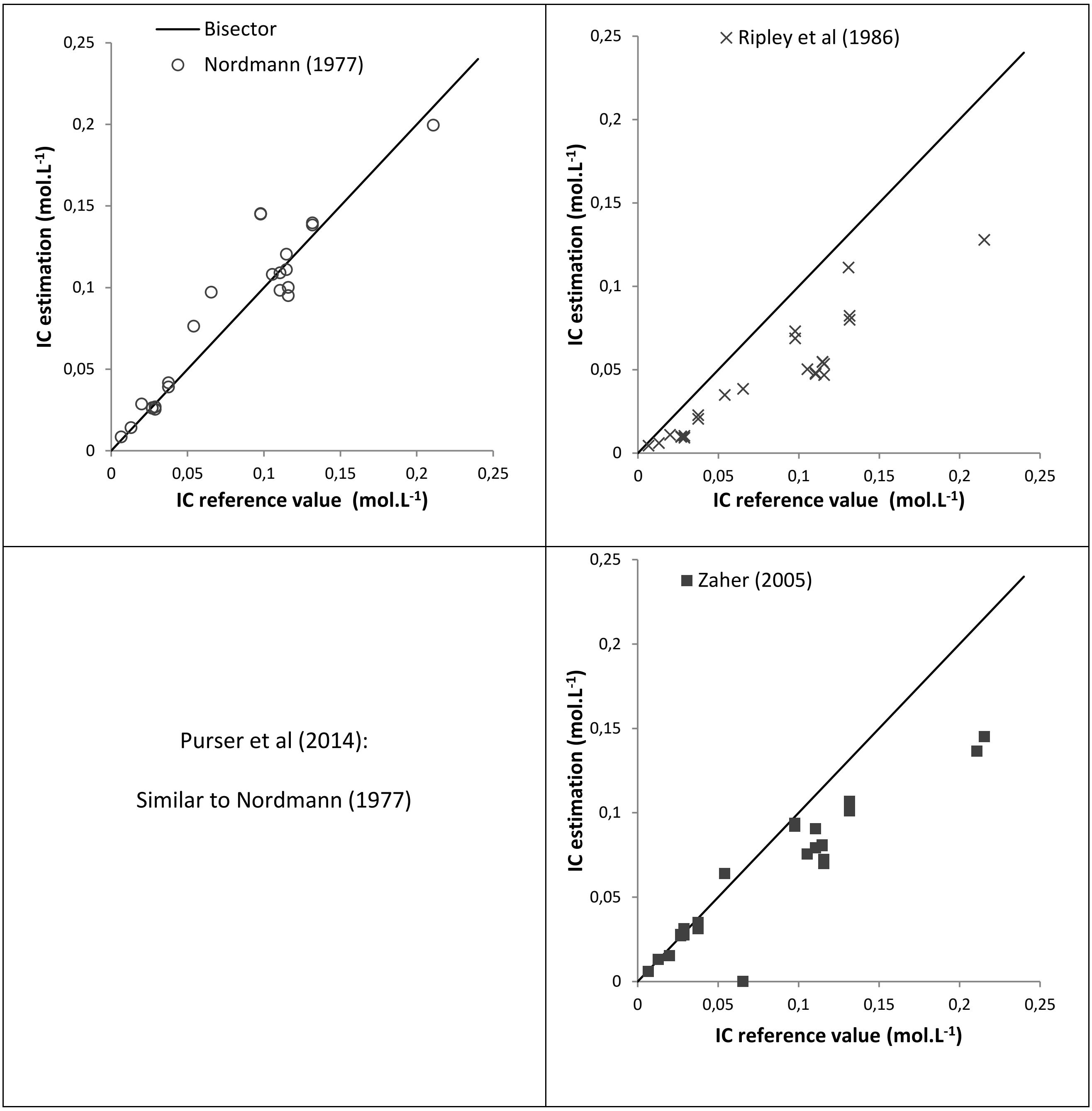


Figure 13: Estimation of IC concentrations by the method of Nordmann (1977), Ripley et al (1986), Purser et al (2014) and Zaher (2005)

VFA has usually a low impact in the titrant volume. Thus, its estimation is easily disturbed by other compounds and the complexity of the samples induced errors when estimating VFA concentration. A determination factor of $R^2=0.02$ were found for equation of Nordmann (1977) and $R^2=0.05$ for equation of Ripley et al (1986) between the GC values of VFA concentration and the estimations provided by these methods (Figure 14). The recalibrations proposed by Purser et al (2014) provided more accurate estimations (equation 3.1: $R^2=0.03$; equation 3.2: $R^2=0.03$; equation 3.3: $R^2=0.11$). But it seems that the equations were not calibrated for this kind of sample. Indeed Purser et al., (2014) calibrated their equations on samples containing VFA in a range of 0-4.5 g/L when our samples were in a range from 0-3 g/L. Moreover, their equations were adapted to one kind of plant sample when we used a huge variety of plant samples. The method developed by Zaher (2005) has been tested and has shown more accurate results in the 24 AD plant samples. The correlation between the estimated value and the GC value is about $R^2=0.58$ for VFA.

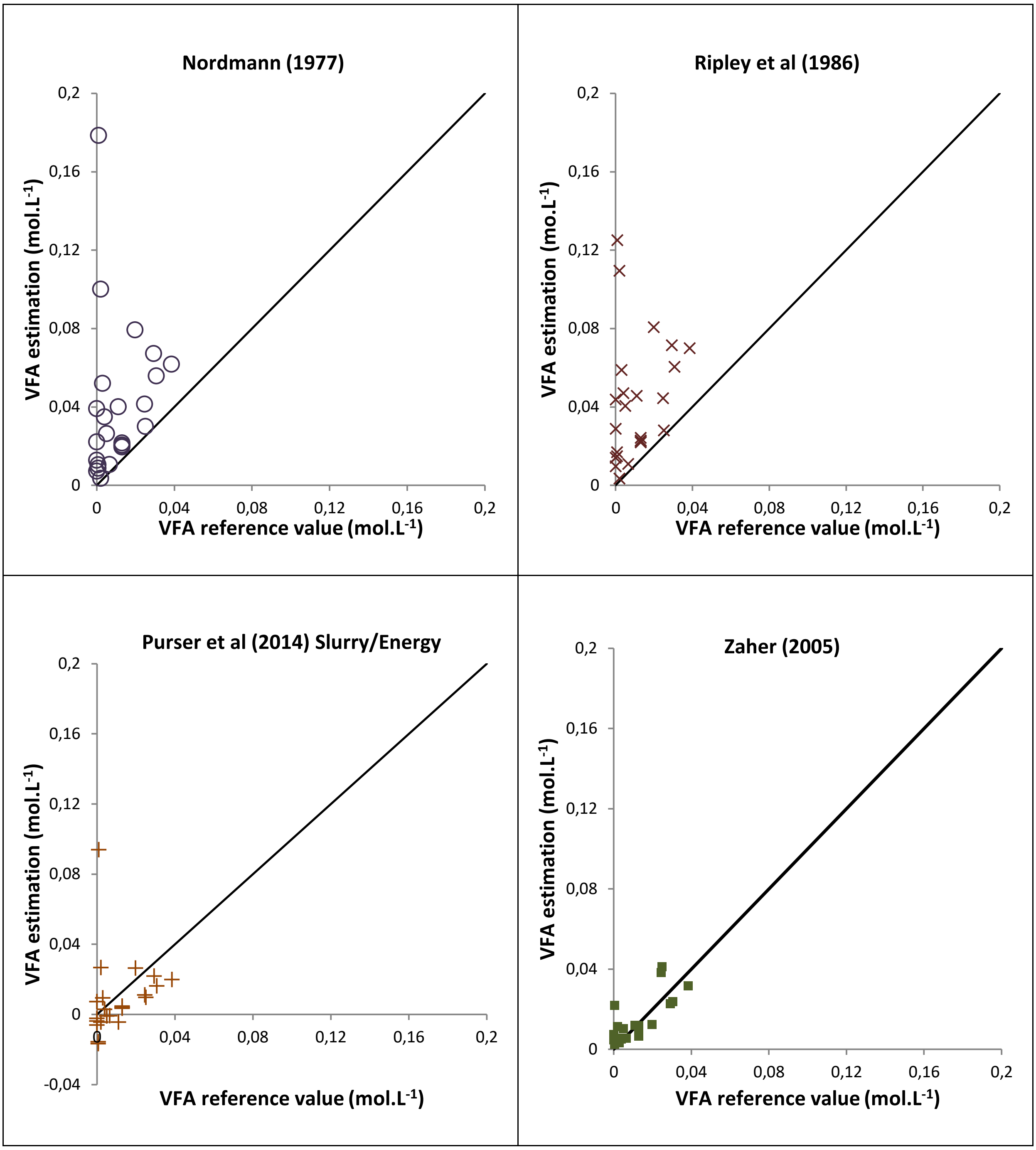


Figure 14: Estimation of VFA concentrations by the method of Nordmann (1977), Ripley et al (1986), Purser et al (2014) and Zaher (2005). The bisector is represented by a straight line on the graphics.

These results tend to show that a method based on the analysis of the titrant volume on key points cannot be accurate for a wide variety of complex samples and that non-linear resolution on the buffer capacity, as used by Zaher (2005), increases the accuracy of the estimation. Indeed, in case of key point analysis, the same equation is applied independently from the number of components in the sample, thus assuming strong hypotheses on its composition. For instance, lactic acid which has a pKa=3.9 will influence the estimation of VFA using equation developed by Nordmann (1977), Ripley et al (1986) or Purser et al (2014). The buffer intensity of lactic acid, IC and VFA is represented in Figure 15. The buffering effect of water is observed from pH 3. It can be seen that lactic acid has a non-zero buffer intensity between the VFA thresholds (pH 4.4 and 5) where only VFA and IC are supposed to influence the titration. Concretely, according to the Nordmann (1977) formula, the buffer intensity of lactic acid will be considered as the one of VFA, overestimating VFA concentration.

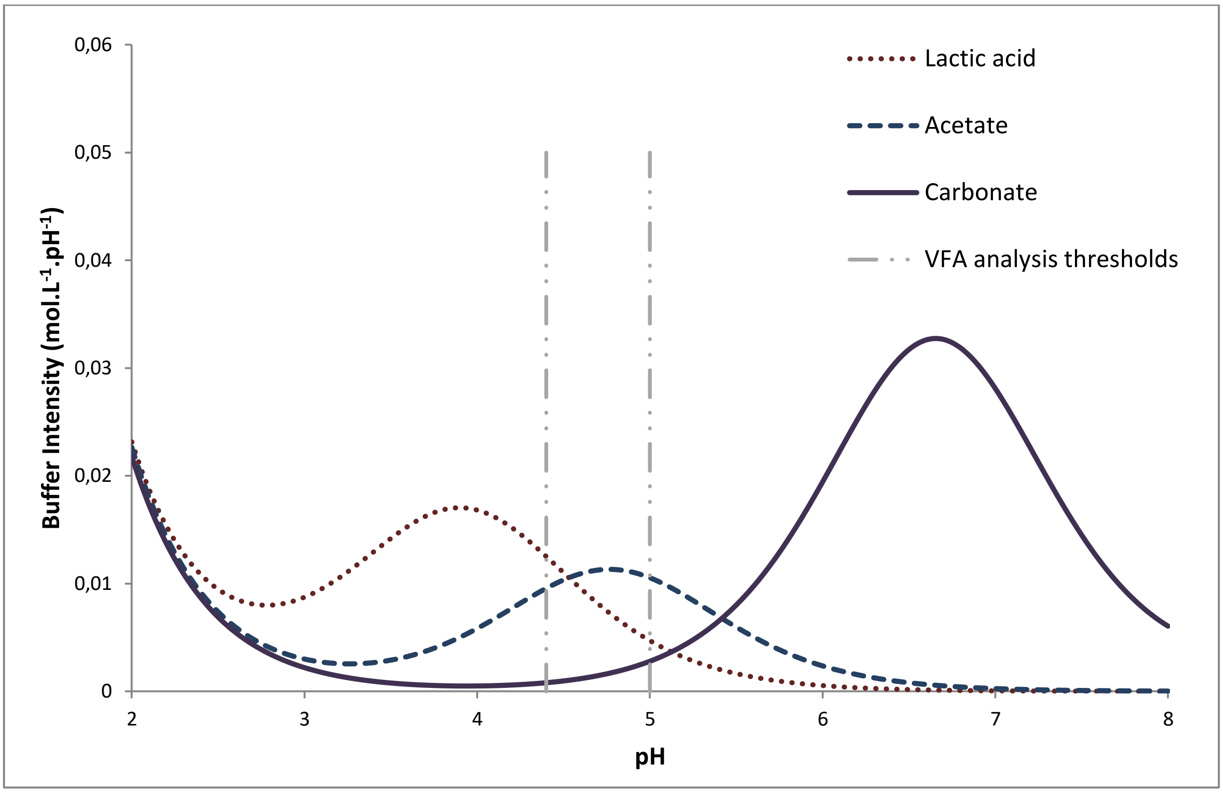


Figure 15: Buffer intensity of lactic acid, acetate and carbonate and their influences on the estimation of VFA and IC by Nordmann (1977). The VFA analysis thresholds indicate the range of pH in which only IC and VFA are supposed to influence the titration using Nordmann (1977) equations.

Nonetheless, the methods developed by Nordmann (1977) and Ripley et al. (1986) are widely used to monitor industrial AD plants. A commercial argument of the firm commercializing the associated devices is that it can be a powerful tool to follow the variation of VFA and IC in an AD plant supposing that only VFA and IC vary in time. Since in Figure 15, it has been shown that lactic acid concentration but also IC concentration impact the estimation of VFA using Nordmann (1977) method, this hypothesis has been checked running a simulation in which only VFA and IC vary in time during the AD process with sinusoidal variations. This purely theoretical simulation allows us to follow the evolution of the estimation of IC and VFA with wide variations. A simulation was run on 100 days with a constant concentration of lactic acid (0.03 mol.L^{-1}), variation of IC from 0 to 0.1 mol.L^{-1} and variation of acetate from 0 to 0.033 mol.L^{-1}. Estimation of VFA and IC was performed using Nordmann (1977) method and compare to the theoretical data (Figure 16).

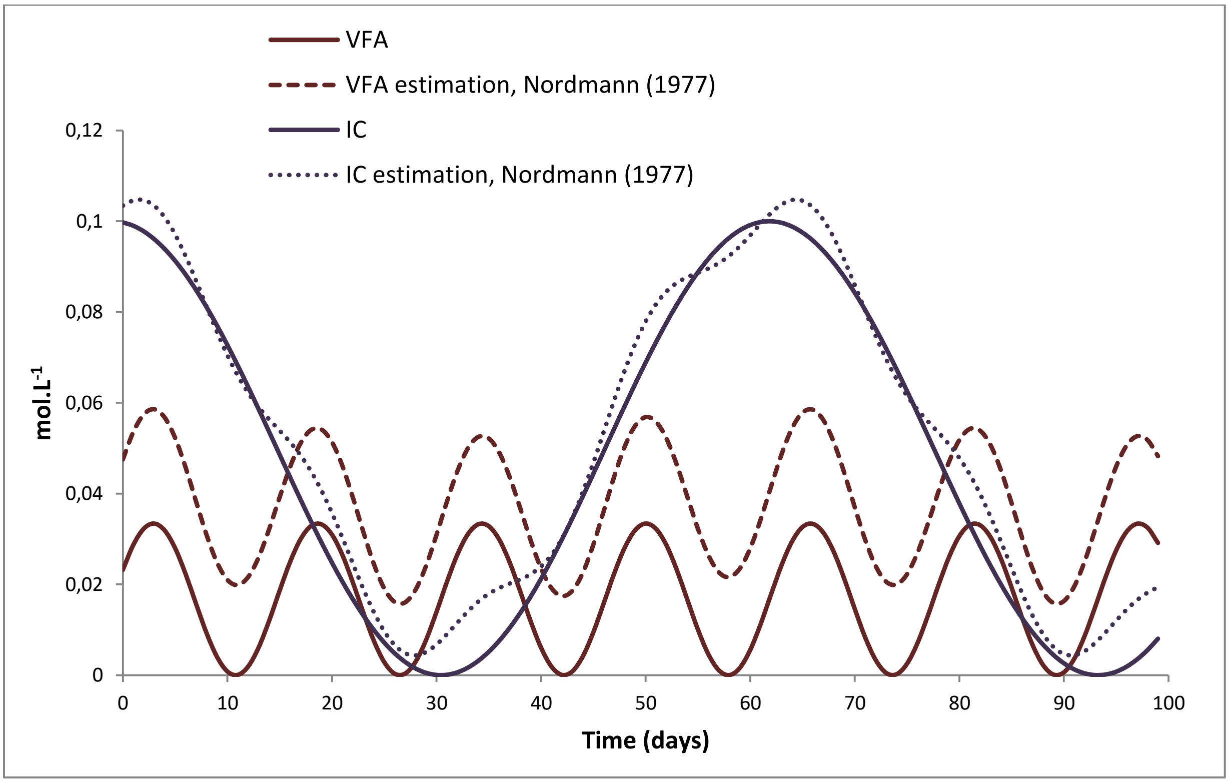


Figure 16: Simulation of Nordmann (1977) estimation for IC and VFA on 100 days supposing a constant concentration of lactic acid (0.03 mol.L-1)

The use of different phase for the sinusoidal variations of IC and VFA highlights slight effects of VFA and IC concentrations on the estimation of each other. Nonetheless, it demonstrates that, as supposed, even if the absolute estimation of VFA is wrong, it allows us to follow the trend of the evolution of VFA and IC if the composition of the other components is stable. A limit of this approach is the stability assumption of the other components that cannot be verified by the device. The use of this assumption must be made carefully, evermore if the characteristics of the substrates to be digested changes.

3.5.2 Estimation of TAN Using Non-Linear Resolution

The method developed by Zaher (2005) was also tested to estimate TAN on the 24 plant samples. The coefficient of determination between the estimated values and the TKN values is about $R^2=0.82$ for TAN. Despite the good correlation, the accuracy of the estimation of TAN concentration decreases with the concentration of TAN (Figure 17, left part).

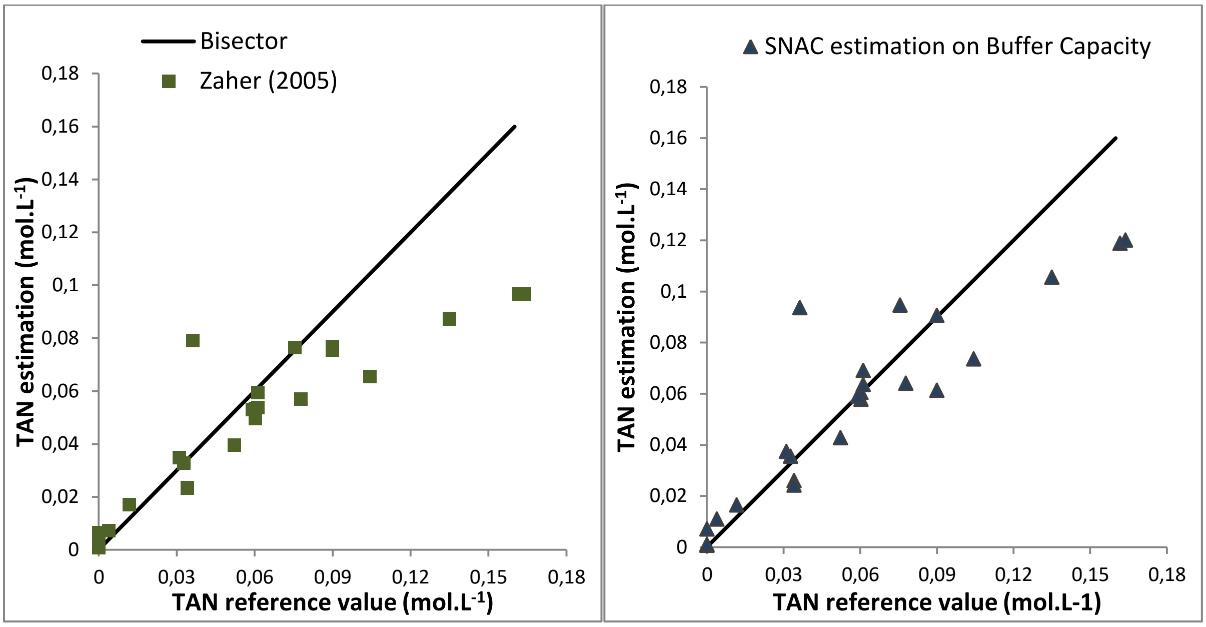


Figure 17: TAN concentrations estimation based on buffer capacity using the method developped by Zaher (2005) on the left part and SNAC on the right part.

The method was limited by two main factors. First, it assumes that the volume does not change during the titration, which is not a good assumption when the concentration is high. The adding titrant volume can represent 20% of the total volume of the solution which dilutes the TAN concentration. This phenomenon is illustrated by the estimation of TAN concentration on the highest values (Figure 17). It could be solved diluting the samples before the titration in order to reduce the amount of TAN. However, the problem is that the accuracy on VFA estimation will decrease due to a lower VFA impact on the signal and thus, a higher sensibility to noise. The second solution is to increase the molarity of the titrant, even if the experiment is carried out at 1 M and Zaher, (2005) recommended 0.5 M. It could be applied but in case of low concentration of TAN and VFA, it would drastically reduce the amount of data recorded during the titration and impact the accuracy of the estimation. Another limit of the resolution on buffer capacity is that the buffer capacity is not charge-sensitive. Thus, for instance, the dissociation of a salt will be seen as the acidification of an acid on the buffer capacity curve. And we noticed that this phenomenon happened many times in our experiments, compromising the use of buffer capacity alone.

The previous model (section 2.2.4.2) was modified to correct the drawbacks highlighted. First of all, we assumed a thermodynamic equilibrium between gaseous and liquid phase for TAN and IC. This hypothesis relies on the fact that titration is made in a closed reactor at constant pressure (equation 6, section 2.2.4.3). Then the possible variation of volume during the titration is taken into account (equation 7, section 2.2.4.3). These modifications corrected the effect of dilution observed using the Zaher (2005) method as shown in Figure 17.

3.6 Accurate Estimation of VFA, IC and TAN Concentrations by Coupling pH and Electrical Conductivity Measurement.

If buffer capacity was used for down titration, up titration was more challenging. Indeed both buffer capacity and buffer conductivity provide interesting data and a decision tree had to be designed. A first estimation was made on buffer capacity and a second one on buffer conductivity for up titration (Figure 18).

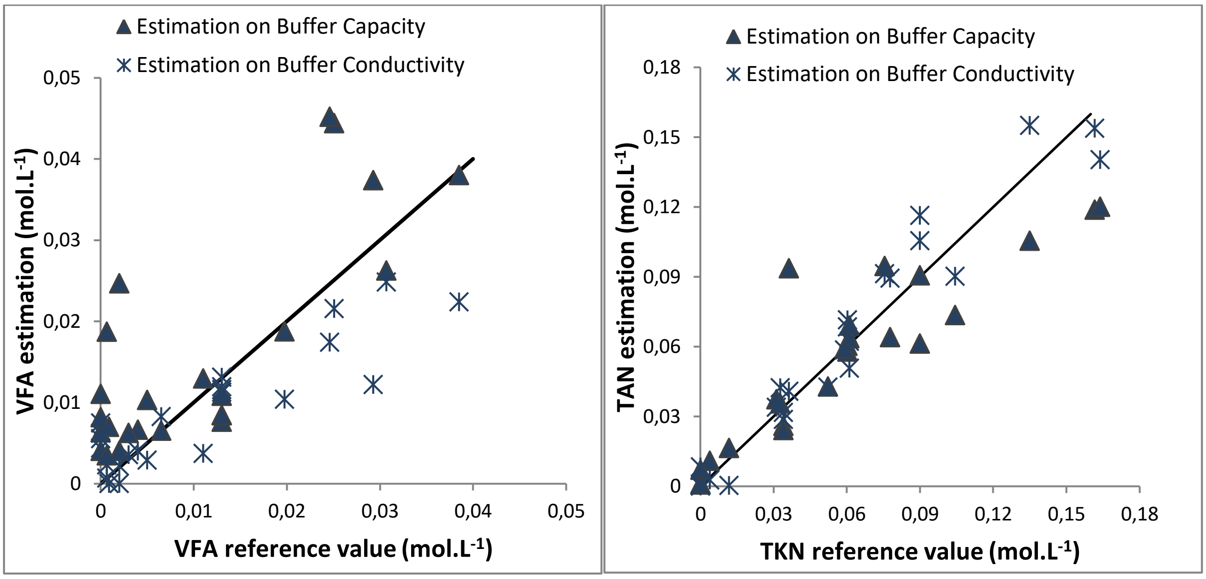


Figure 18: Comparison of the estimation on buffer conductivity and buffer capacity for VFA and TAN concentration on up titration

The Figure 18 shows that the estimations on buffer capacity overestimate VFA concentrations while the estimations based on buffer conductivity underestimate VFA concentrations. Actually, these data are complementary. As explained in "Materials and Methods" section 2.2.5.2, one measurement is charge-sensitive and the other is not. To go in details with this, an example is provided in the following. During up titration, VFA and a complex salt E-D are assumed to be present in the sample. Before up titration starts, under acidic conditions, the salt is dissociated to E^+ and D^-. The addition of sodium hydroxide induces the complexation of these salts into E-D, thus, two charges are lost decreasing the electrical conductivity. This complexation consumes sodium hydroxide and buffers the solution. Around the same pH value, VFAH is titrated to VFA^- increasing the electrical conductivity. This transformation from VFAH to VFA^- consumes sodium hydroxide too and buffers the solution. Hence, on one hand, buffer capacity data results from the addition of the effect of these two base-consumer components. On the other hand, the conductivity buffering data is the average of their opposite effects. Thus, working on both buffer capacity and buffer conductivity, we can manage to get back on the response of VFA only. It explains why buffer capacity and conductivity buffering together could be chosen to estimate the component concentration in case of complex solutions with many components. This effect was not clearly observed on TAN estimations. Indeed, even if the same effects were expected, the high concentration of TAN and low impact of other components around the pKa of ammonium made it not observable.

3.6.1 Building of a Decision Tree

Three unknown concentrations (*i.e.,* TAN, VFA and IC concentrations) have to be estimated by the algorithm. TAN and VFA concentrations are calculated with the buffer capacity and conductivity buffering on up titration. The estimations made on buffer capacity and conductivity buffering are compiled to remove the disturbance of other components. Eventually, the estimation of VFA concentration on down titration is discarded and IC is estimated from down titration. Thus, the algorithm provides an estimation of IC, VFA and TAN concentrations following the decision tree of Figure 19.

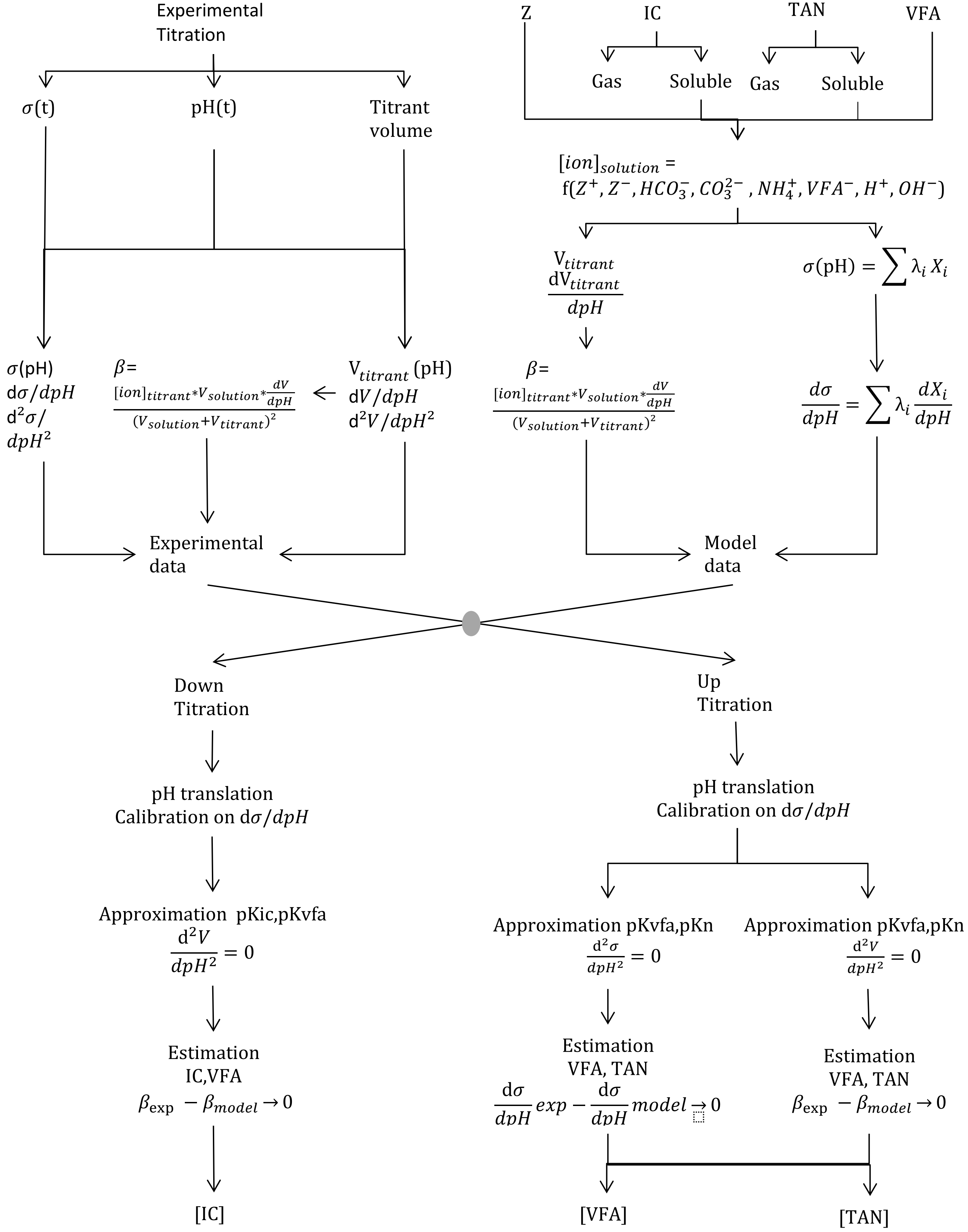

Experimental Titration
σ(t)
pH(t)
Titrant volume
σ(pH)
$d\sigma/dpH$
$d^2\sigma/dpH^2$
$\beta = \frac{[ion]_{titrant} * V_{solution} * \frac{dV}{dpH}}{(V_{solution}+V_{titrant})^2}$
$V_{titrant}$ (pH)
dV/dpH
d^2V/dpH^2
Experimental data
Z
IC
TAN
VFA
Gas
Soluble
Gas
Soluble
$[ion]_{solution} = f(Z^+, Z^-, HCO_3^-, CO_3^{2-}, NH_4^+, VFA^-, H^+, OH^-)$
$V_{titrant}$
$\frac{dV_{titrant}}{dpH}$
$\sigma(\mathrm{pH}) = \sum \lambda_i X_i$
$\beta = \frac{[ion]_{titrant} * V_{solution} * \frac{dV}{dpH}}{(V_{solution}+V_{titrant})^2}$
$\frac{d\sigma}{dpH} = \sum \lambda_i \frac{dX_i}{dpH}$
Model data
Down Titration
Up Titration
pH translation Calibration on $d\sigma/dpH$
pH translation Calibration on $d\sigma/dpH$
Approximation pKic,pKvfa
$\frac{d^2V}{dpH^2} = 0$
Approximation pKvfa,pKn
$\frac{d^2\sigma}{dpH^2} = 0$
Approximation pKvfa,pKn
$\frac{d^2V}{dpH^2} = 0$
Estimation IC,VFA
$\beta_{exp} - \beta_{model} \to 0$
Estimation VFA, TAN
$\frac{d\sigma}{dpH} exp - \frac{d\sigma}{dpH} model \to 0$
Estimation VFA, TAN
$\beta_{exp} - \beta_{model} \to 0$
[IC]
[VFA]
[TAN]

Figure 19: SNAC decision tree

3.6.2 Example of a Titration Using SNAC

Here is shown the example of the titration of a sample from a plant treating an industrial effluent from a paper mill wastewater. Observing when the derivative of buffer capacity and conductivity buffer goes to 0, it reveals 3 components: VFA, TAN and also one other component with pKa at 6.2 as indicated by the derivative of conductivity buffering and buffer capacity. The titration of this component induces a decrease of electrical conductivity interpreted as a loss of charge. Around pH 4.7, the titration of VFA increases the electrical conductivity. Thus, it increases the corresponding peak of buffer capacity and decreases the one of conductivity buffering (Figure 20). Around pH 9.2, the titration of TAN induces a decrease of electrical conductivity and an increase of the buffer capacity. The simulation of VFA and TAN with the reference values was performed. It can be observed that the curve representing the simulation made with the estimation of SNAC reproduces the signal of VFA and TAN without being disturbed by the third component. Figure 20 shows that combining buffer capacity and buffer conductivity, it is possible to estimate accurately the concentration of TAN and VFA in complex media with many other components.

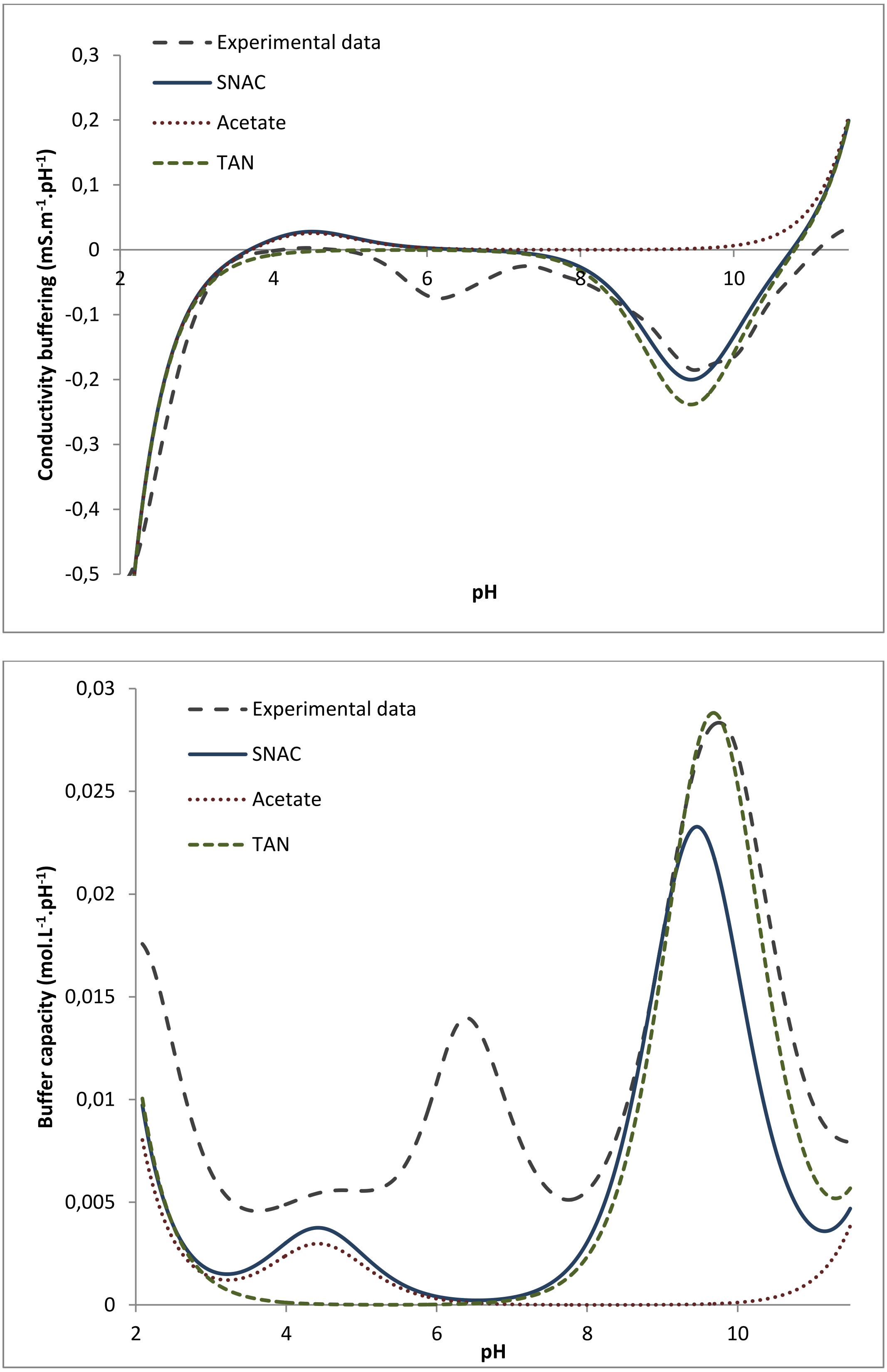


Figure 20: Titration of an industrial effluent, comparison of buffer capacity and buffer conductivity

3.6.3 Discussion of the Advantages between the Existing Methods and SNAC

It has been shown that key point analysis on the titrant volume is accurate to measure the main components which are mainly IC but does not provide accurate values in case of complex solutions to estimate the concentration of secondary components such as VFA because of its lower effect on the titration. Nonetheless, these methods are widely used because, with the assumption that the other components do not change, it provides a cost efficient tool to monitor the dynamics of VFA and IC concentrations (Figure 16). The use of the method developed by Van Vooren (2000), Van De Steene et al. (2002) and improved by Zaher (2005) provides more accurate values. Nonetheless, using buffer capacity alone, it is hard to distinguish and estimate all the components that could be present in a solution. In case of complex samples with many components, their signals overlapped each other too much to estimate them accurately. Combining electrical conductivity and pH sensors provides new information and allows improving the robustness and accuracy of titration sensors used in wide range of plant samples (Figure 21).

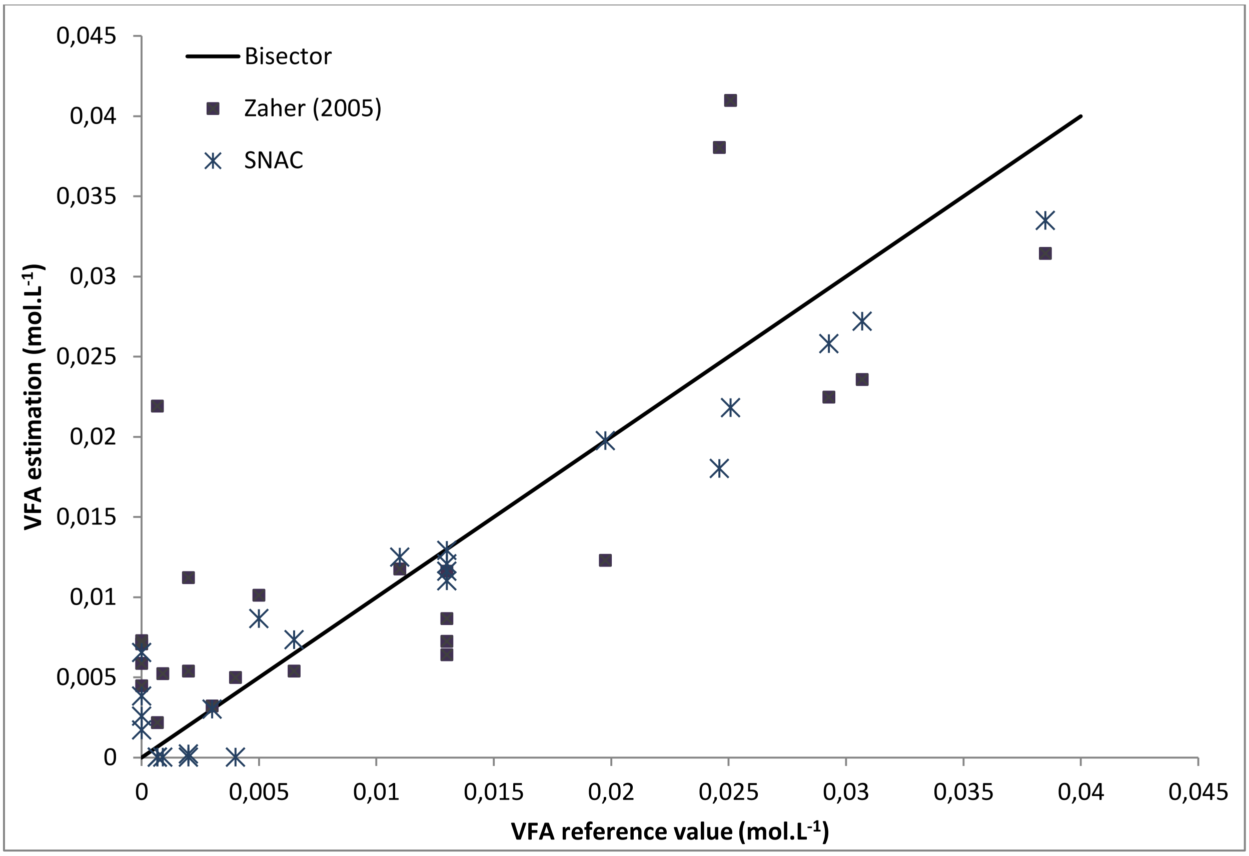


Figure 21: Comparison of VFA estimation using buffer capacity, Zaher (2005), or buffer capacity coupled to buffer conductivity (SNAC)

The fact that buffer conductivity is charge-sensitive and buffer capacity is not, allows the creation of a tool that provides accurate estimation of IC, VFA and TAN concentration for every AD plant without previous recalibration. As shown in Table 4, the standard errors were reduced to 0.003 mol.L^{-1} for VFA, 0.012 mol.L^{-1} for TAN and 0.015 mol.L^{-1} for IC concentration. It has been shown that implementing IC, VFA and TAN in the model coupling buffer capacity and buffer conductivity improved the accuracy of the estimation. Indeed as shown in Figure 20, coupling these two variables, the model estimates VFA and TAN without being disturbed by the other components. Nonetheless, improvements have to be made. First of all, the volume of sample titrated must be reduced to consume less acid and base during the titration. Finding the ideal volume can be challenging, because it must be as small as possible but still big enough to be representative from a non-homogenous plant. Second, if the improvement of the experimental signal accuracy succeeds, it would be interesting to come back on the model and implement other possible components or reactions to recreate with precision the experimental signal.

Table 4: comparison of accuracy of the estimation between the existing methods and SNAC on volatile fatty acids, inorganic carbon and total ammonia nitrogen

	Nordman (1977)	Ripley et al (1986)	Purser et al (2014) slurry energy	Zaher (2005)	**SNAC** (this study)
AGV					
R^2	0.02	0.05	0.03	0.59	**0.95**
Maximal error	0.177	0.124	0.092	0.021	**0.005**
Standard error	0.049	0.045	0.023	0.008	**0.003**
Standard error (% of the average)	458.5%	419.7%	212.3%	75.5%	**31.7%**
IC					
R^2	0.91	0.87	ND	0.87	**0.94**
Maximal error	0.048	0.088	ND	0.074	**0.002**
Standard error	0.019	0.044	ND	0.033	**0.015**
Standard error (% of the average)	8.8%	20.5%	ND	15.4%	**7.2%**
TAN					
R^2	ND	ND	ND	0.82	**0.95**
Maximal error	ND	ND	ND	0.067	**0.017**
Standard error	ND	ND	ND	0.026	**0.012**
Standard error (% of the average)	ND	ND	ND	16.1%	**7.3%**

3.7 SNAC Prototype

The aim of the SNAC method is to develop a sensor that enables an accurate estimation of the VFA, IC and TAN concentrations in an AD plant. To enable the use of the SNAC method at industrial scale, the method has been patented (n°15/59767) and a prototype has been done and is currently tested in an industrial plant. The prototype estimates the temperature, pH, conductivity and alkalinity ratio and VFA/IC/TAN concentrations of the plant. SNAC prototype has been designed to be simple of use (Figure 22). It is composed of:

- A titration vessel including a beaker easily transportable and a lid containing the probes and pipes
- 1 vial of acide and 1 vial of base
- 1 gas trap and 1 gas filtration vial, filtrating the gas and imposing a constant pressure
- 2 peristaltic pumps
- 1 fully automated acquisition and control system
- 1 tablet with a web interface designed in Rshiny that enables to control the titration

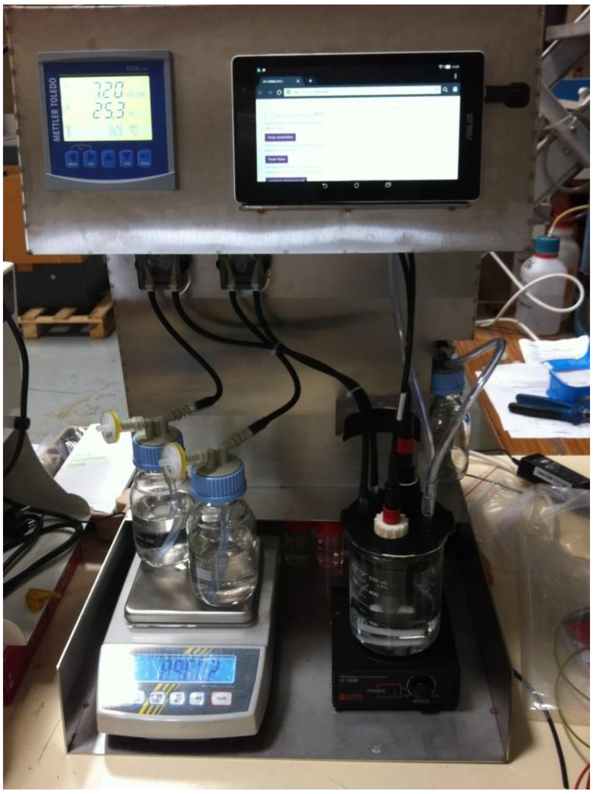
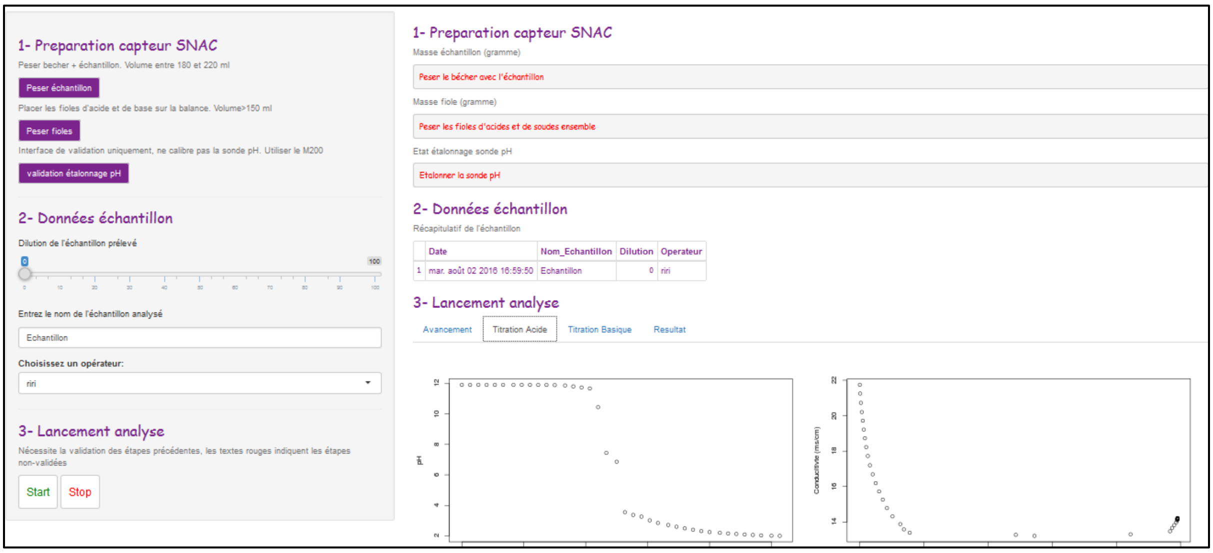


Figure 22: SNAC prototype, front picture.

The web interface (Figure 23) is available from tablets, smartphones or computers and is used to collect the information related with the titration like the sample name, date, volume, operator name…etc. It also allows launching and following the titration. Eventually, a titration report is automatically generated in pdf at the end of the titration which sum up the information related with the titration (Figure 24).


1- Preparation capteur SNAC

Peser becher + échantillon. Volume entre 180 et 220 ml

Placer les fioles d'acide et de base sur la balance. Volume>150 ml

Interface de validation uniquement, ne calibre pas la sonde pH. Utiliser le M200

2- Données échantillon

Dilution de l'échantillon prélevé: 0

Entrez le nom de l'échantillon analysé: Echantillon

Choisissez un opérateur: riri

3- Lancement analyse

Nécessite la validation des étapes précédentes, les textes rouges indiquent les étapes non-validées

1- Preparation capteur SNAC

Masse échantillon (gramme): Peser le bécher avec l'échantillon

Masse fiole (gramme): Peser les fioles d'acides et de soudes ensemble

Etat étalonnage sonde pH: Etalonner la sonde pH

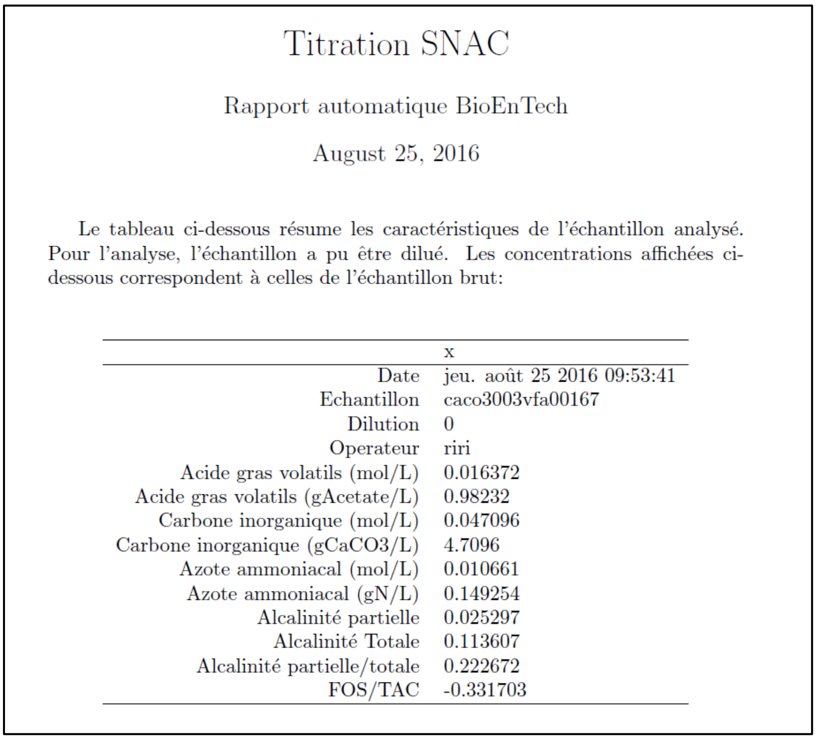
2- Données échantillon

Récapitulatif de l'échantillon

	Date	Nom_Echantillon	Dilution	Operateur
1	mar. août 02 2016 16:59:50	Echantillon	0	riri

3- Lancement analyse

Figure 23: SNAC web interface designed in Rshiny


Titration SNAC

Rapport automatique BioEnTech

August 25, 2016

Le tableau ci-dessous résume les caractéristiques de l'échantillon analysé. Pour l'analyse, l'échantillon a pu être dilué. Les concentrations affichées ci-dessous correspondent à celles de l'échantillon brut:

	x
Date	jeu. août 25 2016 09:53:41
Echantillon	caco3003vfa00167
Dilution	0
Operateur	riri
Acide gras volatils (mol/L)	0.016372
Acide gras volatils (gAcetate/L)	0.98232
Carbone inorganique (mol/L)	0.047096
Carbone inorganique (gCaCO3/L)	4.7096
Azote ammoniacal (mol/L)	0.010661
Azote ammoniacal (gN/L)	0.149254
Alcalinité partielle	0.025297
Alcalinité Totale	0.113607
Alcalinité partielle/totale	0.222672
FOS/TAC	-0.331703

Figure 24: Part of the SNAC titration report in French

3.8 Conclusion

The chapter draws four main conclusions:

- Estimations of VFA and IC based on key point analyses can be worth it to follow AD plants but provide non-accurate absolute estimation.
- Non-linear estimation based on buffer-capacity are more accurate than key point analyses and less sensitive to the composition of the sample.
- The combination of electrical conductivity and pH sensors has improved the accuracy estimating VFA, IC and TAN concentration with non-linear methods.
- Accurate estimation helps decision support tools but constant bias in the estimation still provide an accurate estimation of the plant dynamics.

Thanks to the prototype, SNAC method is currently tested on an industrial unit. Along with the biogas production and composition, it should enable us to estimate the biological state of the plant which is a required condition for plant performances prediction and feeding strategy optimization.

A missing step for feeding strategy optimization is the anticipation of the substrates AD performances. Hence, the next chapters will first focus on analyzing the relations between organic matter composition and AD performances; and on investigating fast methods of characterization for the estimation of AD performances.

Chapter 4. Relations between Biochemical Composition and Anaerobic Digestion Performances of Solid Waste

Chapter 4. Relations between Biochemical Composition and Anaerobic Digestion Performances of Solid Waste 81

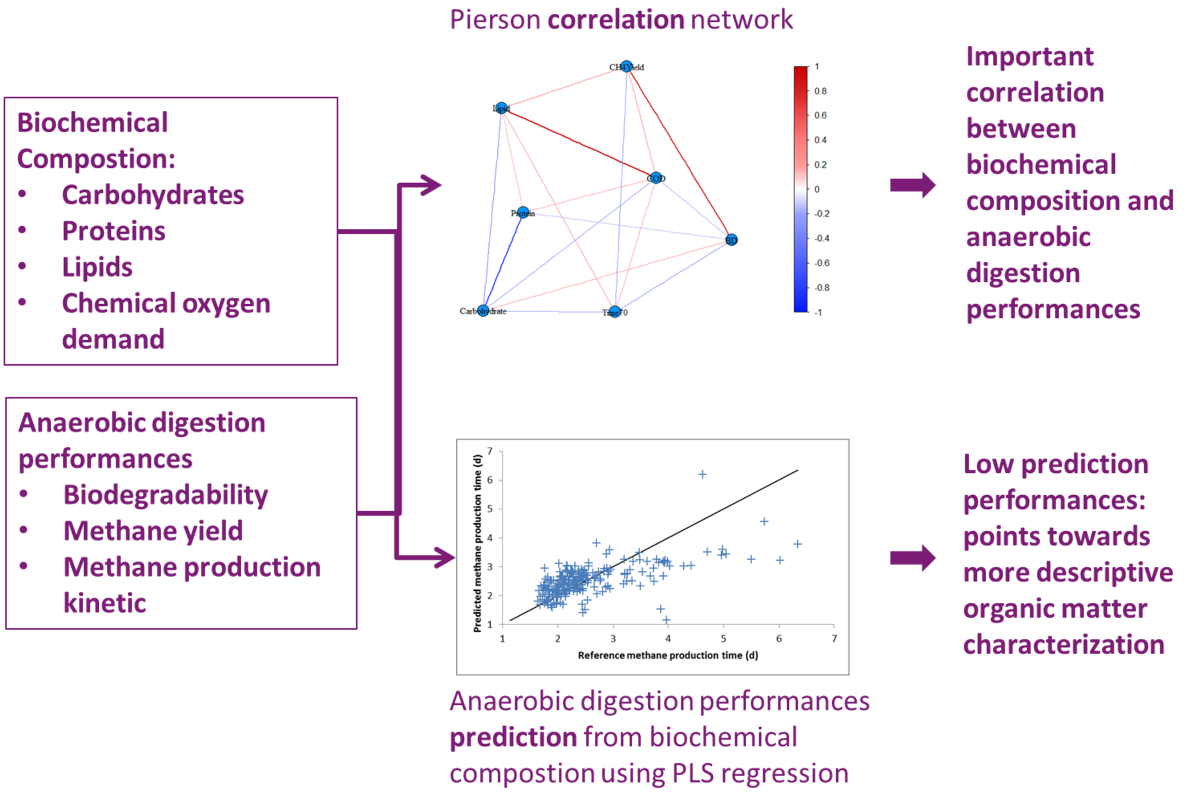
4.1 Chapter Guidelines 83
4.2 Technical and Graphical Abstracts 85
4.3 Introduction of the Chapter 87
  4.3.1 Analytical Methods for Biochemical Characterization of Substrates 87
  4.3.2 Biodegradability and Methane Production Rate Estimation 88
4.4 Materials and Methods Related to this Chapter 91
4.5 Validation of Biochemical Analyses on Solid Waste 91
4.6 Dataset Exploration Using a Principal Component Analysis 94
4.7 Variables Correlation Exploration Using a Principal Component Analysis 95
4.8 Variables Correlation Analysis using Pearson Correlation Coefficients 98
4.9 Discussion about Variables Correlation Analysis 99
4.10 Predictability of the Kinetics of Methane Production from Biochemical Composition 102
4.11 Conclusion on the Existing Relation between the Biochemical Composition and the Methane Production Performances 103

4.1 Chapter Guidelines

In Chapter 3, an accurate solution to estimate VFA and TAN in a plant sample has been proposed. Coupled with gas flowrate and its composition, it allows estimating the biological state of the plant, which is the first step of AD process performances prediction. A second key step consists in anticipating the AD performances of the substrate. It requires the assessment of both biochemical composition and methane production features of the substrate. Thus, this chapter presents, in section 4.3.1, the state of art on solid waste characterization and, in section 4.3.2, the existing methods for methane yield and kinetics estimation. Then a short reference to the “Materials and Methods” section is made in section 4.4. The existing methods about biochemical composition analyses are then tested and adapted to solid waste in section 4.5. In section 4.7 to section 4.9, correlations between AD performances and biochemical composition of the substrates are investigated using different methods, in order to help us understanding their links. Eventually, with the aim to reduce the time of analysis, the prediction performances of methane production kinetics based on biochemical composition are assessed (section 4.10).

4.2 Technical and Graphical Abstracts

Substrates for anaerobic digestion are composed of heterogeneous and complex organic matter. General parameters of the organic matter can be used to describe its composition such as carbohydrate, protein and lipid contents, Chemical Oxygen Demand (COD), biodegradability, methane yield and kinetics of methane production. These parameters are required for the monitoring of digesters but their characterization are time consuming and expensive. Moreover, laboratory techniques are not yet fully adapted to the analyses of solid waste. Thus, these parameters are rarely assessed all together. We first validated biochemical analyses on solid waste. Then, we investigated the existing correlations between COD, methane yield, biodegradability, kinetics of methane production and carbohydrate, protein and lipid contents on 222 samples using principal component analysis and Pearson's correlation coefficient. Our work aims at providing a support to human operators when only a partial characterization of the substrates is available. Positive correlation between methane yield and methane production kinetics but also unexpected independence of COD and methane yield are shown. Under the experimental conditions, carbohydrates, proteins and lipids modulated differently the anaerobic digestion performances. Carbohydrates are correlated with readily biodegradable organic matter while proteins are slightly anti-correlated with biodegradability. Lipids are correlated with high methane yield but slow methane productions. Finally, partial least square regression points towards more descriptive organic matter characterizations that could be used to predict the kinetics of methane production.



Pierson **correlation** network

Biochemical Compostion:
- **Carbohydrates**
- **Proteins**
- **Lipids**
- **Chemical oxygen demand**

Anaerobic digestion performances
- **Biodegradability**
- **Methane yield**
- **Methane production kinetic**

Important correlation between biochemical composition and anaerobic digestion performances

Low prediction performances: points towards more descriptive organic matter characterization

Anaerobic digestion performances **prediction** from biochemical compostion using PLS regression

4.3 Introduction of the Chapter

Among the parameters influencing AD processes, many result from the type of organic matter used and from the applied organic loading rate. Thus, a smart monitoring of the AD performances of a plant passes through its feeding strategy optimization. For example, ammonia is produced by the degradation of nitrogenous matter present in the substrates (eg., proteins). It has been seen in section 3.3.1 that a 3g.L^{-1} of TAN inhibit the process but besides, concentrations below 200 mg/L are beneficial to AD processes (Chen et al., 2008). The feeding of nitrogenous matter should therefore be controlled. Lipid substrates such as fat, oil and grease (FOG) also affect the performances of the process. Due to their high Biochemical Methane Potential (BMP), FOGs are usually reported as a way to improve digester performances (Long et al., 2012). However, digestion of waste with high lipid content causes inhibition of acetoclastic and methanogenic micro-organisms (Long et al., 2012). Operational problems also appear, such as flotation, linked to the low density of lipids. Knowing that the parameters that influence the biogas production are numerous, the characterization of carbohydrate, protein and lipid contents as well as the COD, methane yield and methane production kinetics are part of the main parameters required for optimizing digesters (Batstone et al., 2002).

Besides biogas production, digestate is the object of a rising interest due to its high added-value when used as fertilizer for crop farming (Madsen et al., 2011a). Its valorization is indeed being today integrated into agriculture as a key factor of AD profitability and environmental record. Organic matter in the digestate affects the mineralization and humification processes in soils (Lashermes et al., 2009). The fertilizing quality of the digestate therefore mainly depends on its organic matter and nitrogen composition (Massaccesi et al., 2013). Recently, environmental and economic problems were reported, related to digestate spreading into fields due to the incomplete degradation of the organic matter in the reactor before storage (Monlau et al., 2015). AD plant operators would thus benefit in maximizing the methane production yield while concomitantly reducing the biodegradable COD of the digested material (Ward et al., 2008). This can be achieved *via* a better management of the plant but also through the selection of appropriate substrates in the context of co-digestion (Alburquerque et al., 2012).

4.3.1 Analytical Methods for Biochemical Characterization of Substrates

Nowadays, methods for the estimation of the biochemical composition are still difficult to find (Jimenez et al., 2015b). Moreover, the recovery of COD from solid waste has been reported to be challenging. Different methods can however be found in the literature. The total lipid content is obtained after drying, which correspond to errors of approximately 10% depending on the laboratory. Howeverand grinding to powder (<1 mm), by soxhlet extraction as the percentage of heptane extractable materials (HEM) (Girault et al., 2012). The total protein content is measured with Folin reagent (Souza et al., 2013) following the protocol described in Lowry et al. (1951) or by the Kjeldahl procedure (Madsen et al., 2011). The latter is robust for the determination of nitrogen but tricky for the conversion of nitrogen into protein because the nitrogen content of protein depends on its nature (Izhaki, 1993; Jimenez et al., 2014). The total carbohydrate content is estimated by colorimetric measurements for which it is recommended to use the protocol described in DuBois et al., (1956) (Jimenez, 2012; Ugwuanyi et al., 2005). Finally, the estimation of COD using the international standard method ISO15705:2002 has been adapted for the estimation of soluble COD from solid waste (eg., freeze drying, grinding and suspending the substrate in distilled water before the COD analysis (Buffiere et al., 2008)).

The heterogeneity of the methods proposed in the literature to assess the contents of carbohydrate, protein, lipid or COD in solid samples suggests that many of them are substrate-specific and that the development of a robust and accurate method for a wide range of samples can be challenging. Similar problems are reported for the estimation of the methane production performances.

4.3.2 Biodegradability and Methane Production Rate Estimation

Methane production performances are commonly assessed from the Biochemical Methane Potential (BMP) which is obtained anaerobically by digesting a substrate in batch under controlled optimal conditions and over 30 to 50 days (Hansen et al., 2004). A solution containing the microbial inoculum and a medium composed of nutrients, micro-nutrients and vitamins is prepared. The solution is degassed and the inoculum activity is checked using ethanol or acetic acid. Once the inoculum activity is validated, the substrate to analyze is grounded with a controlled size and added to the solution (Angelidaki et al., 2009). The details of the BMP method vary, inducing large fluctuations in the results. For example, Raposo et al., (2011) reported an international inter-laboratory benchmarking study with measured methane yields of 350 ± 33, 350 ± 29, 380 ± 42, 370 ± 36 and 370 ± 35 mL $CH_4.g^{-1}{}_{VSadded}$ for starch, cellulose, gelatine and mung bean respectively, no significant impact of the inoculum concentration, inoculum age, inoculum source, working volume, temperature, mixing, headspace gas used to flush the bottles, pH and nutrient media were found. Despite several attempts of standardization, variations of ± 10% remain on the BMP value estimation, and recently, Bellaton et al., (2016) even estimated errors of ± 25%.

From the measurements of COD and BMP, the biodegradable part of the organic matter is calculated as the ratio between the theoretical and the experimental methane production as described in equation 18. This formula is also reported in (Jimenez et al., 2014) with X_{BD} the content of biodegradable organic matter.

$$X_{BD} = \frac{BMP}{COD*350} \tag{18}$$

Nevertheless, kinetics estimation remains a challenge and only few methods are available (Jensen et al., 2011). From the BMP digestion experiments described in (Angelidaki et al., 2009) or (Hansen et al., 2004), the methane production rate (MPR) can be calculated and the kinetics estimated. A procedure of kinetics estimation consist in optimizing the hydrolysis and disintegration kinetic parameter so that the ADM1 model fits the MPR (Yasui et al., 2008). However, Batstone et al., (2009) suggested that this approach underestimates the kinetics of methane production in continuous reactors because kinetics depends on the micro-organisms adaptation to the substrate (Figure 25). In order to address this issue, García-Gen et al. (2015) proposed a very interesting alternative running several successive batches using the same substrate and wait until MPR curves converge towards similar trends and values before estimating the kinetics.

No matter the methods used, the estimation of methane production kinetics requires long term experiments which makes it as one of the most difficult parameters to obtain. The procedures described in the literature, indeed, tend to be tedious and time-consuming, thus restricting their use at the industrial scale.

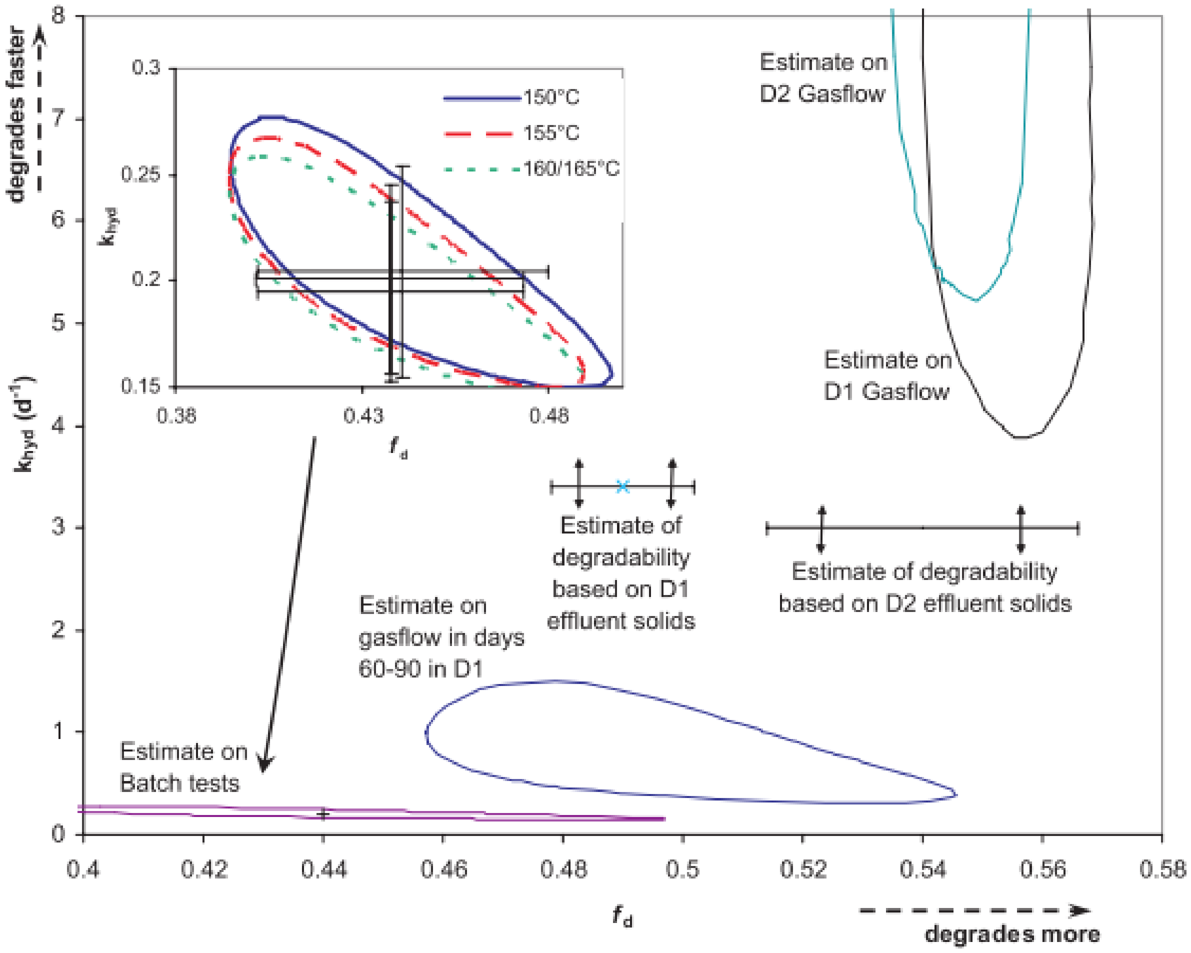


Figure 25: Confidence regions for k_{hyd} (day^{-1}), and f_d (frac) using different fitted data sets, including whole full-scale gasflow data sets (480 measurements for each digester), effluent solids (450 measurements for each digester), one example of a 30-day gasfllow data set (30 measurements), and the BMP tests (10 gas measurements in triplicate batches). The arrows around estimates on solids data indicate that the confidence regions are unbounded in k_{hyd}. The inset (top left) indicates a magnified view of the batch test confidence region *(Batstone et al, 2009)*.

According to the difficulty of estimation of the biochemical composition reported in section 4.3.1, this chapter firstly focuses on the development of repeatable and accurate analyses of the biochemical composition of a wide range of solid samples such as fruits, vegetables, manure, cereals, oils, fats, meat and fishes. Since no relevant solutions are available to shorten the estimation of the methane production kinetics, the second part of this chapter investigates the relation between the biochemical composition and the kinetics in order to better anticipate the methane production. In that aim, Principal Component Analysis (PCA) and Pearson's correlation were carried out to understand the existing correlations between carbohydrate, protein or lipid contents, COD, methane yield and the kinetics of methane production. Finally, PLS regression was carried out to evaluate the feasibility of methane production kinetics prediction from the biochemical composition of the substrates to facilitate plant monitoring.

4.4 Materials and Methods Related to this Chapter

Detailed information about the materials and methods related to this chapter are provided in the "Materials and Methods" sections 2.3.1, 2.3.3.1, 2.4.1, 2.5.1, 2.5.2, 2.6.1, 2.6.2, 2.6.3.1.

4.5 Validation of Biochemical Analyses on Solid Waste

First, the protocols described by Lowry et al., (1951) and Dubois et al., (1956) with a suspension of solid a sample in distilled water, regarded as standard methods, were applied to straw and potatoes. Table 5 suggests that these protocols under-estimate the concentration of carbohydrates in straw. All results were expressed in $mgO_2.gTS^{-1}$ equivalent in order to work with the same units for all methods. Grams of glucose were multiplied by 1065 and grams of BSA by 750, the calculations being based on their biochemical composition (Haag et al., 2003a). Grams of nitrogen were multiplied by 6.25*1.42 times the COD sample (Shi et al., 2014). Giving a closer look at these measurements, the carbohydrates were estimated at 185 $mgO_2.gTS^{-1}$ for straw out of a total COD of 1200 $mgO_2.gTS^{-1}$, thus representing about 15 percent of the COD. However, straw is mainly composed of carbohydrates such as cellulose and hemicellulose. According to the literature, its composition varies between 65 and 80% of the total COD (Bauer et al., 2009) (Motte et al., 2014). Contrastingly, the carbohydrate content in potatoes can reach up to 90%. These results lead to the assumption that colorimetric methods are not sufficiently robust. Samples with soluble and easily hydrolysable carbohydrates have been correctly estimated while others have been clearly underestimated. When using a suspension in water (Buffiere et al., 2008), the solubility of the sample in water is fundamental to ensure repeatable sampling. Moreover, the Dubois et al., (1956) method was developed on monosaccharides and readily hydrolysable polysaccharides such as starch but not on cellulose with glucose linked in β1-4 which are much more difficult to hydrolyze. Consequently, the solubilisation and hydrolysis of the solid sample seemingly needed to be improved. An efficient way to solubilize solid organic matter, in particular cellulose and hemicellulose (Jimenez et al., 2015a), is to induce an acid hydrolysis of the polymers using sulfuric acid. Therefore, in order to test its efficiency and accuracy, the protocol described by DuBois et al., (1956) with a pre-hydrolysis in sulfuric acid, was tested on glucose at 16 different known concentrations, straw, cellobiose and potatoes. Figure 26 illustrates how acid hydrolysis does not affect the estimation of glucose concentration.

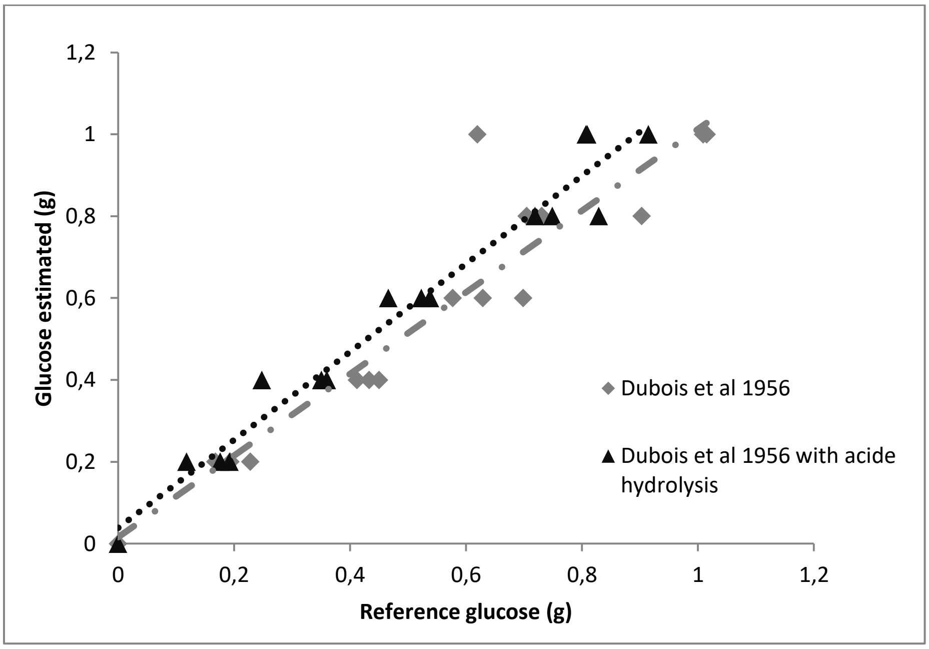


Figure 26: Effect of pre-sulfuric acid hydrolysis on glucose estimation

It has also been validated on cellobiose and starch, with carbohydrate concentration respectively of 94 and 104 % of their theoretical value. The results on straw and potatoes in Table 5 present a good repeatability of the analyses and results that are coherent with the known composition of straw (Bauer et al., 2009) (Motte et al., 2014).

Table 5 also presents repeatable and similar estimations of the proteins content using either the Lowry et al., (1951) protocol or TKN. TKN results were expressed in $mgO_2.gTS^{-1}$ equivalent in order to be comparable with Lowry et al, (1951) estimations. Both methods provided similar results. According to the previous results using colorimetric methods for carbohydrates estimations, there is a risk that proteins could be underestimated in certain samples using Folin reagent in the Lowry's method, mainly depending upon the solubility of the protein in water. Protein analysis using Folin reagent requires a really specific pH and is not compatible with sulfuric acid hydrolysis. Hence, to avoid any risk of protein content underestimation, the TKN method was preferred to the Lowry's protocol.

Finally, lipid estimation as HEM was validated with a less than 3% error, using commercial table oil for which the composition is known.

Table 5: Comparison of analysis methods for the estimation of proteins and carbohydrates. Selected methods are in bold.

	Carbohydrates: Dubois et al., (1956)	Carbohydrates: Dubois et al., (1956) with acid hydrolysis	Proteins : Lowry et al., (1951)	Proteins: Total Kjedhal nitrogen
Straw				
Mean value ($mgO_2.gTS^{-1}$)	185 ± 37%	**925 ± 10%**	55 ± 5%	**56 ± 3%**
Potatoes				
Mean value ($mgO_2.gTS^{-1}$)	922 ± 9%	**832 ± 6%**	76 ± 5%	**82 ± 10%**

For COD determination on solid samples, the experimental work was focused on the suspension of the solid samples in distilled water or sulfuric acid. First, distilled water suspension was used for COD analysis on six samples (manures, cereals, vegetables), with at least two replicates for each sample. On these samples, the Standard Error of Laboratory (SEL) was assessed at 200 $mgO_{2.}gTS^{-1}$. As the repeatability was not acceptable, the suspension of the solid samples in distilled water was assumed to lead to non-homogeneous sampling for COD analyses. Thus suspension in sulfuric acid was tried. The accuracy of the measurement was verified on glucose, cellobiose and starch for which the theoretical COD is known (Haag et al., 2003b). The measurement accuracies were respectively 100.5%, 99.7% and 99.9% of the theoretical values with a SEL on duplicates of 14 $mgO_2.gTS^{-1}$. It thus demonstrated that suspension in sulfuric acid does not denature the sample and that COD measurements after sulfuric acid suspension were accurate and repeatable.

Owing to these results, the choice of reference methods for the present study was directed towards the analysis of solid samples according to Dubois et al., (1956) with acid pre-hydrolysis for carbohydrates ($mgO_2.gTS^{-1}$). TKN was used for nitrogen content ($g.gTS^{-1}$) determination and heptane extractible matter for lipid content determination ($g.gTS^{-1}$). The choice was also made to use the suspension in sulfuric acid for the measurement of COD in all the samples. Once the methods validated, TKN, carbohydrates and COD were measured in duplicate on the whole dataset describing in section 2.5.1, the standard error of laboratory (SEL) was assessed on the duplicates using the same method as SDr (section 2.4.1) to $1.7*10^{-3}$ $g.gTS^{-1}$ and 63 $mgO2.g^{-1}$, 60 $mgO_2.\ gTS^{-1}$ for TKN, carbohydrates and COD respectively. The SEL corresponds for instance to 4.9% of the mean value of analyzed samples. The accuracy and repeatability of the protocols were thus assumed to be validated.

4.6 Dataset Exploration Using a Principal Component Analysis

Once the reference laboratory analyses validated, our work has been focused on investigating the existing relation between the biochemical and the AD performances of a substrate. A principal component analysis has been carried out on the dataset ("Materials and Methods" section 2.5.2) to explore the potential correlations between carbohydrate, protein and lipid contents, COD concentration, methane yield, biodegradability and methane production times which have been seen in the introduction (0) to be the most relevant parameters to characterize a substrate. We decided to focus on the three first principal components (PC) since they take into account more than 90% of the total variance. Figure 27 shows the samples of the dataset represented on the three PC of the PCA. Spheres has been colored depending on their biochemical composition, turquoise spheres have a lipid content above 50%, yellow spheres have a protein content above 50% and red spheres have carbohydrates content above 50%, grey spheres represent the remaining data. Four different groups of samples can be observed:

- The majority of the samples rich in carbohydrates are represented in the Ellipse Hoteling's T2 (95%).
- Another group is composed by two samples with a high score on PC1 and PC2. It corresponds to the FOG samples, characterized with a high content of lipids.
- Seven samples composed a third group, with high score on PC1 and PC3 but low on PC2, it matches the farm waste samples rich in carbohydrates and proteins.
- Finally, samples present a low score on PC1 and PC2 but a high score on PC3, corresponding to protein rich substrate like meat and fish extracts and micro-algae.

The three first components differentiate farm waste, FOG and meat and fish extracts or micro-algae from cereal, fruits and vegetables. Samples are fairly represented on the three PCs and samples seem to be classified depending on their carbohydrate, protein and lipid contents, with farm waste rich in proteins and carbohydrates; meat, fish or micro-algae rich in proteins; and FOG rich in lipids.

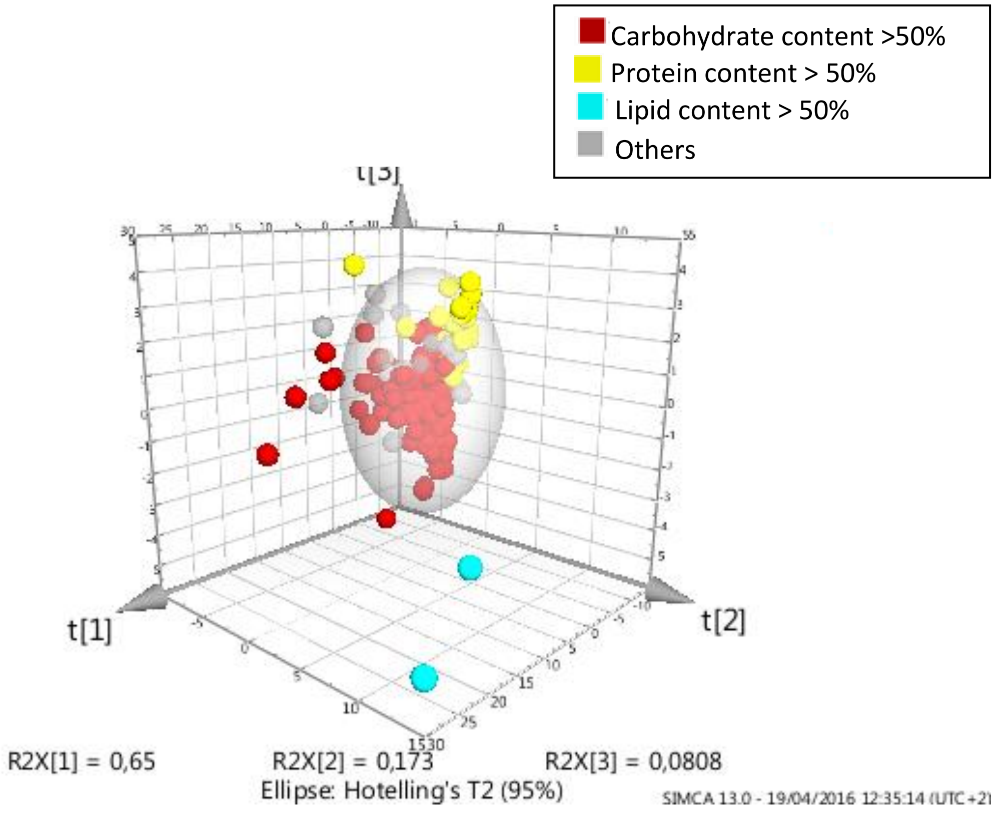
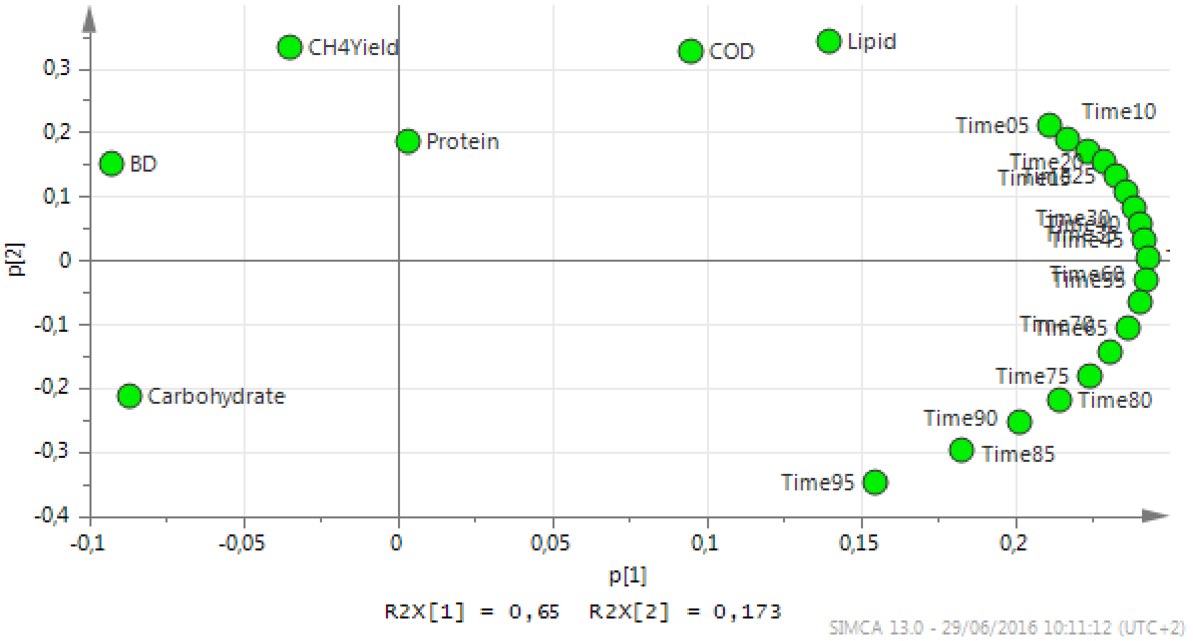


Figure 27: 3D representation of the dataset on the three first principal components of the PCA. Spheres are colored depending on their biochemical composition, turquoise spheres: lipid content up to 50%, yellow spheres: protein content up to 50%, red spheres: carbohydrate content up to 50%, grey spheres represent the remaining data.

4.7 Variables Correlation Exploration Using a Principal Component Analysis

To further investigate the correlations, PCA variables have been plotted on the PC1/ PC2 and PC1/PC3 (Figure 28). PC1, accounting for 65 % of the variance, is mainly supported by the methane production time. It shows correlations between the methane production times. These correlations make sense since the production time at a certain percentage of the methane yield is necessarily above the production times corresponding to lower percentages of the methane yield. Lipids and COD in a smaller manner also participate to PC1 and seem correlated with methane production times while carbohydrate content and biodegradability are in opposition with the methane production times in PC1. In PC2, methane yield, COD, early kinetics of methane production and lipid content are opposed to carbohydrate content and late kinetics of methane production. PC3 is supported by protein content which is in opposition to biodegradability, carbohydrate content and methane yield. To sum up the variable analysis on PCA, methane production times are represented as correlated to each other and represent, with lipid content, most of the variance. PC2 represents a smaller part of the variance and opposes proteins and carbohydrates. PC3 opposes biodegradability and carbohydrate content to protein content. Biodegradability is opposed to carbohydrate content on PC2. Their correlation on PC3 indicates that PC3 does not contain any discriminating information between carbohydrates and biodegradability. Biodegradability is shown as anti-correlated with lipid content and COD but correlated with methane yield on PC1-2-3.

Relations found using PCA will be validated by a PCC analysis in the part 4.8 before being discuss in part 4.9.



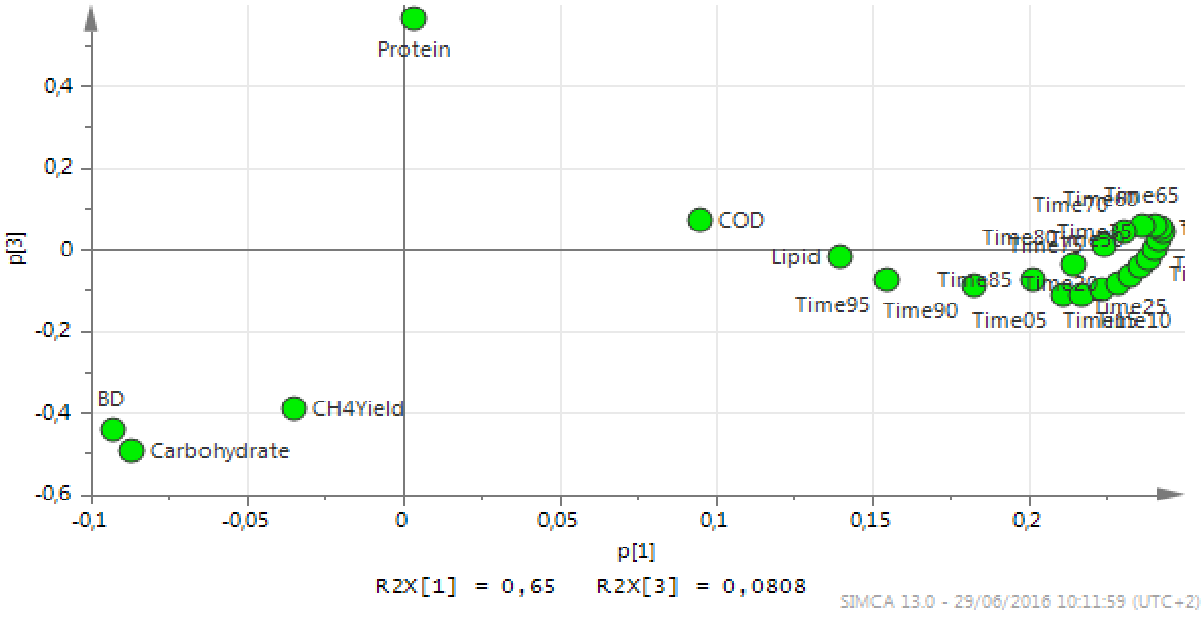


Figure 28: PCA variables representation on PC1/PC2 (up part) and PC1/PC3 (down part). BD and CH4Yield stand respectively for biodegradability and methane yield. Time 05 to 95 represent the times needed to reach a percent of the methane yield from 5 to 95%.

4.8 Variables Correlation Analysis using Pearson Correlation Coefficients

PCA has been used as an exploratory tool to assess potential correlations between carbohydrate, protein and lipid contents, COD, methane yield, biodegradability and methane production times expressed in days needed to reach a percent of the methane yield from 5 to 95 %. To validate these correlations, Pearson Correlation Coefficients (PCC) must be used (details about PCC are provided in "Materials and Methods" section 2.6.1). This led to a correlation matrix which has been plotted as a heatmap (Figure 29). From the p-values analysis, correlations revealed as non-significant are shown as non-existing correlation (i.e. PCC=0). Thus, every correlation shown in Figure 29, even low, is significant. As it can be seen in Figure 29 and PCA analysis, methane production times are correlated to each other. Nonetheless, weak correlations, around 0.2, are found between early and late methane production times. This indicates that the methane production is represented by at least two distinct kinetics as validated on sludge or fruits and vegetables AD by Jimenez et al., (2014) and García-Gen et al., (2015).

Relationships between biochemical composition and methane production times can be observed in Figure 29. Lipid content is correlated with methane production time from 5 to 70% of the methane yield with decreasing PCC respectively from 0.81 to 0.21. Carbohydrate content is anti-correlated with the methane production times from 5 to 75 % of the methane yield with PCC from -0.20 to -0.38. Protein content is anti-correlated with methane production times from 80 to 95 % of the methane yield with PCC of respectively -0.24, -0.28, -0.29. COD is correlated with methane production times, with decreasing PCC from 5 to 70 % of the methane yield in a range from 0.61 to 0.19, and anti-correlated with methane production time of 95 % of the methane yield (PCC: -0.23). Methane yield is weakly correlated with methane production times, with decreasing PCC from 5 to 15 % of the methane yield in a range from 0.21 to 0.13, and anti-correlated with methane production time, with decreasing PCC from 50 to 95 % of the methane yield in a range from -0.15 to -0.47. Biodegradability is anti-correlated with methane production times, with decreasing PCC from 5 to 95% of the methane yield in a range from -0.14 to -0.44.

Correlations and anti-correlations between the indicators of the biochemical composition of the organic matter are also shown in Figure 29. Lipid content is correlated with the methane yield, weakly with protein content and anti-correlated with the carbohydrate content with PCC of respectively 0.44, 0.16, -0.49. Carbohydrate content is anti-correlated with lipid and protein content and COD with PCC of respectively -0.49, -0.85, -0.40. Proteins content is weakly correlated with lipid content and COD with PCC of respectively 0.16, 0.19. Biodegradability is anti-correlated with COD with a PCC of -0.23. Eventually, methane yield is correlated with COD with a PCC of 0.35.

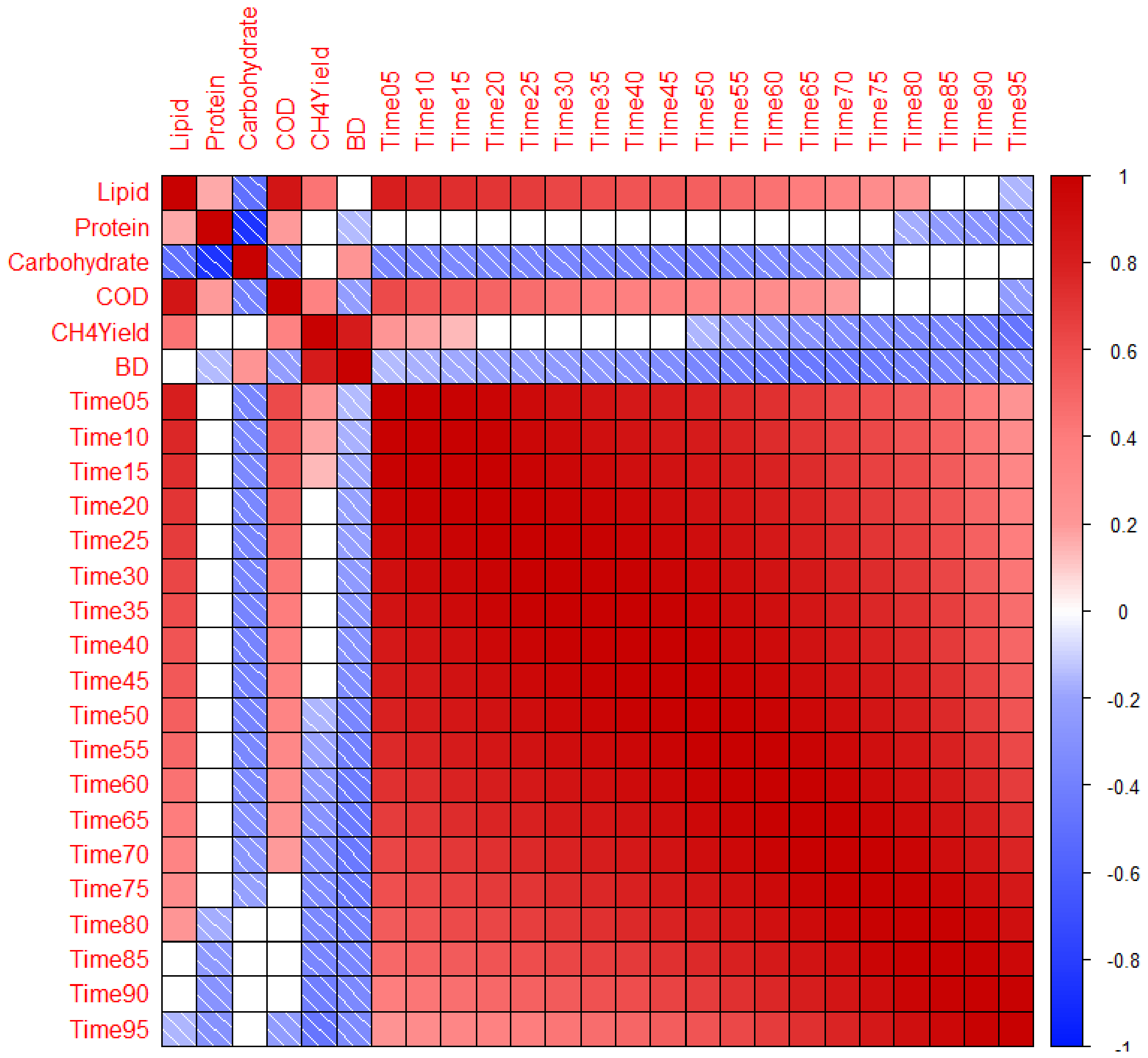


Figure 29: Matrix of Pearson's correlation coefficients between lipid, protein and carbohydrate contents, COD, methane yield (CH4YIELD) and biodegradability (BD), methane production time expressed in days needed to reach a percent of the methane yield from 5 to 95%.

4.9 Discussion about Variables Correlation Analysis

From PCA and Pearson's correlation coefficients, a first conclusion can be drawn about the existing correlations between lipid, protein and carbohydrate contents. Part of the organic matter such as lignin is not analyzed as carbohydrates, proteins or lipids. Thus, one hypothesis could be that biodegradable organic matter represented by carbohydrates, proteins and lipids are correlated between each other and as a consequence, anti-correlated with the non-biodegradable organic matter (eg., lignin). However, protein and lipid contents were correlated indicating that high-protein content in a sample is often correlated with important lipid content. Meat or micro-algae are good examples of the correlation between protein and lipid contents. But lipid and protein content were both anti-correlated with the carbohydrate content, which could be explained by the difference between animal cells with high protein content and vegetal cells with high carbohydrate content.

From PCA and PCC analyses, COD is strongly correlated with biochemical composition in carbohydrate, protein and lipid contents. Carbohydrates are shown to be highly biodegradable but with a lower COD than proteins which has a lower COD than lipids. COD and methane yield are both indicators of the energy contained in the organic matter and their ratio represents the biodegradability. Nonetheless, correlations found on COD cannot be extended to methane yield. The low correlation between COD and methane yield and the anti-correlation between COD and biodegradability observed indicate that methane yield and COD are quite independent parameters. Indeed, supposing a constant methane yield, the higher the COD, the lower the biodegradability. Hence, COD cannot be used to estimate the methane yield.

Correlations between methane production time and biochemical composition are complex. From Figure 6 (“Materials and Methods” section 2.5.1), it can be observed that the highest standard deviation between methane production time is observed at 70% of the methane yield. Thus, the time needed to reach 70% of the methane is the most informative point about methane production kinetics. In order to simplify correlation analyses, a correlation network between methane production time at 70% of the methane yield and carbohydrate, protein and lipid contents, COD, methane yield and biodegradability has been plotted in Figure 30. It reveals that methane production time at 70% of the methane yield is mainly correlated with lipid and COD and anti-correlated with carbohydrates, biodegradability and methane yield. This indicates that, within the set of waste analyzed, readily degradable organic matter is highly biodegradable and rich in carbohydrates and *vice versa*. Lipids and high COD compounds, which are strongly correlated, are found to be mainly slowly degradable. This conclusion was also drawn on the PCA even though, from Figure 27, the correlations with lipid content where shown to be supported by only two substrates. Therefore, these correlations have to be validated on a dataset containing more FOG.

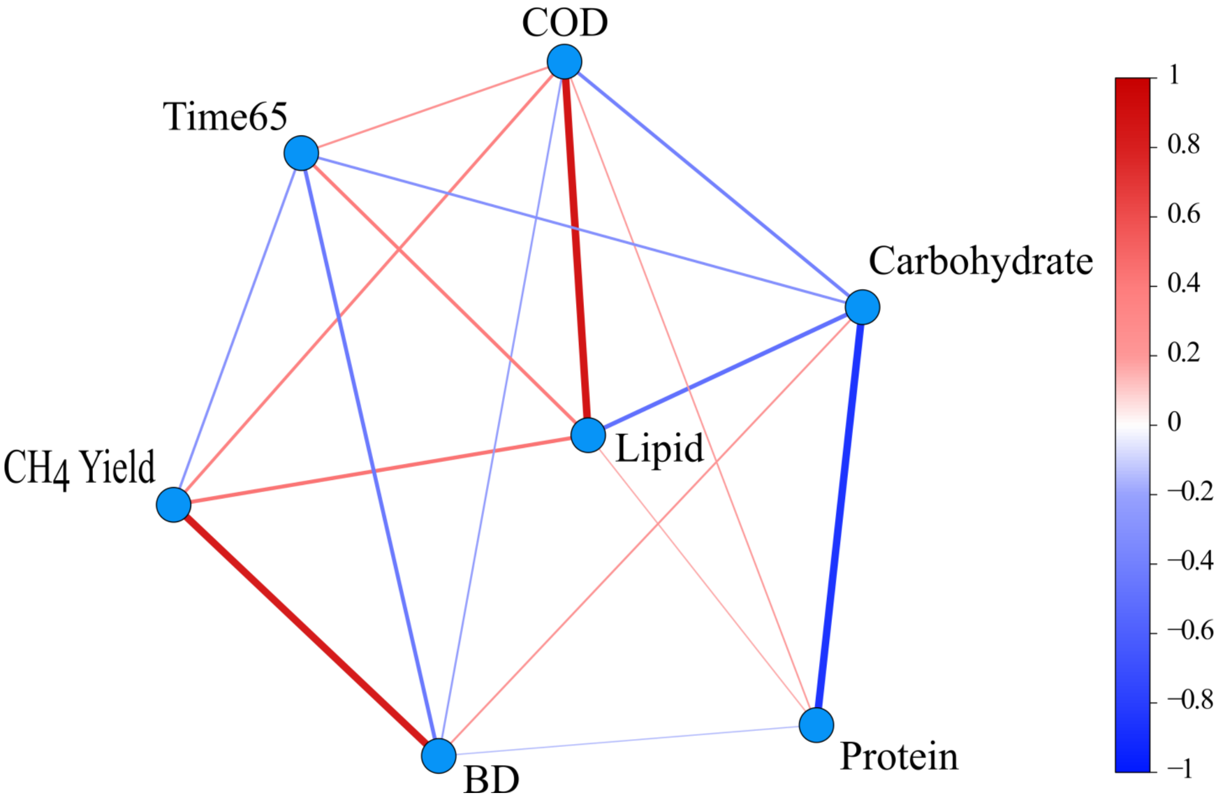


Figure 30: Correlation network between lipids, proteins and carbohydrates content, COD, methane yield (CH4YIELD) and biodegradability (BD), methane production time expressed in days needed to reach 70 percent of the methane yield. The level correlation is represented by the intensity and thickness of the link.

Figure 29 provides detailed information about kinetics of methane production. Methane yield, lipid content and COD have a negative impact on early methane production kinetics (*i.e.*, positive correlation with the time). Carbohydrate content affects positively the early methane production kinetics. Protein content is correlated with the time needed to reach the methane yield. Only biodegradability impacts positively the whole kinetics of methane production (i.e. the higher biodegradability, the lower methane production times). In conclusion, most of the substrates with a high content of carbohydrates were composed of readily degradable organic matter, such as glucose or starch that account for approximately 70% of the methane yield, and slowly degradable organic matter that account for the last 30% of the methane yield. According to Alibardi and Cossu, (2016), readily biodegradable carbohydrates are carbohydrates reserve of fruits and vegetables while slowly biodegradable carbohydrates are structuring carbohydrates such as cellulose. The impact of lipids on the methane production time could be explained by a first phase in which the anaerobic digestion process is inhibited by volatile fatty acids and long chain fatty acids, followed by a quick methane production once the inhibitions is overcome.

4.10 Predictability of the Kinetics of Methane Production from Biochemical Composition

An important question arises from the correlation observed between the biochemical composition and methane production time: could methane production kinetics be anticipated from the biochemical composition? In the introduction (4.3.2), kinetics estimation has been reported as labor-intensive and time-consuming, whereas biochemical composition was said to be simpler to assess. Thus, prediction of the kinetics from the biochemical analysis is of great interest. A PLS regression has been carried out in order to predict the time required to reach 70% of the methane yield that was demonstrated to be the most informative methane production time. Carbohydrate, protein and lipid contents together with COD, methane yield and biodegradability were used as input variables of the PLS regression model. The Standard Error of Cross Validation (SECV) reaches a minimal plateau from 4 to 6 latent variables. Thus 4 latent variables were selected for the PLS regression model. The PLS regression provides a SECV of 0.65 days and a Standard Error of Calibration (SEC) of 0.62. The results between estimated and reference production time can be observed in Figure 31. The coefficient of determination R^2 was about 0.38. These results show that even if correlations exist between methane production times and biochemical composition of the organic matter, no prediction of the methane production times or kinetics can be made from the biochemical composition indicators that are currently used. Indeed, it can be expected that macro-indicators of the biochemical composition such as lipid, protein and carbohydrate contents or COD, methane yield and biodegradability are not precise enough. From the correlation analysis of the variables, the study already revealed that, depending on the types of carbohydrates, they can be either slowly or readily biodegradable. It can thus be supposed that, with a more detailed organic matter method of characterization, the methane production time could be predicted from the biochemical composition analysis.

Near InfraRed spectroscopy (NIRs) could help in this direction. NIRs is indeed known to give insights about the organic matter composition and the quantitative prediction of the organic matter biochemical composition in carbohydrates, proteins and lipids (Núñez-Sánchez et al., 2016). NIR absorption bands are composed of overtones and combinations of the molecular vibrations of C–H, N–H, and O–H bonds which are correlated with molecules such as phenol, acids, alkane, amines, amides and celluloses content (Ward, 2016; Williams and Norris, 2001). Moreover, NIR spectroscopy is widely used for methane yield prediction (Lesteur et al., 2011). It is a powerful tool for anaerobic digestion process monitoring (Jacobi et al., 2012; Krapf et al., 2013). Thus, it can be assumed that NIR spectroscopy could provide an additional and accurate description of the organic matter. This will be assessed in the next chapter, trying to correlate NIR spectra with methane production kinetics.

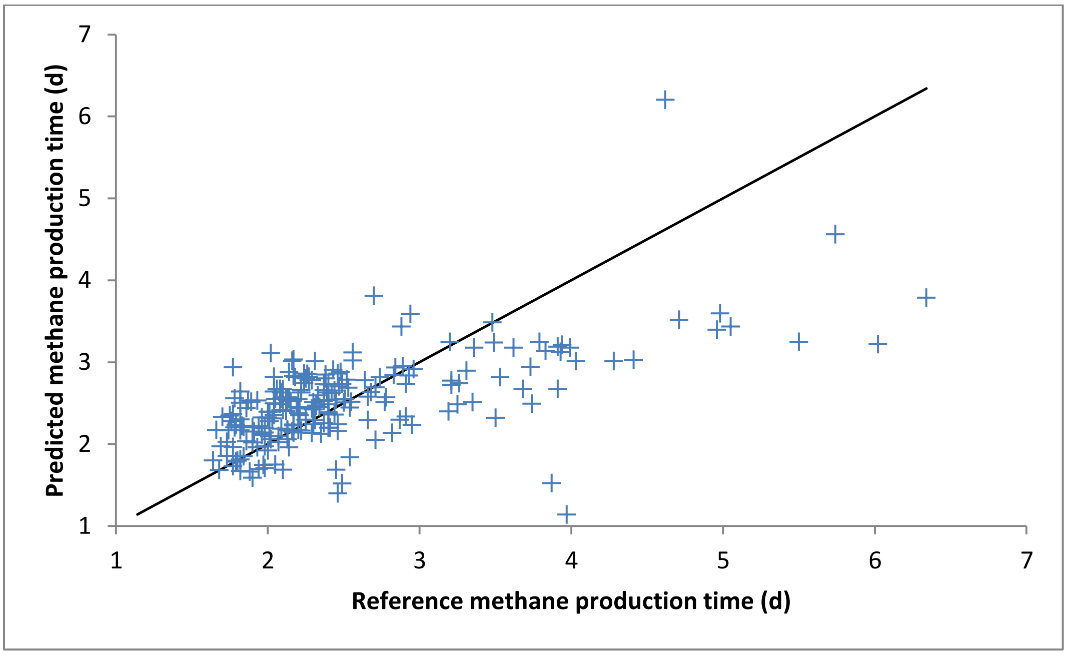


Figure 31: Prediction of methane production time needed to reach 70% of the methane yield using carbohydrate, protein, lipid contents, COD, biodegradability and methane yield as variables of PLS regression model.

4.11 Conclusion on the Existing Relation between the Biochemical Composition and the Methane Production Performances

Correlations between methane production times, methane yield, COD, biodegradability, carbohydrate, protein and lipid contents have been assessed using PCA and Pearson correlation coefficients analysis. It revealed that, among the 222 samples analyzed, carbohydrates are mainly readily biodegradable while lipids are correlated with high methane yields but also with slow methane production kinetics which could be related to process inhibition in the early methane production times. A negative correlation between COD and biodegradability indicates that COD cannot replace methane yield analysis. Eventually, correlations between biochemical composition and kinetics of methane production were highlighted leading to the idea that the kinetics of methane production could be predicted from the biochemical composition. However, the indicators currently used to assess biochemical composition provide a vague description of the organic matter which is not sufficient to accurately predict kinetics of methane production using PLSr. Nonetheless, it points towards more descriptive organic matter characterizations such as NIR spectroscopy that could be used to predict the kinetics of methane production.

Chapter 5. Characterization of the Solid Waste Biochemical Composition and Anaerobic Digestion Performances Using Spectroscopic Analyses

Chapter 5. Characterization of the Solid Waste Biochemical Composition and Anaerobic Digestion Performances Using Spectroscopic Analyses 105

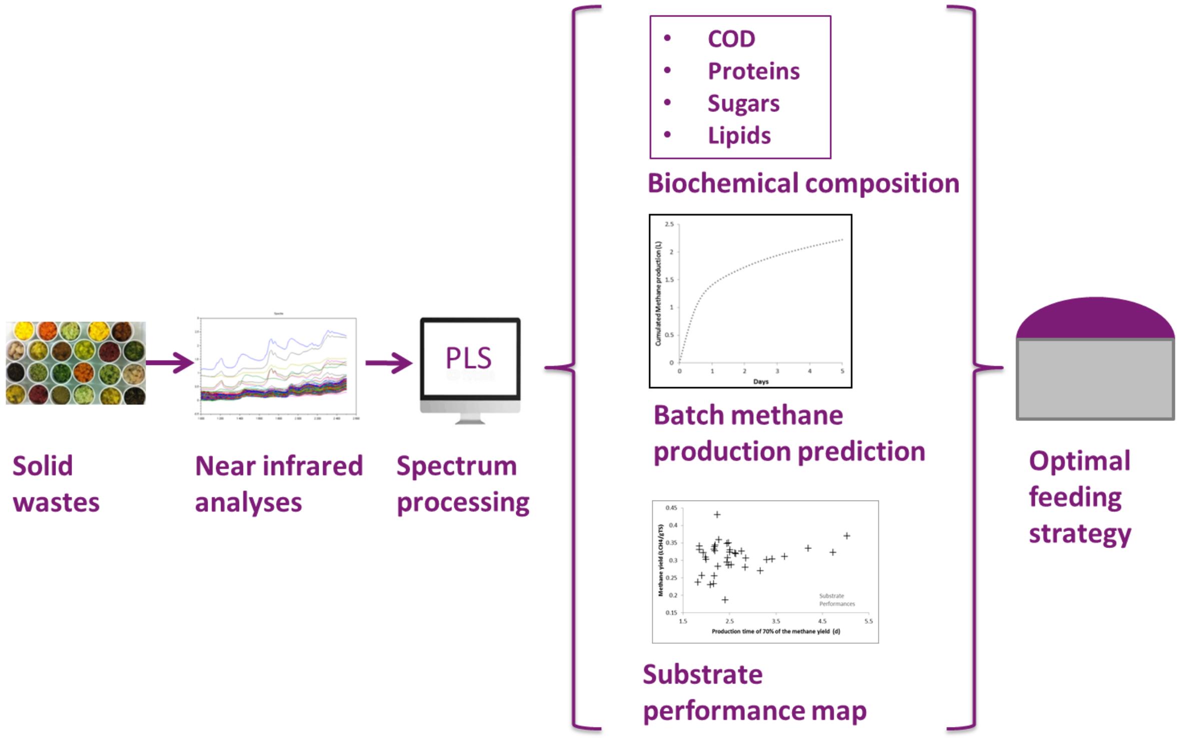
5.1 Chapter Guidelines 107
5.2 Technical and Graphical Abstracts 109
5.3 Introduction of the Chapter 111
  5.3.1 BMP Estimation on Early Biogas Production 111
  5.3.2 Spectroscopic Analyses 114
5.4 Materials and Methods Related to this Chapter 116
5.5 Comparison of 3D Fluorescence and Near Infrared Spectroscopy for Biodegradability Prediction 117
5.6 Fast Characterization of Solid Waste Biochemical Composition with Near Infrared Spectroscopy 121
  5.6.1 Combinations of the Calibration Model Using PLS Regression 121
  5.6.2 Model Validation on an Independent Dataset 124
  5.6.3 Discussion Regarding the Selected Model 124
5.7 Prediction of the Methane Production Rate in Batch Conditions Using NIR Spectroscopy 127
  5.7.1 Prediction of the Methane Yield 127
  5.7.2 Prediction of the Methane Production Time 127
  5.7.3 Methane Production Rate Prediction 132
  5.7.4 Simple Indicator of Methane Production Performances 132
5.8 Confidence Interval Determination Using Spectrum Distance from the Calibration Set 133
5.9 Conclusion on NIRs Characterization 139

5.1 Chapter Guidelines

This chapter follows the work started in Chapter 4 in terms of assessing methane production performances in a shorter time. Chapter 4 highlights that a correlation between biochemical composition and AD performances exists. It supposes that the prediction of methane production kinetics based on biochemical composition of the substrate is feasible but requires a deep characterization of the organic matter. The state of art developed in section 0 and the results of Chapter 4 point towards spectroscopic methods as new methods for a fast characterization of the substrate. A short reference to the “Materials and Methods” section is made in section 5.4. The following section, 5.5, compares the performances of prediction of the different reported spectroscopic methods on a reduced dataset. Based on the results of section 5.5, the performances of prediction of the most promising methods are assessed for the prediction of the carbohydrate, protein and lipid contents, COD (section 5.6) and methane production yield and kinetics (section 5.7).

5.2 Technical and Graphical Abstracts

The development of anaerobic digestion involves both co-digestion of solid waste and optimization of the feeding recipe. Within this context, substrate characterisation is an essential issue. The biochemical composition in carbohydrates, proteins and lipids, the chemical oxygen demand and both methane production yield and kinetics are required characteristics of the inputs to optimize the feeding strategy. Fast estimation of methane production yield has been developed using near infrared spectroscopy. Although widely used, the methane production yield alone is not sufficient to optimize the operation of anaerobic digestion plants. Moreover, the estimation of the biochemical composition and especially of methane production kinetics is labor-intensive and requires time-consuming analyses. Here we show that near infrared spectroscopy allows a fast and efficient estimation of the contents in carbohydrate, lipid and nitrogen, of the chemical oxygen demand and of methane production yield and kinetics, in solid waste. We build a Partial Least Square regression model to predict the biochemical composition and the methane production time, on which methane production kinetics is calculated. The model is calibrated with 275-295 solid waste samples and validated with an independent set of 43 samples across a wide range of solid waste found in anaerobic digestion units. The standard errors of cross-validation are 90 mgO$_2$.gTS^{-1} carbohydrates, 2.5*10^{-2} g.gTS^{-1} lipids, 7.2*10^{-3}g.gTS^{-1} nitrogen, 99 mgO$_2$.gTS^{-1} chemical oxygen demand and 0.34 days methane production time. The standard errors of prediction are 35 mgO$_2$.gTS^{-1} carbohydrates, 9.4*10^{-3} g.gTS^{-1} lipids, 5.2*10^{-3}g.gTS^{-1} nitrogen, 39 mgO$_2$.gTS^{-1} chemical oxygen demand and 0.24 days methane production time. These results show that near infrared spectroscopy is a new fast and cost-efficient way to characterize solid waste content and improve their anaerobic digestion monitoring.


COD
Proteins
Sugars
Lipids
Biochemical composition
PLS
Solid wastes
Near infrared analyses
Spectrum processing
Batch methane production prediction
Cumulated Methane production (L)
Days
Methane yield (LCH4/gTS)
Production time of 70% of the methane yield (d)
Substrate Performances
Substrate performance map
Optimal feeding strategy

5.3 Introduction of the Chapter

Standard characterization methods have been carried out on carbohydrate, protein and lipid contents and COD for solid organic waste in Chapter 4. Nonetheless, it has been seen that the estimation of the methane production yield and kinetics remains complicated and time consuming. Hence, shorten their estimation is a hindrance to overcome for feeding strategy optimization in an industrial purpose (Batstone et al., 2015).

5.3.1 BMP Estimation on Early Biogas Production

Fast biological tests have been developed for BMP or biodegradability estimation, hypothesizing a relationship between aerobic and anaerobic biodegradability. Since aerobic reactions are faster than anaerobic ones, several aerobic tests have been indeed used to quickly estimate the anaerobic biodegradability of the substrate (Lesteur et al., 2010). For instance, Cossu and Raga, (2008) successfully correlated (R^2=0.80) the aerobic respiration at 4 days index (RI_4) with the anaerobic biogas production at 21 days (GB_{21}) on landfill waste and concluded that RI_4 is faster and has to be preferred for biodegradability estimation (Figure 32).

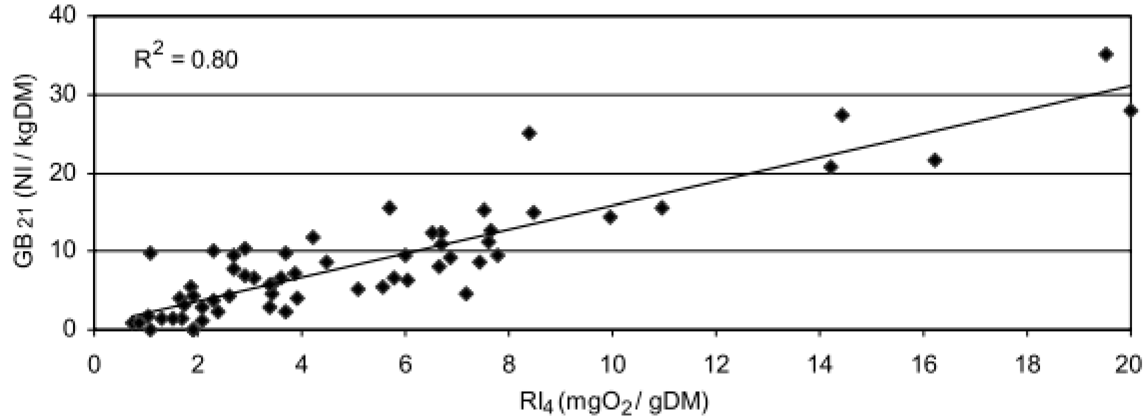


Figure 32: Correlation between respiration index (RI4) and biogas production (GB21) measured in samples of excavated waste from landfills *(Cossu and Raga, 2008)*.

Other tests are based on the assessment of the BMP through the early micro-organism activity during the anaerobic digestion of the sample. A first method assumes that the batch methane production follows a specific kinetics model. Model parameters are calibrated on a reliable set of initial data using first order, Monod or Gompertz kinetic terms. Then, for each new sample, the model previously calibrated is fitted on the first 3 days of methane production to extrapolate the total methane production (Figure 33). BMP and other key information such as substrate toxicity can then be calculated (Strömberg et al., 2015).

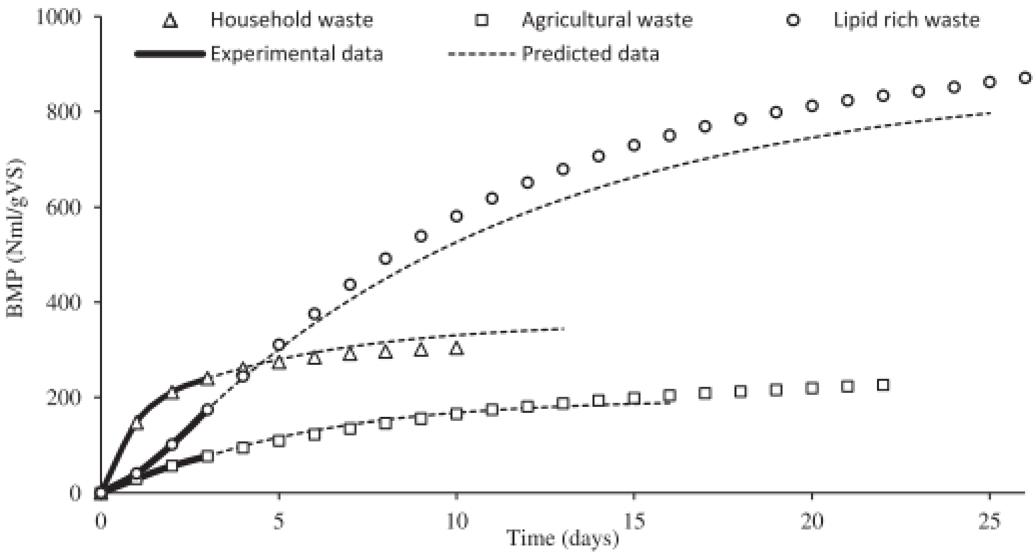


Figure 33: Accumulated specific methane production for three different samples (i.e. mixed food waste as a household waste, banana stem as an agricultural waste, and butter residues as a lipid rich waste). The experimental data is represented by the solid line, waste, and butter residues as a lipid rich waste). The experimental data is represented by the solid lines and the symbols, while the predicted gas profile is shown by the dashed line *(Strömberg et al., 2015)*.

A second method – the Envital® kit developed by the French company Envolure – correlates linearly BMP values with early microbial activity (Figure 34). A component named resazurin is introduced in the digestion broth and reduced by the micro-organisms into resofurin, which fluoresces at 600 nm. The fluorescence is recorded every 30 min for 48h. Bellaton et al., (2016) confirmed the correlation between the kinetics of resazurin reduction and the metabolic activity of anaerobic sludge. BMP measurements from Enivtal® kit and Automatic Methane Potential Test System II (AMPTS II, Bioprocess Control AB, Lund, Sweden) were compared on 18 wastewater sludge samples and similar results were obtained, supposing however 25% of error on AMPTS II estimations (Bellaton et al., 2016).

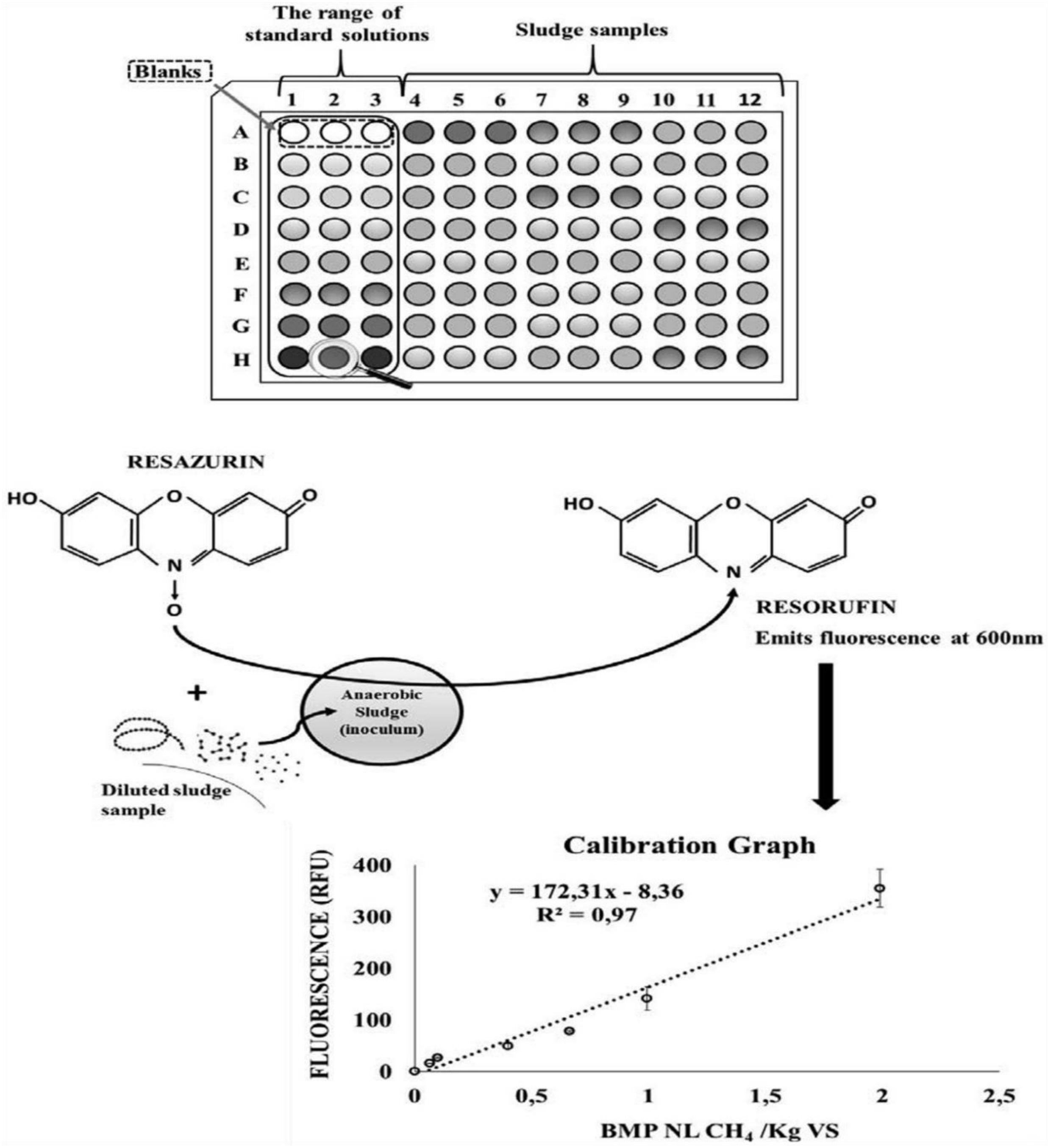

The range of standard solutions
Sludge samples
Blanks
1 2 3 4 5 6 7 8 9 10 11 12
A
B
C
D
E
F
G
H
RESAZURIN
HO
O
N
RESORUFIN
Emits fluorescence at 600nm
+
Anaerobic Sludge (inoculum)
Diluted sludge sample
Calibration Graph
y = 172,31x - 8,36
R² = 0,97
FLUORESCENCE (RFU)
400
300
200
100
0
0
0,5
1
1,5
2
2,5
BMP NL CH4 /Kg VS

Figure 34: Scheme of an example of Envital® assay in microplate based on fluorescence measurement *(Bellaton et al., (2016))*.

5.3.2 Spectroscopic Analyses

Several other methods have been proposed using spectroscopic measurements to obtain a fast estimation of BMP values. For example, Muller et al., (2014) developed an informative and quick method for the determination of the complexity of organic matter using chemical sequential extraction and 3D fluorescence spectroscopy. These authors calibrated their method by decomposing the 3D spectra into seven zones (zone I to VII) corresponding to families-like molecules (Figure 35).

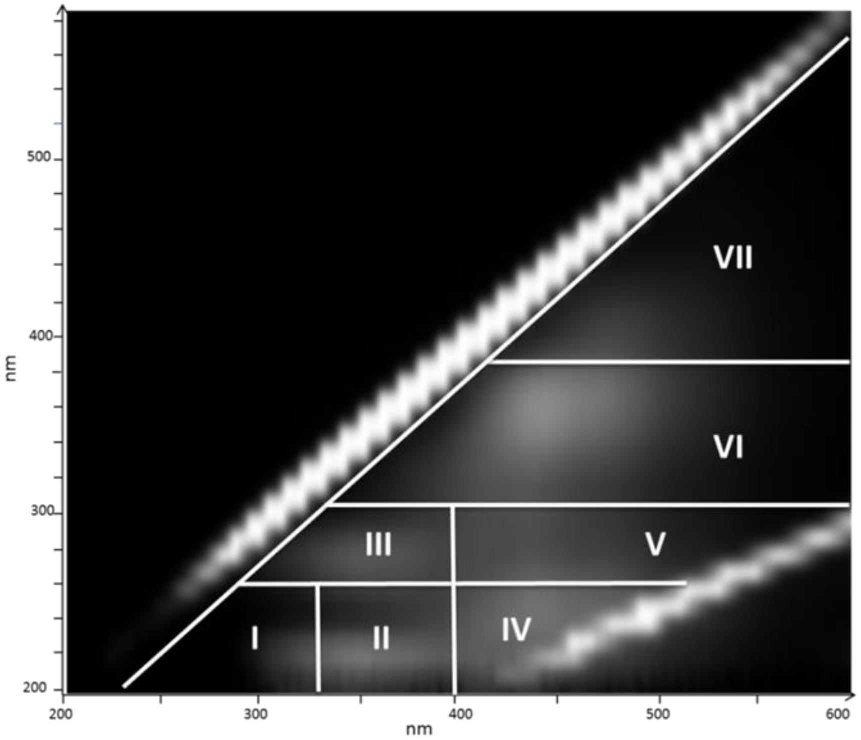


Figure 35: Molecule families-like in 3D spectra of manure based on (Muller et al, 2014). I: Tyrosine-like II: Tryptophan-like III: Microbial product protein-like IV: Fulvic acid-like V: Glycolated protein-like VI: Lignocellulose-like VII: Humic acid-like

Jimenez et al., (2014) predicted the biodegradability of municipal sludge (Figure 36) from the mean intensity of the seven zones corresponding to the families-like molecule shown in Figure 35, and their bioaccessibility from the chemical sequential extraction. Jimenez et al., (2015a) added a new step in the chemical sequential extraction for the characterization of a wider range of organic substrates. They suggested that chemical sequential extraction coupled with 3D fluorescence spectroscopy could be a powerful tool for the prediction of both biodegradability and bioaccessibility, thus reducing the time of analysis from 30-100 days to one week. Since fluorescence emissions are mainly due to aromatic cycle, this work could be correlated with Buffiere et al., (2006) who suppose a link between lignin content, composed of many aromatic cycles and biodegradability. 3D fluorescence spectroscopy was also used by Wan et al., (2012) to monitor the COD removal of an AD plant.

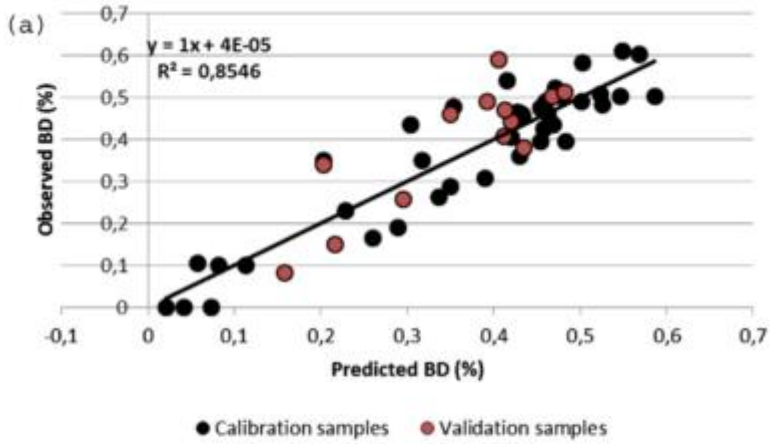


Figure 36: Observed versus predicted biodegradability using chemical sequential extraction and 3D fluorescence *(Jimenez et al., 2014)*.

Another fast method has been recently proposed using NIR spectroscopy (Lesteur et al., 2011, Jacobi et al., 2012), reducing the time of BMP measurements from 30-100 days to a matter of minutes (Ward, 2016). BMP indeed depends on the bio-chemical composition of the organic matter (Buffiere et al., 2006) (Appels et al., 2011) and NIR spectroscopy (NIRs) is already used in the agro-food industry to predict the biochemical composition of samples such as milk, cereals, fruits or vegetables (Bagchi et al., 2016, Núñez-Sánchez et al., 2016). It has therefore been hypothesized that protein, carbohydrate and lipid contents of waste could be assessed by NIR spectra, together with COD concentration.

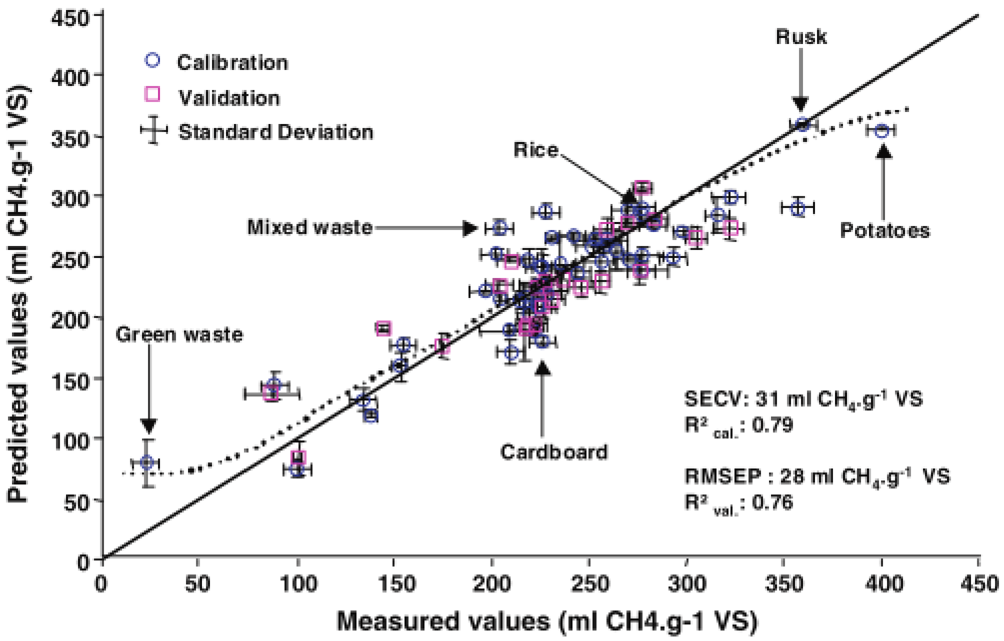


Figure 37: Predicted using NIR spectra vs measured BMP values. The solid line indicates the 1:1 relationship. For each point, horizontal lines indicate the Standard Deviation of the BMP test of the duplicates samples and the vertical lines

indicate the Standard Deviation of the NIR measurement of the triplicates samples. The dotted line represents the Dunne effect. Calibration data set are represented by blue circles and validation data set by purple squares *(Lesteur et al., 2011)*.

The challenge associated to NIRs analysis in AD is the complexity of the spectrum resulting from both chemical and physical interactions. This needs to be correlated with the biochemical composition of a large diversity of samples. Mathematical treatment of the spectrum has to be carried out (Bertran et al., 2001) . Detrending, Savitzky–Golay filtering and Standard Normal Variate normalization (SNV) are widely used for spectrum processing (Nicolai et al., 2007). Detrending subtracts the best straight-line fit from the spectra, while SNV centers and scales individual spectrum using their average and standard deviation and Savitzky–Golay filtering smooths and derives the signal in order to remove noise and baseline or additive effects. For AD waste characterization, SNV, in addition to detrending, could be an efficient treatment but the robustness should increase using SNV and Savitzky–Golay filtering (Lesteur et al., 2011).

To conclude, the most promising method for organic solid waste is NIRs or 3D fluorescence coupled with chemical sequential extraction despite a tedious and time consuming calibration phase needed before their first use. 3D fluorescence coupled with chemical sequential extraction has the advantage to predict both bioaccessibility and biodegradability but is expensive in term of experimentation time compared to NIRs. Beside, NIRs has been successfully used to predict the gas production kinetic parameters of forage digestibility, organic matter composition and inhibitory components (Lesteur et al., 2010). NIRs has the advantage of being very easy to perform, very fast, cost efficient and non-disruptive.

This chapter aims to define a fast, accurate and robust method of substrate composition characterization. Since both sequential chemical extractions coupled with 3D fluorescence spectroscopy or NIRs have been reported efficient for the estimation and qualification of the biodegradable organic matter, this chapter first assess the advantages and drawbacks of each method on a preliminary dataset composed of 28 samples. Eventually the most promising method is extended to a wider range of solid waste such as fruits, vegetables, manure, cereals, oils, fats, meat and fishes to create an important dataset with the aim to predict carbohydrate, protein and lipid contents, COD, methane yield and the Methane Production Rate (MPR) with fast spectroscopic analyses.

5.4 Materials and Methods Related to this Chapter

Detailed information about the materials and methods related to this chapter are provided in the "Materials and Methods" sections 2.3.1, 2.3.2, 2.3.3.1, 2.4, 2.5.1, 2.6.3.1, 2.6.3.3, 2.6.3.4.

The literature reports drawbacks in the conversion of TKN into protein using a conversion factor of 6.25. This factor has been observed to depend on the type of sample (Girault et al., 2012). Since this factor is uncertain and the protein content of a substrate is primarily used for monitoring the nitrogen release in the digester (Chen et al., (2008), Rajagopal et al., (2013)), in the case of fast spectroscopic analyses, the decision was taken not to convert TKN into protein and to work with nitrogen content per grams of dry matter.

5.5 Comparison of 3D Fluorescence and Near Infrared Spectroscopy for Biodegradability Prediction

As seen in the introduction of this chapter (0), 3D fluorescence spectroscopy coupled with chemical sequential extractions and NIRs has been successfully used to predict biodegradability (calculated from the BMP to COD ratio) and BMP respectively on different kinds of samples. Thus, the preliminary dataset was used to compare 3D fluorescence spectroscopy coupled with chemical sequential extractions and NIRs for the prediction of both methane yield and COD. The preliminary dataset was composed of a calibration set of 24 substrates and a validation set of 4 substrates. Table 6 summarizes the preliminary dataset features with minimal (min), maximal (max) and mean (mean) values together with the standard deviation (SD).

Table 6: Dataset features used for methane yield and COD prediction

		Calibration set	Validation set
Methane yield	Max ($L.gTS^{-1}$)	0.717	0.495
	Mean ($L.gTS^{-1}$)	0.33	0.311
	Min ($L.gTS^{-1}$)	0.141	0.199
	SD ($L.gTS^{-1}$)	0.119	0.130
	Number of samples	24	4
COD	Max ($mg0_2.gTS^{-1}$)	2334	1445
	Mean ($mg0_2.gTS^{-1}$)	1251	1221
	Min ($mg0_2.gTS^{-1}$)	935	1001
	SD ($mg0_2.gTS^{-1}$)	291	182
	Number of samples	24	4

Preliminary experiments have been made to assess the advantages and drawbacks of both methods. Prediction of methane yield and COD were carried out using the 21 indicators of chemical sequential extraction coupled with 3D fluorescence (7 fluorescence zones on 3 extracted fractions) or NIR spectra treated by pseudo absorbance and SNV for COD and pseudo-absorbance, SNV and Savitzky–Golay filter for methane yield prediction. The results are summed up in Table 7. Using 3D fluorescence coupled with chemical sequential extraction, SEC and SECV are higher than the standard deviation of the calibration set and R^2 is below 0.1 for both COD and methane yield predictions. SEC, SECV and R^2 indicate that 3D fluorescence coupled with chemical sequential extraction is poorly adapted to the prediction of methane yield and COD on a wide variety of samples. SEC and SECV are around 3 times lower for COD prediction and around 5 times lower for methane yield prediction using NIRs. These results tend to show that methane yield and COD predictions are more accurate using NIRs than 3D fluorescence but this hypothesis needs to be confirmed using an independent validation set.

Table 7: PLS model performance on cross validation and validation set for COD and methane yield prediction using 3D fluorescence spectroscopy coupled with chemical sequential extraction (FLUO) or near infrared spectroscopy (NIRs).

		FLUO	NIRs
Methane yield	SEC (L.gTS^{-1})	0.122	0.023
	SECV (L.gTS^{-1})	0.314	0.057
	R^2 cross-validation	0.07	0.76
	Latent variables	3	6
	SEP (L.gTS^{-1})	0.155	0.092
COD	SEC (mg0$_2$.gTS^{-1})	285	70
	SECV (mg0$_2$.gTS^{-1})	324	120
	R^2 cross-validation	0.00	0.84
	Latent variables	4	7
	SEP (mg0$_2$.gTS^{-1})	459	193

The independent set of 4 samples was used for the model validation and it confirmed that the predictions using NIR are more accurate than using 3D fluorescence coupled with chemical sequential extraction with SEP of respectively 193 mg0$_2$.gTS^{-1} and 459 mg0$_2$.gTS^{-1} for COD prediction and SEP of respectively 0.092 L.gTS^{-1} and 0.155 L.gTS^{-1} for methane yield prediction. Calibration and validation results are plotted in Figure 38 which confirmed that NIR have higher performances. 3D fluorescence shows low performances of prediction compared to the literature (Jimenez et al., 2014). 3D fluorescence was accurate analyzing organic matter containing many aromatic compounds such as biological sludge but is almost blind analyzing fruits, vegetables, meat, fat which contains a lower content in aromatic compounds. Thus, 3D fluorescence is not adapted to the analysis of a large range of solid waste. Nonetheless, 3D fluorescence coupled with sequential chemical extraction has the advantages to characterized well aromatic compounds like lignin. Most of all, it also indicates the accessibility of the organic matter (Jimenez et al., 2014). However, NIR analysis lasts for few minutes on a dry sample while 3D fluorescence coupled with chemical sequential extraction lasts for 5 days. It was thus decided to focus on NIR spectroscopy in the following of this study.

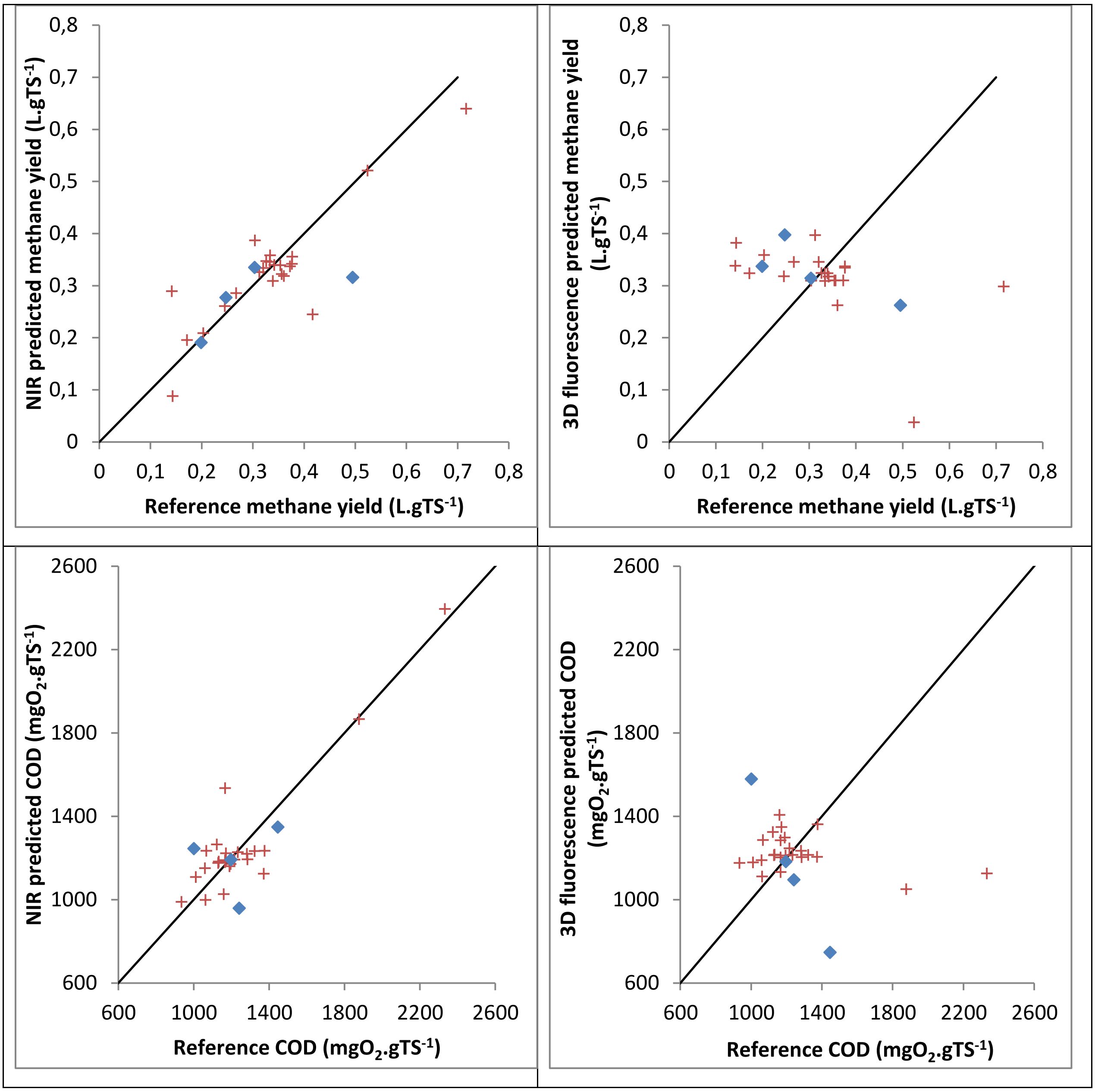


Figure 38: Comparison of cross-validation and validation PLS performances of NIR (left) and 3D fluorescence coupled with chemical sequential extraction (right) for the prediction of methane yield (up) and COD (bottom). The blue squares (◆) represent the prediction of the independent validation set whereas the red crosses (+) represent the estimation of the calibration set. The line represents the bisector between reference and predicted or estimated values.

5.6 Fast Characterization of Solid Waste Biochemical Composition with Near Infrared Spectroscopy

In order to further investigate the potential of NIRs, we developed a dataset of 338 samples as described in the “Materials and Methods” section 2.5.1. The following work focuses on the elaboration of a PLSr model to predict the biochemical composition of the substrate.

5.6.1 Combinations of the Calibration Model Using PLS Regression

Chemometric pretreatments like absorbance, SNV, Savitzky–Golay and Detrend have been applied to each spectra (see also the “Materials and Methods” section 2.4.1). The aim was to remove additive and multiplicative effects on the NIR spectra. The efficiency of this pretreatment is difficult to assess due to the great diversity in samples. The efficiency of the treatment was therefore evaluated using PLS performances. PLS regression was carried out to correlate the treated spectra with carbohydrate, nitrogen and lipid contents and COD independently. Table 3 summarizes the results of the calibration of the PLSr models. First of all, as expected in Lesteur et al., (2011), all treatments provided acceptable results for cross validation and calibration without any need to remove outliers. For instance, all determination coefficients (R^2) reached 0.6. Thus, the choice of the pretreatment focused on the estimation of either the accuracy or the robustness of the PLSr model. For nitrogen content, the Absorbance/SNV/Detrend appears to provide the best estimation, mean, minimum and maximum values in accordance with the experimental data. Standard error of calibration SEC is indeed equal to $3.7*10^{-3}$ $g.gTS^{-1}$ and standard error of cross-validation SECV to $5.3*10^{-3}$ $g.gTS^{-1}$, which is more than 10 times less than the mean value of prediction. The NIR repeatability was also excellent with a SDr of $1.4*10^{-3}$ $g.gTS^{-1}$. Nonetheless, eighteen latent variables were yet required. This constraint is high and may lead to over-fitting on the calibration dataset. Hence, Absorbance/SNV/Savitzky–Golay pretreatment was preferred, providing similar results with SEC value equal to $5.5*10^{-3}$ $g.gTS^{-1}$ and SECV values equal to $7.2*10^{-3}$ $g.gTS^{-1}$ and only ten latent variables.

For the calibration of carbohydrate content prediction, based on the same criteria, Absorbance/SNV/Detrend pretreatment was preferred with nine latent variables for a standard error of calibration SEC equal to 77 $mg0_2.gTS^{-1}$, a cross-valibration SECV equal to 90 $mg0_2.gTS^{-1}$ and excellent repeatability with a SDr of 14 $mg0_2.gTS^{-1}$. The main weakness of the model was the minimal predicted value at -170 $mg0_2.gTS^{-1}$ for a minimal real value at 0 $mg0_2.gTS^{-1}$.

Absorbance/SNV and Absorbance/SNV/Detrend pretreatments for the calibration of lipid content prediction provided equivalent performances, respectively SEC equal to $1.6*10^{-2}$ $g.gTS^{-1}$ for both, SECV equal to $2.5*10^{-2}$ $g.gTS^{-1}$ and $2.3*10^{-2}$ $g.gTS^{-1}$, SDr equal to $6.2*10^{-3}$ $g.gTS^{-1}$ and $6.4*10^{-3}$ $g.gTS^{-1}$, mean at $6.0*10^{-2}$ $g.gTS^{-1}$ and $6.1*10^{-2}$ $g.gTS^{-1}$, R^2 of 0.98 for both. However, Absorbance/SNV was preferred as it presented a slightly better accuracy at low values of lipid contents, as it is the case for many substrates (data not shown).

For COD prediction calibration, Absorbance/SNV and Absorbance/SNV/Detrend pretreatments presented very similar performances respectively with SEC of 89 and 82 $mgO_2.gTS^{-1}$, SECV of 101 and 99 $mgO_2.gTS^{-1}$, R^2 of 0.88 and 0.89. Minimal, maximal and mean estimated values were also similar. Absorbance/SNV/Detrend was preferred as it required one latent variable less and a slightly better SEC, however both treatments still remained acceptable.

Table 8: Influence of spectra treatment on PLSr performances, selected model values are in bold. For calibration dataset details, see Materials and Methods section 2.5.1.

	Treatment	Absorbance / SNV	Absorbance / SNV / Savitzky–Golay	Absorbance / SNV / Detrend
Nitrogen	Max (g.gTS^{-1})	0.15	**0.14**	0.14
	Mean (g.gTS^{-1})	2.8*10^{-2}	**2.8*10^{-2}**	2.8*10^{-2}
	Min (g.gTS^{-1})	-3.0*10^{-2}	**-1.0*10^{-2}**	-2.0*10^{-2}
	SEC (g.gTS^{-1})	6.5*10^{-3}	**5.5*10^{-3}**	3.7*10^{-3}
	SECV (g.gTS^{-1})	6.0*10^{-3}	**7.2*10^{-3}**	5.3*10^{-3}
	R^2	0.96	**0.94**	0.97
	SDr (g.gTS^{-1})	1.4*10^{-3}	**1.6*10^{-3}**	1.4*10^{-3}
	Latent variables	18	**10**	18
	Number of samples	295	**295**	295
Carbohydrates	Max (mg0$_2$.gTS^{-1})	1122	1180	**1161**
	Mean (mg0$_2$.gTS^{-1})	721	726	**721**
	Min (mg0$_2$.gTS^{-1})	-148	-141	**-170**
	SEC (mg0$_2$.gTS^{-1})	71	70	**77**
	SECV (mg0$_2$.gTS^{-1})	84	102	**90**
	R^2	0.91	0.87	**0.9**
	SDr (mg0$_2$.gTS^{-1})	15	17	**14**
	Latent variables	14	10	**9**
	Number of samples	295	295	**295**
Lipids	Max (g.gTS^{-1})	**1.068**	0.982	1.111
	Mean (g.gTS^{-1})	**6.0*10^{-2}**	5.6*10^{-2}	6.1*10^{-2}
	Min (g.gTS^{-1})	**-2.2*10^{-2}**	-4.6*10^{-2}	-2.0*10^{-2}
	SEC (g.gTS^{-1})	**1.6*10^{-2}**	2.0*10^{-2}	1.6*10^{-2}
	SECV (g.gTS^{-1})	**2.5*10^{-2}**	7.0*10^{-2}	2.3*10^{-2}
	R^2	**0.98**	0.86	0.98
	SDr	**6.2*10^{-3}**	7.9*10^{-3}	6.4*10^{-3}
	Latent variables	**10**	5	8
	Number of samples	**193**	193	193
COD	Max (mg0$_2$.gTS^{-1})	2912	2765	**2916**
	Mean (mg0$_2$.gTS^{-1})	1225	1211	**1225**
	Min (mg0$_2$.gTS^{-1})	943	426	**947**
	SEC (mg0$_2$.gTS^{-1})	89	83	**82**
	SECV (mg0$_2$.gTS^{-1})	101	179	**99**
	R^2	0.88	0.63	**0.89**
	SDr (mg0$_2$.gTS^{-1})	10.89	15.79	**11.94**
	Latent variables	8	7	**7**
	Number of samples	296	296	**296**

5.6.2 Model Validation on an Independent Dataset

The model previously developed for carbohydrate, nitrogen and lipid contents and COD prediction was tested on an independent dataset, as described in the "Materials and Methods" section 2.6.3.1. Results are presented in Table 9. No relevant bias has been observed between the predicted and reference data, which indicates that the model does not over-estimate or under-estimate the values. SEP are low. In details, assuming a normal error distribution, the confidence intervals for 68% of the estimated values were: ± $1.4*10^{-2}$g.gTS^{-1} of the predicted value of nitrogen, 70 mgO$_2$.gTS^{-1} for carbohydrates, ± $1.88*10^{-2}$g.gTS^{-1} for lipids and ±78 mgO$_2$.gTS^{-1} for COD. The predictions on the validation data set were excellent (see Table 9). It can be noticed that SEP can be below SECV. Two explanation are possible. First, in the calibration set, which is wider than the validation set, some sample are poorly predicted because of reference analyses not accurate or strong differences with the other samples of the calibration set. As a reminder, not outlier was removed of the calibration set. Second, the samples of the validation set could have some correlation between each other.

Table 9: Validation of the model on an independent dataset for nitrogen, carbohydrate and lipid contents and COD prediction. For validation dataset details, see Materials and Methods section 2.5.1.

Parameters	Nitrogen	Carbohydrate	Lipid	COD
Unit	g.gTS^{-1}	mg0$_2$.gTS^{-1}	g.gTS^{-1}	mg0$_2$.gTS^{-1}
Min	$1.2*10^{-2}$	329	$8*10^{-3}$	1125
Mean	$3.0*10^{-2}$	691	$3.3*10^{-2}$	1299
Max	$8.3*10^{-2}$	927	$12.6*10^{-2}$	1667
R^2	0.92	0.95	0.79	0.91
SEP	$5.2*10^{-3}$	35.2	$0.94*10^{-3}$	39
Bias	0.00	18.8	0.01	14
Number of samples	43	43	38	42

5.6.3 Discussion Regarding the Selected Model

The ideal calibration and validation set should have a mean value representing 33% of protein, 33% of carbohydrate and 33% of lipid which correspond to 0.05 g.gTS^{-1} for nitrogen, 351 mgO2.gTS^{-1} for carbohydrates and 0.33 g.gTS^{-1} for lipids. However, the dataset was chosen to rather represent the substrates usually treated in AD plants. This consequently represents a mean calibration value respectively set to about 0.03 g.gTS^{-1}, 723 mgO$_2$.gTS^{-1} and 0.06 g.gTS^{-1} for nitrogen, carbohydrate and lipid contents. The mean values of the validation set are respectively 0.03 g.gTS^{-1}, 672 mgO$_2$.gTS^{-1} and 0.02 g.gTS^{-1} for nitrogen, carbohydrates and lipids (Table 9). In both calibration and validation data sets, carbohydrates were over-represented while lipids were under-represented. Nonetheless, extreme values in the calibration set define the maximal potential range. In the validation set, the lipid maximal value represents 13% of the maximal existing value. The validation set is in agreement with the range of potential sample values for carbohydrates and proteins whereas lipids agree too poorly with a validation set covering 0 to 13 % of the potential sample values. The samples selected for the study represent a large range of the waste used for AD. Unsurprisingly, most samples were observed to be rich in carbohydrates while only a few were rich in lipids. Thus, current development focuses on samples with high lipid content to improve the robustness of lipid content prediction. Figure 39 illustrates the wide range of values for carbohydrates, with a relatively homogeneous distribution between 0 and 1200 $mgO_2.gTS^{-1}$ and with a higher density between 800 and 1000 $mgO_2.gTS^{-1}$.

Estimated values from the validation set also indicate a good dispersion and homogeneous distribution. Based on this chart and on the criteria in Table 9, the constructed PLS model can be expected to produce significant estimations of carbohydrate contents. The nitrogen content was also represented over a wide range of values, continuously between 0 and 0.15 $g.gTS^{-1}$. A high density of samples ranging between 0 and 0.03 $g.gTS^{-1}$ was observed. It matched with samples containing fruits, vegetables, cereals or a mixture of these. The samples with the highest contents in nitrogen have been reported to be meat or fishes, whereas samples with medium-values were micro-algae or diverse mix waste. As neither meat nor fish were part of the validation set, the validation set did not comprise the highest measured values. Nonetheless, the observed dispersion was equivalent to the calibration set. COD was distributed between 876 and 2884 $mgO_2.gTS^{-1}$. 876 $mgO_2.gTS^{-1}$ matched oxygenated samples while 2884 $mgO_2.gTS^{-1}$ matched fat samples with low-oxygen content which are good representatives of the variety of samples that can be found in AD. The 1600 to 2200 $mgO_2.gTS^{-1}$ range does not correspond to any samples. It rather characterises protein-rich samples for the maximal value and lipid-rich samples for the minimal one. It could be worthwhile to add samples within this range that might contain a mixture of lipids and proteins or/and carbohydrates. However, the miscibility of hydrophilic and lipophilic compounds might be an issue for the analysis of this kind of sample. Indeed, samples in this study were generally fresh, containing of an important amount of water which is not miscible with lipids. Samples with similar contents in hydrophilic and lipophilic compounds were thus split into two sets, one with a low lipid content and one with a high lipid content. The lack of intermediate values remains an experimental problem that is still difficult to solve. Hence, the samples used for calibration and validation represent experimental means for accurately determining the diversity of samples that can be used for AD.

Based on this chart and the previous data, the constructed PLS model can be considered as relevant for carbohydrate, nitrogen and lipid contents and COD estimations over a wide range of sample types.

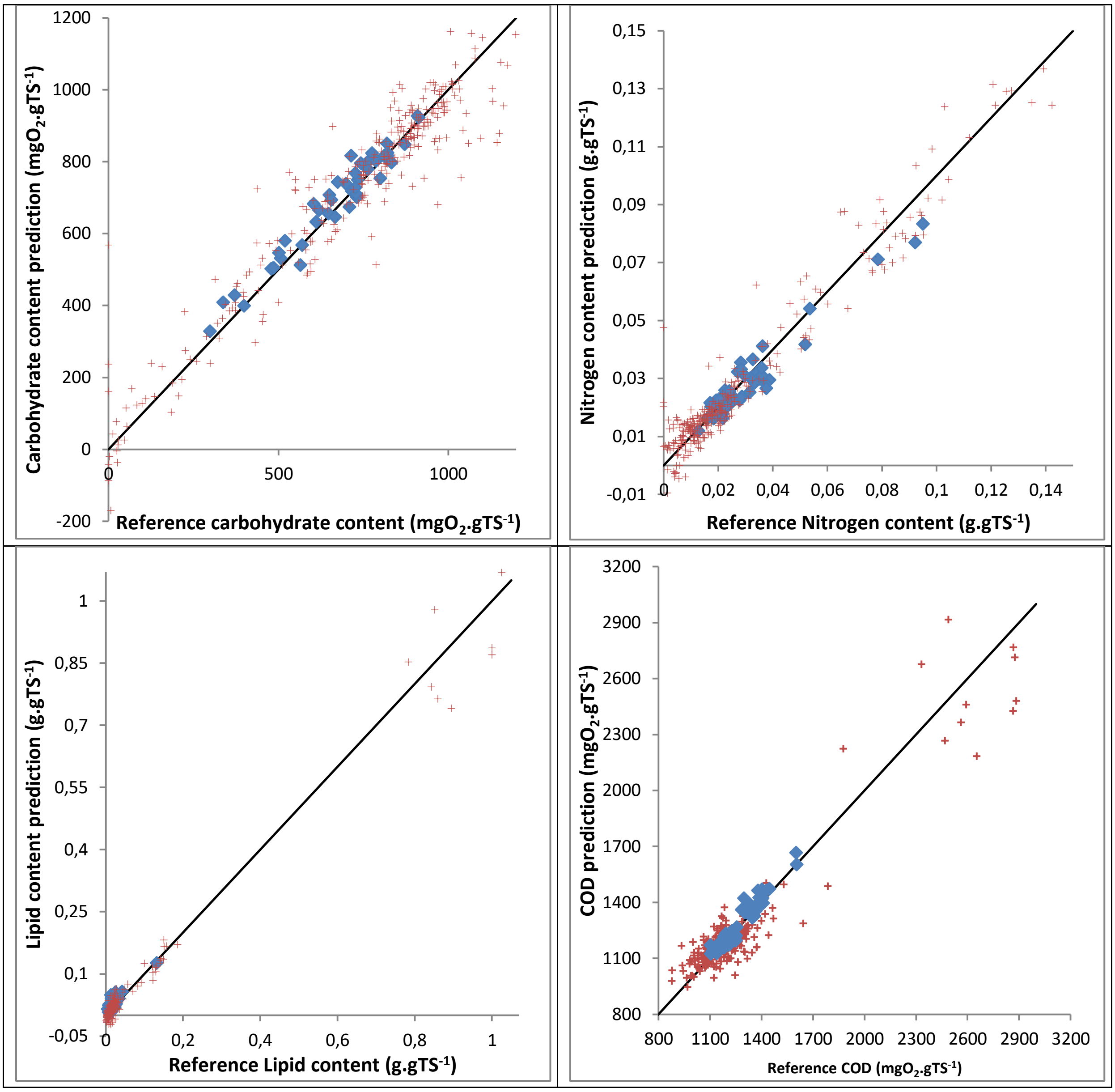


Figure 39: Prediction performances of carbohydrate, nitrogen and lipid contents and COD, respectively on the top left, top right, bottom left and bottom right. The blue squares (♦) represent the prediction of the independent validation set whereas the red crosses (+) represent the estimation of the calibration set. The line represents the bisector between reference and predicted values.

5.7 Prediction of the Methane Production Rate in Batch Conditions Using NIR Spectroscopy

As just shown, NIRs can successfully predict the biochemical composition of solid waste. But can it also provide information on methane production yield and dynamics of these solid waste during anaerobic digestion?

5.7.1 Prediction of the Methane Yield

The methane yield, defined in "Materials and Methods" section 2.3.3.1, was predicted according to (Lesteur et al., 2011) using chemometric treatment Absorbance/SNV/Savitzky–Golay using the dataset described in section 2.5.1. Based on the cross-validation performances, the PLS regression model was calibrated using 13 latent variables. On the calibration set, the SEC was about 0.019 $L.gTS^{-1}$, a SECV about 0.026 $L.gTS^{-1}$ and R^2 about 0.92. The model was validated on the independent validation set with SEP of 0.025 $L.gTS^{-1}$.The SEP found on the methane yield prediction is comparable with the SEP of 0.028 $L.gVS^{-1}$ reported in (Lesteur et al., 2011). The prediction of the methane yield using NIRs will especially be used in Chapter 7 to set the parameters needed for ADM1 simulation.

5.7.2 Prediction of the Methane Production Time

First, the durations required to reach from 5% to 95% of the methane yield with step increases of 5% have been considered as independent values. The choice of the chemometric treatment was based on the PLS performance on cross-validation and calibration. SEC, SECV and R^2 between predicted and reference values in cross-validation are represented in Figure 40. For all pretreatments, SEC and SECV increased until 70% of the methane yield and were even multiplied by approximately 6 from 5% to 70%. Then both decreased. The variations of SEC and SECV are correlated with the variance within the dataset which increases with the methane yield to reach a maximal variance at 70% as suggested in Figure 40. The experimental standard deviation and the standard deviation of the predicted values in cross-validation have been also plotted in Figure 40. The standard deviation reaches a maximum of 1.24 day at 70% of the methane yield. Thus, the most important difference between the samples is reached at 70% of the methane yield. This explains the increase of both SEC and SECV up to this point. Then, the variance decreased indicating a homogenization of the samples. 70% of the methane yield is, thus, the most discriminant production time. It assumes a strong decrease in the methane flow rate and a slow degradation speed above 70% of the methane yield which suggests that the remaining organic matter is slowly degradable.

About the chemometric treatment, both SEC and SECV were minimized using treatment Absorbance/SNV/Savitzky–Golay at any ratio of the methane yield with respectively values from 0.05

to 0.37 days and 0.07 to 0.55 days. R^2 was maximal using treatment Absorbance/SNV/Savitzky–Golay until 70% of the methane yield whereas treatment Absorbance/SNV/Detrending showed slightly improved R^2 after 70% of the methane yield compared to treatment Absorbance/SNV/Savitzky–Golay. Treatment Absorbance/SNV/Savitzky–Golay maximized the PLS regression performances for production time prediction and was chosen as the reference chemometric treatment for this purpose.

PLS regression was conceived to treat also the production time as a vector of correlated values. The advantages of this approach are to acknowledge correlation between the production times at any ratio of the methane yield. The disadvantage is the necessity to set a unique number of latent variables for every ratio of the methane yield. Using treatment Absorbance/SNV/Savitzky–Golay, both approaches were compared considering either the time of production as a correlated vector or as independent values. SEC was not accessible using MixOmics (MixOmics is used to predict the time as a correlated vector, see “Materials and Methods” section 2.6.3.1) and is thus absent in Figure 40. Q^2 was used to set the number of latent variables to 5 which corresponds to the first Q^2 value below 0.0975 with a local minimum. The number of latent variables considering production time as independent values varies from 6 to 12. A reduction of the number of latent variables is usually acknowledged as a gain in robustness. Nonetheless, SECV increased considering production time as a vector with values from 0.07 to 0.60 days and R^2 decreased with values from 0.79 to 0.47. On average, SECV increased by 8.5% and R^2 decreased by 6.1% considering the production time as a vector. From these results, it can be concluded that the performances are maximized considering the time of production as independent values with treatment Absorbance/SNV/Savitzky–Golay.

The resulting PLS regression model was then tested on the validation set. SEP is presented in Figure 41. The same pattern as for SECV and SEC estimation was observed for SEP which a reached maximal value at 70% of 0.47 day. The standard deviation of the validation set also confirmed an increase of the dispersion until 70% of the methane yield (Figure 41). SEP is approximately 2 times lower than the standard deviation of the validation set which proves that PLSr was able to discriminate the production time of the samples. It also shows that the errors of prediction are low, for instance at 65% of the methane yield, the SEP was 0.47 day to predict values in a range from 1.15 to 4.75 day with a mean value of 2.32 day. The model previously built on the calibration set can thus be considered as suitable to predict the times of degradation at any methane yield ratio.

The cross-validation and validation results using Absorbance/SNV/Savitzky–Golay pretreatment and considering the times required to reach a certain percentage of the methane yield as independent values are detailed at the bottom of Figure 41. In these two figures, the methane production times are represented without any distinction on the methane yield ratio. Figure 41 confirms the accuracy of the PLSr model but also shows that the error of prediction decreases with the predicted time. The methane production times above 6 days are usually representing a methane production above 70% of the methane yield. It has been previously observed that the variance decreased after 70% and this goes with a decrease of SECV and SEP which explains the decrease in the error of prediction. On the cross-validation results, one sample is strongly under-predicted at several methane yield ratios. It is actually the only sample represented by oil. It shows that oil is under-estimated on the current dataset of calibration. This error is not found on validation because no oil was contained in the validation set. Nonetheless, from these results, it can be concluded that the prediction of production time is feasible using NIR spectroscopy and PLS model, even if an improvement of the dataset on oil could be made.

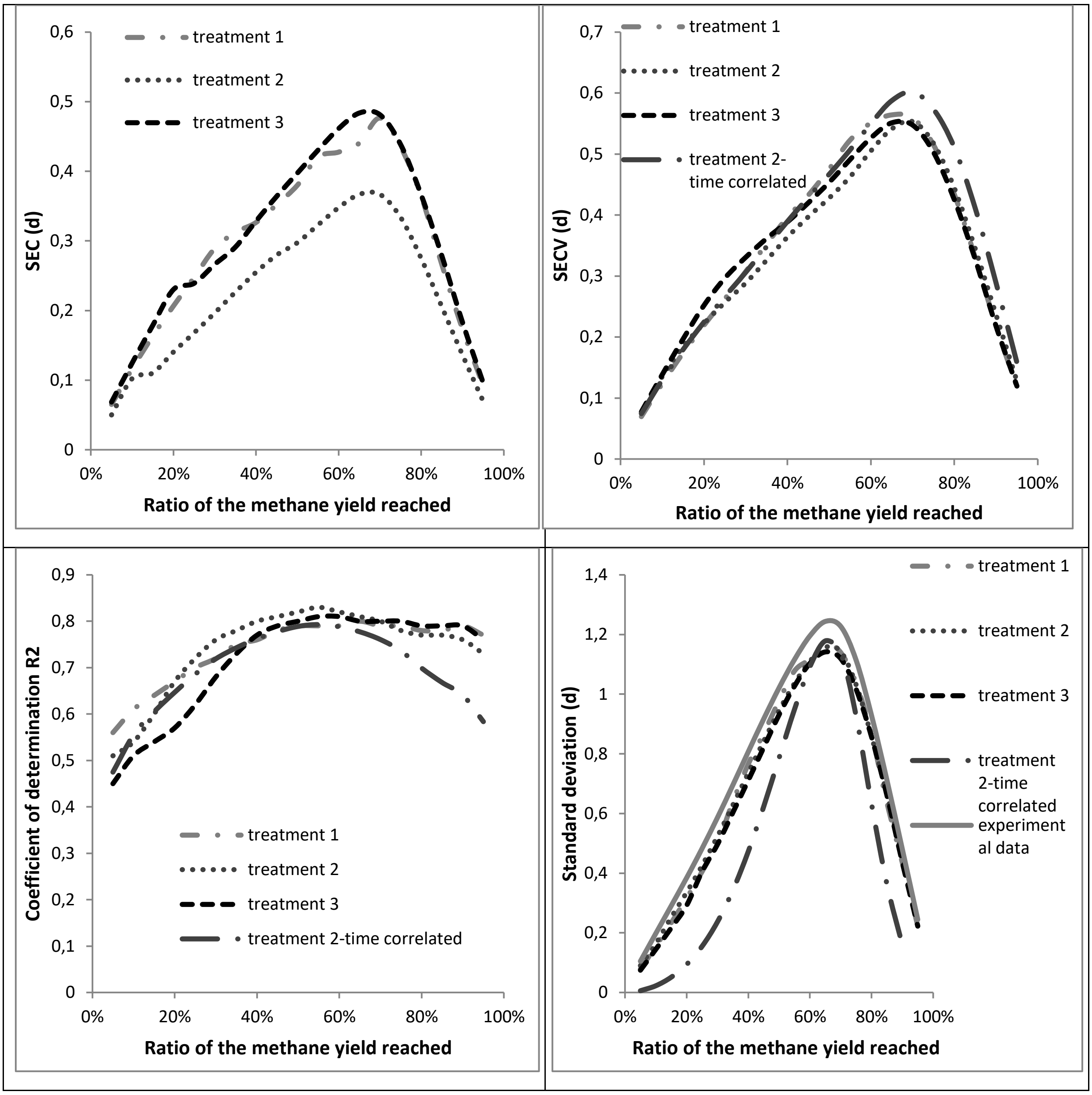


Figure 40: PLS regression performances on cross-validation with 3 different chemometric treatments and considering the methane production time as correlated or independent values. Treatment 1 corresponds to absorbance and SNV, treatment 2 to absorbance, SNV and Savitzky-Golay filter and treatment 3 to absorbance, SNV and detrending.

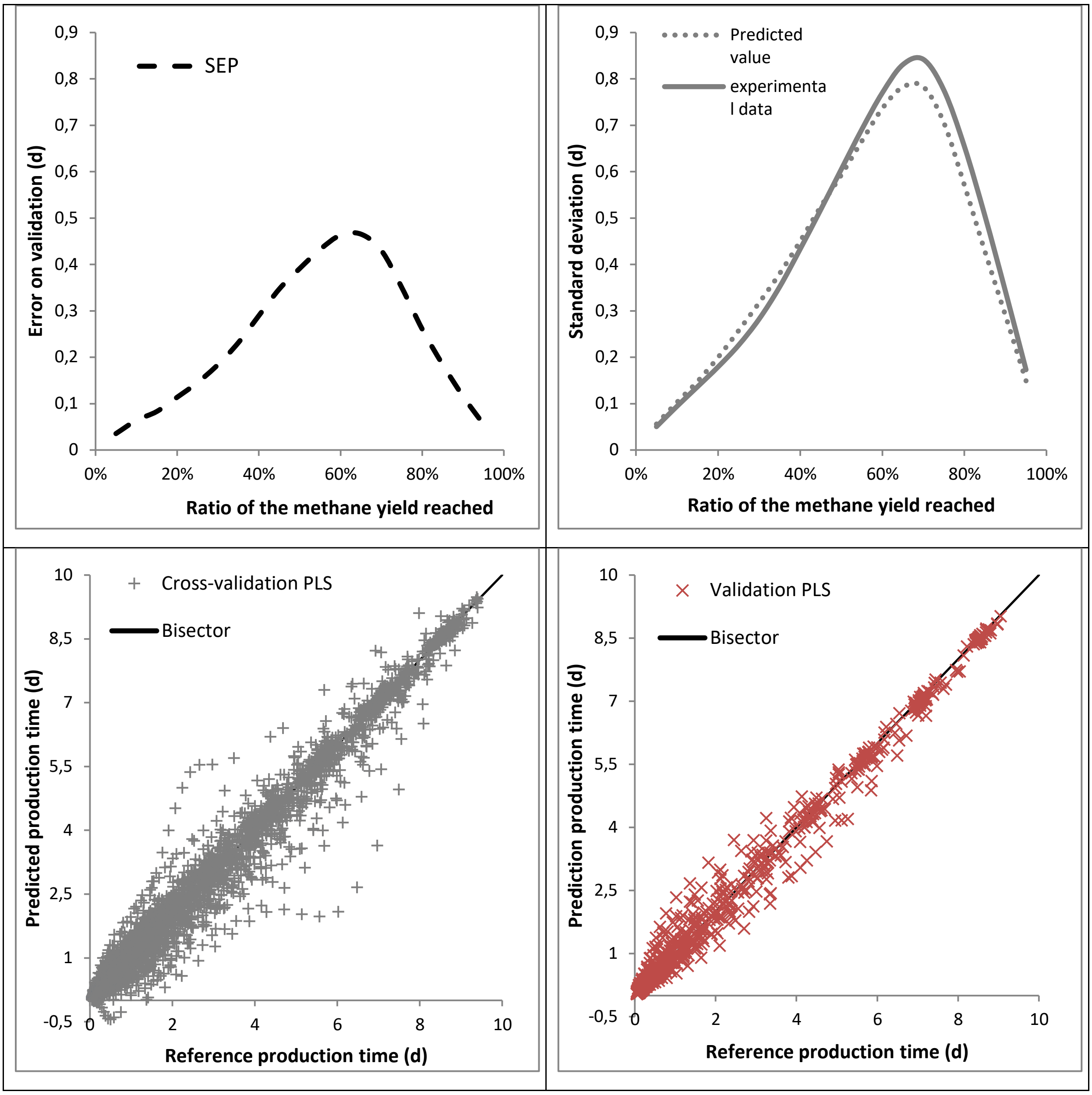


Figure 41: PLS performances on the validation set (up) and representation of the predicted value of both cross-validation and validation (bottom).

5.7.3 Methane Production Rate Prediction

A case study was carried out on two samples from the validation set – salad and cauliflower – considered as representative of bio-wastes, to assess the feasibility of the methane production rate (MPR) prediction. As explained in the “Materials and Methods” section 2.7.2, the MPR was predicted using both methane production times and methane yield predicted by NIR analyses. Then MPR was interpolated from these 19 points. These curves are represented as “NIR prediction” in Figure 42 and compared to the experimental cumulated methane production. It shows that NIR prediction represents properly the production of methane. Both methane yield and production time were correctly predicted. The errors of prediction on the methane yield were about 6.7% on salad and 3.5 % on cauliflower. On the experimental data, the methane flow decreased after 0.4 and 1 day and increased again after 1 and 1.5 day for respectively salad and cauliflower. These behaviors were not accurately predicted by NIR. One possible explanation is an inhibition of the readily degradable substrate on the slowly degradable substrate that is not properly predicted by NIR. Another hypothesis leading to similar results can be found in Yasui et al., (2008): the readily accessible organic matter is degraded first followed by the slowly accessible organic matter, keeping in mind that slowly accessible can be readily degradable.

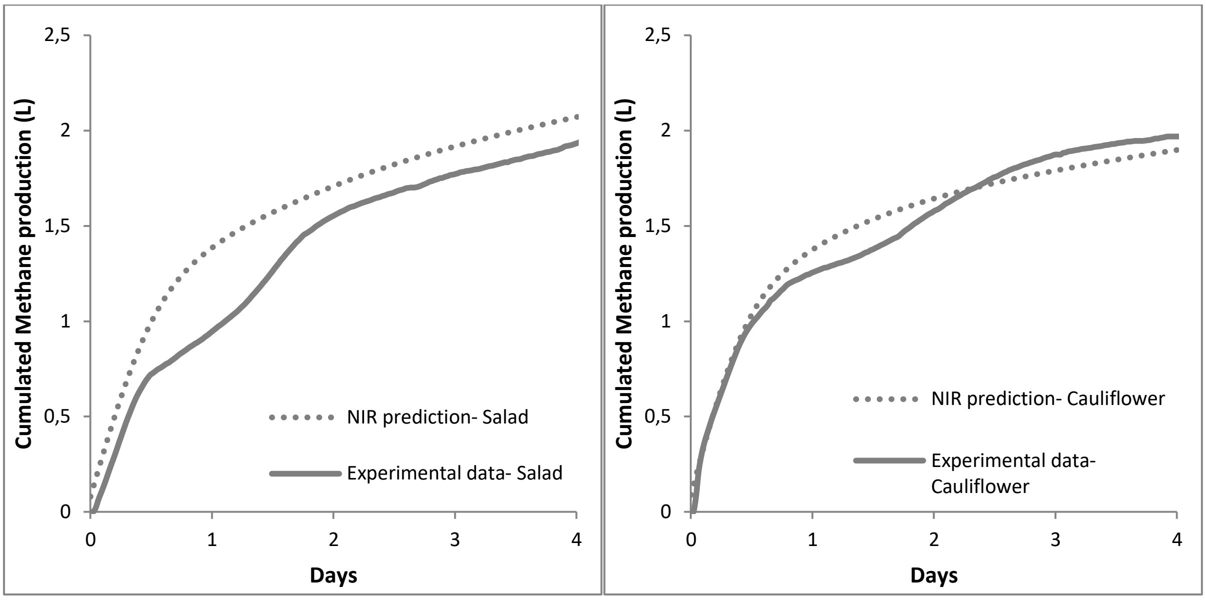


Figure 42: Cumulated Methane production curve for anaerobic of salad (left) and cauliflower (right) in batch. Comparison of predicted with reference data

5.7.4 Simple Indicator of Methane Production Performances

MPR is a relevant indication that allows to differentiate the methane production performances of two substrates, but MPR analysis becomes difficult when the number of samples to compare increases. The time needed to reach 70% of the methane yield has been reported in section 5.7.2 to be the most discriminant time and it allows differentiating substrates with fast kinetics of methane production. In Figure 43, the samples of the validation set (Materials and Methods section 2.5.1) have been plotted depending on their predicted methane production time needed to reach 70% of the methane yield and their predicted methane yield. Figure 43 shows that the NIR spectroscopy analysis of the substrate provides the human operator with a fast and unique opportunity to select the most relevant substrates on a simple graph, based on methane yield and kinetics of methane production.

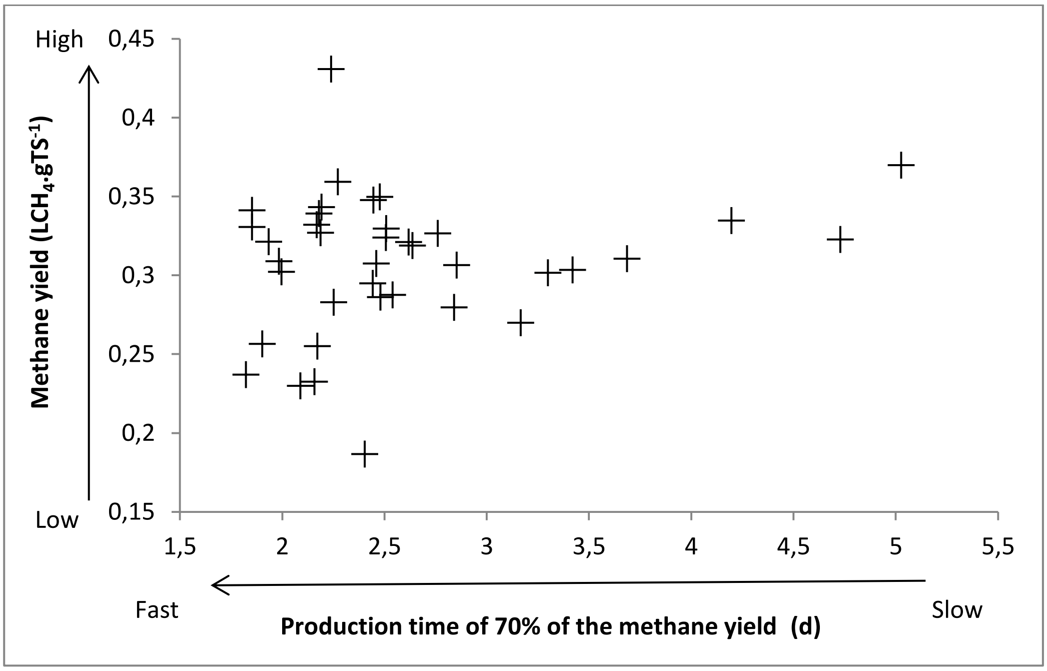


Figure 43: Estimation of the validation set substrates performances expressed as a function of the methane yield and the production time of 70% of the methane yield. Abscissa arrow indicates the methane production kinetics and ordinate arrow the methane yield.

5.8 Confidence Interval Determination Using Spectrum Distance from the Calibration Set

Determination of methane yields is a rising application of NIR spectroscopy. Many different calibrations have been developed in the literature (see for example Lesteur et al., (2011), Jacobi et al., (2012), Ward, (2016)) focusing, either on wide or narrow variation range of substrates. However, Dardenne (2010) reported that NIR prediction accuracy depends on the dataset and that the distance between a new sample and the calibration dataset must be low. However creating a dataset that accounts for any kind of possible samples is a difficult task. Moreover at an industrial scale, a wide diversity of samples can be used and in some cases, the samples lie outside the calibration dataset and the estimations become misleading. It reduces drastically the application of NIR spectroscopy and induces reticence towards the use of NIR at industrial scales. To overcome this problem, expertise on the calibration set is needed and indicators stating for which the estimation is acceptable must be assessed.

Bellon-Maurel et al., (2010) reported in a review that the confidence interval of a NIRs soil quality attributes prediction can be estimated based on the Malahanobis distance between the spectrum of the new sample and the centre of the calibration set. Since this indicators depends on the diversity of samples used and since it was never validated on the prediction of solid waste composition, Euclidian distance from the nearest spectrum or from the mean spectrum and Mahalanobis distance were compared on processed spectra but also on the spectra projected on the space of PLS loading vectors of the calibration set. Taking the example of methane yield prediction, Figure 44 shows, as reported by Bellon-Maurel et al., (2010), that the confidence interval enlarges with the distance. Nonetheless, in the specific case of solid waste composition, where an important diversity of sample is found, the most noteworthy pattern was obtained when correlating the minimal Euclidian distance between the samples and the calibration dataset. Comparing the minimal Euclidian distance between the samples and the calibration set and the minimal Euclidian distance between the samples projected into the PLS space and the calibration set projected into the PLS space, it can be observed that the projection into the PLS space reduces to 0 the distances of samples which used to be far from the calibration set. Indeed the projection into the PLS space removes the part of the spectrum orthogonal to the PLS space, hence the projection of the spectrum which were previously far from the calibration set are found to be near to the projected calibration set. Thus, it can be concluded that the projection into the PLS space for distance analysis is not relevant. Similar results were obtained for carbohydrate, protein and lipid contents, COD and methane production times and it was thus decided to focus on the minimal Euclidian distance. The relation between the error of prediction and the minimal Euclidian distance between the samples and the calibration dataset is plotted in Figure 45 for all the predicted parameters. A relation between the Confidence Interval (CI), representing at least 95% of the samples, and the Minimal Euclidian Distance (MED) is observed for the prediction of all variables. As does the distance, the relation between the MED and the CI seems to be strongly affected by the chemometric pretreatment.

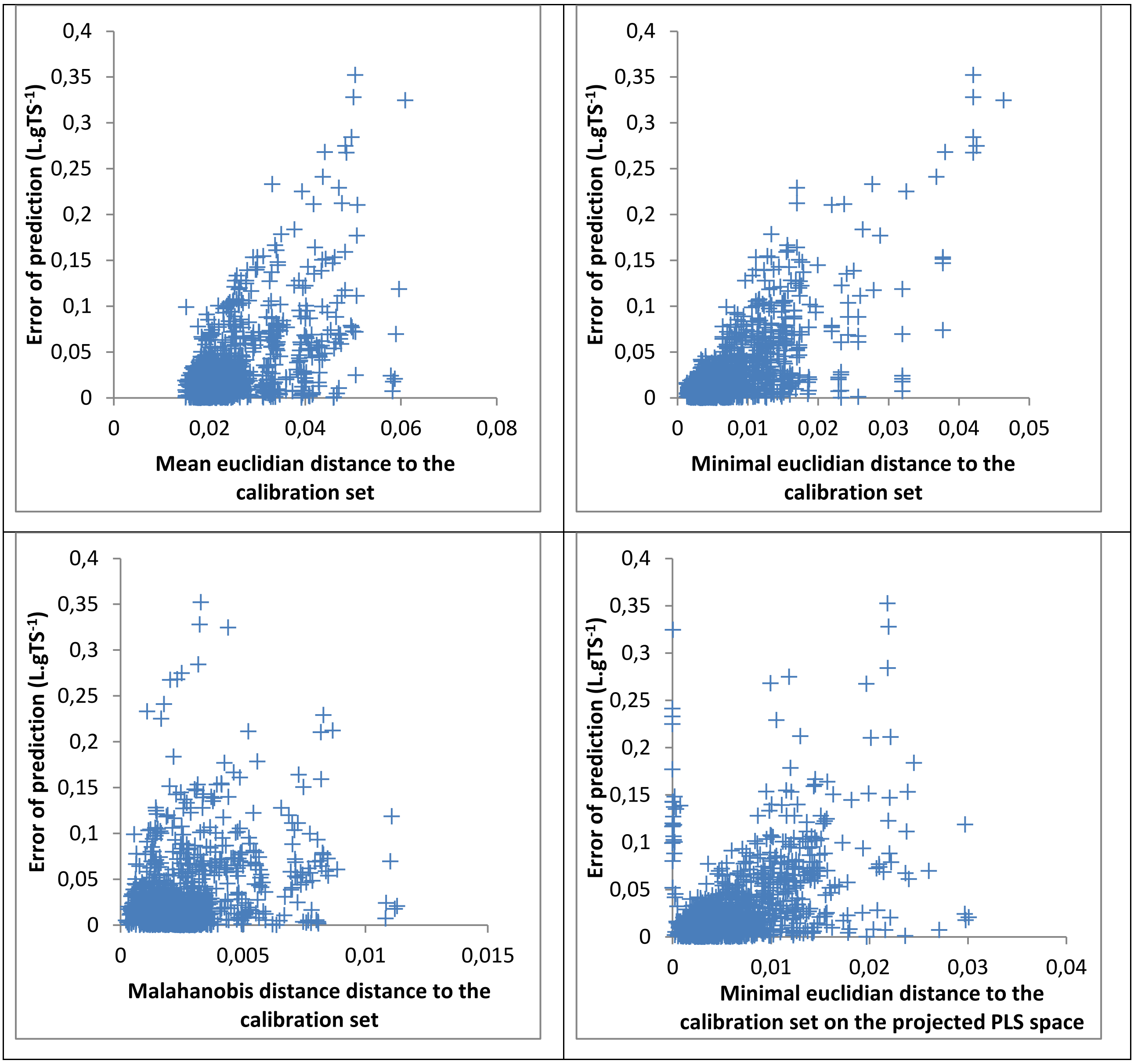


Figure 44: Relations between the distance of a spectrum from the calibration set and its error of prediction. Example of the methane yield prediction

To assess the relation between the CI and the MED, CI has been modelled according to “Materials and Methods” section 2.6.3.4 as a function of the MED. Figure 45 shows that the relation between MED and CI can be modelled by the equation:

$$Confidence\ interval = a * MED^{n} + \|b\|$$

Details of the parameters *a*, *b* and *n* are provided in Table 10. As it can be seen, the parameter *b* is always equal to 0, indicating that when MED is equal to 0, there is no error of prediction. Thus, it demonstrates that every spectrum is specific: a unique value of the variables is associated to each spectrum. Moreover, for each chemometric treatment, similar orders of magnitude can be found for both *n* and the ratio *a*/SECV. It shows that *a* depends on the SECV and on the chemometric pretreatment. Using a bigger dataset, further investigation could be carried out to highlight the relation between CI, SECV, MED and the chemometric treatment to propose an equation based on SECV and MED to predict CI depending on the chemometric treatment.

Table 10: Relation between the Minimal Euclidian Distance (MED) and the Confidence Interval (CI) for carbohydrate, nitrogen and lipid contents, COD, methane yield and methane production time prediction at 70% of the methane yield. In the table, the variables are ranked by chemometric treatment.

Predicted variable	**Chemometric treatment**	**Equation:** $\boldsymbol{CI} = a * MED^{n} + \|b\|$	**Ratio a/SECV**
Lipid content ($g.gTS^{-1}$)	Absorbance / SNV	$CI = 0.012 * MED^{1.54}$	0.523
Carbohydrate content ($mgO2.gTS^{-1}$)	Absorbance / SNV / Detrend	$CI = 193 * MED^{0.46}$	2.14
COD ($mgO2.gTS^{-1}$)	Absorbance / SNV / Detrend	$CI = 173 * MED^{0.63}$	1.74
Nitrogen content ($g.gTS^{-1}$)	Absorbance / SNV / Savitzky–Golay	$CI = 0.52 * MED^{0.70}$	72.0
Methane yield ($L.gTS^{-1}$)	Absorbance / SNV / Savitzky–Golay	$CI = 4.40 * MED^{0.84}$	169
Methane production time at 70% of the methane yield (days)	Absorbance / SNV / Savitzky–Golay	$CI = 43.47 * MED^{0.68}$	79.0

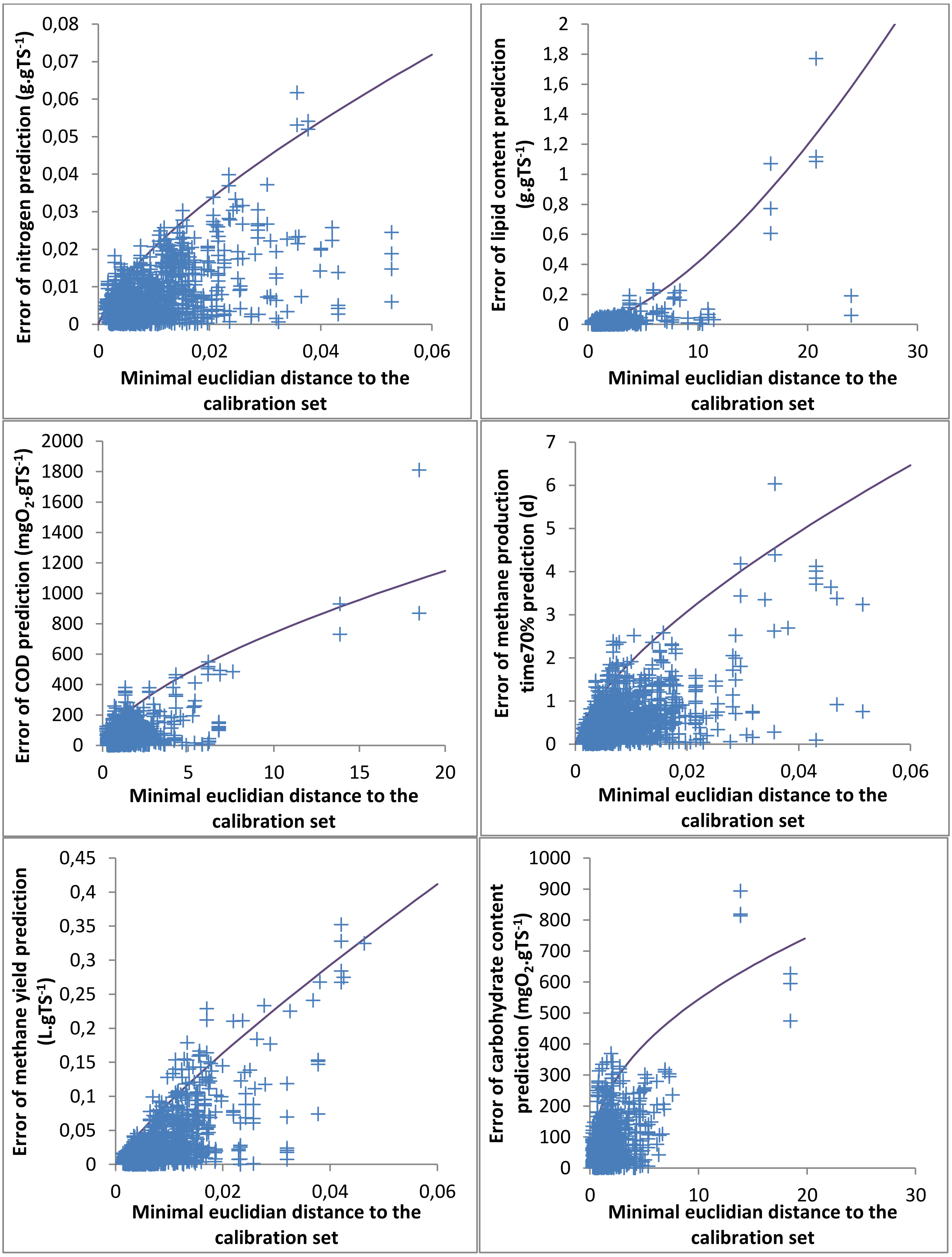


Figure 45: Predictions confidence interval depending on the minimal Euclidian distance from the spectrum to the calibration dataset for carbohydrate, nitrogen and lipid contents, COD, methane yield and methane production time at 70% of the methane yield. The purple line models the confidence interval.

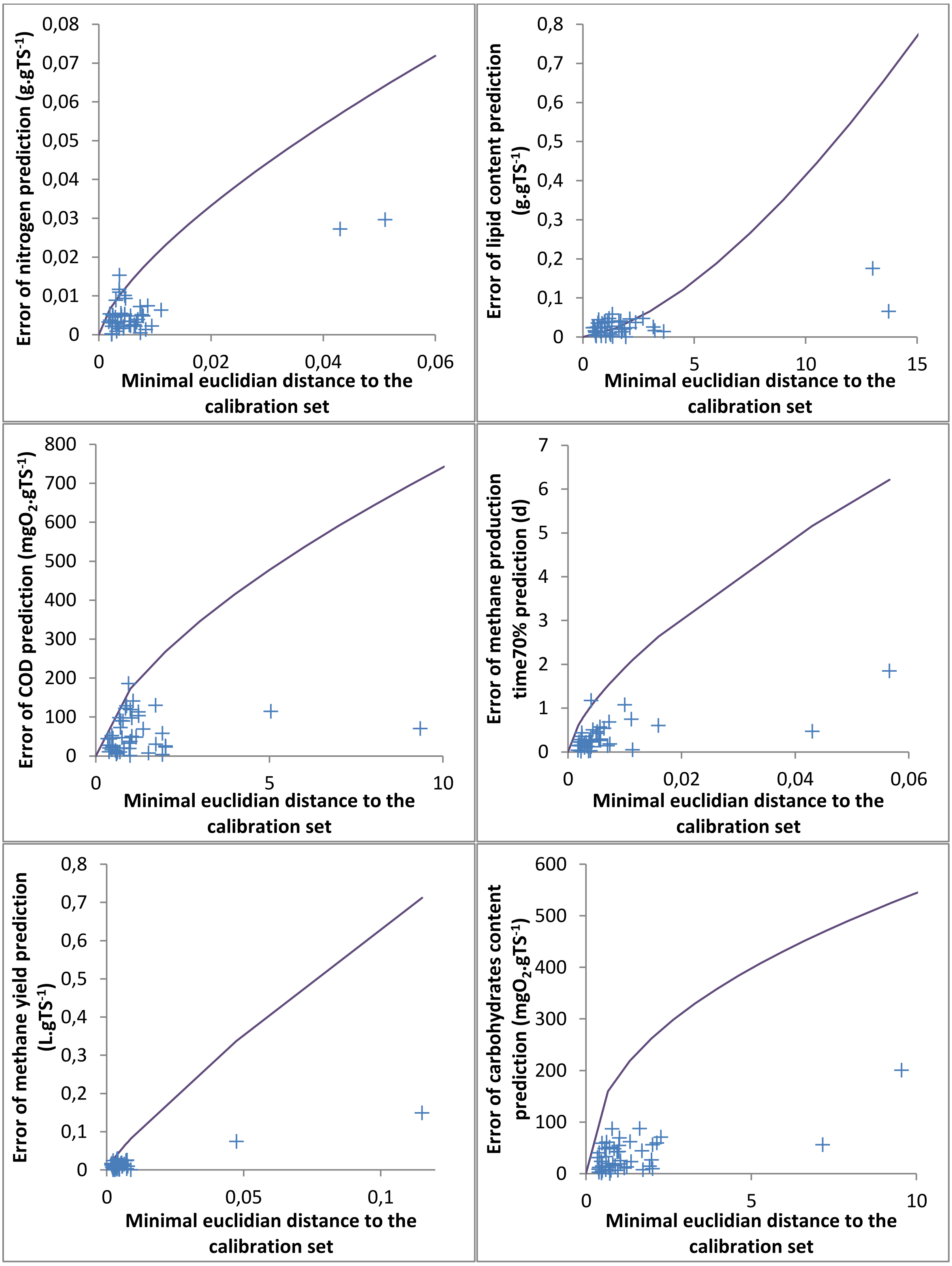


Figure 46: Prediction errors on the validation set depending on the minimal Euclidian distance from the spectrum to the calibration dataset for carbohydrate, nitrogen and lipid contents, COD, methane yield and methane production time at 70% of the methane yield. The purple line corresponds to the equations of the confidence interval established on the calibration set (Table 10).

The relations established on the calibration set between the CI and the MED (Table 10) were validated on the validation set. Figure 46 demonstrates that these relations remain valid since 95% of the samples are included in the confidence range, expect for lipid content prediction. Lipid content prediction in samples with less than 5% of lipids has low performances as shown in Figure 39, Figure 45, Figure 46. Unfortunately the reference method being known as unprecise for low lipid content, it is difficult to improve the performances of prediction below this threshold.

5.9 Conclusion on NIRs Characterization

In the previous chapter, correlation between biochemical composition and methane performances were highlighted, pointing towards more descriptive organic matter characterizations such as NIR or 3D fluorescence spectroscopy. This hypothesis was investigated in this chapter. To do so, both laboratory analyses for the estimation of carbohydrate, nitrogen, lipid contents and COD in solid samples and batch experiments using adapted micro-organisms have been performed to create a dataset for NIR prediction. Based on the PLSr performances, models were established and optimized. They show excellent accuracy of prediction for both cross validation and validation with an independent data set. Moreover, the accuracy of predictions can be warrantied according to CI based on the distance between the sample and the calibration dataset. If the sample should be too far from the calibration dataset, the prediction should not be used and the sample ought to be analysed with reference methods. Furthermore, the sample should ideally be added to the calibration dataset. This would increase its size and relevance. It would also reduce the cost of dataset improvement, selecting only the most relevant sample (i.e., with an important distance from the calibration set). These results on the correlation between distance and CI are a worthwhile approach for improving NIR confidence for industrial applications.

All these results proved that NIR spectroscopy is a new and efficient way to characterize solid waste content for industrial-scale applications. The black box models developed are also a unique opportunity to investigate the link between the composition of the organic matter and the process performances. In the next chapter, the PLS models will be deeply investigated to better understand the molecules contribution to the prediction of each parameter. Nonetheless, the anticipation of the methane performances of the unit based only on a detailed characterization of substrate would be unsatisfactory. Indeed, dynamic models taking into account both biological state of the unit and substrate composition have to be used. Thus Chapter 7 will investigate the use of the PLSr developed in this chapter to implement a dynamic model for process performances prediction.

Chapter 6. Parallel Study of PLS b Coefficients and Near Infrared Wavelengths to Assess Molecules Contribution to Chemical Oxygen Demand, Methane Production Yield and Kinetics.

Chapter 6. Parallel Study of PLS *b* Coefficients and Near Infrared Wavelengths to Assess Molecules Contribution to Chemical Oxygen Demand, Methane Production Yield and Kinetics. 141

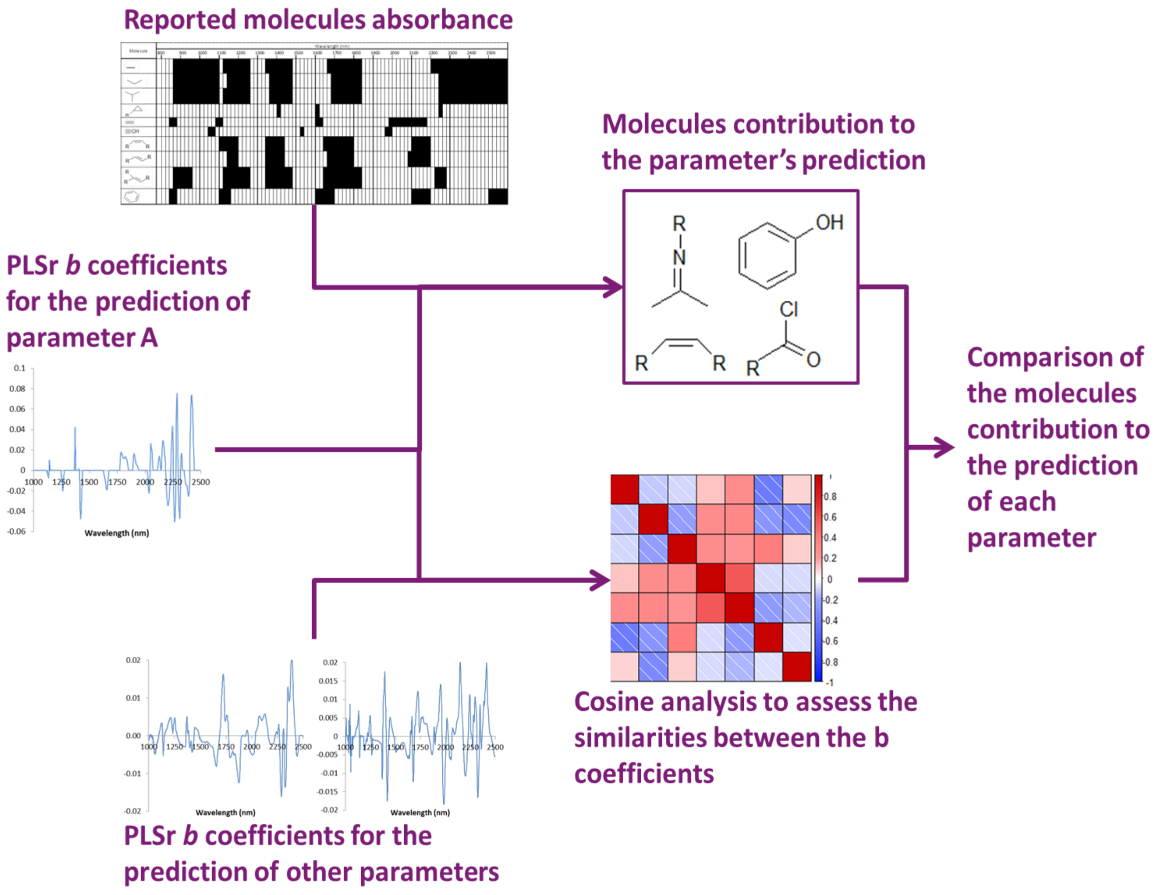
6.1 Chapter Guidelines 143
6.2 Technical and Graphical Abstracts 145
6.3 Introduction of the Chapter 147
6.4 Materials and Methods Related to this Chapter 151
6.5 Homogenization of the Chemometric Treatment 151
6.6 *b* Coefficients Analysis and Molecules Contribution to the Predicted Parameters 152
6.7 Relations between the *b* Coefficients of the Models 157
6.8 Conclusion 160

6.1 Chapter Guidelines

The novelty of this chapter consists mainly on its approach. The classical approach for predictive models consists in collecting knowledge in order to create an expert model that links the predictors to the predicted values. In this chapter, the approach is different in the sense that a black box model does not require any previous knowledge. Thus, the black box model is investigated to understand the existing relations between the predicted values and the predictors. In the current situation, the models previously developed in Chapter 5 are deeply investigated to give clues about the molecules involved in the prediction of AD performances. It specially focuses on the *b* coefficients of the models. The introduction of the chapter (0) recalls the principles of the PLS and *b* coefficients before reporting the known absorbance of molecules in near infrared (NIR). A short reference to the “Materials and Methods“ is made in section 6.4. Then section 6.6 compares the *b* coefficients to the reported molecules absorbance to assess the contribution of the molecules to both biochemical composition and the prediction of AD performances. Eventually, section 6.7, investigates the relations between the different models to go further in determining the contribution of each molecule to the prediction. Since the analysis of the *b* coefficients and the comparison of the models needs a deep understanding of the correlations existing between the different predicted values, this chapter often refers to Chapter 4.

6.2 Technical and Graphical Abstracts

Near infrared coupled with partial least square regression can be used to predict biochemical parameters such as carbohydrate, protein and lipid contents but also chemical oxygen demand, methane production yield and kinetics. The aim of this chapter is to interpret the partial least square *b* coefficients associated to the parameters predictions. Parallel study of the *b* coefficients and near infrared wavelengths provides an insight on molecules contribution to these parameters. It highlights either the role of phenol in non-biodegradable fraction or the importance of alkane and alkene on the methane yield of a substrate. In a second step, the *b* coefficients of the partial least square regression models were compared to each other. The relations were analyzed and confirmed the link of lipid content with methane yield or chemical oxygen demand. It also highlights relations that affect plant monitoring strategy as, for instance, anti-collinearity between the *b* coefficients used for the prediction of chemical oxygen demand and the one used for the prediction of biodegradability, or the independence of the methane production yield and kinetics. Eventually, this study reveals that proteins and biodegradable molecules are positively related with methane production kinetics and thus readily biodegradable. The interest of this work is twofold, it evaluates the molecules contribution to the methane production performances and it validates the use of the *b* coefficients as an investigation tool.


Reported molecules absorbance
Molecules contribution to the parameter's prediction
R
N
OH
Cl
R
R
R
O
PLSr *b* coefficients for the prediction of parameter A
0.1
0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
1000
1250
1500
1750
2000
2250
2500
Wavelength (nm)
Comparison of the molecules contribution to the prediction of each parameter
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
Cosine analysis to assess the similarities between the b coefficients
0.02
0.01
0.00
-0.01
-0.02
0.015
0.005
-0.005
-0.015
Wavelength (nm)
PLSr *b* coefficients for the prediction of other parameters

6.3 Introduction of the Chapter

Indicators such as methane yield, COD or dynamics of methane production provide macro-information useful to monitor anaerobic digestion plants but are poorly informative on the composition and structure of the organic matter. As it was previously shown in Chapter 5, PLSr models between NIR spectra and carbohydrate, protein and lipid contents, COD, methane yield or kinetics of methane production can be efficiently built to estimate these variables. Thus, the features selected by the PLSr can be trusted as indicators of the molecules involved in the determination of these parameters.

PLSr is a multivariate projection based method that performs local regression. Working with NIRs, PLSr maximizes the covariance between the spectrum and the parameters to predict, seeking for linear combinations of the variables from both sets. To do so, PLSr performs successive regressions using projection onto latent vectors. It results in a linear combination with the spectra intensities of each wavelength. Considering Y the parameter to predict:

$$Y = b_0 + \sum_{i=1}^{p} b_i . A_i + R \tag{19}$$

with b_0 the intercept, b_i, the PLSr *b* coefficient for wavelength *I*, A_i, the intensity value of wavelength *I* and R the model residuals.

In these models, the wavelengths contribution, through the analysis of the *b* coefficients, can provide additional information of importance. NIR absorption indeed corresponds to vibration of the atomic and molecular bounds. Thus, the PLSr model provides information on the wavelengths contribution regarding carbohydrate, protein or lipid contents and COD, methane yield or kinetics of methane production prediction.

A particular interest of the *b* coefficients is found with the Net Analyte Signal (NAS). NAS is the part of the measured signal that a calibration model relates to the analyte concentration. It has been demonstrated that NAS could be estimated by the *b* coefficients (Ferré et al., 2001) and thus, the *b* coefficients can be interpreted as the spectrum of the properties of interest. Hence, the *b* coefficients from each PLS regression model can be compared with spectra of known components.

PLSr *b* coefficients take both positive and negative values; their analysis gives an insight on the contribution of the wavelengths into the prediction. One of the difficulties of *b* coefficients analysis in NIRs is the use of continuous spectra with correlation between the wavelengths instead of independents variables. Negative *b* coefficients on a wavelength can be interpreted as the removal of an undesired molecule absorbance which absorbs in the same wavelength as relevant molecules or as the effect of a correlation within the dataset. *b* coefficients interpretation complexity can be reduced using of variable selection to help assessing the most relevant *b* coefficients. A solution is to use variable selection but the information on the *b* coefficients shape and value are altered. To overcome this problem, sparse-PLS is preferred, performing variable selection by Lasso penalization of the loading vectors (Le Cao et al., 2008).

No strict rules establish a link between the wavelengths and the molecules which increase the complexity of molecules contribution assessment through *b* coefficients analysis. Indications on the potential link between wavelengths and several groups of organic molecules has been provided (Osborne and Fearn, (1986), Williams and Norris, (2001), BenDor et al., (1997)). Table 11 to Table 14 sum up the impact of the molecules on the spectra at different wavelengths.

Table 11: Influence of alkane and alkene on NIR spectra. The black marks represent the range of absorbance

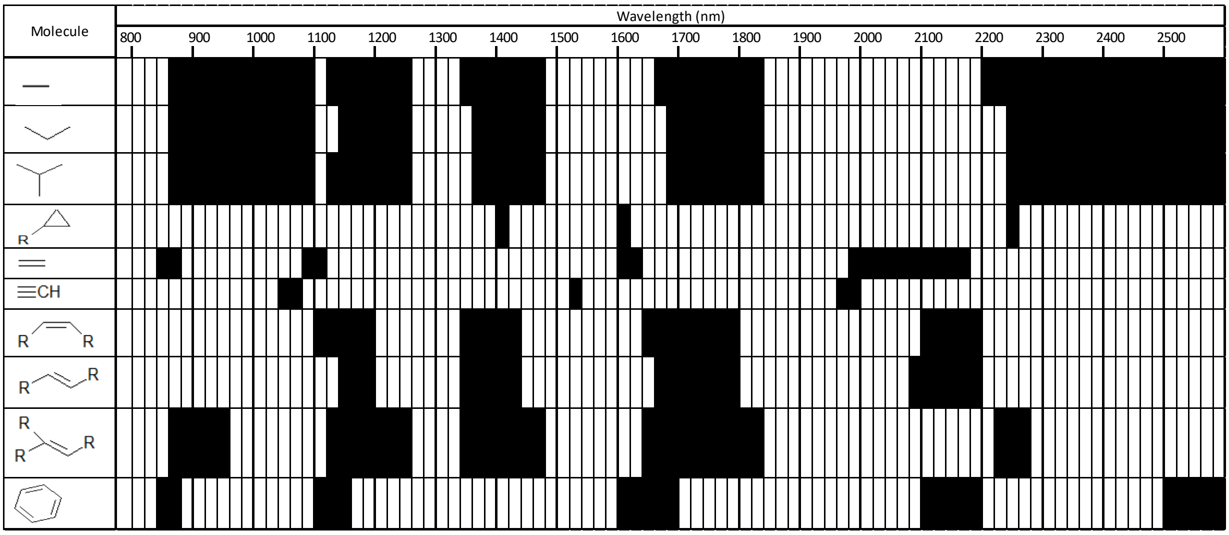

Molecule
Wavelength (nm)
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
2100
2200
2300
2400
2500
R
≡CH
R
R
R
R
R
R
R

Table 12: Influence of alcohol, ketone, ester, epoxide, aldehyde and carboxylic acid on NIR spectra. The black marks represent the range of absorbance

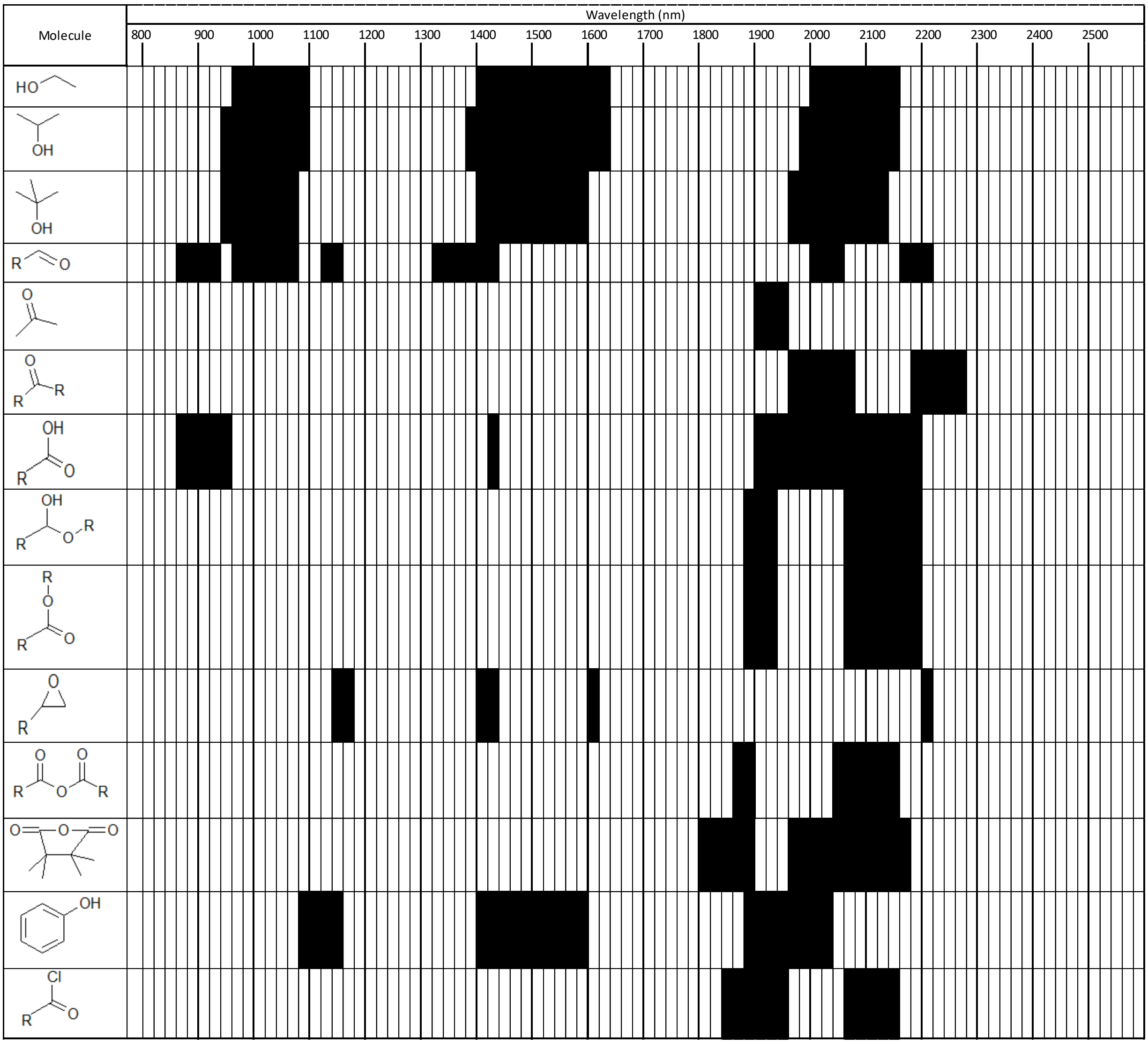


Table 13: Influence of organic molecules containing nitrogen on NIR spectra. The black marks represent the range of absorbance

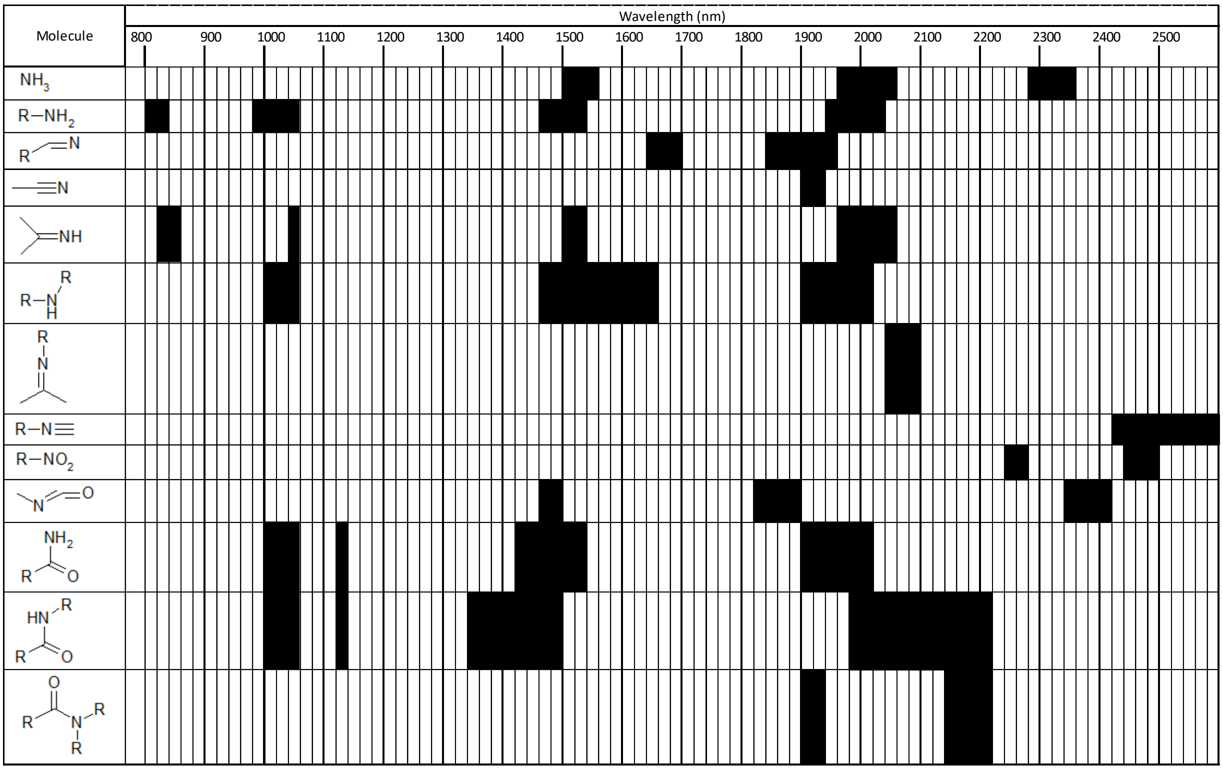
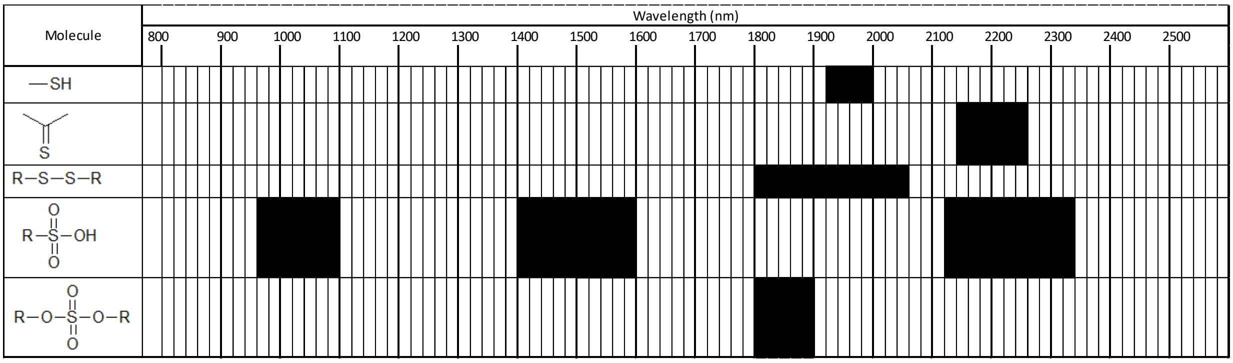


Table 14: Influence of organic molecules containing sulfur on NIR spectra. The black marks represent the range of absorbance



Despite the difficulty of interpretation, *b* coefficients analysis is considered as a relevant tool to better understand the link between the organic matter composition and the indicators used for anaerobic digestion. In the first instance, the current study will assess the contribution of wavelengths involved in the estimation of AD indicators through the *b* coefficients analysis. The relation between these wavelengths and the potential molecules represented from Table 11 to Table 14 will be established and discussed. Eventually cosine between the *b* coefficients of different PLSr model calibrations will be estimated to highlight the relations between the models.

6.4 Materials and Methods Related to this Chapter

Detailed information about the materials and methods related to this chapter are provided in the “Materials and Methods“ sections 2.3.1, 2.3.3.1, 2.4.1, 2.5.1, 2.6.1, 2.6.3.1, 2.6.3.2.

6.5 Homogenization of the Chemometric Treatment

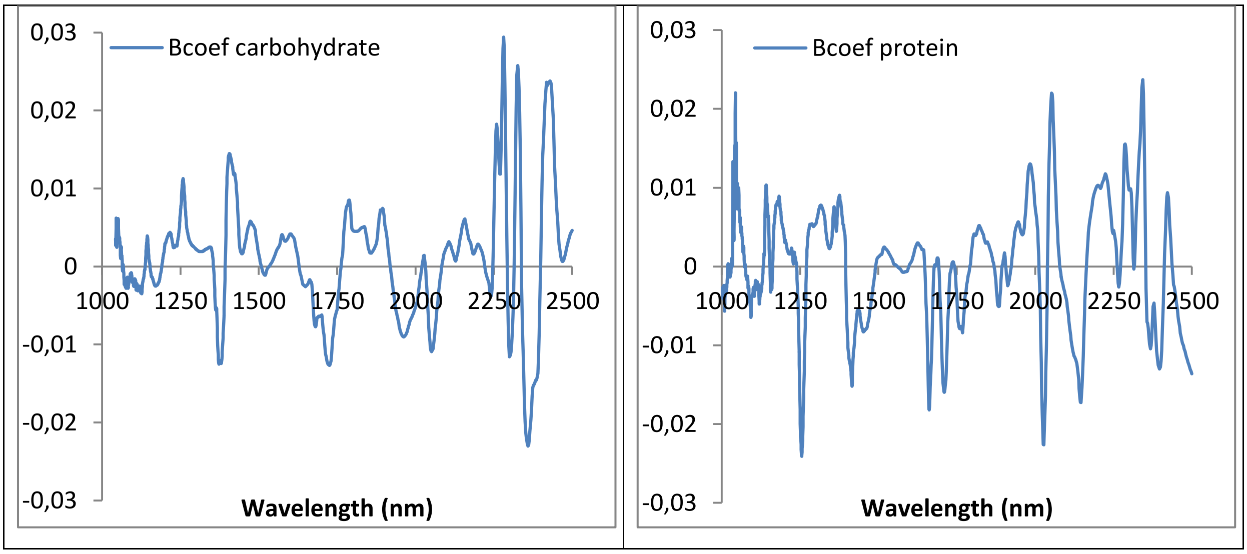
As does the spectrum, *b* coefficients are strongly impacted by the chemometric treatment selected and it is difficult to compare the *b* coefficients selected on different chemometric treatments. However it is relevant to assess the relation between the *b* coefficients selected as the NAS of COD, methane yield and production time or carbohydrate, protein and lipid contents. To overcome the problem of homogeneity between the *b* coefficients, it was decided to use the same chemometric treatment for every PLSr model. According to Lesteur et al., (2011) and Chapter 5, if only one chemometric treatment had to be applied before every PLSr model, absorbance/SNV/detrending is the most appropriate one, providing the best accuracy. The prediction using this pretreatment and sparse PLSr performances were assessed in order to insure the validity of the associated models (see “Materials and Methods” section 2.6.3.2). Table 15 sums up the number of latent variables, the number of variables to keep in *X*-loadings (KeepX), Standard Error of Cross Validation (SECV) on the calibration set and the Standard Error of prediction (SEP) on the validation set. The values given for the production time correspond to the quadratic mean value for the 19 points predicted along the methane production. The performances of the sparse PLSr models prediction were comparable to the performances reported in Lesteur et al., (2011) and in Chapter 5. From the cross-validation and validation performances shown in Table 15, sparse PLSr established using absorbance, SNV and detrending as a chemometric pretreatment can be considered as accurate for the prediction of all variables. The resulting *b* coefficients of sparse PLS are considered as relevant to work on.

Table 15: Sparse PLSr performances. Methane production time is expressed as a mean quadratic error for the 19 time predicted along the methane production curve from 5 to 95% of the methane yield.

	Carbohydrate	Protein	Lipid	Methane yield	Biodegradability	COD	Methane production time
Latent variable	11	13	15	10	11	8	5
KeepX	38	51	24	4	20	36	41
SECV	0.070 $mgO_2.mgO_2^{-1}$	0.031 g. gTS^{-1}	0.008 g. gTS^{-1}	0.041 l. gTS^{-1}	0.116 $mgO_2.mgO_2^{-1}$	83.1 $mgO_2.gTS^{-1}$	0.31 d
SEP	0.034 $mgO_2.mgO_2^{-1}$	0.025 g. gTS^{-1}	0.011 g. gTS^{-1}	0.027 l. gTS^{-1}	0.087 $mgO_2.mgO_2^{-1}$	44.5 $mgO_2.gTS^{-1}$	0.31 d

6.6 *b* Coefficients Analysis and Molecules Contribution to the Predicted Parameters

In order to select only the most relevant variables involved for the prediction, *b* coefficients of sparse-PLS were considered. Because sparse-PLS selects the variables with a maximal covariance between wavelength intensities and values to predict, these variables can be considered as fundamental and linked with molecules strongly related with the parameters to predict. The complexity of the *b* coefficients in Figure 47 indicates that many wavelengths are considered as fundamental by the sparse PLSr. The high number of peaks could be related to the diversity of substrates used and the complexity of the parameters to predict. The *b* coefficients were analyzed and compared to the assignments found in the literature summed up from Table 11 to Table 14.



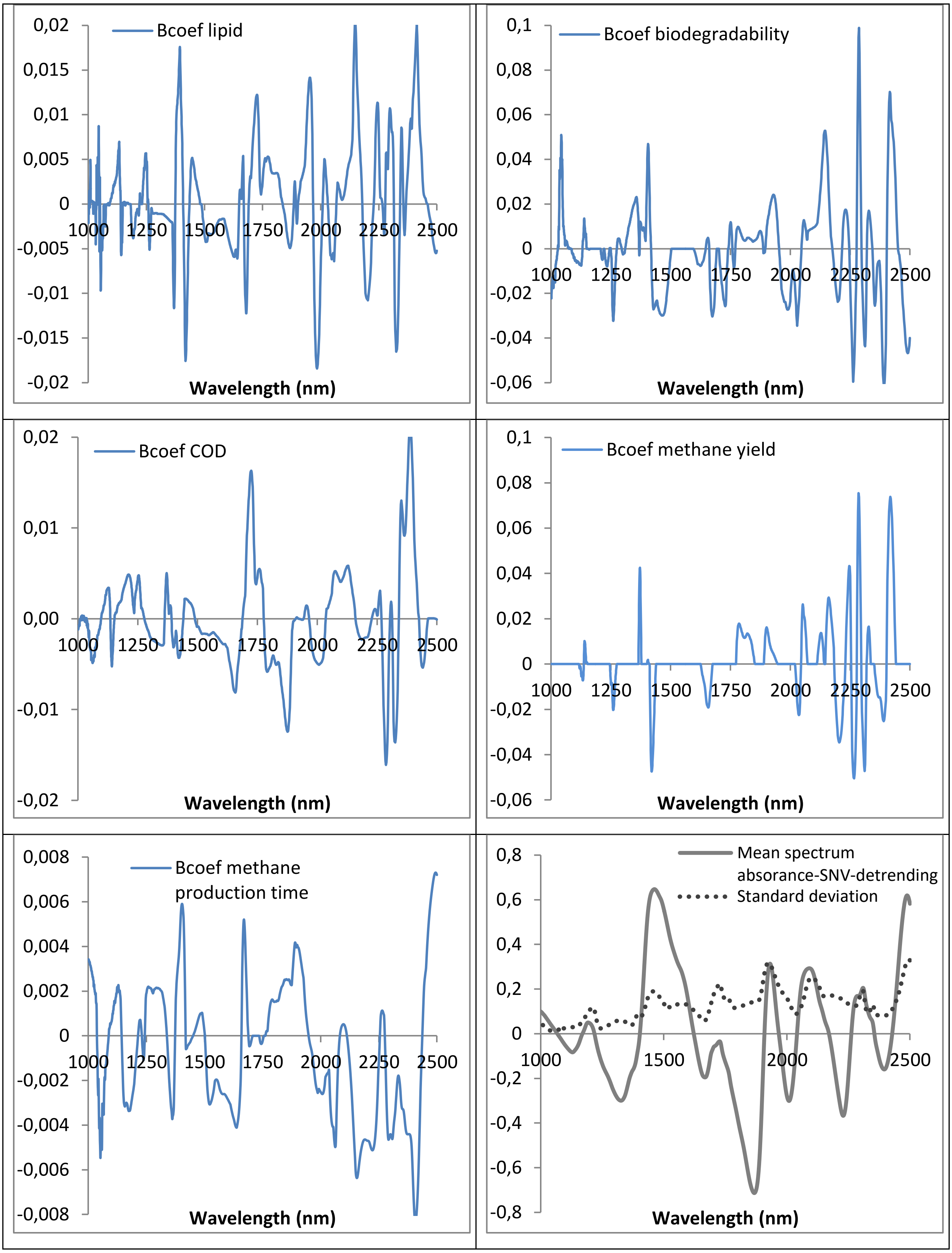


Figure 47: *b* coefficients (Bcoef) of carbohydrate, protein and lipid contents, biodegradability, COD, methane yield and methane production time prediction from spectra pretreated by absorbance/SNV/detrending

b coefficients involved in the prediction of carbohydrate content revealed positive peaks at 1045-1142-1250-1422-1486-1607-2034-2256-2327-2417-2424 and from [1764-1906] nanometers. It revealed negative peaks at 1380-1720-1970-2297-2358. It can be observed that one wavelength often implies many molecules. Indeed, NIR signal results from the interaction of both molecule bounds and matrix effect. It is sometimes difficult to separate the influence of each family of molecules for the *b* coefficients analysis. Nonetheless, for carbohydrate content prediction, comparing data from Table 11 to Table 14, to the *b* coefficients, it can be seen that alcohols, ketone, aldehyde, carboxylic acid, phenol and epoxy are positively involved while alkane and alkene are both positively and negatively contributing to the carbohydrate content prediction. Ammoniac is positively and negatively contributing to the carbohydrate content while positive contribution of amine and amide seems to be observed. All in all, the prediction of carbohydrate content seems to be in strong relation with molecules containing oxygen (details in Annex of this chapter, table 1). These observations make sense since carbohydrates are composed of alcohols and often of ketone, carboxylic acid or aldehyde. The contribution of amide to the prediction of carbohydrate content could be explained by its presence in carbohydrates but also by the similar absorbance of amide and alcohols that makes the separation of each family of molecules difficult.

b coefficients involved in the prediction of protein content indicate that proteins prediction is associated with positive peaks at 1043-1182-1315-1395-1520-1980-2052-[2180-2220]-2287-2343-2424 nanometers. Wavelengths at 1256-1418-1456-1662-1770-1883-2022-2150-2262-2400-2500 nanometers contribute negatively. Alkane, alkene, benzene, alcohol, aldehyde, ketone, carboxylic acid, ester, epoxy, phenol, sulfonic acid and disulfide were associated with many of the *b* coefficients. The study reveals an equivalent number of positive and negative contributions for each of them. In addition, a majority of positive *b* coefficients were associated with the molecules containing nitrogen such as ammoniac, amines, ketamine and amides. Many of the positive *b* coefficients in relation with molecules containing proteins were also in relation with alkane, alkene, benzene, alcohol, aldehyde, ketone, carboxylic acid, ester, epoxy, phenol, sulfonic acid or disulfide. These positive *b* coefficients are counterbalanced by negative ones in the case of molecules non containing nitrogen (details in Annex of this chapter, table 2). Thus, it can be concluded that the sparse PLSr model targeted molecules containing nitrogen and subtracted the contribution of molecules rich in oxygen with negative *b* coefficients. Molecules rich in oxygen have a negative impact on the protein content prediction since they are strongly correlated with the prediction of carbohydrate content which is itself anti-correlated with protein content (see Chapter 4, Figure 30).

b coefficients involved in the prediction of lipids indicate that lipids prediction is associated with positive peaks at 1000-1050-1395-1663-1722-1770-1950-2145-2242-2329-2402 nanometers and with negative peaks at 1260-1418-1678-2050-2198-2210-2268-2323-2394 nanometers. Alcohols, benzene, aldehyde, ketone, carboxylic acid, anhydric acid, amides, thione and sulfonic acid are either positively or negatively contributing to the lipid content prediction. Unexpectedly, carboxylic or ester function were ignored by the sparse PLSr model even if present in many fatty acids. A first explanation would be its low relative contribution to the fatty molecules and a second one, its contribution to both lipids and carbohydrates which are anti-correlated (see Chapter 4, Figure 30). Ammoniac seems to contribute negatively to lipid content but amines contribute positively. These two contributions could be due to absorption wavelengths shared with molecules of interest for the lipid prediction. Eventually, it can be observed that the positive *b* coefficients are mainly associated with alkane and alkene, which represent the main part of fat molecules (details in Annex of this chapter, table 3).

Overall, the analysis of the *b* coefficients shows that they are often associated with many known molecules. It has been seen that molecules rich in oxygen contribute to carbohydrate content prediction, while molecules rich in nitrogen contribute to protein content prediction and alkane and alkene contribute to lipid content prediction. These relations are in accordance with the biochemical composition of carbohydrates, proteins and lipids. Negative *b* coefficients could be explained by the anti-correlation in the calibration set between carbohydrate content and; protein and lipid contents reported in Chapter 4. It shows that the relation deriving from the *b* coefficients analysis is reliable.

b coefficients analysis has been extended to COD, methane yield, biodegradability and methane production time in which less background on the expected contribution is available. *b* coefficients involved in the prediction of methane yield indicated that methane prediction is associated with positive peaks at 1143-1372-1780-1900-2054-2133-2160-2248-2285-2418 nanometers and with negative peaks at 1420-1657-2264-2390 nanometers. A limited number of variables have been selected by the sparse PLSr, strongly related with methane yield prediction. Alkane, alkene, alcohols, carboxylic acid, aldehyde, benzene sulfonic acid, amine and amide were mainly associated with positive peaks (details in Annex of this chapter, table 4). The current analysis of *b* coefficients suggests that alkane, alkene, alcohols, carboxylic acid, aldehyde, amine and amide are biodegradable and contribute to the methane yield.

b coefficients involved in the prediction of COD indicate that COD prediction is associated with positive peaks at 1150-1210-1450-1676-1721-1770-2064-2134-2308-2351-2390 nanometers and with negative peaks at 1000-[1045-1100]-1256-1400-1660-1880-2030-2287-2325 nanometers. The wavelengths associated with the *b* coefficients imply many molecules. Alkane, benzene, ketone, carboxylic acid, ester, anhydric acid, amides and sulfonic acid are either positively or negatively contributing to the COD prediction. As expected, alcohols are negatively related with COD prediction.

Indeed the oxygen contained in alcohols reduces the COD as shown by the negative correlation between COD and carbohydrate content (see Chapter 4, Figure 30). The analysis revealed a negative contribution of amines to COD (details in Annex of this chapter, table 5). It could be explained by a relation between amines and oxygen-containing groups as for instance in glycoproteins. Alkene was related with positive *b* coefficients. Indeed, alkene has an important COD, nonetheless, the contribution of alkane to COD prediction was expected to be more significant (Long et al., 2012).

b coefficients involved in the prediction of biodegradability indicate that peaks at 1355-1406-1658-2052-2145-2228-2287-2415 nanometers contribute positively and peaks at 1257-1456-1673-1728-1982-2035-2262-2314-2347-2390 nanometers contribute negatively to the prediction of biodegradability. Either positive or negative *b* coefficients were associated with alkane, alkene, alcohols, ketone, carboxylic acid, ammoniac, amines, ketimine, amides, disulfide and sulfonic acid. Positive *b* coefficients were associated with aldehyde, anhydric acid. Negative *b* coefficients were associated with phenol and isocyanate (details in Annex of this chapter, table 6). The presence of isocyanate in the sample studied can be questioned because it is a toxic and reactive compound. Nonetheless its toxicity could explain a negative contribution to biodegradability. The negative contribution of phenol was expected. Phenol is, indeed, highly represented in lignin which represents a non-degradable organic matter (Buffiere et al., 2006). Many of the molecules with either negative or positive contributions to biodegradability prediction have been reported with a positive effect on both COD and methane which could explain their low impact on biodegradability prediction since biodegradability is correlated with methane yield but anti-correlated with COD (see Chapter 4, Figure 30).

b coefficients involved in the prediction of methane production-time indicate that peaks at 1000-1125-[1250-1310]-1400-1670-1908-2497 nanometers contribute positively and peaks at 1052-1174-1368-1520-1642-1984-2062-2157-2224-2291-2406 nanometers contribute negatively. Either positive or negative *b* coefficients were associated with aldehyde, ester, epoxy, phenol, disulfide and sulfonic acid. Negative *b* coefficients were associated with alkane, alkene, alcohols, ketone, carboxylic acid, anhydric acid, ammoniac, amines and amides. Positive *b* coefficients were mainly associated with nitrile indicating a positive contribution of nitrile (details in Annex of this chapter, table 7). One reaction that can possibly occur on nitrile is its reduction in amine. This reduction would use the reduction potential needed for the synthesis of methane, inhibiting its production. The contribution found on the *b* coefficients analysis shows that alkane, alkene, alcohols, ketone, carboxylic acid, anhydric acid, ammoniac, amines and amides decreased the degradation time. These results indicate that these molecules are readily bio-degradable.

From the study of the *b* coefficients, interesting contributions have been highlighted. Nonetheless, in many cases the complexity of the spectrum resulting from NIRs analyses made the relations difficult to interpret. The wavelength related with each molecule also depends on the matrix and the molecular environment (Nicolai et al., 2007). Thus the data found in the literature are not specific enough to draw some exact conclusions on the molecules related with the parameters predicted by the PLSr.

6.7 Relations between the b Coefficients of the Models

b coefficient from the PLS regression models established can be compared with each other as the NAS of the property of interest by analyzing the cosine of *b* coefficients of each PLS regression model. It makes it possible to compare, for instance, the similarity between features used to predict the biochemical composition of the substrate and the features used to predict the methane yield or production time. A matrix of collinearity was established between the *b* coefficients of the PLSr model for the prediction of carbohydrate, protein and lipid content, biodegradability, methane yield, COD and methane production time using the cosine between the *b* coefficients, as detailed in “Materials and Methods” section 2.6.1. The matrix is drawn as a heatmap in Figure 48.

The analysis of cosine reveals that carbohydrate content predictors (i.e. *b* coefficients of the model) were positively related with biodegradability and methane yield predictors, while negatively related with the COD predictors (Figure 48). It shows a negative impact of the molecules involved in carbohydrate content predictors on COD. Indeed, carbohydrates contain oxygen that reduces the COD. On the other hand, carbohydrates are mainly composed of biodegradable compounds that affect positively both methane yield and biodegradability as it was demonstrated experimentally by Labatut et al., (2011).

Protein content predictors were related with methane yield and biodegradability predictors too. Molecules rich in nitrogen contributing to protein content predictors can thus be considered as highly biodegradable. In section 3.3.1, is has been seen that total ammonia nitrogen (TAN) inhibits the AD process. Thus, the high biodegradability of the protein we highlighted indicates that the majority of the organic nitrogen introduce in the plant will be biodegraded to TAN which may leads to process inhibition. Nitrogen is rarely found in lipids or carbohydrates as highlighted by the negative relation between protein and carbohydrate or lipid contents predictors. Proteins predictors were also negatively related with COD predictors. It indicates that molecules with a high COD such as lipids were not acting on proteins prediction.

b coefficients involved in lipid content prediction presented a strong similarity with the one involved in COD prediction. Lipids are poor in oxygen and mainly composed of C and H which implies a high COD. The same conclusion can be drawn for methane yield supposing the COD of the substrate as biodegradable. Nonetheless, the relation between lipid content and methane yield predictors is weaker than between lipids and COD predictors. This indicates that an important part of the lipids are biodegradable and contribute to both methane yield and COD prediction. However, it also indicates that some lipids contribute only to COD prediction and are thus considered as not biodegradable.

Methane yield predictors and biodegradability predictors have important similarity. Biodegradability is the result of the ratio methane yield to COD (Equation 18, section 4.3.2). Thus a high methane yield implies a high biodegradability. It can be noticed that methane yield and COD *b* coefficients are very poorly anti-collinear. This absence of relation is explained by the independency of methane yield and COD. It is quite an important result. Indeed, in many AD plants, COD is followed instead of methane yield because COD is faster and easier to obtain (Jimenez et al., 2015b). But, this implies that COD and methane yield are correlated. However, in our study, the independence of the *b* coefficients used for methane yield and COD prediction and their low correlation (see Chapter 4) shows that this hypothesis is far from being always true. This could explain why plant performances are often below the operator expectation. The independence of methane yield and COD also explains the anti-collinearity between COD and biodegradability predictors and confirms that NIRs and PLSr enable to distinguish biodegradable and non-biodegradable molecules. Based on equation 18 (section 4.3.2), if the COD increases and the methane yield remain constant, biodegradability decreases.

The relations between the *b* coefficients of the methane production time and the *b* coefficients of the methane yield, COD, biodegradability, carbohydrate, protein and lipid contents provide some clues about the impact of different molecules on the methane production time. It reveals very low similarity between the methane production time and; carbohydrate and lipid content, COD and methane yield predictors. Carbohydrates and lipids represent huge families of molecules. In these families, the kinetics of degradation varies a lot. For instance, the degree of polymerization, the crystalline structure and the type of bonds between the monomers impact a lot the process of biodegradation (Resch et al., 2014).

The current results relate negatively the methane production time with biodegradability and protein content predictors, which indicates collinearity with the kinetics of methane production. The higher the biodegradability or protein content predicted, the faster the production of methane will be. It means that the predictors of protein content, and thus of nitrogen rich molecules, are associated with readily degradable matter. In vegetal organic matter, proteins or nitrogen rich molecules usually have a role of energetic reserve. The role of proteins in animals is pretty different and they ensure many key functions. Proteins are commonly found in polymers for example muscle fibers. The similarity between kinetics and proteins predictors indicates that molecules rich in nitrogen considered either as reserves or fibers are readily bio-degradable. Non-degradable organic matter predictors are linked with complex polymeric structure such as lignin as detailed in section 3.2. The hydrolysis of such structures is slow and incomplete because of the difficulty for enzymes and micro-organisms to access and degrade the organic matter while soluble organic matter is often readily accessible and degradable. Thus, the more biodegradable the organic matter, the faster its degradation.

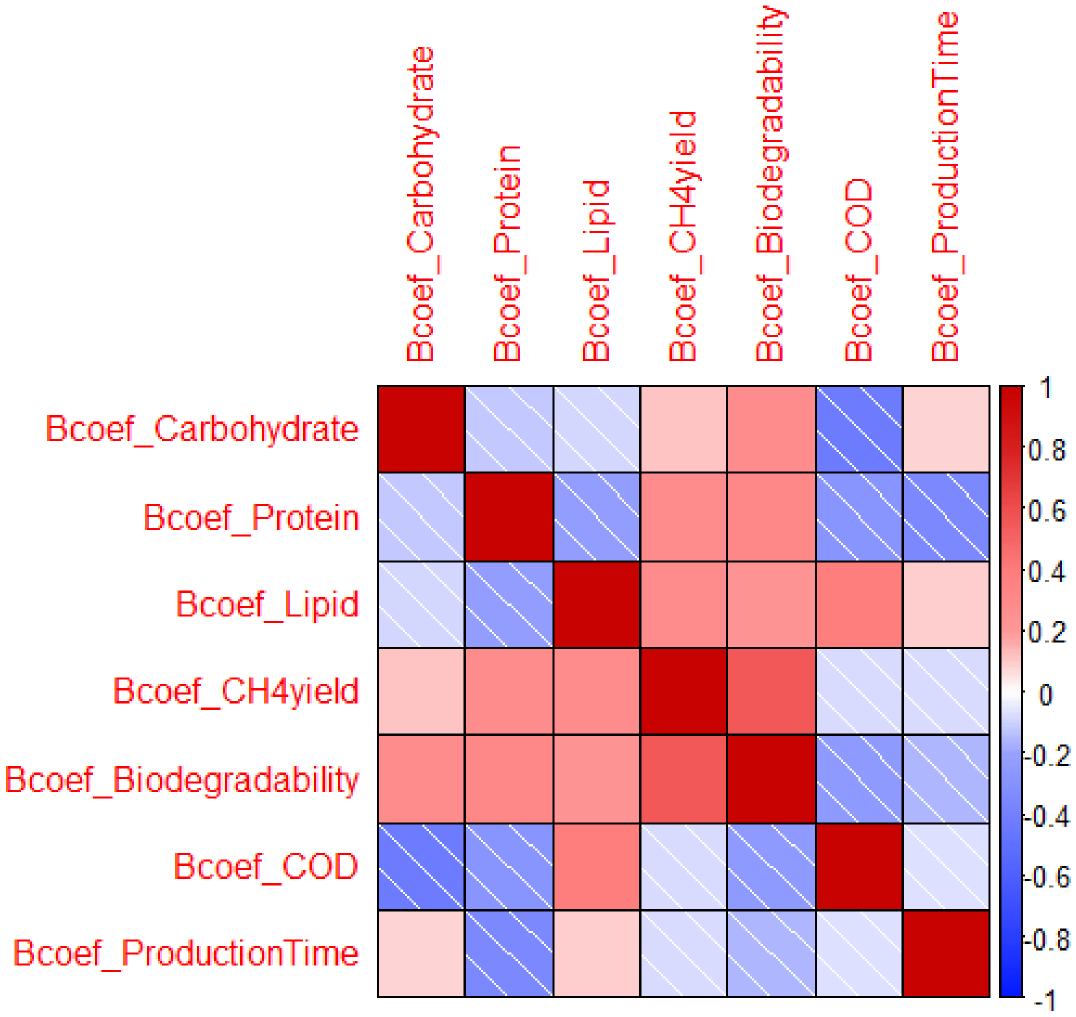


Figure 48: *b* coefficients (Bcoef) cosine matrix of carbohydrate, protein and lipid contents, methane yield (CH4Yield), COD, biodegradability and methane production time (ProductionTime).

6.8 Conclusion

This study analyzes the *b* coefficients of PLSr models for the prediction of carbohydrate, protein, lipid contents and COD, biodegradability and methane production yield and kinetics. The simultaneous analysis of the *b* coefficients for these 7 parameters provides the opportunity to compare the *b* coefficients with the reported absorbance of molecules on NIR spectra but also to analyze the correlations between *b* coefficients for different predictive models. It revealed that:

- Carbohydrate content predictors are related to oxygenated molecules content such as alcohols. The carbohydrate content impacts negatively the COD prediction but positively the methane yield and biodegradability.
- Protein content predictors are related to any molecules containing nitrogen. The associated molecules are readily and highly biodegradable.
- Lipid content predictors are related with alkane and alkene content. Moreover, the *b* coefficients implied in the lipids content prediction have important similarities with the one implied in COD and methane yield prediction.
- Biodegradability is negatively related with lignin and positively with kinetics of methane production.
- In contradiction to the mainstream thought, *b* coefficients implied in the COD and methane yield prediction are anti-collinear and revealed low relation with methane kinetics predictors.

This highlighted the relations between the parameters mainly used to pilot AD plant. It is hoped that it will help human plant operators to determine the optimal feeding of the digesters with a better understanding of the links between the biochemical composition, COD, methane production yield and kinetics. This study also proposed the use of NIRs and *b* coefficients analysis to get an insight of the organic matter composition of complex samples.

Annex of the chapter

Table 1: Relation between *b* coefficients and molecules involved in the prediction of carbohydrate content

Molecule	Wavelength reported with positive b coefficients (nm)	Wavelength reported with negative b coefficients (nm)
Alkane	1045-1142-1250-1422-2256-2327-2417-2424	1380-1720-2297-2358
Alkene	1142-1422	1380-1720
Alkyne	1045	1970
Benzene	1142-1607	
Primary alcohol	1045-1422-1486-1607	
Secondary alcohol	1045-1422-1486-1607	1380-1970
Tertiary alcohol	1045-1422-1486-1607	
Aldehyde	1045-1142-1422-2034	1380
Ketone	2034-2256	1970
Carboxylic acid	2034	1970
Ester	2034	
Epoxy	1142-1422-1607	
Anhydric acid		
Cyclo anhydric acid	[1764-1906]	
Phenol	1142-1422-1486-2034	1970
Acyl halide		
Ammoniac	2034-2327	1970-2297-2358
Primary amines	1045-1422-1486-1607-2034	1970
Primary aldimine		
Nitrile	2424	
Primary ketimine	1045	1970
Secondary amine	1045	1970
Secondary ketimine		
Nitro	2256	
Isocyanate	2256-2417	2358
Primary amides	1045-1422-2304	1970
Secondary amides	1045-1422-2034	1380
Tertiary amides		
Sulfhydryl		1970
Thione	2256	
Disulfide	2034	1970
Acid sulfonic	1045-1422-2056	2297
Diester sulfonic	[1764-1906]	

Table 2: Relation between *b* coefficients and molecules involved in the prediction of protein content

Molecule	Wavelength reported with positive b	Wavelength reported with negative b

	coefficients (nm)	coefficients (nm)
Alkane	1043-1182-1395-[2180-2220]-2287-2343-2424	1256-1418-1456-1662-1770-2262-2400-2500
Alkene	1182-1395-1980-2052--[2180-2220]	1418-1456-1662-1770-2022-2050-2262
Alkyne	1980	
Benzene	[2180-2220]	1662-2050-2500
Primary alcohol	1043-1520-2052	1418-1456-2022-2050
Secondary alcohol	1043-1395-1520-1980-2052	1418-1456-2022-2050
Tertiary alcohol	1043-1520-1980-2052	1418-1456-2022
Aldehyde	1395-2052--[2180-2220]	1418-1456-2022
Ketone	1980-2052-[2180-2220]	2022-2262
Carboxylic acid	1980-2052-[2180-2220]	2022-2050
Ester	[2180-2220]	1883-2050
Epoxy	[2180-2220]	1418
Anhydric acid		2050
Cyclo anhydric acid	1980-2052	2050
Phenol	1520-1980	1418-1456
Acyl halide		1883-2050
Ammoniac	1520-1980-2287	2022
Primary amines	1043-1520-1980-2052-2343	2022
Primary aldimine		1662-1883
Nitrile	2424	2500
Primary ketimine	1520-1980-2052	2022
Secondary amine	1043-1520-1980	
Secondary ketimine	2052	
Nitro		
Isocyanate	2343	1883-2262-2400
Primary amides	1043-1520-1980-[2180-2220]-	1456
Secondary amides	1043-1395-1980-2052-[2180-2220]-	1418-1456-2022-2050
Tertiary amides		
Sulfhydryl	1980	
Thione	-[2180-2220]-	2262
Disulfide	1980-2052	1883-2022
Acid sulfonic	1043-1520-[2180-2220]-2287	1418-1456-2050-2262
Diester sulfonic		

Table 3: Relation between *b* coefficients and molecules involved in the prediction of lipid content

Molecule	Wavelength reported with positive b coefficients (nm)	Wavelength reported with negative b coefficients (nm)
Alkane	1000-1050-1134-1395-1722-1770-2242-2402	1370-1418-2323
Alkene	1134-1395-1663-1722-1770-2145-2242	1370-1678-1982-2198
Alkyne	1050-2145	1982
Benzene	1663	1678-2198
Primary alcohol	1000-1050-2145	1418-2050
Secondary alcohol	1000-1050-1395-2145	1418-1982-2050
Tertiary alcohol	1000-1050	1418-1982-2050
Aldehyde	1000-1050-1134-1395	1370-1418-2050-2198
Ketone	1950-2242	1982-2050
Carboxylic acid	1950-2145	1982-2198
Ester	2145	2198
Epoxy		1418
Anhydric acid	2145	2050
Cyclo anhydric acid	2145	1982-2050
Phenol	1134-1950	1418-1982
Acyl halide	1950	
Ammoniac		1982-2050-2323
Primary amines	1000-1050-1950	1982
Primary aldimine	1663-1950-2145	
Nitrile		
Primary ketimine	1050	1678-1982-2050
Secondary amine	1000-1050-1950	1982
Secondary ketimine		
Nitro	2242	
Isocyanate	2402	
Primary amides	1000-1050-1134-1950	1982-1982
Secondary amides	1000-1050-1134-1395-2145	1370-1418-2050-2198
Tertiary amides		2198
Sulfhydryl	1950	1982
Thione	2145-2242	2198
Disulfide	1950	1982-2050
Acid sulfonic	1000-1050-2145-2242	1418-2198-2323
Diester sulfonic		

Table 4: Relation between *b* coefficients and molecules involved in the prediction of methane yield

Molecule	Wavelength reported with positive b coefficients (nm)	Wavelength reported with negative b coefficients (nm)
Alkane	1143-1372-1780-2248-2285-2418	1420-2264-2390
Alkene	1143-2054-1372-1780-2133-2160-2248	1420-1657
Alkyne		
Benzene	1143-2133-2160	1657
Primary alcohol	2054-2133-2160	1420
Secondary alcohol	2054-2133-2160	1420
Tertiary alcohol	2054-2133-2160	1420
Aldehyde	1143-1372-2054	
Ketone	1900-2054-2248	2264
Carboxylic acid	1900-2054-2133-2160	2264
Ester	1900-2133-2160	
Epoxy	1143	1420
Anhydric acid	2054-2133-2160	
Cyclo anhydric acid	2133-2160	
Phenol	1143-1900	
Acyl halide	1900-2133-2160	
Ammoniac	2054-2285	
Primary amines		
Primary aldimine	1900	1657
Nitrile	1900	
Primary ketimine	2054	
Secondary amine	1900-2054	1657
Secondary ketimine		
Nitro		2264
Isocyanate	2418	2390
Primary amides	1900	
Secondary amides	1372-2054-2133-2160	1420
Tertiary amides	1900-2133-2160	
Sulfhydryl		
Thione	2160-2248	2264
Disulfide	1900-2054	
Acid sulfonic	2160-2248-2285	2264
Diester sulfonic		

Table 5: Relation between *b* coefficients and molecules involved in the prediction of COD

Molecule	Wavelength reported with positive b coefficients (nm)	Wavelength reported with negative b coefficients (nm)
Alkane	1150-1210-1450-1721-1770-2308-2390	1000-1050-1144- 1256-1400-2287-2325
Alkene	1150-1210-1450-1676-1721-1770-2134	1144-1400-1660
Alkyne		1050
Benzene	1676	1660
Primary alcohol	1450-2064-2134	1000-1050- 1400-2030
Secondary alcohol	1450-2064-2134	1000-1050- 1400-2030
Tertiary alcohol	1450-2064-2134	1000-1050- 1400-2030
Aldehyde	1150-2064	1000-1050-1144- 1400-2030
Ketone	2064	2030
Carboxylic acid	2064	2030
Ester	2134	1880
Epoxy	1150	1144-1400
Anhydric acid	2064-2134	1880
Cyclo anhydric acid	2064-2134	1880-2030
Phenol	1150-1450	1050-1144- 1400-2030
Acyl halide	2134	
Ammoniac	2064	2030-2287
Primary amines		1000-1050- 2030
Primary aldimine	1676	1660
Nitrile		
Primary ketimine	2064	2030
Secondary amine		1660
Secondary ketimine		
Nitro		
Isocyanate	2390	1880
Primary amides	1450	1000-1050
Secondary amides	1450-2064-2134	1000-1050- 2030
Tertiary amides		
Sulfhydryl		
Thione		
Disulfide	2064	1880-2030
Acid sulfonic	1450-2308	1000-1400-2287-2325
Diester sulfonic		1880

Table 6: Relation between *b* coefficients and molecules involved in the prediction of biodegradability

Molecule	Wavelength reported with positive b coefficients (nm)	Wavelength reported with negative b coefficients (nm)
Alkane	1355-1406-2287-2415	1257-1456-1728-2262-2314-2347-2390
Alkene	1406-2052-2145	1456-1673-1728-1982
Alkyne		1982
Benzene	2145	
Primary alcohol	1406-2052-2145-2228	1456-2035
Secondary alcohol	1406-2052-2145-2228	1456-1982-2035
Tertiary alcohol	1406-2052-2228	1456-1982-2035
Aldehyde	1355-1406-2228	2035
Ketone	2052-2228	1982-2035--2262
Carboxylic acid	2052-2145-2228	1982-2035
Ester	2145	
Epoxy	1406	
Anhydric acid	2052-2145-2228	1982
Cyclo anhydric acid	2052-2145-2228	2035
Phenol	1406	1456-1982-2035
Acyl halide	2145	
Ammoniac	2228-2287	1982-2035-2314
Primary amines	2052-2228	1982-2035
Primary aldimine		
Nitrile		
Primary ketimine	2052-2228	1982-2035
Secondary amine		1982
Secondary ketimine		
Nitro		
Isocyanate	2415	1673-2262-2347-2390
Primary amides		1456-1982
Secondary amides	1355-1406-2052-2145-2228	1456-1982-2035
Tertiary amides		
Sulfhydryl		1982
Thione	2145	
Disulfide	2052-2228	1982-2035
Acid sulfonic	1406-2145-2287	1456-2262-2314
Diester sulfonic		

Table 7: Relation between *b* coefficients and molecules involved in the prediction of methane production time

Molecule	Wavelength reported with positive b coefficients (nm)	Wavelength reported with negative b coefficients (nm)
Alkane	1000-1125-1400-1670-2497	1052-1174-1368-2224-2291-2406
Alkene	1125-1400-1670	1174-1368-1642-1984-2062-2157-2224
Alkyne		1052-1984
Benzene	1125	1642-2157
Primary alcohol	1000-1400	1052-1520-2062-2157
Secondary alcohol	1000-1400	1052-1520-1984-2062-2157
Tertiary alcohol	1000-1400	1052-1520-1984-2062
Aldehyde	1000-1125-1400	1368-2062
Ketone		1984-2062-2224
Carboxylic acid	1908	1984-2062-2157
Ester	1908	2157
Epoxy	1400	1174
Anhydric acid		2062-2157
Cyclo anhydric acid		1984-2062-2157
Phenol	1400-1908	1520-1984
Acyl halide	1908	2157
Ammoniac		1520-1984-2062
Primary amines	1000	1052-1520-1984-2291
Primary aldimine	1670-1908	1642
Nitrile	1908-2497	
Primary ketimine		1052-1520-1984
Secondary amine	1000-1908	1052-1520-1642-1984
Secondary ketimine		2062
Nitro	2497	
Isocyanate		2406
Primary amides	1000-1125-1908	1052-1520-1984
Secondary amides	1000-1125-1400	1052-1368-1984-2062-2157
Tertiary amides	1908	2157
Sulfhydryl		1984
Thione		2224
Disulfide	1908	1984-2062
Acid sulfonic	1000-1400	1052-2157-2224-2291
Diester sulfonic		

Chapter 7. Challenging the Simultaneous Hydrolysis Concept and Fast Implementation of ADM1 for the Anaerobic Co-digestion Performances Simulation

Chapter 7. Challenging the Simultaneous Hydrolysis Concept and Fast Implementation of ADM1 for the Anaerobic Co-digestion Performances Simulation 169

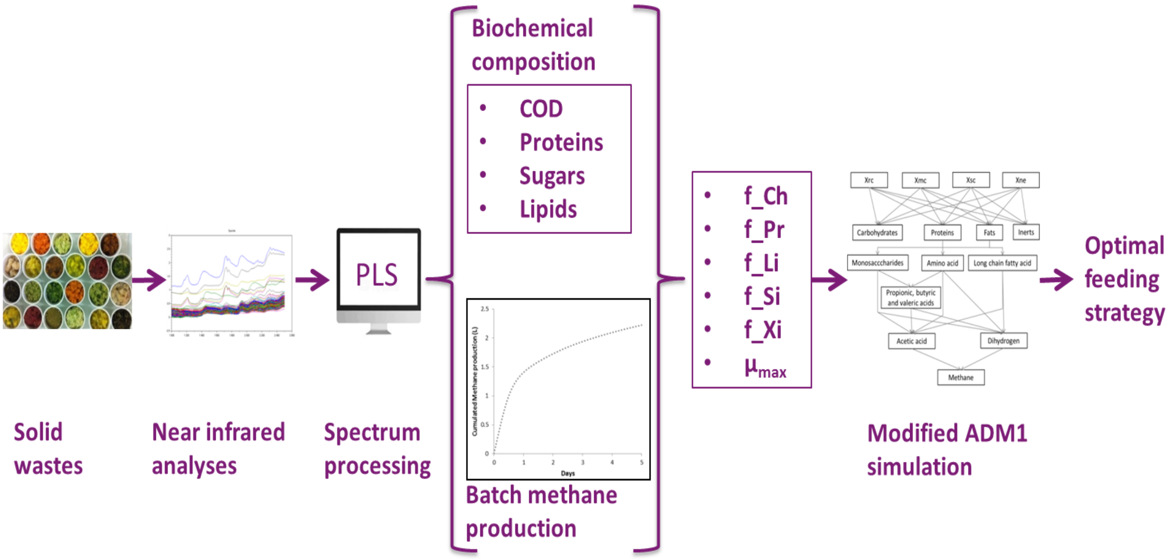
7.1 Chapter Guidelines 171
7.2 Technical and Graphical Abstracts 173
7.3 Introduction of the Chapter 175
7.4 Materials and Methods Related to this Chapter 181
7.5 Modelling Hydrolysis: Simultaneous versus Sequential Biodegradation of the Hydrolysable Fractions 182
    7.5.1 Definition of Simultaneous and Sequential Concepts 182
    7.5.2 Results Obtained on Several Substrates 183
    7.5.3 Fractions Biodegradability: Case Studies on Apple and Wheat Straw 186
    7.5.4 Case of Continuous Reactor 188
    7.5.5 Conclusion on Hydrolysis Modelling 190
7.6 ADM1 Implementation Assisted by Near Infrared Spectroscopy for the Anaerobic Digestion Performances Simulation 191
    7.6.1 ADM1 Calibration from NIR Prediction 191
    7.6.2 Discussion on the MPR Prediction Using NIR 194
7.7 Example of Industrial Application 194
7.8 Conclusion 195

7.1 Chapter Guidelines

The aim of this PhD is to develop an approach that allows predicting the AD process performances based on the feeding strategy. To do so, in Chapter 3, new performant methods for the biological state estimation have been developed. Then in Chapter 5, a fast estimation of the substrate composition using NIRs has been validated. The last step is thus to develop a dynamic model taking into account both biological state estimation and substrate composition. As it will be developed in the introduction of this chapter (7.3), AD has been modelled over the years but solutions remain unsuitable to solid waste co-digestion, mainly because of too simplistic hydrolysis models. The first part of this chapter (7.5) focuses on adapting the hydrolysis model of ADM1 to fit co-digestion of solid waste. Then, fast implementation of ADM1 is made using the PLSr models developed in Chapter 5 and the ability of the ADM1 model to predict experimental methane production is discussed.

7.2 Technical and Graphical Abstracts

Optimization of feeding strategy is an essential issue of co-digestion and requires both fast substrate characterization and predictive models. The hydrolysis mechanism is mechanistically not well understood yet. Consequently, models available in the literature are sometimes not appropriate to describe experimental results. In particular, when multiple fractions of organic matter are present, a model that considers simultaneous degradation of the substrates may not have the resolution to separate the different kinetics. This study first proposes a discussion of structural modelling of organic matter hydrolysis (in aerobic and anaerobic bioprocesses). We assess sequential hydrolysis of the multiple fractions of organic matter as an alternative to simultaneous hydrolysis. In a second part, the fast and efficient estimation of carbohydrate, protein and lipid contents and COD by near infrared spectroscopy is used to implement a modified Anaerobic Digestion Model n°1. Then, the disintegration kinetic parameters are fitted on the methane production rate predicted by near infrared spectroscopy. The development of this fast spectroscopic characterization reduces the time of analysis from 30 days to a matter of minutes. The modified Anaerobic Digestion Model n°1 is then used to simulate the performance of anaerobic digestion batch processes. This approach, validated by simulation of different experiments, aims to a powerful predictive tool for advanced control of anaerobic digestion plants and feeding strategy optimization.


Biochemical composition
COD
Proteins
Sugars
Lipids
PLS
f_Ch
f_Pr
f_Li
f_Si
f_Xi
μ_{max}
Xrc
Xmc
Xsc
Xne
Carbohydrates
Proteins
Fats
Inerts
Monosaccharides
Amino acid
Long chain fatty acid
Propionic, butyric and valeric acids
Acetic acid
Dihydrogen
Methane
Optimal feeding strategy
Cumulated Methane production (L)
Days
Solid wastes
Near infrared analyses
Spectrum processing
Batch methane production
Modified ADM1 simulation

7.3 Introduction of the Chapter

In Chapter 5 and Chapter 6, a fast NIRs characterization of the waste for the estimation of its biochemical composition and methane production performances has been developed and discussed. NIRs analysis of the waste allows a good anticipation of the associated process performances, but it is not sufficient for plant monitoring. Co-digestion development from a large diversity of waste occurs simultaneously with the need for advanced dynamic models like Anaerobic Digestion Model 1 (ADM1) (Batstone et al., 2015). ADM1 consists in a detailed modeling taking into account the main reactions of the AD process (Figure 49) (Batstone et al., 2002). ADM1 has been extensively used and analyzed in both academic and practical applications (Batstone et al., 2006). The use of such models requires a detailed biochemical characterization of the substrates (Jimenez et al., 2014). The first limit encountered while using such a complex model is that many parameters have to be set and depend either on the substrate or the process (Batstone et al., 2015). To set these parameters, a characterization of the substrate as input variables describing the substrate in protein, lipid, carbohydrate contents and Chemical Oxygen Demand (COD), biodegradability and kinetics of hydrolysis is needed. Unfortunately, they are not often measured all together in industrial plants and then, simpler models have to be used (Lauwers et al., 2013). However, the NIR analyses developed in Chapter 5 are a promising solution to solve the matter, giving the required information to implement ADM1 within minutes.

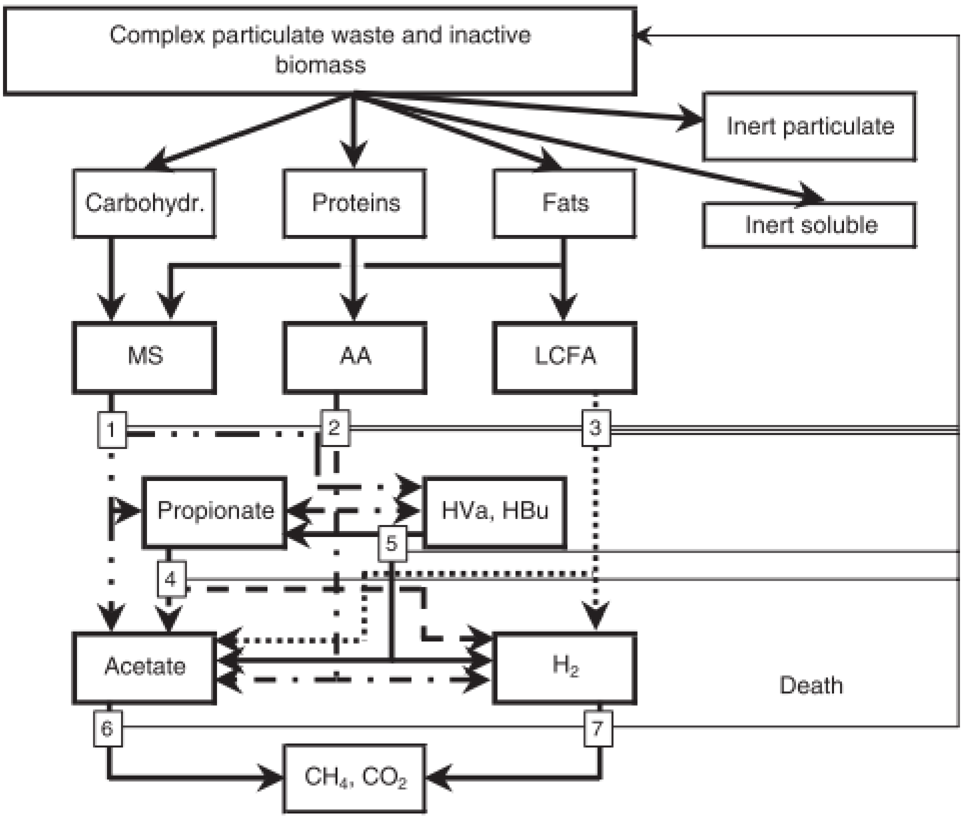


Figure 49: The ADM1 model with the following biochemical processes: (1) acidogenesis from monosaccharides (MS), (2) acidogenesis from amino acids (AA), (3) acetogenesis from long chain fatty acids (LCFA), (4) acetogenesis from propionate, (5) acetogenesis from butyrate (HBu) and valerate (HVa), (6) aceticlastic methanogenesis, and (7) hydrogenotrophic methanogenesis *(Batstone et al., 2002)*.

The second limit is that hydrolysis is recognized as the rate-limiting step (Appels et al., (2008), Vavilin et al., (2008)) and its modelling remains challenging. Hydrolysis is the breakdown of a polymer into monomers by addition of water molecules (Brock et al., 1991). This is a very general definition of the processes occurring in practice. In mathematical models, hydrolysis must be adequately described to be able to predict spatial and temporal availabilities of organic substrate for nutrient removal processes (Morgenroth et al., 2002). As process modelling needs to properly model the limiting step, a focus on hydrolysis models in waste biodegradation is required.

According to a review made by Morgenroth et al. (2002), hydrolysis stoichiometry and kinetics in waste treatment applications are not well understood and first order processes are applied as an aggregate approximation (Eastman and Ferguson, 1981).

Different approaches to hydrolysis have been applied for aerobic and anaerobic models. In the case of aerobic process models, since the first developed activated sludge models (ASM) (Henze et al., 1987), the hydrolysis concept has been challenged, with first order processes generally dominating in applications due to difficulties in identifying higher order models. However, with ASM models, two biodegradable fractions were considered, according to a biological response and not to a physical separation (Ekama et al., 1986; Henze et al., 1987). The associated kinetics was based on a surface-limited equation and one micro-organism population. Ekama and Marais, (1979) divided the wastewater into two biodegradable fractions that are degraded at two different rates, a readily biodegradable fraction consisting mainly of soluble organic matter and a slowly biodegradable fraction that consists of large molecules, colloids and particles that have to be hydrolyzed before degradation. The distinction between these two fractions was also determined by experimental biological response analysis (Ekama et al., 1986; Spérandio and Paul,(1999)).

In the case of anaerobic process models, only one biodegradable fraction was first considered. This biodegradable fraction was then split into biochemical fractions (i.e. carbohydrates, lipids, proteins, inerts) after a disintegration process as in the ADM1 (Batstone et al., 2002). Other previous studies such as (Shimizu et al., 1993) considered that the input variables are the biochemical fractions. First order equations have generally been applied.

Despite these differences, when model simulations do not match with the experimental data, the strategy is the same for both aerobic and anaerobic process modelers: (i) increase the number of hydrolysable fractions (Orhon et al., 1998; Sperandio and Paul, 1999; Yasui et al., 2008; Mottet et al., 2013; García-Gen et al., 2015), (ii) replace the first order kinetics by surface limitation equations (i.e. Contois, Vavilin et al., 2008; Mottet et al., 2013), or (iii) include a particle size distribution model (Dimock and Morgenroth, 2006; Sanders et al., 2000; Yasui et al., 2008).

Vavilin et al. (2008) compared several kinetics of hydrolysis to predict anaerobic batch tests results. They concluded that the use of Contois kinetics on two phases (colonization and hydrolysis) were more effective than first order or Monod equations. Indeed, Mottet et al. (2013) introduced two hydrolysable fractions into ADM1 and considered Contois kinetics for hydrolysis and disintegration to predict biodegradation of a secondary sludge, in a continuous pilot scale digester (Figure 50). Results showed a significant improvement of biogas production fit using these modifications in comparison with the classical ADM1.

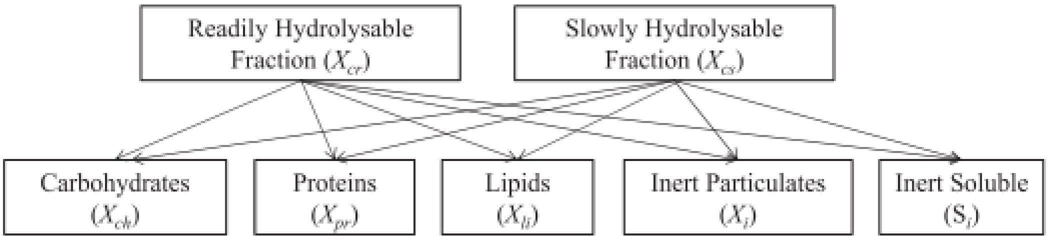

Readily Hydrolysable Fraction (X_{cr})
Slowly Hydrolysable Fraction (X_{cs})
Carbohydrates (X_{ch})
Proteins (X_{pr})
Lipids (X_{li})
Inert Particulates (X_i)
Inert Soluble (S_i)

Figure 50: Disintegration COD flux in the modified model including two different fractions of composite, Xcr and Xcs *(Mottet et al., 2013)*.

Based on a common conclusion that the hydrolysis rate decreases when particle size increases, particle breakup models (Dimock and Morgenroth, 2006; Yasui et al., 2008) and shrinking particle models (Sanders et al., 2000) have been utilized (Figure 51). For example, Sanders et al. (2000) used surface-based kinetics, where the rate of hydrolysis is proportional to the available surface area of slowly biodegradable organic matter. This particle breakup can result in an increase of the available surface area while hydrolysis progresses (Dimock and Morgenroth, 2006).

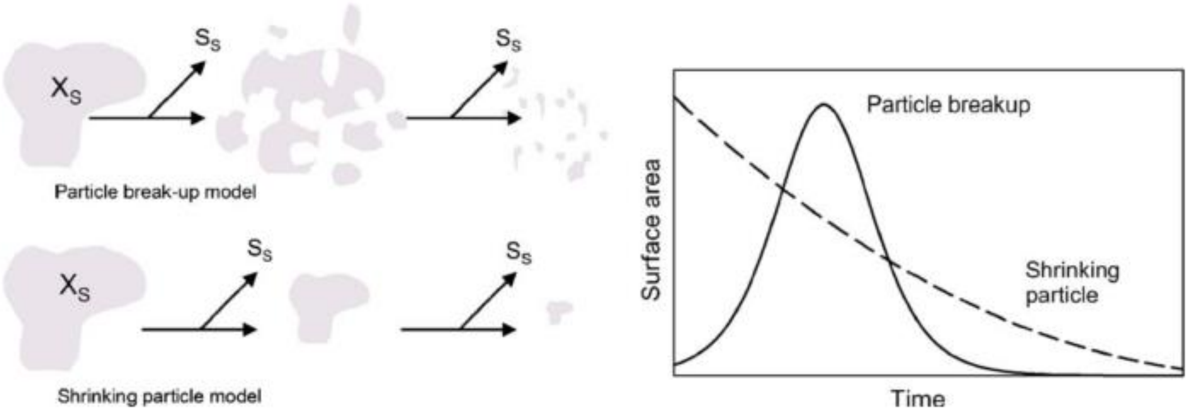

X_S
S_S
S_S
Particle break-up model
X_S
S_S
S_S
Shrinking particle model
Particle breakup
Shrinking particle
Surface area
Time

Figure 51: Conceptual comparison of surface-based kinetics with and without particle breakup *(Dimock and Morgenroth, 2006)*.

When dealing with several hydrolysable fractions, some authors considered parallel degradation (Lagarde et al., 2005; Orhon et al., 1998; Mottet et al., 2013; Garcia-Gen et al., 2015)(Figure 52) and others recognized sequential degradation (Confer and Logan, 1997; Lagarde et al., 2005; Spérandio and Paul., 1999, Yasui et al., 2008)(Figure 53). There is no consensus on the concept involved in the hydrolysis of several fractions, leading to inconsistency.

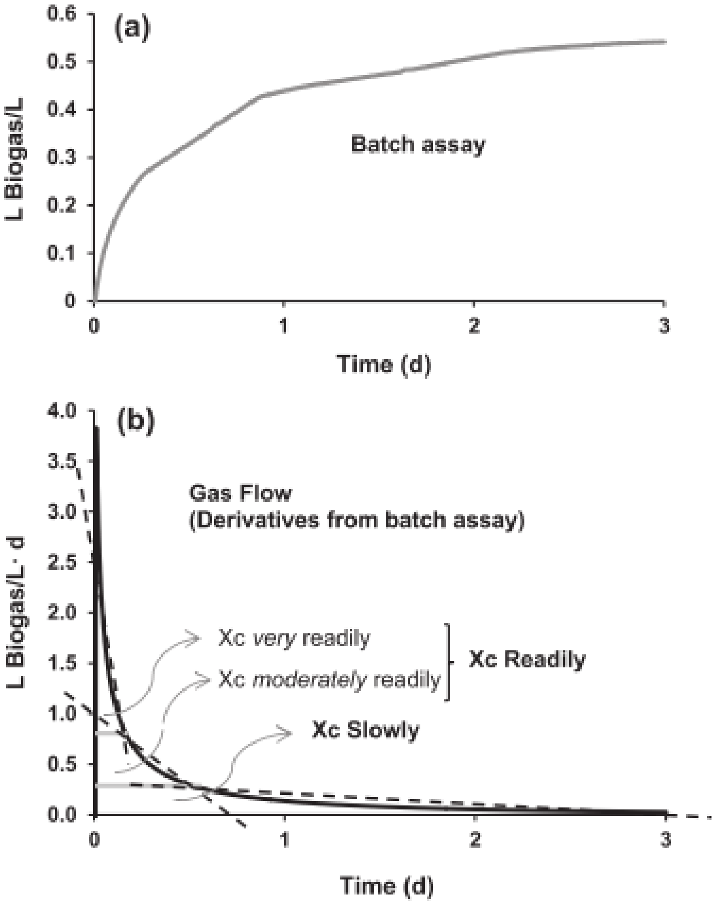


Figure 52: Experimental batch assay of mango. (a) Cumulative biogas production per reactor volume. (b) Gas production rate (computed from cumulative derivative) *(García-Gen et al., 2015b)*.

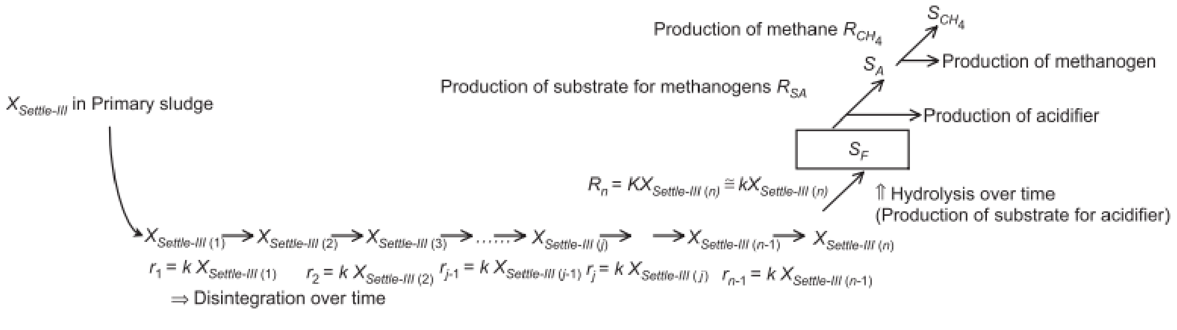

S_{CH_4}
Production of methane R_{CH_4}
S_A
Production of methanogen
Production of substrate for methanogens R_{SA}
Production of acidifier
$X_{Settle\text{-}III}$ in Primary sludge
S_F
$R_n = KX_{Settle\text{-}III\,(n)} \cong kX_{Settle\text{-}III\,(n)}$
⇑ Hydrolysis over time
(Production of substrate for acidifier)
$X_{Settle\text{-}III\,(1)} \rightarrow X_{Settle\text{-}III\,(2)} \rightarrow X_{Settle\text{-}III\,(3)} \rightarrow \ldots \rightarrow X_{Settle\text{-}III\,(j)} \rightarrow \quad \rightarrow X_{Settle\text{-}III\,(n-1)} \rightarrow X_{Settle\text{-}III\,(n)}$
$r_1 = k\,X_{Settle\text{-}III\,(1)}$ $r_2 = k\,X_{Settle\text{-}III\,(2)}$ $r_{j-1} = k\,X_{Settle\text{-}III\,(j-1)}$ $r_j = k\,X_{Settle\text{-}III\,(j)}$ $r_{n-1} = k\,X_{Settle\text{-}III\,(n-1)}$
⇒ Disintegration over time

Figure 53: Structure of simplified particle break-up model. Illustration of one hypothesis leading to sequential degradation based on the particle size *(Yasui et al., 2008)*.

While it is possible to evaluate multiple fractions simultaneously in a single batch test (Wang et al., 2013), this is only effective when the fractions have very different kinetics. Some attempts have been made to describe this by particle breakup models (Dimock et al., 2006; Yasui et al., 2008) based on particle size characterization. Experimental approaches can also be found. Morgenroth et al. (2002) defined four experimental approaches quantifying (1) enzymes, (2) hydrolytic intermediates, (3) bulk parameters, or (4) bacterial kinetics (Morgenroth et al., 2002). According to these authors, the first two approaches allow studying mechanisms involved in hydrolysis but it restricted their work to specific substrates (e.g. starch). The latter two approaches evaluate the overall processes but not the specific mechanisms involved in the hydrolysis process. Modelling oxygen uptake rate or methane production rate (MPR) curves often does not identify underlying kinetics due to a lack of resolution. Therefore, the challenge is to find an experimental way to study the mechanisms of hydrolysis in details.

If we consider the definition of hydrolysis, it refers to the breakdown of organic substrate into smaller products that can subsequently be taken up and biodegraded by bacteria. But it also refers to all mechanisms that make slowly biodegradable substrate available for bacterial growth. In this latter definition, the key word “available” leads considering two majors concepts: bioaccessibility and biodegradability. Hydrolysis is mainly governed by the first two. Due to the complex organization of some organic residues, the concept of bioaccessibility was introduced and defines access to the molecule. It depends on the process duration, the hydrolytic activity or the pre-treatment applied. The biodegradability represents the part of the organic matter consumed by the microorganisms.

While biodegradability tests are well known (Angelidaki and Sanders, 2004), characterization of bioaccessibility remains a challenge. Recently, a new promising methodology of organic matter characterization based on chemical sequential extraction (Figure 54) has been successfully developed to describe the organic matter bioaccessibility of organic residues (Jimenez et al., 2015a). This methodology concept is based the fact that the chemical accessibility simulates the bioaccessibility.

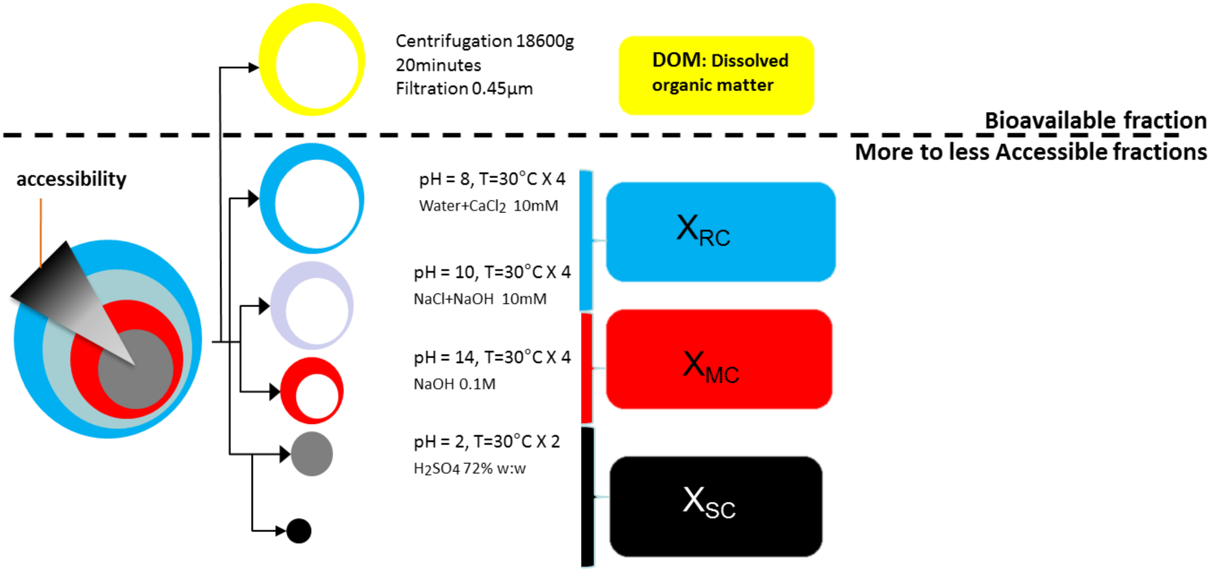

Centrifugation 18600g
20minutes
Filtration 0.45µm
DOM: Dissolved organic matter
Bioavailable fraction
More to less Accessible fractions
accessibility
pH = 8, T=30°C X 4
Water+CaCl2 10mM
pH = 10, T=30°C X 4
NaCl+NaOH 10mM
pH = 14, T=30°C X 4
NaOH 0.1M
pH = 2, T=30°C X 2
H2SO4 72% w:w
X_{RC}
X_{MC}
X_{SC}

Figure 54: Illustration of the chemical sequential extraction (Jimenez et al., 2015a) to estimate the bioaccessibility. X_{RC}, X_{MC} and X_{SC} stand respectively for readily, moderately and slowly accessible organic matter.

This methodology is able to provide new organic matter variables for particulate organic matter. Jimenez et al., (2015a) showed that coupling this approach with 3D fluorescence analysis could be successfully used to implement a modified ADM1 model predicting biogas production performances and digestate quality of an anaerobic digester fed by sludge.

The objective of this chapter is to provide a modelling approach that predicts the methane production performances. First, the classical parallel concept of hydrolysis of ADM1 is challenged to provide a model applicable to different substrates, particularly to solid substrates. The use of a new fractionation methodology, describing bioaccessibility, could indeed help to better understand and describe this phenomenon. The fractionation developed is used as anaerobic digestion model input variables for a plant wide organic waste treatment objective. The simultaneous or sequential hydrolysis of these fractions is tested to properly fit the experimental MPR. In a second objective, this study aims at using the NIR analysis developed in Chapter 5 to reduce the difficulties of ADM1 parameters estimation. The predicted MPR and biochemical composition of the substrates provided by NIRs are used to implement ADM1. This approach reduces the time needed to estimate ADM1 parameters from 30 days to a matter of minutes, making it suitable for industrial application. Finally, the accuracy of this approach has been estimated based on two case studies.

7.4 Materials and Methods Related to this Chapter

Detailed information about the materials and methods related to this chapter are provided in the "Materials and Methods" sections 2.3.2, 2.3.3, 2.4.1, 2.7.1, 2.7.2.

7.5 Modelling Hydrolysis: Simultaneous versus Sequential Biodegradation of the Hydrolysable Fractions

7.5.1 Definition of Simultaneous and Sequential Concepts

First, a definition of both concepts, i.e. simultaneous versus sequential, has to be given. Figure 55 gives a schematic overview of these definitions. In the case of the simultaneous concept, all the fractions X_{RC}, X_{MC}, X_{SC} and X_{NE} are degraded simultaneously. The first degradation period corresponds to the degradation of the sum of all the fractions, it ends when one fraction is completely biodegraded. Then, the second degradation period starts corresponding to the degradation of the sum of all the remaining fractions, it ends when a new fraction is totaly biodegraded, and so on until all the fractions are biodegraded. Consequentely, when the total degradation rate is studied, it is very difficult to distinguish the kinetics of each fraction.

In the case of the sequential concept, the most accessible fraction (i.e. X_{RC}) is degraded first. This fraction acts as a protection layer for the following accessible fraction (i.e. X_{MC}), the latter protecting the following accessible fraction (i.e. X_{SC}) itself protecting the least accessible fraction (i.e X_{NE}). Consequentely, during the first period of degradation, X_{RC} is the only fraction consumed, before the degradation of X_{MC} , then of X_{SC} and finally of X_{NE}. In this concept, each kinetics is independant. Thus it is possible to have the fraction X_{SC} faster than X_{MC} , which is impossible using the simultaneous concept as it is demonstrated on Figure 55.

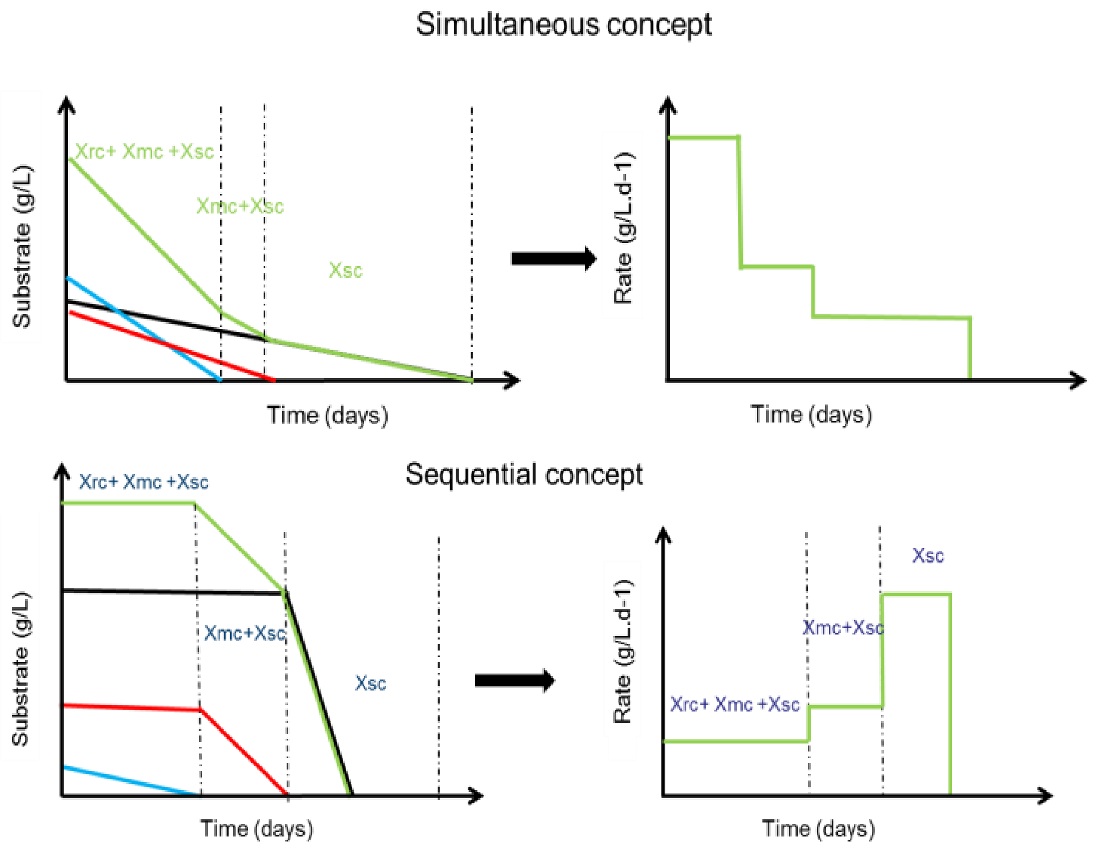

Simultaneous concept
Substrate (g/L)
Xrc+ Xmc +Xsc
Xmc+Xsc
Xsc
Time (days)
Rate (g/L.d-1)
Time (days)
Sequential concept
Substrate (g/L)
Xrc+ Xmc +Xsc
Xmc+Xsc
Xsc
Time (days)
Rate (g/L.d-1)
Xsc
Xmc+Xsc
Xrc+ Xmc +Xsc
Time (days)

Figure 55: Schematic definition of simultaneous and sequential concepts. Xrc, Xmc and Xsc stands respectively for the readily biodegradable, the moderately biodegradable and the slowly biodegradable fractions. To simplify the figure, Xne is not represented.

7.5.2 Results Obtained on Several Substrates

Several substrates were tested on successive batches until the MPR has converged to similar curves to remove the microbial innoculum effect (see “Materials and Methods” section 2.3.3.1). Experimental biogas production resulting from these batch tests are summarised in Figure 56. Simulations (using ADM1 described in Materials and Methods section 2.7.1) obtained with the simultaneous (i.e. the switching function is always equal to 1 in the model) and sequential models (i.e. switching function varies) are also presented in Figure 56. The initial condition effect on simulation was removed running simulation of successive batches using the same approach as on experimental batches. The details of the parameters adjusted for each substrate are reported on Table 16. The parameters related to the biochemical composition were estimated by the laboratory analyses described in the “Materials and Methods” section 2.3.1, the COD per fraction was estimated by chemical sequential extraction (“Materials and Methods” section 2.3.2). The value of the switching function and the kinetic parameters were adjusted by trail and error to fit the methane produciton curve.

Table 16: Parameters of simulation of the batches using simultaneous model or sequential model

			Carrot	Cauliflower	Lettuce	Wheat straw	Potatoes
COD per fraction ($kgDCO.m^{-3}$)		X_{RC}	0.59	0.47	0.63	0.51	0.33
		X_{MC}	0.21	0.21	0.08	0.56	0.11
		X_{SC}	0.25	0.33	0.11	1.26	0.90
		X_{NE}	0.16	0.32	0.40	1.61	0.00
Biochemical composition (%)		f_X_{RC}_si	0.00	0.00	0.00	0.00	0.00
		f_X_{RC}_xi	0.00	0.20	0.25	0.51	0.00
		f_X_{RC}_ch	0.73	0.55	0.37	0.02	0.77
		f_X_{RC}_pr	0.05	0.19	0.09	0.04	0.23
		f_X_{RC}_li	0.22	0.01	0.18	0.43	0.00
		f_X_{MC}_si	0.00	0.00	0.00	0.00	0.00
		f_X_{MC}_xi	0.22	0.70	0.66	0.80	0.19
		f_X_{MC}_ch	0.31	0.16	0.03	0.06	0.19
		f_X_{MC}_pr	0.05	0.13	0.03	0.02	0.00
		f_X_{MC}_li	0.43	0.01	0.28	0.12	0.62
		f_X_{SC}_si	0.00	0.00	0.00	0.00	0.00
		f_X_{SC}_xi	0.43	0.19	0.42	0.47	0.03
		f_X_{SC}_ch	0.33	0.52	0.17	0.40	0.79
		f_X_{SC}_pr	0.01	0.14	0.06	0.01	0.01
		f_X_{SC}_li	0.23	0.16	0.35	0.13	0.17
Sequential	Switch (no unit)	K_{IS}	0.01	0.005	0.005	0.005	0.005
	Kinetics (d^{-1})	K_{DIS}_X_{RC}	3.50	4.00	4.00	10.00	5.00
		K_{DIS}_X_{MC}	2.50	2.00	1.00	3.50	1.00
		K_{DIS}_X_{SC}	2.50	1.00	9.00	3.50	1.50
		K_{DIS}_X_{NE}	0.50	0.50	7.00	0.80	0.50
Simultaneous kinetics (d^{-1})		K_{DIS}_X_{RC}	2.00	4.00	4.00	5.00	2.00
		K_{DIS}_X_{MC}	1.67	2.00	2.00	0.50	1.00
		K_{DIS}_X_{SC}	1.67	1.00	1.00	0.80	0.75
		K_{DIS}_X_{NE}	0.33	0.50	1.00	0.50	0.25

When a desceding order of kinetics correspond to a descending order of bioaccessibility, the modifications of the kinetics with the sequential concept had no impact, as illustrated by the results for carrots (Figure 56a). However, others were better modelled with the sequential model than with the simultaneous one. This is the case for cauliflower, lettuce and wheat straw (respectively Figure 56b, Figure 56c, Figure 56d). In these cases, it can be observed that kinetics of methane production of intermediate fractions is slower than the one of the last fraction, which is impossible using simultaneous concept. Consequently, the use of the sequential concept for all substrates is a solution in order to reach a good fit of all MPR curves for all substrates, above all, when a desceding order of kinetics is not associated with a descending order of bioaccessibility. However, neither the sequential model nor the simultaneous model fitted experimental data of the potato biodegradation (Figure 56e).

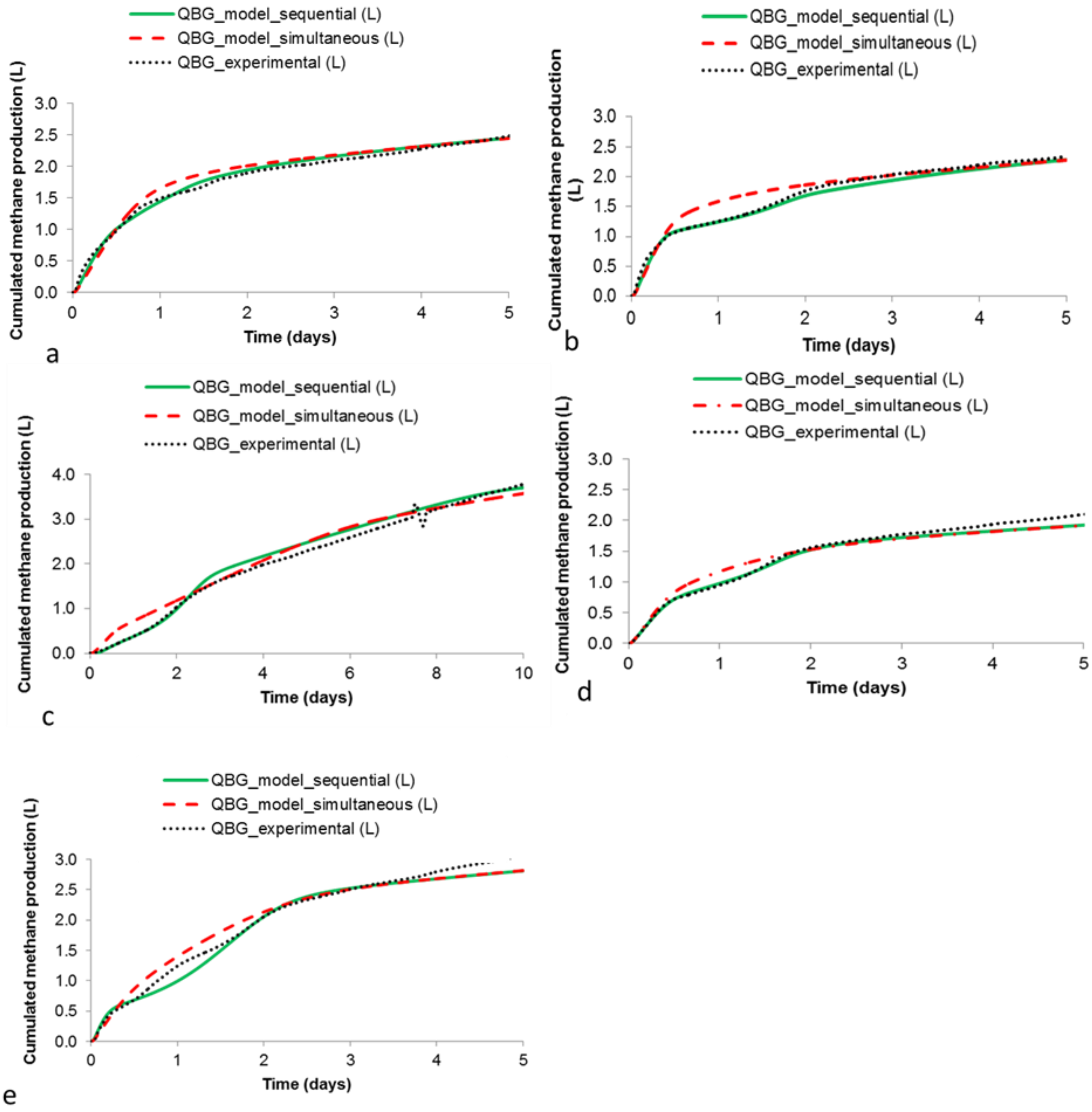


Figure 56: Methane Production Rate curves (QBG) obtained experimentally and by the simulations with simultaneous model and with the sequential model (a: carrot, b: cauliflower, d: lettuce, e: wheat straw, e: potatoes).

In this case, the hypothesis proposed to explain the observation is that the number of biodegradable fractions (4 fractions in these models) is insufficient.

7.5.3 Fractions Biodegradability: Case Studies on Apple and Wheat Straw

Using chemical sequential extraction, at each extraction, a liquid fraction is extracted (REOM corresponding to X_{RC}, MEOM to X_{MC} or SEOM to X_{SC}) and a pellet remains, containing the previous fractions minus the extracted fraction. X_{NE} corresponds to the solid non extractible fraction NEOM which remains after all the extractions. Thus, after chemical sequential extraction, 3 pellets are created containing, $X_{MC}/X_{SC}/X_{NE}$, X_{SC}/X_{NE} and X_{NE} (extraction protocole is detailled in “Materials and Methods” section 2.3.2). X_{NE} was not used for the experiments because the extraction of X_{SC} using sulfuric acid made the pellet containing X_{NE} not suitable for AD experiments. The others remaining pellets, $X_{MC}/X_{SC}/X_{NE}$ and X_{SC}/X_{NE} from an apple and a wheat straw (model substrate for lignocellulosic waste) were digested according to BMP tests protocole (Materials and Methods section 2.3.3.2). For each substrate, three cumulated biogas production curves were obtained from the digestion of the total substrate and the pellets: total substrate (X_{RC} /$X_{MC}/X_{SC}/X_{NE}$), total substrate after REOM extractions ($X_{MC}/X_{SC}/X_{NE}$), and total substrate after REOM and MEOM extractions (X_{SC}/X_{NE}). Results obtained are presented in Figure 57a and Figure 57b.

In the case of the apple, as for carrot, the kinetics of biodegradability decreased with bio-accessibility. The kinetics of each fraction could be obtained by subtraction and the simultaneous concept can be applied.

In the case of wheat straw, the X_{RC} fraction was not very important which explains the similarity that is found between biodegradability curves of total substrate and of total substrate minus X_{RC}. However, when the BMP test was performed on the total substrate after REOM and MEOM extractions (X_{SC} /X_{NE} fractions only), the rate increased, compared with the total substrate after REOM extractions ($X_{MC}/X_{SC}/X_{NE}$).

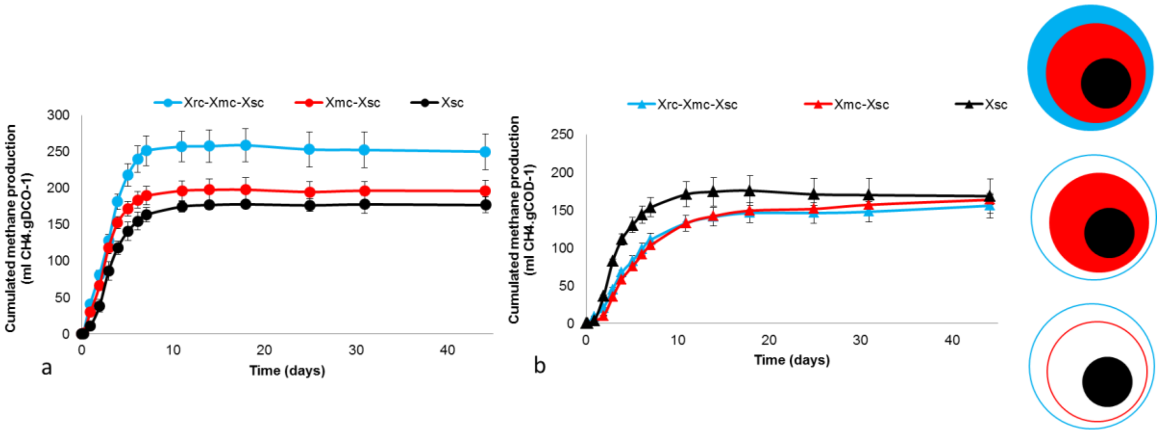


Figure 57: Anaerobic biodegradation of the apple (a) and the wheat straw (b) fractions

The biodegradation kinetics associated the fractions X_{RC}, X_{MC} and X_{SC} are assessed from these experiments. Supposing a simultaneous model, their are estimted by subtraction. For instance, the MPR associated to the fraction X_{RC} was recreated substracted the MPR of the pellet $X_{MC}/X_{SC}/X_{NE}$ to the MPR of the total substrate ponderated by the importance of $X_{MC}/X_{SC}/X_{NE}$ into the total substrate. Figure 58 shows the results of this approach on wheat straw experiments. It can be observed that the kinetics obtained for the X_{MC} fraction is negative. This result is due to the fact that all hydrolysable fractions are considered to be hydrolysed in a simultaneous way. Another scenario, corresponding to sequential hydrolisis, is to consider that fraction *n* is not hydrolysed until fraction *n-1* reaches a sufficiently low concentration. That corresponds to the second scenario of Figure 58. In this way, the proposed switching function has to be used in the hydrolysis modelling of each fraction.

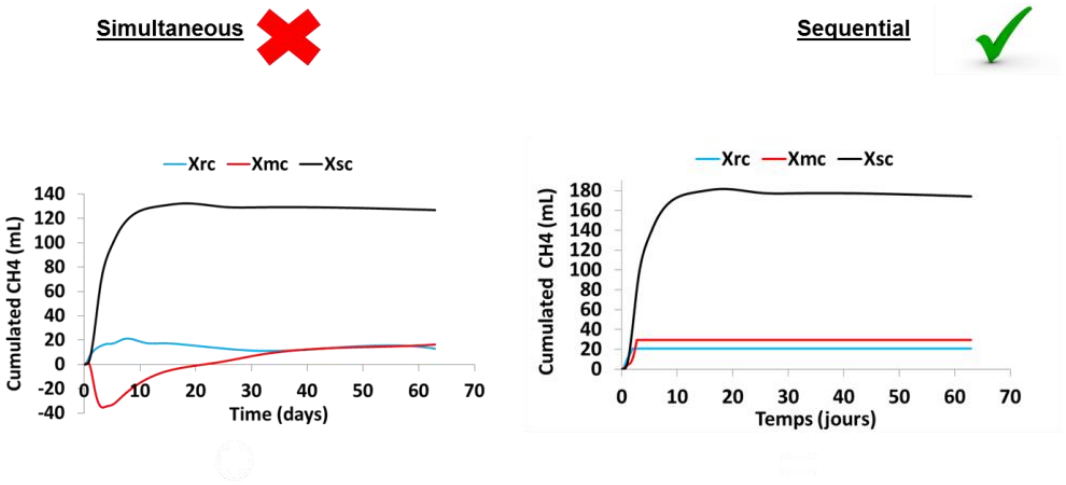


Figure 58: Conceptual scenarios proposed to explain accessible fractions hydrolysis kinetics and fed-batch digestion of wheat

The extraction step to remove the MEOM fraction acts like an alkaline pre-treatment, allowing the partial solubilization recalcitrants in the wheat straw. Indeed, vegetables are generally composed of a recalcitrant wall which protects the plant. In the wheat straw case, a wax layer protects another layer containing cellulose and pectins (pectin is water soluble). MEOM extraction removed the wax layer and allows a quicker biodegradation of the X_{SC} and X_{NE} fractions (i.e. hemicellulose and cellulose).

This means that the non-accessible feature of X_{SC} and X_{NE} limits hydrolysis despite the high biodegradable potential of X_{SC}. Similar results were obtained by (Rincker et al., 2013) after pre-treatments applied on lignocellulose-like substrates. According to the authors, the lag phase could correspond to a colonization process. This colonization phase was also observed in the case of cellulosic fibers with low lignin content (toilet paper) found in primary sludge (Ginestet et al., 2002). Yasui et al. (2008) also observed such phenomenon between 2 fractions of organic matter from several primary sludge. As no inhibition was noticed, the authors proposed a sequential model of hydrolysis based on the particle size to explain this observation.

7.5.4 Case of Continuous Reactor

Up to now, only batch data were used in order to calibrate the models. It has been shown that sequential approach using a switching function between the fractions allows a better fit on the experimental data. If we consider continuous reactor, sequential approach in batch can be interpreted by two different models:

A. The hydrolysis mechanisms tend to a simultaneous biodegradation approach, after feeding substrates to the bioreactor, as illustrated in Figure 59. It induces, as for simultaneous model, to set the switching function to 1, but keeping the same kinetic parameters as fited in batch, which simulate usually a faster hydrolisis than using the parameter fitted using a simultaneous model
B. The value of the switching function remains the same, which mechanistically correspond to an inhibition of the most bioaccessible fractions on the other ones. Thus in case of a continous reactor, the hydrolisis of the poorly accessible fraction such as X_{SC} and X_{NE} is drastically slown down, and they accumulate in the reactor.

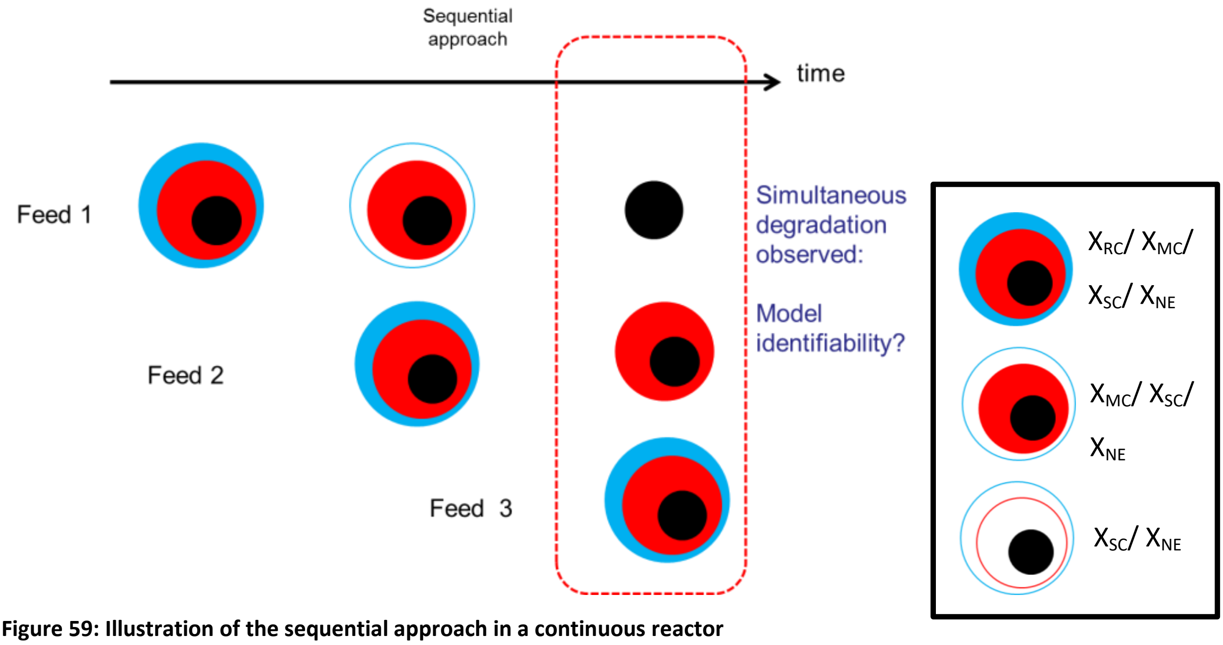

Sequential approach
time
Feed 1
Feed 2
Feed 3
Simultaneous degradation observed:
Model identifiability?
X_{RC}/ X_{MC}/ X_{SC}/ X_{NE}
X_{MC}/ X_{SC}/ X_{NE}
X_{SC}/ X_{NE}

Figure 59: Illustration of the sequential approach in a continuous reactor

To illustrate the differences between these models, a 6m^3 digestor fed with a constant inflow of 0.05 m^3.d^{-1} of wheat straw at a concentration of 157 kgCOD.m^{-3} was simulated. The same simulation parameters as described in Table 16 were used for biochemical composition, kinetic parameters, switching function and COD repartition per fraction. The results of the simulation are shown in Figure

60 for methane flow rate (top left), total VFA concentration (top right), not digested COD of the input (bottom right) and COD of the fraction in case of sequential model B (bottom left). The hypothesis A of sequential hydrolysis, based on bio-accessibility limitation, is represented by the *“sequential access”* curves while the hypothesis B is represented by the *“sequential inhibition”* curves. It can be seen that all the models give very different results. In case of sequential approach with the hypothesis B, only the fraction X_{RC} and X_{MC} are degraded; and X_{SC} and X_{NE} accumulate in the digestor (bottom right of Figure 60). The simultaneous approach and the sequential approach with the hypothesis A were more similar. Nonetheless, the hydrolisis of the fractions using the sequential approach (hypothesis A) is faster, due to higher μ_{max} fited on the batch data (see Table 16), which induce first a more important methane production but then an earlier acidification of the reactor.

The simulation results of Figure 60 help us to understand the differences of the sequential and simultaneous approaches. Differentiating the two different sequential models is feasible, even if experimental data are not yet available, it seems that hypothesis B is not realistic. However, identify the sequential model based on bioaccessibility limitation (hypothesis A) or simultaneous model is really difficult since it differs only by the μ_{max} of the hydrolisis function. Thus it could be tricky to validate if the differences of kinetics observed between batch or continuous experiments come either from a sequential process or just a biomass gowth and adaptation.

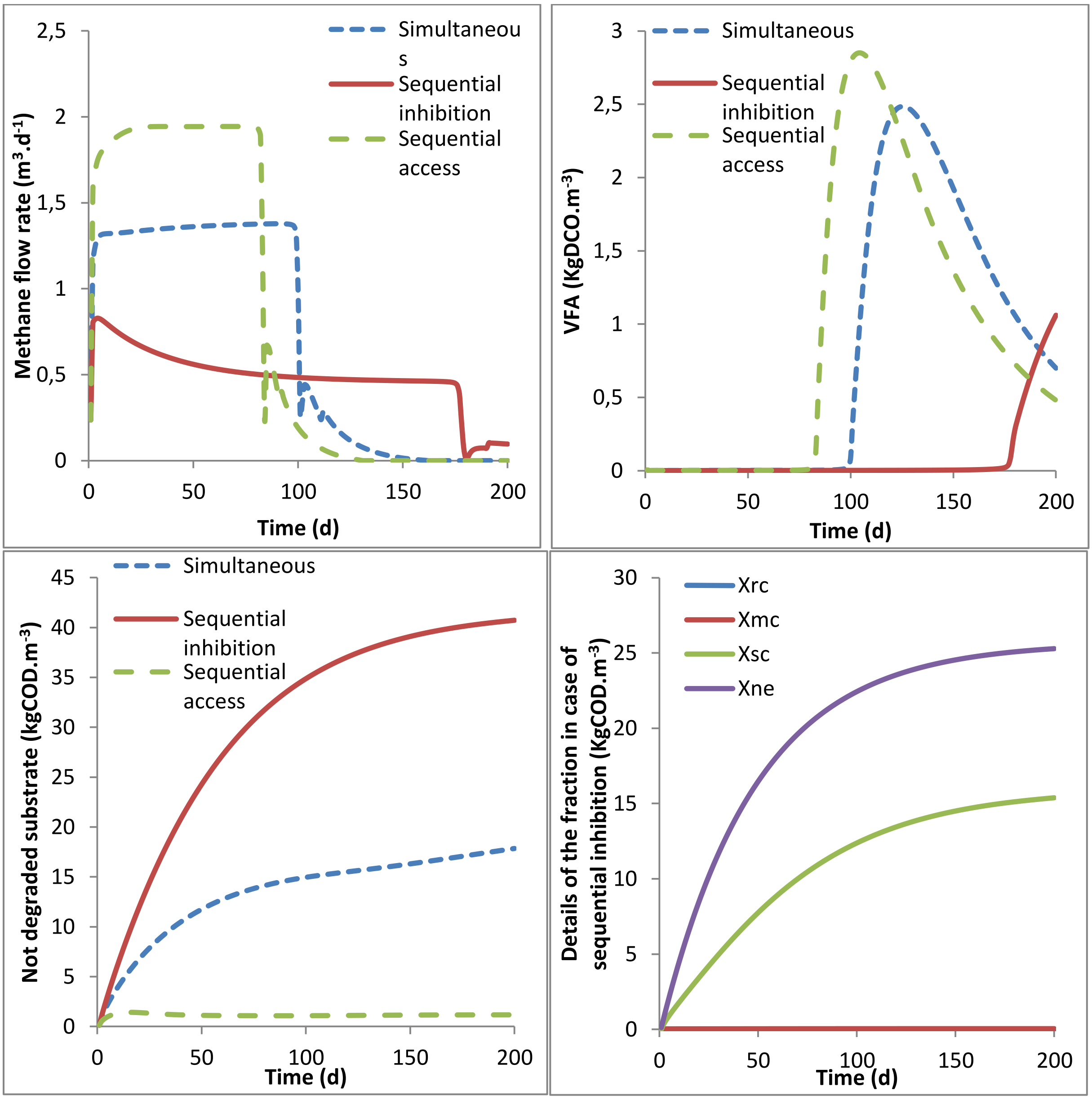


Figure 60: Continuous reactor simulation using simultaneous model or sequential model based either on bio-accessibility limitation (access) or inhibition.

Finally, the question of how these mechanisms (successive degradation, colonization) should be taken into account when modelling continuous systems still needs to be debated. Solid particles have a distribution of retention times in such systems, with different histories and degrees of degradation. Population balance models seem appropriate to describe this heterogeneity in hydrolysis (Lebaza et al., in press) but simplified approaches would probably also be suitable.

7.5.5 Conclusion on Hydrolysis Modelling

The objective was to discuss about hydrolysis modelling concepts, i.e. simultaneous versus sequential approaches. To illustrate the discussion and challenge the simultaneous approach, some experimental case studies were modelled to test both concepts. Also, a new fractionation methodology was introduced in order to study the hydrolysis mechanisms in detail. This methodology, based on accessibility characterization, was successfully used to replace the ADM1 variables describing particulate organic matter. The methodology also allowed testing simultaneous versus sequential hydrolysis. The sequential approach fitted all substrates biodegradation while the simultaneous approach was sometimes not applicable (like for wheat straw or lettuce biodegradation, Figure 57). While the ADM1 model with sequential hydrolysis allows effective fit to data, it only represents a hypothesis. Further proof needs to be obtainedand an experimental set-up should be designed, with a specific accessibility fractions study.

Nethertheless, sequential hydrolysis model was considered to be more flexible and adapted for batch experiment modelling and was thus prefered to model the MPR estimated by NIR. Sequential chemical extraction gives an accurate definition of the fraction based on bio-accessibility but is time-consumming. Since every fraction has its own kinetics, it is possible to roughly estimate the fraction based on the MPR. Thus, in order to increase the applicability of the modified ADM1 for plant performances prediction, the fraction definition made by chemical sequential extraction was replaced by a slope analysis according to the methodology detailled in “Materials and Methods” section 2.7.2. Fast implementation using NIR estimation was applied on two case studies. The results are discussed on the second part of this chapter.

7.6 ADM1 Implementation Assisted by Near Infrared Spectroscopy for the Anaerobic Digestion Performances Simulation

7.6.1 ADM1 Calibration from NIR Prediction

In Chapter 5, it has been proven that NIR spectroscopy can be used to estimate the carbohydrate, protein and lipid contents, COD, methane yield and MPR in batch conditions. It turns out to be also a powerful method to implement an ADM1 model able to predict the methane production performances of a substrate. The approach was applied to the case studies of salad and cauliflower, illustrating co-substrates used in farm plants. The curve predicted by NIR was used as a reference for the μ_{max} optimization of the kinetic parameters according to the “Materials and Methods” section 2.7.2.4.

The ADM1 model using only one particulate input as firstly described by Batstone et al., (2002) was tested. Respectively for salad (Figure 61.A) and cauliflower (Figure 61.B), μ_{max} were estimated at 1.92 and 4.33 d^{-1}. As envisaged previously in the first part of this chapter (7.5), it can be seen that, after optimizing μ_{max}, the model does not represent accurately the predicted MPR. Indeed, it tends to under-estimate the kinetics at the beginning of the production and to over-estimate it in the second part. Two different approaches can be used to fit the data, first, the other kinetic parameters of the ADM1 model can be modified, second more kinetic parameters describing the hydrolysis process can be set to represent efficiently the curve.

Since the first part of this chapter shows that a sequential hydrolysis model with four different fractions is more accurate and flexible to represent batch digestion, it was decided to focus on the hydrolysis modeling. A second optimization was done using the modified ADM1 described in "Materials and Methods" section 2.7.1, with four fractions, four distinct μ_{max} kinetic parameters and a sequential degradation. It can be seen that the use this model improved the fit between the modeled cumulated biogas production and the predicted cumulated biogas production. The degradation of X_{RC} is faster than when considering a single input while the degradation of X_{MC}, X_{SC} and X_{NE} are slower, which fits the predicted data. It can be seen in Figure 61.A-B that the NIR predicted data allows calibrating a modified ADM1 that accurately represents the experimental batch data.

Since using NIR, as demonstrated in section 5.8, it is possible to predict the confidence interval of the prediction, it is also possible to predict a confidence range of simulation. If we consider a normal distribution of the error and a sample nearby the calibration set, 68.2 % of the reference values are contained in a range of ± 1 SEP of the predicted value. Thus, as detailed in Materials and Methods section 2.7.2, an error of ± 1 SEP was applied to each variables predicted by NIR (i.e each methane production time, methane yield, COD, carbohydrate, protein and lipid content) in order to obtain an envelope where 68.2 % of the possible scenario of methane production are included. Since COD, carbohydrate, protein and lipid contents affect parameters which were not recorded such as inert content, long chain fatty acids or ammoniacal nitrogen, the predicted confidence range on this data cannot be experimentally validated. However, applying an error of ± 1 SEP to each methane production time and to the methane yield provide an envelope where 68.2 % of the possible MPR curves are included. μ_{max} of the Contois kinetic parameters were optimized to fit the most optimistic MPR, minus one SEP on methane production time and plus one SEP on methane yield, and to fit the most pessimistic MPR, plus one SEP on methane production time and minus one SEP on methane yield by trail and error. The confidence range of the simulations is presented in Figure 61.C-D. It can be seen that the experimental MPR is included in the simulated range of confidence of the MPR. It shows that the predicted confidence range represents accurately the experimental MPR. This approach is interesting in order to predict the performance of the plant depending on the feeding strategy. The introduction of potential errors of prediction in the model also provided an estimation of the risk that the plant operator is taking (Južnič-Zonta et al., 2012). Thus it could provide key information to the human operator to optimize safely the production of a plant.

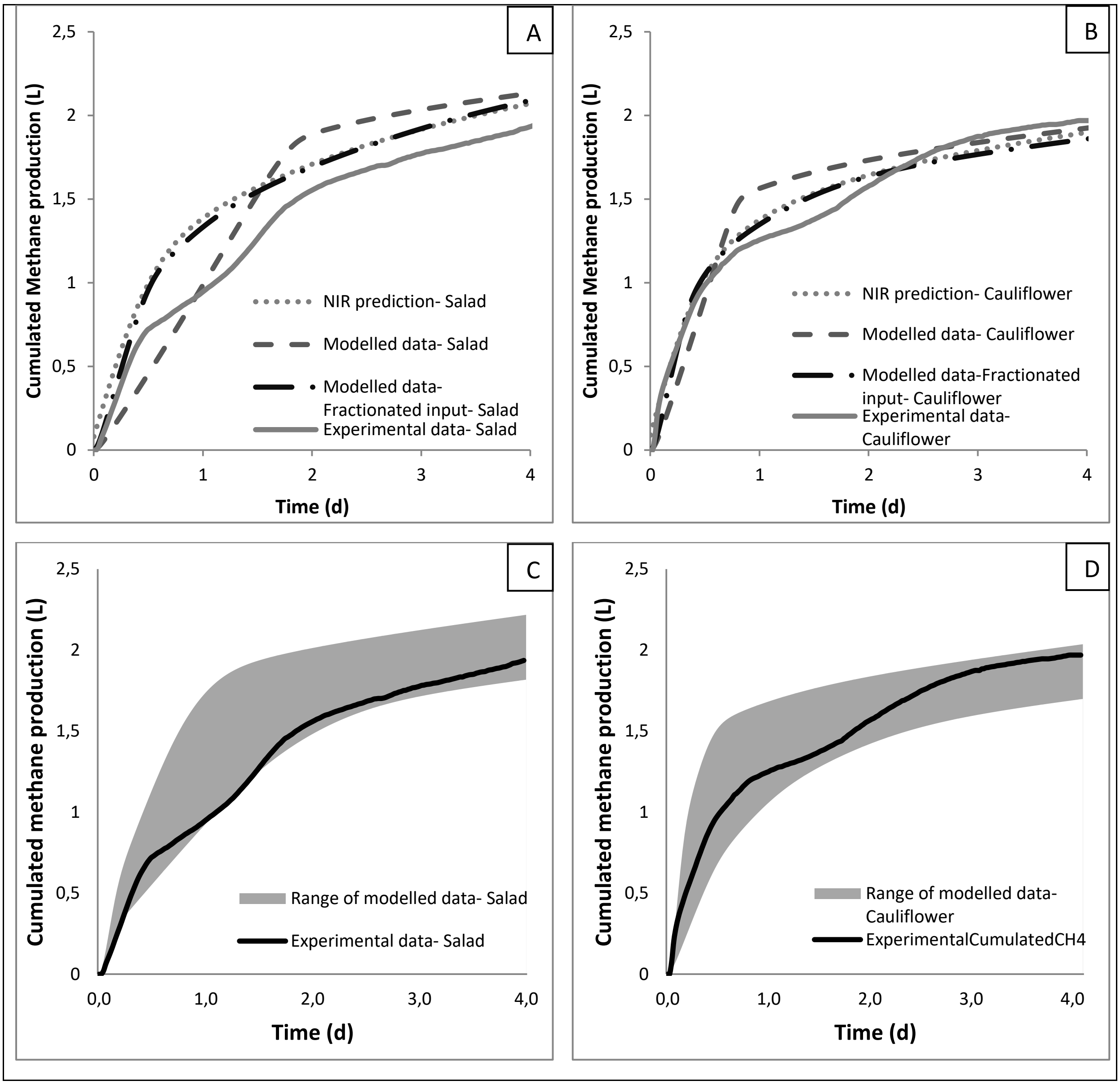


Figure 61: Cumulated methane production curve for anaerobic of salad (left) and cauliflower (right) in batch. A and B, comparison of predicted and modelled data using ADM1 considering the particulate organic matter as a single input or fractioning the particulate organic matter in four inputs with their own kinetics of disintegration for salad and cauliflower respectively. B. C and D, modelling of the methane cumulated production for salad and cauliflower AD taking into account the error of prediction for salad and cauliflower respectively. Methane production times and methane yield were chosen at ± 1 SEP of the predicted value. The grey aera represents the range of confidence.

7.6.2 Discussion on the MPR Prediction Using NIR

Batstone et al., (2009) have demonstrated that hydrolysis kinetic parameters estimated in BMP test translated to very poor model performance when BMP-estimated values were used in the continuous model. On the other hand, García-Gen et al., (2015) demonstrated that the calibrated ADM1 kinetic parameters from batch assays that allow micro-organisms to adapt to individual substrates can be used to model semi-continuous co-digestion operation treating simultaneously 5 waste. In our case, we used the same experimental protocol as García-Gen et al., (2015) and thus the model parameters can be assumed to be valid for continuous operation of the digester. The main drawback of the batch experiments with adapted micro-organisms is the time needed, from 6 to 12 weeks, which reduces its use. This study also shows that NIR spectrometry can be efficiently used to predict experimental batch MPR on adapted micro-organisms. Kinetic parameters of the ADM1 model calibrated on the predicted batch MPR can be used for continuous co-digestion simulation. This approach gives the opportunity to replace time-consuming biological experiments by fast NIR analysis for feeding strategy optimization.

7.7 Example of Industrial Application

The approach NIRs plus predictive model is tested to monitor the anaerobic digester of a STEP unit. This unit is equipped with a supervision system MeMo® (BioEnTech, 74 avenue Paul Sabatier, 11100, Narbonne, France. Website: http://bioentech.eu) which calibrates ADM1 in real time. Thanks to lab analyses and on-line sensors data, ADM1 is calibrated on the calibration period and validated on the current period. In Figure 62, an example of the benefit of the NIR characterization is shown. NIRs has been used to analyze the substrates digested in the unit. From the information collected by MeMo and NIRs, simulation using ADM1 are run to predict the plant performances. A simulation of the addition of 8000 $kgO_2.d^{-1}$ of a new substrate characterized by NIRs is run using the calibrated ADM1. Based on this simulation, variables such as biogas production and composition or COD removal efficiency are estimated. More complex indicators such as diagnostic of the plant (indicating risk of overload, toxicity and washout) are also predicted using the simulated data. It aims to assist the operator and help him taking the right decision.

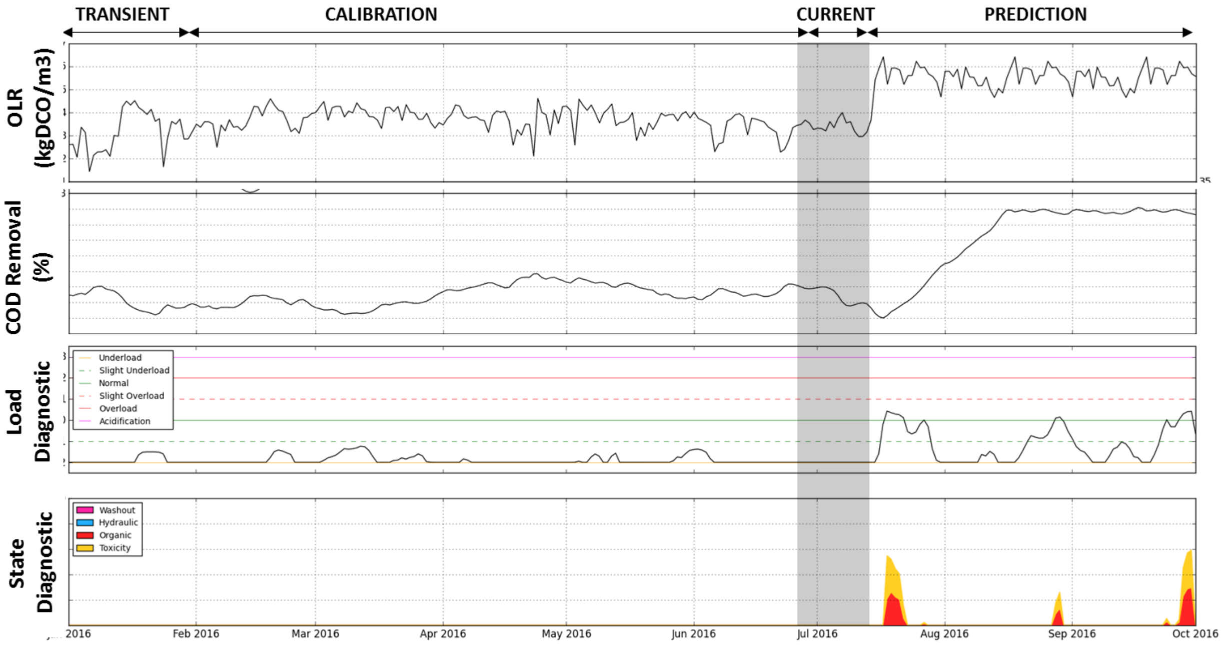


Figure 62: Plant performances prediction using ADM1 plus NIRs

This simulation is a case study that shows the potentiality of the approach. The next step will be to validate the prediction results comparing them with real performances.

7.8 Conclusion

The current study shows that:

- Even if the biological mechanism of hyrolysis are not fully understood, the sequential hydrolysis model fits all substrates biodegradation while the simultaneous hydrolisis model is sometimes not applicable.
- NIRs enables a fast estimation of the parameters describing the substrate needed to implement a modified ADM1. Thanks to this fast parameters estimation, it becomes possible to predict the optimal performances of degradation of any input.
- This study demonstrates that the use of NIRs coupled with PLSr models and ADM1 modeling can help the human operators to monitor the plant. This approach can be used as a powerful tool for plant monitoring and feeding strategy optimization.

Chapter 8. General Conclusion and Outlooks

8.1 Conclusion

Anaerobic digestion development has to go through the development of co-digestion and diversification of the substrates used. The quantity of organic waste available as AD resources gives bright prospects for the sector but its novelty and the lack of industrial solutions adapted to the diversity of organic waste penalizes its profitability and development. Co-digestion plant conception needs to be more cost-efficient in term of investment and operational expenditures to improve the plant profitability and give the right signals to financiers. Moreover, AD plants suffer from under-optimized performances related to poor feeding strategy optimization. To answer this issue, sensors, methods and models needed to optimize the feeding strategy have to answer three questions:

1. What is the current biological state of the reactor?
2. What is the optimal feeding strategy and which performances can be expected from it under optimal conditions?
3. From the biological state and the feeding strategy, what are the operational conditions allowing to reach the optimal production and how to prevent drifting?

The state of art on anaerobic digestion monitoring reveals that many sensors, methods and models have been developed over years to optimize plant performances but many of the solutions developed are not fully adapted to co-digestion. In 2002, the development of the ADM1 was a major step forward AD optimization. It is an accurate model for plant performances modeling which takes into accounts both biological state and substrates performances. Yet it suffers from the many parameters that have to be calibrated for an accurate use and its simplistic hydrolysis process modeling. Thus estimation of plant biological state and substrates features are critical factors of plant monitoring. However, sensors, substrates characterization and process monitoring must be adapted to monitor co-digestion. They need to take into account the complexity of the AD biological reactions and the diversity of feeding strategy in order to secure the plant feeding and to optimize its global performances. While carbon dioxide and methane production analyses are matured, the available solutions estimating the soluble process-intermediates (i.e. volatile fatty acids, inorganic carbon and ammoniacal nitrogen concentration) are not suitable for co-digestion of a wide diversity of substrates with frequent changes in the feeding. The lack of sensors estimating the soluble process-intermediates restricts the biological state estimation which is a first break of feeding strategy optimization. A second break is the difficulties of substrates characterization. Analyses of organic waste AD performances are labor-intensive and time-consuming which limits their use. The poor waste characterization goes with high risk of inhibition in case of changes in the feeding strategy. It restricts the flexibility on the inputs and results in feeding optimization difficulties. The low flexibility on the feeding strategy is a growing concern for AD plants since stress on resources will increase with the number of biogas plants.

With the current context of AD in mind, this PhD developed three axes (Figure 63):

- Accurate estimation of process-intermediates (i.e. volatile fatty acids, inorganic carbon and ammoniacal nitrogen concentration) in the plant to assess its biological state
- Solid organic waste characterization with the development of fast, repeatable and accurate analyses for the estimation biochemical composition and AD performances in a wide range of solid samples such as fruits, vegetables, manure, cereals, oils, fats, meat and fishes.
- Dynamic modeling of the AD process challenging the classical concept hydrolysis and facilitating the implementation of ADM1.

On the one hand, our investigation on process-intermediates estimation reveals that current sensors using titration methods for VFA and IC estimation based on the analysis of the titrant volume on key points provide non-accurate absolute estimations. In case of limited changes in the feeding strategy, constant bias in the estimation is present. On the other hand, it can be worth it to follow AD plants since they provide an correct estimation of the dynamics of VFA, IC and TAN concentrations, but recalibrating sensors is required at any change in the feeding recipe. Therefore this type of sensors is not adapted to monitor co-digestion. A more promising solution is the non-linear estimation of VFA, IC and TAN based on buffer-capacity analysis on the whole titration process. Non-linear analyses of the buffer capacity are more accurate than titrant volume analyses on key-points and less sensitive to the composition of the sample. Thus, these analyses are more suitable to monitor co-digestion. Nonetheless, buffer-capacity, based on pH and titrant volume rapidly reaches a limit to distinguish between the numerous soluble compounds found on AD plants, which may lead to incorrect estimations. The combination of electrical conductivity and pH sensors facilitate the discrimination of the different compounds. Thus, it has improved the accuracy of VFA, IC and TAN concentration estimation using non-linear methods. This work showed that accurate estimation of soluble process-intermediates such as VFA, IC and TAN are possible using titration methods, combining electrical conductivity and pH sensors, based on non-linear methods. It led to the development of the System of titration for total ammoniacal Nitrogen, volatile fatty Acids and Inorganic carbon (SNAC). Combined with gas flow and composition analysis, SNAC is aimed to provide a cost-efficient sensor for the biological state estimation.

Solution for solid organic waste characterization has been investigated to predict AD performances. Current COD and carbohydrate content measurement methods are not adapted to solid waste characterization and require an acid hydrolysis of the organic matter before using the usual methods. Relations between biochemical composition and AD performances have then been studied. Correlations between methane production time, methane yield, COD, biodegradability, carbohydrate, protein and lipid contents using PCA and Pearson correlation coefficients analysis reveals that a relation between the organic matter composition and the AD performances existed. Carbohydrates are mainly readily biodegradable while lipids have a high methane yields but a slow methane production kinetics. Due to its simplicity, COD analysis often replace BMP test, but a negative correlation between COD and biodegradability indicates that COD cannot replace methane yield analysis. The highlighted correlations between carbohydrate, protein and lipid contents, COD, methane yield, biodegradability and kinetics of methane production leads to the idea that kinetics of methane production, which is tedious to estimate, could be predicted from the biochemical composition. However, a PLS regression demonstrates that these indicators provide a vague description of the organic matter which are not sufficient to accurately predict methane production kinetics. Nevertheless, we could observe a relation between biochemical composition and AD process performances which points towards more descriptive organic matter characterizations that could be used for AD process performances prediction.

NIR and 3D fluorescence spectroscopy coupled with chemical sequential extraction have been previously reported as two fast methods that provide a sufficient description of the organic matter to predict its biodegradability. Their advantages and drawbacks for COD and methane yield prediction were assessed on a wide variety of samples. 3D fluorescence coupled with chemical sequential extraction has the advantage to predict the bio-accessibility of the organic matter but NIRs is more accurate and time-efficient for the prediction of biodegradability. Hence NIRs has to be preferred for substrate characterization. Both biochemical analyses and batch experiments have been employed to create a dataset for NIR prediction of carbohydrate, nitrogen and lipid content, COD, methane yield and methane production times and PLSr models have been established. The PLSr predictive models demonstrate an excellent accuracy of prediction for both cross validation and validation with an independent data set. As suspected from the correlation found between the biochemical composition and the AD performances of a substrate, NIRs give a detailed overview of the organic matter which enables a complete, robust and fast estimation of both biochemical composition and AD process performances. A drawback of NIRs coupled with PLSr model is the accuracy of the model which depends on the distance from the analysed sample to the calibration set. The distance is tricky to estimate for the operator. To ensure the industrial applicability of the NIRs analyses, indicators of the confidence interval of prediction have been developed based on the minimal Euclidian distance between each sample and the calibration set. Thanks to the PLSr models developed, NIRs is foreseen as a new and efficient way to characterize solid waste contents for large-scale applications.

Carbohydrate, protein and lipid content, methane yield, COD or methane production time are useful indicators for feeding strategy optimization but are poorly informative on the link between the molecular composition of organic matter and AD performances. Since NIR spectra of pure compounds are available in the literature, the analysis of the *b* coefficients of the PLSr models considered as the Net Analyte Signal (NAS) of these indicators is a unique opportunity to deeply investigate the relations between the organic matter composition and these indicators. The wavelengths contribution assessed through the simultaneous analysis of the *b* coefficients associated with carbohydrate, protein and lipid content and methane yield, COD, biodegradability and methane production time prediction were compared with the reported absorbance of molecules on NIR spectra. Carbohydrate content prediction is related with the content of oxygenated molecules such as alcohols. Protein content prediction is related nitrogen containing molecules which makes sense since protein content prediction is calibrated on protein content estimation from TKN. Lipid content prediction is related with alkane and alkene. Biodegradability prediction is negatively impacted by phenolic compounds such as lignin. The similarities between *b* coefficients of each predictive model were also investigated. The carbohydrate content impacts negatively the COD prediction but positively the methane yield and biodegradability. Important similarities between the lipid content prediction and COD and methane yield prediction exist. *b* coefficients used for biodegradability or kinetics of methane production prediction are collinear. In contradiction to the mainstream thought, COD and methane yield prediction were anti-collinear and revealed low relation with methane kinetics prediction. The relations found are consistent with the correlation assessed by Pierson correlations between the biochemical composition and the AD process performances. Only the prediction of protein content relates with readily and highly biodegradable molecules while protein content did not correlate with methane production kinetics or methane yield using Pearson Correlation Coefficients (PCC) which could be explained by a slight difference on the dataset used.

Eventually, predictive ADM1 model taking into account both biological state and substrates characterization was modified. The sequential hydrolysis model fits all substrates biodegradation while the simultaneous hydrolysis model is sometimes not applicable. Sequential hydrolysis is preferred for solid waste hydrolysis modelling but the biological mechanism of hydrolysis remains not fully understood. A major drawback of ADM1 was its parameters estimation. NIR spectroscopy enables a fast estimation of the parameters describing the substrate which are needed to implement a modified ADM1. It facilitates the use of ADM1 to predict the optimal performance of degradation of any characterized input.

Overall, this study demonstrates that NIR spectroscopy coupled with PLS regression enables a fast implementation of the parameters describing the substrate in ADM1 while SNAC estimates the process-intermediates for the biological state estimation of the reactor, which could be used to fit the process parameters of ADM1. The work detailed in this PhD aims to help the human operators monitor their plants. This approach can be used as a powerful tool for plant monitoring and feeding strategy optimization.

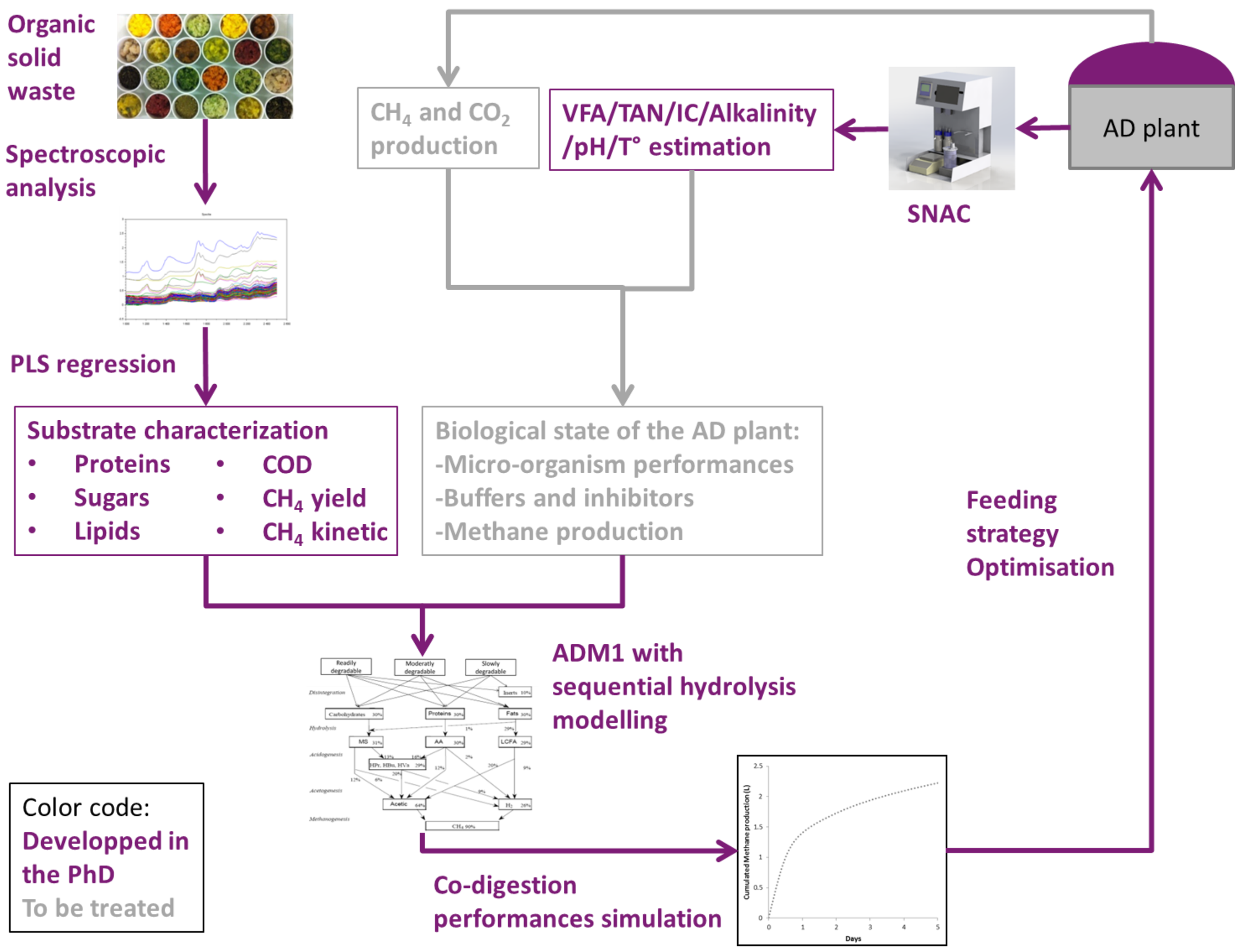

Organic solid waste
Spectroscopic analysis
PLS regression
Substrate characterization
Proteins
Sugars
Lipids
COD
CH_4 yield
CH_4 kinetic
CH_4 and CO_2 production
VFA/TAN/IC/Alkalinity /pH/T° estimation
SNAC
AD plant
Biological state of the AD plant:
-Micro-organism performances
-Buffers and inhibitors
-Methane production
Feeding strategy Optimisation
ADM1 with sequential hydrolysis modelling
Co-digestion performances simulation
Color code:
Developped in the PhD
To be treated
Days

Figure 63: Feeding strategy optimization, current and future development

8.2 Outlooks

Nonetheless, for industrial application, further developments must be done to improve the approach, methods and tools developed (Figure 63).

- The SNAC method, validated for the analysis of plant samples in laboratory must be validated on plant monitoring. For that purpose, a sensor using the SNAC method has been prototyped. The prototype estimates the temperature, pH, conductivity and alkalinity ratio and VFA/IC/TAN concentrations of the plant. Along with the biogas production and composition, it aims to estimate the biological state of the plant. It will allow assessing for instance the inhibitory compounds concentration and micro-organism activity. The second step will be to use SNAC and gas flow rate and composition to set the ADM1 variable related to the process.
- The current database used for NIR calibration will have to be increased for industrial application. Particularly FOG and sludge are not enough represented in the database
- NIR analysis can be done on dry samples only. Three disadvantages are linked to drying: volatile compounds cannot be analyzed, the drying process considerably slow down the analysis and on-line application cannot be considered. In large scale application, the high change frequency of the inputs and the heterogeneity of the substrates are a threat for substrate characterization for plant performances prediction. The approach developed on dry samples using NIRs enables the characterization of various samples from an heterogeneous substrate in a shorter period of time which fits more the industrial expectancy. Nonetheless, a further interesting development or improvement of this technology would be to work with wet samples and to install online NIR sensors able to characterize in real time any input coming into the plant. The obstacle is the absorption of water using NIR that disturbs the analysis. Mayer (2015) provides for example some hints about the estimation of methane yield on wet samples using NIR
- PLSr model is calibrated on NIRs to predict the MPR under non-limiting batch conditions in which the hydrolysis is the limiting step. For a better understanding of the potential inhibitory effects, it would be interesting to get further investigations on the process-intermediate kinetics. Ammoniacal nitrogen, LCFA and VFA should be followed during the AD batch experiments to get more detailed information about the ammoniacal nitrogen release, LCFA and VFA production kinetics.

- Monitoring and feeding strategy optimization coupling SNAC, ADM1 and NIRs must be validated on industrial plant
- pH could also be a problem of AD process performances prediction. Indeed, in the current study, the inputs influence the pH through VFA, IC and TAN production supposing a neutral pH of the inputs. But the inputs could be acid or basic, strongly affecting the [H^+] concentration by their feeding before the production of VFA, IC and TAN production. Since the organic loading of the batch experiments is very low, the pH of the inputs affects only slightly the batch pH. But at industrial scale with a more important load, this has to be taken into account.
- Other compounds such as sulfur, phosphorous and antibiotics affect the plant performances but are not yet taken into account neither in NIRs prediction nor in ADM1. Taking the example of sulfur, its estimation could be challenging because of the high sensitivity of the process to low variation in sulfur concentration.
- Digestate encounters a growing interest for land field application and homologation of the digestates as fertilizers is regularly discussed in conferences. Since the fertilizing substances contained in the digestate are strongly related with the feeding strategy, monitoring the inputs to optimize the fertilizing substance concentration in the digestate will be an important challenge.
- Eventually, the rheology of the inputs is a major concern of the plant operator since it strongly affects the pumpability of the inputs. For its simplicity, rheology is mainly assessed from the TS analysis even if recent studies decorrelate TS from rheology (Liu et al., 2015)).

8.3 Projected development

Projects are currently in the pipe to handle on these further developments. SNAC prototype has been recently installed in a farm unit equipped with supervision system to validate its advantages for plant monitoring.

The approach NIRs plus SNAC and predictive model for feeding optimization should be tested on units equipped with supervision system. The results, currently under acquisition, are expected to be presented in the conference “Journées de la methanisation industrielle et agricole”, from the 6-8 December 2016 in Chambery.

A project bringing together two Small and Medium-sized Enterprise (SME), one multinational and two public institutions has been submitted to ADEME. If the project is accepted, it will accelerate the industrialization of the solutions developed in this PhD. The project aims to:

- facilitate the connections between waste producers, anaerobic digestion plants and digestate users
- Collect and centralize the information related with organic wastes, AD plant demand and digestate on a platform, accessible by the actors of AD
- Optimize the feeding strategy to increase the AD plant profitability

To do so, one part of the preliminary budget is allocated to the deployment of the solution developed in this PhD. SNAC sensor will be produce in small series (10-20 sensors) to equip 10-20 units and to assist the estimation of their biological states. A second part is allocated on the improvement of the dataset used for the calibration of the PLSr using NIRs on a wide range of waste with the aim to characterize the organic waste produced at territory scale. Eventually, the approach NIRs plus SNAC and predictive model (MeMo®) will be tested to monitor 5-6 units to optimize their feeding strategy and profitability.

Another part of the project focuses on the extension of the work done to predict the performances of biogas production on digestate valorization. It first consists in the development of a fast method of characterization of the digestate fertilizing substances based on spectroscopic analysis to optimize their valorization path. Second, as ADM1 was used for the prediction of the biogas production, models should be created to anticipate the agronomical quality depending on the feeding strategy for a better valorization of the digestate.

Eventually, extension of fast spectroscopic methods of prediction on wet samples, investigation on the prediction of the process-intermediate kinetics and pH modelling could be the basis of a new PhD project.

Chapter 9. References

Aceves-Lara, C.-A., Latrille, E., Conte, T., Steyer, J.-P., 2012. Online estimation of VFA, alkalinity and bicarbonate concentrations by electrical conductivity measurement during anaerobic fermentation. Water Sci. Technol. 65, 1281–1289. doi:10.2166/wst.2012.703

Alburquerque, J.A., de la Fuente, C., Ferrer-Costa, A., Carrasco, L., Cegarra, J., Abad, M., Bernal, M.P., 2012. Assessment of the fertiliser potential of digestates from farm and agroindustrial residues. Biomass and Bioenergy 40, 181–189. doi:10.1016/j.biombioe.2012.02.018

Alibardi, L., Cossu, R., 2016. Effects of carbohydrate, protein and lipid content of organic waste on hydrogen production and fermentation products. Waste Manag. 47, 69–77. doi:10.1016/j.wasman.2015.07.049

Al-Kandari, N.M., Jolliffe, I.T., 2005. Variable selection and interpretation in correlation principal components. Environmetrics 16, 659–672. doi:10.1002/env.728

Angelidaki, I., Alves, M., Bolzonella, D., Borzacconi, L., Campos, J.L., Guwy, A.J., Kalyuzhnyi, S., Jenicek, P., Van Lier, J.B., 2009. Defining the biomethane potential (BMP) of solid organic wastes and energy crops: A proposed protocol for batch assays. Water Sci. Technol. 59, 927–934. doi:10.2166/wst.2009.040

Angelidaki, I., Sanders, W., 2004. Assessment of the anaerobic biodegradability of macropollutants. Rev. Environ. Sci. Biotechnol. 3, 117-129. doi:10.1007/s11157-004-2502-3

Appels, L., Baeyens, J., Degrève, J., Dewil, R., 2008. Principles and potential of the anaerobic digestion of waste-activated sludge. Prog. Energy Combust. Sci. 34, 755–781. doi:10.1016/j.pecs.2008.06.002

Appels, L., Lauwers, J., Degrève, J., Helsen, L., Lievens, B., Willems, K., Van Impe, J., Dewil, R., 2011. Anaerobic digestion in global bio-energy production: Potential and research challenges. Renew. Sustain. Energy Rev. 15, 4295–4301. doi:10.1016/j.rser.2011.07.121

Aquino, S.F., Chernicharo, C.A.L., Soares, H., Takemoto, S.Y., Vazoller, R.F., 2008. Methodologies for determining the bioavailability and biodegradability of slduges. Environ. Technol. 29 (8), 855-862.

ATEE Club Biogaz, 2011. Etat des lieux de la consommation des « produits fermiers » en France. Available on http://atee.fr/sites/default/files/fichiers/2011-09_etatdeslieux_methanisation_clubbiogaz_v3.pdf

Bagchi, T.B., Sharma, S., Chattopadhyay, K., 2016. Development of NIRS models to predict protein and amylose content of brown rice and proximate compositions of rice bran. Food Chem. 191, 21–27. doi:10.1016/j.foodchem.2015.05.038

Bastide, G., 2015. Methanisation Fiche technique- ADEME Angers. Available on http://www.ademe.fr/sites/default/files/assets/documents/expertsie_dechets_-_fiche_technique_methanisation.pdf

Batstone, D.J., Keller, J., Angelidaki, R.I., Kalyuzhnyi, S.V., Pavlostathis, S.G., Rozzi, A., Sanders, W.T.M., Siegrist, H., Vavilin, V.A., 2002. Anaerobic Digestion Model No.1. Scientific and Technical Report No. 13. IWA Publishing, London, UK.

Batstone, D.J., Keller, J., Steyer, J.P., 2006. A review of ADM1 extensions, applications, and analysis: 2002–2005. Water Sci. Technol. 54, 1. doi:10.2166/wst.2006.520

Batstone, D.J., Puyol, D., Flores-Alsina, X., Rodríguez, J., 2015. Mathematical modelling of anaerobic digestion processes: applications and future needs. Rev. Environ. Sci. Biotechnol. 14, 595–613. doi:10.1007/s11157-015-9376-4

Batstone, D.J., Tait, S., Starrenburg, D., 2009. Estimation of hydrolysis parameters in full-scale anerobic digesters. Biotechnol. Bioeng. 102, 1513–1520. doi:10.1002/bit.22163

Bauer, A., Bösch, P., Friedl, A., Amon, T., 2009. Analysis of methane potentials of steam-exploded wheat straw and estimation of energy yields of combined ethanol and methane production. J. Biotechnol. 142, 50–55. doi:10.1016/j.jbiotec.2009.01.017

Bellaton, S., Guérin, S., Pautremat, N., Bernier, J., Muller, M., Motellet, S., Azimi, S., Pauss, A., Rocher, V., 2016. Early assessment of a rapid alternative method for the estimation of the biomethane potential of sewage sludge. Bioresour. Technol. 206, 279–284. doi:10.1016/j.biortech.2016.01.139

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.M., McBratney, A., 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. TrAC - Trends Anal. Chem. 29, 1073–1081. doi:10.1016/j.trac.2010.05.006

BenDor, E., Inbar, Y., Chen, Y., Ben-Dor, E., 1997. The reflectance spectra of organic matter in the visible near infrared and short wave infrared region(400-2500nm) during a controlled decomposition process. Remote sens.Environ (61), pp 1-15.

Bertran, E., Iturriaga, H., Maspoch, S., Montoliu, I., 2001. Effect of orthogonal signal correction on the determination of compounds with very similar near infrared spectra. Anal. Chim. Acta 431, 303–311. doi:10.1016/S0003-2670(00)01328-3

Boe, K., Batstone, D.J., Angelidaki, I., 2007. An innovative online VFA monitoring system for the anerobic process, based on headspace gas chromatography. Biotechnol. Bioeng. 96, 712–721. doi:10.1002/bit.21131

Boe, K., Steyer, J.P., Angelidaki, I., 2008. Monitoring and control of the biogas process based on propionate concentration using online VFA measurement. Water Sci. Technol. 57, 661–666. doi:10.2166/wst.2008.046

Bouvier, J.C., Steyer, J.P., Delgenes, J.P., 2002. On-line titrimetric sensor for the control of VFA and / or alkalinity in anaerobic digestion processes treating industrial vinasses. IWA VII Latin American Workshop and Symposium on Anaerobic Digestion, Merida, Mexico, 22–25 October, 2002, 65–68.

Brock, T.D., Madigan, M.T., Martinko, J.M., Parker, J., 1991. Biology of Microorganisms, 6th Edition, Prentice Hall, New Jersey. 874.

BSGA, EBA, ECN, ESPP, 2015. Bio-Nutrients and Organic Carbon Recycling are essential components of a Circular Economy 1–3.Available on http://www.compostnetwork.info/wordpress/wp-content/uploads/2015/11/151123_BSAG_EBA_ECN_ESPP-Jont-position-paper-circular-economy.pdf

Buffiere, P., Frederic, S., Marty, B., Delgenes, J., 2008. A comprehensive method for organic matter characterization in solid wastes in view of assessing their anaerobic biodegradability. Water Sci. Technol. 1783–1788. doi:10.2166/wst.2008.517

Buffiere, P., Loisel, D., Bernet, N., Delgenes, J.-P., 2006. Towards new indicators for the prediction of solid waste anaerobic digestion properties. Water Sci. Technol. 53, 233. doi:10.2166/wst.2006.254

Charnier, C., Latrille, E., Lardon, L., Miroux, J., Steyer, J.P., 2016. Combining pH and electrical conductivity measurements to improve titrimetric methods to determine ammonia nitrogen, volatile fatty acids and inorganic carbon concentrations. Water Res. 95, 268–279. doi:10.1016/j.watres.2016.03.017

Chen, Y., Cheng, J.J., Creamer, K.S., 2008. Inhibition of anaerobic digestion process: a review. Bioresour. Technol. 99, 4044–4064. doi:10.1016/j.biortech.2007.01.057

Confer, D.R., Logan, B.E., 1997. Molecular weight distribution of hydrolysis products during the biodegradation of model macromolecules in suspended and biofilm cultures. II. Dextran and dextrin. Water Res. 31, 2137–2145. doi:10.1016/S0043-1354(97)00050-X

Cossu, R., Raga, R., 2008. Test methods for assessing the biological stability of biodegradable waste. Waste Manag. 28, 381–388. doi:10.1016/j.wasman.2007.01.014

Costello, R.C., Sullivan, D.M., 2013. Determining the pH Buffering Capacity of Compost Via Titration with Dilute Sulfuric Acid. Waste and Biomass Valorization 5, 505–513. doi:10.1007/s12649-013-9279-y

Dardenne, P., 2010. Some considerations about NIR spectroscopy: Closing speech at NIR-2009. NIR news 21, 8-14. doi:10.1255/nirn.1165

Dimock, R., Morgenroth, E., 2006. The influence of particle size on microbial hydrolysis of protein particles in activated sludge. Water Res. 40, 2064–2074. doi:10.1016/j.watres.2006.03.011

DuBois, M., Gilles, K.A., Hamilton, J.K., Rebers, P.A., Smith, F., 1956. Colorimetric method for determination of sugars and related substances. Anal. Chem. 28, 350–356. doi:10.1021/ac60111a017

Eastman, J. a, Ferguson, J.F., 1981. Solubilization of particulate organic carbon during the acid phase of anaerobic digestion. J. Water Pollut. Control Fed. 53(31), 352–366.

EBA, 2015. Contribution of anaerobic digestion to the European Circular Economy. Available on http://european-biogas.eu/wp-content/uploads/2015/08/2015-08-19-circular-economy-paper_final1.pdf

EBA, 2014. Biogas report, 1-29. Available on http://european-biogas.eu/2014/12/16/4331/.

E-cube, 2015. Etat des lieux de la fillière biogaz 1–109. Available on http://atee.fr/sites/default/files/2015-11-26_etat-des-lieux-biogaz_rapport_final.pdf

Ekama, G.A., Dold, P.L., Marais, G. v. R., 1986. Procedures for determining influent COD fractions and the maximum specific growth rate of heterotrophs in activated sludge systems. Water Sci. Technol. 18, 91–114.

Ekama, G.A., Marais, G. v R., 1979. Dynamic Behaviour of the Activated Sludge Process. Water Pollut. Control Fed. 51, 534–556.

EurObserv'er, 2014. Biogas Barometer. Available on http://www.eurobserv-er.org/biogas-barometer-2014/

FAO, 2015. World fertilizer trends and outlook to 2018, Food and Agriculture Organization of United Nations. Rome 2015. ISBN 978-92-5-108692-6

Feitkenhauer, H., von Sachs, J., Meyer, U., 2002. On line titration of volatile fatty acids for the process control of anaerobic digestion plants. Water Res. 36, 212–218.

Ferré, J., Brown, S.D., Xavier Rius, F., 2001. Improved calculation of the net analyte signal in inverse multivariate calibration. J. Chemom. 15, 537–553. doi:10.1002/cem.647

Fourati, S., Cristescu, R., Loboda, A., Talla, A., Filali, A., Railkar, R., Schaeffer, A.K., Favre, D., Gagnon, D., Peretz, Y., Wang, I., Beals, C.R., 2016. Pre-vaccination inflammation and B-cell signalling predict age-related hyporesponse to hepatitis B vaccination. Nature Communication. 1–12. doi:10.1038/ncomms10369

G.Bastide, 2014. Suivi technique ,économique, environnemental et social d'installations de méthanisation, ADEME 1–23. Étude réalisée pour le compte de l’ADEME par Apesa et Biomasse Normandie D’après les suivis réalisés par S3D, Méthaneva et le LDAR.

García-Gen, S., Rodríguez, J., Lema, J.M., 2015a. Control strategy for maximum anaerobic co-digestion performance. Water Res. 80, 209–216. doi:10.1016/j.watres.2015.05.029

García-Gen, S., Sousbie, P., Rangaraj, G., Lema, J.M., Rodríguez, J., Steyer, J.-P., Torrijos, M., 2015b. Kinetic modelling of anaerobic hydrolysis of solid wastes, including disintegration processes. Waste Manag. 35, 96–104. doi:10.1016/j.wasman.2014.10.012

Ginestet, P., Maisonnier, A., Spérandio, M., 2002. Wastewater COD characterization: Biodegradability of physico-chemical fractions. Water Sci. Technol. 45, 89–97.

Girault, R., Bridoux, G., Nauleau, F., Poullain, C., Buffet, J., Steyer, J.-P., Sadowski, a G., Béline, F., 2012. A waste characterisation procedure for ADM1 implementation based on degradation kinetics. Water Res. 46, 4099–110. doi:10.1016/j.watres.2012.04.028

Haag, J.E., Vande Wouwer, a., Queinnec, I., 2003a. Macroscopic modelling and identification of an anaerobic waste treatment process. Chem. Eng. Sci. 58, 4307–4316. doi:10.1016/S0009-2509(03)00272-0

Haag, J.E., Van de Wouwer, a., Queinnec, I., 2003b. Macroscopic modelling and identification of an anaerobic waste treatment process. Chem. Eng. Sci. 58, 4307–4316. doi:10.1016/S0009-2509(03)00272-0

Hansen, T.L., Schmidt, J.E., Angelidaki, I., Marca, E., Jansen, J.L.C., Mosbæk, H., Christensen, T.H., 2004. Method for determination of methane potentials of solid organic waste. Waste Manag. 24, 393–400. doi:10.1016/j.wasman.2003.09.009

Hejnfelt, A., Angelidaki, I., 2009. Anaerobic digestion of slaughterhouse by-products. Biomass and Bioenergy 33, 1046–1054. doi:10.1016/j.biombioe.2009.03.004

Henze, M., Grady, C.P.L., Gujer, W., M., G.v.R., Matsuo, T., 1987. Activated Sludge Model No. 1, IAWPRC Scientific and Technical Reports, No.1. IWA Publishing, London, UK.

International Energy Agency, 2015. Energy and Climate Change. World Energy Outlook Special Report. 1–200. doi:10.1038/479267b

International Energy Agency, 2014. World Energy Outlook 690 pp. doi:10.1787/20725302

Izhaki, I., 1993. Influence of nonprotein nitrogen on estimation of protein from total nitrogen in fleshy fruits. J. Chem. Ecol. 19, 2605–2615. doi:10.1007/BF00980695

Jacobi, H.F., Ohl, S., Thiessen, E., Hartung, E., 2012. NIRS-aided monitoring and prediction of biogas yields from maize silage at a full-scale biogas plant applying lumped kinetics. Bioresour. Technol. 103, 162–172. doi:10.1016/j.biortech.2011.10.012

Jenkins, S.R., Morgan, J.M., S.C.L., 1983. Measuring anaerobic sludge digestion and growth by a simple alkalimetric titration.pdf. J. WPCF 55, 451–453.

Jensen, P.D., Ge, H., Batstone, D.J., 2011. Assessing the role of biochemical methane potential tests in determining anaerobic degradability rate and extent. Water Sci. Technol. 64, 880–886. doi:10.2166/wst.2011.662

Jimenez, J., 2012. Organic matter characterization with 3D fluorescence spectroscopy for anaerobic digestion modeling of wastewater treatment sludge. PhD thesis Université de Montpellier II.

Jimenez, J., Aemig, Q., Doussiet, N., Steyer, J.-P., Houot, S., Patureau, D., 2015. A new organic matter fractionation methodology for organic wastes: bioaccessibility and complexity characterization for treatment optimization. Bioresour. Technol. doi:10.1016/j.biortech.2015.07.037

Jimenez, J., Gonidec, E., Cacho Rivero, J.A., Latrille, E., Vedrenne, F., Steyer, J.-P., 2014. Prediction of anaerobic biodegradability and bioaccessibility of municipal sludge by coupling sequential extractions with fluorescence spectroscopy: towards ADM1 variables characterization. Water Res. 50, 359–72. doi:10.1016/j.watres.2013.10.048

Jimenez, J., Latrille, E., Harmand, J., Robles, A., Ferrer, J., Gaida, D., Wolf, C., Mairet, F., Bernard, O., Alcaraz-Gonzalez, V., Mendez-Acosta, H., Zitomer, D., Totzke, D., Spanjers, H., Jacobi, F., Guwy, A., Dinsdale, R., Premier, G., Mazhegrane, S., Ruiz-Filippi, G., Seco, A., Ribeiro, T., Pauss, A., Steyer, J.P., 2015. Instrumentation and control of anaerobic digestion processes: a review and some research challenges. Rev. Environ. Sci. Biotechnol. 14, 615–648. doi:10.1007/s11157-015-9382-6

Južnič-Zonta, Ž., Kocijan, J., Flotats, X., Vrečko, D., 2012. Multi-criteria analyses of wastewater treatment bio-processes under an uncertainty and a multiplicity of steady states. Water Res. 46, 6121–6131. doi:10.1016/j.watres.2012.08.035

Karthikeyan, O.P., Visvanathan, C., 2013. Bio-energy recovery from high-solid organic substrates by dry anaerobic bio-conversion processes: A review. Rev. Environ. Sci. Biotechnol. 12, 257–284. doi:10.1007/s11157-012-9304-9

Koch, K., Lübken, M., Gehring, T., Wichern, M., Horn, H., 2010. Biogas from grass silage - Measurements and modeling with ADM1. Bioresour. Technol. 101, 8158–8165. doi:10.1016/j.biortech.2010.06.009

Krapf, L.C., Nast, D., Gronauer, A., Schmidhalter, U., Heuwinkel, H., 2013. Transfer of a near infrared spectroscopy laboratory application to an online process analyser for in situ monitoring of anaerobic digestion. Bioresour. Technol. 129, 39–50. doi:10.1016/j.biortech.2012.11.027

Labatut, R. a, Angenent, L.T., Scott, N.R., 2011. Biochemical methane potential and biodegradability of complex organic substrates. Bioresour. Technol. 102, 2255–2264. doi:10.1016/j.biortech.2010.10.035

Lagarde, F., Tusseau-Vuillemin, M.H., Lessard, P., Héduit, A., Dutrop, F., Mouchel, J.M., 2005. Variability estimation of urban wastewater biodegradable fractions by respirometry. Water Res. 39, 4768–4778. doi:10.1016/j.watres.2005.08.026

Lahav, O., Loewenthal, R.E., 2000. Measurement of VFA in anaerobic digestion: The five-point titration method revisited. Water SA 26, 389–392.

Lahav, O., Morgan, B.E., 2004. Titration methodologies for monitoring of anaerobic digestion in developing countries - A review. J. Chem. Technol. Biotechnol. 79, 1331–1341. doi:10.1002/jctb.1143

Lashermes, G., Nicolardot, B., Parnaudeau, V., Thuriès, L., Chaussod, R., Guillotin, M.L., Linères, M., Mary, B., Metzger, L., Morvan, T., Tricaud, a., Villette, C., Houot, S., 2009. Indicator of potential residual carbon in soils after exogenous organic matter application. Eur. J. Soil Sci. 60, 297–310. doi:10.1111/j.1365-2389.2008.01110.x

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. Prog. Energy Combust. Sci. 39, 383–402. doi:10.1016/j.pecs.2013.03.003

Le Cao, K.-A., Rossow, D., Robert-Granié, C., Besse, P., 2008. A Sparse PLS for Variable Selection when Integrating Omics data. Stat. Appl. Genet. Mol. Biol. 7, Article. 35. doi:10.2202/1544-6115.1390

Lesteur, M., Bellon-Maurel, V., Gonzalez, C., Latrille, E., Roger, J.M., Junqua, G., Steyer, J.P., 2010. Alternative methods for determining anaerobic biodegradability: A review. Process Biochem. 45, 431–440. doi:10.1016/j.procbio.2009.11.018

Lesteur, M., Latrille, E., Maurel, V.B., Roger, J.M., Gonzalez, C., Junqua, G., Steyer, J.P., 2011. First step towards a fast analytical method for the determination of Biochemical Methane Potential of solid wastes by near infrared spectroscopy. Bioresour. Technol. 102, 2280–2288. doi:10.1016/j.biortech.2010.10.044

Liu, G.J., Liu, Y., Wang, Z.Y., Lei, Y.H., Chen, Z.A., Deng, L.W., 2015. The effects of temperature, organic matter and time-dependency on rheological properties of dry anaerobic digested swine manure. Waste Manag. 38, 449–454. doi:10.1016/j.wasman.2014.12.015

Long, J.H., Aziz, T.N., Reyes, F.L.D.L., Ducoste, J.J., 2012. Anaerobic co-digestion of fat, oil, and grease (FOG): A review of gas production and process limitations. Process Saf. Environ. Prot. 90, 231–

245. doi:10.1016/j.psep.2011.10.001

Lowry, O.H., Rosebrough, N.J., Farr, A.L., Randall, R.J., 1951. Protein measurement with the Folin phenol reagent. J. Biol. Chem. 193, 265–275. doi:10.1016/0304-3894(92)87011-4

Lützhøft, H.C.H., Boe, K., Fang, C., Angelidaki, I., 2014. Comparison of VFA titration procedures used for monitoring the biogas process. Water Res. 54, 262–272. doi:10.1016/j.watres.2014.02.001

Madsen, M., Holm-Nielsen, J.B., Esbensen, K.H., 2011b. Monitoring of anaerobic digestion processes: A review perspective. Renew. Sustain. Energy Rev. 15, 3141–3155. doi:10.1016/j.rser.2011.04.026

Massaccesi, L., Sordi, A., Micale, C., Cucina, M., Zadra, C., Di Maria, F., Gigliotti, G., 2013. Chemical characterisation of percolate and digestate during the hybrid solid anaerobic digestion batch process. Process Biochem. 48, 1361–1367. doi:10.1016/j.procbio.2013.06.026

Mata-Alvarez, J., Macé, S., Llabrés, P., 2000. Anaerobic digestion of organic solid wastes. An overview of research achievements and perspectives. Bioresour. Technol. 74, 3–16. doi:10.1016/S0960-8524(00)00023-7

Mayer, F., 2015. Biomethane yield of energy crops and prediction of their biochemical methane potential with near infrared spectroscopy. PhD thesis, Université de Louvain.

Moletta, R., 2011. La méthanisation, (2° Éd.). ed. lavoisier.

Molina, F., Castellano, M., García, C., Roca, E., Lema, J.M., 2009. Selection of variables for on-line monitoring, diagnosis, and control of anaerobic digestion processes. Water Sci. Technol. 60, 615–622. doi:10.2166/wst.2009.379

Monlau, F., Sambusiti, C., Ficara, E., Aboulkas, A., Barakat, A., Carrère, H., 2015. New opportunities for agricultural digestate valorization: current situation and perspectives. Energy Environ. Sci. 8, 2600–2621. doi:10.1039/C5EE01633A

Moosbrugger, R.E., Wentzel, M.C., Ekama, G.A., Marais v., G.R., 1993. A 5 pH point titration method for determining the carbonate and SCFA weak acid/bases in anaerobic systems, in: Water SA, 19(1) , 29-40

Morgenroth, E., Kommedal, R., Harremoes, P., 2002. Processes and modeling of hydrolysis of particulate organic matter in aerobic wastewater treatment—a review. Water Sci. Technol. 45, 25–40.

Motte, J.-C., Escudié, R., Beaufils, N., Steyer, J.-P., Bernet, N., Delgenès, J.-P., Dumas, C., 2014. Morphological structures of wheat straw strongly impacts its anaerobic digestion. Ind. Crops Prod. 52, 695–701. doi:10.1016/j.indcrop.2013.11.038

Mottet, a., Ramirez, I., Carrère, H., Déléris, S., Vedrenne, F., Jimenez, J., Steyer, J.P., 2013. New fractionation for a better bioaccessibility description of particulate organic matter in a modified ADM1 model. Chem. Eng. J. 228, 871–881. doi:10.1016/j.cej.2013.05.082

Muller, M., Jimenez, J., Antonini, M., Dudal, Y., Latrille, E., Vedrenne, F., Steyer, J.-P., Patureau, D., 2014. Combining chemical sequential extractions with 3D fluorescence spectroscopy to characterize sludge organic matter. Waste Manag. 34, 2572–2580. doi:10.1016/j.wasman.2014.07.028

Nicolai, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., Lammertyn, J., 2007. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy : A review. Postharvest Biol. Technol. 46, 99–118. doi:10.1016/j.postharvbio.2007.06.024

Nordmann, W., 1977. Die Überwachung der Schlammfaulung. Informationen für das Betriebspersonal. Beilage zur Korrespondenz Abwasser 3, 77.

Núñez-Sánchez, N., Martínez-Marín, A.L., Polvillo, O., Fernández-Cabanás, V.M., Carrizosa, J., Urrutia, B., Serradilla, J.M., 2016. Near Infrared Spectroscopy (NIRS) for the determination of the milk fat fatty acid profile of goats. Food Chem. 190, 244–252. doi:10.1016/j.foodchem.2015.05.083

Orhon, D., Ubay Çokgör E., Sözen, S., 1998. Dual hydrolysis model of the slowly biodegradable substrate in activated sludge systems. Biotechnol. Tech. 12, 737–741. doi:10.1023/A:1008860517183

Osborne, B., Fearn, T., 1986. Near infrared spectroscopy in food analysis, Food Chemistry. Harlow, UK. p 200.

Purser, B.J.J., Thai, S.-M., Fritz, T., Esteves, S.R., Dinsdale, R.M., Guwy, a J., 2014. An improved titration model reducing over estimation of total volatile fatty acids in anaerobic digestion of energy crop, animal slurry and food waste. Water Res. 61, 162–70. doi:10.1016/j.watres.2014.05.020

R.Lide, D., 2003. Handbook of Chemistry and Physics, CRC Press. p 2620

Rajagopal, R., Massé, D.I., Singh, G., 2013. A critical review on inhibition of anaerobic digestion process by excess ammonia. Bioresour. Technol. 143, 632–41. doi:10.1016/j.biortech.2013.06.030

Ramirez, I., Mottet, A., Carrère, H., Déléris, S., Vedrenne, F., Steyer, J.-P., 2009. Modified ADM1 disintegration/hydrolysis structures for modeling batch thermophilic anaerobic digestion of thermally pretreated waste activated sludge. Water Res. 43, 3479–3492. doi:10.1016/j.watres.2009.05.023

Raposo, F., Fernandez-Cegri, V., de la Rubia, M.A., Borja, R., Béline, F., Cavinato, C., Demirer, G., Fernandez, B., Fernandez-Polanco, M., Frigon, J.C., Ganesh, R., Kaparaju, P., Koubova, J., Mendez, R., Menin, G., Peene, A., Scherer, P., Torrijos, M., Uellendahl, H., Wierinck, I., de Wilde, V., 2011. Biochemical methane potential (BMP) of solid organic substrates: Evaluation of anaerobic biodegradability using data from an international interlaboratory study. J. Chem. Technol. Biotechnol. 86, 1088–1098. doi:10.1002/jctb.2622

Resch, M.G., Donohoe, B.S., Ciesielski, P.N., Nill, J.E., Magnusson, L., Himmel, M.E., Mittal, A., Katahira, R., Biddy, M.J., Beckham, G.T., 2014. Clean Fractionation Pretreatment Reduces Enzyme Loadings for Biomass Saccharification and Reveals the Mechanism of Free and Cellulosomal Enzyme Synergy. ACS Sustain. Chem. Eng.

Rincker, M., Diara, A., Peu, P., Badalato, N., Girault, R., Carrère, H., Bassard, D., 2013. Anaerobic respirometry as a tool to evaluate the effect of pretreatment on anaerobic digestion efficiency. 13th World Congress on Anaerobic Digestion, Santiago de Compostela, Spain. June 25 -28.

Ripley, L.E., Boyle, W.C., Converse, J.C., 1986. Improved Alkalimetric Monitoring for Anaerobic Digestion of High-Strength Wastes. Water Pollution Control Federation 58(5), 406–411.

Rosén, C., Jeppsson, U., 2008. Aspects on ADM1 Implementation within the BSM2 Framework. k. Tech. Report no. LUTEDX/(TEIE-7224). Department of Industrial Electrical Engineering and Automation, Lund University, Lund, Sweden. 1-35.

Ruiz, G., Molina, F., Steyer, J., Vanrolleghem, P., Zaher, U., Roca, E., 2005. Industrial scale validation of a new titrimetric sensor for anaerobic digestion processes : comparison of methodologies. Proceedings 2nd IWA Conference on Instrumentation, Control and Automation (ICA2005), Busan, Korea, May 2005.

Sanders, W.T.M., Geerink, M., Z., G. and Lettinga, G., 2000. Anaerobic hydrolysis kinetics of particulate substrates. Water Sci. Technol. 41, 17–24.

Shi, X.-S., Yuan, X.-Z., Wang, Y.-P., Zeng, S.-J., Qiu, Y.-L., Guo, R.-B., Wang, L.-S., 2014. Modeling of the methane production and pH value during the anaerobic co-digestion of dairy manure and spent mushroom substrate. Chem. Eng. J. 244, 258–263. doi:10.1016/j.cej.2014.02.007

Shimizu, T., Kudo, K., Nasu, Y., 1993. Anaerobic waste-activated sludge digestion-a bioconversion mechanism and kinetic model. Biotechnol. Bioeng. 41, 1082–1091. doi:10.1002/bit.260411111

Souza, T.S.O., Carvajal, A., Donoso-Bravo, A., Peña, M., Fdz-Polanco, F., 2013. ADM1 calibration using BMP tests for modeling the effect of autohydrolysis pretreatment on the performance of continuous sludge digesters. Water Res. 47, 3244–3254. doi:10.1016/j.watres.2013.03.041

Spanjers, H., Lier, J.B. va., 2006. Instrumentation in anaerobic treatment – research and practice. Water Sci. Technol. 53, 63-76. doi:10.2166/wst.2006.111

Spérandio, M., Paul, E., 1999. Estimation of wastewater biodegradable COD fractions by combining respirometric experiments in various So/Xo ratios. Water Res. 34, 1233–1246. doi:10.1016/S0043-1354(99)00241-9

Steyer, J.P., Bernard, O., Batstone, D.J., Angelidaki, I., 2006. Lessons learnt from 15 years of ICA in anaerobic digesters. Water Sci. Technol. 53, 25–33. doi:10.2166/wst.2006.107

Steyer, J.-P., Buffière, P., Rolland, D., Moletta, R., 1999. Advanced Control of Anaerobic Digestion. Water Res. 33, 2059–2068. doi:10.1016/S0043-1354(98)00430-8

Strömberg, S., Nistor, M., Liu, J., 2015. Early prediction of Biochemical Methane Potential through statistical and kinetic modelling of initial gas production. Bioresour. Technol. 176, 233–241. doi:10.1016/j.biortech.2014.11.033

Ugwuanyi, J.O., Harvey, L.M., McNeil, B., 2005. Effect of digestion temperature and pH on treatment efficiency and evolution of volatile fatty acids during thermophilic aerobic digestion of model high strength agricultural waste. Bioresour. Technol. 96, 707–719. doi:10.1016/j.biortech.2004.06.027

Van De Steene, M., Van Vooren, L., Ottoy, J.P., Vanrolleghem, P. a, 2002. Automatic buffer capacity model building for advanced interpretation of titration curves. Environ. Sci. Technol. 36, 715–723.

Vannecke, T.P.W., Lampens, D.R. a, Ekama, G. a, Volcke, E.I.P., Lützhøft, H.C.H., Boe, K., Fang, C., Angelidaki, I., 2014. Evaluation of the 5 and 8 pH point titration methods for monitoring anaerobic digesters treating solid waste. Water Res. 36, 861–869. doi:10.1016/j.watres.2014.02.001

Vavilin, V.A., Fernandez, B., Palatsi, J., Flotats, X., 2008. Hydrolysis kinetics in anaerobic degradation of particulate organic material: An overview. Waste Manag. 28, 939–951. doi:10.1016/j.wasman.2007.03.028

Vooren, L. Van, 2000. Buffer capacity based multipurpose hard- and software sensor for environmental applications. PhD thesis Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen.

Wan, S., Xi, B., Xia, X., Li, M., Lv, D., Wang, L., Song, C., 2012. Using fluorescence excitation-emission matrix spectroscopy to monitor the conversion of organic matter during anaerobic co-digestion of cattle dung and duck manure. Bioresour. Technol. 123, 439–444. doi:10.1016/j.biortech.2012.04.001

Wang, W., Xie, L., Luo, G., Zhou, Q., Angelidaki, I., 2013. Performance and microbial community analysis of the anaerobic reactor with coke oven gas biomethanation and in situ biogas upgrading. Bioresour. Technol. 146, 234–239. doi:10.1016/j.biortech.2013.07.049

Ward, A.J., 2016. Near-Infrared Spectroscopy for Determination of the Biochemical Methane Potential: State of the Art. Chem. Eng. Technol. 39, 611–619. doi:10.1002/ceat.201500315

Ward, A.J., Hobbs, P.J., Holliman, P.J., Jones, D.L., 2008. Optimisation of the anaerobic digestion of agricultural resources. Bioresour. Technol. 99, 7928–7940. doi:10.1016/j.biortech.2008.02.044

Williams, P., Norris, K., 2001. Near Infrared Technology in the Agricultural and Food Industries, Near-infrared technology in the agricultural and food industries, (2nd ed., pp. 145–169). St. Paul, MN:American Association of Cereal Chemists

Yasui, H., Goel, R., Li, Y.Y., Noike, T., 2008. Modified ADM1 structure for modelling municipal primary sludge hydrolysis. Water Res. 42, 249–59. doi:10.1016/j.watres.2007.07.004

Zaher, U., 2005. Modelling and monitoring the anaerobic digestion process in view of optimisation and smooth operation of WWTP's. PhD thesis, Falculteit Bio-ingenieurswetenschappen

Zaher, U., Bouvier, J.C., Steyer, J.P., Venrolleghem, P.A., 2004. Titrimetric Montoring of Anaerobic Digestion: VFA, Alkalinities and More, in: 10th World Congress of Anaerobic Digestion (AD10), Montreal, Canada, vol. 1, 2004. p. 330–6, 29 August–2 September 2004.

ANNEXES

1.1 *Guidelines of the annexes*

The annexes present supplementary results and materials. The section 1.2 presents the repartition of the COD fraction in proteins, carbohydrates, lipids and inert. The section 1.3 shows some 3D fluorescence spectra. The section 1.4 shows the NIR spectra. Eventually the section 1.5 provides some results of prediction with and without mixed samples into the dataset, taking the example of the methane production time and COD prediction.

1.2 *Repartition of the COD fractions of the substrate*

In this section are reported the sum of the COD fractions of 222 substrates included in the calibration set of NIR prediction with a full characterization in proteins, carbohydrates, lipids, biodegradability and COD. It mainly shows that the sum of the protein, carbohydrate and lipid fractions is around 90-100% whereas the sum of the protein, carbohydrate, lipid and inert fractions overpass strongly 100%. It leads to the hypothesis that a part of the inert fraction is composed of proteins, carbohydrates and lipids.

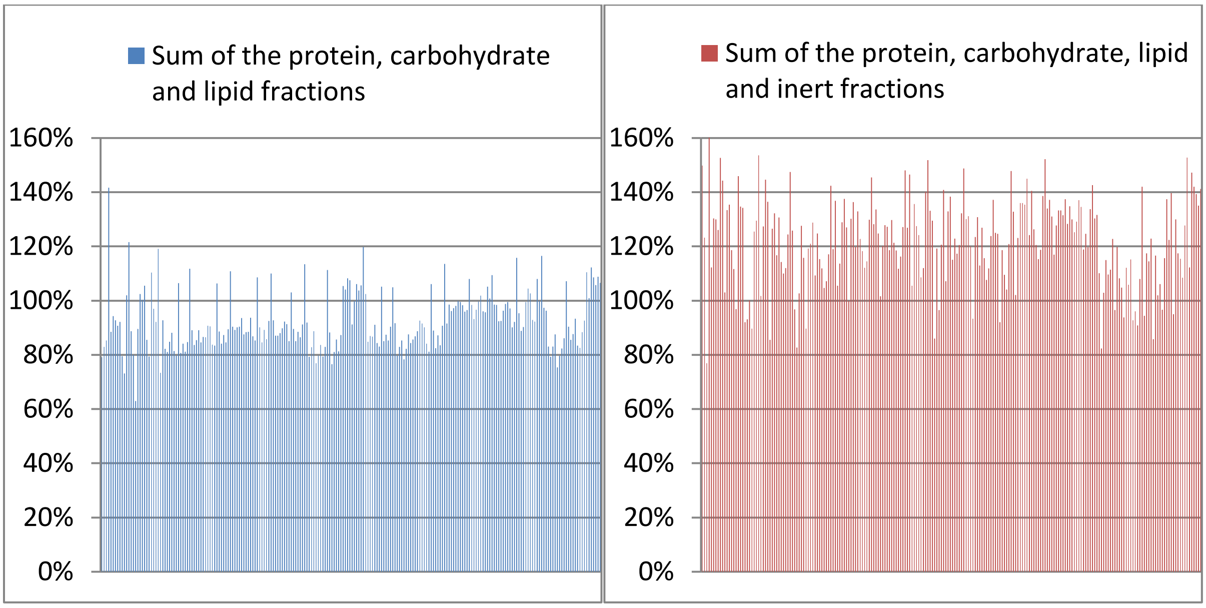


Figure 1: sum of the COD fraction on 222 substrates of the NIR calibration set with a full characterization in proteins, carbohydrates, lipids, COD and biodegradability.

1.3 Spectrum of 3D fluorescence

In the section are reported some of the 3D fluorescence spectrum analyzed after chemical sequential extraction. They provide an illustration of the spectrum obtained for some substrates. In the abscissa is represented the absorption wavelength and in the ordinate, the emission wavelength.

1.3.1 Micro-algae

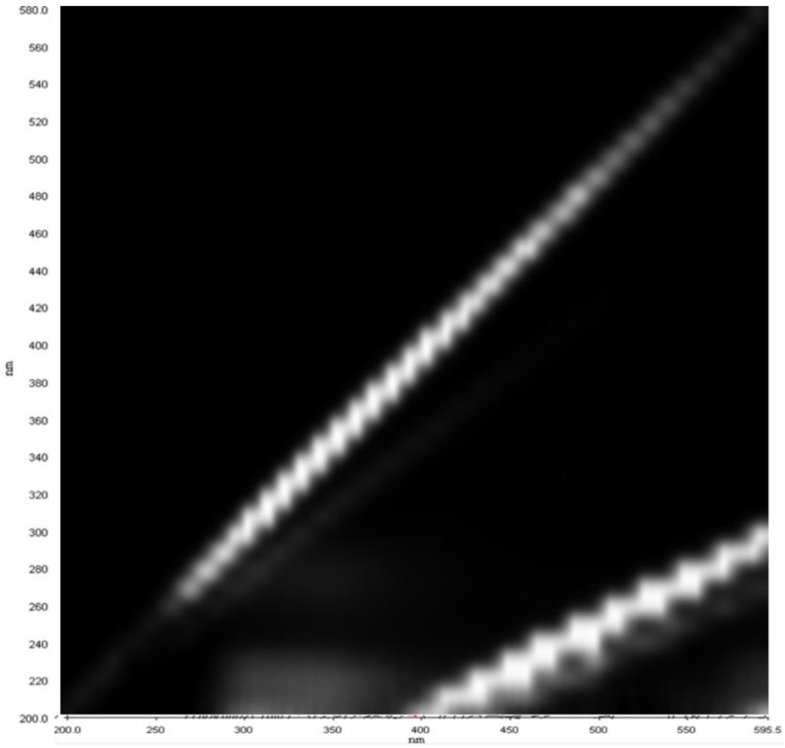


Figure 2: 3D fluorescence spectrum from micro algae. Fraction REOM, dilution 1000.

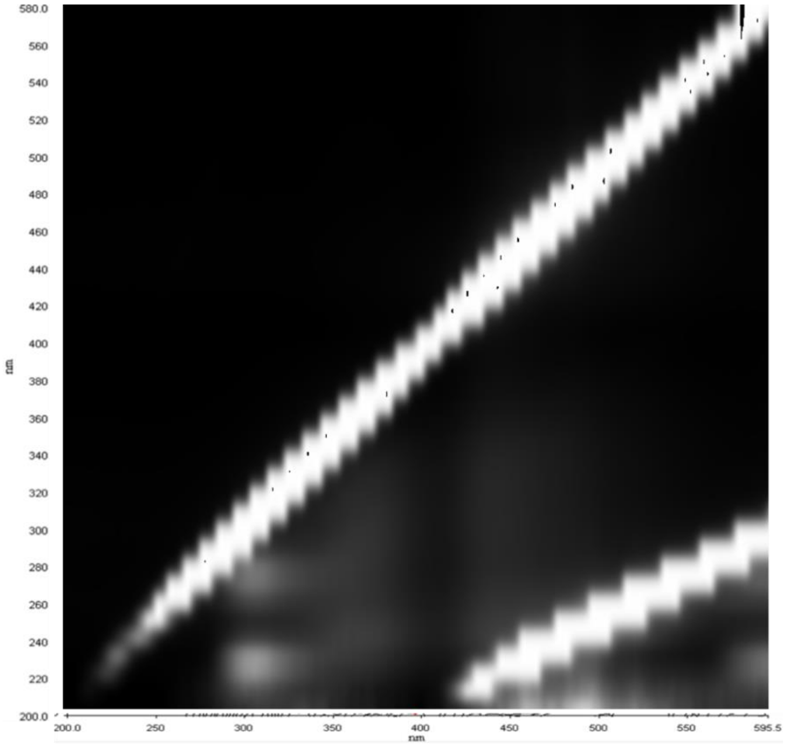


Figure 3: 3D fluorescence spectrum from micro algae. Fraction MEOM, dilution 100.

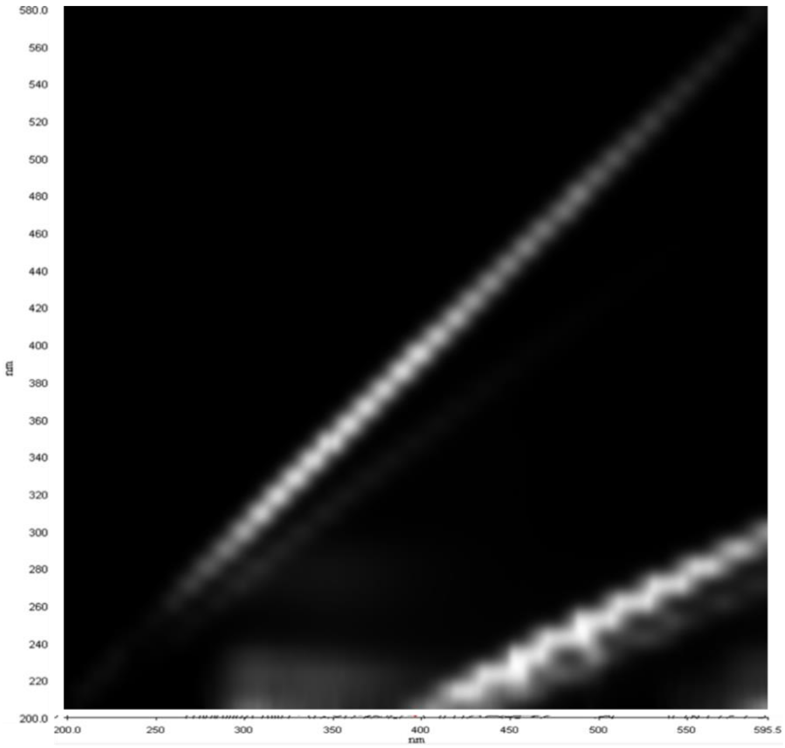


Figure 4: 3D fluorescence spectrum from micro algae. Fraction SEOM, dilution 1000.

1.3.2 *Straw*

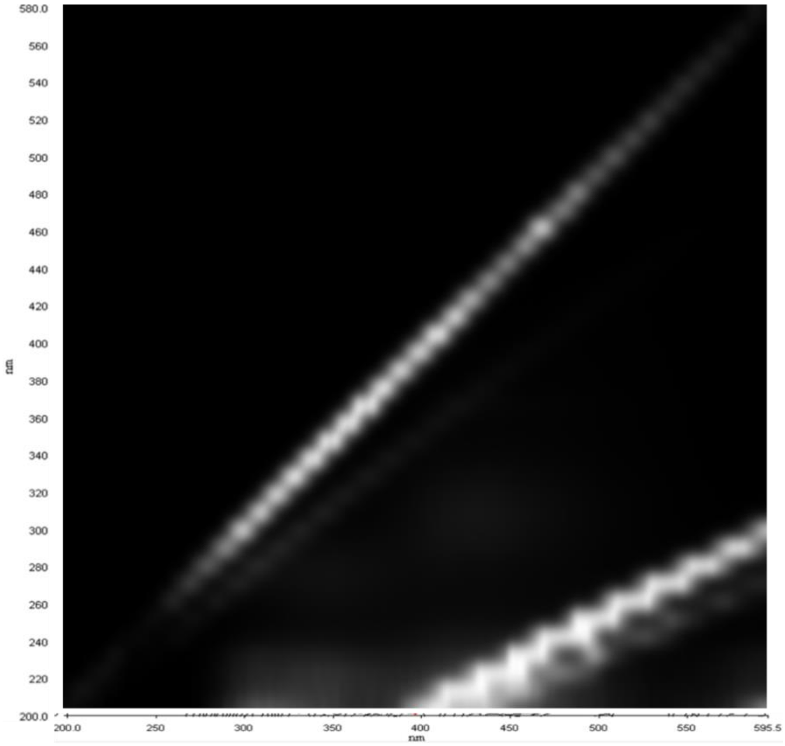


Figure 5: 3D fluorescence spectrum from straw. Fraction REOM, dilution 2000.

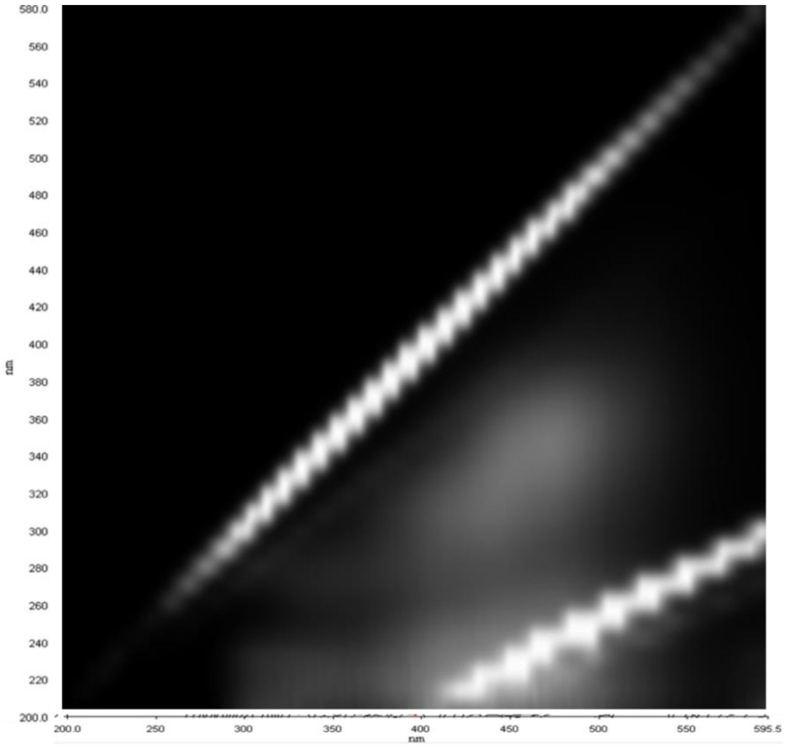


Figure 6: 3D fluorescence spectrum from straw. Fraction MEOM, dilution 200.

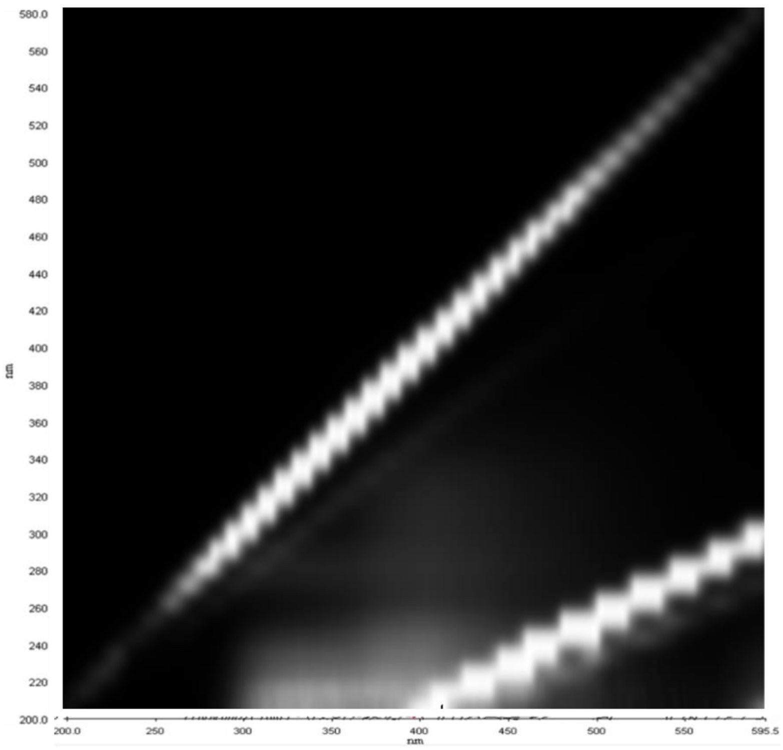


Figure 7: 3D fluorescence spectrum from straw. Fraction SEOM, dilution 200.

1.3.3 Carrot

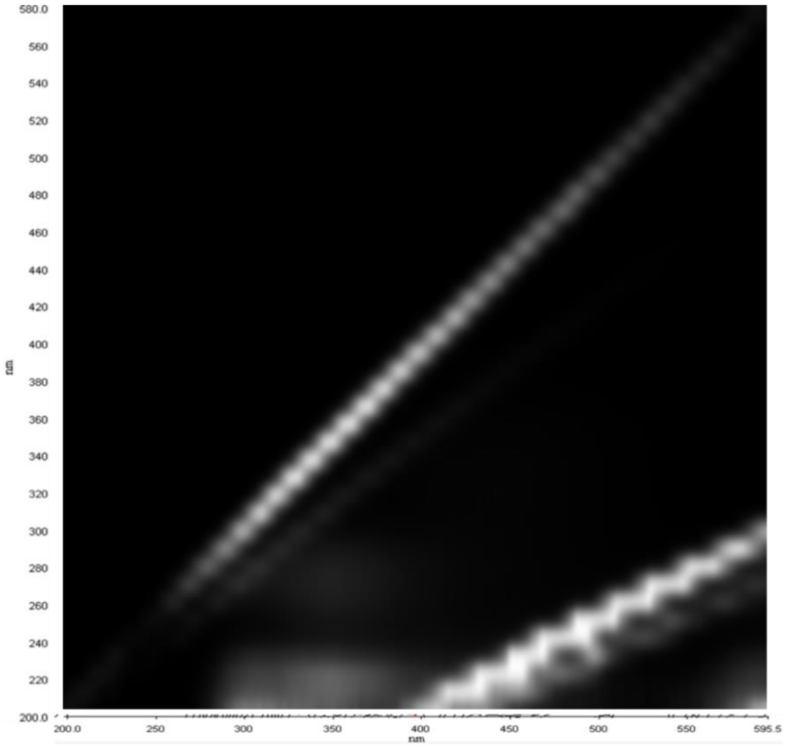


Figure 8: 3D fluorescence spectrum from carrot. Fraction REOM, dilution 500.

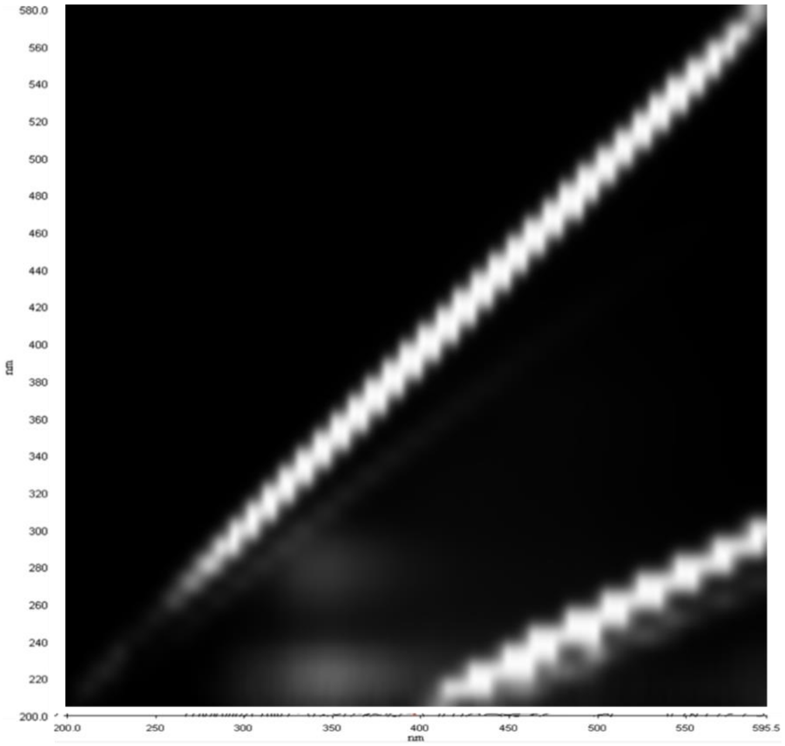


Figure 9: 3D fluorescence spectrum from carrot. Fraction MEOM, dilution 100.

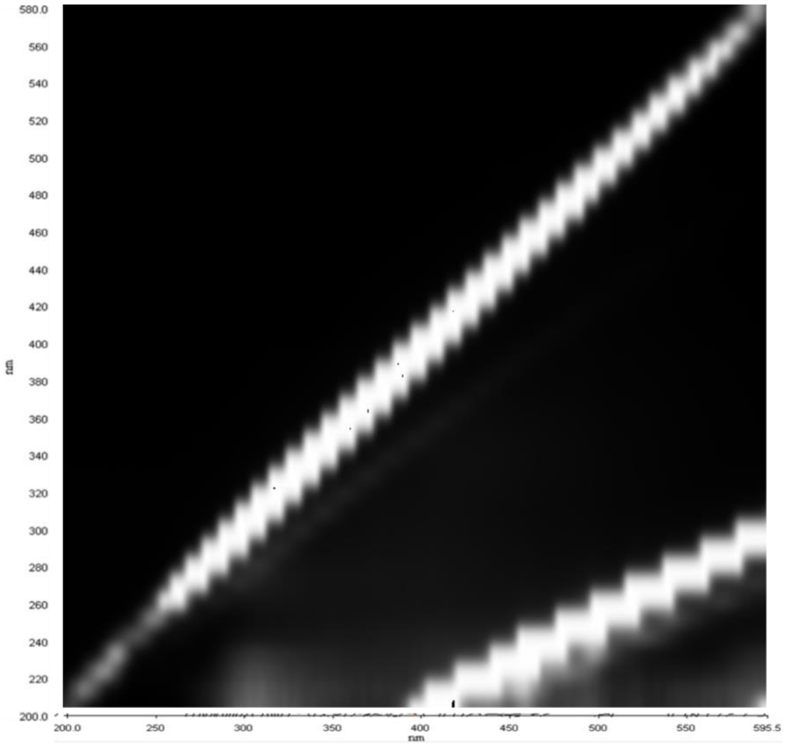


Figure 10: 3D fluorescence spectrum from carrot. Fraction SEOM, dilution 50.

1.4 Near infrared spectrum

In this section are reported some the NIR spectrum analyzed by reflectance. They provide an illustration of the spectrum obtained. In the abscissa is represented the wavelength in nanometer and in the ordinate, the reflectance.

1.4.1 Spectrum of the whole database

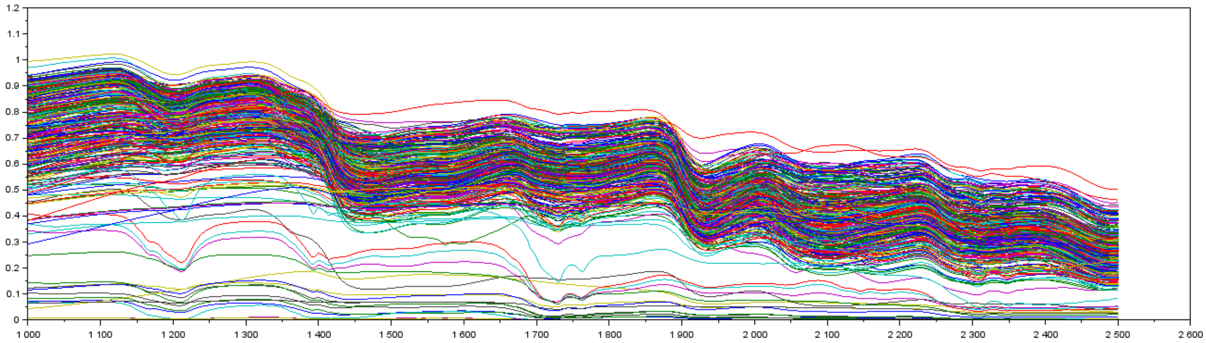


Figure 11: Spectrum of the whole database in reflectance. Reflectance in ordinate and wavelength (nm) in abscissa.

1.4.2 Example of some spectrum: carrot, straw, micro-algae, meat extract and table oil.

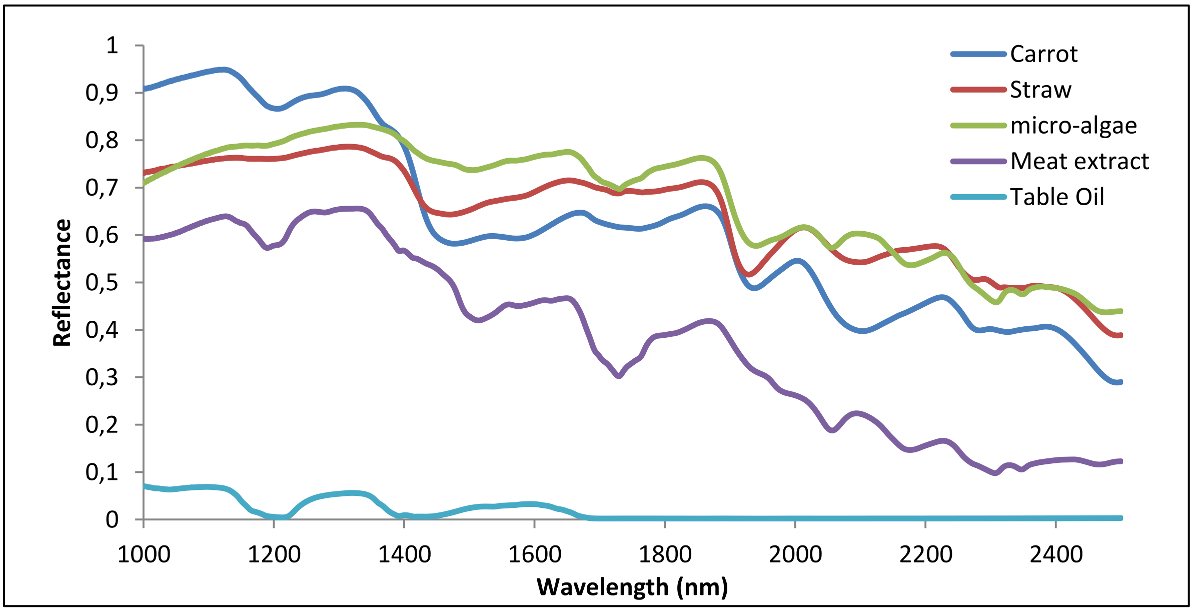


Figure 12: Spectrum of the some samples of the dataset in reflectance.

1.5 Prediction results with and without mixed samples

In this section is presented the results of cross-validation obtained for the methane production time at 70% of the BMP and the COD on a calibration set composed of only raw samples (without any mixture of samples) and with mixed samples. It shows that the performances of prediction are similar. Mixing the samples increased the number of different spectra and thus increases the robustness of the PLS model but the accuracy of the model can't be attributed to the presence of mixed samples into the calibration set.

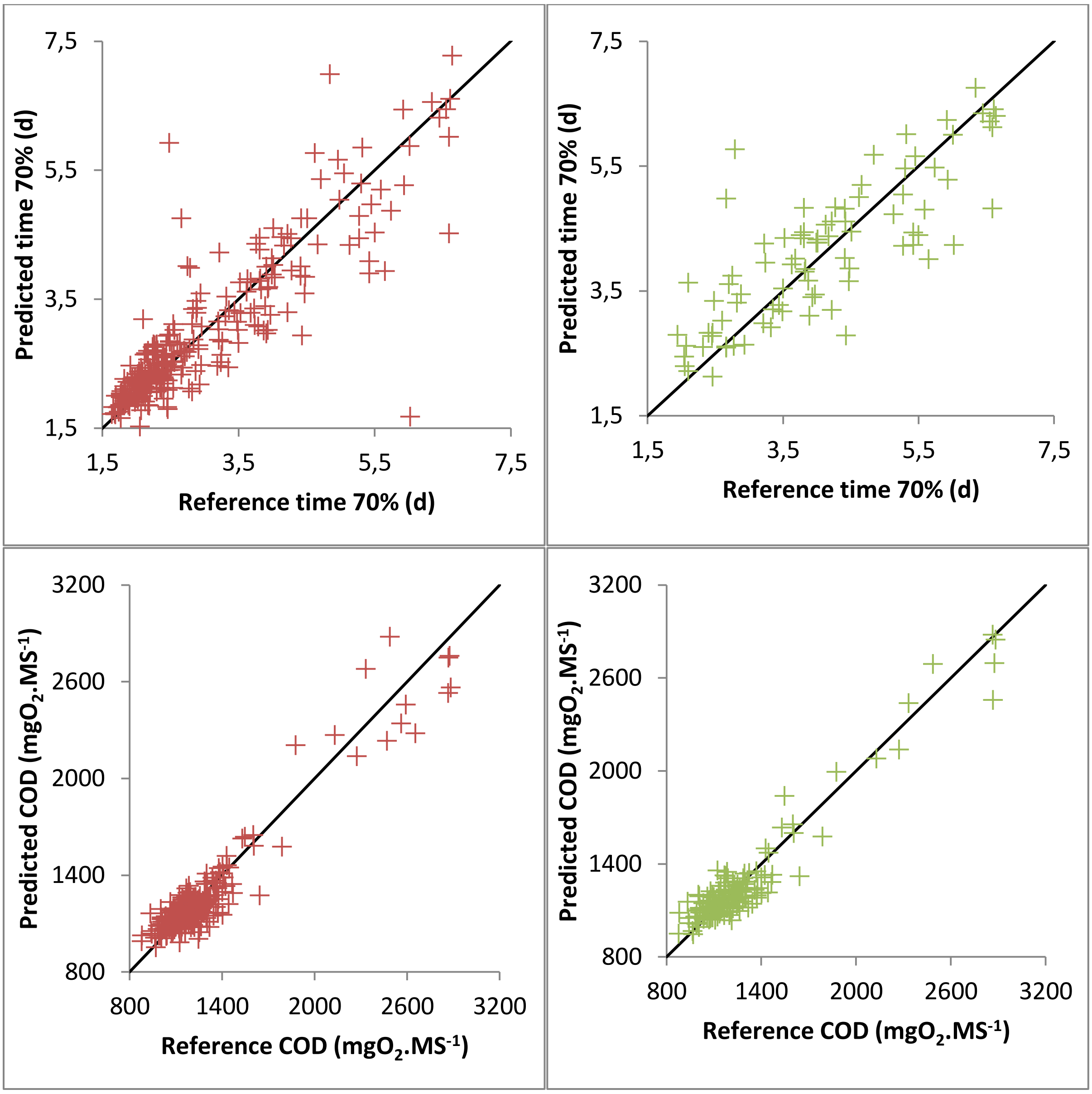


Figure 13: Cross validation results for methane production time at 70% of the BMP (up) and COD (down) on calibration sets composed of only raw samples (left and green) or raw and mixed samples (right and red).

FAST CHARACTERIZATION OF THE ORGANIC MATTER, INSTRUMENTATION AND MODELLING FOR THE ANAEROBIC DIGESTION PROCESS PERFORMANCES PREDICTION

Anaerobic digestion is an important pillar of the European circular economy, producing methane and organic fertilizers from waste. The development of the anaerobic digestion sector goes through co-digestion and feeding strategy optimization. Its development requires the biological state estimation of the plant, substrate characterization and predictive models simulating the plant performances, for which current solutions are not suitable. In this thesis, a titration sensor coupling pH and electrical conductivity for the estimation of volatile fatty acids, inorganic carbon and ammonia has been designed, improving the accuracy on volatile fatty acids estimation by 14.5 compared to current sensors. Along with biogas analyses, it allows estimating the biological state of the unit. Besides, fast near infrared spectroscopic analysis, estimating in a matter of minute carbohydrate, protein and lipid contents, chemical oxygen demand, methane production yield and kinetics, has been developed. Fast substrate characterization is then used to implement a modified Anaerobic Digestion Model n°1 which predicts the performances of a plant under optimal condition. Coupling biological state estimation to this approach enables to correct the prediction with the current state of the plant. This approach provides a powerful predictive tool for advanced control of anaerobic digestion plants and feeding strategy optimization.

DEFENDED ON 25TH NOVEMBER 2016 AT :

WITH THE FINANCIAL SUPPORT OF :

INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE
Unité de recherche (UR0050) - Laboratoire de Biotechnologie de l'Environnement

102, avenue des étangs
F-11100 Narbonne
France

Tél. : + 33 4 68 42 51 51
Courriel: lbe-contact@supagro.inra.fr
www.montpellier.inra.fr/narbonne