An evaluation model for the deep dewatering potential of municipal sludge based on PLS regression

P R Yu\textsuperscript{1}, Y Luo\textsuperscript{1}, F Lin\textsuperscript{1}, J G Li\textsuperscript{1}, X L Zhu\textsuperscript{1}, Z B Zhang\textsuperscript{1} and M R Liu\textsuperscript{1,2}

\textsuperscript{1}State Key Laboratory of Pulp and Paper Engineering, School of Light Industry and Engineering, South China University of Technology, 510640, China

E-mail: lmr@scut.edu.cn

Abstract. For constructing a new convenient and reliable system to evaluate the dewatering performance of sludge, the relationship between the water content of dewatered municipal sludge and its physicochemical properties was comprehensively investigated. Then, an evaluation and prediction model was established based on partial least squares regression (PLS), which could accurately predict the solid content of municipal sludge treated by high-pressure deep dewatering. Furthermore, the main factors affecting the sludge dewaterability were explored by the means of principal component analysis (PCA), and it was concluded that first principal component (PC1), containing most composition of extracellular polymeric substances (EPS), accounted for 53.1\% of the X variation and 62.9\% of information Y variation. Both protein in soluble EPS (PNS) and polysaccharide in a tightly bound EPS (PST), with the variable importance in the projection (VIP) values more than 1, were regarded as the most decisive factors to the deep dewatering performance of municipal sludge. Additionally, the validation results demonstrate the accuracy and practicability of the obtained model, specifically fit for a conditioning system for cationic polyacrylamide (CPAM) and lysozyme either separately or jointly.

1. Introduction

With the increasing sewage disposal ratio and technology, the output of municipal sewage sludge an inevitable by-product sharply rises. The water content of sludge is as high as 95\%-99.5\% and remains at 75\%-85\% after mechanical dewatering, which not only occupies a huge volume but also brings about many difficulties in transport and subsequent disposal. In this case, deep dewatering after conditioning has exhibited its comparatively high dewatering efficiency at a lower operational cost [1,2].

Since a long time, both the SRF [3,4] and capillary suction time (CST) [5] which focus on measuring sludge filtration performance, have been used to judge the dewaterability. Bound water (BW) [6] and dry solids (DS) content [7] is also used to evaluate the effect of sludge dewatering. As the major component of bio-floc, extracellular polymeric substances (EPS) account for 50\%-90\% of the total organic matter, vital for dewatering and settling performance of sludge [8]. The diversity of evaluating indicators is not convenient for the comparison of sludge dewatering performance, even existing some discrepancies for that conditioned or dewatered using different methods. Moreover, a simple correlation analysis only could obtain some one-sided or isolated information between single indexes as the real sludge dewatering performance is determined by many interacting factors. Recently, multivariate statistical analysis, including partial least squares (PLS) [9] and parallel factor analysis (PARAFAC) [10], was applied to investigate the effect of various factors on sludge
dewaterability, and a multiple linear model [11] between CST and certain fractionized compounds was even built. However, these all overly focused on the relation between EPS and CST, without taking the others into account.

In view of the ambiguous relation between the physicochemical properties and dewaterability of sludge, the objective is to establish a new model to predict and evaluate its deep dewatering potential, which could comprehensively understand the influence of multiple factor interactions.

2. Experimental

2.1. Materials

2.1.1. The source and characteristics of municipal sludge. The original secondary sludge was collected from a local municipal wastewater treatment plant in Guangzhou, China, where the wastewater was treated by modified anaerobic-anoxic-oxic (A/A/O).

2.1.2. Chemical reagents. CPAM was purchased from the Snf Floerger with molecular weight of 10 to 12 million and different cationic degree. Lysozyme with Enzyme activity of 2000U/mg was provided by Amresco. Chitosan hydrochloride with degree of deacetylation \( \geq 90\% \) was produced by Cool Chemistry. Both the reaction buffer and micrococcus lysodeikticus cell suspension were purchased from Sigma.

2.2. Methods

2.2.1. Sludge conditioning process. The sludge sample was treated with separate or combined conditioning. For separate conditioning, 100 mL suspension of sludge was prepared in a 250 mL beaker with the addition of conditioner, and then was stirred rapidly for 30 s at 200 rpm before slow mixing for 2 min at 50 rpm using a coagulation agitator. For the combined conditioning, lysozyme conditioning was carried out according to the above procedure before CPAM conditioning. The basic parameters of the conditioner and the specific conditioning methods used are listed in Table 1.

| Sample label | conditioner | Dose of conditioner (mg/g DS of Chemical flocculants orx10^6U/g of lysozyme) |
|--------------|-------------|---------------------------------------------------------------------------------|
| 1            | -           |                                                                                 |
| 2-7          | CPAM with cationic degree of 20% (FO4196) | 1, 1.5, 2, 2.5, 3 and 4                                                         |
| 8-13         | CAPM with cationic degree of 60% (FO 4650) | 1, 1.5, 2, 2.5, 3 and 4                                                         |
| 14-18        | CPAM with cationic degree of 80% (FO 4680) | 1, 2, 2.5, 3 and 4                                                              |
| 19-24        | chitosan hydrochloride                    | 2.5, 5, 10, 15, 20 and 30                                                       |
| 25-27        | lysozyme                                  | 1.6, 2.4 and 3.2                                                                |
| 28-32        | lysozyme and CPAM with cationic degree of 60% | 2.4/1, 2.4/2, 2.4/2.5, 2.4/3 and 2.4/4                                         |

2.2.2. The determination of the sludge water content and the its sludge characters. 300 mL sludge was centrifuged at 1800 rpm for 5 min. Then, the remaining precipitant was transferred to a lab-scale deep-dewatering device developed by ourselves, and the solid-liquid separation was realized by the static pressure, of 5.4 MPa ~ 6 MPa for 5 min. Then the water content of the dewatered sludge cake (W) was examined using a HX204 halogen moisture analyser.

The pH value of the sludge samples was measured with a pH5-3C pH metre and the zeta potential was detected using a zetasizer nano-ZS90. The particle sizes of the sludge were measured by a mastersizer 3000. CST and SRF were evaluated with a 304M CST instrument and a custom SRF determinator, respectively. The turbidity degree (TD) and transmittance (TM) of the supernatant were measured by a Hach 2100N turbidity metre and a TU-1080 UV-vis spectrophotometer separately. In
addition, sludge volume index (SVI) was tested by measuring the height variation in 30 min of the sludge-liquid interface in a 100 mL cylinder, and computed using the dry solid from 100 mL sludge.

2.2.3. EPS extraction and analysis. The EPS extraction from the sludge is carried out according to the modified heat extraction method [12]. The soluble EPS (S-EPS), loosely bound EPS (LB-EPS) and tightly bound EPS (TB-EPS) were separated using a series of methods that comprised centrifugation, ultrasonication and thermal extraction. The protein (PN) concentration in the extracted EPS was measured by the Coomassie brilliant blue method with bovine serum albumin (BSA) as a standard [13], and the polysaccharide (PS) concentration was determined using the anthrone method with a glucose standard [14].

2.2.4. Statistical analysis.

• Pearson correlation analysis

  The Minitab 17 software was used to analyse the degree of correlation between the deep dewatering performance of sludge and its physicochemical properties, and P value signifies the “significance degree” of the linear correlation degree.

• PLS regression

  PLS regression was conducted in combination with the Minitab 17 software to establish the evaluation model. The optimal number of PLS principal components was determined by leave-one-out cross validation, aiming at the minimum predicted residual error sum of squares (PRESS). To evaluate the model’s predictive ability, the root mean square of the prediction (RTMSEP) [15] and mean relative error (MRE) were calculated according to equations (1) and (2), respectively, where n denotes the number of sludge samples, and y as well as ȳ signify the actual and predicted values of the water content, respectively. In addition, the determination coefficient (R²) [16] as shown describes the proportion of the variation accounting for in the total variation.

\[
RMSEP = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}}
\]

\[
MRE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|
\]

\[
R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}
\]

The concept of VIP was also introduced to analyse the interpreting ability of the independent variables (xⱼ) for the dependent variable (y) as shown in equation (4).

\[
VIP_j = \frac{p}{\sum_{h=1}^{m} Rd(y; t_j) w_{hj}^2}
\]

3. Results and discussion

3.1. The characteristics of the municipal sludge before and after conditioning

The fundamental characteristics of the raw sludge are listed in table 2. And the characteristic of the
dewatered sludge conditioned with CPAM, chitosan hydrochloride or lysozyme were analysed in the following sections. As shown in figure 1, the dependent variables, such as W and NCST, were greatly affected by the chemical conditioning method with similar change tendencies when referring to chemical conditioning (C1, C2, C3 and C4). Additionally, the lysozyme treatment provides a green and efficient method to destroy the cellular structure of the activity sludge, which benefits deep dewatering [17]. Here the PN in the soluble EPS (PNS) was increased rapidly after lysozyme conditioning (C5 and C6), whereas those corresponding to chemical conditioning remained almost constant. All the results manifested that the W value was affected by a number of many-sided factors.

**Table 2.** The characteristics of the municipal sludge used in this study.

| Parameter            | Value               |
|----------------------|---------------------|
| W (%)                | 72.4±0.8            |
| CST (s)              | 61.2±2.1            |
| SRF×1012/ (m·kg-1)   | 7.67±0.23           |
| SVI (mL/g DS)        | 33.1±0.4            |
| TD (NTU)             | 7.49±0.06           |
| D₅₀ [μm]             | 40.9±0.1            |
| Zeta potential       | -16.2±1.1           |
| pH                   | 6.60±0.20           |
| S-EPS PN (mg/g DS)   | 0.982±0.131         |
| S-EPS PS (mg/g DS)   | 1.01±0.05           |
| LB-EPS PN (mg/g DS)  | 0.956±0.140         |
| LB-EPS PS (mg/g DS)  | 1.30±0.12           |
| TB-EPS PN (mg/g DS)  | 1.36±0.21           |
| TB-EPS PS (mg/g DS)  | 6.04±0.55           |

**Figure 1.** Water content (a), NCST values (b) and of PNS concentrations (c) for all the sludge samples.
3.2. Correlation between W and the sludge physicochemical properties

| W    | NCST | SRF | SVI | TM | TD | Dv[50] |
|------|------|-----|-----|----|----|--------|
| W    | 0.226| 0.187| -0.0180| -0.0570| 0.293| 0.165 |
|      | 0.669b| 0.569b| 0.0630| -0.227| 0.229| -0.221| -0.0160| |
|      | 0.251| 0.474| 0.464| -0.316| 0.242| |
| NCST | 0.226| —   | **0.853**b| **0.547**a| **-0.666**c| **0.502**b| **-0.543**b| **-0.497**b| **-0.838**c|
|      | 0.668a| 0.824a| 0.434b| -0.598a| 0.507b| |
|      | 0.251| 0.952a| 0.908a| -0.885a| 0.713b| |
| SRF  | 0.187| **0.853**b| —   | **0.595**a| **-0.602**a| **0.662**a| **-0.562**a|
|      | 0.569b| 0.824a| 0.470b| -0.491b| 0.717a| -0.519a|
|      | 0.474| 0.952a| 0.953a| -0.878a| 0.778b| -0.823b|
| SVI  | -0.0180| **0.547**a| —   | **-0.669**a| **0.644**a| **-0.554**a|
|      | 0.0630| 0.434b| 0.470b| —   | -0.566a| 0.603a| -0.551a|
|      | 0.464| 0.908b| 0.953a| -0.871a| 0.808b| -0.821b|
| TM   | -0.0570| -0.666c| **-0.602**c| **-0.669**c| —   | **-0.735**c| **0.728**b| **0.835**c|
|      | -0.227| -0.598a| -0.491b| -0.566a| —   | -0.693a| 0.643|
|      | -0.316| -0.885a| -0.878a| -0.871a| —   | -0.899a| |
| TD   | 0.293| **0.502**b| **0.662**c| **0.644**c| **-0.735**c| —   | **-0.495**c|
|      | 0.229| 0.507b| 0.717a| 0.603a| -0.693c| —   | -0.676c|
|      | 0.242| 0.713b| 0.778b| 0.808b| -0.899a| —   | -0.669|
| Dv[50]| 0.165| **-0.543**c| **-0.562**c| **-0.554**c| **0.728**b| **-0.495**c| —   |
|      | -0.221| 0.497b| 0.519a| 0.551a| **-0.835**a| **-0.676**a| —   |
|      | -0.0160| -0.838c| -0.823b| 0.821b| 0.643| -0.669| |

| PNS  | PNL  | PNT  | PSS  | PSL  | PST  |
|------|------|------|------|------|------|
| PNS  | —    | **0.786**b| **0.325**| **0.474**| **0.750**b| **0.869**b|
|      | 0.0300| 0.539b| 0.546b| 0.869b| |
|      | 0.325| -0.208| 0.644a| 0.518b| 0.166|
|      | 0.474| -0.268| -0.166| 0.147| 0.384|
| PNL  | 0.786b| —    | **-0.129**| **0.660**b| **0.301**b| **0.472**b|
|      | 0.325| -0.649a| -0.0560| -0.124| -0.473b|
|      | 0.474| 0.650| 0.722b| 0.151| -0.541|
| PNT  | **0.0300**| **-0.129**| —    | **0.413**b| **-0.0270**| **0.150**|
|      | -0.208| -0.649a| 0.302| 0.0630| 0.654a|
|      | -0.268| 0.650| 0.936a| -0.324| -0.868a|
| PSS  | **0.539**b| **0.660**b| -0.0560| **0.413**b| —    | **0.134**b| **0.256**|
|      | 0.644a| 0.722b| 0.302| 0.611a| 0.715a|
|      | -0.166| 0.936a| -0.345| -0.902b| |
| PSL  | **0.546**b| **0.301**| **-0.0270**| **0.134**| —    | **0.618**b|
|      | 0.518b| -0.124| 0.0630| 0.611a| 0.535b|
|      | 0.147| -0.151| -0.324| -0.345| 0.318|
| PST  | **0.869**b| **0.472**b| **0.150**| **0.256**| —    | **0.618**b|
|      | 0.166| -0.473b| 0.654a| 0.715a| 0.535b|
|      | 0.384| -0.541| -0.868a| -0.902b| 0.318|

*indicates P<0.01.
**indicates P<0.05.

Three groups of correlation coefficients corresponding to samples of 1~32 (bold), 1-24 (conditioned with chemical flocculants) and 25-32 (conditioned with lysozyme) are listed from top to bottom in table 3. There was a strong positive correlation between W and NCST ($r_5 =0.668$, $p < 0.01$), accompanied by a moderate correlation between W and SRF ($r_5 =0.569$, $p < 0.01$), when the municipal sludge was conditioned by chemical flocculants. In contrast, these correlations were not statistically significant ($p > 0.05$) when using lysozyme alone or combined with CPAM. For all the sludge samples, there was an extremely strong negative correlation between W and PNS, W and the protein in loosely bound EPS (PNL) as well as W and the polysaccharides in a tightly bound EPS (PST).
separately. In addition, multi-collinearity correlations existed among different independent variables, such as NCST and SRF ($r_p = 0.853$, $p < 0.01$), so PLS regression was chosen to analyse the data and establish the model in view of the small sample capacity and amounts of independent variables.

3.3. Variable selection

Excessive independent variables would lead to the mathematical models too complex and inconvenient. Therefore, the selection of independent variables followed the principle of ‘essential and as few as possible’ so as to facilitate the application of a prediction model. Distinct from the general variable selection process, the number of independent variables was minimized to ensure the acceptable regression effect and model stability.

The process of variable selection was realized by introducing the concept of backward elimination in stepwise regression. Then, a model was established based on PLS regression, consisting all variables with the expression of $W_{pre} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \ (p=1,2,\ldots,12)$ . The regression coefficient, constant term and the VIP value of variables are listed in table 4. Afterwards, the variable with the minimum VIP was eliminated, and the remaining dependent variables were occupied to establish a new model to determine the values of $R^2$ and PRESS. The steps above were then repeated.

| Independent variables | Regression coefficients $\beta$ | Standardized coefficients | VIP  |
|-----------------------|---------------------------------|---------------------------|------|
| Constant term         | 74.8                            | 0                         | -    |
| NCST                  | 0.165                           | 0.586                     | 1.02 |
| SRF$\times 10^{12}$   | -0.374                          | -0.189                    | 0.762|
| SVI                   | -0.197                          | -0.207                    | 0.118|
| TM                    | -0.00580                        | -0.0185                   | 0.416|
| TD                    | 0.105                           | 0.0778                    | 0.477|
| Dv[50]                | 0.0112                          | 0.139                     | 0.408|
| PNS                   | -0.625                          | -0.628                    | 1.67 |
| PNL                   | -1.78                           | -0.386                    | 1.31 |
| PNT                   | -1.24                           | -0.257                    | 0.171|
| PSS                   | 4.78                            | 0.666                     | 1.13 |
| PSL                   | 0.0575                          | 0.00836                   | 1.07 |
| PST                   | -0.686                          | -0.424                    | 1.68 |

As presented in figure 2, there were eventually 7 dependent variables left during the screening process extending down to line f, including NCST, SRF, PNS, PNL, PSS, PSL and PST. When the number of principal components was 4, the minimum value of PRESS of 83.8 was obtained and the value of $R^2$ was 0.846, which was in accordance with the basic modelling requirement. If the process of backward elimination continued until line g or h, the value of $R^2$ would reduce to 0.836 or 0.751, respectively.

3.4. The establishment of a prediction model for deep dewatering potential of municipal sludge

In PLS regression, if there is an apparent linear relation between the first principal component $t_1$ from independent variable $X$ and $u_1$ from dependent variable $Y$, it is reasonable to draw a linear relation between $Y$ and $X$. The results exhibited in figure 3 certainly applied it due to the strong positive correlation ($r_p =0.793$, $p<0.01$) between $t_1$ and $u_1$.

In this study, 5 sludge samples were randomly selected for use as the test data and the others were used as training data. Then, the PLS prediction model for municipal sludge deep dewaterability was established as $W_{pre} = b_0 + b_1 NCST + b_2 SRF + b_3 PNS + b_4 PNL + b_5 PSS + b_6 PSL + b_7 PST$ . All the constant terms and regression coefficients are listed in table 5.
Figure 2. Model select during the process backward elimination.

Figure 3. The relationship between $t_1$ and $u_1$.

Table 5. The regression of the independent variables and constant terms of the model.

| Independent variables | Unit          | Regression coefficients $\beta$ | Standardized coefficients |
|-----------------------|---------------|---------------------------------|---------------------------|
| Constant term         | -             | 70.3                            | 0                         |
| NCST                  | s/g DS        | 0.124                           | 0.439                     |
| SRF×10^(-12)          | m·kg⁻¹        | -0.318                          | -0.160                    |
| PNS                   | mg/g DS       | -0.604                          | -0.607                    |
| PNL                   | mg/g DS       | -0.869                          | -0.189                    |
| PSS                   | mg/g DS       | 4.28                            | 0.597                     |
| PSL                   | mg/g DS       | 0.436                           | 0.0634                    |
| PST                   | mg/g DS       | -0.986                          | -0.611                    |

As table 5 shows, RMSEP and MRE increase by 44.2% and 44.8%, respectively, after variable selection, which causes a decrease in the prediction accuracy. In combination with the analysis of
residual error, it was concluded that the prediction error was due to sludge sample 24.

The response diagram for the model in figure 4 shows that the sample points were all distributed near the diagonal line with the exception of those conditioned by chitosan hydrochloride (samples 22, 23 and 24). After further statistical analysis of the residual error, the absolute errors of $w$ for samples 20/32 and 28/32 were less than 1 and 2, respectively, which indicated the fitting and prediction accuracy of the PLS model meets the requirements for the actual application.

3.5. The determining factors for the dewatering potential of municipal sludge

PCA was conducted to determine the factors influencing sludge deep dewatering performance. As shown in figure 5, the first principal component (PC1) explained 53.1% of the X variation and 62.9% of information Y variation, and the second principal component (PC2) similarly explained 28.9% and 11.9%, respectively. These two together covered most of the X and Y variation.

With respect to PC1, the absolute value of the eigenvector projection corresponding to the organic compounds in the EPS was rather large. In particular, the PNS and PST had relatively higher...
eigenvector projection values of -0.529 and -0.513, respectively. For PC2, the most positive influencing variable was NCST, and the next was SRF. As PC1 accounted for 62.9% information of the Y variation, the most important factors were PNS and PST. Further analysis revealed that the VIP values of them were 1.27 and 1.28 separately, which was significance in explaining the Y variation.

4. Conclusions
Based on understanding the relation between the physicochemical properties and sludge dewaterability, the influence of multiple factor interactions was comprehensively investigated, with the objective to establish a new model to predict and evaluate the potential of its deep dewatering performance. A prediction model was established based on PLS regression: 

\[ W = 70.30 + 0.12 \text{NCST} - 0.32 \times 10^{-12} \text{SRF} - 0.87 \text{PNS} + 4.28 \text{PSS} + 0.44 \text{PSL} - 0.99 \text{PST}. \]

The determining factors for the deep dewaterability of municipal sludge were PNS and PST, as concluded by PCA in combination with a comparison of independent variable’s VIP. The obtained model can accurately predict the solid content of municipal sludge treated by high-pressure deep dewatering and is applicable to CPAM and lysozyme conditioning either separately or jointly.

References
[1] Zhu F, Zhang Z and Jiang H 2012 J. Procedia Environ. Sci. 16 363-7
[2] Yu L, Yu Y and Jiang W 2015 J. Environ. Sci. Pollut. Res. 22 2599-609
[3] Liu J, Yang Q and Wang D 2016 J. Bioresource Technology 206 134-40
[4] Zhang J, Yang K and Wang H 2015 J. Desalin. Water Treat. 57 14424-32
[5] Hong C, Wang Z and Si Y 2016 J. Appl. Microbiol. Biotechnol. 101 1-8
[6] Wang H F, Wang H J and Hu H 2017 J. RSC Adv. 7 30274-82
[7] Dai Y, Huang S and Liang J 2017 J. Chem. Eng. J. 321 123-38
[8] Wilen B M, Jin B and Lant P 2003 J. Water Res. 37 2127-39
[9] Lü F, Zhou Q and Wu D 2015 J. Chem. Eng. J. 262 932-8
[10] Yu G H, He P J and Shao L M 2010 J. Water Res. 44 797-806
[11] Xiao K, Chen Y and Jiang X 2017 J. Chemosphere. 170 233-41
[12] Cao B, Zhang W, Wang Q, 2016 J. Water Res. 105 615-24
[13] Bradford M 1976 J. Anal. Biochem. 72 248–54
[14] Raunkjær K, Hvitved-Jacobsen T and Nielsen P H 1994 J. Water Res. 28 251-62
[15] Al-Harrasi A, Rehman N U and Mabood F 2017 J. Spectrochim. Acta, Part A. 184 277-85
[16] Sampaio P S, Soares A and Castanho A 2017 J. Food Chem. 242 196-204
[17] Chen Z, Zhang W and Wang D 2016 J. Water Res. 103 170-81