Data-driven polynomial chaos expansions for characterization of complex fluid rheology: Case study of Phosphate slurry

Nabil El Moçayd[a,*], Mohammed Seaid[b]

[a]*International Water Research Institute, University Mohammed VI Polytechnic, Benguerir, Morocco*

[b]*Department of Engineering, University of Durham, South Road, Durham DH1 3LE, UK*

Abstract

Mine transportation through hydraulic pipelines is increasingly used by various industries around the world. In Morocco, this has been implemented for the case of phosphate transportation. This allows to increase the production and reduce the transportation cost. Given the vital importance of phosphate in the global food security and regarding the huge amount of phosphate rock reserves in Morocco, it is detrimental to assess the reliability, to optimize and to increase its transportation in a safe manner. Usually hydraulic transportation of such fluids is fully quantified with a full characterization of its rheology related to its non-Newtonian behavior. The rheology allows to know the viscous and the elastic properties of a fluid exhibiting viscoelastic properties. In the case of water-phosphate slurry this behavior is not well-documented and classical constitutive laws for the rheology are of limited used, because of the high variability of different physico-chemical components of the slurry. The present work aims at quantifying the sensitivity of the water-phosphate slurry rheology to these components. In order to achieve this objective, a data-driven model based on polynomial chaos expansions (PCE) is developed and investigated. The choice of this class of models is motivated by the simplicity to conduct sensitivity analysis with the PCE and the limited amount of data available as the water-phosphate slurry pipeline is very new. In order to alleviate further the impact of the limitation given by the available data, we introduce the bagging technique which is an Ensemble based data-driven model using the PCE. Results presented in this study demonstrate that the bagging allows to reduce the validation error of the model by up to two orders of magnitude. Thus, it reduces considerably the variability on the estimation of hyperparameters in the model. Moreover, the sensitivity analysis shows that the variability on the elasticity coefficient is mainly due to the variability of the slurry density and the solid rate. Viscosity on the other side is not affected by the heterogeneity of the granulation distribution.

Keywords: Data-driven modeling, Polynomial chaos expansion, Rheology, Bagging, phosphate slurry, Machine learning, Sensitivity analysis.

*Corresponding author

Email addresses: nabil.elmocayd@um6p.ma (Nabil El Moçayd), m.seaid@durham.ac.uk (Mohammed Seaid)

1. Introduction

Phosphate is a key element for the plant growth and other agricultural purposes, see for example [1]. Therefore, it urges for phosphate industries around the world to increase their operations efficiency. The phosphate is a non renewable resource and it is classically extracted from phosphate rocks. Authors in [2] analyzed the worldwide reserve-to-production that determines the lifetime of phosphate rocks. The conclusion of this work shows that because of the upcoming socio-economical stresses, the global phosphorous security will rely on a single country: Morocco. In fact, Morocco holds 77% of global phosphate reserve in the world and needs to increase its production by around 700% by 2075 in order to meet with the global increasing demands. Achieving this productivity value in such time would require to reach an optimal level of performance in every single operation within the industry. In Morocco, one could classify the main activities of the phosphate production into the following four classes [3]:

1. *Extraction and washing:* This activity is mainly linked to extracting phosphate from rocks. The phosphate is then enriched with water and other minerals in order to be transported.
2. *Processing:* This activity aims at transforming the crude phosphate obtained from extraction into phosphoric acid or fertilizers.
3. *Transportation:* This activity aims at transporting the phosphate from the mine sites to the processing platforms.
4. *Distribution and sales:* This activity is responsible for distributing the phosphate-based products in the world.

Hence, the transport of phosphate is a major component in the industry. Historically, this was achieved by carrying the phosphate over railways. However, for sustainability reasons, the industry moved to a transport logistic using slurry pipeline. This allows to save 3×10^6 m^3 of water and to reduce considerably the carbon footprint of the country [1]. Yet this operation is far to be described as optimal for reliability reasons that ought to be detailed hereafter.

Mine transport has begun to receive much attention especially regarding their importance to reduce the cost of processing in mining industry. The hydro-transport could be carried when mixing the minerals with water. The resulting fluid can be then transported through pipelines. Such hydrodynamic process is mainly controlled by two major features: the head losses and the flow friction. While it is easy to quantify such physical constraints for classical fluids such as water, mineral slurry usually exhibits a non-Newtonian behavior that makes them very hard to assess [4]. In fact, a mineral slurry could be classified as a visco-plastic fluid in such a way that it has a solid behavior when the shear stress is under a threshold (the elasticity coefficient τ_y). Moreover, its apparent viscosity depends on the time and the flow regime considered. Consequently, it is very hard to model the behavior of such fluids and quantify their dynamics [5]. The study of such physical parameters is known in the literature as rheology. The study of rheology for slurry allows not only to assess the head loss of the hydrodynamics but also to predict turbulent flows [6]. In fact, the rheology of a fluid has an effect of slumping which can impact the transport of the mineral slurry. One

[1] Based on the information available on the website `www.ocpgroup.ma`

has to make sure that the hydrodynamic is sufficiently well establish to prevent sedimentation and not too over sized to prevent pipe failure and unnecessary energy loss. Hence, modeling such hydrodynamics is a key tool to support operational decisions. Selecting models that accurately describe the rheology of a slurry is of most importance in order to initialize numerical simulation of non-Newtonian flows [7]. The literature provides a considerable amount of rheological models known as constitutive models aiming at describing the behavior of a non-Newtonian fluid, see [8, 9] among others. These equations allows to extrapolate the rheology estimated from measurements using a finite set of experimental measurements or field data. In the case of mineral slurry, many previous works studied the applicability of these constitutive models . These works revealed some major difficulties in the modelling and numerical simulation of slurry flows. In the framework of phosphate slurry flow in pipelines, the use of classical non-Newtonian models such as Bingham or Hershley & Bukley to represent the rheology has already been addressed in [10] among others. Results show that these constitutive laws fail to correctly model the hydrodynamics for all ranges of the flow, we refer the reader to [10] for more details. In fact, there are several reasons why fitting and extrapolating some constitutive models to describe rheological behavior of a slurry is inaccurate. For instance, authors in [11] linked the complex rheology of a cement slurry to the irreversible and temporal nature of the fluid. Moreover, several works highlighted that rheological parameters are very sensitive to the physico-chemical composition of the fluid mixture. For example, the study reported in [12] showed that the bottom ash slurry is sensitive to the chemical composition and the solid concentration. The same conclusion was drawn for coking coal and water slurry in [13]. Similarly, authors in [14] showed that the rheology of silica sand based suspension is primarily driven by the particle size and the solid concentration. The recent work in [15] highlighted also the effect of the particle concentration, density, chemical properties of the mixture, temperature and pressure on the rheology of a cement slurry.

In the case of phosphate slurry, authors in [10] showed that despite the applicability of some constitutive models to describe its rheology, these models lack generalization. The emphasis was on the great variability of the rheological parameters using real data from the pipeline. An emerging idea to accurately model the rheological behavior consists on using Machine Learning (ML) methods. Such methodology was implemented for Iron-Ore slurry using artificial neural network, see [16]. The ML maps the nonlinear relationship existing between the apparent viscosity and the the physical properties of the fluid with very high accuracy. The model allows also to derive the physical characteristics allowing to understand the rheology and to assess the parameters to which the rheological parameters are sensitive. There are several ML methods allowing to perform the regression analysis including the linear regression, neural networks, kernel methods (such as Gaussian processes), support vector machine, graphical models, among others. We also refer the reader to [17] for a full overview of these methods. Choosing an adapted regression method is often a very hard task since these methods are still subject to open issues [18]. One common approach used in different applications that allows to choose the appropriate ML tools consists on comparing different approaches and choosing the method displaying the best accuracy based on a considered metric [19]. Another method that may be used is through a deep knowledge of the problem nature, the quality of the data and the purposes for which the model will be used. In the present study, the ML method is used to predict the rheology of phosphate slurry and to assess

the sensitivity of this latter to different physico-chemical components of the slurry. This would lead to optimize the operational management and to design future transportation models. In order to achieve these two objectives, operating data have been made available. These data display some non-negligible uncertainty. Thus the selected ML method should be able to handle and propagate the uncertain data [20]. It should be stressed that, the exploitation of the phosphate pipeline started in 2014 but full data accompanied with rheological measurements are available only for the period 2017-2018. This means that only few data are available to build the ML model. This is also a well-known issue when building regression-based ML models [21]. Finally, the resulting ML model should be able to carry sensitivity analysis for physical interpretation. Recent studies revealed that a ML regression model based on data-driven polynomial chaos expansion (PCE) allows to overcome these previous challenges, compare [22, 23] among others. The PCE has largely been used in the context of uncertainty quantification (UQ), see for example [24, 25]. In the context of hydraulic simulations it allows to achieve good accuracies in quantifying the uncertainty of the friction parameter, compare [26, 25]. In general, there are many ways a PCE could be built but the most common consists on using it as a metamodel (a surrogate model, a surface response, etc) which allows to mimic the behavior of a numerical model. Several numerical techniques allow to build a PCE surrogate, compare [27]. However, the most appealing one and relevant with regards to its applicability in the ML context remains the regression method [28, 29]. The main objective of the UQ is to assess the propagation of uncertainty supposed on the inputs through a numerical model on its response. The PCE allows to replace the expensive numerical model by a polynomial decomposition following a spectral approach [30]. As it has been highlighted in [23], the main difference between the use of the PCE in ML and in UQ is that the model is unknown in the case of the ML and the statistics of the model output are available. Another issue with the use of the PCE in ML is that, the probability density function of the inputs should be derived from the data rather than assumed as it is generally used in the UQ. Furthermore, when using a spectral approach, the input parameters are supposed to be statistically independent. An assumption that could not be always verified using the data. Hopefully, some recent studies succeeded in overcoming those two challenges. In fact, some studies allowed to derive the probability density function from the data and therefore to build an appropriate PCE, see [31, 32]. On the other hand, authors in [33] suggested a general framework to address dependencies in inputs parameter using vine copulas. For all these reasons and based on the appealing results published in [23], the present work addresses the prediction of the phosphate slurry rheology using data-driven PCE regression method as a ML tool. Another advantage of the use of the PCE in the UQ context lies in the fact that the uncertainty information could be derived analytically from the spectral coefficient, see [34]. One of the important information easily derived from the PCE in the UQ context is what are the contribution of uncertainty of each input parameter on the resulting uncertainty namely, the sensitivity analysis [35]. There are several purposes for which sensitivity analysis could be run. An overview on sensitivity analysis methods and its application can be found in [36, 37]. For the present case, the sensitivity analysis will help to determine what are the most important physico-chemical parameters that drive the rheology of the phosphate slurry, as in [38]. This would help in the operational to control few parameters and therefore, optimizing the transport of phosphate with a minimal cost. Indeed the variance-based methods aim at

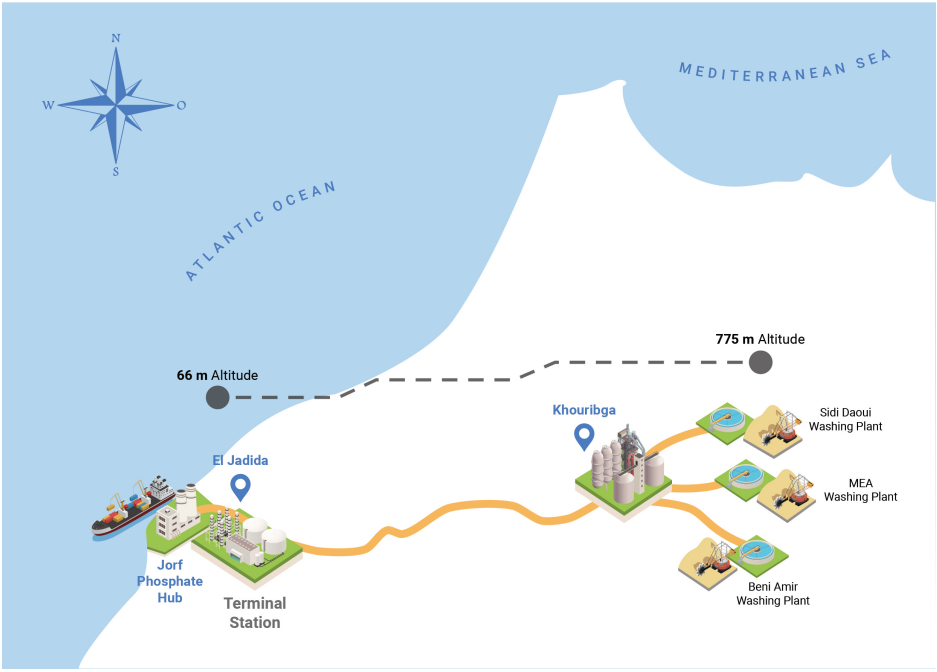

Figure 1: A general illustration of the phosphate slurry pipeline between Khouribga and Jorf Lasfar in Morocco.

quantifying the impact of the variation of parameters on the variation of a quantity of interest. In the present case, sensitivity analysis would help to control few parameters whose variation have a considerable impact on the variation of the rheology.

In the current work, a robust ML regression-based method is developed using a data-driven PCE to predict the rheology of phosphate slurry. The present model is built using available field data and the use of PCE will help to carry the sensitivity analysis. The objective of this study is to understand what are the physico-chemical parameters whose variability drives the variability of the rheology. The results from the sensitivity analysis will be used to deliver some recommendations related to the parameters which have significant impact on the variability observed in the transport of the phosphate slurry in pipelines. This paper is organized as follows. In Section 2 we present details of the phosphate slurry pipeline and the different data provided in the field. The ML model methodology along with the PCE techniques are described in Section 3. Section 4 outlines the performance of the proposed method along with the results obtained for the sensitivity analysis. Finally, concluding remarks are included in Section 5.

2. Data for the phosphate slurry

Minerals slurry pipeline is of an increasing interest as it aims to transport minerals hydraulically instead of the traditional railway transportation. Thus, it allows to decrease considerably the cost of supply chains in the corresponding industry. Morocco has for example, implemented such technology to transport phosphate in the form of

phosphate-water slurry. First the phosphate is mixed with water then transported in pipelines from the head station at Khouribga city to the terminal station at Jorf Lasfar port over 187 Km, see Figure 1. It is worth mentioning here that the head station is also fed with phosphate through secondary pipelines from the washing station. Indeed, the three operational mine sites are located at 50 Km away from the head station (see Figure 1). The phosphate is then gathered in one location before the transport happens. This system allowed to increase the phosphate production with a minimum investment. Although the economic value of this process and its impact on the supply chain was highly appreciated by the industry, the transport has still not achieved its optimized form. The principal reason behind this statement is mainly due to the huge variability in the rheology of the slurry mixture. In fact, one of the main features driving the motion of such fluid is its rheology, especially for the non-Newtonian fluid like phosphate slurry [39]. In the framework of fluid mechanics, the rheology governs the ability of the fluid to go against its motion. Generally, the classification of a fluid (Newtonian or non-Newtonian) is made based on the dependence of the viscosity μ to the mechanical properties of the material constituting the fluid. The viscosity defines the relation between the shear stress $\dot{\gamma}$ (normal stresses on all the fluid sides) and the the shear rate τ (deformation due to the fluid motion). In practice, when this relationship is linear the fluid is referred to be Newtonian and therefore, the viscosity is assumed to be a scalar coefficient. However, when the relationship is nonlinear the fluid is categorized as non-Newtonian and in this case the viscosity is no longer a single coefficient but needs to be identified with a function of the velocity and refereed to as a constitutive equation.

In general, standard approaches to derive a constitutive law for the rheology consists on fitting some well-known equations using some finite numbers of experiments [40, 41]. Usually, the viscosity is measured using a viscometer by filling the fluid in an annular thin space between two concentric cylindrical surfaces. The cylinder located inside the container rotates with an adjustable speed and the viscosity is then obtained using the information about the torque imposed on the cylinder and the angular velocity. We refer the reader to [5] for more details on the viscometer. In many situations, the viscometer is also equipped with some well-established constitutive equations where the experiments are used to fit the hyperparameters of these equations. At the Khouribga site, the viscometer used to estimate the rheology of the phosphate slurry is equipped with the Casson law defined as

$$\sqrt{\tau} = \sqrt{\tau_c} + \sqrt{\mu\dot{\gamma}}, \tag{1}$$

where τ_c refers to the elasticity coefficient and μ is the viscosity. It is evident that the Casson law (1) is a nonlinear equation that involves an elasticity coefficient which translates the limit of the exerted efforts from which the material changes its behavior from solid to fluid. In the case of phosphate slurry, this experiment is carried out every day on the fluid during two years, in order to determine the elasticity and the viscosity of the slurry before its transport in the primary pipeline. All this database was made available for the purpose of the present work. It should also be stressed that, results reported in [10] have shown that this constitutive equation is not suitable for the case of phosphate slurry fluid. Other different constitutive equations known in the literature have also assessed in[10]. The authors conclude that there is no a single equation that is able to model the rheological behavior of the phosphate slurry.

Table 1: List of all physico-chemical parameters estimated before the transport of the phosphate slurry in the primary pipeline.

Parameter	Designation
Al2O3	Allumine content
BPL	Phosphate content
CO2	Carbon dioxide content
MgO	Manganese content
SiO2	Sillice content
SiO2 R	Residual Sillice dioxide content
$< 37\ \mu m$	Proportion of particles with a size lower than 37 μm
$< 40\ \mu m$	Proportion of particles with a size lower than 40 μm
$< 44\ \mu m$	Proportion of particles with a size lower than 44μm
$< 53\ \mu m$	Proportion of particles with a size lower than 53 μm
$< 74\ \mu m$	Proportion of particles with a size lower than 74 μm
$< 160\ \mu m$	Proportion of particles with a size lower than 160 μm
$< 210\ \mu m$	Proportion of particles with a size lower than 210 μm
$< 250\ \mu m$	Proportion of particles with a size lower than 250 μm
$< 315\ \mu m$	Proportion of particles with a size lower than 315 μm
$< 400\ \mu m$	Proportion of particles with a size lower than 400 μm
$< 500\ \mu m$	Proportion of particles with a size lower than 500 μm
SR	Solid rate
ρ	Density of the slurry
Volume Beni Amir	Volume of phosphate token from Beni Amir washing station
Volume MEA	Volume of phosphate token from MEA washing station
Volume Daoui	Volume of phosphate token from Daoui washing station

Note that along with this experimental work, there are also several experiments that should be run before the transport occurs for security reasons. The objective of these experiments is to identify the physico-chemical composition of the slurry which includes the parameters identified in Table 1. Given the considerable variability in the estimation of the constitutive equation for the phosphate slurry, we suggest in this study to use a Machine Learning (ML) based algorithm that allows to predict the elasticity and the viscosity coefficients. The ML is trained from the available field experiment results from the last two years in order to improve the accuracy in the estimation of the phosphate

rheology. Since there is a direct link between the friction at the pipeline and the physico-chemical composition of the slurry, we consider the parameters described in Table 1 as inputs for the proposed ML algorithm.

3. Data-driven polynomial chaos expansions

In this section we describe the general methodology used to develop a ML based model to predict the rheology of the phosphate slurry. We briefly discuss techniques employed for data-driven polynomial chaos expansions. Details on the application of these tools for predicting the rheology are also included in this section. In the current work, the proposed ML tools are also used to derive sensitivity indexes.

3.1. Polynomial chaos expansions

The Polynomial chaos Expansion (PCE) has been intensively used as a surrogate model in the context of uncertainty quantification. It aims at reproducing the global behavior of a simulation following a polynomial decomposition. The multivariate polynomials that form the basis are chosen according to the probability density function of the considered stochastic input variables as defined for example in [42]. Indeed, the PCE is a spectral method designed to find a certain deterministic function between the input random variables $\mathbf{X} \in \mathbb{R}^{\mathbf{d}}$ and a target random variable $\mathcal{Y}$, where d is the number of input parameters. In the present study, the variables contained in the vector $\mathbf{X}$ are the variables described in the Table 1 whereas, $\mathcal{Y}$ represents the apparent viscosity and elasticity coefficients. Thus, one can write

$$\mathcal{Y} = \mathcal{M}(\mathbf{X}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^{\mathbf{d}}} \beta_{\boldsymbol{\alpha}} \boldsymbol{\Psi}_{\boldsymbol{\alpha}}(\mathbf{X}), \tag{2}$$

where $\beta_{\boldsymbol{\alpha}}$ are the spectral coefficients of the decomposition to be determined and $(\boldsymbol{\Psi}_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathbb{N}^d}$ are the polynomial basis. Here, the component α_i in the multi-index $\boldsymbol{\alpha} \in \mathbb{N}^d$ is the polynomial degree of $(\boldsymbol{\Psi}_{\boldsymbol{\alpha}})$ in the ith element of $\mathbf{X}$. Hence, the total polynomial degree of $\boldsymbol{\Psi}_{\boldsymbol{\alpha}}$ is $|\boldsymbol{\alpha}| = \sum_{i=1}^{d} \alpha_i$. In order to ensure the convergence of the decomposition (2), all the random variables considered in the study (including $\mathcal{Y}$ and all the components of $\mathbf{X}$) need to be of finite variance, see for instance [43]. Note that building a PCE requires three main ingredients: (i) choosing the right orthonormal basis, (ii) truncate the infinite sum in (2) and (iii) compute the spectral coefficients.

The selection of the orthonormal polynomial basis in the context of PCE is driven by the probability density function of the inputs variables. Originally, this has been achieved by using homogeneous chaos theory considering Hermite polynomials as the orthogonal basis because the inputs were considered to follow a Gaussian law, compare [30] among others. Following ideas proposed in [44], the homogeneous chaos theory has been generalized for other classical distributions, see for instance [42]. This approach is known in the community as the *generalized polynomial chaos (gPC)*. Yet, for many applications where data are available, their density probability functions do not follow classical distributions. For this purpose, several research studies have generalized the PCE for arbitrary distributions allowing therefore a data-driven PCE, see for example [45, 46]. Other methods used to derive data-driven PCE have

also been intensively presented and discussed in [23]. In the present work and given the small size of the available data, various distributions are derived from the data using a kernel density estimation allowing to have a non-parametric probability density function as discussed in [47]. Indeed, for a given set of n observations $\left\{X_i^{(1)}, \ldots, X_i^{(n)}\right\}$ for a random parameter $X_i \in \mathbf{X}$, the kernel density defines an estimation $\hat{f}_{X_i}$ of the probability density function f_{X_i} inferred from this set as

$$\hat{f}_{X_i}(x) = \frac{1}{nh} \sum_{j=1}^{n} k\left(\frac{x - x_j^{(i)}}{h}\right), \tag{3}$$

where h is an appropriate bandwidth and $k(\cdot)$ is the kernel function. There exists many kernel functions used in practice but only the Gaussian kernel is used in the present work as it has been demonstrated to fit correctly complex probability density functions. Once $\hat{f}_{X_i}$ is given for each component of $\mathbf{X}$, one can build a corresponding orthogonal polynomial basis $\left(\phi_{\alpha_i}^{(i)}\right)_{\alpha_i \in \mathbb{N}}$ with respect to the estimated probability density function using Stiltjes or Gram-Schmidt orthogonalization, compare [31, 45] among others. Once the univariate polynomial basis is obtained for each component of $\mathbf{X}$, the multivariate polynomial basis is reconstructed using the tensor product of the d univariate polynomials as

$$\boldsymbol{\Psi}_{\boldsymbol{\alpha}}(\boldsymbol{X}) = \prod_{i=1}^{d} \phi_{\alpha_i}^{(i)}(X_i)\,. \tag{4}$$

Note that this results comes from the fact that all the random variables are independent. This assumption should be verified upon building the PCE and in the case of mutually dependent inputs, the PCE could still be achieved, we refer the reader to [23] for a deep discussion about these methods. In practice, the sum in (2) is truncated to a finite series as

$$\mathcal{Y} \approx \sum_{\boldsymbol{\alpha} \in \mathcal{A}^{d,p} \subset \mathbb{N}^d} \beta_{\boldsymbol{\alpha}} \boldsymbol{\Psi}_{\boldsymbol{\alpha}}(\mathbf{X})\,. \tag{5}$$

Here, the set $\mathcal{A}^{d,p}$ is usually defined based on a total polynomial degree threshold such that the total polynomial degree of the sum in (5) does not exceed the value p. The classical way to define this set is

$$\mathcal{A}^{d,p} = \left\{\boldsymbol{\alpha} \in \mathbb{N}^d : \qquad |\boldsymbol{\alpha}| \leq p\right\}. \tag{6}$$

Note that, when defining the set of the multi-indices following this strategy and for a fixed value of the maximum polynomial degree p, $\mathcal{A}^{d,p}$ contains $\binom{d+p}{p}$ elements following the binomial law. The main issue with this truncation scheme is that when the value of p is increased, many terms are needed compared to the very few available data which will lead to overfitting. Alternatively when the value of p is too small this may lead to underfitting. Fortunately, a hyperbolic truncation technique as defined in [29] could be applied to overcome this drawback. Following this method the set $\mathcal{A}^{d,p}$ is redefined as

$$\mathcal{A}^{d,p,q} = \left\{\boldsymbol{\alpha} \in \mathbb{N}^d : \qquad \|\boldsymbol{\alpha}\|_q = \left(\sum_{i=1}^{d} \alpha_i^q\right)^{1/q} \leq p\right\}, \tag{7}$$

where $q \in [0, 1]$. Note that this truncation scheme is especially useful when only few interactions influences the response and it helps to achieve a sparse polynomial expansion. Thus, estimating the right value of the hyperparameters q and p is achieved using a cross validation algorithm as in [29].

3.2. Estimation of the spectral coefficients

The determination of a PCE is therefore conditioned by the estimation of the spectral coefficients $\beta_{\boldsymbol{\alpha}}$. In the context of uncertainty quantification, there exist many methods used in the literature to compute the spectral coefficients, we refer the reader to [48, 24, 27] for detailed discussions on these methodologies. However in the context of ML, only the regression method is considered, see [23]. One can also use the most efficient sparse decomposition when the value of d is very high ($d \leq 5$). The regression method is based on solving a least-square (LS) minimization problem in some ℓ_2-norm to estimate the coefficients $y_{\boldsymbol{\alpha}_i}$, see for instance [49, 28]. In practice, we begin by defining an error $\boldsymbol{\epsilon}$ as the distance between the known value of the parameter to be predicted (the training set of $\mathcal{Y}$) and the PCE using the training set of $\mathbf{X}$. Here, we suppose that the size of the training set is N_{ls} and we define

$$\boldsymbol{\epsilon} = \boldsymbol{\mathcal{Y}} - \boldsymbol{\beta}^{\top}\boldsymbol{\Psi}, \tag{8}$$

where $\boldsymbol{\mathcal{Y}} = \left(\mathcal{Y}^{(1)}, \ldots, \mathcal{Y}^{(N_{ls})}\right)^{\top}$ is the vector of the output parameters in the training set. We also define $\boldsymbol{\beta} = \left(\beta_{\boldsymbol{\alpha}_0}, \ldots, \beta_{\boldsymbol{\alpha}_{N_{PC}-1}}\right)^{\top}$ as the vector of the $N_{PC} = \mathrm{Card}\left(\mathcal{A}^{d,p,q}\right)$ unknown coefficients (where $\mathrm{Card}(X)$ is the number of elements in the set X) and $\boldsymbol{\Psi}$ is the $N_{PC} \times N_{ls}$-valued matrix assembling the values of all orthonormal polynomials at the input training values $\boldsymbol{\Psi}_{ik} = \Psi_i\left(\boldsymbol{\zeta}^{(k)}\right)$, with $i = 0, 1, \ldots, N_{PC-1}$ and $k = 1, 2, \ldots, N_{ls}$. Estimating the set of coefficients $\boldsymbol{\beta}$ following the ordinary least-square functional (8) which is equivalent to minimize the following function

$$\mathrm{J}(\boldsymbol{\beta}) = \boldsymbol{\epsilon}^{\top}\boldsymbol{\epsilon} = \left(\boldsymbol{\mathcal{Y}} - \boldsymbol{\beta}^{\top}\,\boldsymbol{\Psi}\right)^{\top}\,\left(\boldsymbol{\mathcal{Y}} - \boldsymbol{\beta}^{\top}\,\boldsymbol{\Psi}\right). \tag{9}$$

This yields a standard well-established linear algebraic solution as

$$\boldsymbol{\beta} = \left(\boldsymbol{\Psi}^{\top}\,\boldsymbol{\Psi}\right)^{-1}\,\boldsymbol{\Psi}^{\top}\,\boldsymbol{U}. \tag{10}$$

Note that the equation (9) has a solution only when $N_{ls} \geq N_{PC}$ and the solution given by (10) is the solution of the problem:

$$\boldsymbol{\beta} = \underset{\tilde{\beta}}{\mathrm{argmin}}\ \mathrm{J}(\boldsymbol{\beta}).$$

It has also been proven that in the context of PCE, solving this problem may add a lot of coefficients while their respective values are negligible and a sparse regression is preferred in this case, see for instance [29]. Indeed, one can write the previous problem into its regularized form

$$\boldsymbol{\beta} = \underset{\tilde{\beta}}{\mathrm{argmin}}\ \mathrm{J}(\boldsymbol{\beta}) + \lambda \left\|\tilde{\boldsymbol{\beta}}\right\|_1.$$

Notice that the last term of this equation forces the regression to put all the samples energy such that only the most meaningful coefficients are retained. Thus, a sparse expansion is achieved and yet, the coefficient λ needs to be estimated. The Least Angle Regression (LAR) is an algorithm that allows to solve the previous equation without requiring an explicit optimization with respect to λ, see for example [50]. In the present study, as only few hundreds

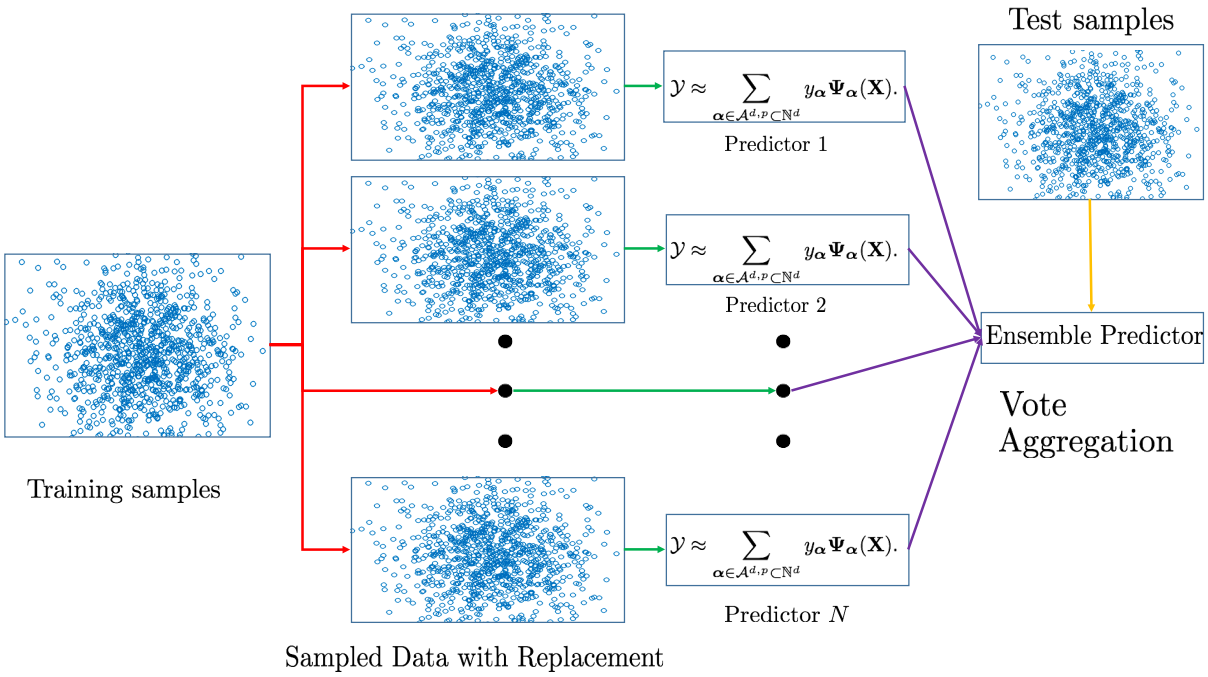

Figure 2: A schematic representation of the principles used for bagging in machine learning.

of samples of data are available, the LAR is considered along with a hyperbolic truncation strategy in order to estimate the spectral coefficients as described in [29].

It should be stressed that it is very common when using small data size, the accuracy of a ML method is very sensitive to the training sample. In these cases, when a change in the training sample gives ML outputs that are significantly different, the so-called bagging principle could be very useful, see [51] among others. In the present work, we propose to apply a bagging method over a data-driven PCE predictor. The main idea of this method is illustrated in Figure 2. First, multiple training sets are formed by making bootstrap replicates from the initial training sample. For each new training set, a predictor is obtained using the data-driven PCE. Finally, the ensemble predictor is estimated by aggregating the multiple versions of the predictors. This class of techniques has helped to define some highly accurate ML models such as the *random forests* which are basically regression trees to which bagging has been applied [52]. In the present work, the aggregation phase is obtained by averaging the results for the different predictors. Note that the selection of this method is motivated by its suitability for the operational requirements given its straightforward implementation.

3.3. Sensitivity analysis

In the uncertainty quantification, the Sensitivity Analysis (SA) is usually used to assess the relative importance of each input parameter on the output. There exist two different methods to assess the sensitivity namely, the local SA and

the global SA. The local SA aims at estimating the sensitivity when the value of the parameter is slightly perturbed. This is mainly achieved using the gradient or the adjoint methods. Alternatively, the global SA aims at assessing the resulting sensitivity when the input parameters are subject to the uncertainty. Only the global SA is retained for the purpose of this work. Again, there are several methods reported in the literature to assess the global SA of parameters, see [53] and further references are therein. The variance-based SA remains the most popular method. In fact, it is based on the analysis of the variance, also known as ANOVA. Hence, we write

$$Var(\mathcal{Y}) = V\left(\mathcal{Y}|X^1\right) + \cdots + V\left(\mathcal{Y}|X^n\right) + V\left(\mathcal{Y}|X^1, X^2\right) + \cdots + V\left(\mathcal{Y}|X^{n-1}, X^n\right) + V\left(\mathcal{Y}|X^1, \cdots X^n\right), \tag{11}$$

where $V\left(\mathcal{Y}|X^1\right)$ denotes the variance part of $\mathcal{Y}$ explained by the regressor X^1. Basically, this equation means that the variance of $\mathcal{Y}$ could be explained by adding a different contribution of different factors which we denote them by U. Moreover, the total variance is the sum of principals and the different interactions of the factors. Therefore, one can define a Sensitivity Index SI for a single or multiple factors as

$$SI_{X^U} = \frac{Var\left(\mathcal{Y}|X^U\right)}{Var(\mathcal{Y})}. \tag{12}$$

It is evident that the sensitivity index SI is bounded by 0 and 1 because the variance explained by a single or a set of factors is lower than the total variance of $\mathcal{Y}$. It is also clear that following (11), the sum over all indexes is equal to 1. The factors with the highest value of SI are the factors that have the highest impact on the response $\mathcal{Y}$. Consequently, these factors are used for monitoring as they are the most important in the ML model. Generally, when computing the sensitivity index one can either include only the principal effects of X^U (in this case the sensitivity index is referred to as a first order sensitivity index) or it could be computed using all the interactions in which X^U contributes (in this case it is referred to as a total sensitivity index). We refer the reader to [54] for a full review on the sensitivity index and on the methodologies used for their estimations. In the current study, in order to analyze the sensitivity we use the Sobol' indices [55]. The choice of this method is motivated by the fact that the spectral coefficients of the PCE allow to compute analytically the Sobol' indices as reported in [34, 56]. In fact, the spectral coefficient of the PCE can account for direct statistics of the model under study. Therefore, using some finite combinations of the spectral coefficients one can read the Sobol indices as well, see for instance [34]. Finally, since a data-driven PCE with bagging is used in this study, one expects to have as many estimations of the sensitivity indexes as the available predictors in the bagging, see Figure 2. In this situation, the estimation of these indexes follows the same aggregation rules as those used for the model. Therefore, the obtained sensitivity indices are the average of the Sobol' indices analytically obtained from each predictor. Here, we make use of the sensitivity analysis in order to emphasize the main factors from the physico-chemical composition of the phosphate slurry whose variation are driving the variation of the rheology and therefore, the uncertainty in its hydrodynamical properties. Once correctly estimated, these parameters need to be controlled during the transport of phosphate. It should stressed that all the computations used here were performed using UQLAB library [57] which allows to perform probabilistic and statistical modeling.

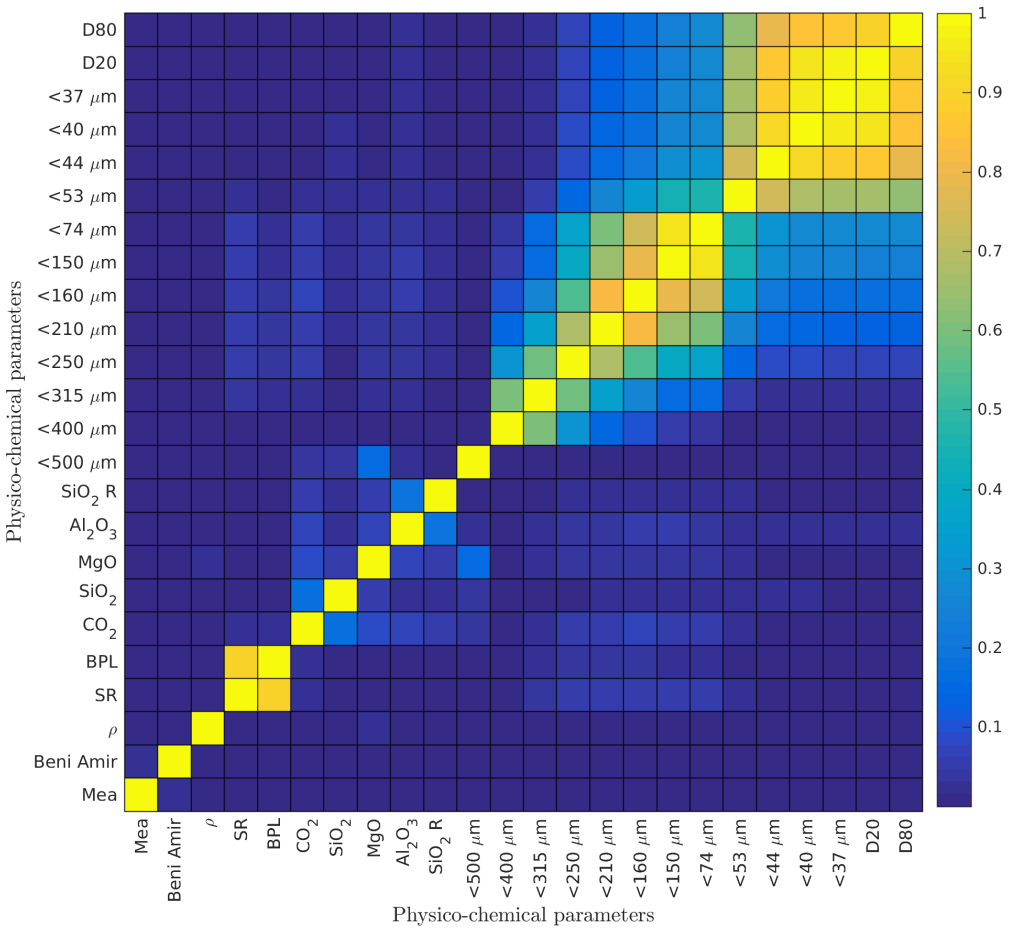

Figure 3: Correlation matrix between the different inputs in the present work.

4. Numerical results and discussions

We present results obtained for the available data with only 592 samples collected on the pipe for the past two years for the 25 features reported in Table 1. At the operational site, only one measurement is taken every day which explains the reduced number of the data available for the present study. These data are generally described as unbalanced data and may lead to the well-known problem of *overfitting*, see for example [58]. In this case, the selection of features and vectors is of utmost importance to build cost-effective models, compare [59] among others. Here, building a data-driven PCE using the LAR method to compute the spectral coefficient would guarantee an automatic vectors selection. In order to improve the machine learning method further, the selection of features is used here. Literature provides different methodologies used for the selection of features, see [60] for an overview. It is worth mentioning here that this topic is still subject to many open questions. However, given the small size of data used in this study, we choose to focus on the Spearman correlation coefficient. Indeed, the correlation analysis has been proven to provide relevant features in the case of machine learning, see for instance [61]. The choice of the Spearman correlation rather

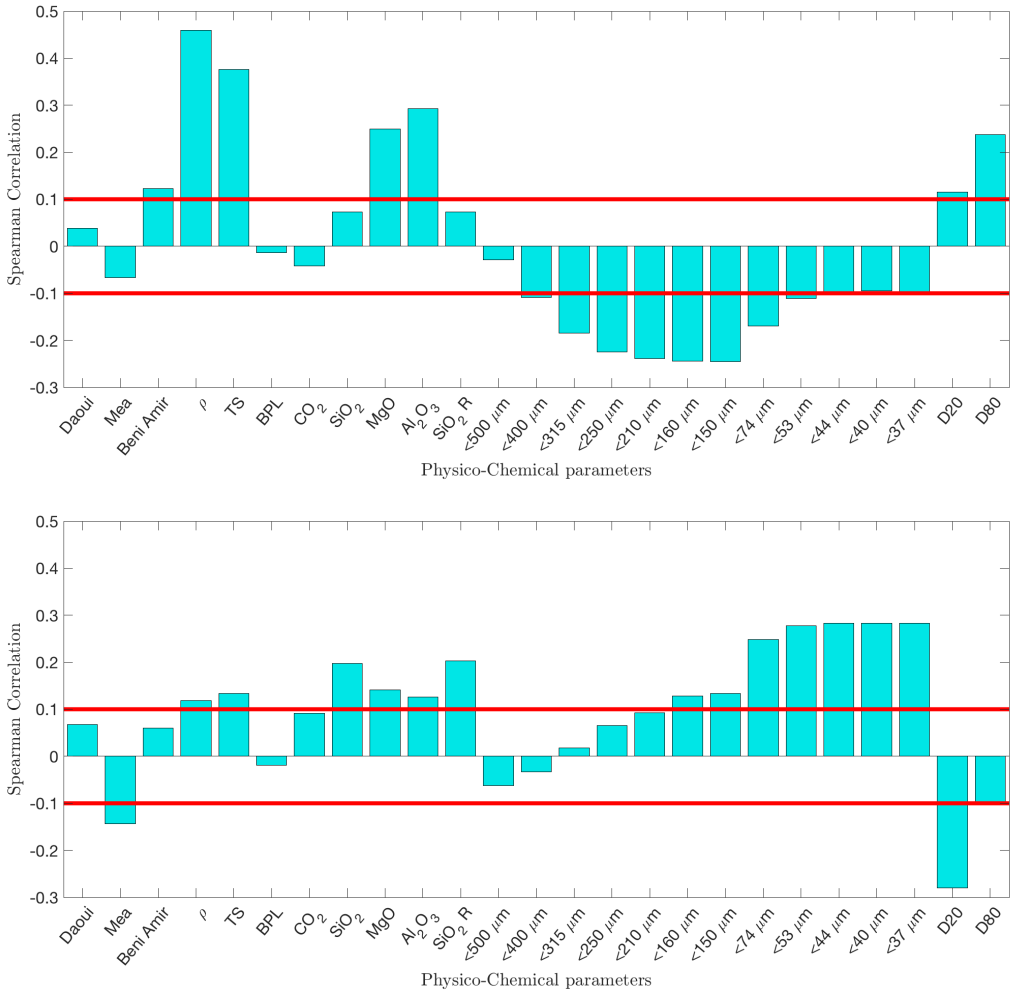

Figure 4: Spearman correlation coefficient for each of the features described in Table 1 computed for the elasticity (top plot) and for the viscosity (bottom plot) parameters as an output. Here, red lines represent the considered threshold for the feature selection.

than the Pearson is due to the fact that the relationship between the rheological parameters and the physico-chemical components is rather nonlinear. First the correlation between the different features is examined for the considered data and Figure 3 depicts the correlation matrix between the different inputs. Generally, the features could be considered as independent variables except for the grain size description.

In Figure 4 we present the results obtained for the Spearman correlation coefficient in both the elasticity and the viscosity parameters. Here, 60% of the selected features to predict the elasticity regression have a correlation coefficient higher than 0.2, while only 31% of these parameters have a correlation coefficient higher than 0.2 for the viscosity coefficient. This suggests that the ML models would work more efficiently for the elasticity than the viscosity coefficients. Furthermore, this analysis allows to reduce the number of features that one can use as regressors for the

Table 2: The selected features for each of the considered output parameters along with their values of the Spearman correlation coefficient.

Elasticity		Viscosity	
Parameter	Correlation coefficient	Parameter	Correlation coefficient
Volume Beni Amir	0.12	Volume MEA	-0.14
ρ	0.46	ρ	0.11
SR	0.38	SR	0.13
MgO	0.25	SiO_2	0.2
Al_2O_3	0.29	MgO	0.14
$< 400\ \mu m$	-0.1	Al_2O_3	0.13
$< 315\ \mu m$	-0.18	SiO_2 R	0.20
$< 250\ \mu m$	-0.22	$< 160\ \mu m$	0.12
$< 210\ \mu m$	-0.24	$< 150\ \mu m$	0.13
$< 160\ \mu m$	–0.24	$< 74\ \mu m$	0.25
$< 150\ \mu m$	-0.24	$< 53\ \mu m$	0.28
$< 74\ \mu m$	-0.16	$< 44\ \mu m$	0.28
$< 53\ \mu m$	-0.11	$< 40\ \mu m$	0.28
D20	-0.11	$< 37\ \mu m$	0.28
D80	-0.24	D20	-0.18
		D80	-0.10

PCE. It should also be noted that these regressors are different depending on the output parameter considered. Here, the selected features are those whose the absolute value of the Spearman coefficient exceeds 0.1, represented in Figure 4 by the red lines. In the case of the elasticity parameter, only 15 features are retained instead of the 25 described in Table 1. For completeness, these features are summarized in Table 2 with their corresponding correlation values. For the viscosity coefficient, only 16 features are retained instead of the 25 initial features and they are also included in Table 2. The results of this feature selection step was compared to the modern HISC method investigated in [62] among others. The obtained results reveal that the elasticity coefficient depends on 9 parameters namely: Volume Daoui, Volume Mea, Volume Beni Amir, ρ, MgO, Al_2O_3, SiO_2 R, $< 500\ \mu m$, $< 37\ \mu m$, D20. On the other hand, the viscosity coefficient depends on the parameters: Volume Daoui, Volume Mea, Volume Beni Amir, SR, $CO2$,SiO_2, MgO, Al_2O_3, SiO_2 R, $< 500\ \mu m$, $< 37\ \mu m$, D20. Overall, the results are comparable with those obtained using the Spearman correlation. However, the absence of the density for the viscosity coefficient and the solid rate for

the elasticity coefficient is something not reasonable from a dynamical perspective. This may be due to the lack of sufficient data. For this reason we choose to continue the present analysis with the feature selected using the Spearman correlation coefficient.

Since the PCE relies on the probabilistic description of the input parameters, the probability density function is estimated using the kernel smoothing technique in (3) for all the selected features given in Table 2. Note that evaluating these probability density functions helps in identifying the most suitable orthogonal polynomial basis for which the PCE is performed. These probability density functions are shown in Figure 5. In the present case, these functions do not have a common behavior such as the Gaussian or uniform measure (see the probability density function of Al_2O_3 and $< 315\ \mu m$ among others). This confirms that the ML techniques are very sensitive to the polynomial basis which can lead to further *overfitting* problems. For this reason, arbitrary PCE are used to build the ML model.

4.1. The assessment of data-driven PCE

Next, we consider the ML tools for the selected features based on the data-driven PCE. Given the small size of data available, we assess the sensitivity of different hyperparameters on the accuracy of the surrogate model. As mentioned in section 3, the sparse polynomial representation is preferred using a LARS method. The feature selection step helped reducing the number of hyperparameters needed here (including the spectral coefficients) but one of the main drawbacks of the PCE is related to this point especially, when the dimension is relatively large. For example, in the present step the feature selection helped into reducing the dimensionality to 15 and 16 for the elasticity coefficient and for the viscosity coefficient, respectively. This means that, if a polynomial approximation of degree 5 is needed, the total number of hyperparameters that needs to be estimated is $15,504$ and $20,349$, respectively. Given the small size of the considered data (549), the use of sparse polynomial approximation is required. Furthermore, we employ a hyperbolic truncation over the classical truncation scheme. This truncation scheme depends on the parameter q which describes the level of hyperbolicity as described in section 3.2. Thus, decreasing the value of q allows to have a sparse polynomial representation which help to overcome the problem of *Overfitting*. However, this means also that interactions between the different features are negligible on the output.Estimating a good value of q is very important as its value needs to be sufficiently high to include the complexity of the relationship and not too high to avoid the problem of *Overfitting*.

In order to assess the level of hyperbolicity, four different tests are carried out using $q \in \{0.25, 0.5, 0.75, 1\}$. For each test, the maximum polynomial degree is set using a cross validation by varying its values from 1 to 10. The available samples are split randomly such that 70% are used for training and 30% are used for validation. Moreover, each PCE is repeatedly built 20 times to assess the robustness of the model. The results of theses simulations are reported in Figure 6. We used a box-plot in order to illustrate the different statistics given by the 20 models estimated in each test. As discussed in section 3, two metrics of errors can be used to assess the accuracy of the model namely, the validation error and the Leave-One-Out (LOO) error. Both metrics are used for this study and the obtained results

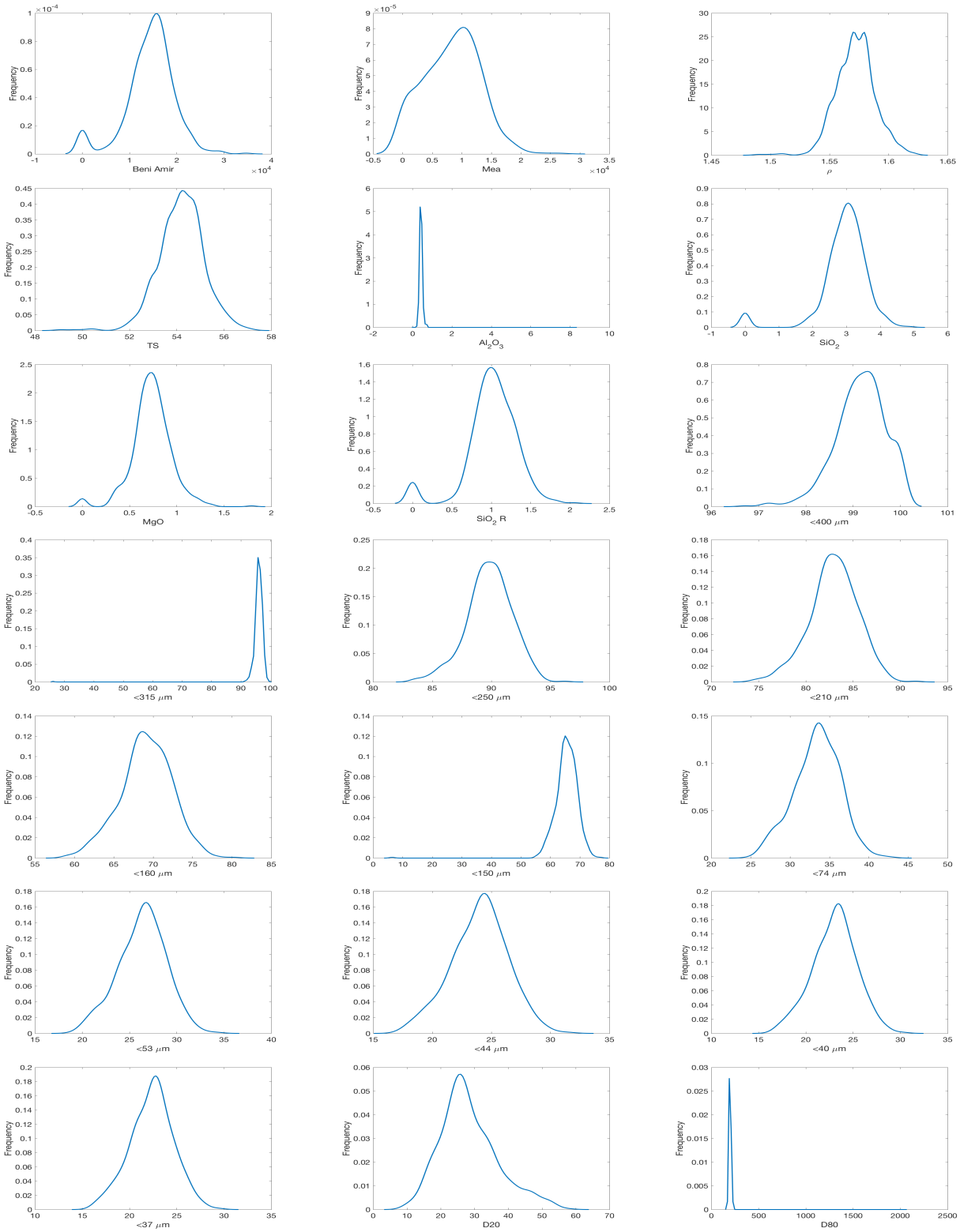

Figure 5: Probability density functions for all the features selected as regressors of the ML for the estimation of both elasticity and viscosity parameters in the present study.

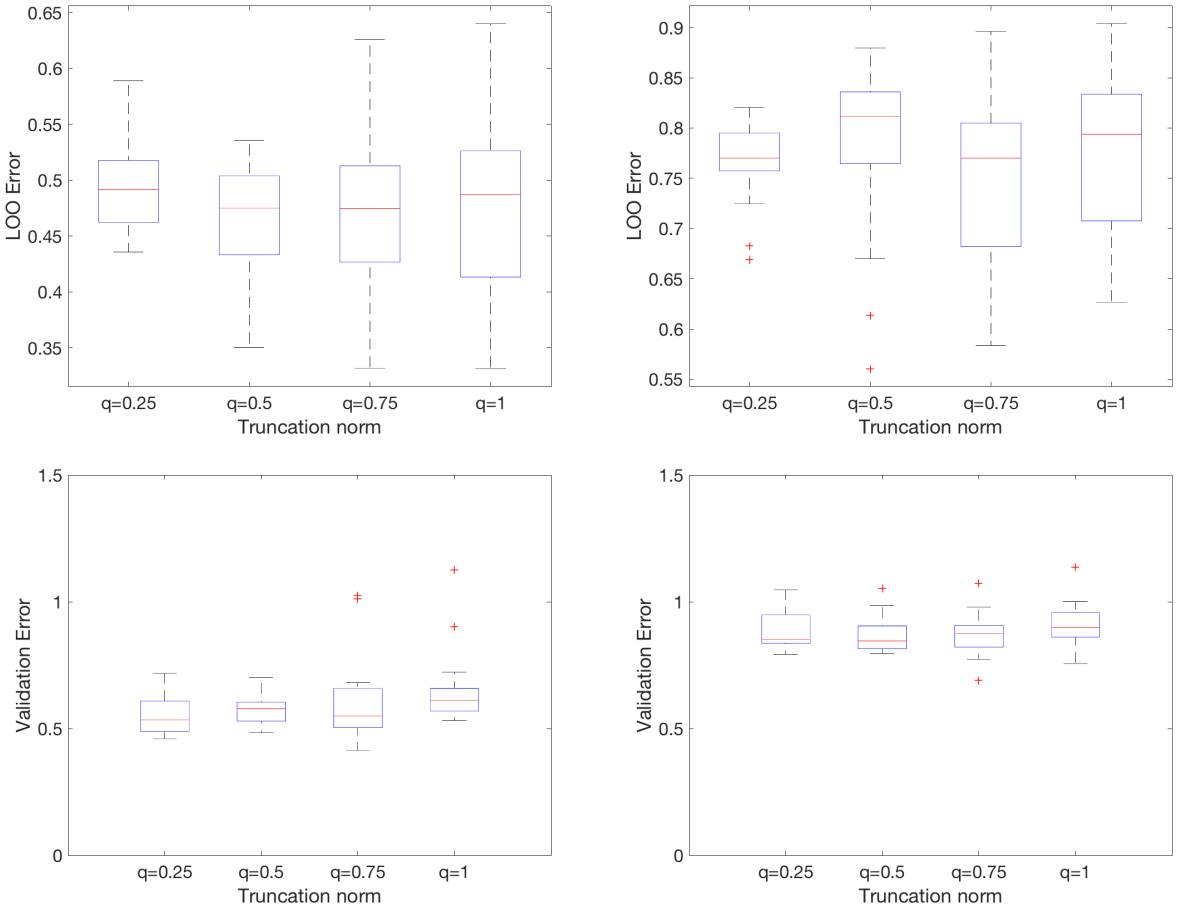

Figure 6: Sensitivity of the accuracy of the ML to the level of hyperbolic truncation using the validation error (bottom plot) and the LOO error (top plot) for the estimation of the elasticity coefficient (left plot) and viscosity coefficient (right plot).

reveal that the variance in errors of the different models built is very high. This is mainly attributed to the small size of data provided for this analysis. For the considered tests, the truncation norm $q = 0.75$ seems to give the most accurate results but it also displays the most variance in the results. In addition, one can see that the validation error and the LOO error seem to give the same error values. Finally, one can assess that the small size of the available data is real challenge here as the variability of the errors is important. Therefore, in order to reduce this effect, we suggest the use of the well-established Ensemble methods in the ML. More precisely, given the high value of the variance in the model performances in the present work, we use the well-established bagging technique.

In order to use a bagging based data-driven PCE as a ML tool, it is import to assess the ensemble size, referred to as N in Figure 2. Figure 7 exhibits the validation error function of the number N. One can see that the optimal Ensemble size is achieved for $N = 20$ since the error starts to have an oscillatory behavior after $N = 20$ which can be justified by the small size, the strong non-linearities or the heterogeneities of the available data. In general, bagging

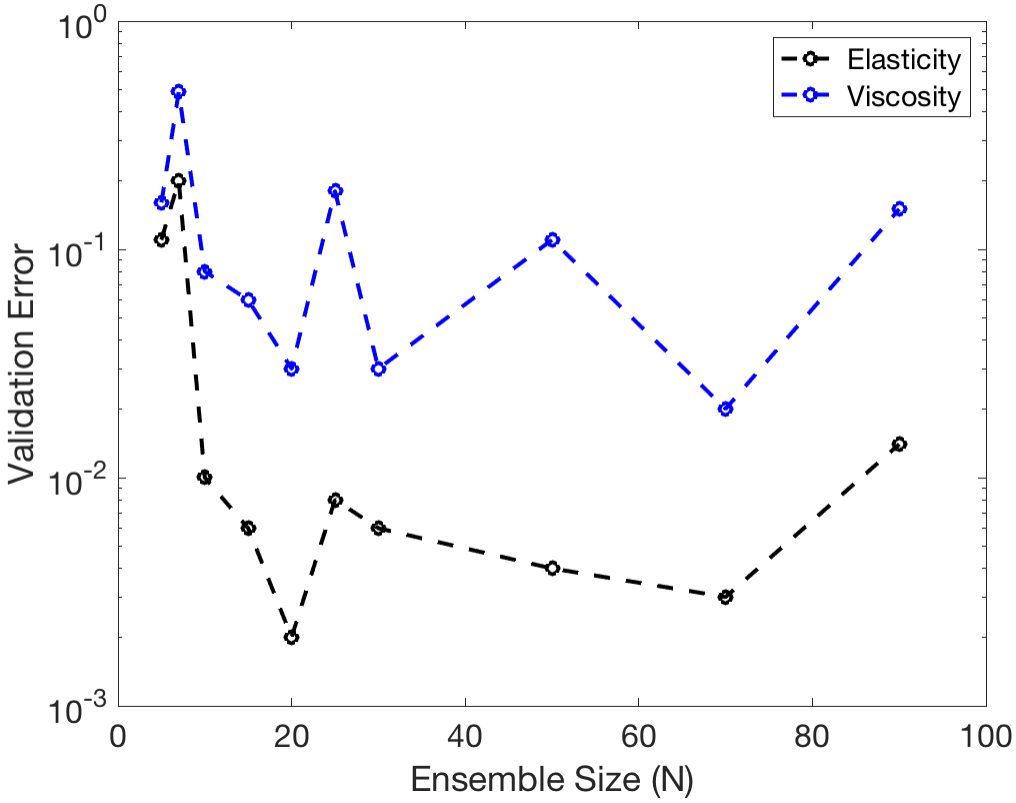

Figure 7: Sensitivity of the model accuracy to the ensemble size for the estimation of viscosity and elasticity coefficients.

allows to improve the accuracy of the ML model. In the case of elasticity estimation, bagging decreases the error by three order of magnitudes (around 0.5 in Figure 6 without bagging and nearly 10^{-3} in Figure7 with bagging) while, for the viscosity estimation it allows a decrease of two order of magnitudes (around 0.5 in Figure 6 without bagging and nearly 10^{-2} in Figure7 with bagging). It is also clear that the error for the viscosity coefficient is larger than the error for the elasticity coefficient. This supports our conclusions regarding the results on the selection of features as from the provided data, the correlation indicates a higher rate for the elasticity than the viscosity. It should also be noted that one way to improve the accuracy in this study is by the mean of increasing the observations. This yields to study the characterization of the rheology in the next section.

4.2. Sensitivity analysis of rheology

In this section we turn our attention to the sensitivity analysis of both hyperparameters of rheology namely, the elasticity and the viscosity, to the features selected above. Results of the Sobol' index are presented in Figure 8. Here, the top figure presents the first order Sobol' index while the bottom plot represents the total index. As it can be seen from this figure, both results exhibit similar behaviors which confirm that there are no significant effects due to

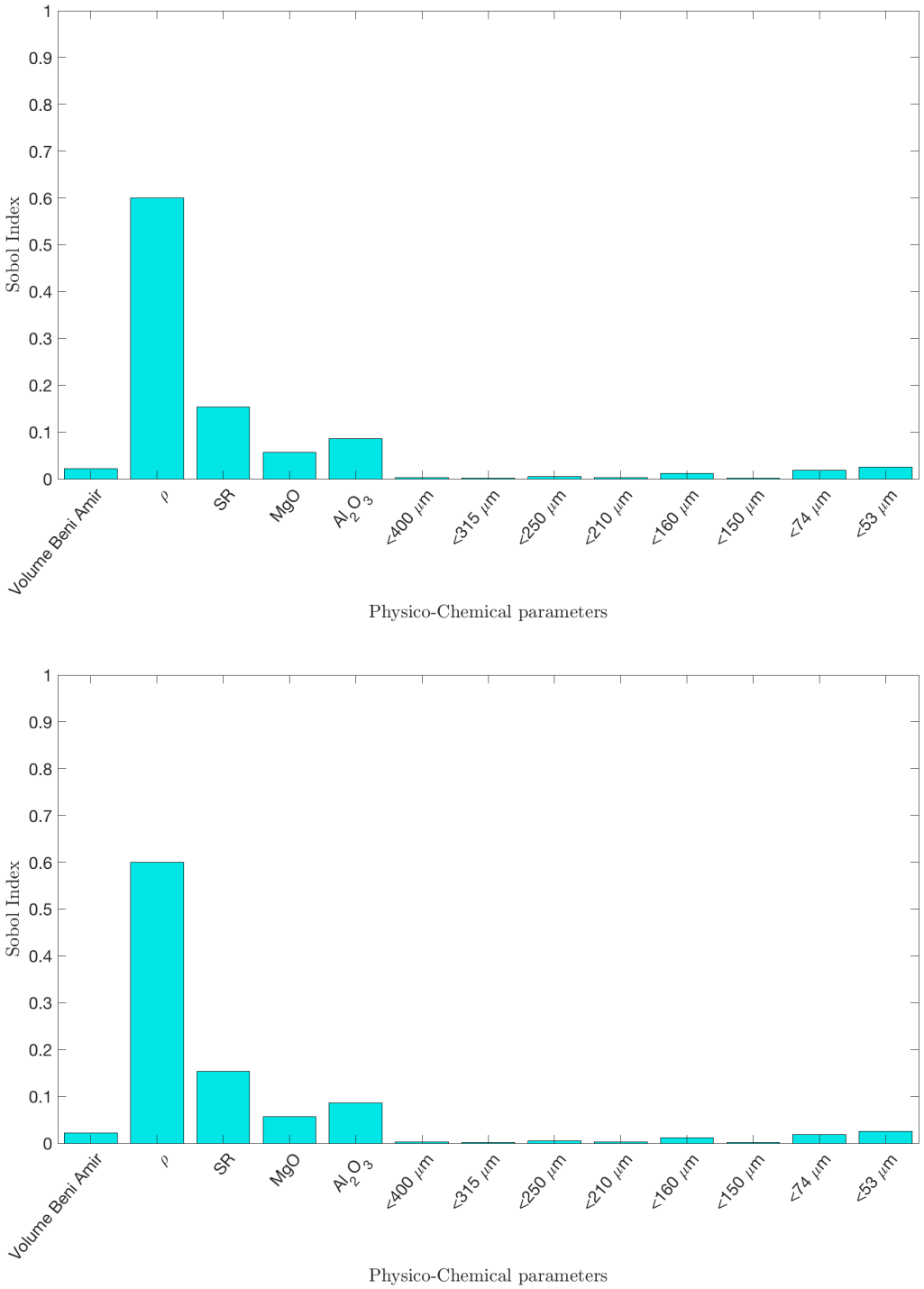

Figure 8: First-order Sobol index (top plot) and the total index (bottom plot) for the elasticity coefficient.

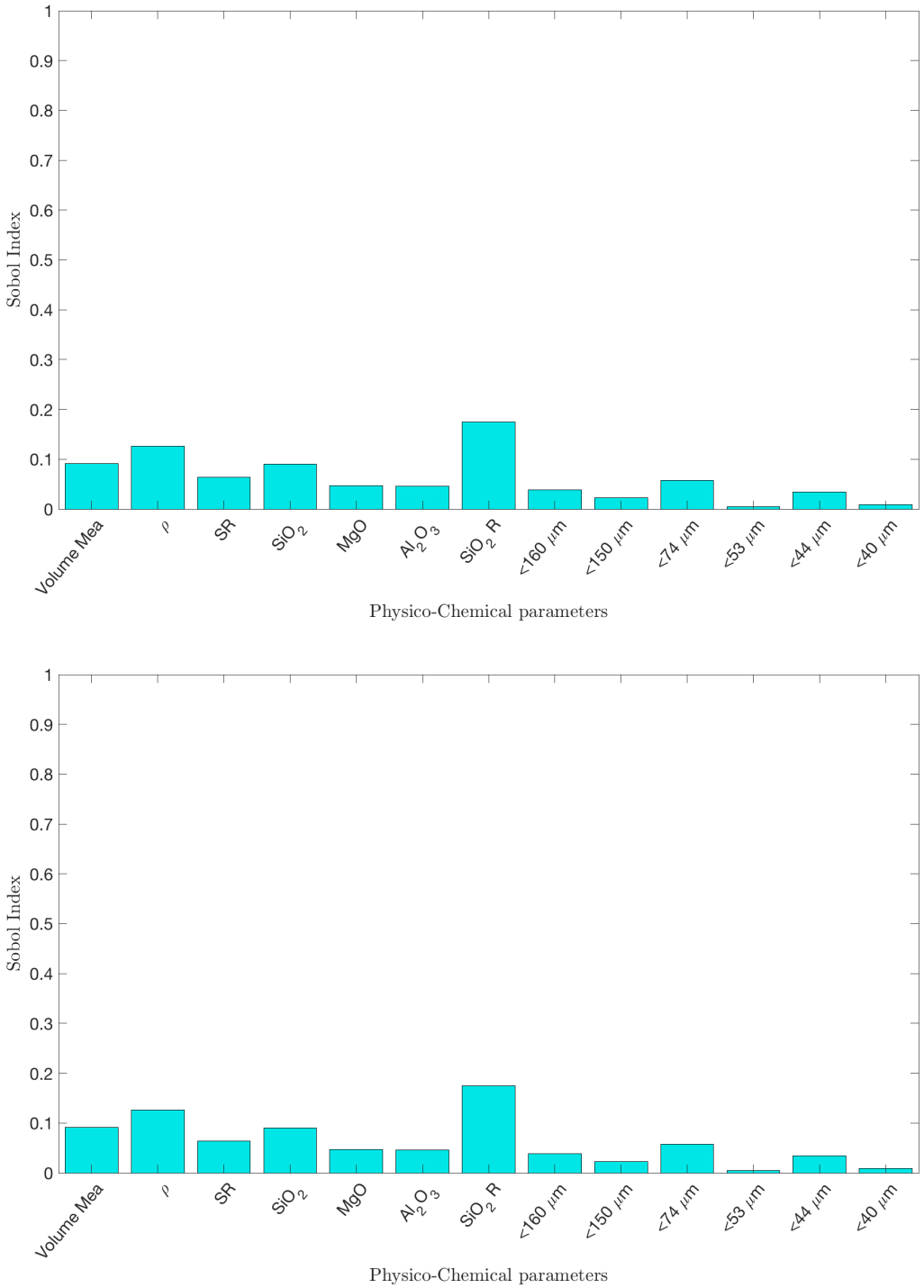

Figure 9: First-order Sobol index (top plot) and the total index (bottom plot) for the viscosity coefficient.

the interactions of the features. Furthermore, this sensitivity analysis demonstrates that the variation of the elasticity coefficient is mainly due to four parameters namely, the density of the fluid, the solid rate, the concentration of MgO and the concentration of Al_2O_3. Based on this analysis, one can state that the variability observed in the elasticity coefficient is mainly due to the variability in the density and the solid rate in the second position. In fact, the elasticity coefficient translates the dynamic effort needed to change the slurry from a solid phase to a fluid phase. This also explains the variability of the elasticity due to the variation of the density, the solid rate and the concentration of metal-based chemical elements. For the particular case of the Phosphate slurry, increasing the density by injecting more solid particles and the solid rate will lead to recover a solid mechanical behavior. In summary, the sensitivity analysis shows that to ensure a fluid dynamic like motion it is important to monitor the density and the solid rate of the slurry.

Similar sensitivity analysis is performed for the viscosity parameter and the obtained results for the first-order Sobol index and the total index are shown in Figure 9. As in the analysis of the elasticity coefficient, there is no significant change between Sobol index and the total index for the viscosity coefficient. These results demonstrate that the impact of the interaction between the physico-chemical parameters is negligible in the variability on the viscosity parameter. Unlike the previous case, there is no serious impact of a single parameter over the others that drives the variability on the viscosity coefficient. However, it is evident that the sensitivity indexes for the physical and the chemical parameters are more important than the granular structure of the slurry. As the viscosity parameter represents the loss of the dynamic due to the pipe friction, the chemical constitution of the fluid along with the density and the solid rate drive the energetic cost of the water-phosphate slurry. The present work paves the way into improving the understanding of complex fluids flows in pipelines. There are many ways one could think of in order to improve the results suggested in the present work. For instance, one can for example use the ML model as tools to predict the rheological laws and couple it with a hydrodynamic model. This will allow to understand how the other fixed parameters impacts the flow fields, such as the size of pipelines, the pipeline materials, the hydraulic machinery used at the operational site. Despite the results of the other rheological laws allowing to represent a non-Newtonian fluid and especially for the Phosphate slurry (see for example [10]), one could build the ML model based on these laws. Finally, a direct approach to improve the present work consist on looking for other parameters that are described with some stochasticity and were not taken into account here as added features in the ML model.

5. Conclusions

In the present work, a data-driven model has been used to study the variability impact of different physico-chemical parameters on the rheology of phosphate slurry in pipelines. The obtained results help to understand and identify the main components driving the dynamics of the phosphate slurry. Two rheological parameters were considered namely, the elasticity and the viscosity coefficients. In the presence of fluid with suspension, two mechanical behaviors can arise: fluid and solid. Elasticity parameter helps to define the threshold that makes the behavior move from solid to

fluid dynamics whereas, the viscosity parameter helps to understand the constraints that the fluid imposes to itself during the dynamics. Hence, studying these two properties is of enormous importance to optimize the transport of phosphate in pipelines and to reduce the energy cost as well as the water demand needed for the slurry. In order to achieve these objectives, a polynomial chaos based data-driven is built using a set of 512 samples subject to 25 features with their corresponding estimation of the elasticity and the viscosity as measured in the experimental laboratory at the industrial site. This constitutes a major challenge as the number of samples is very limited to build reliable data-driven models. First, the selection of features is carried out in order to reduce the number of important features using the Spearman correlation identification. Next, in order to further reduce overfitting, a vector selection is used to reduce the number of hyperparamters to the data-driven model using the least angle regression with a hyperbolic truncation scheme. This leads to build sparse polynomial expansions and yet, the results demonstrate that despite using all these advanced methods, the data-driven model exhibits tremendous variability on the output results. For this reason we introduced bagging which is a well-established ensemble method used in Machine Learning. This method is known to be efficient and easy to implement technique to reduce models variability. The introduction of bagging in the present case has helped to achieve reliable data-driven model and the validation error decreased by two orders of magnitude for the case of elasticity estimation and one order of magnitude in the case of viscosity estimation. This is mainly attributed to the fact that the viscosity is a very variable parameter and it also depends on the fluid dynamics occurring in the pipeline.

A sensitivity analysis has also been performed in this study supporting the selection of the PCE data-driven model over other ML methods. The sensitivity analysis is carried out for the considered data and the Sobol indices are analytically computed from the spectral coefficients. The computed results show that the variability of the elasticity parameter is mainly explained by the fluid density, the solid rate and the heavy metal chemical components (Magnesium and Allumine). For the viscosity parameter, there is not a single element that outperfoms the others regarding the sensitivity analysis. However, one major conclusion here is that the variability of this parameter is not highly affected by the variability in the granulometry distribution. In summary, this work paves the way towards understanding the dynamics of the phosphate slurry in pipelines. It should also be stressed that while the results of the present study seem to be encouraging, having more data will surely reduce the volatility of the ML tools. There are several ways in order to improve the current model when additional data are obtained such as the use of reinforcement learning. The use of the PCE based methods on arbitrary distributions relaxes the challenges given by heterogeneous data, the use of advanced methods of bagging can also definitely help into tackling this issue. One of the main perspectives of the present study includes the coupling of this data-driven model with hydrodynamical models in order to represent the flow taking into account the operational settings (such as the air compressibility, the pipe material and the pipe diameter among others). These parameters were not taken into account in the present model as they do not exhibit aleatory behavior. Furthermore, the present methodology could be adopted in several industrial activities such as metallurgy, plastic-based industry, food-processing, mine extraction, etc. Especially when data in materials, properties, parameters, exhibit a considerable amount of stochasticity. Data-driven PCE could be adopted in order to control the

stochastic parameters whose variability considerably impacts the variability of the quantity of interest.

Acknowledgment

Financial support provided by the Office Chérifien des Phosphates (OCP) Group is gratefully acknowledged. The authors wish to thank Mr. Nabil Al Atlassi from OCP Khouribga for providing experimental data. Discussions with operational engineers at the industrial platform OCP in Khouribga are also highly appreciated. The authors would also like to thank anonymous referees for giving very helpful comments and suggestions that have greatly improved this paper.

References

[1] C. Poza-Carrión, J. Paz-Ares, When nitrate and phosphate sensors meet, Nature plants 5 (4) (2019) 339.

[2] J. Cooper, R. Lombardi, D. Boardman, C. Carliell-Marquet, The future distribution and production of global phosphate rock reserves, Resources, Conservation and Recycling 57 (2011) 78 – 86.

[3] J. Elamrani, M. Lemtaoui, Creating shared value in moroccan companies: a case study of ocp sa, Transnational Corporations Review 8 (4) (2016) 280–288.

[4] S. Y. Avksentiev, P. Makharatkin, Influence of rheology on pressure losses in hydrotransport system of iron ore tailings, Journal of Industrial Pollution Control (2017) 741–748.

[5] F. Irgens, Rheology and non-newtonian fluids, Springer, 2014.

[6] A. Bartosik, Application of rheological models in prediction of turbulent slurry flow, Flow Turb. Combust. 84 (2010) 277–293.

[7] S. Boyaval, A finite-volume discretization of viscoelastic saint-venant equations for fene-p fluids, 2017, pp. 163–170.

[8] J. Favero, A. Secchi, N. Cardozo, H. Jasak, Viscoelastic flow analysis using the software openfoam and differential constitutive equations, Journal of non-newtonian fluid mechanics 165 (23-24) (2010) 1625–1636.

[9] R. Larson, Constitutive equations for thixotropic fluids, Journal of Rheology 59 (3) (2015) 595–611.

[10] H. Belbsir, K. El-Hami, A. SOUFI, Study of the rheological behavior of the phosphatewater slurry and search for a suitable model to describe its rheological behavior, International Journal of Mechanical & Mechatronics Engineering 18 (2018) 73.

[11] F. H. Marchesini, R. M. Oliveira, H. Althoff, P. R. de Souza Mendes, Irreversible time-dependent rheological behavior of cement slurries: Constitutive model and experiments, Journal of Rheology 63 (2) (2019) 247–262.

[12] K. P. Singh, A. Kumar, D. R. Kaushal, Experimental investigation on effects of solid concentration, chemical additives, and shear rate on the rheological properties of bottom ash (ba) slurry, International Journal of Coal Preparation and Utilization 0 (0) (2019) 1–14.

[13] H. P. Singh, A. V. Singh, D. R. Kaushal, Model for the rheological parameters of coking coal and water slurry with calcium hydroxide additive, Particulate Science and Technology 0 (0) (2019) 1–10.

[14] N. Mangesana, R. Chikuku, A. Mainza, I. Govender, A. van der Westhuizen, M. Narashima, The effect of particle sizes and solids concentration on the rheology of silica sand based suspensions, Journal of the Southern African Institute of Mining and Metallurgy 108 (2008) 237 – 243.

[15] C. Tao, B. Kutchko, E. Rosenbaum, M. Massoudi, A review of rheological modeling of cement slurry in oil well applications, Energies 13 (2020) 570.

[16] S. Kumar, M. Singh, J. Singh, J. P. Singh, S. Kumar, Rheological characteristics of uni/bi-variant particulate iron ore slurry: Artificial neural network approach, Journal of Mining Science 55 (2) (2019) 201–212.

[17] I. H. Witten, E. Frank, M. A. Hall, Practical machine learning tools and techniques, Morgan Kaufmann (2005) 578.

[18] F. Hussain, S. A. Hassan, R. Hussain, E. Hossain, Machine learning for resource management in cellular and iot networks: Potentials, current solutions, and open challenges, IEEE Communications Surveys Tutorials (2020) 1–1.

[19] F. Elmaz, zgn Ycel, A. Y. Mutlu, Predictive modeling of biomass gasification with machine learning-based regression methods, Energy 191 (2020) 116541.

[20] K. S. Kasiviswanathan, K. P. Sudheer, Methods used for quantifying the prediction uncertainty of artificial neural network based hydrologic models, Stochastic Environmental Research and Risk Assessment 31 (7) (2017) 1659–1670.

[21] G. Forman, I. Cohen, Learning from little: Comparison of classifiers given little training, in: J.-F. Boulicaut, F. Esposito, F. Giannotti, D. Pedreschi (Eds.), Knowledge Discovery in Databases: PKDD 2004, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 161–172.

[22] B. Sudret, S. Marelli, C. Lataniotis, Sparse polynomial chaos expansions as a machine learning regression technique, in: International Symposium on Big Data and Predictive Computational Modeling, ETH Zürich, 2015.

[23] E. Torre, S. Marelli, P. Embrechts, B. Sudret, Data-driven polynomial chaos expansion for machine learning regression, Journal of Computational Physics 388 (2019) 601 – 623.

[24] D. Xiu, G. E. Karniadakis, Modeling uncertainty in flow simulations via generalized polynomial chaos, Journal of Computational Physics 187 (1) (2003) 137–167.

[25] N. El Moçayd, S. Ricci, N. Goutal, M. C. Rochoux, S. Boyaval, C. Goeury, D. Lucor, O. Thual, Polynomial surrogates for open-channel flows in random steady state, Environmental Modeling & Assessment 23 (3) (2018) 309–331.

[26] N. Goutal, C. Goeury, R. Ata, S. Ricci, N. El Mocayd, M. Rochoux, H. Oubanas, I. Gejadze, P.-O. Malaterre, Uncertainty quantification for river flow simulation applied to a real test case: The garonne valley, in: Advances in Hydroinformatics, Springer, 2018, pp. 169–187.

[27] O. Le Maitre, O. Knio, Spectral Methods for Uncertainty Quantification, Springer, 2010.

[28] M. Berveiller, B. Sudret, M. Lemaire, Stochastic finite element: a non intrusive approach by regression, European Journal of Computational Mechanics/Revue Européenne de Mécanique Numérique 15 (1-3) (2006) 81–92.

[29] G. Blatman, B. Sudret, Adaptative sparse polynomial chaos expansion based on Least Angle Regression, Journal of Computational Physics 230 (6) (2011) 2345–2367.

[30] R. G. Ghanem, P. D. Spanos, Stochastic finite elements: a spectral approach, Courier Corporation, 2003.

[31] C. Soize, R. Ghanem, Physical systems with random uncertainties: Chaos representations with arbitrary probability measure, SIAM Journal on Scientific Computing 26 (2) (2004) 395–410.

[32] S. Oladyshkin, W. Nowak, Incomplete statistical information limits the utility of high-order polynomial chaos expansions, Reliability Engineering & System Safety 169 (2018) 137 – 148.

[33] E. Torre, S. Marelli, P. Embrechts, B. Sudret, A general framework for data-driven uncertainty quantification under complex input dependencies using vine copulas, Probabilistic Engineering Mechanics 55 (2019) 1 – 16.

[34] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, Reliab Eng Syst Safety 93 (7) (2008) 964–979.

[35] S. Dubreuil, M. Berveiller, F. Petitjean, M. Salaün, Construction of bootstrap confidence intervals on sensitivity indices computed by polynomial chaos expansion, Reliability Engineering and System Safety 121 (2014) 263 – 275.

[36] B. Iooss, P. Lemaître, A review on global sensitivity analysis methods, in: Uncertainty management in simulation-optimization of complex systems, Springer, 2015, pp. 101–122.

[37] E. Borgonovo, E. Plischke, Sensitivity analysis: A review of recent advances, European Journal of Operational Research 248 (3) (2016) 869 – 887.

[38] O. Asserin, A. Loredo, M. Petelet, B. Iooss, Global sensitivity analysis in welding simulationswhat are the material data you really need?, Finite Elements in Analysis and Design 47 (9) (2011) 1004 – 1016.

[39] D. Barthés-Biesel, A. Acrivos, The rheology of suspensions and its relation to phenomenological theories for non-newtonian fluids, International Journal of Multiphase Flow 1 (1) (1973) 1–24.

[40] E. Hinch, L. Leal, Constitutive equations in suspension mechanics. part 1. general formulation, Journal of Fluid Mechanics 71 (3) (1975) 481–495.

[41] A. Leonov, Constitutive equations for viscoelastic liquids: Formulation, analysis and comparison with data, in: D. Siginer, D. D. Kee], R. Chhabra (Eds.), Advances in the Flow and Rheology of Non-Newtonian Fluids, Vol. 8 of Rheology Series, Elsevier, 1999, pp. 519 – 575.

[42] D. Xiu, G. E. Karniadakis, The wiener–askey polynomial chaos for stochastic differential equations, SIAM Journal on Scientific Computing 24 (2) (2002) 619–644.

[43] O. G. Ernst, A. Mugler, H.-J. Starkloff, E. Ullmann, On the convergence of generalized polynomial chaos expansions, ESAIM: Mathematical Modelling and Numerical Analysis 46 (2012) 317339.

[44] R. Askey, J. A. Wilson, Some basic hypergeometric orthogonal polynomials that generalize Jacobi polynomials, Vol. 319, American Mathematical Soc., 1985.

[45] X. Wan, G. E. Karniadakis, Beyond wiener–askey expansions: Handling arbitrary pdfs, Journal of Scientific Computing 27 (1) (2006) 455–464.

[46] H. Lei, J. Li, P. Gao, P. Stinis, N. A. Baker, A data-driven framework for sparsity-enhanced surrogates with arbitrary mutually dependent randomness, Computer Methods in Applied Mechanics and Engineering 350 (2019) 199 – 227.

[47] E. Parzen, On estimation of a probability density function and mode, The annals of mathematical statistics 33 (3) (1962) 1065–1076.

[48] D. Xiu, Numerical Methods for Stochastic Computations: A Spectral Method Approach, Princeton University Press, 2010.

[49] S.-K. Choi, R. V. Grandhi, R. A. Canfield, C. L. Pettit, Polynomial Chaos expansion with Latin Hypercube Sampling for estimating response variability, AIAA journal 42 (6) (2004) 1191–1198.

[50] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, The Annals of Statistics 32 (2) (2004) 407–499.

[51] L. Breiman, Bagging predictors, Machine Learning 24 (2) (1996) 123–140.

[52] L. Breiman, Random forests, Machine learning 45 (1) (2001) 5–32.

[53] R. Ghanem, D. Higdon, H. Owhadi, Handbook of uncertainty quantification, Vol. 6, Springer, 2017.

[54] C. Prieur, S. Tarantola, Variance-based sensitivity analysis: Theory and estimation algorithms, Handbook of uncertainty quantification (2017) 1217–1239.

[55] I. M. Sobol', On sensitivity estimation for nonlinear mathematical models, Matematicheskoe modelirovanie 2 (1) (1990) 112–118.

[56] T. Crestaux, O. L. Matre], J.-M. Martinez, Polynomial chaos expansion for sensitivity analysis, Reliability Engineering & System Safety 94 (7) (2009) 1161 – 1172, special Issue on Sensitivity Analysis.

[57] S. Marelli, B. Sudret, UQLab: a framework for uncertainty quantification in MATLAB, in: Proc. 2nd Int. Conf. on Vulnerability, Risk Analysis and Management (ICVRAM2014), Liverpool, United Kingdom, 2014.

[58] V. López, A. Fernández, S. García, V. Palade, F. Herrera, An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics, Information Sciences 250 (2013) 113 – 141.

[59] A. Ng, Feature selection, l 1 vs. l 2 regularization, and rotational invariance, Proceedings of the Twenty-First International Conference on Machine Learning.

[60] R. Sheikhpour, M. A. Sarram, S. Gharaghani, M. A. Z. Chahooki, A survey on semi-supervised feature selection methods, Pattern Recognition 64 (2017) 141 – 158.

[61] S. Ji, U. Baek, M. Shin, Y. Goo, J. Park, M. Kim, Best feature selection using correlation analysis for prediction of bitcoin transaction count, in: 2019 20th Asia-Pacific Network Operations and Management Symposium (APNOMS), 2019, pp. 1–6.

[62] B. Iooss, A. Marrel, Advanced methodology for uncertainty propagation in computer experiments with large number of inputs, Nuclear Technology.