HZ University of Applied Sciences

Modelling Phosphorus Recovery from Urban Wastewater

Research report

Version 1.0

Maria van Schaik
As in-company supervisor

Bram Verkruysse
As first examinee

Tamarindo Metekohij
6-5-2020

Table of Contents

List of figures

List of tables

Abstract

Calcium phosphate and struvite crystallization has been widely studied for phosphate removal and recovery from urban wastewater. In this study, calcium phosphate and struvite crystallization were carried out in a fluidized-bed reactor. Experimental values of the P concentration ranged from 117-1000 mg P/l in feed water and 37-457 mg P/l in the bottom of the reactor. Black-box approach to model the two processes was conducted using multiple linear regression. Pearson correlation analysis describes the correlations of experimental data generated for all parameters. With regard of inflow, chemicals dosing, pH, related ion concentrations, and molar ratio (Ca/P and NH_4/P) were selected as independent variables. They are found to be significant through Pearson's correlation analysis and knowledge of crystallization process. The output variable comprises P crystallization fraction. Consequently, the information used to generate model equations. Both models, calcium phosphate and struvite acquired R^2 higher than 0.7 and low Std. Error (<10) allowing good prediction models. Experimental data and generated predicted values are close to each other. In addition, if applied in a realistic setting, the models show over-estimated values, which need to be considered in the economic analysis.

Abbreviations

ACP = amorphous calcium phosphate

DCPD = dicalcium phosphate dihydrate

OCP = octacalcium phosphate

TCP = tricalcium phosphate

HAP = hydroxyapatite

MAP = magnesium ammonium phosphate

IAP = Ion activity product

RSA = Response Surface Analysis

α = ion activity

β = supersaturation

Ksp = solubility product

1. Introduction

1.1. Introduction on nutrient recovery

There has been a change in current societal production systems driven by limits in resource availability. The focus has shifted from residues treatment, such as wastewater treatment, toward resource recovery (Puyol et al., 2017). For example, there is a great potential to recover nutrients, such as nitrogen and phosphorus from human waste and transform them into renewable resources with a consistent supply. Human urine contributes 80% of the total nitrogen content and 40–50% of the total phosphate contribution to urban wastewater (Wilsenach & van Loosdrecht, 2006). Generally, phosphorus in wastewater is in accessible and usable forms, such as phosphate (PO_4). Phosphorus (P) is important for maintaining all forms of life (Kok et al., 2018). Major processes such as photosynthesis, transmission of energy, cell division, development and fruit and crop quality rely on the availability of this nutrient. Although phosphorous may be replenished over geological time by the fluxes of the global biogeochemical cycle, its major source, phosphate rock, is a non-renewable resource. The latter because of its faster usage rate relative to its creation. Eventually, to stop the failure of agriculture, P will need to be recycled on large scales.

The NEREUS Interreg2Seas project engages in regional challenges by stimulating the development of a green economy and the re-use of resources. More specifically, the project aims to increase the reuse of resources from the city's wastewater (*NEREUS Project Interreg 2 Seas*, 2018). In an effort to speed up the reuse of resources, the project performs pilot tests for the recovery of water, energy and nutrients (nitrogen and phosphorus) from urban wastewater. The final goal of the NEREUS project is to develop a decision support tool (DST) to facilitate the multi-criteria, including technical, economic, environmental and social evaluation of resource recovery alternatives (processes and sequences of these). For this, the processes could potentially be used for resource recovery and need to be well understood.

Mathematical models can facilitate the understanding of a process, since they enable the prediction of their behavior when process conditions change (Estrada-Flores et al., 2006). When constructing a mathematical model of a given physical process, different ways of tackling the problem can be used (Bohlin, 1991). The two most common ways are: 1) process modelling based on complete knowledge of the system (white-box models) and 2) process identification on statistical information (black-box models). For many industrial processes, some exist, but incomplete knowledge concerning the system (Sohlberg, 1998). This implies that there is a grey zone between the white-box models and the black-box models, which gives a third way of making models of engineering systems. Grey-box models apply systematic use of partial priori knowledge of the process and experimental data (Sohlberg, 1998). Based on the availability of the knowledge and data, the step by step approach can be decided.

In this thesis, black-box approach will be used to predict phosphorus recovery efficiency from urban wastewater. The approach requires significant parameters identification from Pearson correlation analysis and modelling via regression analysis. The modelled processes are calcium phosphate ($CaPO_4$) and struvite ($MgNH_4PO_4$) crystallization, tested at the NEREUS pilot plant, operated by Southern Water, in United Kingdom. The phosphorus recovery is done by using a technology developed by RHDHV (Royal Haskonig DHV), namely, The Crystalactor®.

The technology is a fluidized bed type crystallization technology for selective removal and recovery of components from water and wastewater. The Crystalactor® is a cylindrical reactor, partially filled with a suitable seed material like sand or minerals (Figure 1). The phosphate containing wastewater is pumped in an upward direction, maintaining the pellet bed in a fluidized state. By dosing a calcium or magnesium salt to the water and adjusting the pH, the solubility of calcium phosphate or struvite, also called magnesium ammonium phosphate (MAP) is exceeded and subsequently phosphate is crystallized on the seed material. The chemistry of the crystallization process is comparable to the conventional precipitation.

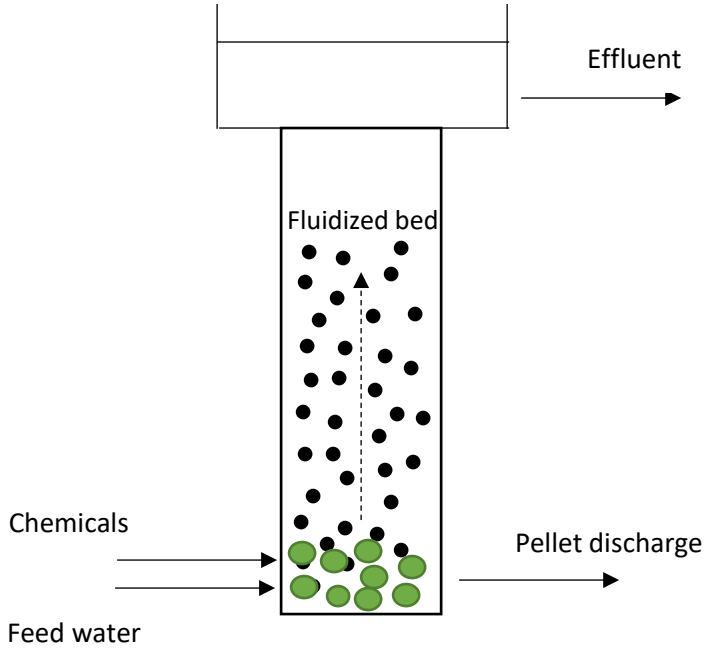


Figure 1: Representation of the Crystalactor®

The pellets grow and move towards the reactor bottom. The phosphate covered grains are discharged from the bottom of the bed and replaced intermittently by fresh sand grains. The data used for identification represent process performance as well as recovered product quality provided by Southern Water.

1.2. Research Questions

How well can struvite and calcium phosphate recovery efficiency be modelled via a black-box approach?

a. Which parameters are significant to black-box model struvite and calcium phosphate precipitation?
b. How accurate are the obtained models (R^2, Std. err.)?
c. How promising is the best model to be used under realistic setting?

1.3. Scope of the research

In this research, mathematical models on both processes of calcium phosphate ($CaPO_4$) and struvite ($MgNH_4PO_4$) crystallization are performed. Black-box modelling is applied by using regression analysis based on experimental data from pilot test carried out by Southern Water in United Kingdom.

1.4. Scientific and societal relevance of the research project

In the future, the society will face double environmental crises coming from lack of resources and excess waste, if current production and consumption patterns continue (Dobbs et al., 2011). The danger will be more severe because of climate change and growing population. For example, water scarcity becomes more problematic with nearly half the global population already living in potential water scarce areas for at least one month per year (Burek et al., 2016). According to Burek et. Al (2016) this could increase to some 4.8–5.7 billion in 2050. This will threaten the quality of the environment and the health of human beings. Transforming the management methods to solve sustainability questions is necessary.

Water is a resource transport medium besides being a resource of its own. As people use drinking water and generate wastewater, many constituents such as chemicals and energy are added. Therefore, there are many opportunities for resource recovery and closed cycles in the urban water chain, particularly in wastewater (Van Der Hoek et al., 2016). Radical changes are required in the way waste and resources streams are organized. Although there is the growing persuasion that future regulations for the treatment of wastewater must be based on the principles of a circular economy. The technology is available (Neczaj & Grosser, 2018; Puyol et al., 2017; Velenturf & Purnell, 2017), still decisionmakers hesitate to make use of this because of economic, environmental and social uncertainties.

In this case, resources recovery is part of the circular economy. Nutrients such as phosphorus are crucial for agriculture and currently have not been regained from wastewater. As a result, these resources cannot be re-used in a very useful way (Agudelo-Vera et al., 2012).

This thesis is intended to provide support needed to help understanding the processes of recovering phosphorus from wastewater. Eventually, in large scale, it can serve as decision support tool (DST) to help advancing the implementation of resources recovery. May this thesis be useful for readers and further research about similar topic or any related field.

2. Experimental Section

2.1. Theory

Nowadays phosphorus removal from wastewater has been of great concern. The recognized phosphorus removal technologies include chemical precipitation, biological phosphorus (P) removal, crystallization, tertiary filtration and ion exchange (Morse et al., 1998). Among them, crystallization processes stand out because they do not only achieve high P removal but also recover P from wastewater as useful products, including struvite and calcium phosphates. However, to predict the crystallization process is still a complex task. It is controlled by a combination of factors, such as thermodynamics of liquid-solid equilibrium, phenomena of mass transfer between solid and liquid phases, kinetics of reactions, and several physiochemical parameters (Corre et al., 2005).

A crystallization or precipitation process is composed of two simultaneous processes: nucleation and crystal growth. Nucleation involves the formation of small units that later act as centers of crystallization. There are two kinds of nucleation, homogeneous (precipitation forms spontaneously) and heterogeneous (initiated by the presence of other particles). More practically, crystallization processes are operated with or without seed. In seeded process heterogeneous nucleation dominates, while in unseeded process homogeneous substantially accounts for nucleation (Song et al., 2006). The seed for a crystallization process acts as a substrate, on which nucleation and the following crystal growth can proceed. In addition, the system must be in a supersaturated state to allow the crystallization to transpire. Supersaturation can be expressed by the saturation index (SI),

$$SI = \log\left[\frac{IAP}{Ksp}\right]$$

where IAP is the ion activity product and Ksp is the solubility product. The calculation of SI of a solution is important to understand the crystallization process.

Under proper physical and chemical environment, different kinds of calcium phosphates, such as ACP (amorphous calcium phosphate), DCPD (dicalcium phosphate dihydrate), OCP (octacalcium phosphate), TCP (tricalcium phosphate) and HAP (hydroxyapatite) may precipitate from saturated solutions (Song et al., 2002) (Table 1). Thus, the crystallization of calcium phosphate is complicated because it depends on the supersaturation and Ca/P of the various species (Montastruc et al., 2004) as well as concerns with the formation of several possible precursors and their transformation (Song et al., 2006).

Table 1: Different calcium phosphate forms

Calcium phosphate precipitate	Chemical formula	Ca/P ratio	Environmental pH[a]	Precipitate stage formation
Amorphous calcium phosphate (ACP)	$Ca_3(PO_4)2$	1.18-2.5	5-11	ACP ⟶ ANY[b]
Dicalcium phosphate dihydrate (DCPD)	$CaHPO_4{\cdot}2H_2O$	1.0	<6.5	DCPD ⟶ OCP
Tricalcium phosphate (TCP)	β-$Ca_3(PO_4)_2$	≐ 1.5	5-9	DCPD ⟶ β-TCP
Octacalcium phosphate (OCP)	$Ca_4H(PO_4)_2{\cdot}5H_2O$	1.3-1.5	6.5-8	DCPD ⟶ OCP ⟶ HAP
Hydroxyapatite (HAP)	$Ca_5(PO_4)_3OH$	1.67	6.8-9	ACP ⟶ OCP ⟶ HAP

a pH refers to the acidity/alkalinity of the environmental conditions required for precipitate formation to occur
b where ANY is any other calcium phosphate precipitate other than ACP

If phosphate is crystallized as hydroxyapatite (HAP), $Ca_5(PO_4)_3OH$ from a solution,

$$5Ca^{2+} + 3PO_4^{3-} + OH^- \Leftrightarrow Ca_5(PO_4)_3OH$$

the SI with respect to HAP is defined as

$$SI(HAP) = \log\left[\frac{IAP}{Ksp\ (HAP)}\right]$$

$$SI(HAP) = \log\left[\frac{(\alpha Ca^{2+})^5 \times (\alpha PO_4^{3-})^3 \times (\alpha OH^-)}{K_{sp}\ (HAP)}\right]$$

where α_x refers to the activity of ion x and Ksp(HAP) is the solubility product of HAP. Similarly, the SI for the precipitation of other calcium phosphate species will depend on the ions.

As for struvite, it usually precipitates as stable white orthorhombic crystals in a 1:1:1 molar ratio according to the equation below,

$$Mg^{2+} + NH_4^+ + + PO_4^{3-} + 6H_2O \rightarrow MgNH_4PO_4 + 6H_2O$$

and the saturation can be expressed as

$$SI\ (MAP) = \frac{\alpha Mg^{2+} \times\ NH_4^+ \times \alpha PO_4^{3-}}{K_{sp}(MAP)}$$

Thermodynamically, the metastable zone is defined as the critical zone of supersaturation of solution where crystallization is not governed by nucleation and thus avoids rapid and/or spontaneous precipitation. Crystallization in the metastable zone is heterogeneous (Ali & Schneider, 2006). In this pilot study using pellet reactor, the metastable zone technique is practiced. Primary nucleation was effectively minimized by the particular construction of the reactor and the choice of the appropriate degree of supersaturation. Dissolved phosphate and suspended micro-crystals from primary nucleation undesirably influence phosphorus recovery. The dissolved phosphate concentration is fixed by the solubility product, the ionic concentration of reagent and the process pH. In short, it can be said that the overall performance of crystallization is a function of supersaturation, pH and concentration of reactants.

At a given pH and overdose, the degree of supersaturation depends only upon the phosphate concentration in the wastewater. The phosphate concentration at the bottom of the reactor has to be maintained below a critical value in order to prevent primary nucleation. Moreover, the mechanical strength of the crystals decreases with increased supersaturation. In practice, it has been observed that negligible nucleation occurs at a phosphate concentration of 25-125 mg/l P. This concentration is obtained in the pellet reactor by the correct selection of the circulation flow, irrespective of the phosphate concentration in the wastewater.

2.2. Materials and Methods

Phosphorus recovery studied in this paper is based on calcium phosphate and struvite crystallization processes using fluidized bed reactor (The Crystalactor®) from urban wastewater. The unit consists of a pellet reactor with feed pump, an operational buffer tank, dosing equipment, a stripper tower and a dual media filter. Stripper tower is used to remove carbonic acid (H_2CO_3) in calcium phosphate route, thus no stripping tower would be used for struvite route. The pellet reactor is designed based on the basics presented in Table 2.

Table 2: Basic designs of pellet reactor, The Crystalactor®

pH range feed water	**6.0-8.0**
pH stripping tower	<4.0
pH range effluent reactor	7-11
Feed flow	8-24 l/h
Recirculation flow	72-56 l/h
Temperature	Ambient

The experimental data acquired from the pilot plant of Southern Water include among others feed flow, recirculation flow, P concentration, chemicals dosing flow, pH and temperature. During the pilot trial samples were taken twice every day the reactor runs. The measurements were done at the inlet, bottom and outlet of the reactor, but details are not presented in this paper.

The high [P] concentration ranged from 117-1000 mg P/l in feed water and 37-457 mg P/l at the bottom of the reactor. Trials were done by changing the process conditions for both processes as can be seen in Table 3. Variances were focused in changing the inflow (feed flow and recirculation) for calcium phosphate route and changing $MgCl_2$ dosing flow and pH for struvite route.

Table 3: Experimental conditions of controlled crystallization in Southern Water pilot plant

	Calcium phosphate	Struvite
Conditions	Feed flow 8 l/h Recirculation flow 55 l/h	Normal operations
	Feed flow 8 l/h Recirculation flow 70 l/h	Normal operations resumed (at 50% $MgCl_2$ dosing 500 ml/h)
	Feed flow 5.6 l/h Recirculation flow 70 l/h	Operation without MgCl2 dosing and pH 8 in reactor
	Feed flow 6.8 l/h Recirculation flow 70 l/h	Operation without $MgCl_2$ dosing and pH 7.5 in (sludge reduction)
	Feed flow 10 l/h Recirculation flow 68 l/h	Excess $MgCl_2$ dosing at 25 rpm (218 ml/h) to recover excess P at pH 7.5 in reactor
	Feed flow 10 l/h Recirculation flow 70 l/h	New experiment: improve P recovery at higher pH >8 in reactor (no excess $MgCl_2$ dosing)
		Slight excess $MgCl_2$ dosing at higher pH>8 in reactor (to improve P recovery)
Total data	76 data points (only 49 are valid for observation)	37 data points (only 26 are valid for observation)
Chemicals dosing	• 30% HCl • 35% $CaCl_2$ • 50% NaOH	• 30% $MgCl_2$ • 50% NaOH

2.2.1. Regression Analysis

Regression methods have been widely used for predictions and forecasting purposes (Bonen et al., 1979; Catalina et al., 2013; Pauly, 1989; Tulleken, 1993). They are a set of statistical methods used for the estimation of relationships between a dependent variable and one or more independent variables. Specifically, correlation analysis is performed to provide insight to the key factors (parameters) influencing the phosphorus recovery. For these reasons, statistical analysis was chosen to be the technique for crystallization prediction of this study. SPSS® is used to determine the appropriate

model for the system, identify significant process parameters and their interactions. The independent variables used were:

- inflow (adding feed flow and recirculation flow)
- chemical reagents dosing flow: NaOH, $CaCl_2$ for calcium phosphate route and $MgCl_2$ for struvite route
- initial and final concentrations of P, Ca, Mg and NH_4
- Ca/P and Mg/P ratio
- pH feed and effluent.

The dependent variable would be phosphorus crystallization fraction. Correlation analysis of Pearson was performed to all the parameters to identify the significance. Pearson's correlation coefficient is the statistics test that measures the statistical relationship, or association, between two continuous variables. Pearson's correlation produces a statistic that ranges from -1, indicating a perfect negative correlation, to +1, indicating a perfect positive correlation. A value of zero indicates no correlation at all. It is known as the best method of measuring the association between variables of interest because it is based on the method of covariance. The information from Pearson analysis describe the importance of the parameters to be taken into consideration.

Later, the independent variables for the models were picked based on the theory behind crystallization process as well as the analysis from Pearson. Based on that information, multiple regression models could be generated. There are several methods to accomplish this by examining the interaction between the independent variables with respect to prediction. In this study, hierarchical regression was selected. The latter recognizes that theory should drive the statistical model and that the decision of what terms enter the regression model should be determined by theoretical concerns. Hierarchical regression adds terms (parameters) to the regression model in stages. At each stage, an additional term or terms are added to the model and the change in R^2 is calculated.

The appropriateness of a model can be further assessed by the R^2, Adjusted R^2 and Std. Error. In general, the higher the R^2, the better the model fits the data. However, chasing a high R^2 value can lead to reduced precision and a lessened ability to make predictions. Other conditions to look at are the Adjusted R^2 and Std. Error. The adjusted R^2 is used to compare models with different numbers of predictors and Std. Error is wished to be minimal.

In addition, when running a Multiple Regression, there are several assumptions (Table 4) that need to be checked, the data met for the analysis to be reliable and valid. The models qualify if they pass the assumptions check.

Table 4: Assumptions in Multiple linear regression

	Assumption	Validity
1	Linearity between independent variables and dependent variables	Relationship between the independent variables and the dependent variable can be characterized by a straight line.
2	There is no multicollinearity in your data	Variance Inflation Factor (VIF) scores to be well below 10 Tolerance scores to be above 0.2
3	The values of the residuals are independent	Durbin-Watson value between 1 and 3 or close to 2
4	The variance of the residuals is constant	Residuals should look like a random array of dots
5	The values of the residuals are normally distributed	The closer the dots lie to the diagonal line, the closer to normal the residuals are distributed
6	There are no influential cases biasing the model	Cook's Distance to be less than 1

Next, model validation was performed by comparing the generated predicted values and actual values. Validation examines the model fit to the actual experimental data. Conclusions could be drawn by observing the model and residual summary.

3. Results and Discussions

3.1. Correlation analysis

Pearson's correlation analysis describes the most important effects of the parameters to be considered in the models. Table 5 shows the correlation coefficients (r) and identifications of the significance (determined by p-values). p-values are not explicitly presented but value less than 0.05 and 0.01 identified as * and ** respectively are shown, indicating that the parameters in relation to P crystallization are statistically significant. Correlations analysis between the parameters was also performed and presented in Appendix 1 for calcium phosphate and Appendix 2 for struvite.

Table 5: Pearson's correlation coefficients of different parameters with P crystallized fraction in calcium phosphate and struvite processes

	P crystallized fraction	
	Calcium Phosphate	**Struvite**
Feed flow	-0.255	.[a]
Recirculation flow	-0.269	0.373
Inflow	-.311*	0.373
NaOH dosing flow	0.050	0.303
$CaCL_2$ dosing flow	-0.163	-
$MgCl_2$ dosing flow	-	-0.195
[Ca] in feed filtered	0.011	-
[Mg] in feed total (unfiltered)	-	.489*
[NH_4] in feed total (unfiltered)	-	-0.110
[P] in feed filtered / unfiltered	0.265	0.225
[Ca] in effl. reactor filtrated	-0.067	-
[Mg] in effl. Filter total (unfiltered)	-	-.465*
[NH_4] in effl. Filter total (unfiltered)	-	-0.252
[P] in effl. reactor filtrated / unfiltered	-.835**	-.634**
[Ca][/P] ratio feed + [Ca] dosing	-.298*	-
NH_4/P ration total (unfiltered)	-	-0.296
pH feed	-0.010	0.330
pH effluent reactor	.777**	.412*

**. Correlation is significant at the 0.05 level (2-tailed).*
***. Correlation is significant at the 0.01 level (2-tailed).*
.[a] Cannot be computed because at least one of the variables is constant

From Table 5, it can be said that Inflow, [P] in effluent, Ca/P ratio and pH effluent are significant for calcium phosphate crystallization process. As for struvite, Mg in feed, Mg in effluent, P in effluent and pH effluent are important. However, the parameters in the effluent can be disregarded as they are the outcome of the system and cannot be controlled, except for pH effluent. The pH of the effluent in this case is the same as the pH at the bottom of the reactor, in which the condition when the crystallization happens. That leaves Inflow, Ca/P ratio and pH effluent for calcium phosphate and Mg in feed and pH effluent for struvite process.

Other parameters to observe are explained in this section. The overall performance of crystallization is a function of supersaturation, pH, and concentration of reactants. Therefore, it is logical to include chemicals dosing flow in the models. NaOH as pH adjusting agent and calcium or magnesium salts as the reactants. Above all, ion concentrations of P, Ca, NH_4 Mg in feed are the key determination of the whole crystallization process and must be incorporated in the models. They may as well have an inflow (feed flow and recirculation flow) as it plays an important role in the crystallization performance. The up-flow velocity in the reactor pushed the sand in a fluidized state. Correlations between these parameters are also supported by the Pearson analysis in Appendix 1 and Appendix 2. It is important to note that Ca/P ratio can represent the total Ca and P in the reactor, thus, the same as $CaCl_2$ dosing, Ca in feed and P in feed together. The same holds for NH_4/P ratio, it represents NH_4 in feed and P in feed. Those conclude the predictors for the models with two groups for each process (Table 6). After considering what parameters are significant to P crystallization, multiple linear regression models were developed.

Table 6: Predictors used for calcium phosphate and struvite process

Group	Calcium phosphate	Struvite
1	• Inflow • NaOH dosing flow • $CaCl_2$ dosing flow • Ca/P ratio	• Inflow • NaOH dosing flow • $MgCl_2$ dosing flow • [Mg] in feed

	• pH feed • pH effluent	• NH_4/P ratio • pH feed • pH effluent
2	• Inflow • NaOH dosing flow • $CaCl_2$ dosing flow • [P] in feed • [Ca] in feed • pH feed • pH effluent	• Inflow • NaOH dosing flow • $MgCl_2$ dosing flow • [Mg] in feed • [NH_4] in feed • [P] in feed • pH feed • pH effluent

3.2. Regression models

3.2.1. Calcium Phosphate

Initially, predictors used are inflow, Ca/P ratio and pH effluent (Group 1) or same predictors with replacing Ca/P with $CaCl_2$ dosing, Ca in feed and P in feed (Group 2). In the second model, variable of NaOH dosing is entered and the third model, pH feed is added. Table 7 is summary of the results of the different models. It shows that model 6 from group 2 has the highest R^2 and lowest Std. Error to be selected as the best model.

Table 7: Model summary of given two groups for calcium phosphate

Model Summary

Group	Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
						R Square Change	F Change	df1	df2	Sig. F Change
1	1	.736	.542	.510	10.803	.542	16.948	3	43	.000
	2	.782	.611	.574	10.069	.069	7.501	1	42	.009
	3	.810	.656	.614	9.590	.044	5.296	1	41	.027
2	4	.737	.543	.487	11.052	.543	9.734	5	41	.000
	5	.855	.731	.690	8.585	.188	27.948	1	40	.000
	6	.889	.790	.752	7.681	.059	10.966	1	39	.002

However, model 6 cannot directly be used because its validity further tested based on multiple linear regression assumptions in Table 4. The outcome appeared that assumptions of 1, 2 and 6 could not be met. Thus, adjustments were necessary for improvements. Assumption 1 does not strongly affect the reliability of the model but need to be pointed out that the relationship of the independent variables and dependent variable is not all linear. Assumption 2 concerns the highly correlated variables and found to be Inflow, P in feed and pH feed. Removing one of the variables or multiplying the variables could be done. Lastly, assumption 6 identified two outliers in the data, which need to be removed. The combinations of removing and multiplying highly correlated variables can be seen in Table 8 and followed by the generated models (Table 9).

Table 8: Predictors of model improvements for model 6

Model	Predictors
6.1	Inflow, chemicals dosing, ca, pH effluent, (P feed*pH feed)
6.2	Chemicals dosing, p feed, ca, pH effluent, (inflow*pH feed)
6.3	Chemicals dosing, ca, pH feed, pH effluent, (inflow*P feed)
6.4	Chemicals dosing, ca, pH effluent, (P feed*pH feed*inflow)
6.5	Inflow, chemicals dosing, ca, p, pH effluent
6.6	Chemicals dosing, ca, pH effluent, (inflow*P feed)

Table 9: List of model improvements for model 6

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
6.1	.872	.760	.724	8.104	1.208
6.2	.877	.769	.735	7.950	1.217
6.3	.877	.768	.734	7.962	1.208
6.4	.875	.766	.738	7.901	1.167
6.5	.855	.731	.690	8.585	1.247
6.6	.844	.713	.678	8.762	1.213

Looking at how the R^2 are relatively high and similar, it can be difficult to determine which model is the best. In this case, the lowest Std. Error is considered to be the most important, so model 6.4 was selected. The model equation is shown below, and the full model descriptions along with its assumptions check are provided in Appendix 3.

$$\text{P crystallization } (CaPO_4) = -116.733 - 53.268x_1 + 10.519x_2 - .065x_3 + 27.504x_4 + .001x_5$$

The variables, x_1, x_2, x_3, x_4 and x_5 in the equation denote actual values of respectively NaOH dosing, $CaCl_2$ dosing, Ca in feed, pH effluent and (P feed*pH feed*inflow).

3.2.2. Struvite

Same procedures were applied to struvite process. In the first stage, predictors used are inflow, Mg in feed, NH_4/P ratio and pH effluent (Group 1) or same predictors with replacing NH_4/P with NH_4 in feed and P in feed (Group 2). In the second and third models, chemicals dosing of NaOH and $MgCl_2$ and pH feed are entered respectively. Table 10 is a summary of the results of the different models. It shows that model 2 from group 1 has the highest R^2 and lowest Std. Error, as such to be selected as the best model.

Table 10: Model summary of given two groups for struvite

Model Summary

Group	Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
						R Square Change	F Change	df1	df2	Sig. F Change
1	1	.839	.704	.648	9.919	.704	12.490	4	21	.000
	2	.873	.762	.686	9.362	.057	2.289	2	19	.129
	3	.874	.764	.672	9.571	.002	.177	1	18	.679
2	4	.847	.718	.647	9.924	.718	10.179	5	20	.000
	5	.877	.768	.678	9.479	.050	1.962	2	18	.169
	6	.885	.783	.681	9.444	.014	1.133	1	17	.302

In the validity check, assumption 1 and 5 may be violated but only need to be interpreted with caution as they do not badly influence the model. Full model descriptions and assumptions check are also provided in Appendix 4. There was no multicollinearity in the variables and no outliers were found in the model. The equation for model 2 is expressed as:

$$\text{P crystallization (struvite)} = -243.273 + .554x_1 - 10.204x_2 + 17.539x_3 + .083x_4 - 7.348x_5 + 31.251x_6$$

Where, x_1, x_2, x_3, x_4, x_5 and x_6 are inflow, NaOH dosing, $MgCl_2$ dosing, Mg in feed, NH_4/P ratio and pH effluent respectively.

3.3. Validation

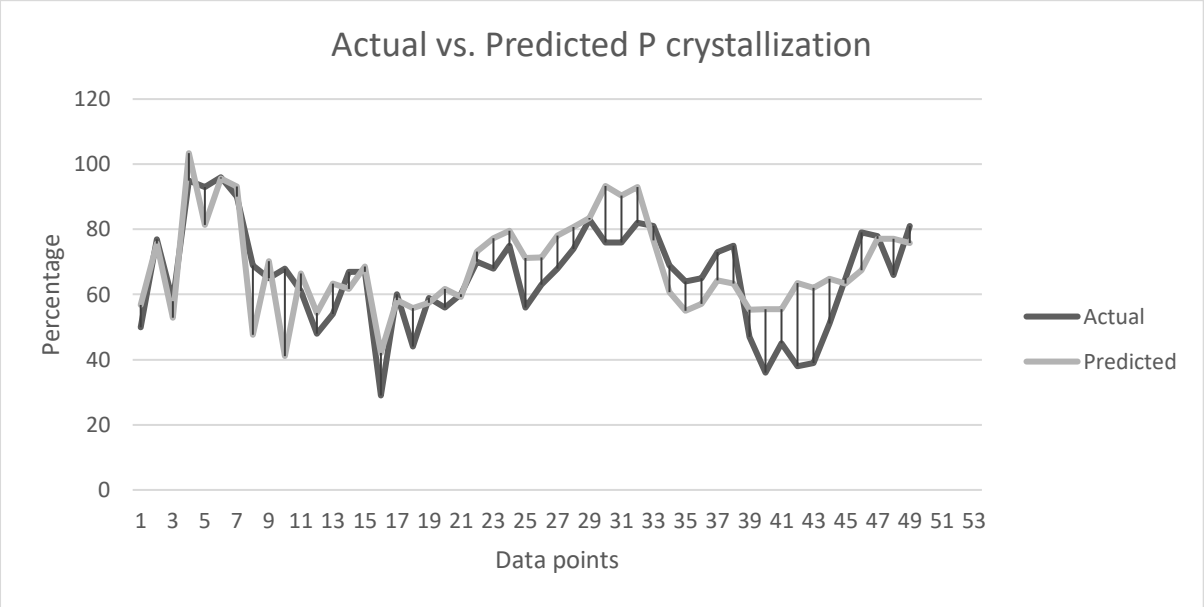
The high coefficients of determination (R^2 > 0.7) and low Std. Error (<10) of the chosen models confirming the relationship between the models and the dependent variable. Experimental data and generated predicted values are close to each other which is evident in Figure 2 for calcium phosphate (a) and struvite (b). In ANOVA tables of both processes (Appendix 3 and Appendix 4), the p-value (Sig.) associated with the F-value is very small (0.000). It can be concluded that the group of the independent variables reliably predict the dependent variable.

The residuals (actual subtracts predicted values for each data point) of the models can provide information on how promising the models are. Positive residual indicates over-estimated number and in contrast, negative residual specify under-estimated number. In Table 11, both models give higher numbers of over-estimated over under-estimated values. Under-estimated prediction of crystallized phosphate in the industry is preferred as it does not give high expectation of the actual recovered product. In economic analysis, this should be taken into account.

Table 11: Residuals summary

	Calcium phosphate	Struvite
Max	26.8965	32.08225
Min	-25.4803	-16.4339
Under-estimated count	20	12
Over-estimated count	29	14

a)



b)

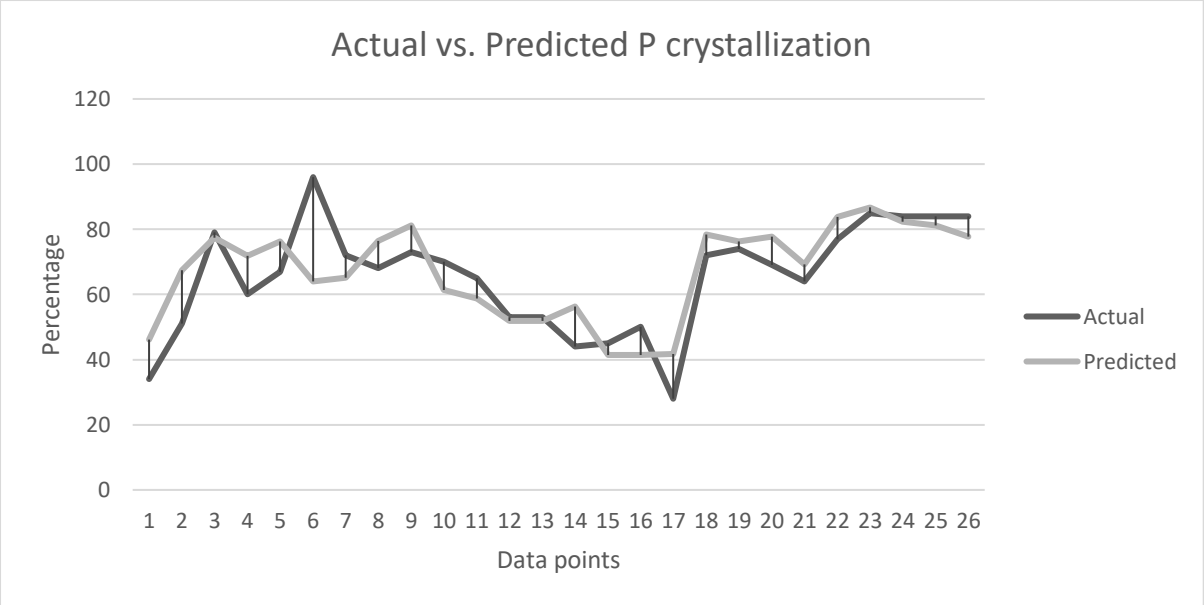


Figure 2: Distribution of experimentally determined data and model predicted values of P crystallization for a) calcium phosphate and b) struvite

Conclusion

Black-box modelling of calcium phosphate and struvite crystallization in a fluidized-bed reactor was carried out using multiple linear regression analysis. The predictors are inflow, chemicals dosing, pH, related ion concentrations and molar ratio (Ca/P and NH_4/P). They are found to be significant through Pearson's correlation analysis and knowledge of crystallization process. Trial and error were done to obtain the best model for each process. Both models acquired R^2 higher than 0.7 and Std. Error lower than 10 allowing good prediction models. Besides that, the models show over-estimated values if applied in realistic setting, which need to be considered in the economic analysis.

Recommendations

There are still many approaches one could do to draw coefficient for making prediction model. One example is by using response surface analysis (RSA). The model in this study might have weaknesses as not all the assumptions for multiple regression are met. RSA may generate better models for crystallization. Most of the time, after creating model, optimization is also performed. SPSS does not have this option. If there is another opportunity, Design Expert is recommended to do RSA and optimization. The software is specifically dedicated to performing design of experiments.

In addition, making a grey-box models for crystallization may be possible when considering the combination of factors, such as thermodynamic and kinetic process. There are three theoretical approaches to model crystallizations: classical nucleation theory (CNT), kinetic approach, and the third category comprising of theories not directly derived from CNT or kinetic theory (Ostapienko et al., 2019). Depending on the researcher's main interest, those approaches can be applied. The complexity of the model increases with applying the knowledge of crystallization for calcium phosphate and struvite.

Acknowledgement

The author would like to thank Maria van Schaik for her support and assistance in this project. Also, the cooperation with Seda Sucu for her helpful advice on statistical issues examined in this study. As

well, words of gratitude to Bram Verkruysse for his useful advice and comments. This project has received funding from the Interreg 2 Seas program 2014-2020 co-funded by the European Regional Development Fund under subsidy contract No 2S03-011. For further information about the NEREUS Project, see https://www.nereus-project.eu/

References

Agudelo-Vera, C. M., Leduc, W. R. W. A., Mels, A. R., & Rijnaarts, H. H. M. (2012). Harvesting urban resources towards more resilient cities. *Resources, Conservation and Recycling*, *64*, 3–12. https://doi.org/10.1016/j.resconrec.2012.01.014

Ali, M. I., & Schneider, P. A. (2006). A fed-batch design approach of struvite system in controlled supersaturation. *Chemical Engineering Science*, *61*(12), 3951–3961. https://doi.org/10.1016/j.ces.2006.01.028

Bohlin, T. (1991). Interactive System Identification: Prospects and Pitfalls. In *Springer-Verlag Berlin Heidelberg* (1st ed.). Springer-Verlag Berlin Heidelberg. https://doi.org/10.1017/CBO9781107415324.004

Bonen, A., Campbell, C. J., Kirby, R. L., & Belcastro, A. N. (1979). A multiple regression model for blood lactate removal in man. *Pflügers Archiv European Journal of Physiology*, *380*(3), 205–210. https://doi.org/10.1007/BF00582897

Burek, P., Satoh, Y., Fischer, G., Kahil, M. T., Scherzer, A., Tramberend, S., Nava, L. F., Wada, Y., Eisner, S., Flörke, M., Hanasaki, N., Magnuszewski, P., Cosgrove, B., & Wiberg, D. (2016). Water Futures and Solution. Fast Track Initiative. *International Institute for Applied Systems Analysis*, *May*, 1–113.

Catalina, T., Iordache, V., & Caracaleanu, B. (2013). Multiple regression model for fast prediction of the heating energy demand. *Energy and Buildings*, *57*, 302–312. https://doi.org/10.1016/j.enbuild.2012.11.010

Dobbs, R., Oppenheim, J., Thompson, F., Brinkman, M., & Zornes, M. (2011). *Resource Revolution : Meeting the world ' s energy , materials , food , and water needs* (Issue November). https://www.mckinsey.com/~/media/McKinsey/Business Functions/Sustainability/Our Insights/Resource revolution/MGI_Resource_revolution_full_report.ashx

Estrada-Flores, S., Merts, I., De Ketelaere, B., & Lammertyn, J. (2006). Development and validation of "grey-box" models for refrigeration applications: A review of key concepts. *International Journal of Refrigeration*, *29*(6), 931–946. https://doi.org/10.1016/j.ijrefrig.2006.03.018

Kok, D. J. D., Pande, S., Van Lier, J. B., Ortigara, A. R. C., Savenije, H., & Uhlenbrook, S. (2018). Global phosphorus recovery from wastewater for agricultural reuse. *Hydrology and Earth System Sciences*, *22*(11), 5781–5799. https://doi.org/10.5194/hess-22-5781-2018

Montastruc, L., Azzaro-Pantel, C., Pibouleau, L., & Domenech, S. (2004). Use of genetic algorithms and gradient based optimization techniques for calcium phosphate precipitation. *Chemical Engineering and Processing: Process Intensification*, *43*(10), 1289–1298. https://doi.org/10.1016/j.cep.2003.12.002

Morse, G. K., Brett, S. W., Guy, J. A., & Lester, J. N. (1998). Review: Phosphorus removal and recovery technologies. *Science of the Total Environment*, *212*(1), 69–81. https://doi.org/10.1016/S0048-9697(97)00332-X

Neczaj, E., & Grosser, A. (2018). Circular Economy in Wastewater Treatment Plant–Challenges and Barriers. *Proceedings*, *2*(11), 614. https://doi.org/10.3390/proceedings2110614

NEREUS Project Interreg 2 Seas. (2018). https://www.nereus-project.eu/about/

Ostapienko, B. I., Lopez, D., & Komarova, S. V. (2019). Mathematical modeling of calcium phosphate precipitation in biologically relevant systems: scoping review. *Biomechanics and Modeling in Mechanobiology*, *18*(2), 277–289. https://doi.org/10.1007/s10237-018-1087-7

Pauly, D. (1989). A multiple regression model for predicting the food consumption of marine fish populations. *Marine and Freshwater Research*, *40*(3), 259–273. https://doi.org/10.1071/MF9890259

Puyol, D., Batstone, D. J., Hülsen, T., Astals, S., Peces, M., & Krömer, J. O. (2017). Resource recovery from wastewater by biological technologies: Opportunities, challenges, and prospects. *Frontiers in Microbiology*, *7*(JAN), 1–23. https://doi.org/10.3389/fmicb.2016.02106

Sohlberg, B. (1998). Supervision and Control for Industrial Processes. In *Springer*. Springer. https://doi.org/10.1017/CBO9781107415324.004

Song, Y., Hahn, H. H., & Hoffmann, E. (2002). Effects of solution conditions on the precipitation of phosphate for recovery: A thermodynamic evaluation. *Chemosphere*, *48*(10), 1029–1034. https://doi.org/10.1016/S0045-6535(02)00183-2

Song, Y., Weidler, P. G., Berg, U., Nüesch, R., & Donnert, D. (2006). Calcite-seeded crystallization of calcium phosphate for phosphorus recovery. *Chemosphere*, *63*(2), 236–243. https://doi.org/10.1016/j.chemosphere.2005.08.021

Tulleken, H. J. A. F. (1993). Grey-box modelling and identification using physical knowledge and bayesian techniques. *Automatica*, *29*(2), 285–308. https://doi.org/10.1016/0005-1098(93)90124-C

Van Der Hoek, J. P., De Fooij, H., & Struker, A. (2016). Wastewater as a resource: Strategies to recover resources from Amsterdam's wastewater. *Resources, Conservation and Recycling*, *113*, 53–64. https://doi.org/10.1016/j.resconrec.2016.05.012

Velenturf, A. P. M., & Purnell, P. (2017). Resource recovery from waste: Restoring the balance between resource scarcity and waste overload. *Sustainability (Switzerland)*, *9*(9). https://doi.org/10.3390/su9091603

Wilsenach, J. A., & van Loosdrecht, M. C. M. (2006). Integration of processes to treat wastewater and source-separated urine. *Journal of Environmental Engineering*, *132*(3), 331–341. https://doi.org/10.1061/(ASCE)0733-9372(2006)132:3(331)

Appendices

Appendix 1

Pearson correlation analysis between different parameters for calcium phosphate route

	Feed flow	Recirculation flow	Inflow	NaOH dosing flow	CaCL2 dosing flow	[P] in feed filtered	[Ca] in feed filtered	[P] in effl. reactor filtrated	[Ca] in effl. reactor filtrated	[Ca][/P] ratio feed + [Ca] dosing	pH feed	pH effluent reactor
Feed flow	1	0.093	.290*	.381**	.688**	-0.120	-0.165	0.158	0.259	0.248	0.149	-0.145
Recirculation flow	0.093	1	.980**	-.391**	0.224	-.853**	0.181	-0.086	.589**	.693**	.843**	-.348*
Inflow	.290*	.980**	1	-.300*	.353*	-.844**	0.138	-0.051	.621**	.719**	.841**	-.364**
NaOH dosing flow	.381**	-.391**	-.300*	1	0.145	.560**	-.358*	0.132	-0.089	-.505**	-.373**	.339*
CaCL2 dosing flow	.688**	0.224	.353*	0.145	1	-0.226	-.289*	0.050	.370**	.566**	0.209	-0.231
[P] in feed filtered	-0.120	-.853**	-.844**	.560**	-0.226	1	-0.139	0.188	-.463**	-.795**	-.625**	.344*
[Ca] in feed filtered	-0.165	0.181	0.138	-.358*	-.289*	-0.139	1	-0.124	0.029	.283*	.422**	0.054
[P] in effl. reactor filtrated	0.158	-0.086	-0.051	0.132	0.050	0.188	-0.124	1	-0.153	-0.106	-0.228	-.730**
[Ca] in effl. reactor filtrated	0.259	.589**	.621**	-0.089	.370**	-.463**	0.029	-0.153	1	.537**	.680**	-0.094
[Ca][/P] ratio feed + [Ca] dosing	0.248	.693**	.719**	-.505**	.566**	-.795**	.283*	-0.106	.537**	1	.577**	-.365*
pH feed	0.149	.843**	.841**	-.373**	0.209	-.625**	.422**	-0.228	.680**	.577**	1	-0.152
pH effluent reactor	-0.145	-.348*	-.364**	.339*	-0.231	.344*	0.054	-.730**	-0.094	-.365*	-0.152	1
[P] crystallized fraction	-0.255	-0.269	-.311*	0.050	-0.163	0.265	0.011	-.835**	-0.067	-.298*	-0.010	.777**

**. Correlation is significant at the 0.05 level (2-tailed).*

***. Correlation is significant at the 0.01 level (2-tailed).*

Appendix 2

Pearson correlation analysis between different parameters for struvite route

	Feed flow	Recirculation flow	Inflow	NaOH dosing flow	MgCl2 dosing flow	[Mg] in feed total (unfiltered)	[NH4] in feed total (unfiltered)	[P] in feed total (unfiltered)	NH4/P ration total (unfiltered)	[Mg] in effl. Filter total (unfiltered)	[NH4] in effl. Filter total (unfiltered)	[P] in effl. Filter total (unfiltered)	pH in feed	pH in effluent reactor
Feed flow	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]	.[a]
Recirculation flow	.[a]	1	1.000**	0.067	-0.103	0.100	0.147	0.025	0.247	0.113	-0.024	-0.062	0.154	-0.097
Inflow	.[a]	1.000**	1	0.067	-0.042	0.114	0.016	0.151	-0.165	-0.293	0.193	0.007	0.338	-0.176
NaOH dosing flow	.[a]	0.067	0.067	1	-0.094	.651**	-0.062	.392*	-.481*	0.006	-0.161	0.155	0.307	-0.185
MgCl2 dosing flow	.[a]	-0.103	-0.042	-0.094	1	-0.010	-0.358	-0.309	0.036	0.242	-0.265	-0.338	-.613**	-.597**
[Mg] in feed total (unfiltered)	.[a]	0.100	0.114	.651**	-0.010	1	-.508**	-0.171	0.031	-0.007	-0.037	-0.259	0.099	-0.126
[NH4] in feed total (unfiltered)	.[a]	0.147	0.016	-0.062	-0.358	-.508**	1	.633**	-0.041	0.128	-0.137	0.223	0.041	.392*
[P] in feed total (unfiltered)	.[a]	0.025	0.151	.392*	-0.309	-0.171	.633**	1	-.728**	-.446*	0.073	0.336	.545**	0.174
NH4/P ration total (unfiltered)	.[a]	0.247	-0.165	-.481*	0.036	0.031	-0.041	-.728**	1	.746**	-0.334	-0.280	-.570**	0.047
[Mg] in effl. Filter total (unfiltered)	.[a]	0.113	-0.293	0.006	0.242	-0.007	0.128	-.446*	.746**	1	-0.325	-0.061	-.544**	-0.231
[NH4] in effl. Filter total (unfiltered)	.[a]	-0.024	0.193	-0.161	-0.265	-0.037	-0.137	0.073	-0.334	-0.325	1	.476*	.387*	-0.143
[P] in effl. Filter total (unfiltered)	.[a]	-0.062	0.007	0.155	-0.338	-0.259	0.223	0.336	-0.280	-0.061	.476*	1	0.366	-.407*
pH in feed	.[a]	0.154	0.338	0.307	-.613**	0.099	0.041	.545**	-.570**	-.544**	.387*	0.366	1	0.202
pH in effluent reactor	.[a]	-0.097	-0.176	-0.185	-.597**	-0.126	.392*	0.174	0.047	-0.231	-0.143	-.407*	0.202	1

**. Correlation is significant at the 0.05 level (2-tailed).*

***. Correlation is significant at the 0.01 level (2-tailed).*

.[a] Cannot be computed because at least one of the variables is constant.

Appendix 3

Full model descriptions for calcium phosphate

Model Summary

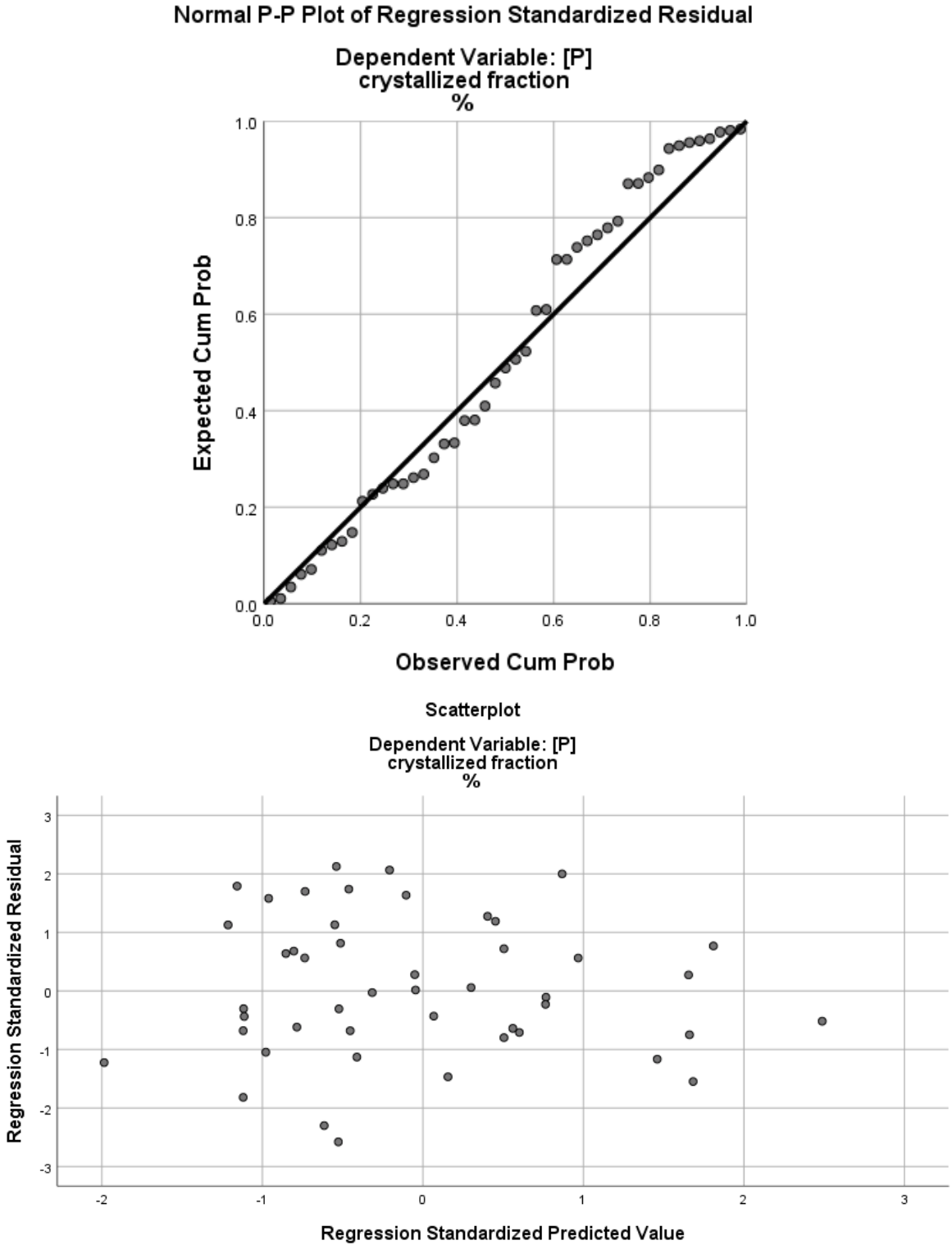
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
6.4	.875[a]	.766	.738	7.901	1.167

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	8392.549	5	1678.510	26.887	.000
Residual	2559.519	41	62.427		
Total	10952.068	46			

Coefficients

	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
(Constant)	-116.733	17.947		-6.504	.000		
NaOH dosing flow	-53.268	8.972	-.795	-5.937	.000	.318	3.147
CaCL2 dosing flow	10.519	5.042	.215	2.086	.043	.536	1.865
[Ca] in feed filtered	-.065	.020	-.302	-3.265	.002	.666	1.502
pH effluent reactor	27.504	2.740	1.240	10.039	.000	.374	2.676
PfeedxpHfeedxinflow	.001	.000	.376	4.150	.000	.694	1.440


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: [P]
crystallized fraction
%
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0
Scatterplot
Dependent Variable: [P]
crystallized fraction
%
Regression Standardized Residual
Regression Standardized Predicted Value
-3
-2
-1
0
1
2
3

Appendix 4

Full model descriptions for struvite

Model Summary

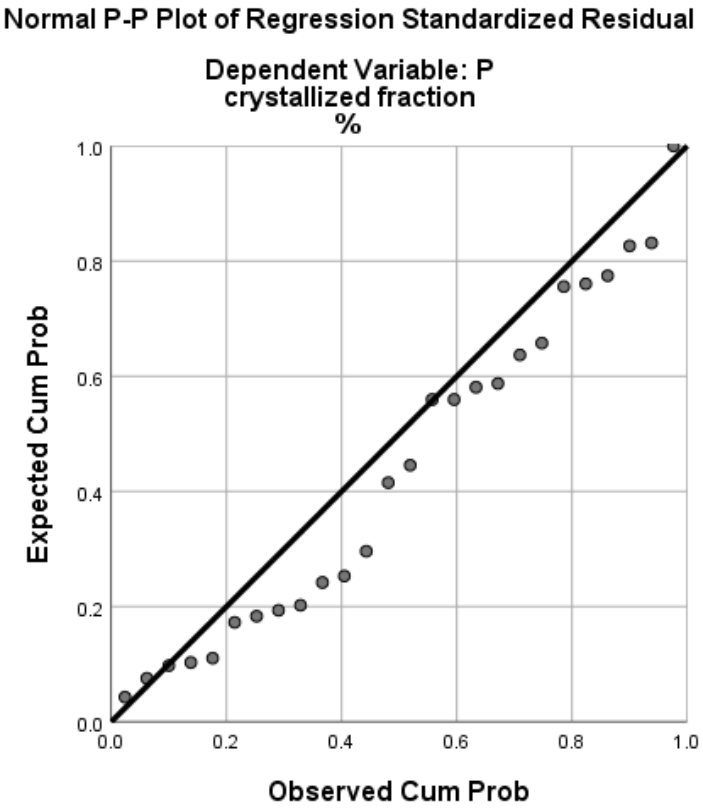
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
2	.873[a]	.762	.686	9.362	1.730

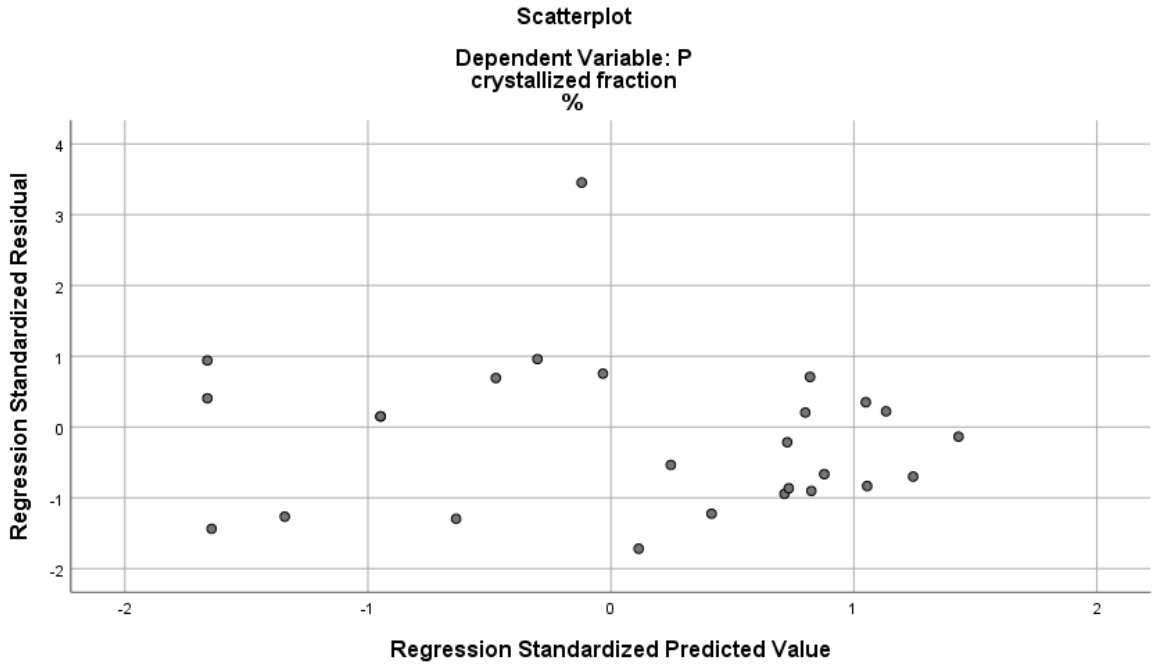
ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	5317.179	6	886.197	10.112	.000
Residual	1665.167	19	87.640		
Total	6982.346	25			

Coefficients

	Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
(Constant)	-243.273	61.365		-3.964	.001		
Inflow l/h	.554	.181	.372	3.071	.006	.853	1.172
NaOH dosing flow l/h	-10.204	10.326	-.209	-.988	.335	.279	3.580
MgCl2 dosing flow l/h	17.539	11.610	.227	1.511	.147	.558	1.793
[Mg] in feed total (unfiltered) mg Mg/l	.083	.022	.683	3.812	.001	.391	2.555
NH4/P ration total (unfiltered) mol/mol	-7.348	2.902	-.396	-2.532	.020	.512	1.952
pH in effluent reactor	31.251	7.094	.679	4.405	.000	.529	1.890


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: P crystallized fraction %
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0


Scatterplot
Dependent Variable: P crystallized fraction %
Regression Standardized Residual
Regression Standardized Predicted Value
-2
-1
0
1
2
3
4