Modelling of agro-zootechnical anaerobic co-digestion for full-scale applications
Digital Supplementary material

Davide Carecci[1], Giulia Quarta[1], Arianna Catenacci[2], Gianni Ferretti[1], Elena Ficara[2]

I. The "high-fidelity" model: agri-AcoDM as an extended ADM1

The majority of the extensions are described in detail in other works [5], [4], [11] and are briefly lister hereafter:

- the disintegration step was removed;
- three carbohydrates (X_{ch}) and two proteins (X_{pr}) with different hydrolyzing behaviors (readily (r), mildly (m), and slowly (s) available) have been added [4]. Although more state variables were added, hydrolysis was still modeled as first-order process;
- dead bacterial biomass (X_{dec}) and lignin (X_{lig}) have been added, as they present a very different chemical oxygen demand over volatile solids (COD/VS) stoichiometric ratio with respect to the original sole non-biodegradable particulate (X_i);
- precipitation/dissolution processes have been modeled similarly to [2], and 4 salts have been added to grasp the fate of total ammoniacal nitrogen N (TAN, S_{in}), inorganic dissolved carbon C (S_{ic}) and phosphorous P (S_{ip}), modeled as in [11];
- to effectively grasp precipitation processes, activity corrections for the non-ideality of the liquid bulk were added for the acid-base physico-chemical equations;
- the accumulation of VFAs in the bulk is the main cause for methanogenesis inhibition, due to the pH reduction it may provoke. In addition, high VFA concentration was also proved to be inhibiting [9]. The large delay in the VFA to pH dynamics, especially in well-buffered systems, makes pH less-likely to be the main cause for inhibition. For this reason, Monod-like functions for the methane production kinetics were modified to Haldane-like ones [8].

The overall differential algebraic equation system is defined as in (1), where $\mathbf{x}$ are dynamic variables, $\mathbf{z}$ are algebraic variables, $\mathbf{u}$ are inputs, θ are fixed parameters and $\mathbf{y}$ are measurable outputs. Functions $\mathbf{f}$ and $\mathbf{g}$ are highly non-linear (and in some cases $\in C^0$).

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x},\mathbf{z},\mathbf{u},\theta) \\ \mathbf{0} = \mathbf{g}(\mathbf{x},\mathbf{z},\mathbf{u},\theta) \\ \mathbf{y} = \mathbf{h}(\mathbf{x},\mathbf{z},\mathbf{u},\theta) \end{cases} \tag{1}$$

The vector of parameters θ is partitioned into $\theta_{u,i}$, representing the concentration (kg m^{-3}) of state variables in the i^{th} co-feedstock (for each of the n co-feedstocks in the input flux), and θ_p, namely stoichiometric, physico-chemical and kinetic parameters (actually a function of process temperature).

An example of ODE for the liquid-solid mixture mass balances is Eq.2, i.e.;

$$\begin{aligned} \dot{\mathbf{x}} &= \frac{\sum_{i=1}^{n} u_i}{V}(\mathbf{x_{in}} - \mathbf{x}) + \mathbf{K}_G \mathbf{r}(\mathbf{x}, \theta_p, \mathbf{z}) \\ \mathbf{x_{in}} &= \frac{\sum_{i=1}^{n} u_i \theta_{u,i}}{\sum_{i=1}^{n} u_i} \end{aligned} \tag{2}$$

where $\mathbf{x}$ are the state variables, V (m^3) is the reactor volume, u_i (m^3 d^{-1}) is the volume flowrate of the i^{th} input co-feedstock, $\mathbf{K}_G$ is the transposed stoichiometric Gujer matrix that guarantees COD and mass conservation (C, N, P), extended and modified according to the new variables and processes introduced and $\mathbf{r}$ the vector of process kinetic rates. An example of model non-linearities is shown in Eq. (3), that defines the reaction rate r_{ac} of the acetoclastic methanogens population (X_{ac}), i.e.,

$$\begin{aligned} r = k_{\max,ac} f_{\mathrm{pH}} \frac{K_{inh,nh3}}{K_{inh,nh3} + S_{nh3}} \frac{S_{ac}}{K_{s,ac} + S_{ac} + \frac{S_{ac}^2}{K_{inh,ac}}} \\ \min_j \left(\frac{S_j}{S_j + K_{s,j}} \right) X_{ac} \end{aligned} \tag{3}$$

where $k_{\max,ac}$ is the maximum specific process rate and f_{pH} is the function that describes pH dependency [3]. The inhibition constant for free ammonia concentration (S_{nh3}) is $K_{inh,nh3}$ and $K_{s,j}$ is the half-saturation constant for the limiting nutrient S_j ($j = \{P,N\}$) [3]. The inhibition ($K_{inh,ac}$) and the half-saturation ($K_{s,ac}$) constants outline the Haldane-like function for total acetate concentration (S_{ac}) uptake. For gaseous states, Eq. (2) still holds and has a null influent contribution, V is the reactor head-space volume and u_i is the biogas outflow, regulated by a SISO PI controller that keeps it slightly over atmospheric pressure. The gas/liquid transfer kinetic rates are described according to the Fick's law. The much faster acid-base physico-chemical reactions have been modeled as an algebraic equation system ($\mathbf{z} \in \mathbb{R}^{30}$) [2].

The model, named agri-AcoDM, was implemented using Modelica, an open-source, high-level, declarative and object-oriented modeling language, and employing the free OpenModelica interpreter [12].

II. THE "REDUCED-ORDER" MODELS: AM2HN AND AM2HNTAN AS EXTENDED AM2

The ordinary differential equation system is defined as in 4 (with rational non-linear $\mathbf{f}$ and $\mathbf{g}$) and described hereafter.

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, \theta) \\ \mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{u}, \theta) \end{cases} \tag{4}$$

$$\dot{\mathbf{X}}_{\mathbf{h}} = D(\mathbf{X}_{\mathbf{h,in}} - \alpha \mathbf{X}_{\mathbf{h}}) - \mathbf{k}_{\mathbf{h}}\mathbf{X}_{\mathbf{h}} \tag{5}$$

$$\dot{X}_1 = (\mu_1 - \alpha D - k_{d,1})X_1 \tag{6}$$

$$\dot{X}_2 = (\mu_2 - \alpha D - k_{d,2})X_2 \tag{7}$$

$$\dot{S}_1 = D(S_{1,in} - S_1) - k_1\mu_1 X_1 + \frac{\mathbf{COD}}{\mathbf{VS}}\mathbf{k}_{\mathbf{h}}\mathbf{X}_{\mathbf{h}} \tag{8}$$

$$\dot{S}_2 = D(S_{2,in} - S_2) + k_2\mu_1 X_1 - k_3\mu_2 X_2 \tag{9}$$

$$\dot{Z} = D(Z_{in} - Z) + \mathbf{N}_{\mathbf{xh}}\mathbf{k}_{\mathbf{h}}\mathbf{X}_{\mathbf{h}} - - N_x((\mu_1 - k_{d,1})X_1 + (\mu_2 - k_{d,2})X_2) \tag{10}$$

$$\dot{C} = D(C_{in} - C) + k_4\mu_1 X_1 + k_5\mu_2 X_2 - qC \tag{11}$$

$$\dot{N} = D(N_{in} - N) + \mathbf{N}_{\mathbf{xh}}\mathbf{k}_{\mathbf{h}}\mathbf{X}_{\mathbf{h}} - - N_x((\mu_1 - k_{d,1})X_1 + (\mu_2 - k_{d,2})X_2) \tag{12}$$

$$y_1 = qM = k_6\mu_2 X_2 \tag{13}$$

$$y_2 = qC = kLa(CO_2 - k_H P_c) \tag{14}$$

where:

$$\mu_1 = \mu_{max,1}\frac{S_1}{S_1 + K_{s,1}} \tag{15}$$

$$\mu_2 = \mu_{max,2}\frac{S_2}{S_2 + K_{s,2} + \frac{S_2^2}{KI2}}\left(\frac{K_{inh,nh3}}{K_{inh,nh3} + S_{nh3}}\right) \tag{16}$$

$$CO_2 = C + S_2 - Z \tag{17}$$

$$\phi = CO_2 + k_H P_t + \frac{qM}{kLa} \tag{18}$$

$$Pc = \frac{\phi - \sqrt{\phi^2 - 4k_H P_t CO_2}}{2k_H} \tag{19}$$

$$pH = -\log_{10}\left(\frac{k_b CO_2}{Z - S_2}\right) \tag{20}$$

$$S_H^+ = \left(\frac{1}{10}\right)^{pH} \tag{21}$$

$$S_{nh3} = \frac{K_{a,NH_4}N}{K_{a,NH_4} + S_H^+} \tag{22}$$

$$x_{in} = \frac{\sum_{i=1}^n u_i x_{in,i}}{\sum_{i=1}^n u_i} with\ x \in \{S_1, S_2, C, N, Z\} \tag{23}$$

$$\mathbf{X}_{\mathbf{h,in}} = \frac{\sum_{i=1}^n u_i BD_{vs,i} VS_i}{\sum_{i=1}^n u_i} \tag{24}$$

In the above equations, $\mathbf{X}_{\mathbf{h}}$ (g_{VS} L^{-1}) is a vector $\in \mathbb{R}^n$ that describes the dynamics of the biodegradable particulate fraction of each co-feedstock. Both alkalinity (Z) and inorganic carbon (C) states are needed to simulate the carbon dioxide flowrate (qC) and exploit its measurement. For the small dilution ratios typical of the operations of interest, bacterial decay and inorganic N release were considered too. The $\frac{\mathbf{COD}}{\mathbf{VS}}$ ratios, the N concentration in $\mathbf{X}_{\mathbf{h}}$ ($\mathbf{N}_{\mathbf{xh}}$) and the biodegradability (BD_i) of the i^{th} feedstock volatile solids (VS_i) are fixed as constant and computed from its average chemical characterization [5]. Similarly, all x_{in} are set from feedstock characterization, but they can be time-varying. For the description of all other quantities the reader is referred to the original AM2 manuscript. Values of other terms, such as the bacteria N content (N_x) and physico-chemical equilibrium constants, were taken from the original ADM1 values (corrected with the process temperature). α (i.e. the ratio between the particulate and soluble matter retention times) was set to unity. The unit of measure of each quantity is usually the same of the original manuscript, and where it is not, it can be intended from Table I (and Table II in the main manuscipt, where parameters are reported). In the 'AM2HN' version, the N state equation is not present and the μ_2 expression is reduced to the sole original Haldane function.

Constants: $\frac{\mathbf{COD}}{\mathbf{VS}}$ = [1.27, 1.49, 1.39] g_{COD} g_{VS}^{-1}; $\mathbf{N}_{\mathbf{xh}}$ = [0.89, 1.53, 7.34] $mmol_N$ g_{VS}^{-1}; N_x = 8.58 $mmol_N$ g_{VS}^{-1}, $K_{a,nh4}$, $K_{a,co2}$ = 1.85e^{-9}, 5.08e^{-7} mmol L^{-1}; k_H = 22.7 mmol L^{-1} atm^{-1}.

III. PARAMETER ESTIMATION

A. Parameter subset selection (PSS)

the dataset The agri-AcoDM simulation was started from 25/10/2023 to discard the first period of inoculum adaptation out of the training/test datasets. In addition, the initial conditions were set equal to steady-state values from a simulation with constant initial diet feeding.

In addition to the 24 most uncertain kinetic $\theta_{\mathbf{p}}$, other 18 $\theta_{u,i}$ ($i = 1 \ldots n$) were selected to conduct the sensitivity analyses, for a total of 78 parameters ($n = 3$). Since a good organic characterization was carried out at least once for each of the i^{th} co-feedstock, no $\theta_{u,i}$ was considered for estimation given their much lower range of uncertainty with respect to $\theta_{\mathbf{p}}$'s. Indeed, the overall $BD_{vs,i}$ were computed from the asymptotic value of the BMP tests ($\text{BMP}_{\infty,i}$), whereas literature/hypothesized values were set for its fractioning among the different macromolecule contributions [22].

The global (G.)/multiple-at-a-time dynamic sensitivity analysis (Sobol's first-order indexes method) was computed exploiting the python SALib package [15]. The latter is particularly advised for non-linear models, but two main drawbacks are (i) the high computational burden to guarantee convergence and (ii) the risk of biased results when crossing certain regions of the parameter space (e.g. bifurcation given by Haldane inhibition) [18]. The computation of the sensitivity matrixes and the related Fisher Information Matrix (FIM) based on global sensitivity analysis (GFIM) was previously carried out in [1].

B. Estimation and uncertainty quantification

The Differential Evolution (DE) and Nelder-Mead (NM) algorithms embedded in the python SciPy library [20] were exploited, with default options except from the 'polish' and 'maxiter' options set at 30 and *True* respectively, for the DE. DE is a good way to tackle the issue of local minima, but its slow convergence rate restricts it to offline estima-tion problems. The parameters were always scaled to the (1,10) range to improve the performances of the minimization algorithms and to compute the *arrival cost* in the model adaptation scheme. The average value of data ($\bar{y}$) were used to normalize simulation errors of different magnitudes for the different outputs: it was preferred to the standard L1- and L2-norms as these tend to reduce the relevance of low values of $\mathbf{y}$. The w_y were set to have the same order of magnitude of MRSE% for all the outputs at the first iteration of the minimization algorithm.

Spearman's coefficient of determination (R^2) and Mean Absolute Relative Error (MARE) were used to quantify model performances, in light of the consideration in [14].

For the computation of the FIM and related 95% confidence intervals (CI) of the parameter estimates, hypothesis testing and the linear propagation of uncertainty from the estimates to the model outputs, the procedures in [8], [19], [7] were followed. For the weighing matrix in the computation of the FIM, a constant relative variance was adopted for the error model: 20% S_{ac} and S_{pro}, 2%-5% on gas measurements.

For the AM2HN/tan models, all state initial conditions at 15/12/2024 were set consistently with agri-AcoDM. In the first stage of the AM2HN/tan parameter estimation: (i) all the available correspondences between ADM1- and AM2-like variables were exploited [13], as reported in Table **??**, and (ii) the hightly excinting input was obtained varying, within reasonable operating values, both D (0.015-0.045 d^{-1}) and the diet composition (the fraction of each co-feedstock in the diet changed up to 40%).

IV. Full-scale applicability: the dataset

The aim of the experiments on the pilot-scale reactors was to reproduce the operating conditions of the full-scale reactor from which the pilot-scale reactors (12 L working volume reactor, T = 42°C, initial D = 0.031 d^{-1} and organic loading rate (OLR) = 2.84 g_{COD} d^{-1} L^{-1}) were inoculated on the 25/10/2023. Similar loading rate and diet composition (30, 47 and 23% on OLR basis for maize silage, cow slurry and tomato sauce respectively) were maintained at first. Afterward, the dataset entails the control experiment described in [6]. The *online* control scheme consisted of equipping the methane flow (Q_{ch4}) controller with an override from the biogas composition (i.e., the CO_2/CH_4 ratio) controller, to guarantee less inhibition-prone transients. Indeed, the latter is a fair and cheap online proxy of total VFA concentration [6]. Due to equipment limitations: i) tomato sauce was fed with the available peristaltic pumps as control action; ii) maize silage and cattle slurry were fed manually and impulsively 3 times a week; iii) the gas composition ($\%CH_4, \%CO_2, \%O_2, H_2S$ in ppm) was analyzed every 3 liters of biogas accumulated in gasbags, whereas the biogas flowrate and *pH* were logged every minute. For this reason, both models were extended with mass balance equations to reproduce the delay in the measurement of gas composition given by its accumulation in the gasbag. In general, the experimental conditions were a bit more challenging than actual full-scale applications. Tomato sauce was selected to conduct the automatic control experiment as a pumpable substitute for the OLR portion of other TS-rich agro-products and industrial by-products.

In addition to gas data, it is reasonable to assume that some spot (i.e. offline) VFA and TAN measurements would be available at full-scale, especially during transients of diet change: *offline* digestate measurements (VS, soluble COD, TAN, VFA composition) were taken manually 3 times a week, before reactor manual feeding, and in few other occasions. Butirate and valerate were not exploited as frequently found below the detection limit.

The peculiarity of the entire dataset (15/12/2023-10/04/2024) is that it entails also long reactor free-response (i.e. not-fed reactor/discharge/under-loaded) periods (22/12/2023-08/01/2024; 18-25/03/2024) and impulse responses (in particular, a maize impulse on 26/03/2024), as well as the transient dynamic between two different diet compositions (diet change; with re-spect to the diet applied till 05/02/2024, +53% maize silage feed rate, -27% cow slurry and an average +89% tomato sauce (managed by the controller), were fed af-ter 23/02/2024). Although in practice it is not realistic to have such an informative dataset from continuous full-scale operation, it remains a good general guideline to select a period with different operating conditions for training/calibration, because it tends to prevent overfit-ting itself. Test/validation set was also selected as meaningful to verify for the process 'understanding' of the model. Anyway, if the available dataset is actually 'poorer', a modeler could compensate with more batch tests as described below, with limited additional ex-penditure for plant owners.

The feedstock characterization as in [5] was performed two times throughout the whole period, with the exception of the measurements of the slurry TS/VS, that was repeated for each stored tank, and the characterization of tomato sauce, that was mainly taken from the commercial label and completed with one VFA measurement.

The BMP tests that were run before the beginning of the pilot-plant operation were inoculated from the reference full-scale reactor, and their data was used to derive the information about the feedstock's biodegra-dability (BD_{VS}) and a first guess for $\mathbf{k_h}$. The BMP of cow slurry only was repeated on the 10/04/2024, inoculated from the pilot-scale reactor, and exploited along with the others activity tests in the estimation of the agri-AcoDM $\boldsymbol{\theta_p}$.

All BMP tests were designed according to standardized protocols [16]; the start pH value and the hourly Q_{ch4} were manually and automatically recorded respectively. The AMPTS equipment (from BPC Instruments AB, Sweden) was used to record the hourly Q_{ch4} of all the batch tests.

V. Results and discussions

In the following text and Tables, the vectorization within square brackets refers to the list of co-feedstocks and its ordering is '[maize silage, cow slurry, tomato sauce]'.

The BMP tests conducted before the pilot experimentation resulted in BMP_∞= [408, 214, 386] NmL_{CH_4} g_{VS}^{-1} (BD_{VS}= [82, 45, 79]%). The results of the cow slurry BMP test

TABLE I

VARIABLE CORRESPONDENCE AND CONVERSION BETWEEN AGRI-ACODM AND AM2HN/TAN MODELS

	Variable	agri-AcoDM	Conversion*
State variables	$\sum_{i=1}^{n} X_{h,i}$ $[g_{VS}\ L^{-1}]$*	$X_{ch,r}, X_{ch,m}, X_{ch,s}, X_{pr,r}, X_{pr,s}, X_{li}$ $[g_{COD}\ L^{-1}]$	$\frac{X_{ch,r}+X_{ch,m}+X_{ch,s}}{COD/VS_{ch}} + \frac{X_{pr,r}+X_{pr,s}}{COD/VS_{pr}} + \frac{X_{li}}{COD/VS_{li}}$
	$S_1 [g_{COD}\ L^{-1}]$	S_{su}, S_{aa}, S_{fa} $[g_{COD}\ L^{-1}]$	$S_{su} + S_{aa} + S_{fa}$
	S_2 [mM]	S_{va}, S_{bu}, S_{pro}, S_{ac} $[g_{COD}\ L^{-1}]$	$1000\left(\frac{S_{va}}{208} + \frac{S_{bu}}{160} + \frac{S_{pro}}{112} + \frac{S_{ac}}{64}\right)$
	X_1 $[g_{VS}\ L^{-1}]$	X_{su}, X_{aa}, X_{fa} $[g_{COD}\ L^{-1}]$	$\frac{X_{su}+X_{aa}+X_{fa}}{COD/VS_{bm}}$
	X_2 $[g_{VS}\ L^{-1}]$	$X_{ac}, X_{h2}, X_{c4}, X_{pro}$ $[g_{COD}\ L^{-1}]$	$\frac{X_{ac}+X_{h2}+X_{c4}+X_{pro}}{COD/VS_{bm}}$
	Z [mM]	$S_{va}, S_{bu}, S_{pro}, S_{ac}$ $[g_{COD}\ L^{-1}]$, $S_{hco3}, S_{co3}, S_{nh3}, S_{oh}, S_h, S_{hpo4}, S_{po4}, S_{h3po4}$ [M]	$1000\ (\frac{S_{va}}{208} + \frac{S_{bu}}{160} + \frac{S_{pro}}{112} + \frac{S_{ac}}{64} + S_{hco3} + S_{nh3} + 2S_{co3} + S_{oh} - S_h + S_{hpo4} + 2S_{po4} - S_{h3po4})$
	C [mM]	S_{ic} [M]	$1000\ S_{ic}$
	N [mM]	S_{in} [M]	$1000\ S_{in}$
Calculation variables	CO_2 [mM]	S_{CO2} [M]	$1000\ S_{co2}$
	B [mM]	S_{hco3} [M]	$1000\ S_{hco3}$
	pH [–]	pH [–]	-
	qC $[mM\ d^{-1}]$	$\rho_{T,10}$ $[M\ d^{-1}]$	$1000\ \rho_{T,10}$
	qM $[mM\ d^{-1}]$	$\rho_{T,9}$ $[M\ d^{-1}]$	$1000\ \rho_{T,9}$
	Pc [atm]	$P_{gas,co2}$ [bar]	$\frac{P_{gas,co2}}{P_{gas,co2}+P_{gas,ch4}}$
	S_{nh3} [mM]	S_{nh3} [M]	$1000\ S_{nh3}$

*n = number of co-feedstocks in the diet **COD/VS [ch, pr, li, bm] = [1.18, 1.53, 2.83, 1.41] $g_{COD}\ g_{VS}^{-1}$

TABLE II

SUMMARY OF THE MEASUREMENTS INCLUDED IN THE DATASET AND THEIR DIFFERENT EXPLOITATION

Name	UdM	Type	Experiment type	Ny	Date	Exploitation
$\mathbf{Q_{ch4}}$	$L\ h^{-1}$	*online*	pilot	840	15.12.2023-19.01.2024	training
$\mathbf{Q_{ch4}}$	$L\ h^{-1}$	*online*	pilot	1968	19.01.2024-10.04.2024	test
$\mathbf{Q_{ch4}}$	$L\ h^{-1}$	*online*	batch activity$S_{ac,0}$* = 3 g L^{-1}	408	10.04.2024-27.04.2024	training
$\mathbf{Q_{ch4}}$	$L\ h^{-1}$	*online*	batch activity $S_{ac,0}$ = 10 g L^{-1}	408	10.04.2024-27.04.2024	training
$\mathbf{Q_{ch4}}$	$L\ h^{-1}$	*online*	batch activity $S_{pro,0}$ = 3 g L^{-1}	408	10.04.2024-27.04.2024	training
$\mathbf{Q_{ch4}}$	$L\ h^{-1}$	*online*	batch activity $S_{pro,0}$ = 6 g L^{-1}	408	10.04.2024-27.04.2024	training
$\mathbf{Q_{ch4}}$	$L\ h^{-1}$	*online*	batch BMP cow slurry	408	10.04.2024-27.04.2024	training
$\mathbf{Q_{co2}}$	$L\ h^{-1}$	*online*	pilot	840	15.12.2023-19.01.2024	training
$\mathbf{Q_{co2}}$	$L\ h^{-1}$	*online*	pilot	1'968	19.01.2024-10.04.2024	test
$\mathbf{S_{ac}}$	$mg\ L^{-1}$	*offline*	pilot	10	15.12.2023-19.01.2024	training
$\mathbf{S_{ac}}$	$mg\ L^{-1}$	*offline*	pilot	62	19.01.2024-10.04.2024	test
$\mathbf{S_{pro}}$	$mg\ L^{-1}$	*offline*	pilot	10	15.12.2023-19.01.2024	training
$\mathbf{S_{pro}}$	$mg\ L^{-1}$	*offline*	pilot	62	19.01.2024-10.04.2024	test
TAN	$mg_N\ L^{-1}$	*offline*	pilot	18	15.12.2023-10.04.2024	manual tuning
pH	-	*online*	pilot	2'808	15.12.2023-10.04.2024	test

*The initial concentration of the specific substrate is reported for batch activity tests.

were confirmed by the same test conducted at the end of the pilot operation and that was considered as part of the dataset (∼10% deviation).

A. Agri-AcoDM calibration and performance

Following the results of the PSS, the too rare measurements of TAN resulted in very poor informative content: this implied that no parameters with a strong impact on TAN were selected for estimation. However, even tough TAN was not affected by strong variations during the whole time period, a good fit of such data is important for the overall estimation task, especially when considering operations that are likely to be prone to S_{nh3} active inhibition. For this reason, before automatic parameter estimation, the TAN data were exploited for a preliminary manual tuning of the hydrolysis constant ($k_{hyd,xprr}$) only (from 1 down to 0.7 d^{-1}), to obtain a MARE on TAN below 20%. The introduction of salt precipitation and acid-base equilibrium non-ideality revealed to be particularly beneficial for matching the pH at the beginning of batch activity tests, and this is of particular interest to give a reliable reconstruction of the initial lag-phase typical of these tests. Note that, since precipitation is in place, caution has to be paid in the exploitation of inorganic compound measurements (e.g. TAN and partial/total alkalinity), because the handling of samples (e.g. dilution) highly impacts the results. It is thus advised that modelers do take into account the dilution ratios of these measurements in their models before computing simulation errors and that the model tuning guidelines in [2] are followed.

B. AM2HN/tan calibration and performance

Results from the first-stage estimation (i.e. on syn-thetic data) of the reduced-order models were very satisfactory: R^2 was always higher than 0.6, with X_1 reaching 0.97. AM2HN was better over X_1, whereas AM2HNtan was better over X_2 and S_2.

The 'inner' stoichiometric parameters (k_1, k_2, k_3) were not included in the PSS scheme to be possibly 'refined' on real data: the modeling and estimation approach that was followed indeed trusts agri-AcoDM for the estimation of parameters that strongly affect the trajectories of states for which no validation data are available (namely X_1, X_2 and S_1).

C. 'Online' model adaptation: AM2HN parameter update

In [10], two possibilities of model adaptation schemes are considered if model mismatches increases *online* during operation: (i) when the system input varies substantially with respect to the previous working conditions, a parallel data-driven model of the residuals is added in a parallel hybrid modeling scheme, whereas (ii) if the system input did not change, model parameters are re-estimated. In this work, the AM2HN/tan models were built to have a sufficient prediction capability for different input conditions, so that the '(ii)' case was considered.

In light of model-based control schemes, the aforementioned approach was tested in view of the asynchronous separation between state and parameter *online* estimations, that can be beneficial with respect to a unique exogenous state observer (e.g. EKF with state vector augmented with practically identifiable parameters such as $\mu_{max,2}$); otherwise, the risk is to follow too much the measurement noise or to 'absorb' the measurement information on the parameter instead of on the state estimates, if some collinearity between states and parameters is present on the output measurements (e.g. $\mu_{max,2}$ and X_2 on qM).

The *arrival cost* itself plays a role similar to the one of the PSS scheme in reducing the risk of overfit and it is effective in *online* schemes where the need for *exploration* is generally lower than the one for *exploitation* of the parameter values learned on old data and/or *offline*. Compared to [21], a more naive yet intuitive weighting ($\mathbf{W}_\theta$) based on the local relative-relative sensitivity ranking indexes computed on the k^{th} time window was applied to the *arrival cost*. An affine linear mapping (Eq.25) was selected with user-set intercept and slope (a,b) to move from the local relative-relative sensitivity ranking indexes of each j^{th} element of the parameter vector $\theta_\mathbf{p}$ to its *arrival cost*'s weight. Indeed, $\mathbf{W}_\theta$ is a square diagonal matrix with the j^{th} element defined as:

$$w_{j,j} = a + b r_k(\theta_{p,j}) \tag{25}$$

with r_k being the index of the δ^{msqr} metric of the parameter $\theta_{p,j}$ in the $SI_{rr,k}$ ranking, evaluated on the k^{th} time window. In this work, a and b were set to 10 and 30 respectively to obtain similar order of magnitudes for the different terms of the cost function at the beginning of the optimization on the first time window, if the initial guess of $\theta^*_{\mathbf{p},\mathbf{1}}$ is set randomly from the Gaussian distributions of the $\theta^*_\mathbf{p}$ learned *offline* (*expected values* $\pm$ standard deviations). The exploitation of Sobol's sensitivity computed within narrow bounds from the reference/previous parameter values is expected to lead to more robust results, but it was not selected for *online* purposes due to its high computational burden. The most useful information that can be extracted by this approach is a mapping F between the parameters and the states, outputs and/or inputs ($\theta_\mathbf{p}(t) = F(\mathbf{x}(t), \mathbf{y}(t), \mathbf{u}(t))$), but much more data would be required to obtain a statistically meaningful map.

In the prospective of a nonlinear model predictive controller (NMPC), although 'prior' model trajectories (with $\theta^*_{\mathbf{p},\mathbf{k-1}}$) presented a worse fitting over some windows, the update could be beneficial to prevent overloading the system (e.g. on the 4^{th} and 5^{th} windows, where S_2 accumulation is foreseen); with this in mind, slack variables will be needed to guarantee recursive feasibility if an upper bound on S_2 is enforced as state constraint.

VI. Conclusions

Testing PSS methods that better deal with the true (i.e. not-Gaussian) uncertainty probability distribution (e.g. profile likelihood) will be carried out on the same dataset for a fair comparison [17].

The ultimate goal will be to design and validate robust NMPC schemes able to improve the control performances with respect to the results obtained testing the classical control scheme described in [6], over similar diet change: indeed, good tracking performance and controlled Q_{ch4} improvement was achieved after the period of diet change, but the tracking was poor on the transient because of the inability of the classical control to foresee process inhibition.

References

[1] Juan C. Acosta-Pavas, Carlos E. Robles-Rodríguez, Jérôme Morchain, Claire Dumas, Arnaud Cockx, and César A. Aceves-Lara. Dynamic modeling of biological methanation for different reactor configurations: An extension of the anaerobic digestion model no. 1. *Fuel*, 344:128106, 2023.

[2] Damien Batstone and Xavier Flores-Alsina. *Generalised Physicochemical Model (PCM) for Wastewater Processes*. IWA Publishing, 11 2022.

[3] D.J. Batstone, J. Keller, I. Angelidaki, S.V. Kalyuzhnyi, S.G. Pavlostathis, A. Rozzi, W.T.M. Sanders, H. a Siegrist, and V.A. Vavilin. The iwa anaerobic digestion model no 1 (ADM1). *Water Science and technology*, 45(10):65–73, 2002.

[4] K Bułkowska, I Białobrzewski, E Klimiuk, and T Pokoj. Kinetic parameters of volatile fatty acids uptake in the ADM1 as key factors for modeling co-digestion of silages with pig manure, thin stillage and glycerine phase. *Renewable Energy*, 126:163–176, 2018.

[5] D. Carecci, A. Catenacci, S. Rossi, F. Casagli, G. Ferretti, A. Leva, and E. Ficara. A plant-wide modelling framework to describe microalgae growth on liquid digestate in agro-zootechnical biomethane plants. *Chemical Engineering Journal*, 485:149981, 2024.

[6] Davide Carecci, Arianna Catenacci, Elena Ficara, Gianni Ferretti, and Alberto Leva. Modelling, optimisation and control of full-scale co-digestion biomethane plants. In *Proceedings of the 63rd IEEE Conference on Decision and Control*. IEEE, in press, 2024.

[7] Francesca Casagli, Gaetano Zuccaro, Olivier Bernard, Jean-Philippe Steyer, and Elena Ficara. Alba: A comprehensive growth model to optimize algae-bacteria wastewater treatment in raceway ponds. *Water Research*, 190:116734, 2021.

[8] A. Catenacci, D. Carecci, A. Leva, A. Guerreschi, G. Ferretti, and E. Ficara. Towards maximization of parameters identifiability: Development of the calopt tool and its application to the anaerobic digestion model. *Chemical Engineering Journal*, 499:155743, 2024.

[9] Audrey Chai, Yee-Shian Wong, Soon-An Ong, Nabilah Aminah Lutpi, Sung-Ting Sam, Wei-Chin Kee, and Hock-Hoo Ng. Haldane-andrews substrate inhibition kinetics for pilot scale thermophilic anaerobic degradation of sugarcane vinasse. *Bioresource technology*, 336:125319, 2021.

[10] Laura Boca de Giuli, Alessio La Bella, Giuseppe De Nicolao, and Riccardo Scattolini. Lifelong learning for monitoring and adaptation of data-based dynamical models: A statistical process control approach. In *2024 European Control Conference (ECC)*, pages 947–952, 2024.

[11] Xavier Flores-Alsina, Kimberly Solon, Christian Kazadi Mbamba, Stephan Tait, Krist V. Gernaey, Ulf Jeppsson, and Damien J. Batstone. Modelling phosphorus (P), sulfur (S) and iron (Fe) interactions for dynamic simulations of anaerobic digestion processes. *Water Research*, 95:370–382, 2016.

[12] P. Fritzson, A. Pop, K. Abdelhak, A. Asghar, B. Bachmann, W. Braun, D. Bouskela, R. Braun, L. Rogovchenko-Buffoni, F. Casella, R. Castro, R. Franke, D. Fritzson, M. Gebremedhin, A. Heuermann, B. Lie, A. Mengist, L. Mikelsons, K. Moudgalya, and P. Östlund. The OpenModelica Integrated Environment for Modeling, Simulation, and Model-Based Development. *Modeling, Identification and Control: A Norwegian Research Bulletin*, 41:241–295, 10 2020.

[13] S. Hassam, E. Ficara, A. Leva, and J. Harmand. A generic and systematic procedure to derive a simplified model from the anaerobic digestion model no. 1 (adm1). *Biochemical Engineering Journal*, 99:193–203, 2015.

[14] H. Hauduc, M.B. Neumann, D. Muschalla, V. Gamerith, S. Gillot, and P.A. Vanrolleghem. Efficiency criteria for environmental model quality assessment: A review and its application to wastewater treatment. *Environmental Modelling Software*, 68:196–204, 2015.

[15] Jon Herman and Will Usher. Salib: An open-source python library for sensitivity analysis. *Journal of Open Source Software*, 2(9):97, 2017.

[16] Christof Holliger, Madalena Alves, Diana Andrade, Irini Angelidaki, Sergi Astals, Urs Baier, Claire Bougrier, Pierre Buffière, Marta Carballa, Vinnie de Wilde, Florian Ebertseder, Belén Fernández, Elena Ficara, Ioannis Fotidis, Jean-Claude Frigon, Hélène Fruteau de Laclos, Dara S. M. Ghasimi, Gabrielle Hack, Mathias Hartel, Joern Heerenklage, Ilona Sarvari Horvath, Pavel Jenicek, Konrad Koch, Judith Krautwald, Javier Lizasoain, Jing Liu, Lona Mosberger, Mihaela Nistor, Hans Oechsner, João Vítor Oliveira, Mark Paterson, André Pauss, Sébastien Pommier, Isabella Porqueddu, Francisco Raposo, Thierry Ribeiro, Florian Rüsch Pfund, Sten Strömberg, Michel Torrijos, Miriam van Eekert, Jules van Lier, Harald Wedwitschka, and Isabella Wierinck. Towards a standardization of biomethane potential tests. *Water Science and Technology*, 74(11):2515–2522, 09 2016.

[17] Kevin A. P. McLean and Kim B. McAuley. Mathematical modelling of chemical processes—obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. *The Canadian Journal of Chemical Engineering*, 90(2):351–366, 2012.

[18] George Qian and Adam Mahdi. Sensitivity analysis methods in the biomedical sciences. *Mathematical Biosciences*, 323:108306, 2020.

[19] Alberto Saccardo, Beatriz Felices-Rando, Eleonora Sforza, and Fabrizio Bezzo. Droop model identification via model-based design of experiments to describe microalgae nitrogen uptake in continuous photobioreactors. *Chemical Engineering Journal*, 468:143577, 2023.

[20] Rainer Storn and Kenneth Price. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *J. of Global Optimization*, 11(4):341–359, dec 1997.

[21] Andrea Tuveri, Caroline S.M. Nakama, José Matias, Haakon Eng Holck, Johannes Jäschke, Lars Imsland, and Nadav Bar. A regularized moving horizon estimator for combined state and parameter estimation in a bioprocess experimental application. *Computers Chemical Engineering*, 172:108183, 2023.

[22] Sören Weinrich, Eric Mauky, Thomas Schmidt, Christian Krebs, Jan Liebetrau, and Michael Nelles. Systematic simplification of the anaerobic digestion model no. 1 (adm1) – laboratory experiments and model application. *Bioresource Technology*, 333:125104, 2021.