Supporting Information for

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

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

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

S1. Experimental methods and definitions

The particle size was estimated as a percentage of small particles in the sample being able to sift through the sieve of 45 μm diameter pores. Chemical analysis of metal oxides content was carried out with ICP-OES Varian, Agilent 730. The mass loss by calcination was estimated in a Heraeus M110 oven. The cation interchange capacity (CIC) was measured with NH_4 ion selective electrode, Metrohm 867-801 and Thermo Scientific Orion 960 titrator. The starting and final moisture (RH $\%_{g/g}$) were evaluated as weight loss at 145°C in a HG53 Halogen Moisture Analyser. XRD patterns were recorded with a D8-ADVANCE diffractometer (Bruker), using Cu Kα radiation. The voltage and current sources were set at 40 kV and 30 mA, respectively. Diffractograms were recorded at a goniometer speed of 0.5 s per step between 2° and 70° (2θ). The samples were pre-treated separating the clay fraction (fine size material) by centrifugation and used as oriented aggregate mounts for clay-mineral identification and semi-quantification. The pH of a 10 wt% suspension of raw clay in distilled water was measured with Crison Basic 20 pH-meter, at zero time and after 24 hours keeping the dispersion under stirring at 25°C. The free acidity was determined both by pH measurement of a filtered solution (by Filterlab 1240 with pore diameters of 14-18 μm) of 5 wt% dispersion of raw clay in distilled water after keeping the temperature at 50°C during 5 minutes under stirring.

Physisorption isotherms were measured using Micromeritics Gemini V surface area and pore size analyzer under N_2 gas flow at 77K. The isotherms were obtained after a 18hrs degassing step at 124°C. N_2 was selected as a gas probe following IUPAC recommendations.[1,2] The desorption curves provide important information on the mesoporosity via Barrett-Joyner-Halenda (BJH) method,[3–6] however, the low accuracy impedes quantitative conclusions.[1,7–9] Thus, morphological parameters were calculated in the adsorption branch of N_2 isotherm.[6] Figure S1 shows examples of N_2 physisorption isotherms and t-plot of the 5 clay categories investigated. The main calculated morphological features are reported in Table S1. The BET surface areas of fibrous clays are higher than the laminar shaped minerals, while the natural montmorillonites exhibit the lowest surface areas, similarly to the Mg-rich smectite. The micropores are attributed to the tunnels in sepiolite and palygorskite and to the interlayer space in smectites. While the presence of mesopores and macropores, as revealed by total pore volume and the main pore size, are assigned to the arrangement between fibers and plates themselves.[10,11]

The surface activity of nanoclays in water was assessed in terms of acid-base character (pH), which is indirectly correlated to catalytic power of the active sites as well as to the affinity towards (certain) molecules to be adsorbed at the surfaces.[12,13]

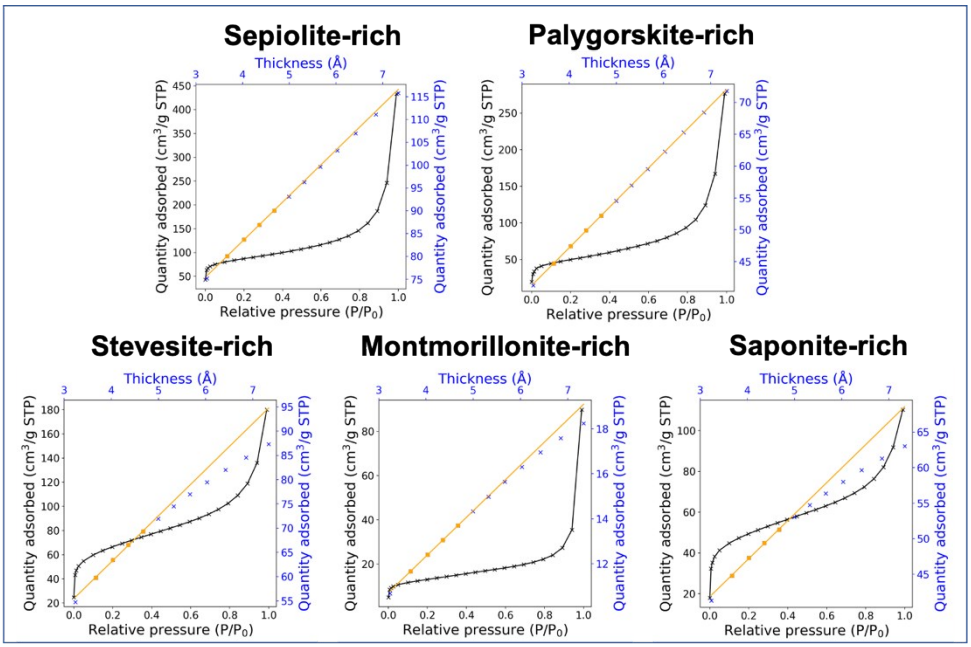

Fig. S1. Nitrogen adsorption isotherm linear plot and t-plots of natural clay minerals rich in sepiolite, saponite, palygorskite, stevensite and montmorillonite.

Table S1. Properties of high-purity natural nanoclays.

Property	Palygorskite	Sepiolite	Montmorillonite	Saponite	Stevensite
BET Surface area (m^2/g)	149	302	43	131	239
External surface area (m^2/g)	115	153	31	96	150
Main pore size (Å)	98	89	121	39	35
Total pore volume (cm^3/g)	0.43	0.67	0.14	0.17	0.22
Micropores ($\%_{v/v}$)	6.3	10.1	4.7	19.9	20.9
pH	6.8	7.2	7.2	7.5	7.6

S2. The feature space coverage investigations with principal component analysis

Principal component analysis (PCA) was performed using Sci-kit learn library. The algorithm was applied to both datasets discussed in the manuscript. The whole vector space (41 descriptors), shown in Table 1, was reduced to 6 principal components (PCs) explaining the 80% and 75% of the total variance in the morphological and surface activity datasets, respectively.

We compared the coverage of the material space achieved by our two datasets with respect to hypothetical well-designed experiments. The latter consisted of the points sampled using the design of experiments (DoE) approach's latin hypercube sampling with multidimensional uniformity. We reduced the whole feature descriptor space to the main six principal components, and we plotted the normalized PC2 vs PC1 assessing the data distribution and the feature space coverage (Figure S2). The collected data points are quite dispersed within the plot space, and our datasets provide significant overlap with the DoE points used as reference. Some small gaps in the point clouds representing out datasets (Fig. S2) can be attributed to experimental factors, which prevented practical realization of the corresponding material space region.

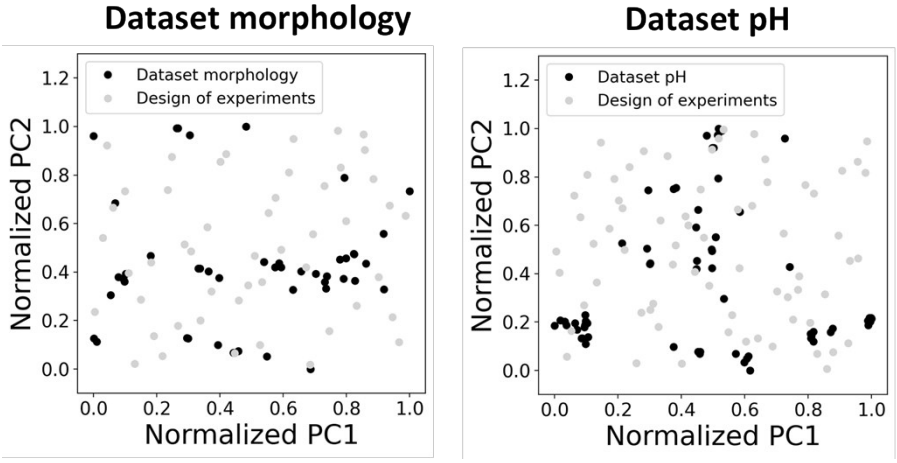

Figure S2. Graphical visualization of data point distribution into normalized PC1 and PC2 descriptors.

The novel prototype materials, i.e. the ones used to validate the outcomes of our approach (P1 and P2), reside in particular regions of the descriptor space, which have enough coverage by data, allowing us to trust in the models predictions. Their location within the space represented by 6 principal components is highlighted in Fig. S3.

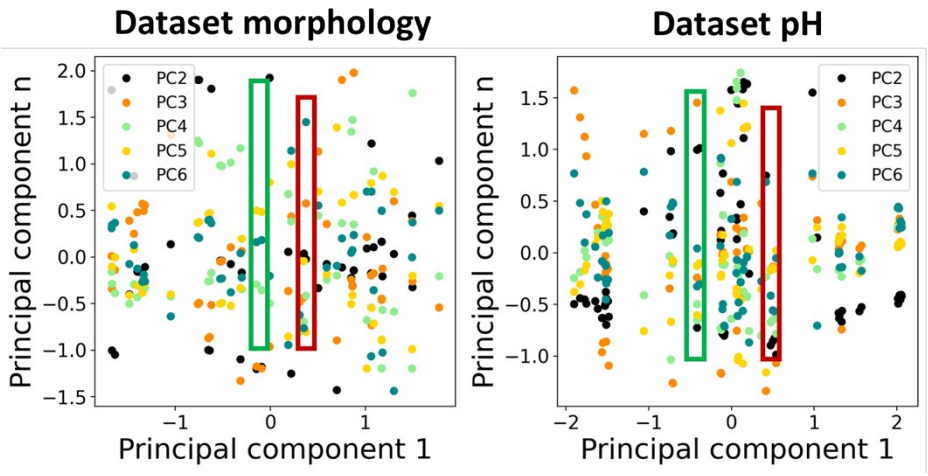

Figure S3. The plots of five principle components PC2-PC6 with respect to the PC1. The ranges of PCn values of prototype P1 is given in green box and P2 in red box.

S3. Hyperparameter optimization

Random Forest and Extremely Randomized Trees (Extra Trees) regressors were tested with a number of different combination of n_estimators in [25, 50, 100, 150, 200, 250, 500, 750, 1000, 1500, 2000, 3000, 4000, 5000], min_samples_split in [1, 2, 3, 4, 10], min_samples_leaf in [1, 2, 3, 4], max_features in [0 to 41] and max_depth in [50, 100, 300, 600, 900]. Decision Tree was fitted tuning min_samples_split, min_sample_leaf, max_features and maximum depths.

Table S2. Cross-validation accuracy performed on trainset with K=10.

Assessment	pSA	pESA	pVol	pMicro	pMS	ppH
R^2	0.83	0.65	0.86	0.59	0.1	0.76
MEA	28	17	0.05	2.7	22	0.6

S4. Dimensionality reduction.

Principal component analysis (PCA) was implemented in Sci-kit learn Python library. The algorithm was applied to both datasets. The vector space which identify the properties of raw clay (20 descriptors) was reduced to 4 principal components (PC (n)) explaining the 92% and 88% of the total variance in the morphological and surface activity datasets, respectively. The individual contributions of the principal components to the variance are shown in Figure S4.

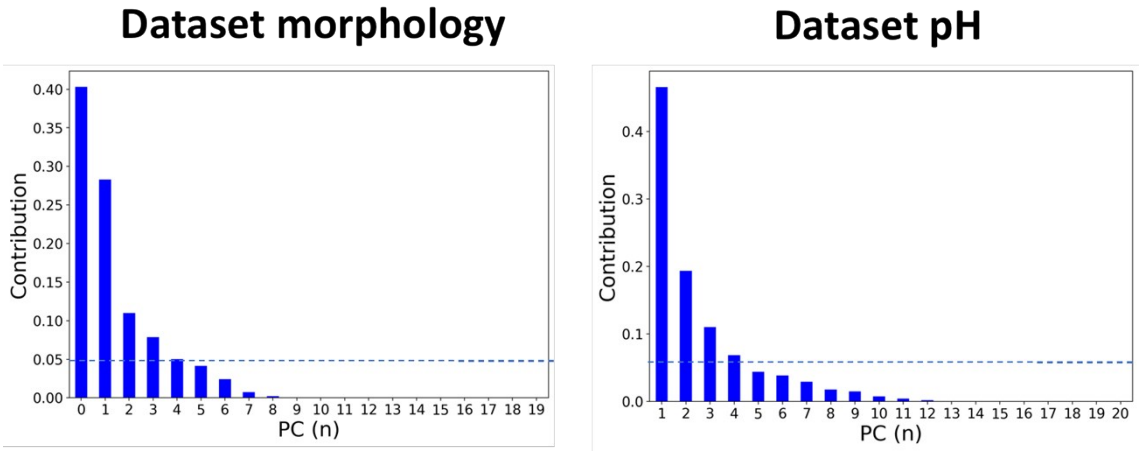

Figure S4. Individual contributions of the principal component in explain the total variance of the employed dataset (morphology and pH). The contribution score of 0.05 was chosen as threshold allowing the feature reduction to 4 main principal components.

The singular contribution of the pristine descriptors (20 features) in morphological dataset associated to the main 4 PC is shown in Figure S5. while the contribution in the surface activity dataset is reported in Figure S6.

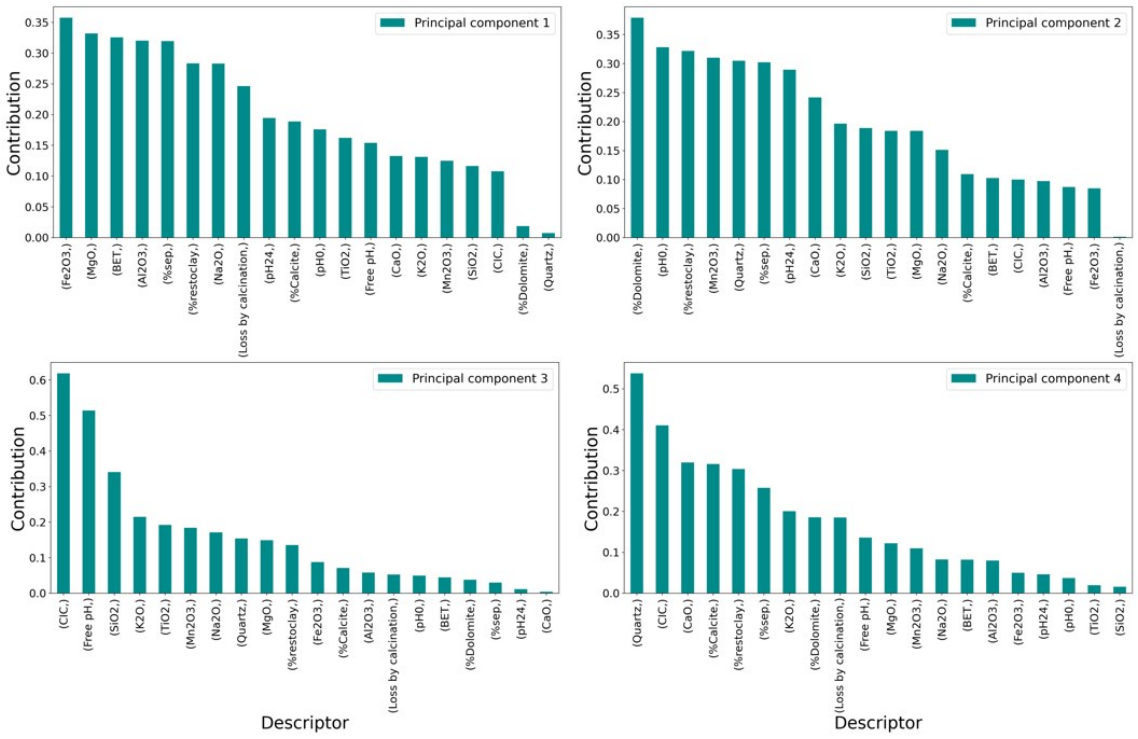

Figure S5. Contribution of the pristine features describing the properties of raw clay materials in the morphological dataset reduced by PCA. The 4 main PC are shown.

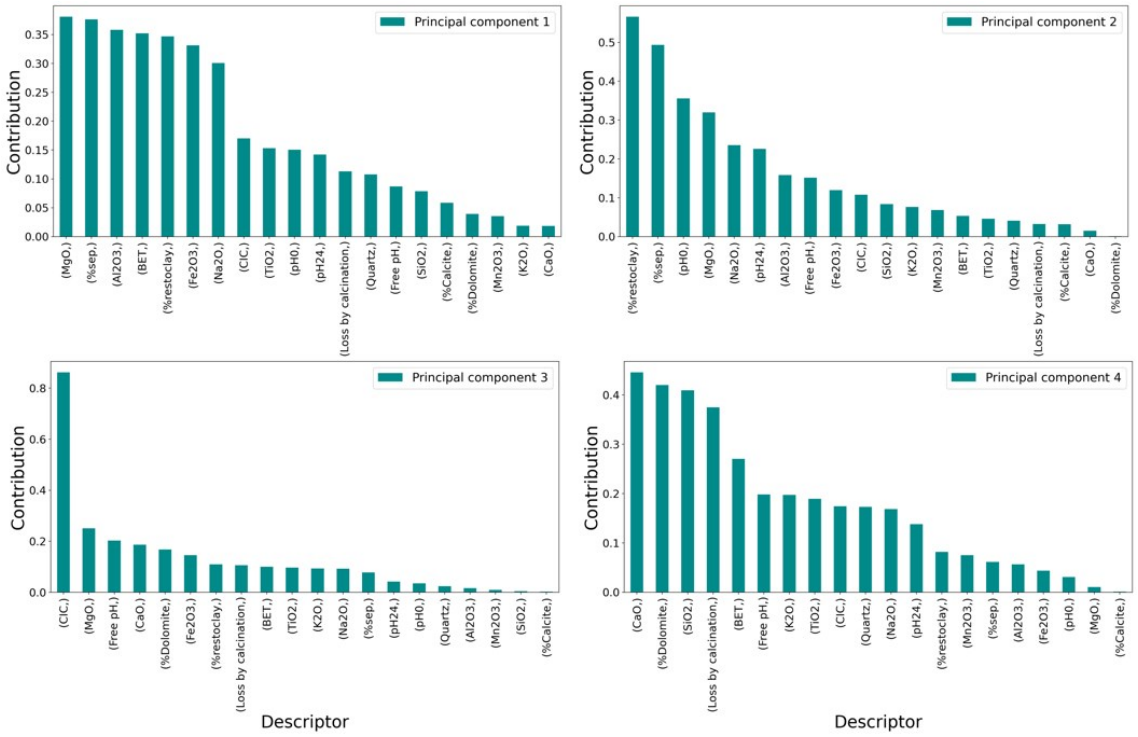

Figure S6. Contribution of the pristine features describing the properties of raw clay materials in the surface activity dataset reduced by PCA. The 4 main PC are shown.

The reduced vector space, together with the group of features of additive and modification process (Table 1) create a space of 25 descriptors which were implemented assessing the effect of the dimensionality reduction on the model predictability. The 6 models predicting pSA, pESA, pMS, pVol, pMicro and ppH were optimized selecting the best hyperparameter on the base of k-fold cross-validation (K=3) on the trainset (85%) and the assessment on the testset (15%) is summarized in Table S3.

Table S3. Model assessment results for the six ML models using the test set. The reduced vector space of 25 descriptors was used in this analysis.

Assessment	pSA	pESA	pMS	pVol	pMicro	ppH
R^2	0.932	0.83	0.87	0.90	0.89	0.92
MAE	14	10.8	3	0.02	1.47	0.45
MSE	332	214	19	0.0008	3.8	0.37

S5. Material preparation and catalytic efficiency test

A new acid nano-catalyzer, referred-to as P1, was prepared using the formulation that maximized the value of the design function DF for palygorskite-based materials. Herein, the palygorskite mineral with purity grade >70% was modified by sulfuric acid. The latter additive was added while stirring the mixture for 15 minutes. The modified material was then milled in a Restch ZM200 equipment tuning the speed from 8000 to 18000 rpm until obtaining the final particle size desired. The H_2SO_4/clay was fixed at optimal value of 8%$_{g/g}$, the acid concentration was 5M, the starting and final RH(%) was fixed to 10 and 16% respectively and the size to 75% of particles <45 μm.

The properties as well as the performance in the catalytic degradation of chlorophyll-a of P1 were compared with the ones associated with the raw palygorskite as well as another clay material, sepiolite, both raw and modified by the same treatment as P1 (the latter is referred-to as P2).

Thermal degradation experiments were conducted using a laboratory-scale equipment, reproducing the industrial process. For each test, 150 mg of lipid media with the proper amount of chlorophyll-a (ca. 3 ppm) was introduced into a glass batch (250 ml) under stirring with controlled pressure and temperature. The

nano-catalyzer materials were added with a dosage of 0.4%$_{g/g}$ keeping pressure at 60 mbar and the temperature at 100°C. The mixture was stirred for 30 minutes, then the nano-catalyzer was separated by filtration over a Buchner funnel with Filterlab 1250 with the pore diameter of 10-13 μm and the resulting amount of chlorophyll-a was photometrically measured using a Lovibond Tintometer Color Scale in a 0.25" (10 mm) glass cell.

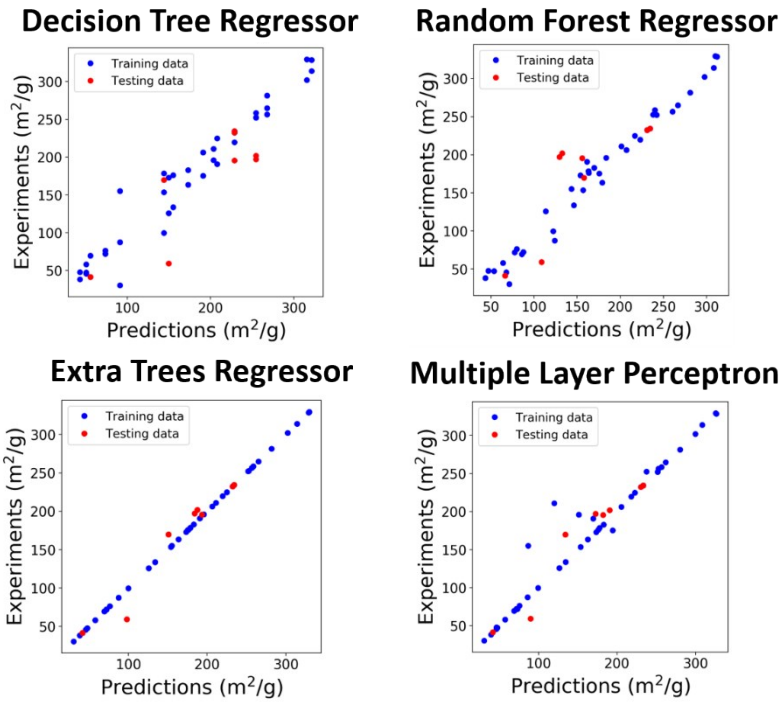

Fig S7. Visual assessment of BET surface area prediction models for modified clay. The dataset includes 49 data points, which were split 85-15% respectively into the training and test sets.

Table S4. Characterization of the BET surface area models based on various machine learning algorithms.

Assessment	Decision Tree Regressor	Random Forest Regressor	Extra Tree Regressor	Multilayer Perceptron
R^2	0.59	0.64	0.943	0.92
MAE	35	33	11	14.9
MSE	2027	1715	276	393

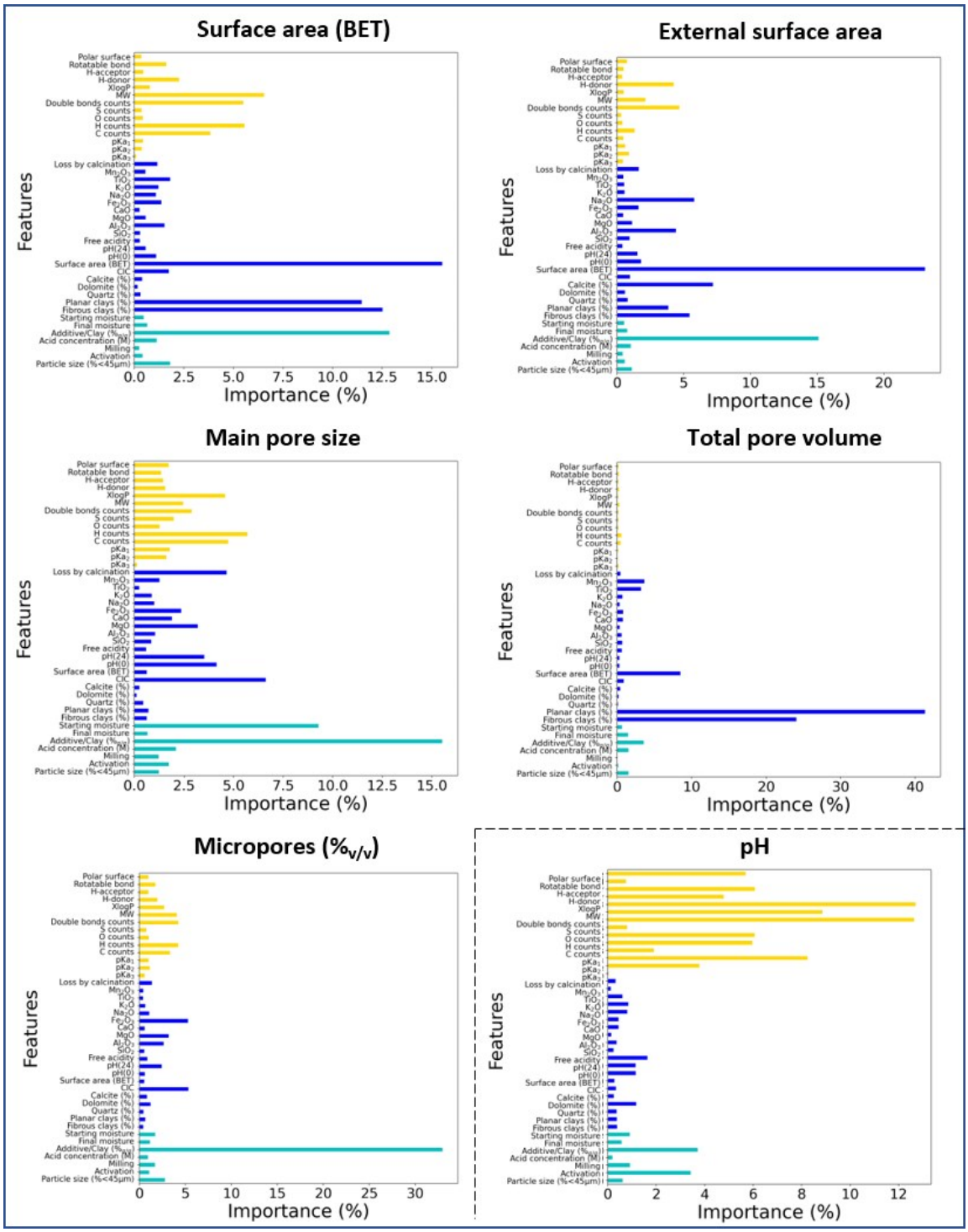

Fig S8. Feature importance scores for the Extra Trees Regressor-based models of properties of modified clay-based materials.

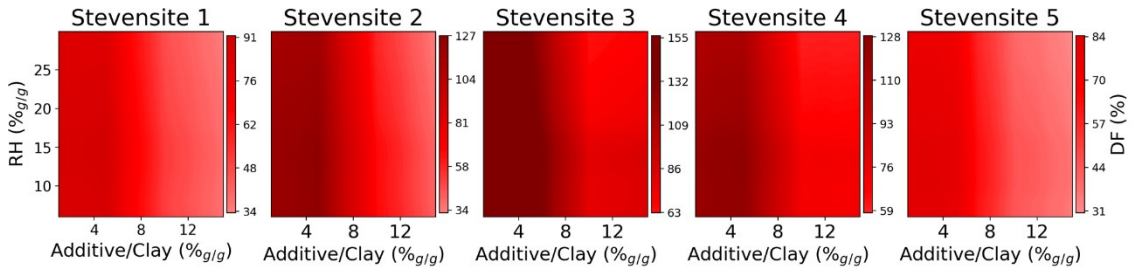

Fig S9. The plots of the design function improvements (in %) in the selected ranges of the additive/clay ratio and starting moisture content (RH $\%_{g/g}$) for five stevensites with different grade of purity (from 80% to 62% of phyllosilicate content).

Table S5. Properties of P1 and P2 materials as predicted by the Extra Tree models and measured experimentally.

Property	Predicted		Experimental	
	P1	**P2**	**P1**	**P2**
BET Surface area (m^2/g)	101	203	106	184
External surface area (m^2/g)	65	105	69	81
Main pore size (Å)	151	87	111	85
Total pore volume (cm^3/g)	0.38	0.51	0.36	0.49
Micropores ($\%_{v/v}$)	5.2	8.6	4.6	9.4
pH	3.8	6.5	4.2	6.8
Design function	5.0	3.6	5.5	2.9

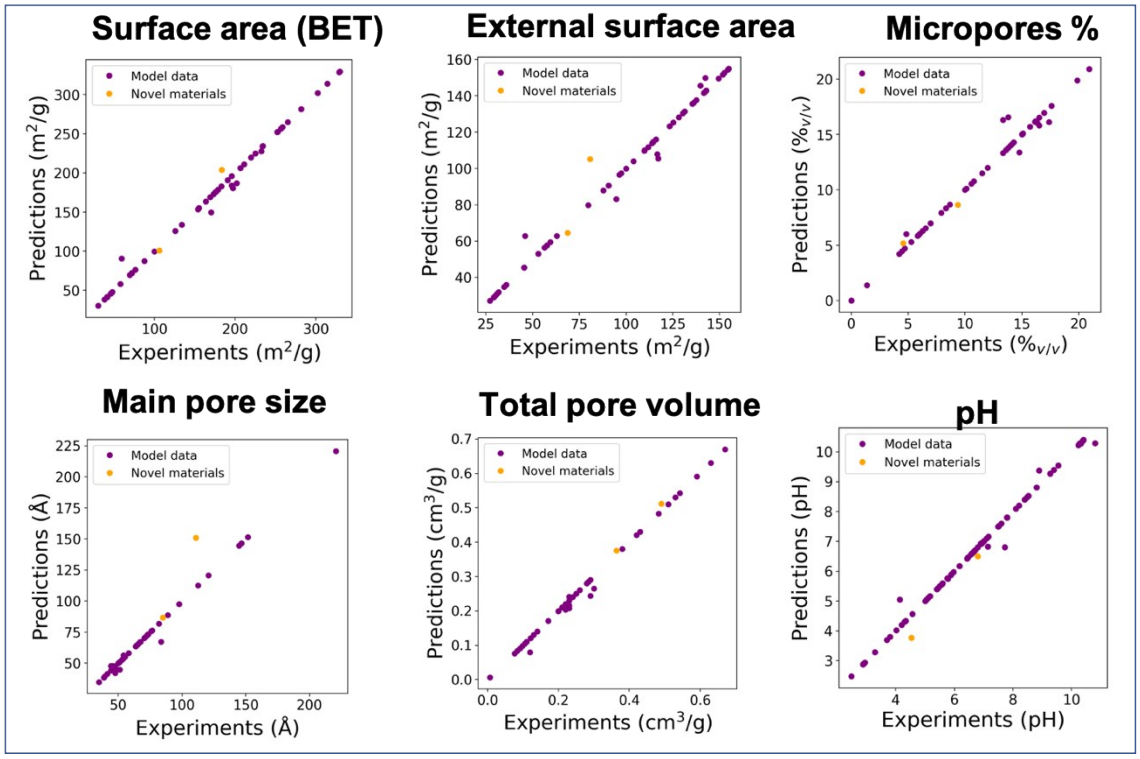

Fig S10. Visual assessment of the quality of Extra Tree model predictions for the novel materials synthesized in this work (P1 and P2).

References

1 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.

2 ISO9277:2010, 2010.

3 S. Lowell, J. E. Shields, M. A. Thomas and M. Thommes, *Characterization of porous solids and powders: surface area, pore size, and density*, 2005, vol. 42.

4 E. P. Barrett, L. G. Joyner and P. P. Halenda, *J. Am. Chem. Soc.*, 1951, **73**, 373–380.

5 J. C. Groen, L. A. A. Peffer and J. Pérez-Ramírez, *Microporous Mesoporous Mater.*, 2003, **60**, 1–17.

6 M. F. De Lange, T. J. H. Vlugt, J. Gascon and F. Kapteijn, *Microporous Mesoporous Mater.*, 2014, **200**, 199–215.

7 J. Rouquerol, F. Rouquerol, P. Llewellyn, G. Maurin and K. Sing, *Adsorption by Powders and Porous Solids, 2nd Edition*, Academic Press, 2nd edn., 2012.

8 M. Thommes, in *Nanoporous Materials: Science and Engineering*, 2004, pp. 317–364.

9 A. Galarneau, F. Villemot, J. Rodriguez, F. Fajula and B. Coasne, *Langmuir*, 2014, **30**, 13266–13274.

10 F. Bergaya, B. K. G. Theng and G. Lagaly, *Handbook of Clay Science*, 2006, vol. 1.

11 M. Suárez and E. García-Romero, *Appl. Clay Sci.*, 2012, **67–68**, 72–82.

12 G. D. A.-O. 1 Jozefaciuk 2002, *Clays Clay Miner.*, 2002, **50**, 647-656 ST-EFFECT OF ACID AND ALKALI TREATMENTS.

13 C. T. Johnston, in *Surface and Interface Chemistry of Clay Minerals*, Elsevier Ltd., 1st edn., 2018, vol. 9, pp. 89–124.