Showcasing research from Maciej Haranczyk's group, IMDEA Materials Institute, Madrid, Spain.

Machine-learning-accelerated multimodal characterization and multiobjective design optimization of natural porous materials

The practice of tinkering with clay materials dates back millennia. It can now receive a significant boost with the involvement of machine learning techniques. Herein, a collection of experimental material characterization results are turned into reliable quantitative predictors of experimental outcomes. Such predictors can then serve as surrogates of real experiments, and can be used in a high-throughput manner to search the material design parameter space quickly and at low cost. A new catalyst for the degradation of chlorophyll-a is achieved by applying the presented approach.

As featured in:

See Maciej Haranczyk *et al.*, *Chem. Sci.*, 2021, **12**, 9309.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/chemical-science

Registered charity number: 207890

Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Sci.*, 2021, **12**, 9309

All publication charges for this article have been paid for by the Royal Society of Chemistry

Machine-learning-accelerated multimodal characterization and multiobjective design optimization of natural porous materials†

Giulia Lo Dico, [abc] Álvaro Peña Nuñez, [c] Verónica Carcelén[c] and Maciej Haranczyk *[a]

Received 9th February 2021
Accepted 1st June 2021

DOI: 10.1039/d1sc00816a

rsc.li/chemical-science

Natural porous materials such as nanoporous clays are used as green and low-cost adsorbents and catalysts. The key factors determining their performance in these applications are the pore morphology and surface activity, which are typically represented by properties such as specific surface area, pore volume, micropore content and pH. The latter may be modified and tuned to specific applications through material processing and/or chemical treatment. Characterization of the material, raw or processed, is typically performed experimentally, which can become costly especially in the context of tuning of the properties towards specific application requirements and needing numerous experiments. In this work, we present an application of tree-based machine learning methods trained on experimental datasets to accelerate the characterization of natural porous materials. The resulting models allow reliable prediction of the outcomes of experimental characterization of processed materials (R^2 from 0.78 to 0.99) as well as identification of key factors contributing to those properties through feature importance analysis. Furthermore, the high throughput of the models enables exploration of processing parameter–property correlations and multiobjective optimization of prototype materials towards specific applications. We have applied these methodologies to pinpoint and rationalize optimal processing conditions for clays exploitable in acid catalysis. One of such identified materials was synthesized and tested revealing appreciable acid character improvement with respect to the pristine material. Specifically, it achieved 79% removal of chlorophyll-*a* in acid catalyzed degradation.

1. Introduction

Porous materials are extensively employed for their adsorption capacity and surface reactivity in a wide variety of industrial applications such as gas capture and separation,[1,2] water treatment,[3] catalysis,[4–7] and others.[8,9] In this material panorama, natural nanoclays, while being less porous, have the advantage of being environmentally friendly and low-cost.[10] Different types of functionalities have been conferred by modifications, among them, acid and alkaline treatments represent effective strategies for the preparation of inorganic nano-catalyzers.[10–12] Those processed materials find employment, *e.g.*, in organic synthesis,[13] biofuel production,[14] and biolubricant synthesis.[15]

Nanoporous clays are a quite diverse family of materials whose morphology and chemical structure vary depending on the natural environments and deposits.[16] The different structural growth patterns give rise to diverse particle aggregates among which, sepiolites and palygorskites exhibit a fibrous shape, and smectites are known to be layered nanosilicates.[10] Fig. 1A provides a photo of a powdered clay together with two structural models of fibrous sepiolite and lamellar smectite. The repeated inverted structure of fibrous nanoclays leads to higher surface area (125–195 $m^2\ g^{-1}$ for palygorskites and 230–320 $m^2\ g^{-1}$ for sepiolites) compared to smectites, which are 30–80 $m^2\ g^{-1}$ for the Al-rich and 150–300 $m^2\ g^{-1}$ for the Mg-rich ones.[10] Moreover, clay morphology is critically affected by the pervasive substitutions in the lattice structure and by exchangeable cations affecting the layer charge and swelling properties.[10,17]

The structural dissimilarities between various clays are reflected in the surface reactivity, which can be approximated by Brønsted and Lewis acid–base characters.[18–20] Generally, the active sites reside predominantly in the particle edges, in the hydroxylic termination of octahedral sheets and on the basal oxygen of tetrahedral sheets.[21,22] Fibrous nanoclays possess higher hydroxyl density compared to smectites whose reactivity is increased by their swelling properties and cation exchange capacity (CEC).[23,24] Furthermore, defects in the chemical

[a]*IMDEA Materials Institute, C/Eric Kandel 2, 28906 Getafe, Madrid, Spain. E-mail: maciej.haranczyk@imdea.org*

[b]*Department of Materials Science and Engineering and Chemical Engineering, Universidad Carlos III de Madrid, Getafe, Spain*

[c]*Tolsa Group, Carretera de Madrid a Rivas Jarama, 35, Madrid, Spain*

† Electronic supplementary information (ESI) available: Experimental methods, hyperparameter optimization and validation. See DOI: 10.1039/d1sc00816a

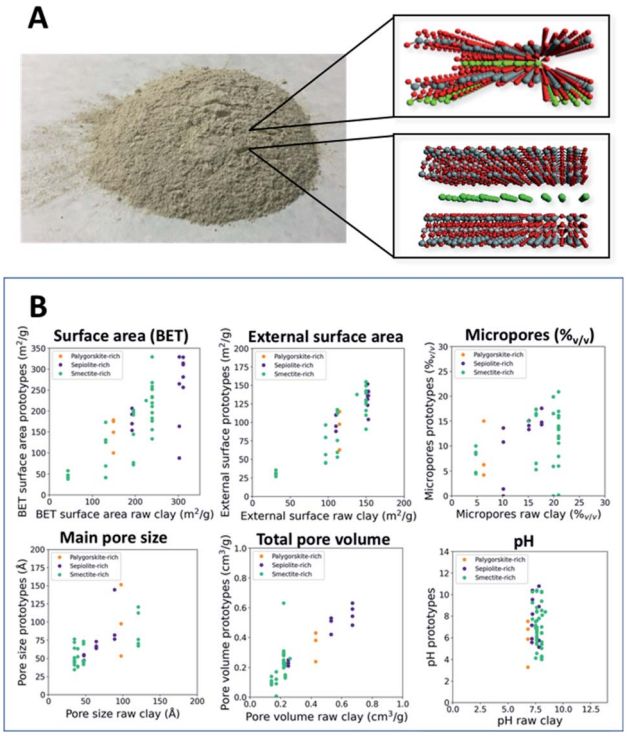

Fig. 1 Photo of a powdered clay together with two examples of fibrous sepiolite (top) and lamellar smectite (bottom) (A). Morphological properties and pH of processed materials (prototypes) compared to the corresponding features of the pristine, natural clays (B). Each raw material represents a discrete value on the horizontal axis of the plots.

structure may cause charge vacancies compensated by diverse ions influencing the catalytic behavior or the affinity with certain molecules.[17] Such a modification of natural nanoclays may drastically change the morphology and surface activity, and in some cases it may also lead to loss of crystallinity.[14,25,26] Fig. 1B further illustrates this tunability by highlighting ranges of property values achieved for various raw clay minerals with such treatments. The properties of processed materials shift from the corresponding ones for pristine materials as a function of modification parameters such as the nature of an additive, the additive-to-clay ratio, additive concentration and the contact time, requiring control and optimization for achieving the desired outcomes.[27] The latter is challenging due to the complexity of the parameter space as well as the cost of experimental work involved. Computer-aided approaches can provide solutions to potentially overcome these challenges.

A number of computational studies have investigated the effect of modification, lattice defects and impurity in specific material structures in the context of adsorption and catalysis.[17,28] However, the wide variability and imprecise nature of (modified) nanoporous clays makes such modeling approaches, which rely on precise atomistic models, ineffective for material screening.[17,29] Similarly, machine learning-supported property prediction approaches based on crystal structures such as the ones used in the discovery of zeolites and metal–organic frameworks cannot be easily adopted.[1,30–32] Instead, advanced design of experiments strategies have been exploited to address process optimization by performing only a small number of experiments.[33] For example, central composite design has been involved in the optimization of HCl treatment on montmorillonite,[34] which explored experimental factors such as temperature, contact time, acid–clay ratio and acid concentration. Similarly, response surface optimization sheds light on the relationships among various experimental process parameters (time, acid concentration, and microwave heating power) and the specific surface area of montmorillonite.[27] The results indicated the optimum parameters for the modification of a specific clay mineral, however, more sophisticated methods are needed to account for the high variability of properties in pristine materials coming from various deposits.

In this contribution, we incorporate machine learning algorithms into surrogate models of experimental characterization outcomes, and then employ the latter to accelerate the exploration of the morphological parameters and surface activity of clay-based materials achievable through processing of raw materials. We demonstrate how the throughput of the models can be harnessed in multiobjective optimization of materials. In particular, we focus on a design function relevant to acid catalysis applications and verify its utility by synthesis of the identified material. The promising acid nano-catalyzer was characterized and tested in terms of performance in the catalytic degradation of chlorophyll-*a*. The latter has scientific and industrial significance, *e.g.*, in pharmaceutical applications and in the production of biofuels.[35,36]

2. Methods

2.1. Dataset, descriptors, and target properties

We collected experimental characterization results of both raw and processed materials based on 9 diverse clays of different grades of purity, *i.e.*, five pure minerals (phyllosilicate > 70%) corresponding to five classes of materials, namely palygorskite, sepiolite, montmorillonite, saponite and stevensite, and four stevensites coming from different deposits exhibiting a lower phyllosilicate content (from 70% to 40%, with the remainder of impurities of various natures). 86 processed materials were prepared through several modifications such as acid, basic or organic treatment. Their morphology and/or surface activity were characterized by obtaining two datasets of 49 and 69 materials, respectively, sharing 32 materials.

The feature vector space for material representation was defined by rational selection and/or implementation of 41 descriptors. To facilitate the analysis, the descriptors were aggregated into three main families, namely, the properties of raw clays, the characteristics of the additives and the parameters describing the modification process (summarized in Table 1). The raw clay features were experimentally measured (details in Section S1, Fig. S1 and Table S1†). The descriptors for the additives were obtained from PubChem, and correspond to the acid dissociation constants (pK_a), molecular weight (MW), molecular formula (C, H, O, S counts), number of hydrogen donor (H-donor) and acceptor (H-acceptor) groups, number of double and rotatable bonds, octanol–water partition

Table 1 Input feature descriptors selected to address the design of hierarchical porous materials

Raw clay	Additive	Modification process
Cation exchange capacity (CEC)	pK_{a1}	Activation
	pK_{a2}	Milling time
Surface area (BET)	pK_{a3}	Additive/clay$\%_{g/g}$
Acid–base character (3 features)[a]	C, H, O, S counts	Additive (M)
	# Double bonds	$RH\%_{g/g}$
Chemical composition (10 binary features)[b]	Molecular weight	Final $RH\%_{g/g}$
	$X \log P$	Particle size
Phyllosilicate composition (5 features)[c]	H-donor	
	H-acceptor	
	Rotatable bond	
	Polar surface	

[a] pH measured at 0 and after 24 h (pH_0, pH_{24}) and free acidity. [b] SiO_2, Al_2O_3, MgO, CaO, Fe_2O_3, Na_2O, K_2O, TiO_2 and Mn_2O_3 and loss by calcination. [c] Relative content of fibrous, planar phyllosilicates, dolomite, calcite, and quartz.

coefficients ($X \log P$) and topological polar surface area.[37] In case a material datapoint corresponded to a raw material, a neutral value of 7 was assigned to the pK_a, while 0 was given to the remaining additive descriptors. The parameters describing the modification process were defined by assigning unique ascending integer values for the type of mixing between additives and raw clays (activation), *i.e.* not activated, by spraying, or by mixing them into a dilute solution or a solid mixture.

The milling time is classified with integer values from 0 to 2 for, respectively, after, before modification and double milling (before and after modification).

Then, the additive concentration in water and the amount with respect to dry clay (additive (M) and additive/clay ($\%_{g/g}$) respectively), as well as the starting ($RH\%_{g/g}$) and final moisture (final $RH\%_{g/g}$), and the particle size ($\%_{g/g}$ of <45 μm particles) were introduced completing the vector space. The two datasets were normalized giving the same weighted contribution at all the information used. The dataset splitting was fixed as 85% of training data and 15% of testing data. Furthermore, we ensured that the datasets are a good representation of the material space captured by the above feature vectors, see the ESI, Section S2, Fig. S2 and S3.†

The targets are represented by the main morphological properties calculated by N_2 physisorption isotherms at 77 K (details in the ESI, Section S1, Table S1 and Fig. S1†) and the surface activity which is herein assessed by measuring the pH.[38,39] Table 2 summarizes the morphological targets, with their corresponding abbreviations, which were experimentally assessed for every natural or processed clay-based material contained in the two datasets.

2.2. Models

Six independent machine learning models were built for prediction of each of the targets listed in Table 2. The models are summarized according to eqn (1):

$$y\_i = f(x_{\text{raw\_clay}}, x_{\text{additive}}, x_{\text{modification\_process}}) \quad (1)$$

where i indicates the target of Table 2 while $x_{\text{raw\_clay}}$, x_{additive}, and $x_{\text{modification\_process}}$ are the aggregated feature vectors corresponding to Table 1. We employed the Extremely Randomized Trees (Extra Trees) regressor algorithm which was assessed against the simplest Decision Tree, Random Forest regressor, and a more complex deep learning model Multiple Layer Perceptron (MLP).[30,40–42] All models were implemented using the Scikit-learn library except for the MLP model that was implemented in TensorFlow and Keras framework. After splitting the dataset, and leaving 15% of the points aside for testing, a detailed grid search for optimal values of hyperparameters was performed involving 10-fold cross-validation (details in Section S3† and Table 3). Table S2† summarizes the cross-validation accuracy in terms of R^2 and mean absolute error (MAE). The obtained scores are reasonable considering the small data size. The lowest scores seen for pMS reflects the contribution of error propagation, *i.e.* being calculated as 4Vol/SA, where Vol and SA are the targets of pore volume and surface area of Table 2. Moreover, since the MS is calculated from the adsorption branch of the N_2 physisorption isotherms, accurate estimation of pore geometry is lost leading to an overestimation or underestimation of the real pore size. On the other hand, the MAE of 22 Å is consistent with the experimental error, observed at high MS values. Once the optimal hyperparameters were identified, the final models were assessed using the 15% datapoints (*i.e.* test set). The accuracy was expressed in terms of the R^2 score, MAE and mean squared error (MSE). The effect of the dimensionality reduction on the model predictability was also

Table 2 Assignment of targets and the corresponding abbreviations

Target class	Target	Abbreviations of the predicted targets
Morphology	Surface area (SA)	pSA
	External surface area (ESA)	pESA
	Micropore content (Micro)	pMicro
	Main pore size (MS)	pMS
	Total pore volume (Vol)	pVol
Surface activity	Free acidity (pH)	ppH

Table 3 Optimized hyperparameters obtained by grid search performed on cross-validation ($K = 10$) on the training set

Hyperparameter	pSA	pESA	pVol	pMicro	pMS	ppH
n_estimators	2000	2000	2000	2000	2000	2000
min_samples_split	2	4	3	3	4	3
min_sample_leaf	1	1	1	1	1	1
max_features	36	9	8	27	18	38
max_depth	100	900	900	600	300	900

tested by reducing the group of features of raw_clay, being the most extended one (20 descriptors), to 4 main principal components (details in ESI Section S4, Fig. S4–S6 and Table S3†). We observed an increment of the predictability for the main pore size (pMS model R^2 from 0.77 to 0.87) while slight variability was observed in the prediction of pSA and ppH. The remaining targets (pESA, pVol, pMicro) suffered a predictability deterioration up to 10% of R^2 score. To facilitate the interpretability of the models, we employed the original vector space of Table 1. Our code and dataset are available at https://doi.org/10.5281/zenodo.4742294.

2.3. Design function and its application workflow

Development of materials towards a specific application often requires finding the balance between competing properties contributing to the overall performance. Herein, our goal is to improve the physico-chemical characteristics suitable for acid catalysis application. In this scenario, it is desirable to maximize the surface area while minimizing the micropores, which are not directly available for the substrate, and the pH of the material. We formulate the corresponding design function (DF) as:

$$\mathrm{DF} = \mathrm{pSA} \times \mathrm{pMicro}^{-1} \times \mathrm{ppH}^{-1} \quad (2)$$

The design function involves predicted properties of BET surface area (pSA), micropore content (pMicro), and pH (ppH). Sulfuric acid was chosen as a promising additive for acid modification. The search for the optimum value of the design function was performed into the entire feature vector space corresponding to the modification process features (7 descriptors). We then simplified the space to two-most important descriptors, *i.e.* H_2SO_4/clay% and RH%, while fixing the other five modification process descriptors to the average of single optimum value for the five material classes. In particular, the models were run on a set of proposed materials considering the 5 clay classes under different starting moisture contents (6–30%) and additive/clay ratios for H_2SO_4. The other parameters were fixed to: 16% of the final moisture, additive–clay activation being achieved using a spray diffuser and double milling (before and after the activation), and the amount of particles smaller then 45 μm was fixed to 75%. The DF improvements with respect to the starting values were calculated as $(\mathrm{DF_{proc}} - \mathrm{DF_{raw}})/\mathrm{DF_{raw}} \times 100$, where $\mathrm{DF_{proc}}$ and $\mathrm{DF_{raw}}$ are the design functions of the processed and raw nanomaterials, respectively. The promising acid nano-catalyzer extracted using DF was prepared and tested in the catalytic degradation of chlorophyll-*a* (experimental details in Section S5†).

3. Results and discussion

3.1. Model assessment

Out of the considered model frameworks, Extra Trees regressor has exhibited the highest accuracy (details in the ESI, Section S3, Fig. S7 and Table S4†), and its results will be discussed within this article. Fig. 2 and Table 4 summarize its predictive ability and accuracy, which are high enough to employ the model in material design. With the exception of pMS, the models exhibit R^2 higher than 0.93. The lower R^2 value for pMS is likely caused by the error propagation, since it is calculated as 4Vol/SA, and the approximation of a cylindrical pore shape.[38] This assumption loses meaning when planar clays are considered, enhancing the errors.

3.2. Feature importance analysis of morphological property models

Fig. 3 summarizes the aggregated group of features by summing up the singular contribution of the 41 descriptors according to Table 1 (the latter being individually displayed in Fig. S8†). Further analysis indicated that just 3 least important descriptors could be removed without appreciable deterioration of predictive power (variation of more than 5% in terms of MSE), implying that the complexity of the problems needs to be defined by the whole vector-space selected. pSA, pESA and pVol mostly depend on the starting properties of the raw minerals, while pMicro and pMS are strongly influenced by the modification process (70.5% and 64.8% respectively). These findings indicate competitive effects occurring within the modification. The activation by strong acids opens the pores but at the same time dissolves organic impurities (*i.e.* loss by calcination descriptor) shifting the main pore size to lower values.[43] In contrast, organic modifiers may cover the smallest pores but also promote the cation exchange and intercalation.[43,44]

Furthermore, Fig. S8† indicates that $X \log P$ (as an indirect measure of the affinity with the polar clay surfaces), the molecular weight with its chemical formula as well as the acid strength are key features of the additive, while the additive/clay ratio, and the starting moisture content and activation are the most important features of the modification process. Surprisingly, the final particle size makes low contribution. Normally, the surface area of natural porous solids increases significantly with the decrease of grain size; nevertheless, when a chemical treatment is involved, this parameter loses importance.[45] The low contribution of the final moisture content is attributed to the sample degassing pre-treatment before recording the N_2

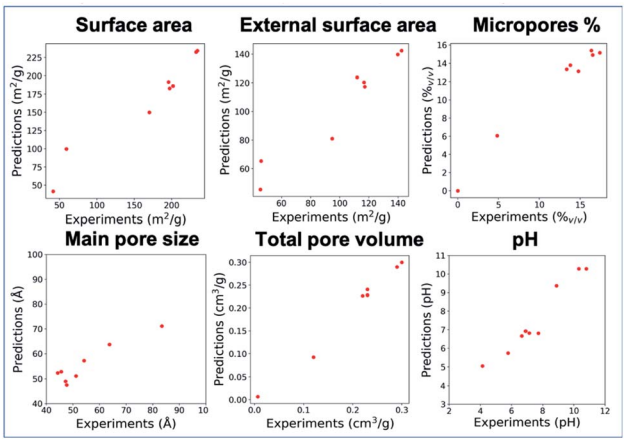

Fig. 2 Graphical model assessment results for the prediction of the labeled data of the test set.

Table 4 Test set-based model assessment results for the six ML models

Assessment	pSA	pESA	pMS	pVol	pMicro	ppH
R^2	0.943	0.93	0.77	0.986	0.954	0.959
MAE	11	6.09	4.09	0.006	0.96	0.33
MSE	276	89.6	38.7	0.00012	1.6	0.19

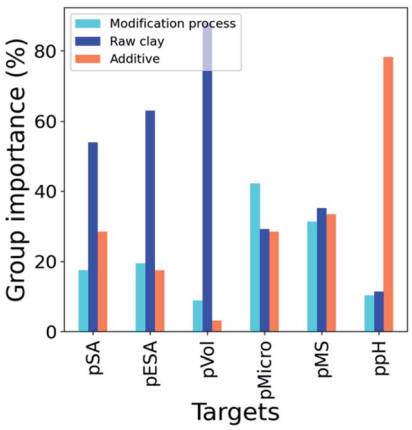

Fig. 3 Summed up importance scores of the features which were grouped according to Table 1.

physisorption isotherms.[38] Phyllosilicate composition in terms of fibrous and planar clay content is the key descriptor in the raw material group displaying the diverse intrinsic behaviour of the two morphological classes.[10,43] The other contributions mark the correlation between the target properties and the chemical structures in terms of substitutions in the theoretical layer composition as well as reversible exchangeable cations. Those substitutions may naturally occur in both octahedral and tetrahedral sheets, even if more energetically costly, complicating the assignment of their roles in the resulting properties.[10,17,19]

3.3. Feature importance analysis for the surface activity model

The overall group of features extracted using the model of surface activity is shown in Fig. 3, while Fig. S8† highlights the singular contribution of each descriptor. The acid–base character is mainly correlated with the additive features. The descriptors identifying the polar characteristics (X log P, polar surface, double bond count, and hydrogen bound donor and acceptor counts) have the biggest impact on ppH. Besides that, the molecular weight of the additive participates with a high importance score due to the fact that it is not completely removed after treatments. These findings establish a competitive effect between the additive's ability to modify the surface and the tendency to cover it. As expected, the raw clay group of features is dominant in the model by virtue of the intrinsic acid–base character (pH_0, pH_{24} and free acidity). Despite the low importance of the cation exchange capacity, the type of ions (K^+, Na^+) has a significant impact, indicating that these cations may interact with the modifier tuning its effectiveness. Finally, additive/clay and activation are the most correlated with the process treatment group.

3.4. Toward tuning clay properties by controlled modification

The models of experimental outcomes discussed in Sections 3.1–3.3 allow us to systematically study the effects of certain modifiers on clay minerals, and exploit the insights in practical material design. We investigated the correlations between the modification process descriptors (7 descriptors) and the targets of Table 2. The analysis of the effects of tuning particular descriptors representing the modification process on the property highlighted that many of them follow similar trends independent of the types of clays involved. In particular, pSA, pESA, pVol, pMicro, and ppH are enhanced and pMS decreased for those processed minerals presenting fine grain size. The additive concentration contributes to a slight property enhancement in the range of 2–7 M. Double milling is recommended while the final moisture should be in the range of 6–15% to appreciate changes in ppH. The activation of raw clays through nebulizing the additive by spray or mixing the clay directly with the acid solution leads to a stronger surface modification compared to a simple solid mixture. In contrast, the effect of additive/clay and starting moisture content is a function of the clay to be modified. This observation enables simplifying the space to two-most important descriptors, *i.e.* H_2SO_4/clays% and RH%, shown in Fig. 4. Each clay mineral has its optimal additive amount and starting water content that is correlated with the availability of the tunnels in fibrous clays or the interlayer space in laminar clays. Fig. 4 reveals the areas leading to increase (in yellow) or decrease (in dark) of the target properties. The dark regions in pSA, pESA, pVol and pMicro start from *c.a.* 3–7% additive/clay representing the threshold of structure collapse concurring with the macropore formations. The starting moisture content highlights a general slow tendency affecting pSA and pESA in fibrous clays and stevensites. In contrast, the starting moisture content is the key parameter for montmorillonite which finds enhancement at low values up to 20%.[46] At high moisture also pMicro decreases while pVol and pMS increase. In fact, the high water content in swelling clays, on coming into contact with modifiers, may delaminate the montmorillonites creating macropores and reducing the surface area.[47,48] The pMS shows two yellow regions separated by a dark region ranging between 10 and 20% which suggests that the interlayer space of dried montmorillonites is not available for modification. Montmorillonite is a unique clay, which increases the surface area, with respect to its starting values of 43 $m^2\ g^{-1}$, through acid modification. The ppH behaves in a different way depending on the starting materials. Palygorskite and montmorillonite acquire an acidic character even at low H_2SO_4 content, while the same treatment fails in conferring significant acid functionalities to the other minerals.

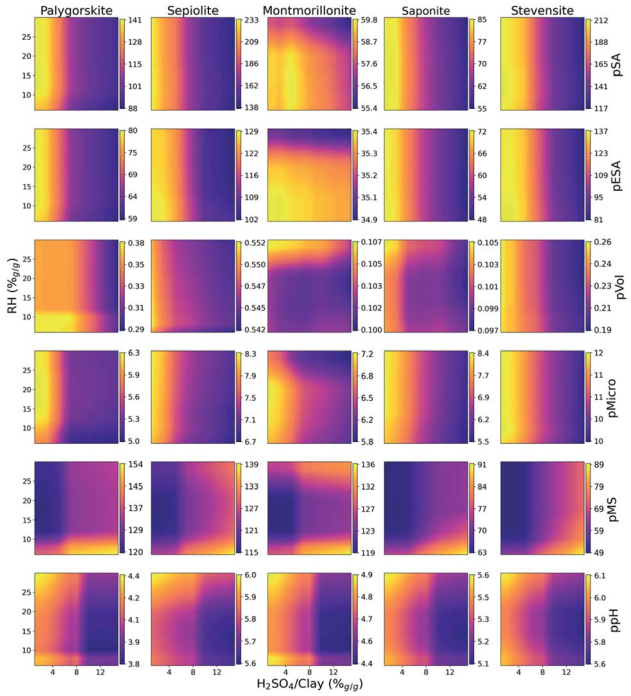

Fig. 4 Effect of the starting moisture content and H_2SO_4/clays% on the predicted target pSA, pESA, pVol, pMicro, pMS and ppH.

3.5. Design function

The surface activity and availability are key features in the design of acid nano-catalyzers.[28] The conflicting modification parameters and the uniqueness of each pristine natural mineral makes the identification of optimal preparation procedures experimentally very costly. The proposed multiobjective function, DF, married with the property models allows us to screen the parameter space to identify the ideal acid nano-catalyst recipes. The search was conducted over the entire space of seven features representing the process treatment. In Section 3.4, we observed a similar correlation between the five minor contributing descriptors allowing vector-space reduction to the two-most important, *i.e.* additive/clay and RH. Thus, to visually illustrate the results we explored the 2d space while the other five were fixed at averaged optimal values identified in Section 3.4.

Five natural materials, corresponding to high-purity palygorskite, sepiolite, montmorillonite, saponite and stevensite were explored. Fig. 5 highlights the optimal improvement areas with respect to the DF of the raw material. An improvement of 125% of DF at low RH% can be noted for saponite. Although

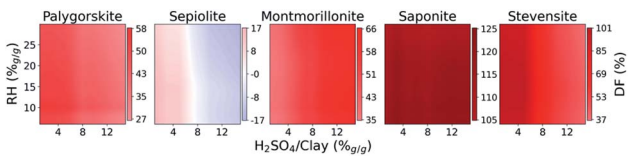

Fig. 5 Trend of design function improvements (%) with different additives/clays ($\%_{g/g}$) and starting moisture contents ($RH\%_{g/g}$).

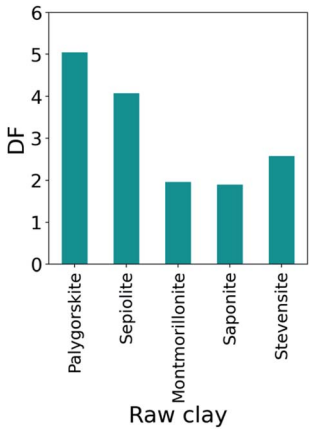

Fig. 6 Design function prediction for promising nano-catalyzers derived from different starting materials modified with sulfuric acid at its optimal amount (8% for palygorskite, 5% sepiolite, 12% montmorillonite, 3% for saponite and 6% for stevensite).

RH% plays a minor role, palygorskite and sepiolite are successfully modified at lower RH values of up to 15%. The optimal limit of additive/clay is identified as 5% for sepiolite but can be extended to 8% for palygorskite. In contrast, montmorillonites find their best conditions at high values of additive/clay. The DF of stevensite is improved until up to 6% of additive/clay, independent of the RH.

The impact of clay impurity on the design function is highlighted in Fig. S9.† Herein, the starting materials were stevensites provided by different deposits and exhibiting a diverse grade of purity, *i.e.* the phyllosilicate content varying from 80% (stevensite 1) to 62% (stevensite 5). The figure indicates that the DF can be generalized to the materials with impurities as the optimal regions for low-purity stevensites are located in the same regions as the high-purity ones. However, differences in the improvement scores can be appreciated following a non-linear trend with the stevensite content. In fact, in many cases natural clays present traces of other phyllosilicates that contribute to facilitating or preventing the modification. Our machine learning-supported multiobjective optimization represents an accelerated and low cost strategy for finding the optimal solutions for such impure materials. Furthermore, DF provides a rapid guide for experimentation and initial material selection by quickly identifying the optimal conditions for each considered clay. Fig. 6 displays the optimized DF values of nano-catalyzers based on five different clays. The plot identifies palygorskite as a promising starting system for the preparation of high-performing acid catalyzer.

3.6. Assessment of properties and catalytic activity of the optimized acid nano-catalyzer

The new nano-catalyzer prototype, P1, for which the maximal value of DF was observed, is a palygorskite-based material. Specifically, P1 was prepared by spraying 8% of sulfuric acid (5 M) onto palygorskite with the starting moisture content of 10%. The final RH (%) was set to 16% and the amount of particles smaller than 45 μm was set to 75%. To check if the above

Table 5 Results of acid catalyzed thermal degradation of chlorophyll-*a* for the promising nano-catalyzer (P1) assessed against the pure natural palygorskite and the corresponding sepiolite-based materials

Material	Remaining chlorophyll-*a* (ppm)	Chlorophyll-*a* removal (%)
P1	0.651	79
Raw palygorskite	1.84	39
P2	1.959	35.5
Raw sepiolite	1.91	37.1

modification process can be applied to improve other clays, a second prototype, P2, was also synthesized based on sepiolite (details in Section S3†).

Table S5 and Fig. S10† summarize the predicted and experimentally measured properties of P1 and P2. Overall, there is a very good agreement between the predicted and measured properties providing further validation of our models.

Similarly, the DF values calculated using the experimentally measured properties of P1 and P2 closely match the corresponding model prediction, Table S5.† Furthermore, the DF value being lower for P2 than P1 confirms that sepiolite does not get the benefit from the modification optimized for palygorskite. In fact, although the two fibrous nanoclays have similar morphologies, sepiolite exhibits a strong reduction of the available surface area together with non-acid character after the treatment.

The catalytic activities of P1 and P2 were tested in the degradation of chlorophyll-*a* in lipidic media following the protocol outlined in Section S5.† A generally accepted reaction mechanism involves replacing Mg^{2+} of the chlorophyll structure with H^{+} at the clay surface.[49] The fragmented molecules are eliminated by the high temperature and vacuum conditions or remain entrapped by physisorption in the porous clay and removed by filtration.[49,50] Generally, a desirable nano-catalyzer removes around 60–75% of the pigments depending on diverse variables (temperature, vacuum and dose).[51] The performances of P1 and P2 as well as the corresponding raw materials are collected in Table 5. P1 was able to remove 78.6% of chlorophyll-*a*, 40% of improvement with respect to the pristine material. The remaining pigments are then removed under extreme conditions of high temperature and vacuum forcing the elimination of volatile compounds and allowing the stripping of 88% of chlorophyll-*a*. Additionally, the catalytical activity of P2 indicates no improvement with respect to the raw sepiolite and that the conditions in which P1 and P2 were prepared are likely non-transferable between different clay types. This observation is in favor of our approach, which allows high-throughput search for optimum conditions for various clays and cutting times and experimental costs. Furthermore, the quality of the underlying property models opens the possibility of smart design of other DFs tailored to specific applications.

4. Conclusions

We have demonstrated an accelerated approach to the characterization and development of processed natural porous materials. Herein, statistical machine learning models are trained on the available experimental results, and then used in place of the real experiments, which are typically necessary in the context of material development. The models were based on a custom 41 feature representation corresponding to (1) raw clay characteristics, (2) additive characteristics and (3) processing conditions. In particular, Extra Tree regressor was selected for its high predictive power ($R^2 > 0.77$) while assessing morphological parameters and the surface activity as well as its reported suitability for working with small datasets. The feature importance analysis of the models shed light onto the features of the raw minerals significantly affecting their internal and external surface area and pore volume, while micropore content and main pore size are strongly influenced by the features corresponding to the modification process and the acidity determined by the feature corresponding to the additive involved. The feature importance was analyzed to disclose roles of real physico-chemical descriptors. Each investigated clay-based material reveals a unique behavior that is affected by structural defects and ion substitutions. Competitive influence of the features on the property targets renders it necessary to find the right compromise for specific requirements. The models were employed to predict the outcomes of a set of proposed materials rich in palygorskite, sepiolite, montmorillonite, saponite and stevensite of high purity screened by the multiobjective optimization.

Furthermore, a design function was proposed to investigate ideal material processing scenarios to achieve high performing hierarchical acidic nano-catalysts. The results suggest that the investigated raw nanoclays are not equally qualified precursors, *i.e.* palygorskite is preferred as a starting natural porous material. By screening through the space of important synthesis and processing parameters, we identified and then synthesized a nano-catalyzer for effective removal of chlorophyll-*a* from lipidic media. This palygorskite-based material achieves 79% degradation of chlorophyll-*a* in an acid catalyzed degradation reaction, outperforming natural minerals with 40% of improvement.

Data availability

Our code and dataset are available at https://doi.org/10.5281/zenodo.4742294.

Author contributions

Giulia Lo Dico and Maciej Haranczyk carried out the computational work. Giulia Lo Dico and Álvaro Peña Nuñez performed the experiments and analysed the data with assistance from Verónica Carcelén. The manuscript was written by Giulia Lo Dico and Maciej Haranczyk. All authors provided input on the manuscript.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors acknowledge the Community of Madrid for its support of this work through the project IND2018/IND-9819. The authors are grateful to anonymous reviewers who offered very helpful suggestions. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract no. DE-AC02-05CH11231.

References

1 C. M. Simon, R. Mercado, S. K. Schnell, B. Smit and M. Haranczyk, *Chem. Mater.*, 2015, **27**, 4459–4475.
2 S. Ma, B. Space, L. Wojtas, M. Eddaoudi and M. J. Zaworotko, *Nature*, 2013, **101**, 1–5.
3 A. Awasthi, P. Jadhao and K. Kumari, *SN Appl. Sci.*, 2019, **1**, 1–21.
4 C. M. A. Parlett, K. Wilson, A. F. Lee, A. F. Lee and K. Wilson, *Chem. Soc. Rev.*, 2013, **42**, 3876–3893.
5 C. Perego and R. Millini, *Chem. Soc. Rev.*, 2013, **42**, 3956–3976.
6 B. Wicklein, M. Darder, P. Aranda and E. Ruiz-hitzky, *ACS Appl. Mater. Interfaces*, 2011, **3**, 4339–4348.
7 G. Lo Dico, B. Wicklein, L. Lisuzzo, G. Lazzara, P. Aranda and E. Ruiz-Hitzky, *Beilstein J. Nanotechnol.*, 2019, **10**, 1303–1315.
8 M. E. Davis, *Nature*, 2002, **417**, 813–821.
9 L. Lisuzzo, B. Wicklein, G. Lo Dico, G. Lazzara, G. del Real, P. Aranda and E. Ruiz-Hitzky, *Dalton Trans.*, 2020, **49**, 3830–3840.
10 F. Bergaya, B. K. G. Theng and G. Lagaly, *Handbook of Clay Science*, 2006, vol. 1.
11 R. A. Schoonheydt, C. T. Johnston and F. Bergaya, in *Surface and Interface Chemistry of Clay Minerals*, 2018, vol. 9, pp. 1–21.
12 D. K. Dutta, in *Surface and Interface Chemistry of Clay Minerals*, Elsevier Ltd., 1st edn, 2018, vol. 9, pp. 289–329.
13 G. Nagendrappa, *Appl. Clay Sci.*, 2011, **53**, 106–138.
14 N. Degirmenbasi, N. Boz and D. M. Kalyon, *Appl. Catal., B*, 2014, **150–151**, 147–156.
15 F. M. T. Luna, J. A. Cecilia, R. M. A. Saboya, D. Barrera, K. Sapag, E. Rodríguez-Castellón and C. L. Cavalcante, *Materials*, 2018, **11**, 6–9.
16 V. D'Ascanio, D. Greco, E. Menicagli, E. Santovito, L. Catucci, A. F. Logrieco and G. Avantaggiato, *Appl. Clay Sci.*, 2019, **181**, 105209.
17 Q. Wang, C. Zhu, J. Yun and G. Yang, *J. Phys. Chem. C*, 2017, **121**, 26722–26732.
18 W. Wang and A. Wang, *Palygorskite Nanomaterials: Structure, Properties, and Functional Applications*, Elsevier Inc., 2019.
19 A. Singer and E. Galan, in *Development in Clay Science*, 2011.
20 L. Novikova, F. Roessner, L. Belchinskaya, M. AlSawalha and V. Krupskaya, *Appl. Clay Sci.*, 2014, **101**, 229–236.
21 M. P. Hart and D. R. Brown, *J. Mol. Catal. A: Chem.*, 2004, **212**, 315–321.
22 C. Briones-Jurado and E. Agacino-Valdés, *J. Phys. Chem. A*, 2009, **113**, 8994–9001.
23 J. Hwang and R. Pini, *Environ. Sci. Technol.*, 2019, **53**, 11588–11596.
24 D. Kim, Y. H. Ahn, S. J. Kim, J. Y. Lee, J. Lee, Y. J. Seo and H. Lee, *J. Phys. Chem. C*, 2015, **119**, 22148–22153.
25 P. Huang, A. Kazlauciunas, R. Menzel and L. Lin, *ACS Appl. Mater. Interfaces*, 2017, **9**, 26383–26391.
26 I. Jeon and K. Nam, *Sci. Rep.*, 2019, **9**, 1–10.
27 S. Petrović, L. Rožić, Z. Vuković, T. Novaković and D. Stanisavljev, *Clays Clay Miner.*, 2012, **60**, 32–39.
28 C. Tournassat, J. A. Davis, C. Chiaberge, S. Grangeon and I. C. Bourg, *Environ. Sci. Technol.*, 2016, **50**, 13436–13445.
29 S. V. Churakov and X. Liu, in *Surface and Interface Chemistry of Clay Minerals*, Elsevier Ltd., 1st edn, 2018, vol. 9, pp. 49–87.
30 K. M. Jablonka, D. Ongari, S. M. Moosavi and B. Smit, *Chem. Rev.*, 2020, **120**, 8066–8129.
31 Z. Jensen, E. Kim, S. Kwon, T. Z. H. Gani, Y. Román-Leshkov, M. Moliner, A. Corma and E. Olivetti, *ACS Cent. Sci.*, 2019, **5**, 892–899.
32 J. Schmidt, J. Shi, P. Borlido, L. Chen, S. Botti and M. A. L. Marques, *Chem. Mater.*, 2017, **29**, 5090–5103.
33 J. Freiesleben, J. Keim and M. Grutsch, *Qual. Reliab. Eng. Int.*, 2020, 1–12.
34 L. Bieseki, F. Bertella, H. Treichel, F. G. Penha and S. B. C. Pergher, *Mater. Res.*, 2013, **16**, 1122–1127.
35 T. Li, J. Xu, H. Wu, G. Wang, S. Dai, J. Fan, H. He and W. Xiang, *Mar. Drugs*, 2016, **14**, 1–19.
36 J. Liang, A. Appukuttan Aachary and U. T. Hollader, *Lipid Technol.*, 2015, **27**, 231–233.
37 S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang and E. E. Bolton, *Nucleic Acids Res.*, 2019, **47**, D1102–D1109.
38 M. Thommes, K. Kaneko, A. V. Neimark, J. P. Olivier, F. Rodriguez-Reinoso, J. Rouquerol and K. S. W. Sing, *Pure Appl. Chem.*, 2015, **87**, 1051–1069.
39 ISO9277:2010, 2010.
40 L. Breiman, *Mach. Learn.*, 2001, **45**, 5–32.
41 G. Louppe, L. Wehenkel, A. Sutera and P. Geurts, *Adv. Neural Inform. Process. Syst.*, 2013, 1–9.
42 P. Geurts, D. Ernst and L. Wehenkel, *Mach. Learn.*, 2006, **63**, 3–42.
43 V. V Krupskaya, S. V Zakusin, E. A. Tyupina, O. V Dorzhieva, A. P. Zhukhlistov, P. E. Belousov and M. N. Timofeeva, *Minerals*, 2017, **4**, 49.
44 P. Komadel and J. Madejová, *Acid Activation of Clay Minerals*, Elsevier Ltd., 2nd edn, 2013, vol. 5.
45 M. E. Hodson, *Geochim. Cosmochim. Acta*, 1998, **62**, 3429–3435.
46 C. D. Hatch, J. S. Wiese, C. C. Crane, K. J. Harris, H. G. Kloss and J. Baltrusaitis, *Langmuir*, 2012, **28**, 1790–1803.
47 M. Holmboe and I. C. Bourg, *J. Phys. Chem. C*, 2014, **118**, 1001–1013.

48 L. Chong and E. M. Myshakin, *Fluid Phase Equilib.*, 2018, **472**, 185–195.
49 N. Koca, F. Karadeniz and H. S. Burdurlu, *Food Chem.*, 2007, **100**, 609–615.
50 A. Van Loey, V. Ooms, C. Weemaes, I. Van den Broeck, L. Ludikhuyze, Indrawati, S. Denys and M. Hendrickx, *J. Agric. Food Chem.*, 1998, **46**, 5289–5294.
51 C. Baroi and A. K. Dalai, *Catal. Today*, 2013, **207**, 74–85.