Development of statistical predictive models for estimating the methane yield of Italian municipal sludges from chemical composition: a preliminary study
A. Catenacci, A. Azzellino and F. Malpei

ABSTRACT

The biochemical methane potential (BMP) of primary and biological sludge varies in a wide range, mostly depending on location, sewer characteristics, wastewater treatment plant design and operating conditions. BMP tests are useful to verify the performance of a full scale digester, but they are not yet a common procedure in the operation of most Italian facilities because of cost and test duration. Changes in the composition of sewage sludge can lead to a high variation of biogas production. Aimed at developing BMP predictive models based on low cost and fast analyses, this study investigated the chemical composition of 20 sludge samples by means of principal component and multiple linear regression analyses. Three preliminary predictive models were developed based on soluble organic nitrogen, volatile solids, carbohydrates, proteins, lipids and an operational parameter, the sludge retention time: the explained variance and the standard errors of prediction of BMP are in the range 77–81% and 21–34 NmLCH4·gVS/C0, respectively. Models were evaluated on five additional samples: errors ranged 2–15% for four samples and about 54% for one sample, collected from a peculiar facility. Further data and variables describing the operation mode of the waterline would certainly improve the reliability and robustness of the models.

Key words | anaerobic digestion, biochemical methane potential, multiple linear regression, principal component analysis, sewage sludge

INTRODUCTION

Within the circular economy concept, wastewater treatment plants (WWTPs) have to become more efficient in the management of energy, resource and waste and, at the same time, they should remain effective in providing purification of wastewater against conventional as well as emerging contaminants (Silvestre et al. 2015).

Efficient anaerobic digestion (AD) of sewage sludge is one of the most effective ways to both reduce significantly the volume of waste to be disposed, and to increase energy self-sufficiency in WWTPs (Chae & Kang 2013). Indeed, municipal WWTPs account for about 3–4% of the global electrical demand and 5% of global greenhouse gas emission (Ngheim et al. 2017); considering only US WWTPs, approximately 6.5 million metric tons (dry weight) of sewage sludge are produced annually. However, if captured and managed efficiently, sludge could yield substantial energy in the form of biogas, potentially turning a WWTP into a net energy producer rather than a consumer (Shen et al. 2015).

Compared to Europe, in Italy the integration of AD in WWTPs is not yet a common practice and digesters are often far from being operated under optimal conditions. According to the Italian Gestore Servizi Energetici (2016), AD is available in only 77 large WWTPs, producing 44.2 MW of electricity, a quite limited fraction (less than 3% in 2015, Eurobserver) of the renewable energy produced from biogas by all sources in Italy.

In this respect, the understanding of the relationships existing between the waterline operation, the quantity and quality of sludge produced, and the expected biogas yield, could help to maximize the energy recovery and to identify poorly operating digesters.

The biochemical methane potential (BMP) test is useful, as a benchmark, to verify the performance of a
full scale digester operation. However, it is not yet a common procedure in most Italian WWTPs, mainly because of high cost and long test duration. Besides, it is ineffective to support a preliminary evaluation of the methane production from sewage sludges that could be achieved at regional scale. Indeed, the BMP of sludges varies in a wide range, mostly depending on location, sewer characteristics, WWTP design and operating conditions and, more generally, on the type of sludge. Elbeshbishy et al. (2012) found a BMP range of 221–283 NL CH4·kgVS⁻¹ for the primary sludge collected in Carson (California), but higher values were also reported: a BMP of 415 NL CH4·kgVS⁻¹ was found by Zhang et al. (2016) for the sludge collected from the primary clarifier in a WWTP in Australia (Brisbane). Gavala et al. (2003) measured a BMP of 475 NL CH4·kgVS⁻¹ for the primary sludge collected in Lyngby (Denmark), and a value of 186 NL CH4·kgVS⁻¹ for the biological sludge. A BMP of 165 NL CH4·kgVS⁻¹ was found by Nielfa et al. (2015) for the biological sludge taken from a plant in Spain. A study on waste activated sludge from six WWTPs located in France reports a range of 206–427 NL CH4·kgVS⁻¹ (Mottet et al. 2010).

As a consequence, there’s a need to develop fast and reliable analytical methods to determine the methane production potential in a shorter time, and many alternative approaches are reported in literature for its estimation: respirometric tests and correlations between aerobic and anaerobic biodegradability indexes (Ponsáfi et al. 2008); linear (Lesteur et al. 2011) and non-linear (Godin et al. 2015) prediction models based on non-destructive analytical methods of the organic matter like the near infrared (NIR) spectrum; tools (e.g. the Envital®kit) based on fluorescence measurement (Bellaton et al. 2016). Also, statistical correlations and regression models between BMP and chemical composition of the substrates are of interest because they are readily applicable and carried out using current BMP equipment. Literature reports studies on different substrates: lignocellulosic biomasses (Xu et al. 2014), grassland plants (Dandikas et al. 2015), livestock manures (Kafle & Chen 2016), a mixture of agro-industrial wastes and the organic fraction of municipal solid waste (Schievano et al. 2008). Lately, some authors demonstrated that the BMP test duration could be reduced if predicting the final gas production at an earlier stage (Strömberg et al. 2015; Da Silva et al. 2018). However, few studies have focused on a statistical correlation between WWTPs sludges composition and anaerobic conversion (Mottet et al. 2010; Appels et al. 2011).

In the frame of an on-going research to update available data on sewer sludges, as to both the anaerobic biodegradability and the composition of the organic matter, this preliminary study aims at identifying the main chemical compounds affecting the methane yield of municipal sludges. Typical parameters known as being potentially influential for methane production (e.g. total and volatile solids, chemical oxygen demand (COD), volatile fatty acids, total Kjeldahl nitrogen (TKN)) were measured. Phosphorus and different forms of nitrogen were also measured since they are essential for microbial metabolism; also, organic nitrogen is related to the presence of proteins, and high ammonium concentrations can inhibit the process (even if not expected for sewer sludges). Besides those, a different approach was used with reference to carbohydrates, proteins and lipids: a distinction among fractions having different bioavailability for bacteria to be degraded was made. As for carbohydrates and proteins, a sequential extraction was applied in order to quantify the most easily and rapidly available forms for microbial degradation: the soluble and the extracellular polymeric substances (EPS) fractions. For lipids, analytical determinations were performed on the liquid and solid fractions of sludge after centrifugation. Since the degradability of biological sludge can be affected by the operating conditions of the oxidation tank in the wastewater treatment line, the sludge retention time (SRT) was also included. Based on a dataset of 20 sludge samples, all characterized for 27 composition variables and digested on laboratory scale to experimentally determine the BMP, the main constituents affecting the BMP of the sludge were identified by means of a principal component analysis (PCA). Building on these information, the stepwise multiple linear regression (MLR) was used: preliminary predictive models were developed in order to provide WWTPs operators with a simple tool to be used to verify the performance of full scale digesters, then reducing cost and time required for the analytical BMP determination.

METHODS

Sludge samples and characterization

A total of 25 sludge samples were collected from 15 WWTPs located in Lombardy region, being representative of about 3.1 million Population Equivalent (PE), one-fourth of the total PE served in Lombardy (around 11 million PE in 2015). In particular, 12 samples were taken from primary and 13 from secondary clarifiers: in the following ‘S1’ and ‘S2’ are used for primary and biological sludge, respectively.
The main characteristics of the WWTPs sampled are reported in Table 1.

The samples were stored at 4 °C prior to the analyses and BMP tests. A total of 17 composition variables were directly measured on samples. Ammoniacal nitrogen (NH₄⁺), nitric nitrogen (NO₃⁻), total soluble nitrogen (Ns), total soluble phosphorus (Ps), volatile fatty acids (VFA), soluble COD (sCOD), soluble carbohydrates (CHs) and proteins (PTs) were determined on the liquid fraction after 0.45 μm filtration. Total solids (TS), volatile solids (VS), TKN, total COD (COD) were measured on fresh matter (FM); the lipid content was determined after centrifugation (3,000 rpm, 15 minutes) in the supernatant (LPs) and in the cake (LPc); the elemental nitrogen (N) was determined on the sample dried at 105 °C. EPS were extracted in order to be analysed for their carbohydrates (CHe) and proteins (PTe) contents.

With the exception of TS and VS (g·kgFM⁻¹), all the other 15 compounds were referred and expressed per unit of VS (mg·gVS⁻¹). A total of 10 more variables were derived as ratio between two variables (VS/TS, VFA/sCOD, sCOD/COD), or defined as algebraic sum of two or more variables, referred to VS, as shown in the following (Equations (1)–(7)):

Soluble organic nitrogen: $N_{org,s} = Ns - NH_4^+ - NO_3^-$ (1)
Total organic nitrogen: $N_{org,t} = TKN - NH_4^+$ (2)
Total carbohydrates: $CH = CHs + CHe$ (3)
Total proteins: $PT = PTs + PTe$ (4)
Total lipids: $LP = LPs + LPc$ (5)
Sum of soluble carbohydrates, proteins and of lipids in the supernatant: $CPLs = CHs + PTs + LPs$ (6)
Sum of total carbohydrates, proteins and lipids: $CPLtot = CH + PT + LP$ (7)

### EPS extraction protocol

EPS were extracted adopting the protocol proposed by Ras et al. (2008): previous to the extraction stage, the sludge samples were diluted to 1 gVS·L⁻¹ with Tris-HCl buffer (10 mmol·L⁻¹, pH = 8) and centrifuged (3,000 rpm, 15 minutes). Dilution is required for the implementation of

### Table 1 | Characteristics of WWTPs sampled

<table>
<thead>
<tr>
<th>WWTP ID</th>
<th>Primary sedimentation</th>
<th>Denitrification</th>
<th>Oxidation</th>
<th>Sludge retention time (SRT)</th>
<th>Secondary sedimentation</th>
<th>Average annual PE (2017)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a</td>
<td>X</td>
<td>X</td>
<td>25</td>
<td>b</td>
<td>96,290</td>
</tr>
<tr>
<td>B</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>15</td>
<td>X</td>
<td>27,910</td>
</tr>
<tr>
<td>C</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>30</td>
<td>X</td>
<td>105,000</td>
</tr>
<tr>
<td>D</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>12</td>
<td>X</td>
<td>143,690</td>
</tr>
<tr>
<td>E</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>8</td>
<td>X</td>
<td>243,660</td>
</tr>
<tr>
<td>F</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>15</td>
<td>X</td>
<td>187,770</td>
</tr>
<tr>
<td>G</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>15</td>
<td>X</td>
<td>37,410</td>
</tr>
<tr>
<td>H</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>23</td>
<td>X</td>
<td>66,190</td>
</tr>
<tr>
<td>I</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>15</td>
<td>X</td>
<td>15,880</td>
</tr>
<tr>
<td>L</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>18</td>
<td>X</td>
<td>37,610</td>
</tr>
<tr>
<td>M</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>25</td>
<td>X</td>
<td>164,650</td>
</tr>
<tr>
<td>N</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>23</td>
<td>X</td>
<td>135,000</td>
</tr>
<tr>
<td>O</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>5</td>
<td>X</td>
<td>650,000</td>
</tr>
<tr>
<td>P</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>16.5</td>
<td>X</td>
<td>120,000</td>
</tr>
<tr>
<td>Q</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>35</td>
<td>X</td>
<td>1,027,000</td>
</tr>
</tbody>
</table>

*aFine screening instead of primary clarifier.
*MBC (membrane bioreactor).
*Design data.
*Post-denitrification + post-aeration.
*High concentrations of heavy metals incoming the facility (chromium for F and zinc for M).
*Data from August to December 2017 and from January to July 2018.
*Existing but not in operation when sampling.
the protocol on sludge with different biomass concentrations and to avoid later analyses interferences with substances like humic acids (Zuriaga-Agusti et al. 2013). The pellet was washed twice with Tris-HCl buffer before EPS extraction and the supernatant was collected for carbohydrate and protein analyses. After that, the pellets were re-suspended (1 gVS·L⁻¹) to extract the EPS by means of a two step treatment (500 rpm, 1 hour) with 0.5% Triton X-100 prepared in Tris-HCl buffer, and then centrifuged (3,000 rpm, 15 min) in order to separate the supernatant containing the EPS, which was collected for subsequent carbohydrate and protein analyses. The soluble and EPS fractions were stored at −20 °C until being analyzed.

Analytical methods

TS and VS were determined in duplicate according to Standard Methods 2540 (APHA 2005). The pH was measured by means of a portable multi-probe meter (Hach-Lange, HQ40D). NH₂, NO₃, Ns, Ps, VFA and sCOD were measured using spectrophotometric test kits (Hach-Lange) containing macro- and micro-nutrients was dosed (OECD 511 Standard), and tap water was added to reach the desired working volume (450 mL). N₂ was flushed for five headspace volume exchanges to ensure initial anaerobic conditions; pH at the beginning of each test was in the range 7.2–7.6. Each substrate was tested in duplicate. Additionally, two bottles serving as a blank were prepared by dosing inoculum (the same amount used for testing the substrate), mineral medium and tap water. The methane yield for each bottle (BMPᵢ, expressed in NmLCH₄·gVS⁻¹) was calculated as follows (Equation (8)):

\[ BMPᵢ = \frac{(V_{CH₄,S,ᵢ} - V_{CH₄,B})}{M_{S,ᵢ}} \]

where: \( V_{CH₄,S,ᵢ} \) (NmLCH₄) is the accumulated methane volume at standard conditions (273 K, 1 atm) from the \( i \)-th bottle with the substrate; \( V_{CH₄,B} \) is the average accumulated methane volume at standard conditions from the blank samples; \( M_{S,ᵢ} \) (gVS) is the mass of substrate in terms of VS dosed in the \( i \)-th bottle. The BMP of each substrate was defined as the average of the test replicates (BMPᵢ).

Statistical analysis

Based on a dataset of 20 samples (10 primary, 10 biological sludges), a matrix was built including the BMP value and all the 27 mentioned variables for all the 20 cases. The SRT was included in the analysis for S2 samples, and set as zero for S1 samples. VFA on S2 samples were measured but not included in the statistical analysis because of low concentrations, always below the detection limit (50 mgCH₃COOH·L⁻¹), and then supposed to be negligible for the development of predictive models.

PCA was used on the created data matrix to investigate the correlations among the chemical compounds and to identify the different components of the variability present in the dataset. Based on the extracted principal components, factor loadings and correlations between variables were
used to define a selection of potential predictors that were further analysed through a stepwise MLR approach. Finally, the accuracy of the predictive models was preliminarily evaluated on five new sludge samples, by comparing the model estimates with the measured BMP values (Afifi & Clark 1996; Rencher 2002).

RESULTS AND DISCUSSION

Anaerobic degradability of sludge samples

Figure 1 reports the final BMP for both types of sludge, referred to VS and to COD, and the theoretical methane expected (ThBMP) based on COD balance (100% degradation; 6% of COD for cellular growth). BMP values ranged 176–318 NmLCH₄·g⁻¹VS (112–200 NmLCH₄·g⁻¹COD) for S1 and 58–190 NmLCH₄·g⁻¹VS (34–139 NmLCH₄·g⁻¹COD) for S2 samples. Considering the ratio between the BMP on COD basis and the ThBMP, the biodegradability of S2 (10–42%) was found to be lower than that of S1 (34–61%). Coefficients of variation (CV) of the tests were always below 10% except for samples from plants ‘F’ (15% and 14% for S1 and S2, respectively) and ‘M’ (11% for both S1 and S2), where high concentrations of chromium and zinc enter the facility, probably determining different inhibition in test replicates. Since the study of heavy metals inhibition goes beyond the purpose of this work, further BMP tests at lower S/I ratios were not run, then maintaining the same test conditions for all samples used for model calibration.

Compared to literature, on average, BMP was lower for both S1 (e.g. samples D-1, G-1 and I-1) and S2 samples (especially C-2 and G-2). With reference to S1, when comparing sludge sampled from different WWTPs located in different regions, an intrinsic variability for primary sludge should be expected, due to the composition of the raw sewage. The chemical characteristics of the influent are affected by the presence of industrial discharges as well as by the sewer extension, where partial hydrolysis of particulate matter may occur. Furthermore, in Italy, the surplus sludge from secondary settlers is often recirculated to primary settlers: this was the case for B-1, D-1, G-1, H-1 and I-1 samples. Along with high retention times in the hoppers, this operation worsens the extent of such phenomena, because of the presence of biomass which is active in promoting hydrolysis and partial degradation of the most rapidly degradable components. Regarding S2 samples, biological sludge is normally less digestible anaerobically than primary sludge, being composed mostly of cellular material and EPS instead of more easily degradable carbohydrates and fats that are typically found in primary sludge.

Figure 1 | Average BMP for sludge samples S1 and S2 (labelled with numbers ‘1’ and ‘2’, respectively, following the capital letter identifying the WWTP). Values are referred to the VS content (white bars) and to the COD (grey bars). The solid line is drawn to indicate the ThBMP. Error bars indicate standard deviations.
Biological sludge consists of complex carbohydrates and proteins, and long chain organic molecules: dead bacteria and their refractory organics constitute its non-degradable fractions. Moreover, with the need to fulfill stringent standards for nutrients besides COD and BOD, Italian WWTPs usually apply high SRT to the biomass in the activated sludge process (>10 days) in order to preserve the nitrification capability of the activated sludge. As a consequence, a partial sludge stabilization occurs during the activated sludge process, thus resulting in low yield. With reference to Table 1, this might be the case of samples A-2, C-2, H-2, M-2 (SRT = 25, 30, 23, 25 days, respectively; degradability = 32%, 10%, 14%, 50%, respectively); however, if considering the SRT of samples G-2 and I-2 (15 days) and their degradability (14% and 20%, respectively), it was not possible to extend the correlation between the SRT and the BMP.

Characterization of sludge samples

As the stepwise regression analysis is based on the assumption of normality, the distribution of the dataset was assessed and all parameters resulted in a symmetrical distribution of normality, the distribution of the dataset was activated. I-2 (15 days) and their degradability (14% and 20%, respectively; degradability 32%, 10%, 14%, 50%, respectively); however, if considering the SRT of samples G-2 and M-2 samples, even when referred to the VS content, being a consequence of the waterline treatment processes. Compared to primary sludge, where the hydrolysis of complex substrates in hoppers brings about a partial solubilization of the particulate organic matter (in the present case the ratio sCOD/COD is averagely 5%), the VS of biological sludge is mainly composed of particulate matter consisting of active and dead bacteria cells, and then resulting in low sCOD/COD ratios (<0.5% in this study). Moreover, a higher COD/VS ratio indicates organic matter that requires more oxygen to decompose than those with lower ratios. A mean of 1.488 gCOD/gVS was found for S2 samples, which is typical for activated sludge and indicates the prevailing presence of proteins (1.42 gCOD/gVS for cellular protein, whose formula can be represented as C3H2NO2). A higher mean value of 1.534 gCOD/gVSS was found for S1 samples, then suggesting a more heterogeneous composition of primary sludge including lipids, proteins and carbohydrates.

On average, values were higher for S1 compared to S2, with the exception of few variables: among them, different forms of nitrogen (e.g. TKN, NO3, Norg,t, N), CHe and PTe. Nitrogen is related to the content of proteins: the ratio Norg,t/Norg was found to be higher for S1 (1.23%) compared to S2 (0.35%), then suggesting that proteins in S2 are associated to particulate active/dead biomass. Figure S1 (Supplementary Material, available online) shows the amounts of carbohydrates and proteins in the soluble and in the EPS fractions measured in the 20 sludge samples: they were measured because they were considered the most available for bacteria to be degraded. As expected, PTe and CHe were higher than PTs and CHs, especially for S2 samples. The order of magnitude of CHe and PTe was similar to what was observed by Ras et al. (2008) (i.e. proteins in the soluble and in the extracted fractions: 44 and 182 mgBSA/gVSS, respectively), and by Zuriaga-Agusti et al. (2015) (i.e. carbohydrates and proteins in the EPS fraction around 10 mgGlu/gVSS and ranging 55–100 mgBSA/gVSS, respectively). Moreover, CHe and PTe in S2 samples were, on average, 6.2 and 7.4 times the corresponding soluble fractions; conversely, the ratios CHe/CHs and PTe/PTs for S1 were lower and equal to 2.8 and 2.6, respectively. As expected, the sums of the two fractions for CH and PT were well below the total amounts of carbohydrates and proteins found by Mottet et al. (2010). Conversely, in the present study, total lipids were measured as the sum of LPs and LPC, and the values found for S2 samples were in the range found by Mottet et al. (2010) (10–90 mg gVSS).

Preliminary bivariate linear regression analyses between BMP and some of the parameters listed in Table 2 are shown in Figure 2 for both sludge types. Not well defined relationships can be observed, except for some variables: positive correlations for S1 with LP, PTa, CPLtot, CPL (R2 = 0.414–0.410–0.414–0.580, respectively), negative correlations for S1 with TKN (R2 = 0.458), and for S2 with PT and CPLtot (R2 = 0.433–0.470, respectively). However, no strong correlations were found, then a bivariate linear regression is not sufficient to predict the BMP.

PCA results

PCA was used to preliminarily evaluate the potential explanatory variables based on their reciprocal correlations and their correlations with BMP.
In the PCA carried out including all cases (S1 + S2), all the parameters were used with the exception of VFA and VFA/sCOD (since VFA was not measured on S2 samples) for a total of 27 variables. The first three principal components (PC) extracted explained respectively 44.5%, 15.1% and 8.8% of the variance, accounting overall for a 68% of the total variance. Factor loadings are reported in Table S2 (Supplementary Material, available online). PC1 was loaded by 15 variables, all with high factor loadings (absolute value above 0.7), including also BMP. The loading plot (Figure 3(a)), which provides a map of the variables loadings on the first two PCs, and the factor scores plot (Figure 3(b)), which provides a map of the cases in the principal component biplot are also shown. With reference to Figure 3(a), each vector from the middle of the outer circle to dots depicts an individual variable: the length of a variable vector represents its weight on each PC and the correlation between any two variables is determined by the cosine value of the angle between the two vectors. Variables are positively or negatively correlated when two vectors are
pointing toward similar or opposite directions, respectively. Two vectors with an angle close to 90° are highly independent. PCA reveals a high level of redundancy between the 27 variables and some clusters of highly correlated variables are clearly identified. The high positive correlation of BMP with lipids confirmed their strong contribution to the methane yield. COD and soluble organics (CHs, PTs, PTLs) were found to have a weak correlation with BMP. This is reasonable for COD, since it doesn’t distinguish between the organic matter which can be easily or scarcely anaerobically degraded. Conversely, the result is unexpected for the soluble organic matter, since it is widely thought to be more accessible for microbial degradation. Finally, the inverse correlation of the two parameters derived from one another (TKN and Norg,t) might be explained considering that nitrogen is mainly associated to less degradable complex proteins inside active/dead bacteria cells, and that is particularly true for the S2 sludge.

The score plot in Figure 3(b), which shows the factor scores of the PC2 versus the scores of the PC1, clearly identifies two groups of variables: components scores for each sample are standardized so that the zero sets the average scores for the two PCs. In this case, S1 samples are always above the average score of PC1, while the S2 samples are always below. Since the first two components account for the 60% of variance explained, groupings of data on the

Figure 2 | Linear correlations between the BMP and main chemical parameters: white and black dots identify S1 and S2 samples, respectively. R squared for S1 and S2 samples are reported in italic and underlined, respectively.
plot may indicate two or more separate distributions in the data. Also, it’s observed that S1 samples are more dispersed around PC2 than S2 ones: this is reasonable considering the higher variability of primary sludge, which reflects the specific variability of the influent wastewater. In consideration of this, and with the aim of improving the predictive capability of the model, the dataset was split into two parts, and S1 and S2 samples were further analyzed separately.

Results of the PCA on S1 samples are shown in the loading plot of Figure 3(c): in this case, VFA and VFA/sCOD were included while SRT was removed from the list of variables used (28 variables). The first three extracted PCs explained 69% of the total variance of the
dataset. Table S5 (Supplementary Material, available online) reports the factor loadings of the PCs. The BMP was negatively correlated with TKN, Norg,t and N, and, yet again, slightly positively correlated with COD; contrarily to the previous case, in this case the resulting BMP was strongly positively correlated with the soluble organics: this could be ascribed to the nature of S1, which contains higher amounts of soluble matter compared to S2, which were instead included in the previous analysis.

With reference to S2 cases, the same 27 variables identified for S1 + S2 samples were used in the PCA. Factor loadings of the PCs are reported in Table S4 (Supplementary Material, available online). The 66% of the total variance of the dataset is explained by the first three PCs. Figure 3(d) shows the loading plot: the absolute BMP factor loading is higher (-0.827) on PC2 compared to PC1 (0.162). In this case the BMP was inversely correlated with SRT, which is expected. Also, the BMP is negatively correlated to PT, PTe and CPLtot: such a result might be related to the reduced accessibility of the organic matter when included in EPS. Indeed, pretreatments applied on biological sludge in order to improve the biogas yield are intended to make accessible degradable proteins and carbohydrates contained in the sludge.

Model development

With reference to S1 samples, based on PCA results, variables in the top left/bottom right quadrants of Figure 3(c), VS/TS and Norg,t were not included in the MLR because of no/low correlation with the BMP, as well as COD (factor loadings on PC1 and PC2 < 0.1); 13 possible predictors were then considered: CPLtot, CPLs, PTs, CHs, LPs, Norg,s, VFA/sCOD, TKN, N, LPc, LP, CH, PT. Results are shown in Table 3 (Model S1): CPLs and the ratio VFA/sCOD were found to be the strongest predictors, with positive correlations with the BMP, then including the main soluble components of the organic matter that can be rapidly and easily degraded.

As for S2 samples, given the high absolute BMP factor loading on PC2, seven variables with absolute factor loadings above 0.6 on PC2 were used as possible predictors in the MLR (SRT, COD, TKN, Norg,t, PT, PTe, CPLtot). Results of the stepwise MLR are shown in Table 3: SRT and PT were selected as predictors of Model S2, with a negative correlation with BMP. The inverse relationship between BMP and SRT is predictable since the application of high SRT in the activated sludge process determines the aerobic degradation of the most biodegradable part, so that an increase in the SRT causes a decrease in the specific gas production. PT includes proteins in both the soluble and the EPS fractions: its inclusion in the model with a high negative standardized coefficient (-0.447) cannot be clearly explained, and a wider dataset and measurement interval would help to confirm and explain this result.

Including all the cases, regardless of whether they are primary or biological sludge, nine variables were selected and then used as possible predictors in the MLR, considering their correlation with the BMP along PC1 (i.e. having factor loadings on PC1 > 0.65 and loadings on PC2 < 0.2): TS, VS, sCOD, Norg,s, LP and LPc are positively correlated to the BMP, while SRT, TKN and Norg,t are inversely correlated. Results of the MLR are shown in Table 3 (Model S1 + S2). The regression selected, LP, as the strongest predictor and Norg,s, which is an indirect measure of soluble proteins, are both positively correlated to BMP.

Model accuracy was good for all the models developed, with the variance explained ($R^2$) ranging 77–81% and the Adjusted $R^2$ from 0.727 to 0.758: higher values

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Coefficients and goodness-of-fit statistics of MLR models for prediction of BMP of sewer sludge</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model (No. of cases)</strong></td>
<td><strong>Predictor</strong></td>
</tr>
<tr>
<td>S1 + S2 (20)</td>
<td>(Constant)</td>
</tr>
<tr>
<td></td>
<td>LP</td>
</tr>
<tr>
<td></td>
<td>Norg,s</td>
</tr>
<tr>
<td>S1 (10)</td>
<td>(Constant)</td>
</tr>
<tr>
<td></td>
<td>CPLs</td>
</tr>
<tr>
<td></td>
<td>VFA/sCOD</td>
</tr>
<tr>
<td>S2 (10)</td>
<td>(Constant)</td>
</tr>
<tr>
<td></td>
<td>SRT</td>
</tr>
<tr>
<td></td>
<td>PT</td>
</tr>
</tbody>
</table>
Pertain to Model S1. The standard error of prediction (SEP) is lower for both S1 and S2 Models (26 and 21 NmL CH4·gVS/C0, respectively), compared to Model S1 + S2 (34 NmL CH4·gVS/C0). All the parameters selected contributed statistically significantly to the models (significance is always less than 0.05, with the exception of the constant for Model S1).

Preliminary models evaluation

The three model accuracy was also evaluated by comparing BMP measurements and model predictions on two S1 and three S2 samples. Results are shown in Table 4. With reference to S1 samples, the BMP values estimated with both models (Model S1 + S2 and Model S1) are very close to the measured values, with errors always below or near 10%.

Concerning the S2 samples, the error on BMP estimates is always below or near 15%, except for case Q-2. Unlike Model S1 + S2, Model S2 significantly underestimates the measured BMP of the Q-2 case. As shown in Table 1, sample Q-2 was taken from a plant without a primary clarifier despite its great potentiality in terms of PE served; Model S2, on the other hand, was built on data measured on biological sludges from plants with both primary and biological treatments (with the exception of A-2). Moreover, its strongest predictor is the SRT, an operational parameter not describing the actual composition of the sludge: the higher measured BMP value compared to the estimate might be related to this peculiarity that, instead, can be described with Model S1 + S2. Finally, the different BMP test condition (S/I ratio) used for samples N-1, N-2 and O-2, compared to samples used for model calibration, is likely to be of negligible influence on BMP estimates; however, extended research is needed to confirm this result.

CONCLUSIONS

BMP of sewer sludge can be predicted from chemical characteristics and operating parameters of the wastewater treatment line using MLR models. As expected, parameters like soluble phosphorus, elemental nitrogen, nitric nitrogen, and ammonium were found to have scarce correlation with BMP. Also, total and soluble COD, VS and TS were found to be weak predictors of the BMP, since they don’t distinguish between anaerobically degradable/non-degradable matter. The SRT, the soluble organic nitrogen and the composition of the organic fraction of sludge in terms of VFA, carbohydrates, proteins and lipids were found to be important parameters for all the models. These parameters are relatively quick to obtain, making the statistical approach used in the present work a fast, cheap and simple tool to predict the BMP supporting the monitoring of wastewater full scale digester. It has to be stressed that a detailed characterization of the organic matter is needed by separating the contribution of fractions having different microbial degradability and accessibility; an in-depth study at different levels of organic matter accessibility would certainly improve the prediction ability of models. A wider dataset would be needed to refine the models and improve their reliability. Moreover, parameters defining the operation mode of the waterline and describing the operating conditions should be collected and included in the statistical analysis, in order to refine the models and to support the management of plants for the selection of the best operating modes and conditions to maximize the energy recovery.

Table 4 | External evaluation of BMP models on two S1 and three S2 samples

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Unit</th>
<th>Sludge samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N-1</td>
</tr>
<tr>
<td>Predictors: measured</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRT</td>
<td>D</td>
<td>0</td>
</tr>
<tr>
<td>LP</td>
<td>mg·gVS⁻¹</td>
<td>91.2</td>
</tr>
<tr>
<td>Norg,s</td>
<td>mg·gVS⁻¹</td>
<td>2.16</td>
</tr>
<tr>
<td>VFA/sCOD</td>
<td>–</td>
<td>0.534</td>
</tr>
<tr>
<td>CPLs</td>
<td>mg·gVS⁻¹</td>
<td>89.3</td>
</tr>
<tr>
<td>PT</td>
<td>mg·gVS⁻¹</td>
<td>–</td>
</tr>
<tr>
<td>BMP measured</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMP (S1 + S2)</td>
<td>NmLCH₄·gVS⁻¹</td>
<td>335</td>
</tr>
<tr>
<td>BMP (S1)</td>
<td>NmLCH₄·gVS⁻¹</td>
<td>306 (8.8%)</td>
</tr>
<tr>
<td>BMP (S2)</td>
<td>NmLCH₄·gVS⁻¹</td>
<td>308 (8.1%)</td>
</tr>
<tr>
<td>BMP measured</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMP (S1 + S2)</td>
<td>NmLCH₄·gVS⁻¹</td>
<td>142 (2.1%)</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

The authors wish to thank Gruppo CAP, Padania Acque S.p.A., Brianza Acque S.r.l., Azienda Servizi Integrati Lambro S.p.A. and MilanoDepur S.p.A. for the helpful cooperation in sampling sludge, and Sabrina Carpino, Nadia Margariti, Silvia Battello and Alessio Cupaioli for their valuable support in the analyses.

REFERENCES


Zhang, T., Wang, Q., Ye, L. & Yuan, Z. 2016 Effect of free nitrous acid pre-treatment on primary sludge


First received 9 November 2018; accepted in revised form 4 February 2019. Available online 19 February 2019