Facilitating development of special clays by incorporation of machine learning techniques.

Giulia Lo Dico

A dissertation submitted by in partial fulfillment of the requirements for the degree of Doctor of Philosophy in

Materials Science and Engineering

Universidad Carlos III de Madrid

Director:

Maciej Haranczyk

Co-director:

Veronica Carcelén

Tutor:

Paula Alvaredo Olmos

March 2023

ACKNOWLEDGEMENT

I would like to express my gratitude to IMDEA Materials, Tolsa, and UC3M for creating and supporting this bold project. I am also thankful to the Community of Madrid for providing funding and resources.

I am deeply grateful to my thesis director and co-director, Maciej and Veronica, and to Juanjo, for their faith in me and for providing me the opportunity to explore and solve problems. I would like to especially thank you, Maciej for your unwavering support and guidance during the most challenging moments.

I am also thankful to Veronica and Antonio for making the implementation of machine learning in all projects of the Research & Technological Innovation department at Tolsa a reality.

Finally, I would like to extend my heartfelt thanks to all the wonderful individuals at IMDEA and Tolsa who have made this experience memorable and filled it with enjoyable moments and exciting adventures. They have become my friends and this experience will always hold a special place in my heart.

Grazie.

PUBLISHED AND PRESENTED CONTENTS

Lo Dico, G.; Nuñez, Á. P.; Carcelén, V.; Haranczyk, M. Machine-Learning-Accelerated Multimodal Characterization and Multiobjective Design Optimization of Natural Porous Materials. *Chem. Sci.* **2021**, 12 (27), 9309–9317. https://doi.org/10.1039/d1sc00816a.

- This work is included as the Chapter 3 of this manuscript.
- In this contribution I implemented data-driven techniques, i.e., machine learning, to accelerate the characterization and design of clay-based materials for catalytic applications. I used a historical dataset collected by Tolsa Company to train Extra Tree Regressor models to predict the morphology and the surface activity of processed nanoclays. The high throughput of the models enabled exploration of processing parameter–property correlations and multiobjective optimization of prototype materials. One of such identified prototypes was experimentally prepared and tested revealing appreciable acid character improvement and 79% removal of chlorophyll-a in a acid-catalyzed degradation.

Lo Dico, G.; Croubles, S.; Carcelén, V.; Haranczyk, M. Machine learning-aided design of composite mycotoxin detoxifier material for animal feed. *Sci. Rep.* **2022**, 12 (1), 1-11. https://doi.org/10.1038/s41598-022-08410-x.

- This work is included as the Chapter 4 of this manuscript.
- In this contribution I applied statistical machine learning methods, trained on an in vitro mycotoxin adsorption dataset, to identify nanoclay composite formulations with high removal capacity and selectivity towards various mycotoxins. The findings were validated with an *in vivo* toxicokinetic study, based on the detection of biomarkers for mycotoxin-exposure in broilers, realized in collaboration with the Faculty of Veterinary Medicine, Ghent University. The optimal detoxifier provided reduction of the oral toxin bio-availability after single bolus administration.

Lo Dico, G.; Muñoz, B.; Carcelén, V.; Haranczyk, M. Data-driven experimental design of rheological clay-polymer composites. *Ind. Eng. Chem. Res.* **2022**, 61 (31), 11455–11463. https://doi.org/10.1021/acs.iecr.2c00936.

- This work is included as the Chapter 5 of this manuscript.
- In this contribution I coupled advanced design of experiments with machine learning to accelerate the design of rheological clay-based additives. I used both historical and novel generated experimental datasets to develop Random Forest Regressors to predict the viscosity of clay-based viscoplastic fluids. The models enabled material screening and Pareto multiobjective optimization, identifying those formulations which reflect a stable gel strength. The top promising prototypes were prepared and characterized finding outstanding performances to be employed in engineering operations such as horizontal directional drilling.

Lo Dico, G.; Lisuzzo, L.; Carcelén, V.; Lazzara, G.; Haranczyk, M. Toward the role of moisture in natural and thermally-treated clay materials. *Microporous and Mesoporous Materials.* **2022** (*Under revision*)

- This work is included as the Chapter 6 of this manuscript.
- This contribution summarizes results of a collaboration between IMDEA Materials, Tolsa and the University of Palermo. The results obtained during my International stay were amplified by running additional experiments in Tolsa's laboratories, analized and converted into a manuscript which is currently under revision by the team of *Microporous and Mesoporous Materials journal.* The study contributs to understand the effect of weak thermal treatment at 150° C of the hydratation properties of various clay materials. In particular, we calculated the activation energy for the removal of the adsorbed water (Ea) from modulated thermogravimetric analysis (MTGA). The thermal treatment enhances the re-humectation in Na-montmorillonite, Smectite, and sepiolite structures with a decrease of the energy

required to remove it, while Ea increases significantly in palygorskite. Then we correlated Ea to the adsorption of two molecules, i.e. aflatoxin B1 (AFB1) and β-carotene, modelling two examples of concrete applications of clays as adsorbent. The findings demonstrate diverse behavior of the five natural clays to the thermal treatment and the importance of the confined water in adsorptive applications.

OTHER RESEARCH MERITS

Other publications

Haranczyk, M.; Lo Dico, G. Book Chapter - Big data science in nanoporous materials. Dataset and descriptors. Book Chapter, Artificial Intelligence-Guided Design and Property Prediction for Zeolites and Nanoporous Materials. (2023) Wiley.

Ruiz-hitzky, E.; Ruiz-garcía, C.; Fernandes, F. M.; Dico, G. Lo; Lisuzzo, L.; Prevot, V.; Darder, M.; Aranda, P. Sepiolite-Hydrogels : Synthesis by Ultrasound Irradiation and Their Use for the Preparation of Functional Clay-Based Nanoarchitectured Materials. *Front. Chem.* **2021**, *9* (August), 1–17. https://doi.org/10.3389/fchem.2021.733105.

Grants and Awards

2021 - RSC Research development grant.

2020 - Chemical Science Outstanding Poster Presentation Prize.

2019 – Industrial Doctorate Fellow (IND2018/IND-9819).

Conference communications

Lo Dico, G.; Carcelén, V.; Haranczyk, M. (2021, December 6-8). *Machine-learning-accelerated experimental characterization and multiobjective design optimization of natural porous materials* [Virtual oral presentation]. MRS Fall Meeting, Symposium DS01-Accelerating Experimental Materials Research with Machine Learning Session DS01.11.02: Closed-Loop and Autonomous IV.

Lo Dico, G.; Carcelén, V.; Haranczyk, M. (2021, November 16-17). *Accelerating design of nanoporous catalyzers assisted by statistical machine learning* [Virtual poster presentation]. RSC Catalysis Science & Technology 10th Anniversary Symposium.

Lo Dico, G.; Croubels S., Carcelén, V.; Haranczyk, M. (2021, September 20-23). *Machine-learning-aided design of composite mycotoxin detoxifier for animal feed* [Virtual oral presentation]. E-MRS Fall Meeting, Symposium H – Modern computational methods and their applications in materials science: Synergy of theory and experiment.

Lo Dico, G.; Croubels S., Carcelén, V.; Haranczyk, M. (2021, July 12-13). *Computer-aided detoxifier material design for animal feed* [Virtual oral and poster presentation]. 15th International conference on materials chemistry (MC15), Session: Materials for changing future.

Lo Dico, G.; Carcelén, V.; Haranczyk, M. (2021, March 11-12). *Accelerating development of promising nanocatalyzers assisted by statistical machine learning* [Virtual oral presentation]. 2nd Webinar on Catalysis & Chemical Engineering

Lo Dico, G.; Carcelén, V.; Haranczyk, M. (2020, September 29-30). *Accelerating development of natural porous materials assisted by statistical machine learning* [Virtual poster presentation]. Chemical Science Symposium 2020: How can machine learning and autonomy accelerate chemistry?

Lo Dico, G.; Carcelén, V.; Haranczyk, M. (2020, July 19-23). *Accelerating development of natural porous materials assisted by statistical machine learning* [Virtual poster presentation]. International Conference on Machine Learning, Optimization, and Data Science (LOD2020) - Sponsored by Springer community.

Workshops

2021, September 24-25 - Computational Methods in Materials Science Fundamentals and Applications (CMMS'21) – Satellite event of the E-MRS Fall Meeting.

2020, July 19 - Biologically Plausible Learning – Satellite event of the International Conference on Machine Learning, Optimization, and Data Science (LOD2020).

2020, July 23 - Integrative Machine Learning – Satellite event of the International Conference on Machine Learning, Optimization, and Data Science (LOD2020).

2020, July 3 - 2nd CRUP/CRUE Open Science Seminar - Career Assessment in the transition to Open Science – UC3M.

Research international stay

July-October, 2020 - Department of Physics and Chemistry, University of Palermo.

Other documents

2021 - First Year Assessment – IMDEA materials

The First Year Assessment (FYA) is required by both IMDEA Materials and UC3M institutions providing a preliminary evaluation of the PhD research project. The FYA has been entirely written by Giulia Lo Dico and reviewed by Maciej Haranczyk. The introduction and methods have been in part incorporated to the Chapter 1 of this PhD dissertation.

Content

Chapter 1 – General introduction and Thesis outline.

1.1. Introduction

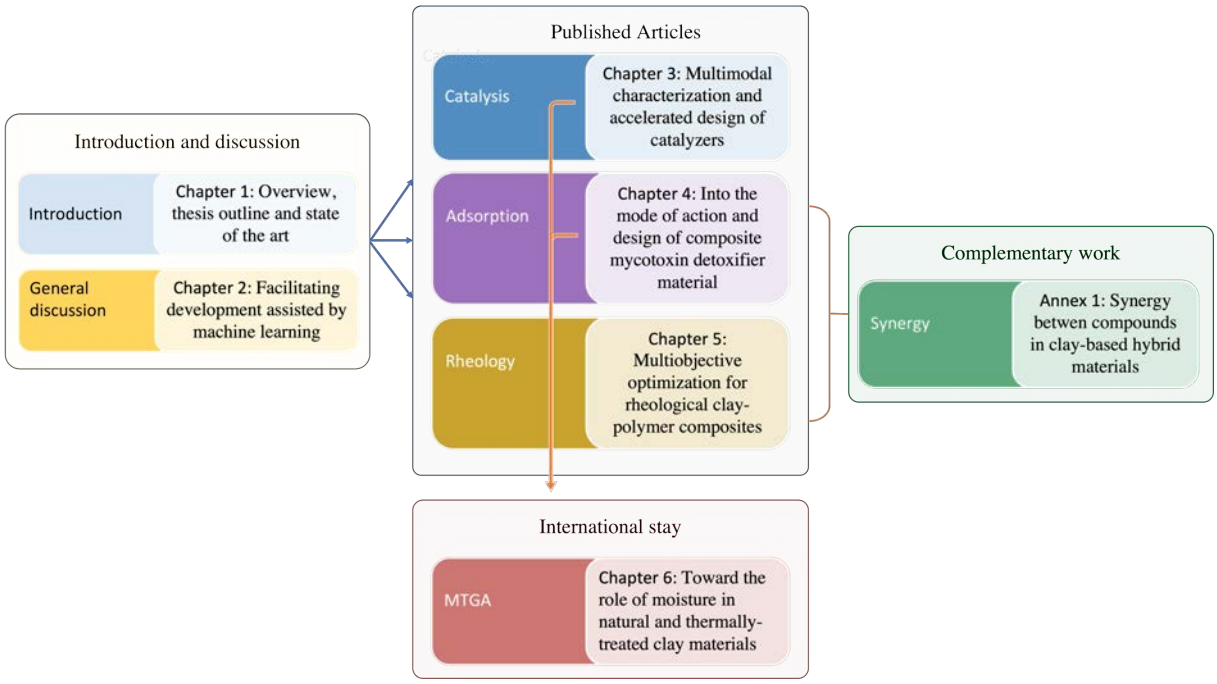
Special clays are natural phyllosilicates arranged in microporous agglomerates.[1] These agglomerates lead to diverse morphologies, e.g., fibrous such as sepiolites, and palygorskites (or attapulgite) and laminar such as smectites. These diverse morphologies correlate with the different physico-chemical features of special clays, and, consequently, with their reactivity. In general, the surfaces of clay particles possess hydrophilic character due to polar groups, e.g., silanols, forming hydrogen bridges.[2] The polar properties of the surfaces, together with a high surface area and available porosity, make clays suitable as catalysts, adsorbents, and rheological additives in both consumer and industrial applications. For example, clay particles are widely implemented as a component of kitty litter formulation[3], as carriers of pesticides and herbicides[4], as adsorbents for water treatment[5] as well as in catalytic applications,[6,7] as reinforcement composite fillers,[8,9] as a fire retardant,[10] and as viscoplastic fluids in horizontal directional drilling operations.[11–13]

Tolsa Company (Tolsa) is the pioneer in research and development of clay-based products. It holds a competitive position thanks to the high quality of raw materials based on sepiolite, palygorskite, and smectites that come from, e.g. mining sites owned and/or exploited for Tolsa. Nowadays, Tolsa expects to exploit its special clays' catalytic, adsorptive, and rheological properties by launching products in new markets and maintaining its competitive position in the existing markets. The conversion from natural sources to final saleable industrial products involves numerous and various treatments, both physical (grinding, micronization, heat treatment) and chemical (surface modification, functionalization, and acid or alkaline treatment).[14–16] The traditional approaches of product development are based on trial-error strategies that require both time and costs. There is therefore a need and an opportunity in adapting new technologies such as data-driven and machine learning techniques to accelerate the development of new clay-based products. This Industrial PhD project supported by a Fellow of the Community of Madrid

(IND2018/IND-9819) has created an opportunity for Tolsa to work together with IMDEA Materials and the University of Carlos III de Madrid to explore the said opportunity.

1.2. Thesis outline

A graphical overview of the organization is presented in Scheme 1.



Scheme 1. Schematic organization of this document.

The initial sections of **Chapter 1** introduce the main goal of this Industrial PhD Thesis and offer an insight on the physico-chemical properties of special clays and their main application fields that are specifically targeted throughout this PhD thesis. The State-of-the-art for machine learning tools in porous materials is also discussed. The latter includes an overview of the datasets, descriptors, and methodology .

Chapter 2 presents this thesis's achievements, giving a general discussion of the three research articles published within this PhD work. The discussion also discloses a novel practice of machine learning models in capturing the synergy between compounds, which is related to the works of

Chapters 4 and 5. The latter is expanded in Annex 1 and is expected to be converted into an additional research publication in the future. A resume of the work realized during the International stay is given and thoroughly presented in Chapter 6.

Chapter 3 includes the "Machine-Learning-Accelerated Multimodal Characterization and Multiobjective Design Optimization of Natural Porous Materials" article published in Chemical Science journal. The contribution demonstrates the first application of machine learning to predict the morphology and surface activity of natural and modified clay-based materials. The suitable proposed material representation allowed us to build predictive algorithms which can be exploited in the design of clay-based acid nanocatalysts.

Chapter 4 moves to "Machine learning-aided design of composite mycotoxin detoxifier material for animal feed", published in Scientific Reports journal. Random forest was educated by a historical set of in vitro adsorption data to screen and identify the optimal formulation. The latter was directly tested in an in vivo trial in collaboration with the Department of Pathobiology, Pharmacology and Zoological Medicine, Faculty of Veterinary Medicine of Ghent University.

Chapter 5 presents the advantages of coupling the advanced design of experiment tools to machine learning in the context of rheological additive design. The work was presented in the article Data-driven experimental design of rheological clay-polymer composites, published in the Industrial & Engineering Chemistry Research journal. The diversity selection algorithm was arranged to identify formulations correlated to the most different outcomes. Three top promising composites were then prepared and tested in their rheological behavior, validating the proposed approach.

Chapter 6 demonstrated the achievements of the work done in collaboration with the Department of Physics and Chemistry of the University of Palermo during my International stay. The work was

completed by running additional experiments in Tolsa laboratories, and the whole story was converted into a manuscript and recently submitted to Microporous and Mesoporous Materials. By using modulated thermogravimetric analysis, the behavior of confined water on the structures of five diverse clays was deeply investigated. The latter allowed us to calculate the activation energy for the removal of the adsorbed water (Ea) and correlate it to adsorption properties.

Annex 1 presents an extension of the use of machine learning models in capturing the synergy between components of nanocomposites. The concept was first explored during the development of the work published and reported in Chapter 4. The predictions of the mycotoxin adsorption followed a non-linear regression within a linear combination of the compounds in the detoxifier formulation. The latter was associated with the ability of machine learning models to consider the synergy between those compounds. The approach was then applied to the work reported in Chapter 5, observing similar behavior. The results will be adapted as a manuscript to be published as a research contribution.

1.3. Morphology and structure of special clays

Clays are traditionally classified as silicon, aluminum, or magnesium oxides. Generally, the silicon resides in the tetrahedron center, equidistant from four oxygens or hydroxyl. The repetition of the tetrahedron in two horizontal directions forms the tetrahedral layer. The aluminum or magnesium cation resides in the octahedral coordination and is connected to the tetrahedron by sharing oxygens. The repetition of one tetrahedral and one octahedral sheet identifies a 1:1 layer structure, while in the 2:1 layer structure, one octahedral sheet connects two tetrahedral sheets. Depending on the organization of such layers, clay minerals are classified as dioctahedral and trioctahedral species. In particular, in trioctahedral structures, the six octahedral sites of the unit cell are occupied, while dioctahedral is referred to structures with only four occupied sites. The natural environment influences particle growth in such a way as to create particle aggregates whose

assembly generates diverse morphology, e.g., lamellar or fibrous. Figure 1 schematizes the 2:1 clay based and resumes their main properties, detailed in the following paragraphs.

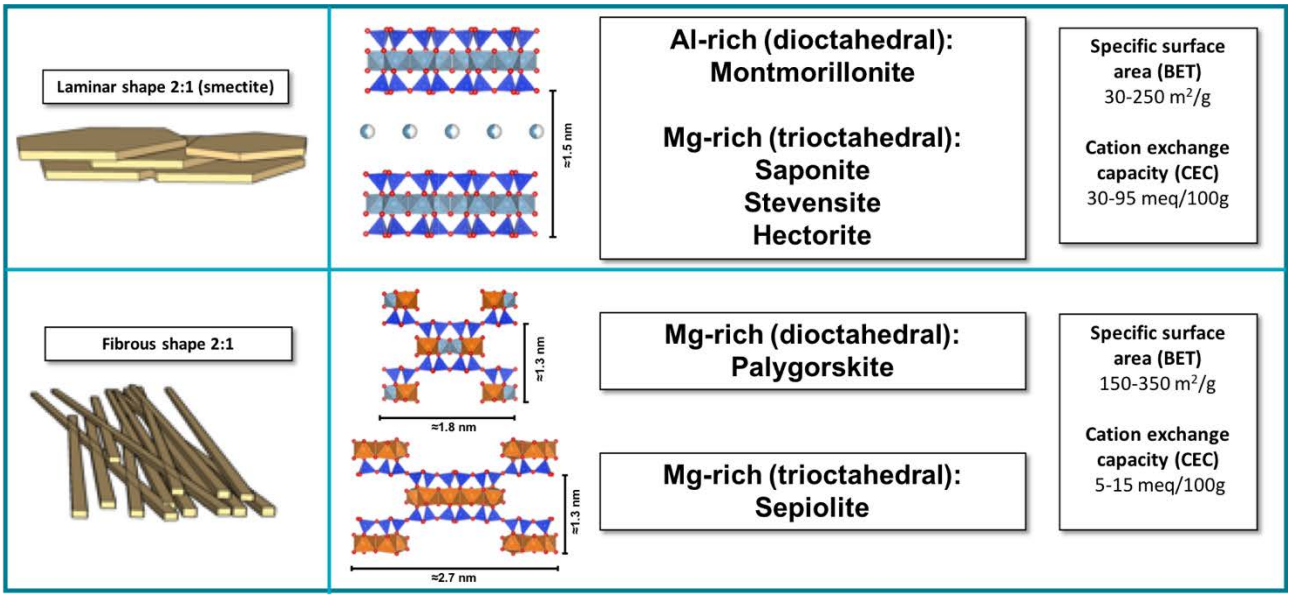


Figure 1. 2:1 clay mineral classification.

1.3.1. Lamellar-shaped 2:1 clays

Smectites, vermiculite, and mica are the principal clay minerals of the 2:1 lamellar-shaped phyllosilicates.[17] The particles possess a variable net negative charge balanced by exchangeable cations (e.g., Na, Ca, and Mg). The layer charge (*X*) of micas is close to 1, and its structure incorporates anhydrous cations between two adjacent layers. On the contrary, smectites and vermiculite possess hydrated cations in the interlayer space. The main difference between them is the substitution of Si^{4+} to Al^{3+} in the tetrahedral sites, which is observed in vermiculite minerals. This phenomenon enhances the layer charge to ~0.6-0.9 compared to one of the smectite groups that range from ~0.2-0.6.[18]. Moreover, although vermiculite is commonly trioctahedral, the octahedral sheet of the smectites can be occupied by trivalent (dioctahedral) or divalent (trioctahedral) cations (Figure 1). The most common smectite mineral is the dioctahedral aluminum montmorillonite. The theoretical composition is SiO_2 66.7% and Al_2O_3 28.3%. However, it is

common to find considerable substitution in the octahedral sheet and some in the tetrahedral sheet.[1] The Mg-rich smectites are generally trioctahedral structures and include saponite, stevensite, and hectorite. Hectorite is a calcium Mg-Li silicate, while saponite and stevensite are sodium Mg silicates. The difference in the chemical composition reflects different properties such as cation exchange capacity (CEC), swelling properties, active surface area, and interactions with inorganic and organic molecules. The layer structure allows intercalating molecules in the interlayer space when clay minerals are in contact with water molecules or salts. This mechanism can be reversible or diffusion-controlled and gives rise to a wide range of possible structures. For example, the intercalation of one or more planes of water molecules expands the interlayer space from ~1 nm (dehydrated layers) to ~1.3 nm (monohydrated layers), ~1.5 nm (bihydrated layers) and ~1.8 nm (trihydrated layers).[17,19] Inorganic cations can be replaced by cationic organic molecules, provoking an expansion of the interlayer space.[19] This phenomenon leads to the swelling of smectite minerals and, in many cases, can cause the delamination or exfoliation of individual silicate layers.[20] The availability of the interlayer space and the control of the modification of the internal surfaces allow the creation of a favorable environment for selective interactions with other molecules.[21]

1.3.2. Acicular fibrous-shaped 2:1 clays

Palygorskite (or attapulgite) and sepiolite are fibrous 2:1 clay minerals generated by the natural transformation of Mg-rich smectites (Figure 1). Their structure consists of ribbons of 2:1 layer linked to the next through an inversion of silica tetrahedron. The inversion of the tetrahedron can be after the repetition of 4 silica units (palygorskite) or 6 silica units (sepiolite), creating channels and tunnels of different sizes. Palygorskites typically possesses a channel size diameter of 4-6 Å, while the sepiolite channels are 5-10 Å. The repeated inverted structure of fibrous clays gives rise to a higher surface area and hydroxyl (OH) density than the 2:1 lamellar shape smectites. Sepiolite shows a trioctahedral structure, while the significant number of vacancies make palygorskite dioctahedral.[22] The ribbons grow in the same direction as the fiber with exchangeable coordinated

cations in the proximity of the oxygen atoms in the octahedral discontinuity. The structure contains two different water molecules, one coordinated to the octahedral cations and the other, named zeolitic water, loosely bonded in the channels. Exposure to a vacuum or the thermal treatment at 100°C removes the zeolitic water. Complete dehydration (>350°C) causes the loss of the coordinated water and, consequently, the structure's folding and the tunnels' disappearance.[23] The zeolitic water is normally lost when the sepiolite is in contact with a vapor or a solution of a guest molecule with an appropriate size. For example, sepiolite can adsorb pyridine molecules or cationic dyes such as methylene blue.[24] The high -OH density on the surfaces of the fibers gives a particular reactivity as solid Brønsted acid. Thus, the cavities can be exploited as a container and a microenvironment for initiating selective reactions such as polymerizing monomers previously adsorbed in the porosity. In that case, the catalytic reactivity can be ascribed to the protons belonging to the coordinated water of the strong Brønsted acid sites. The high reactivity of attapulgite and sepiolite clay minerals may lead to the inactivity of bioactive enzymes requiring a previous functionalization of the clay surfaces.[6,25,26] The size of the fibers mostly depends on the geological environment of the origin.[27] For instance, an aspect ratio of up to 100 and diameters ranging from 10 to 50 nm are usually observed in sepiolite samples from the Taxus Basin (Spain) deposits.[28] The elongated shape of these two minerals results in unique colloidal properties. After ultrasound treatment, they can largely disaggregate in water, creating a percolated network that can sustain co-dispersed compounds or reinforce polymer matrices.[6,8]

1.4. Application of special clays.

Thanks to the wide variety of physicochemical properties of special clays, such as, swelling, cation exchange, accessible surface area, clays are attractive as low cost, eco-friendly and bio-compatible products for catalysis, adsorption, and rheological application. Moreover, clays can accommodate diverse activation and functionalization processes which tune the native properties improving the final performance. According to the aim of this Industrial PhD thesis, we focused the material

design investigation to the three top applications in which Tolsa is well established and desire to increase its portfolio by launching improved products in the markets. The latter are acid nanocatalyzers (i.e., high activated bleaching earths), mycotoxin detoxifier, and horizontal directional drilling fluid. In this section a general overview of the main application of clays is giving with a special emphasis on the ones explored within this thesis.

Catalysis. Clay minerals, even in their natural form, act as catalyzers of various reactions. For example, montmorillonite is one of the active clay for the catalysis of the formation of biopolymers and organic reactions or to catalyze the esterification of fatty acids to produce petroleum-based lubricants.[29–33] Clays can also be exploited as materials to support active nanoparticles with catalytic activity, such as, TiO_2 for photocatalytic degradation of an organic pollulant.[34–37] Their physical and/or chemical modification confer diverse reactivity to the surfaces. For example, attapulgite is a natural catalyst for glucose isomerization to fructose in water and its activity is enhanced by calcination at 500°C. Sepiolite-supported alkaline materials are effective as catalysts for biofuel production via transesterification.[38] Solid acid nanocatalysts can be developed from clay minerals by modification with mineral acids under controlled conditions. The latter leads to a mixture of protonated clay minerals and nanoporous amorphous hydrous silica phases. The increased surface area, porosity, and Brønsted acidity of the modified clays are available for use as catalysts for organic synthesis.[39] For example, acid-modified clays are used ad green catalysts for the hydrolysis of hemicellulosic oligosaccharides.[40] Moreover, clay-based acid nanocatalyzers are commonly used during the oil refining process and known as surface modified bleaching earths (SMBE) or high performance bleaching earths (HPBE). The physical oil refining process includes both bleaching and deodorization/deacidification steps[41]. The modified clays exploit a double action during the bleaching:[42,43]

1) Active adsorbent removing part of the color pigments and many impurities, such as soap, trace metals, phospholipids, oxidation products and polyaromatics.

2) Catalyzer of the acid thermal degradation of large molecules which are then removed during the deodorization step.

Depending on diverse nature of clay minerals, provokes dealumination, or leaching of the Mg from the structure leading high specific surface area and large pore volume.[14,44,45] The resulting catalytic effect of SMBE and HPBE has to be modulated in order to avoid the formation of high level of free fatty acid or other oxidated compounds that can be harmful for human consumption.[43,45–48] The volume of bleaching earths consumed in the world is estimated in over 8MT (OFI, June 2016) and its Market size was valued at USD 3.004 Billion in 2020 and is projected to reach USD 4.22 Billion by 2028, growing at a CAGR of 4.37% from 2021 to 2028 (Verified Market Research, 2020).

Adsorption. Thanks to the available surface area and the active surface, clays are selective adsorbents used in the water treatment, as clarification agents on beverages and juices, and in mitigation of emerging contaminants. In water treatment applications, clays adsorb dispersed organic matter causing turbidity, increase Chemical Oxygen Demand (COD), as well as stabilizing the pH and enhancing the coagulation and flocculation of suspended solids.[49,50] The utilization of clays is common in the clarification of beverages such as wines, musts, ciders, beers, vinegars and juices. The key role of these adsorbents is to remove organic compounds that cause turbidity, as well as other compounds that could cause color, odor, or favor degradations. Clays are also promising materials in the removal of contaminants such as 3-chloropropan-1,2-diol (3-MCPD) esters and oxidates compound which are produced during the oil refining by the extreme temperatures.[51–53] Mycotoxin detoxifiers (MDTs) are another important example of such clay used as selective adsorbent.[54,55] Adsorbents capture mycotoxins which are delivered through the gastrointestinal tract of the animals with their feed, and the mycotoxin-MDT complex is eliminated with feces thus minimizing the absorption in the blood stream.[56] montmorillonites are commonly used for aflatoxins binding by creating complexes with the exchangeable cations, while stevensites exploit their larger surface area for efficient entrapping of ochratoxin A (OTA) and zearalenone (ZEN).[57,58]

Material processing strategies such as the preparation of organo-aluminosilicates have made improvement in the uptake capability of ZEN, OTA and T-2 toxin (T2).[59] Similarly, other porous materials such as activated charcoal (AC) have been recognized as strong adsorbents for several mycotoxins, including deoxynivalenol (DON).[60] However, the required AC doses necessary for a significant detoxification leads to sequestrating of essential micronutrients such as vitamins or minerals.[61] Composite materials based on a mixture of the above may offer to capture and remove much number of mycotoxins while minimizing the interference effect. Besides the regulated mycotoxins, yet-unregulated and emerging mycotoxins occur frequently in agricultural products, thus it is desired that the detoxifier design should address a wide variety of toxic fungal metabolites without compromising the animal health.[62,63] The global market of the mycotoxin detoxifiers is projected to grow at rate of 3.1% from 2020, and reaching a value of USD 3.1 billion by 2027.[64–67]

Rheology. The rheological behavior of clays renders them sustainable for geo-environmental engineering, e.g., for creating a protective layer that impedes the infiltration of water into the underlying soil and limits the outward migration of gases from the underlying waste or contamination.[68] Bentonite-based rheological agents are extensively used to prevent dangerous architectural operations during oil and gas extraction or install underground pipelines, cables, and service conduits through noninvasive horizontal directional drilling operations.[11–13] A typical formulation of a rheological agent is composed of a liquid phase (water, oil, or emulsion) and a clay-based solid phase. The latter may refer to pure clay minerals,[11] organoclays,[69] or multicomponent formulations which incorporate clays,[70,71] polyacrylic or cellulosic polymers,[72–74] surfactants,[75] or hydroxides.[76] Among clays, smectites have been intensively studied for their swelling properties and high cation exchange capacity.[11,70] However, palygorskites and sepiolite are also suitable as thixotropic agents and viscosity controllers.[77,78] The exceptional performance of clay-based rheological agents is typically evaluated by the yield point and the plastic and apparent viscosity. When clays are dispersed in water, they behave as viscoplastic materials, differing from a

Newtonian fluid for their correlation between the shear stress and shear rate. A Newtonian fluid presents a linear correlation giving the shear rate for any shear stress applied, whereas viscoplastic muds start exhibiting a shear rate from the particular stress achieved. The latter is related to clay particle−particle interactions, creating a weak solid network that requires external stress to be broken (reversible process). In particular, laminar clay particles interact predominantly through electrostatic and van der Waals forces through edge-to-face, edge-to-edge, and face-to-face orientations, and the presence of polymeric thickening agents modifies these interactions at the interfaces and, consequently, the viscosity of the fluid.[70] The rheology of a clay dispersion can be described by nonlinear models such as Herschel−Bulkley or can be approximated by the Bingham model, expressed by:

$$\tau = yp + pv^{*} \gamma$$

where τ is the shear stress, yp is the Bingham yield stress, pv is the plastic viscosity, and γ is the shear rate.[11,79,80]

The Market of clays used as rheological additives was valued at US$ 1.62 Bn in 2021 and the size is estimated to grow at a CAGR of 8.1% over the forecast period.

1.5. State of the art

So far, the discovery and optimization of clay-based materials involves trial-and-error methods, while the application of data-driven technics is still poorly investigated. At the same time data-driven approach are extensively explored to accelerate the development of other families of porous materials with great success. This section presents a review of state of the art in the field of the main machine learning and statistical tools, as well as dataset and descriptors frequently used in the context of porous materials. The latter knowledges were the base of the methodologies exploited throughout this PhD thesis.

1.5.1. Data-driven tools in the field of porous materials

In recent years, data-driven approaches and machine learning (ML) techniques have been efficiently implemented in the context of advanced porous crystalline materials, such as synthetic Metallic Organic Frameworks (MOFs), zeolites, or Porous Polymers Networks (PPNs). This strategy provided innovative predictive tools to accelerate the discovery of new materials. The importance of data analytics has been highlighted by the Materials Genome Initiative (MGI)[81] as support for the fast and low-cost development of high-efficient materials. Nowadays, machine learning recovers the younger position in the list of the four paradigms of materials science: 1) empirical trial, 2) physico-chemical laws, 3) computational simulations, and 4) data-driven science.[82] The fourth paradigm perfectly unifies the other three, generating robust Quantitative Structure-Activity/Properties Relationships (QSAR/QSPR) models.[83,84]

1.5.2. Datasets and descriptors of porous materials

Machine learning approaches look for correlations between features and final properties and typically it is a multifactor function. The crucial point in building an efficient predictive model is the quality of the dataset and the definition of the descriptor space. Generally, the datasets for porous materials can be provided by physico-chemical characterization, experimental measurements, or material modeling techniques.[85,86] Various experimental databases are available as open sources such as Mineral Crystal Structure Database[87], which recollects all the crystal structures published in the American Mineralogist; Materials Project[88] and Inorganic Crystal Structure Database (ICSD)[89]. Additional experimental databases of similar porous materials are available such as Computation-Ready, Experimental (CoRE) MOF database[90], which contains synthesized MOF structures from the Cambridge Structural Database (CSD)[91,92], Database of Zeolite Structures[93], as well as databases of predicted MOFs[94], all-silica zeolites[95], PPNs[96], or molecular structures, available from the Crystallography Open Database[97], ZINC[98] and PubChem[99] databases which also contains sequential SMILES (Simplified Molecular Input Line

Entry System) representation. Data quality is a requirement for the machine learning-assisted development of materials science in such a way as to start from a robust and consistent dataset to prevent erroneous information, as underlined by the Materials Genome Initiative.[100] Limited data availability leads to underfitting and high prediction bias in QSAR/QSPR models, and this limitation is intensified if negative examples are rejected.[101,102] Recently, automatic data extraction from materials science literature using natural language processing (NPL) techniques have been implemented to increase the amount of available data, including failed tests.[103,104] For example, Jensen Z. *et al.* presented efficient tools to automatically extract zeolite synthesis and topology data from journal articles to predict synthesis pathways for new zeolite structures.[105]

1.5.3. Machine learning models of structure-property relationships

Machine learning algorithms may be supervised or unsupervised, depending on the nature of the Dataset. Unsupervised algorithms have access to input values (x) and infer patterns without targets or outcomes (y). At the same time, the supervised models derive a function that goes through a specific set of input values predicting the associated y-output values. The first approach effectively analyzes unlabeled data and quickly gives essential information about the relationship among the input data themselves when the output data is unavailable. The second approach investigates the Dataset through a function adjusted by a hyperparameter optimization providing the most accurate methods for screening purposes. The predictions can be achieved by regression if the y varies in a continuous range or by classification if the output values are a discrete set.

Given a set of material structures, their descriptors, and the corresponding set of properties, it is possible to construct a mathematical model of the relationships between descriptors and properties.[106] Statistical machine learning provides numerous frameworks in which the structure-property functions (Eq. 1) can be learned from example data via supervised learning.

$$P = f(x_1, x_2, x_3, ..., x_n) \quad \text{Eq. 1}$$

Where P is the property of interest, f is the function that calculates this property from the descriptors $x_1 \ldots x_n$. The form of $f(x)$ could be linear, multilinear, or nonlinear. More complex machine learning algorithms employ definitions of f based on decision trees, neural networks or other frameworks, the specification of which is expressed by hyperparameters. The latter are usually optimized to achieve the highest quality models given a specific framework. Meanwhile, the framework selection is made by assessing the quality of the resulting models. The models are usually trained and validated using different subsets of the available data (e.g., by splitting the available 80/20). One of the most commonly used frameworks is multilinear regression, random forests, neural networks, and support vector machines. They are implemented in the statistical analysis package R and numerous Python libraries (Sklearn, Scikit-learn, Keras, and TensorFlow). In the following, we will outline the decision tree-based frameworks in more detail as they have a proven track record in nanoporous materials-related applications.

Decision Tree algorithm, introduced by Leo Breiman in 1984, is one of the most common models used in data-mining thanks to its easy interpretability and versatility. Figure 2 depicts an example of a typical workflow of a decision tree classifier that predicts zeolite framework density from synthesis conditions.[107] The tree is a series of top-down questions that build decision rules (branch) starting from the root node, passing to the internal node (decision node) representing the features, and ending to the leaf nodes that correspond to the target classes.

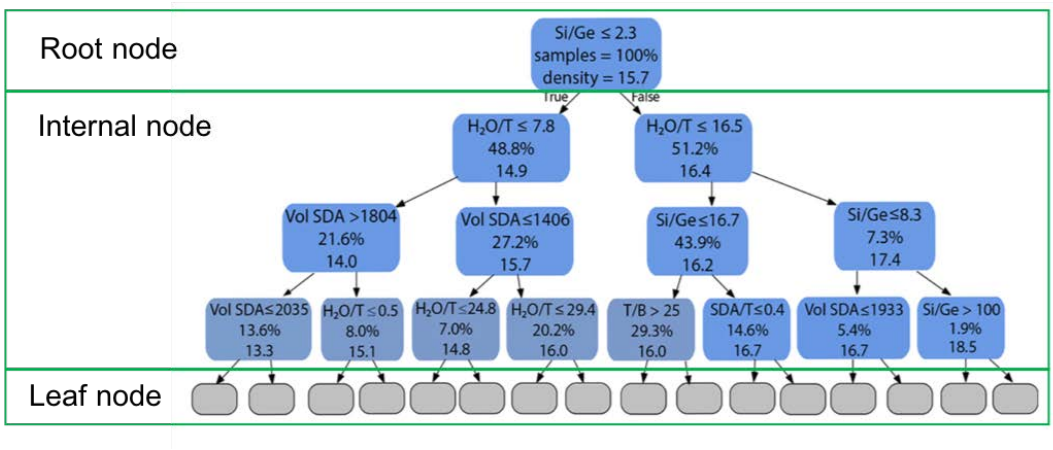


Figure 2. A single decision tree regression model trained to predict zeolite framework density from synthesis conditions (*adapted from Ref. Jensen et al. 2019*).

The core algorithm employs the Gini index (G_i) or Gini impurity (Eq. 2), measuring the degree or probability (p_i) of a particular variable being wrongly classified when randomly chosen.

$$G_i = 1 - \sum_{i=1}^{n} p_i^2 \qquad \text{Eq. 2}$$

In the same way, the Entropy (E_i) is an estimator of homogeneity (Eq. 3), achieving zero value for a homogeneous sample and one if the sample is equally divided. The decision path reduces the level of entropy (Information Gain), selecting the feature/attribute which gives the maximum information about the class and aims. The latter also provides visualization of the most critical features extracting helpful information.

$$E_i = - \sum_{i}^{n} p_i \log_2 (p_i) \qquad \text{Eq. 3}$$

Despite the easy interpretability and fast execution times, decision trees are prone to overfitting, especially when trees are particularly deep.

Alternatively, ensemble learning combines multiple models, named "weak learners", to obtain more accurate and robust predictions. The outcomes depend on the proper architecture of the weak learner and on the method employed to combine them. When a single learning algorithm is used ("homogeneous weak learners"), we can implement (i) bagging, e.g., Random Forest, or (ii) boosting, e.g., Extreme Gradient Boosting, XGBoost, while if we want to combine the diverse type of weak learners ("heterogeneous models"), we may consider (iii) "stacking" approaches. The latter are following explained.

(i) Bagging method is also called **b**ootstrap **agg**regat**ing**. The bootstrap creates a subset by randomly picking objects and duplicating them until the same size as the original Dataset is reached. The remaining objects represent the out-of-bag subset used to assess the model. Once the data is partitioned, *m* models are trained on *m* bootstrap samples, and their outcomes are then aggregated by averaging.

(ii) Boosting method creates the final ensemble model by sequentially adaptively building learners, i.e., the new model depends on the previous ones. The flow starts from a model trained on the training set. Then the following model is created, attempting to correct the errors of the previous one. The learners are added until the training set is perfectly predicted, or the maximum number of models is reached.

(iii) Stacking methods, often used for heterogeneous learners, parallelly trains the models and combine them in a meta-model to compute a prediction based on different weak models' predictions.

A particular case of bagging ensemble learning is the Random Forest algorithm, proposed for the first time by Ho in 1995 and remarked by Breiman in 2001.[108] It represents an ensemble method in which a bunch makes the final prediction of imperfect predictors (i.e., the wisdom of a crowd). Random Forest architecture aims to predict by averaging multiple deep decision trees. The latter are parallelly trained by a random subset of the training set (bootstrapped), hence reducing the variance.

Random forest algorithm applies the general technique of bagging to increase the predictive power, returning the average of *n* individual predictions (regression case) or the most voted class (classification case). A typical workflow can be summarized as follows:

1) Select n random subsets from the training set
2) Build n decision trees splitting each tree by a random subset of features
3) Collect the prediction of the untouched test set from each decision tree
4) Take the mayor voted (classifier) or average the decision trees outcomes (regressor)

The learning process can be controlled by tuning hyperparameters. The most common approach to assess the models is k-fold cross-validation. The data is partitioned into k groups, k-1 of them is used to train the model, which is validated on the remaining fold. The algorithm repeats the process iteratively for k times. The extreme case, for k=n is named leave-one-out cross-validation (LOOCV). The latter configuration ensures that each row of data is allowed to represent the entire testing set and is particularly useful for small amounts of available data. The most critical hyperparameters in Random Forest are:

n_estimators: Number of trees in the forest used to build the model before taking the maximum voting or averages of predictions. Usually, the high number of trees gives accurate performance but makes the code slower.

max_features: Number of features kept in each tree. By increasing this parameter, generally, the model improves its predictability thanks to the many options considered at each node. However, too many features could lead to overfitting.

min_sample_leaf: Minimum number of data points allowed in a leaf node. Smaller leaf helps to capture noise in train data.

min_samples_split: Minimum number of data points placed in a node before splitting the node itself.

max_depth: Maximum number of levels that are considered in each decision tree.

bootstrap: Method for sampling data points (with or without replacement).

random_state: Fixing the reproducibility of model predictions. By setting this parameter, the model always returns the same prediction for a particular dataset.

A common challenge in building structure-property models is inhomogeneous and/or small datasets. In these cases, the selection of the splitting method into training and testing set become crucial to building reasonable predictive models. For example, in our Dataset we have a reduced amount of data exploring a specific descriptor. If that particular information is not given to training our trees, the model would not be able to process an accurate prediction. On the contrary, if the information is exclusively used to train the model, we lose the possibility of evaluating its predictability for that specific data. In these cases, implementing out-of-bag error (oob_score) may be an efficient solution for machine learning models that utilize bootstrap aggregating. When oob_score is applied, the Dataset is not partitioned into training and testing sets. At the same time, the sampling with replacement performed by the bagging process leaves, iteratively, one data to assess the model performance. This approach allows using information about all the available data without leakages.

1.5.4. Material similarity and distance measures

Once the sets of materials structures and their descriptors are available, one can implement basic data analysis functions. Specifically, one can assess the characteristics of materials by comparing the distribution of descriptor values. One can also filter and search structures by comparing the descriptor values of the structures in the Dataset with given query values. Often, the descriptor space is of high dimensionality (e.g., more than 20), and the descriptors are somewhat related. Data analysis tools such as principal component analysis (PCA) can be employed as dimensionality reduction techniques in these cases. In PCA, the original vector space, with n features, is transformed into new k components (k=n). The latter are created by projecting each pristine dimension into the novel components, contributing with different weights. Each k vector

contributes to the total variance with a particular score. Thus, the space may be reduced by removing the least important k component while preserving data information.

More complex analysis of materials often requires a definition of similarity, or distance, between two structures. The motivation is to employ the Similarity Principle, which states that the materials similar in terms of their structure do exhibit similar properties.[109–111] The challenge is implementing the definition of similarity, as the complexity of material structures provides many facets that can be considered. Some domain knowledge is necessary to ensure that the similarity measures consider the important features of the property or the application of interest. The process requires a measure of similarity or distance in a multidimensional descriptor space. For example, for binary vectors, where each bit indicates the presence ("1") or absence ("0") of a particular feature, the Tanimoto (or Jaccard) similarity coefficient has been widely utilized. For features represented by real vectors, Euclidean ($d_{x,y} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$) or Manhattan ($d_{x,y} = \sum_{i=1}^{n} |x_i - y_i|$) distance are commonly used.

The selection of similarity, or distance, a measure typically requires either domain knowledge or some trial-and-error experimentation on smaller sets of examples, i.e., to verify if the considered measure produces results on par with expectations before it is deployed to study large datasets.

Diversity selection. Working with large sets of materials is often required to reduce the size of the sets to facilitate handling. A random sampling of structures to form smaller subsets of the original set is commonly performed. However, when the original set is not homogenous in terms of coverage of the material space, the randomly sampled subset will carry the same characteristics. Instead, one may choose to select subsets that maximize the diversity of the sampled set better to capture the variety present in the original set.

MaxMin maximum-dissimilarity-based selection is a valuable technique for performing this task.[112] Additionally, it does not require the calculation of pairwise distances for all the structures in the original set. The MaxMin approach proceeds as follows:

1) Select a starting structure, e.g., using random selection.

2) Calculate the distance of the remaining structures against every structure which has been selected, storing the minimum observed distance, d.

3) Add the structure that exhibits the highest d to the selected structures.

4) If the specified end criteria still need to be met, go to step 2.

Cluster analysis. An investigated set of materials can be analyzed to identify the groups of similar structures using an unsupervised machine learning technique such as agglomerative hierarchical clustering. The algorithm is unsupervised because it can access input values (descriptors) and infer patterns without target values (i.e., properties). It is a bottom-up approach that can be outlined in the following procedure:

1) Select similarity criterium and compute the distance matrix between the objects in the Dataset.
2) Each object creates its cluster.
3) The pair of clusters identified as the most similar are merged to form a new cluster.
4) The hierarchy is moved up by repeating point 3 until every object is contained in the same cluster.

An alternative, top-down approach, i.e., divisive hierarchical clustering, can also be formulated. One of the advantages of hierarchical clustering is that it can be depicted with a dendrogram in which the cluster contains all objects on one of the sides while on the opposite side of this plot are all clusters containing single elements (singletons). The dendrogram analysis can help identify the appropriate number of clusters based on the distance between clusters. The main disadvantage of hierarchical clustering is that it requires a precalculation of pairwise distances, which can be computationally expensive for large sets of materials. In such cases, alternative approaches, such as k-means clustering, can be employed.[113]

1.6. References

(1) Murray, H. H. Applied Clay Mineralogy: Occurrences, Processing and Application of Kaolins, Bentonites, Palygorskite-Sepiolite, and Common Clays. In *Developments in Clay Science 2*; Elsevier Science, 2006.

(2) Tournassat, C.; Davis, J. A.; Chiaberge, C.; Grangeon, S.; Bourg, I. C. Modeling the Acid-Base Properties of Montmorillonite Edge Surfaces. *Environ. Sci. Technol.* **2016**, *50* (24), 13436–13445. https://doi.org/10.1021/acs.est.6b04677.

(3) Bortzmeyer, D.; Kiefer, J.-C. Animal Litter Composition and Methos for Obtaining Same. 2000.

(4) Sanchez-Martin, M. J.; Rodriguez-Cruz, M. S.; Andrades, M. S.; Sanchez-Camazano, M. Efficiency of Different Clay Minerals Modified with a Cationic Surfactant in the Adsorption of Pesticides: Influence of Clay Type and Pesticide Hydrophobicity. *Appl. Clay Sci.* **2006**, *31* (3–4), 216–228. https://doi.org/10.1016/j.clay.2005.07.008.

(5) Buruga, K.; Song, H.; Shang, J.; Bolan, N.; Jagannathan, T. K.; Kim, K. H. A Review on Functional Polymer-Clay Based Nanocomposite Membranes for Treatment of Water. *J. Hazard. Mater.* **2019**, *379* (April). https://doi.org/10.1016/j.jhazmat.2019.04.067.

(6) Lo Dico, G.; Wicklein, B.; Lisuzzo, L.; Lazzara, G.; Aranda, P.; Ruiz-Hitzky, E. Multicomponent Bionanocomposites Based on Clay Nanoarchitectures for Electrochemical Devices. *Beilstein J. Nanotechnol.* **2019**, *10*, 1303–1315. https://doi.org/10.3762/bjnano.10.129.

(7) Tolsa SA. Reduction Catalysts on Sepiolite Carrier. 1985.

(8) Lisuzzo, L.; Wicklein, B.; Lo Dico, G.; Lazzara, G.; Del Real, G.; Aranda, P.; Ruiz-Hitzky, E. Functional Biohybrid Materials Based on Halloysite, Sepiolite and Cellulose Nanofibers for Health Applications. *Dalt. Trans.* **2020**, *49* (12), 3830–3840. https://doi.org/10.1039/c9dt03804c.

(9) Santaren Rome, J.; Aguilar Diez, E.; Esteban Cubillo, A.; Alvarez Berenguer, A.; Benito

Cano, E.; Garcia Garcia, N.; Guzman Perote, J.; Tiemblo Magro, P. Método Para Producir Un Compuesto Basado En Silicatos Pseudolaminares y El Uso de Los Mismos Como Carga Para Materiales Poliméricos. 2018.

(10) Wicklein, B.; Kocjan, A.; Salazar-Alvarez, G.; Carosio, F.; Camino, G.; Antonietti, M.; Bergström, L. Thermally Insulating and Fire-Retardant Lightweight Anisotropic Foams Based on Nanocellulose and Graphene Oxide. *Nat. Nanotechnol.* **2015**, *10* (3), 277–283. https://doi.org/10.1038/nnano.2014.248.

(11) Zhang, J. R.; Xu, M. D.; Christidis, G. E.; Zhou, C. H. Clay Minerals in Drilling Fluids: Functions and Challenges. *Clay Miner.* **2020**, *55* (1), 1–11. https://doi.org/10.1180/clm.2020.10.

(12) Kroepsch, A. C.; Maniloff, P. T.; Adgate, J. L.; McKenzie, L. M.; Dickinson, K. L. Environmental Justice in Unconventional Oil and Natural Gas Drilling and Production: A Critical Review and Research Agenda. *Environ. Sci. Technol.* **2019**, *53* (12), 6601–6615. https://doi.org/10.1021/acs.est.9b00209.

(13) Gandhi, S. M.; Sarkar, B. C. *Chapter 8 – Drilling*; 2016. https://doi.org/10.1016/B978-0-12-805329-4.00015-6.

(14) Ortiz Niembro, J. A. P.; Solis Santamaria, G.; Thomassiny Villaurratia, E.; Ruf, F. Method of Producing Amorphous Adsorbent and Its Use in the Bleaching of Fats and/or Oils. 2009.

(15) Alvarez Berenguer, A.; Ballesteros Martin, C. A Process for Bleaching and Increasing the Ion Exchange Capability of Sepiolite. *European Patent Application*. 1988. https://doi.org/10.1017/CBO9781107415324.004.

(16) Alvarez Berenguer, A.; Barrasa, C.; Cubillo, A. E.; Gravalos Moreno, J.; Sanchez Rojo, J.; Rome, J. S.; Agullo, J. V. Process for the Preparation of an Additive Comprising Supported and Dispersed TiO2 Particles. *United States Patent*. 2017. https://doi.org/10.2307/1190003.

(17) Brigatti, M. F.; Malferrari, D.; Laurora, A.; Elmi, C. Structure and Mineralogy of Layer Silicates. *Layer. Miner. Struct. their Appl. Adv. Technol.* **2012**, No. January, 1–71.

https://doi.org/10.1180/emu-notes.11.1.

(18) Christidis, G. E. The Concept of Layer Charge of Smectites and Its Implications for Important Smectite-Water Properties. *Layer. Miner. Struct. their Appl. Adv. Technol.* **2012**, No. April, 237–258. https://doi.org/10.1180/emu-notes.11.6.

(19) Vishnu Mahesh, K. R.; Narasimha Murthy, H. N.; Kumaraswamy, E.; Raghavendra, N.; Sridhar, R.; Krishna, M.; Pattar, N.; Pal, R.; Sherigara, B. S. Synthesis and Characterization of Organomodified Na-MMT Using Cation and Anion Surfactants. *Front. Chem. China* **2011**, *6* (2), 153–158. https://doi.org/10.1007/s11458-011-0239-4.

(20) Daab, M.; Eichstaedt, N. J.; Edenharter, A.; Rosenfeldt, S.; Breu, J. Layer Charge Robust Delamination of Organo-Clays. *RSC Adv.* **2018**, *8* (50), 28797–28803. https://doi.org/10.1039/c8ra05318a.

(21) Lu, Y.; Li, Y.; Liu, D.; Ning, Y.; Yang, S.; Yang, Z. Adsorption of Benzene Vapor on Natural Silicate Clay Minerals under Different Moisture Contents and Binary Mineral Mixtures. *Colloids Surfaces A Physicochem. Eng. Asp.* **2019**, *585* (September 2019). https://doi.org/10.1016/j.colsurfa.2019.124072.

(22) Singer, A.; Galan, E. Developments in Palygorskite-Sepiolite Research. In *Development in Clay Science*; 2011.

(23) Bergaya, F.; Theng, B. K. G.; Lagaly, G. *Handbook of Clay Science*; 2006; Vol. 1. https://doi.org/10.1016/S1572-4352(05)01039-1.

(24) Ruiz-Hitzky, E. Molecular Access to Intracrystalline Tunnels of Sepiolite. *J. Mater. Chem.* **2001**, *11* (1), 86–91. https://doi.org/10.1039/b003197f.

(25) Huang, J.; Liu, Y.; Wang, X. Selective Adsorption of Tannin from Flavonoids by Organically Modified Attapulgite Clay. *J. Hazard. Mater.* **2008**, *160* (2–3), 382–387. https://doi.org/https://doi.org/10.1016/j.jhazmat.2008.03.008.

(26) Wicklein, B.; Darder, M.; Aranda, P.; Ruiz-hitzky, E. Phospholipid–Sepiolite Biomimetic Interfaces for the Immobilization of Enzymes. *ACS Appl. Mater. interfaces* **2011**, *3* (11), 4339–

4348.

(27) Ruiz-Hitzky, E.; Aranda, P.; Álvarez, A.; Santarén, J.; Esteban-Cubillo, A. *Advanced Materials and New Applications of Sepiolite and Palygorskite*; 2011; Vol. 3. https://doi.org/10.1016/B978-0-444-53607-5.00017-7.

(28) Ruiz-hitzky, E.; Darder, M.; Alcantara, A. C. S.; Wicklein, B.; Aranda, P. Functional Nanocomposites Based on Fibrous Clays. In *Functional Polymer Composites with Nanoclays*; Lvov, Y., Guo, B., Fakrullin, R. F., Eds.; The Royal Society of Chemistry, 2016; pp 1–53. https://doi.org/10.1039/9781782624226-00001.

(29) Luna, F. M. T.; Cecilia, J. A.; Saboya, R. M. A.; Barrera, D.; Sapag, K.; Rodríguez-Castellón, E.; Cavalcante, C. L. Natural and Modified Montmorillonite Clays as Catalysts for Synthesis of Biolubricants. *Materials (Basel).* **2018**, *11* (9), 6–9. https://doi.org/10.3390/ma11091764.

(30) Boudissa, F.; Mirilà, D.; Arus, V. A.; Terkmani, T.; Semaan, S.; Proulx, M.; Nistor, I. D.; Roy, R.; Azzouz, A. Acid-Treated Clay Catalysts for Organic Dye Ozonation – Thorough Mineralization through Optimum Catalyst Basicity and Hydrophilic Character. *J. Hazard. Mater.* **2019**, *364*, 356–366. https://doi.org/10.1016/j.jhazmat.2018.09.070.

(31) Ferris, J. P. Montmorillonite-Catalysed Formation of RNA Oligomers: The Possible Role of Catalysis in the Origins of Life. *Philos. Trans. R. Soc. B Biol. Sci.* **2006**, *361* (1474), 1777–1786. https://doi.org/10.1098/rstb.2006.1903.

(32) Aldersley, M. F.; Joshi, P. C.; Price, J. D.; Ferris, J. P. The Role of Montmorillonite in Its Catalysis of RNA Synthesis. *Appl. Clay Sci.* **2011**, *54* (1), 1–14. https://doi.org/10.1016/j.clay.2011.06.011.

(33) Ferris, J. P. Mineral Catalysis and Prebiotic Synthesis : Formation of RNA. *Orig. Life* **2005**, 145–150.

(34) Ökte, A. N.; Sayinsöz, E. Characterization and Photocatalytic Activity of TiO2 Supported Sepiolite Catalysts. *Sep. Purif. Technol.* **2008**, *62* (3), 535–543. https://doi.org/10.1016/j.seppur.2008.03.011.

(35) Bakhtiar, A.; Bouberka, Z.; Roussel, P.; Volkringer, C.; Addad, A.; Ouddane, B.; Pierlot, C.; Maschke, U. Development of a TiO2/Sepiolite Photocatalyst for the Degradation of a Persistent Organic Pollutant in Aqueous Solution. *Nanomaterials* **2022**, *12* (19), 3313. https://doi.org/10.3390/nano12193313.

(36) Berenguer, A. A.; Barrasa, A. M. C.; Cubillo, A. E.; Gravalos Moreno, J.; Sanchez Rojo, A. J.; Santaren Rome, J.; Vera Agullo, J. Process for the Preparation of an Additive Comprising Supported and Dispersed TiO2 Particles. 2017.

(37) Liu, R.; Ji, Z.; Wang, J.; Zhang, J. Mesocrystalline TiO2/Sepiolite Composites for the Effective Degradation of Methyl Orange and Methylene Blue. *Front. Mater. Sci.* **2018**, *12* (3), 292–303. https://doi.org/10.1007/s11706-018-0429-9.

(38) Degirmenbasi, N.; Boz, N.; Kalyon, D. M. Biofuel Production via Transesterification Using Sepiolite-Supported Alkaline Catalysts. *Appl. Catal. B Environ.* **2014**, *150–151*, 147–156. https://doi.org/10.1016/j.apcatb.2013.12.013.

(39) Nagendrappa, G. Organic Synthesis Using Clay and Clay-Supported Catalysts. *Appl. Clay Sci.* **2011**, *53* (2), 106–138. https://doi.org/10.1016/j.clay.2010.09.016.

(40) Vilcocq, L.; Spinola, V.; Moniz, P.; Duarte, L. C.; Carvalheiro, F.; Fernandes, C.; Castilho, P. Acid-Modified Clays as Green Catalysts for the Hydrolysis of Hemicellulosic Oligosaccharides. *Catal. Sci. Technol.* **2015**, *5* (8), 4072–4080. https://doi.org/10.1039/c5cy00195a.

(41) List, G. R. Bleaching and Purifying Fats and Oils Theory and Practice.; Academic Press and AOCS Press, 2009.

(42) Zschau, W. Bleaching of Edible Fats and Oils. *Eur. J. Lipid Sci. Technol.* **2001**, *103* (8), 505–551. https://doi.org/10.1002/1438-9312(200108)103:8<505::AID-EJLT505>3.0.CO;2-7.

(43) Sampaio, K. A.; Ceriani, R.; Silva, S. M.; Taham, T.; Meirelles, A. J. A. Steam Deacidification of Palm Oil. *Food Bioprod. Process.* **2011**, *89* (4), 383–390. https://doi.org/10.1016/j.fbp.2010.11.012.

(44) Almeida, E. S.; Carvalho, A. C. B.; Soares, I. O. de S.; Valadares, L. F.; Mendonça, A. R. V.; Silva, I. J.; Monteiro, S. Elucidating How Two Different Types of Bleaching Earths Widely Used in Vegetable Oils Industry Remove Carotenes from Palm Oil: Equilibrium, Kinetics and Thermodynamic Parameters. *Food Res. Int.* **2019**, *121* (January), 785–797. https://doi.org/10.1016/j.foodres.2018.12.061.

(45) Silva, S. M.; Sampaio, K. A.; Ceriani, R.; Verhé, R.; Stevens, C.; De Greyt, W.; Meirelles, A. J. A. Effect of Type of Bleaching Earth on the Final Color of Refined Palm Oil. *LWT - Food Sci. Technol.* **2014**, *59* (2P2), 1258–1264. https://doi.org/10.1016/j.lwt.2014.05.028.

(46) Liang, J.; Appukuttan Aachary, A.; Hollader, U. T. Hemp Seed Oil: Minor Components and Oil Quality. *Lipid Technol.* **2015**, *27* (10), 231–233. https://doi.org/10.1002/lite.201500050.

(47) Tan, C. H.; Ariffin, A. A.; Ghazali, H. M.; Tan, C. P.; Kuntom, A.; Choo, A. C. Y. Changes in Oxidation Indices and Minor Components of Low Free Fatty Acid and Freshly Extracted Crude Palm Oils under Two Different Storage Conditions. *J. Food Sci. Technol.* **2017**, *54* (7), 1757–1764. https://doi.org/10.1007/s13197-017-2569-9.

(48) Sampaio, K. A.; Ayala, J. V; Silva, S. M.; Ceriani, R.; Verhé, R.; Meirelles, A. J. A. Thermal Degradation Kinetics of Carotenoids in Palm Oil. *JAOCS, J. Am. Oil Chem. Soc.* **2013**, *90* (2), 191–198. https://doi.org/10.1007/s11746-012-2156-1.

(49) Uddin, M. K. A Review on the Adsorption of Heavy Metals by Clay Minerals, with Special Focus on the Past Decade. *Chem. Eng. J.* **2017**, *308*, 438–462. https://doi.org/10.1016/j.cej.2016.09.029.

(50) Awasthi, A.; Jadhao, P.; Kumari, K. Clay Nano-adsorbent: Structures, Applications and Mechanism for Water Treatment. *SN Appl. Sci.* **2019**, *1* (9), 1–21. https://doi.org/10.1007/s42452-019-0858-9.

(51) Oey, S. B.; van der Fels-Klerx, H. J.; Fogliano, V.; van Leeuwen, S. P. J. Mitigation Strategies for the Reduction of 2- and 3-MCPD Esters and Glycidyl Esters in the Vegetable Oil Processing Industry. *Compr. Rev. Food Sci. Food Saf.* **2019**, *18* (2), 349–361.

https://doi.org/10.1111/1541-4337.12415.

(52) Yao, Y.; Cao, R.; Liu, W.; Zhou, H.; Li, C.; Wang, S. Molecular Reaction Mechanism for the Formation of 3-Chloropropanediol Esters in Oils and Fats. *Journal of Agricultural and Food Chemistry*. 2019, pp 2700–2708. https://doi.org/10.1021/acs.jafc.8b06632.

(53) Ramli, M. R.; Siew, W. L.; Ibrahim, N. A.; Hussein, R.; Kuntom, A.; Razak, R. A. A.; Nesaretnam, K. Effects of Degumming and Bleaching on 3-MCPD Esters Formation during Physical Refining. *JAOCS, J. Am. Oil Chem. Soc.* **2011**, *88* (11), 1839–1844. https://doi.org/10.1007/s11746-011-1858-0.

(54) Çelik, K. The Efficacy of Mycotoxin-Detoxifying and Biotransforming Agents in Animal Nutrition. *Nanomycotoxicology Treat. Mycotoxins Nano W.* **2019**, 271–284. https://doi.org/10.1016/B978-0-12-817998-7.00012-4.

(55) Pirouz, A. A.; Selamat, J.; Iqbal, S. Z.; Mirhosseini, H.; Karjiban, R. A.; Bakar, F. A. The Use of Innovative and Efficient Nanocomposite (Magnetic Graphene Oxide) for the Reduction on of Fusarium Mycotoxins in Palm Kernel Cake. *Sci. Rep.* **2017**, *7* (1), 1–9. https://doi.org/10.1038/s41598-017-12341-3.

(56) Huwig, A.; Freimund, S.; Käppeli, O.; Dutler, H. Mycotoxin Detoxication of Animal Feed by Different Adsorbents. *Toxicol. Lett.* **2001**, *122* (2), 179–188. https://doi.org/10.1016/S0378-4274(01)00360-5.

(57) Pappas, A. C.; Tsiplakou, E.; Georgiadou, M.; Anagnostopoulos, C.; Markoglou, A. N.; Liapis, K.; Zervas, G. Bentonite Binders in the Presence of Mycotoxins: Results of in Vitro Preliminary Tests and an in Vivo Broiler Trial. *Appl. Clay Sci.* **2014**, *99*, 48–53. https://doi.org/10.1016/j.clay.2014.06.009.

(58) Sohling, U.; Haimerl, A. Use of Stevensite for Mycotoxin Adsorption. WO2006119967A1, 2012.

(59) Schall, N.; Simmler-hibenthal, H.; Feldhaus, H. G. Mycotoxin Adsorbents. 2004.

(60) Devreese, M.; Antonissen, G.; De Backer, P.; Croubels, S. Efficacy of Active Carbon

towards the Absorption of Deoxynivalenol in Pigs. *Toxins (Basel).* **2014**, *6* (10), 2998–3004. https://doi.org/10.3390/toxins6102998.

(61) Kihal, A.; Rodriguez-Prado, M.; Godoy, C.; Cristofol, C.; Calsamiglia, S. In Vitro Assessment of the Capacity of Certain Mycotoxin Binders to Adsorb Some Amino Acids and Water-Soluble Vitamins. *J. Dairy Sci.* **2020**, *103* (4), 3125–3132. https://doi.org/10.3168/jds.2019-17561.

(62) Fraeyman, S.; Croubels, S.; Devreese, M.; Antonissen, G. Emerging Fusarium and Alternaria Mycotoxins: Occurrence, Toxicity and Toxicokinetics. *Toxins (Basel).* **2017**, *9* (7), 1–26. https://doi.org/10.3390/toxins9070228.

(63) Gruber-Dorninger, C.; Novak, B.; Nagl, V.; Berthiller, F. Emerging Mycotoxins: Beyond Traditionally Determined Food Contaminants. *J. Agric. Food Chem.* **2017**, *65* (33), 7052–7070. https://doi.org/10.1021/acs.jafc.6b03413.

(64) European Union. Commission Recommendation 2006/579/EC of 17 August 2006 on the Presence of Deoxynivalenol, Zearalenone, Ochratoxin A, T-2 and HT-2 and Fumonisins in Products Intended for Animal Feeding. *Off. J. Eur. Union L 229/7* **2006**, *L 229/7*, 7–9.

(65) European Union. Commission Regulation (EC) No 1881/2006 of 19 December 2006 Setting Maximum Levels for Certain Contaminants in Foodstuffs. *Off. J. Eur. Union* **2006**, *L 364*, 5–24.

(66) Dohlman, E. Mycotoxin Hazards and Regulations. *Int. trade food Saf. Econ. theory case Stud.* **2003**, 97.

(67) MarketWatch. *Feed Mycotoxin Binders and Modifiers Market Size to Surpass US$ 3.21 Billion by 2027*; 2020.

(68) Shackelford, C. D. Geoenvironmental Engineering. In *Encyclopedia of Physical Science and Technology*; Meyers, R. A., Ed.; Elsevier, 2003; pp 601–621. https://doi.org/10.1016/B0-12-227410-5/00879-6.

(69) Weng, J.; Gong, Z.; Liao, L.; Lv, G.; Tan, J. Comparison of Organo-Sepiolite Modified by

Different Surfactants and Their Rheological Behavior in Oil-Based Drilling Fluids. *Appl. Clay Sci.* **2018**, *159* (October 2017), 94–101. https://doi.org/10.1016/j.clay.2017.12.031.

(70) Jung, Y.; Son, Y. H.; Lee, J. K.; Phuoc, T. X.; Soong, Y.; Chyu, M. K. Rheological Behavior of Clay-Nanoparticle Hybrid-Added Bentonite Suspensions: Specific Role of Hybrid Additives on the Gelation of Clay-Based Fluids. *ACS Appl. Mater. Interfaces* **2011**, *3* (9), 3515–3522. https://doi.org/10.1021/am200742b.

(71) Pang, J.; Liang, Y.; Masuda, Y.; Matsuoka, T.; Zhang, Y.; Xue, Z. Swelling Phenomena of the Nonswelling Clay Induced by CO 2 and Water Cooperative Adsorption in Janus-Surface Micropores. *Environ. Sci. Technol.* **2020**, *54* (9), 5767–5773. https://doi.org/10.1021/acs.est.0c00499.

(72) Li, M.-C.; Wu, Q.; Song, K.; F. De Hoop, C.; Lee, S.; Qing, Y.; Wu, Y. Cellulose Nanocrystals and Polyanionic Cellulose as Additives in Bentonite Water-Based Drilling Fluids: Rheological Modeling and Filtration Mechanisms. *Ind. Eng. Chem. Res.* **2016**, *55* (1), 133–143. https://doi.org/10.1021/acs.iecr.5b03510.

(73) Jain, R.; Mahto, V. Evaluation of Polyacrylamide/Clay Composite as a Potential Drilling Fluid Additive in Inhibitive Water Based Drilling Fluid System. *J. Pet. Sci. Eng.* **2015**, *133*, 612–621. https://doi.org/10.1016/j.petrol.2015.07.009.

(74) Li, M. C.; Wu, Q.; Song, K.; Qing, Y.; Wu, Y. Cellulose Nanoparticles as Modifiers for Rheology and Fluid Loss in Bentonite Water-Based Fluids. *ACS Appl. Mater. Interfaces* **2015**, *7* (8), 5009–5016. https://doi.org/10.1021/acsami.5b00498.

(75) Shettigar, R. R.; Misra, N. M.; Patel, K. Cationic Surfactant (CTAB) a Multipurpose Additive in Polymer-Based Drilling Fluids. *J. Pet. Explor. Prod. Technol.* **2018**, *8* (2), 597–606. https://doi.org/10.1007/s13202-017-0357-8.

(76) Gimba, A. S. B.; Amakhabi, S. O.; Ogolo, O.; waseun Alonge, O.; Nzerem, P.; Okafor, I.; Afolayan, D. Evaluation of the Potential of Calcium Hydroxide Synthesized from Eggshells as a Drilling Fluid Additive. *Pet. Coal* **2020**, *62* (1), 1–10.

(77) Neaman, A.; Singer, A. Rheological Properties of Aqueous Suspensions of Palygorskite. *Soil Sci. Soc. Am. J.* **2000**, *64* (1), 427–436. https://doi.org/10.1016/0144-8617(89)90041-6.

(78) Abdo, J.; AL-Sharji, H.; Hassan, E. Effects of Nano-Sepiolite on Rheological Properties and Filtration Loss of Water-Based Drilling Fluids. *Surf. Interface Anal.* **2016**, *48* (7), 522–526. https://doi.org/10.1002/sia.5997.

(79) Jeong, S. W.; Locat, J.; Leroueil, S. The Effects of Salinity and Shear History on the Rheological Characteristics of Illite-Rich and Na-Montmorillonite-Rich Clays. *Clays Clay Miner.* **2012**, *60* (2), 108–120. https://doi.org/10.1346/CCMN.2012.0600202.

(80) Rehm, B.; Haghshenas, A. Flow Drilling: Underbalance Drilling with Liquid Single-Phase Systems. In *Underbalanced Drilling: Limits and Extremes*; 2012; pp 39–58. https://doi.org/10.1016/B978-1-933762-05-0.50009-7.

(81) National Science and Technology Council. Materials Genome Initiative for Global Competitiveness. **2011**.

(82) Agrawal, A.; Choudhary, A. Perspective: Materials Informatics and Big Data: Realization of the "Fourth Paradigm" of Science in Materials Science. *APL Mater.* **2016**, *4* (5). https://doi.org/10.1063/1.4946894.

(83) Berhanu, W. M.; Pillai, G. G.; Oliferenko, A. A.; Katritzky, A. R. Quantitative Structure-Activity/Property Relationships: The Ubiquitous Links between Cause and Effect. *Chempluschem* **2012**, *77* (7), 507–517. https://doi.org/10.1002/cplu.201200038.

(84) Le, T.; Epa, V. C.; Burden, F. R.; Winkler, D. A. Quantitative Structure-Property Relationship Modeling of Diverse Materials Properties. *Chem. Rev.* **2012**, *112* (5), 2889–2919. https://doi.org/10.1021/cr200066h.

(85) Evans, J. D.; Jelfs, K. E.; Day, G. M.; Doonan, C. J. Application of Computational Methods to the Design and Characterisation of Porous Molecular Materials. *Chem. Soc. Rev.* **2017**, *46* (11), 3286–3301. https://doi.org/10.1039/c7cs00084g.

(86) Gopakumar, A. M.; Balachandran, P. V; Xue, D.; Gubernatis, J. E. Multi-Objective

Optimization for Materials Discovery via Adaptive Design. *Sci. Rep.* **2018**, No. February, 1–12. https://doi.org/10.1038/s41598-018-21936-3.

(87) The American Mineralogist Crystal Structure Database. *American Mineralogist*. 2017. https://doi.org/http://doi.org/10.17616/R32S57.

(88) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Mater.* **2013**, *1* (1). https://doi.org/10.1063/1.4812323.

(89) Technology, N. I. of S. NIST Inorganic Crystal Structure Database. https://doi.org/https://doi.org/10.18434/M32147.

(90) Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q. Computation-Ready, Experimental Metal-Organic Frameworks: A Tool to Enable High-Throughput Screening of Nanoporous Crystals. *Chem. Mater.* **2014**, *26* (21), 6185–6192. https://doi.org/10.1021/cm502594j.

(91) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **2016**, *72* (2), 171–179. https://doi.org/10.1107/S2052520616003954.

(92) Moghadam, P. Z.; Li, A.; Wiggin, S. B.; Tao, A.; Maloney, A. G. P.; Wood, P. A.; Ward, S. C.; Fairen-Jimenez, D. Development of a Cambridge Structural Database Subset: A Collection of Metal-Organic Frameworks for Past, Present, and Future. *Chem. Mater.* **2017**, *29* (7), 2618–2625. https://doi.org/10.1021/acs.chemmater.7b00441.

(93) International Zeolite Association. Database of Zeolite Structures. 2018. https://doi.org/http://doi.org/10.17616/R3HS6N.

(94) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. Large-Scale Screening of Hypothetical Metal-Organic Frameworks. *Nat. Chem.* **2012**, *4* (2), 83–89. https://doi.org/10.1038/nchem.1192.

(95) Pophale, R.; Cheeseman, P. A.; Deem, M. W. A Database of New Zeolite-like Materials. *Phys. Chem. Chem. Phys.* **2011**, *13* (27), 12407–12412. https://doi.org/10.1039/c0cp02255a.

(96) Martin, R. L.; Simon, C. M.; Smit, B.; Haranczyk, M. In Silico Design of Porous Polymer Networks: High-Throughput Screening for Methane Storage Materials. *J. Am. Chem. Soc.* **2014**, *136* (13), 5006–5022. https://doi.org/10.1021/ja4123939.

(97) Graulis, S.; Chateigner, D.; Downs, R. T.; Yokochi, A. F. T.; Quirós, M.; Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P.; Le Bail, A. Crystallography Open Database - An Open-Access Collection of Crystal Structures. *J. Appl. Crystallogr.* **2009**, *42* (4), 726–729. https://doi.org/10.1107/S0021889809016690.

(98) Irwin, J. J.; Shoichet, B. K. ZINC - A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2005**, *45* (1), 177–182. https://doi.org/10.1021/ci049714+.

(99) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem 2019 Update: Improved Access to Chemical Data. *Nucleic Acids Res.* **2019**, *47* (D1), D1102--D1109. https://doi.org/10.1093/nar/gky1033.

(100) Widener, A. Materials Genome Initiative. In *Chemical & Engineering News*; 2013; Vol. 91, pp 25–27.

(101) Cohen, S. Machines Learn Chemical Intuition. *Nature* **2019**, *566* (2015), 464–465.

(102) Raccuglia, P.; Elbert, K. C.; Adler, P. D. F.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J.; Norquist, A. J. Machine-Learning-Assisted Materials Discovery Using Failed Experiments. *Nature* **2016**, *533* (7601), 73–76. https://doi.org/10.1038/nature17439.

(103) Kim, E.; Huang, K.; Saunders, A.; McCallum, A.; Ceder, G.; Olivetti, E. Materials Synthesis Insights from Scientific Literature via Text Extraction and Machine Learning. *Chem. Mater.* **2017**, *29* (21), 9436–9444. https://doi.org/10.1021/acs.chemmater.7b03500.

(104) Moliner, M.; Román-Leshkov, Y.; Corma, A. Machine Learning Applied to Zeolite Synthesis: The Missing Link for Realizing High-Throughput Discovery. *Acc. Chem. Res.* **2019**, *52* (10), 2971–2980. https://doi.org/10.1021/acs.accounts.9b00399.

(105) Jensen, Z.; Kim, E.; Kwon, S.; Gani, T. Z. H.; Román-Leshkov, Y.; Moliner, M.; Corma, A.; Olivetti, E. A Machine Learning Approach to Zeolite Synthesis Enabled by Automatic Literature Data Extraction. *ACS Cent. Sci.* **2019**. https://doi.org/10.1021/acscentsci.9b00193.

(106) Pardakhti, M.; Moharreri, E.; Wanik, D.; Suib, S. L.; Srivastava, R. Machine Learning Using Combined Structural and Chemical Descriptors for Prediction of Methane Adsorption Performance of Metal Organic Frameworks (MOFs). *ACS Comb. Sci.* **2017**, *19* (10), 640–645. https://doi.org/10.1021/acscombsci.7b00056.

(107) Jensen, Z.; Kim, E.; Kwon, S.; Gani, T. Z. H.; Román-Leshkov, Y.; Moliner, M.; Corma, A.; Olivetti, E. A Machine Learning Approach to Zeolite Synthesis Enabled by Automatic Literature Data Extraction. *ACS Cent. Sci.* **2019**, *5* (5), 892–899. https://doi.org/10.1021/acscentsci.9b00193.

(108) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. https://doi.org/10.1007/978-3-662-56776-0_10.

(109) Han, J.; Kamber, M.; Pei, J. *Getting to Know Your Data*; 2012. https://doi.org/10.1016/b978-0-12-381479-1.00002-2.

(110) Leach, A. R.; Gillet, V. J. *An Introduction to Chemoinformatics*; Springer, 2007. https://doi.org/10.1007/978-1-4020-6291-9.

(111) Johnson GMM, M. A. *Concepts and Applications of Molecular Similarity*; John Wiley & Sons, 1990.

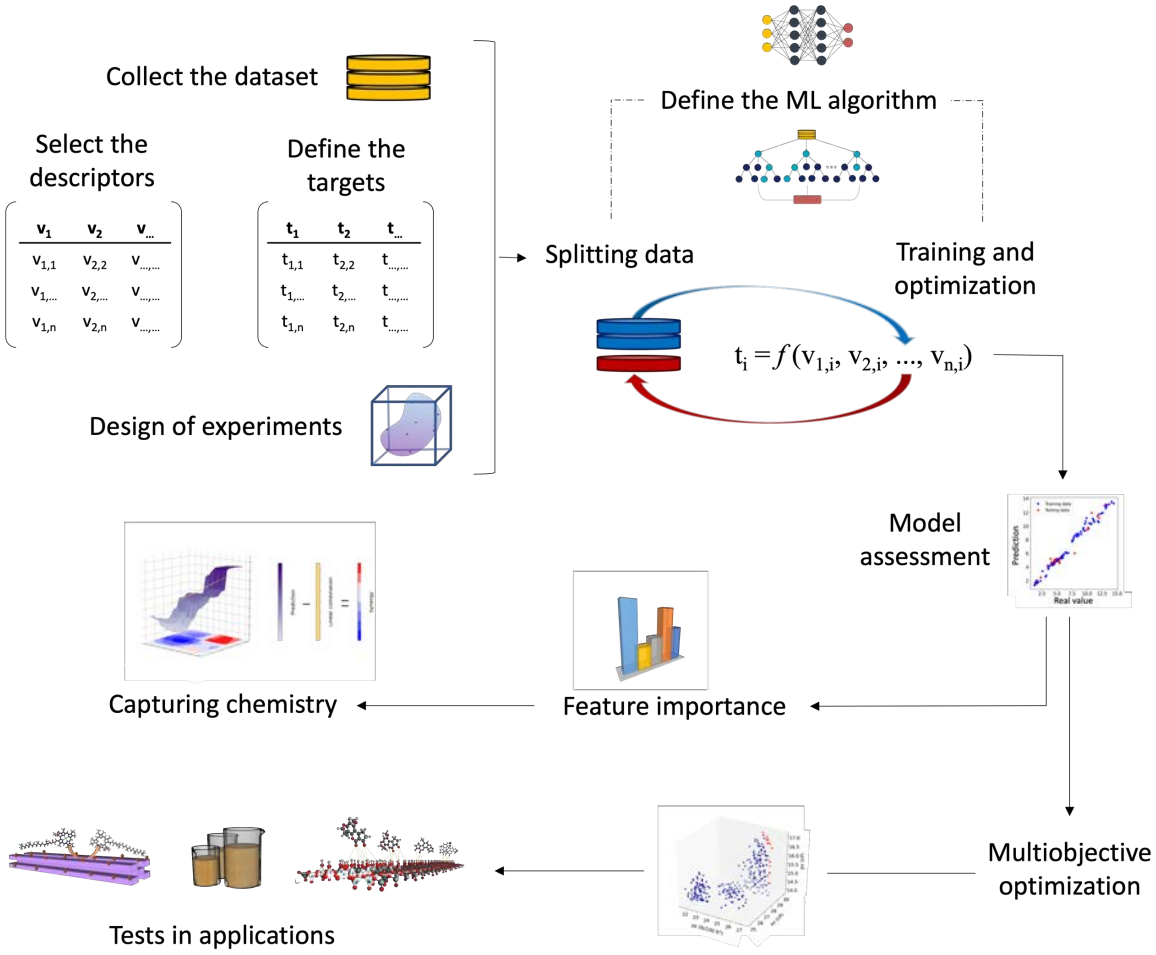
(112) Snarey, M.; Terrett, N. K.; Willett, P.; Wilton, D. J. Comparison of Algorithms for Dissimilarity-Based Compound Selection. *J. Mol. Graph. Model.* **1997**, *15* (6), 372–385. https://doi.org/10.1016/S1093-3263(98)00008-4.

(113) Li, Y.; Wu, H. A Clustering Method Based on K-Means Algorithm. *Phys. Procedia* **2012**, *25*, 1104–1109. https://doi.org/10.1016/j.phpro.2012.03.206.

Chapter 2 – General discussion.

2.1. Discussion and contributions

A graphical representation of the methodology adopted throughout this PhD thesis is presented in Scheme 2.



Scheme 2. Visual methodology created and adopted throughout the principal works of this PhD thesis.

The approach takes advantage of the information generated in our laboratory at Tolsa company during the last six years and throughout this PhD thesis. The basic sample composition and properties data are combined with a proper set of descriptors defining the starting datasets for machine learning to address each problem.

In cases where some area of the descriptor space was poorly explored (e.g., in exploring the formulation of rheological additives of Chapter 5), we populated these areas with points

(corresponding to compositions) with design of experiments matrix algorithms. The new data are incorporated into the historical data and used as a starting point of the research, as presented in the work of Chapter 5. Once the datasets were organized, the data entries are split into training and testing sets, and various predictive algorithms are trained to identify the optimal model architecture to be used as an oracle. The process involves iterative actions of hyperparameter optimization and validation with data out of the dataset. In most cases, the tree-based architecture overperform other algorithms, such as Multilayer Perceptron (MLP). It was preferred not only for accurate predictions but also for the ability to weigh the importance of the descriptors. The feature importance was the tool through which we captured the chemistry while avoiding long experimentation. The optimized models are exploited as surrogates of experiments accelerating the clay-based material characterization. We employ them to run virtual experiments for a large number of possible combinations of chemical and physical modification. The predictions were analyzed with complementary technics of multiobjective optimization, adapting the approach depending on the challenge to be addressed. The methods used are listed below and presented case by case.

In work described in Chapter 3, we proposed a design function (DF) to improve the physico-chemical of acid nanocatalyzers. Thus, we formulate a function that maximizes the surface area while minimizing the micropores, and the pH of the material, defined as follows:

$$DF = pSA \times pMicro^{-1} \times ppH^{-1}$$

The DF involved predicted properties of BET surface area (pSA), micropore content (pMicro), and pH (ppH).

In Chapter 4, the challenge was to find the optimal multicomponents mycotoxin detoxifier (MDT) for multitoxin mitigation. In the MDT formulation, each component interacts with the

others enhancing or damaging the effective adsorption of various mycotoxins. The multiobjective optimization was visualized by proposing the concept of "synergy between compounds" in the mycotoxin detoxifier material formulations. The model predictions were compared with the linear combination of targets of the pure components. The difference between predictions and the linear combination of pristine adsorption efficiency was assigned to the synergistic effect between components. This function allowed us to discriminate formulations in which the components cooperate enhancing the performance of multiple targets. The concept of synergy is amplified in the following section and extended to different problems, i.e., the design of clay-based rheological additives.

Chapter 5 presented a classical approach to multiobjective optimization using the Pareto front which identify the top promising formulations placed on the frontiers of multiple to-be-maximized targets.

Once the top promising prototypes were identified, I returned to the lab, preparing the materials and testing their performance in each application. With the experience of Tolsa, the industrial production of some of my high-performing materials was optimized, and various clients externally validated the scaled-up prototypes.

2.1.1. Employed dataset and descriptors

The initial task in the collaboration between IMDEA and Tolsa was to curate the historical data collected at Tolsa's laboratories build machine-learning- datasets. The targets were defined for each application to be addressed and were expanded by the corresponding selected descriptor space. The latter was searched and adapted for each dataset considering all the aspects of the problems. The Table 1 resumes the characteristics of each dataset.

Table 1. Resume of amount of data, descriptors, and targets used in the works of Chapters 3-5.

Work	Amount of data	Descriptors	Targets
Machine-learning-accelerated multimodal characterization and multiobjective design optimization of natural porous materials (Chapter 3)	49 (Morphology) 69 (surface activity)	Raw clays; Additive; Modification process	Surface area (pSA); External surface area (pESA); Micropore content (pMicro); Main pore size (pMS); Total pore volume (pVol); Free acidity (ppH)
Machine learning-aided design of composite mycotoxin detoxifier material for animal feed (Chapter 4)	51	Material; Mycotoxin; Experimental conditions	Adsorption (ads); Efficiency (eff)
Data-driven experimental design of rheological clay-polymer composites (Chapter 5)	80 (Historical dataset) 140 (updated dataset)	Composite material; Mixing rate	Apparent viscosity (av); Plastic viscosity (pv); Yield point (yp)

The features assigned to the materials of the dataset used in Chapter 3, incorporated three sets of descriptors, i.e., depicting the raw clays, the additives, and the modification process. The crucial point was choosing the ideal set of features to describe the raw clays because, since natural materials, computational modeling often reduces and distorts the material visualization. Thus, we opted for experimental physico-chemical characterization (e.g., XRD, CEC, pH, BET), collecting

20 features. The additives were the agents used to modify the raw clays (e.g., H_2SO_4, NaOH, Na_2CO_3) and were described by 14 features obtained by the PubChem repository. The modification process was defined by seven features that encoded the information of the material preparation done on a lab scale while reproducing the large industrial production. 41 descriptors were collected to define the processed materials and give enough information to the machine learning models when predicting the six properties listed in Table 1. To reduce the descriptor space dimension, we tried to remove the least important descriptors or implement the principal component analysis technics observing the appreciable deterioration of predictive power, implying that the complexity of the problems needs to be defined by the whole vector space selected.

The description of features space used to depict the mycotoxin detoxifiers (MTDs) in Chapter 4 was simplified. Two additional sets of features were included for comprehensive information on the material (MTDs), mycotoxins, and experimental conditions. In this case, we selected eight experimental features for the most representative materials. At the same time, the different chemistry of the mycotoxins was encoded by 15 features obtained by Toxic Exposome Database (T3DB). The targets listed in Table 1 were calculated by *in vitro* adsorption experiments which mimic the MTD's real action in the animals' gastrointestinal tract. Five features were created to describe the experimental conditions of the *in vitro* tests. Our feature selection allowed the building of predictive models with the ability to capture minor differences in the mycotoxin chemical structures. For example, the adsorption capacity of a selected MDT toward three mycotoxins belonging to the same trichothecene group (DON>DAS>T2). T2 joins two acetoxys and one methylbutanoate group in the DON structure, while DAS only shares with T2 the acetoxy groups. The selected features encoded the corresponding slight variations of mycotoxin molecular properties allowing accurate model predictions.

In Chapter 5 we represent the rheological gel structures with six features assigned to the composite formulation and one to the energy employed to disperse it in water (mixing rate). In this case, the physico-chemical features of the raw clays (AlCs) were encoded by a diversity criterium.

We employed Euclidean distances into the space represented by: phyllosilicate and organic impurity content obtained from XRD patterns, oxide composition obtained by chemical analysis, and cation exchange capacity. We then selected the distance score with respect to the AlC1, as shown in Figure 3.

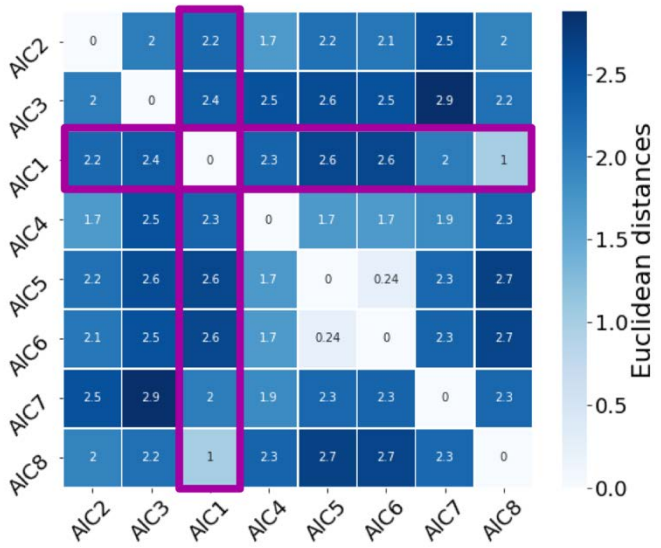


Figure 3. Distance matrix calculated on the base of Euclidean distances in the space of physico-chemical properties of the raw AlCs.

Even we reduced the descriptor space to just seven features. Those contain the full chemistry and gel preparation information to educate the machine learning models to investigate the descriptors-targets relationship. With our approach, we observed that when we apply diversity critiria to the descriptor space we obtain no total correlation to diverse targets. The latter is associated with a non-linear dependence of the descriptor space and the targets challenging to be captured only by human experience.

The diverse expression of the problems addressed throughout this PhD thesis was crucial for the success of the machine learning approach in the context of clay-based materials. It is

expected to be helpful in the case of research on similar natural porous materials. The full details were published and incorporated in Chapters 3-5.

2.1.2. Machine learning models

During this Industrial PhD thesis, the main goal of facilitating the material design of clay-based products was centered on three main applications, i.e., catalysis, adsorption, and rheology, established on the line to the main market of Tolsa products. As anticipated before, once the data were collected and the descriptors were selected, we benchmarked diverse model architectures to identify the best predictive algorithm to be used. The main results are resumed below while detailed in Chapters 3-5.

The design of acid nanocatalyzers was approached by starting from the study of the effect of physical or chemical modification of the morphology and surface activity (pH). The selected targets are listed in Table 1, and six independent machine learning models were built to predict each of the targets listed in Table 1. The models are functions of:

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

where i indicates the target of Table 1 while x_{raw_clay}, $x_{additive}$, and $x_{modification\_process}$ are the aggregated feature vectors presented in the previous section. We assessed the simplest Decision Tree, the ensemble algorithm of the Random Forest regressor, the more randomized Extra Trees regressor, and a more complex deep learning model Multiple Layer Perceptron (MLP). All models were implemented using the Scikit-learn library except for the MLP model implemented in TensorFlow and Keras frameworks. The Extra Tree regressor was the most accurate model architecture, followed by MLP, Random Forest regressor, and Decision Tree. After splitting the dataset, we used 85% to train and optimize the Extra Tree regressor algorithm to predict each of the six

properties of Table 1. The hyperparameters were optimized with 10-fold cross-validation., The predictability was verified by predicting the properties of the 15% of the data points, kept aside as a testing set. The accuracy scores were the R^2 and mean absolute error (MAE) are reasonable considering the small data size (Table 2).

Table 2. Test set-based model assessment results for the six optimized Extra Trees regressors.

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	6	0.96	0.33

The lowest score was seen for main pore size (MS), reflecting the contribution of error propagation. The MS is calculated as 4Vol/ SA, where Vol and SA are the targets of pore volume and surface area of the materials in the dataset. 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 ˚A is consistent with the experimental error observed at high MS values.

To predict the mycotoxin sequestration of various clay-based detoxifiers, we used a dataset of *in vitro* tests containing two outcomes: adsorption (%) and efficiency (%), as defined in Chapter 4.

Considering the good predictability of tree-based architectures that we observed in the context of clay materials, Random Forest regressor (RF) and Extra Trees regressor (EXT) were hyperparameters optimized on the base of k-fold cross-validation with K = 5. Diverse predictive models were benchmarked, i.e., support vector machine, multiple layer perceptron, and K-nearest neighbors. The algorithms were assessed employing R^2, MEA, accuracy, and p-value, which are calculated with the 5×2cv paired t-test. In particular, the Support vector machine and k-nearest

neighbors were implemented using the Scikit-learn library. The Support vector machine was optimized based on k-fold cross-validation, finding the best hyperparameter, i.e., kernel='poly', gamma='scale', degree=3. Multiple layer perceptron (MLP) was implemented in TensorFlow and Keras framework. The network was built with six dense layers with ReLU activation function, 100 neurons in the hidden layers, Adaptive Moment Estimation (Adam) as an optimizer, and 300 epochs. The performances were evaluated by R^2 score, mean absolute value (MAE), p-value and accuracy score. The latter is added to classify how often our regressors make predictions close to the real value by considering the estimated experimental error as tolerance grade. The accuracy is defined as the ratio of the number of correct predictions and the total predictions corresponding to the test set size. The correct prediction was defined considering the estimated experimental error (5.7) and obtained when the difference between the predicted and real target's absolute value is less than 5.7. The computed p-values allowed us to compare RF to the other models to observe a significant difference ($P < 0.05$) in the performances of multiple layer perceptron and K-nearest neighbors and a considerable statistical difference between RF and support vector machine. Out of the results, Random Forest was selected for its outperforming, low variability within the fold switching and clear model interpretability, exhibiting a R^2 of 0.84 ± 0.08 in cross-validation of trainset (K = 5) compared to 0.76 ± 0.14 of EXT. Figure 4a shows the visual assessment of RF when predicting the adsorption (RFads) and the efficiency (RFeff) when the models are trained on 85% of the available dataset and tested in the remaining 15%. The models were further validated by predicting and testing the adsorption and efficiency of a mycotoxin detoxifier (MDT), toward an out-of-trainset trichothecene toxin, i.e., diacetoxyscirpenol (DAS). Figure 4b depicts the ability of RFads and RFeff to run virtually *in vitro* tests enabling fast material screening and multiobjective optimization.

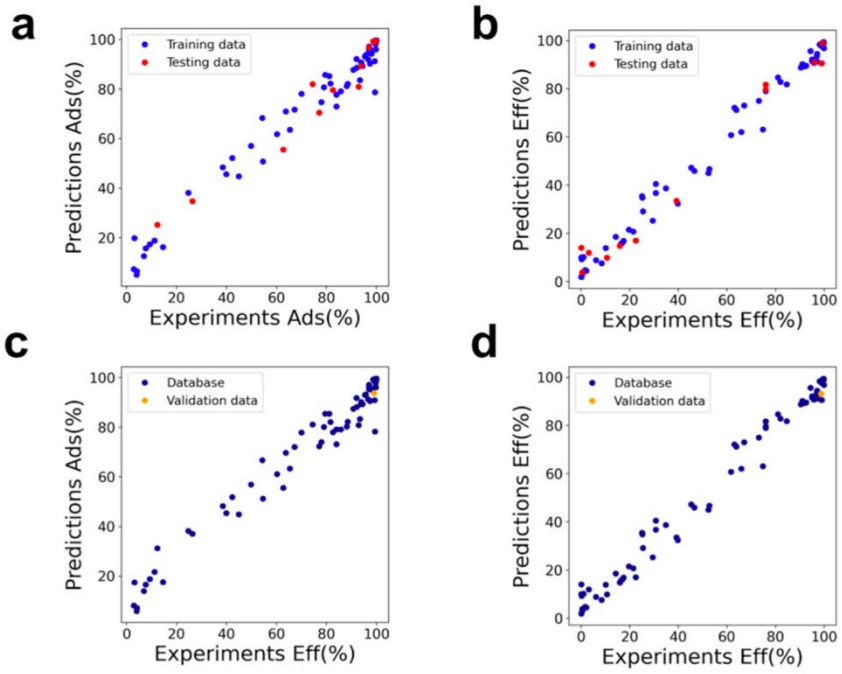


Figure 4. Visual assessment of RFads (a) and RFeff (b) of 6 mycotoxins by 15 MDTs under different experimental conditions. Validation of RFads (c) and RFeff (d) toward one yet-unregulated toxin (DAS).

Random Forest regressor was employed as consistent model architecture in the case of the data-driven experimental design of rheological clay-polymer composites, as included in Chapter 5. The algorithm was trained to predict the three apparent viscosity (RFav2), plastic viscosity (RFpv2), and yield point (RFyp2) on the dataset containing both the historically generated information and the data designed by a design of experiment matrix. The algorithms were hyperparameter optimized and tested on 15% of the data. The assessment results are depicted in Figure 5 and Table 3. The design of the experiments algorithm effectively generated homogeneous data, allowing us to build models with good predictability ($R^2 > 0.91$) to be used as an oracle for material screening purposes. Figure 5 also shows the feature importance associated with each model.

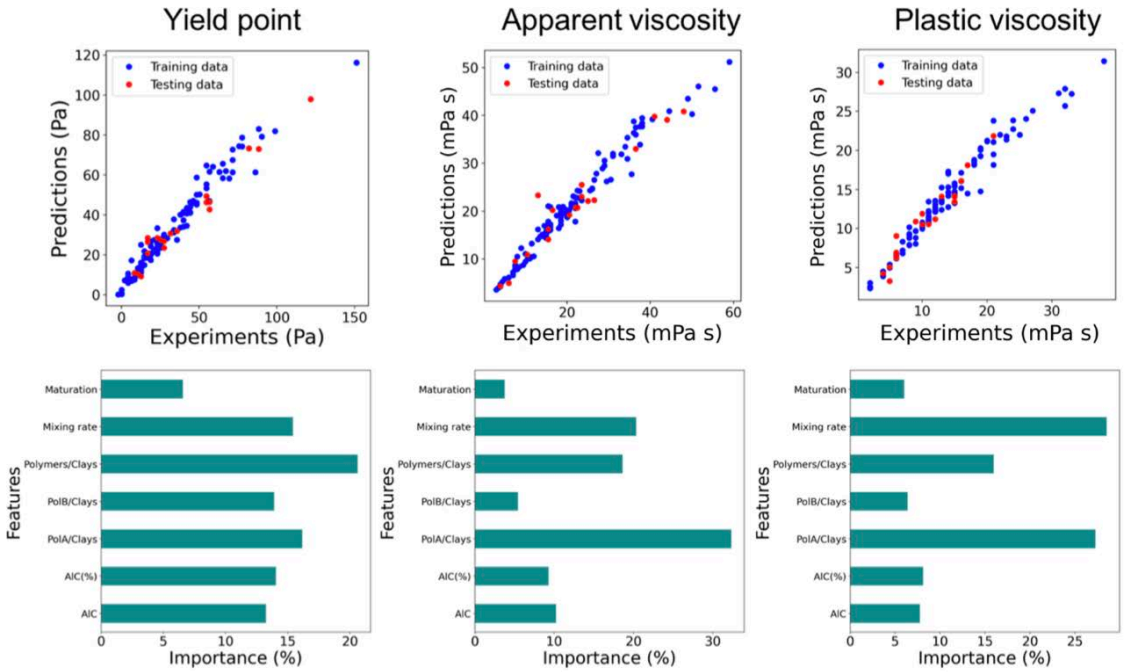


Figure 5. Graphical assessment (top) of RFyp2, RFav2, and RFpv2 and the corresponding feature importance analysis (bottom).

The distribution of the importance reflects a non-equal correlation between descriptors and targets. Thus, the formulations that maximize one target may cause the others to be scarce, requiring advanced tools for multiobjective optimization. The latter is presented in the following section and deeply studied in Chapter 5.

Table 3. Model Assessment Results for the Test Set of DS2

score	RFyp2	RFav2	RFpv2
R^2	0.91	0.91	0.93
MAE	7	2.7	0.99
MSE	84	13.4	1.5

2.1.3. Multiobjective optimization and application tests

Materials development for a specific application often requires balancing competing properties contributing to the overall performance. The multiobjective optimization was the most outgoing objective of this PhD thesis and was approached differently depending on the problem to be faced.

In Chapter 3, our goal was 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. In this case, we incorporate the multiobjective optimization into the proposed design function (DF), which involves predicted properties of BET surface area (pSA), micropore content (pMicro), and pH (ppH). Sulfuric acid was selected as an 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 single optimum value for the five material classes. In particular, the models were run on a set of proposed materials considering the five 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 number of particles more minor than 45 mm was fixed to 75%. The DF improvements for the starting values were calculated as (DFproc - DFraw)/DFraw x 100, where DFproc and DFraw are the design functions of the processed and raw nanomaterials, respectively. Five natural materials were explored, corresponding to high-purity palygorskite, sepiolite, montmorillonite, saponite, and stevensite. Figure 6 highlights the optimal improvement areas for the DF of the raw material. An improvement of 125% of DF at low moisture (RH) can be noted for saponite. Although 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.

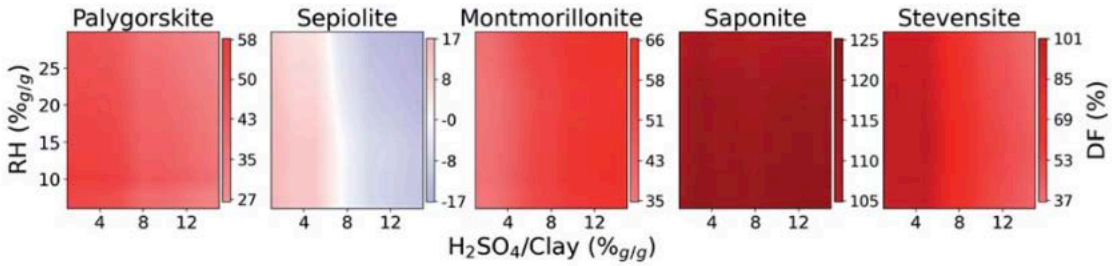


Figure 6. Trend of design function improvements (%) with different additives/clays (%g/g) and starting moisture contents (RH%g/g).

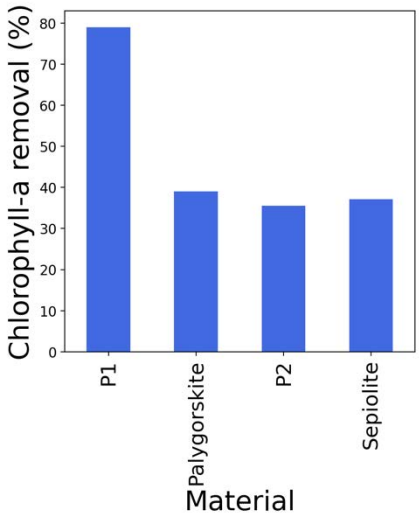


Figure 7. Removal of the chlorophyll-a after acid-catalyzed thermal degradation with the promising material P1, assessed against the sepiolite-based materials.

The promising acid nano-catalyzer extracted using DF was prepared and tested in the catalytic degradation of chlorophyll-a. A second prototype, P2, was also synthesized based on sepiolite to

verify the righteousness of the models, which predicted no improvement for the raw sepiolite. The results, summarized in Figure 7, show that P1 could remove 78.6% of chlorophyll- a, 40% of improvement against the pristine material, and P2.

In Chapter 4 we have proposed an alternative visualization of the multiobjective optimization introducing the concept of synergy between compounds. The MDTs contained in the dataset are formulations of single, two, or three compounds. Each compound's affinity toward different mycotoxins depends on the mycotoxin chemical structures and the physico-chemical features of MDT. The multiobjective optimization concept, in this case, can be rephrased as optimization of the MDT formulation for multi-toxin detoxification, hence search the optimal combination of the MDT compounds which enhance the simultaneous adsorption of a broad set of mycotoxins. Thus, we generate a hypothetical combination of sepiolite, montmorillonite, and activated carbon and run virtual experiments to detoxify 22 different mycotoxins. The optimal MDT combines sepiolite-montmorillonite in a ratio of 1:4, and 2% of AC (SEP/MONT/AC), and its predicted performance is shown in Figure 8. SEP/MONT/AC is expected to achieve an efficiency of at least 94% in all the explored toxin groups, except for the trichothecenes, which are the most heterogeneous group of molecules. Moreover, deoxynivalenol (DON) was found to be the most challenging-to-be-removed toxin by SEP/MONT/AC. The latter information was also used to design an *in vivo* trial to validate the finding of our approach externally. In collaboration with the Department of Pathobiology, Pharmacology, and Zoological Medicine of Ghent University, we observed the DON relative oral bioavailability when the detoxifier is administrated to broiler chickens. The results confirmed a relevant reduction in *vivo* systemic exposure to DON. The identified MDT is currently commercialized under the market name of Atox Nature Silver (ANS), demonstrating great success in the market of additives for animal feed.

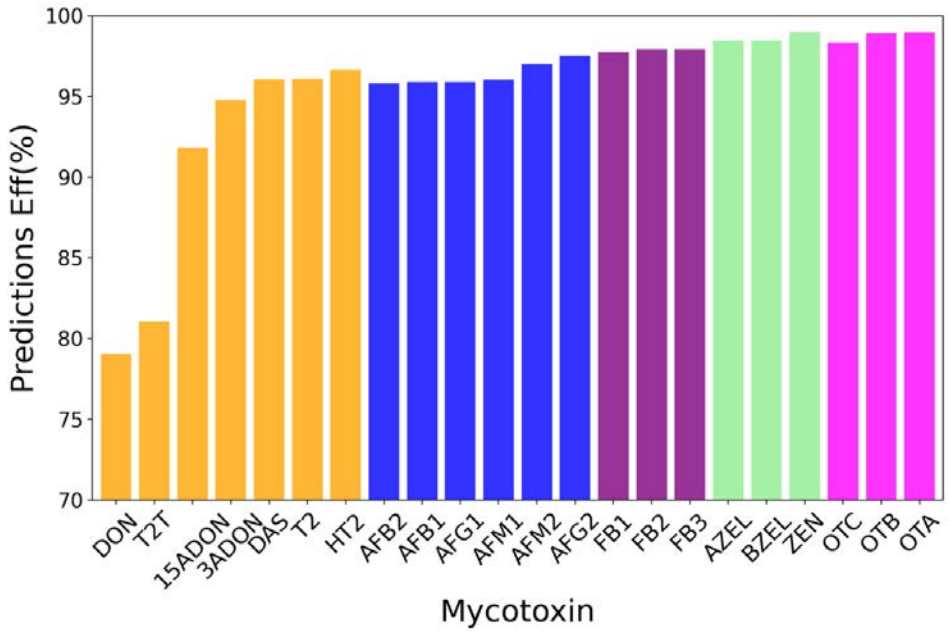


Figure 8. Efficiency prediction of a list of regulated and yet unregulated mycotoxins. In orange, the group of trichothecenes, aflatoxins in blue, fumonisins in purple, zearalenone in green and ochratoxins in magenta. The experimental setting was fixed to 2kg/t of inclusion rate of MDT, and 2 μg/ml of toxin concentration. The pH during the adsorption experiment was fixed to 3, while the desorption pH was 6.5.

During the development of the work incorporated in Chapter 5 on the design of rheological additives, we observed that often slight modification of the composite formulations provokes a dramatic decrease in the viscosities and yield point of the fluids. This observation reflects the non-equal correlation between material features and the final properties. Hence, if we hypothesize that different material formulations show different rheological behavior which follows the role of mixtures, we may get discrepancies and incorrect estimation of the real behavior. Our trained RF models possess the needed information to capture the challenging to-be-supposed correlation between descriptors and targets, enabling material screening based on their predictions. To optimize the creation of a well-organized gel structure, we proposed a set of 7290 material formulations coupling clustering technics to the Pareto algorithm. Screening thresholds for good-

performing clay-based rheological additives were fixed to yp >42 Pa (20 lb/100 ft^2), av >25 mPa s, and pv >15 mPa s. The threshold values correspond to stable gel formation when the rheological agents are dispersed in water. 6730 proposals were excluded by the yp imposed limit, 248 by the apparent viscosity, and 192 by plastic viscosity. The Pareto algorithm then analyzed the set of 120 selected promising prototypes. The latter highlights those descriptor combinations which maximize the three objectives, i.e., yield point (yp), apparent viscosity (av), and plastic viscosity (pv). Figure 9 depicts the predictions of the 120 promising formulations with the Pareto points marked in red. Three prototypes, named P1−P3, were prepared, and the rheological behavior of their water dispersion was characterized. The correspondences between the predictions and the experiments remarked the robust implementation of the Random Forest algorithm in designing novel functional clay-based materials. Moreover, RFav2, RFpv2, and RFyp2 are nowadays on-working models exploited in the industrial development process when Tolsa needs to prototype unique materials for the precise requirement of the clients.

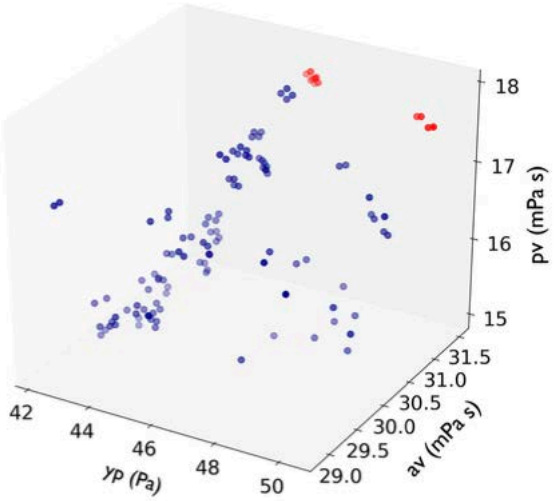


Figure 9. Predicted outcomes of the three targets (yp, av, and pv), marking the Pareto front in red and the nonoptimal formulations in blue.

2.1.4. Capturing synergy between components

During the development of the work of "Machine learning-aided design of composite mycotoxin detoxifier material for animal feed", incorporated in Chapter 4, we have started to explore the possibility of predicting the synergic effect between components and rapidly identifying the optimal formulations. The idea was born by observing the distribution of the importance scores assigned to the features. We also observed a no-linear dependence between the main features and the properties, complicating the understanding of their correlation. However, machine learning models possess the ability to capture this information and give us precise properties estimation. Conversely, the algorithms are complex black boxes challenging to be interpreted.[1–3] We proposed to use the models' predictions as surrogates of real experiments mapping the space of the MDT formulations as a way to extract more information from the black box and visualize the most exciting area of the descriptor space.

Similarly, during the development of the work of Chapter 5, we observed a complex correlation between the features of the rheological additives and their viscosity outcomes. The latter was analyzed by clustering analysis which highlighted no complete correspondences between the groups identified in the descriptor and the ones of the property space. An example is shown in Figure 10. The diversity function based on Euclidean distance was applied to the descriptor space. We then graphed the corresponding properties, assigning the groups' colors identified by cutting the dendrogram at the level of 17 of the distance in the descriptor space. The observed overlapping of the groups results from the complex correlation between features and properties, far from the rule of mixtures.

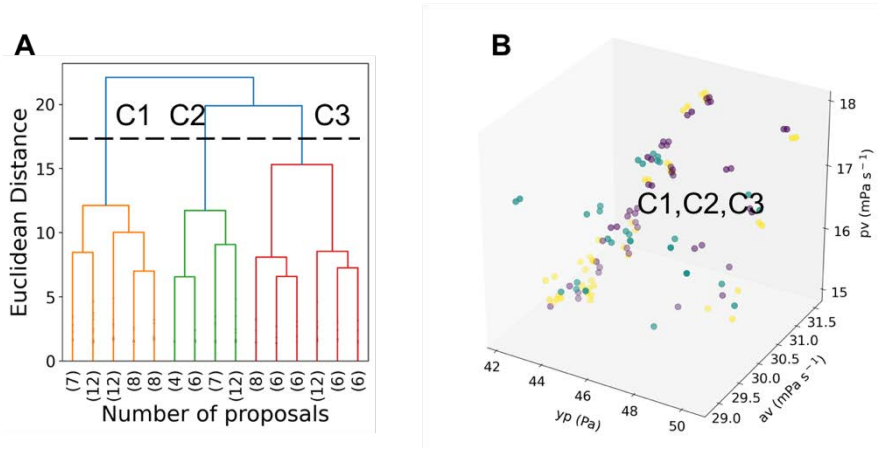


Figure 10. Clustering analysis of the 120 promising formulations. The dendrogram obtained by applying the Euclidean distance into the descriptor space (A) was cut at the level of 17 (dashed line), identifying the three groups (C1-3) represented in the target space (B).

Again, it is challenging to interpret by humans' intuition though accurately captured by machine learning. In this section, we present an extension of the synergetic effect visualization applied to the dataset and models of Chapters 4 and 5. The detailed discussion is available in Annex 1.

Synergistic effect definition. The synergy between components in the hybrid formulations was calculated considering each material descriptor value as a linear combination of the descriptors of the singular compound ($\sum_i a_i x_i$, where a_i is the amount (%) of the i-component in the hybrid formulations and x_i is the descriptor value corresponding to the pure i-component). The machine learning models predicted the targets of the hypothetical composites (y_{pred}). The predictions were compared with the linear combination of targets of the pure components, being calculated as $\sum_i a_i y_i$, where y_i is the predicted target of the pure i-component. The difference between predictions and the linear combination of pristine efficiency was assigned to the synergistic effect between components (Eq. 4).

$$\text{Synergy} = y_{pred} - \sum_i a_i y_i \qquad \text{Eq. 4}$$

Synergic effect in adsorption properties. By exploiting the models and dataset of Chapter 5 we explore the synergic effect in all the possible combinations of ternary formulations composed of sepiolite + montmorillonite + activated carbon (AC). Each descriptor was calculated as a linear combination of the descriptors of the singular compound, and we used RFeff to predict the adsorption efficiency of a target mycotoxin, zearalenone (ZEN). Figure 11 shows the prediction surface in the 3D composition space of the hybrids. The resulting synergy contribution, calculated following Eq. 4, was projected on the x,y plane. The predictions displayed a noticeable displacement from the plane.

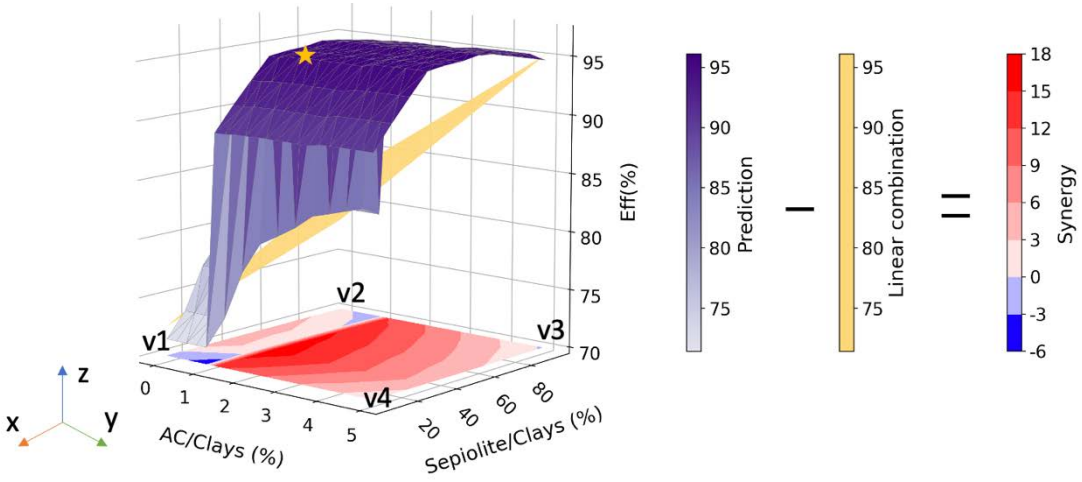


Figure 11. Prediction of efficiency in adsorption of zearalenone (ZEN) mycotoxin (purple surface) in the 3D space of ternary hybrid composed of sepiolite + montmorillonite + activated carbon (CA). The resulting synergic contribution was projected in the x,y plane. The start represents the experimental data point. The experimental settings were fixed to 2 kg/t of inclusion rate, 2 μg/ml of toxin concentration pH 3 during the adsorption step, and pH 6.5 during the desorption step.

Although our models were trained by a dataset that contained just information of pure MDTs and the rule of mixture calculated the descriptor space of the explored hypothetical MDTs, our approach was able to capture the hidden information about specific interactions between the

compounds occurring when they are put working together. Thus, the predictions are out of the plane, creating spots in which the performance is higher (painted in red) or lower (painted in blue) than their estimations under the assumption that the components of the hybrid adsorb the mycotoxin as they do when they act alone. However, the cooperation between clays can be visualized following the x,z plane. The detoxification increases as the content of sepiolite increases. The observed maximum positive cooperation for those formulations incorporated the sepiolite in ranges of 20-60%. The methodology allowed us to map all possible combinations and identify the most interesting area to explore, looking for the top promising MDT for multi-toxin sequestration.

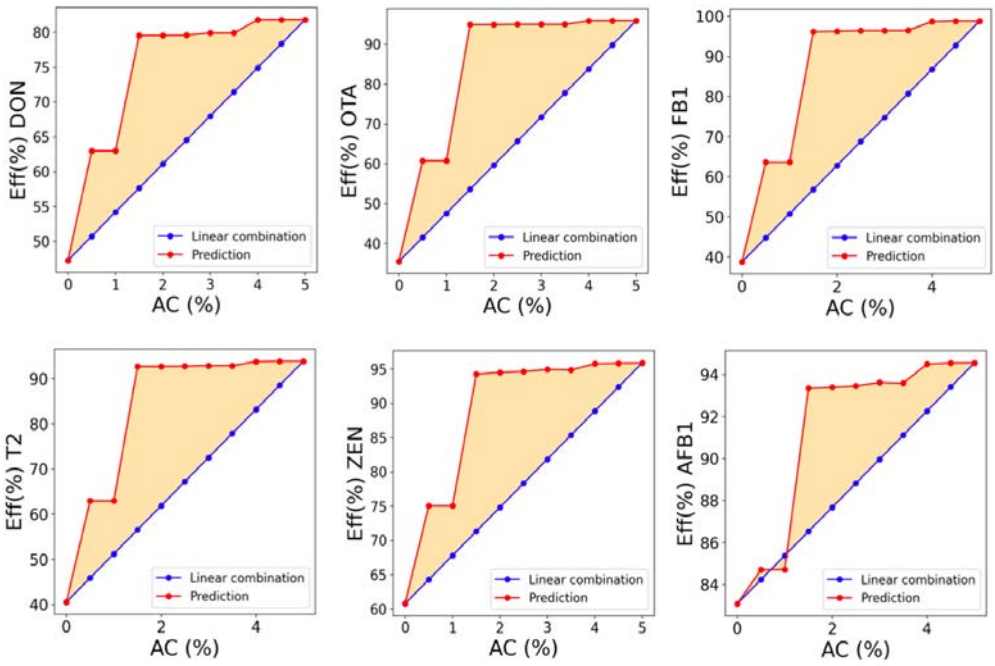


Figure 12. Synergy capturing by RF predicting efficiency for a series of hybrids of sepiolite-montmorillonite- charcoal in which the sepiolite-montmorillonite ratio was fixed to 1/4. The orange area is assigned to the positive synergistic effect. The experimental settings for the uptake of DON, ochratoxin (OTA), T2-toxin, fumonisin B1 (FB1), aflatoxin B1 (AFB1), and zearalenone (ZEN) were fixed to 2 kg/t of inclusion rate of MDT, 2 μg/ml of toxin concentration. The pH during the adsorption experiment was fixed to 3, while the desorption pH was 6.5.

In Chapter 4, we cut the 3D surface fixing the composite sepiolite-montmorillonite in a ratio 1:4 while varying the AC from 0% to 5%. The linearly-combinate descriptors associated with the resulting combinations were used as input to run virtual experiments and identify the MDT which exploits the maximum positive synergy effect (see Figure 12).

The approach was also applied to investigate the behavior of different clays' mixtures and assess our tactic's power in predicting the synergy in various cases. The latter were validated by selecting strategical formulations and running experimental tests. The results are explained in detail in Annex 1.

Synergic effect in rheological properties. We transferred the previously presented approach to investigate the synergy between the components of rheological hybrids (Chapter 5). The validated RFav2 was used as an oracle model to predict the apparent viscosities of the water slurries. The hybrids are mixtures of smectite (MgC), montmorillonite (AlC), acrylic polymer (PolA), and cellulosic polymer (PolB) in various ratios. Figure 13 shows the predictions (purple surface), the calculated linear combination of the vertices, and the resulting synergic contribution. We explored and changed the amount of MgC, AlC, and PolA, while fixing PolB to 1% (Figure 13A) and MgC, AlC, and PolB, while fixing PolA to 1% (Figure 13B). Similarly, to the previous observations, the surface of predictions is displaced from the plane, creating spots of higher viscosity (painted in red) or lower (painted in blue) concerning the viscosity estimated under the no-synergy assumption. The latter highlights the areas of the formulations which create a stable gel architecture and others in which the components flocculate or are not dispersed in a favorable environment. The positive sign of the synergy can be observed starting from the incorporation of 0.25% of acrylic polymer with a maximum placed in 0.75-1%.

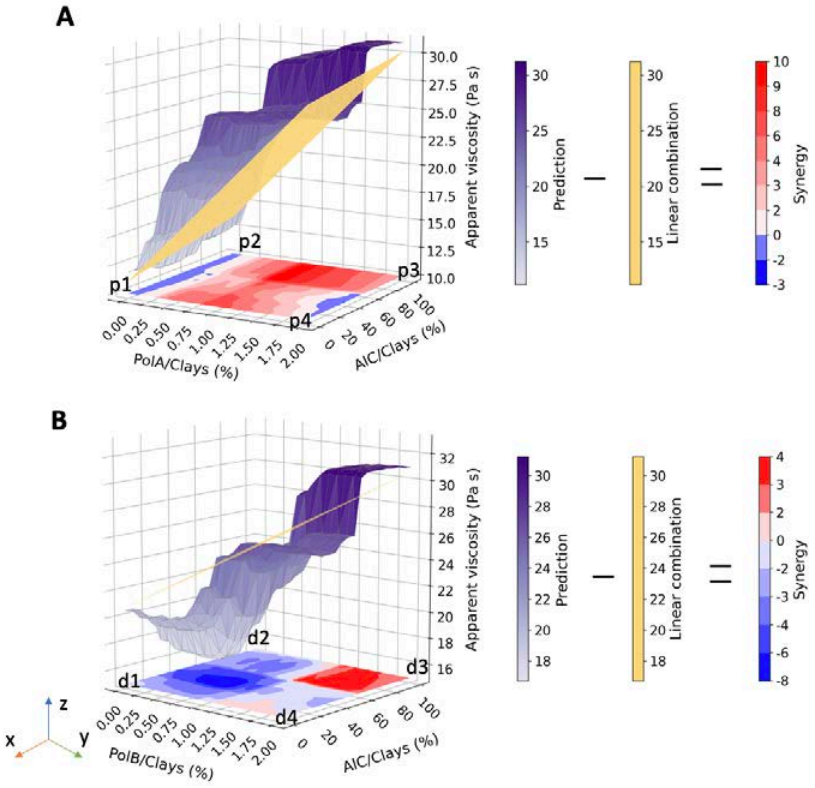


Figure 13. Prediction of the apparent viscosity of rheological hybrid composed by a combination of smectite (MgC) + Na-montmorillonite (AlC) + acrylic polymer (PolA), fixing the cellulosic polymer (PolB) to 1% (A). Prediction of the apparent viscosity of rheological hybrid composed by the combination of smectite (MgC) + montmorillonite (AlC) + cellulosic polymer (PolB), fixing the acrylic polymer (PolA) to 1% (B). The resulting synergic contribution was projected in the x,y plane.

At least 1% of the cellulosic polymer should be added to observe cooperation and, consequently, the gel architecture. Both polymers optimize the electrostatic interparticle interactions with a high amount of Na-montmorillonite and can be appreciated following the x,z planes. On the contrary, the interaction between components in mixtures of smectite and Ca-montmorillonite is unfavorable (negative sign) when incorporated in high amounts. The latter is an effect of the diverse clays' intrinsic properties encoded in our AlC descriptor.

Also, in this case, we extend the approach to the diverse situation to test the capability of our models to capture the synergistic effect. The details are available in Annex 1.

2.1.5. International stay contribution

The work advanced during the starting period of this PhD thesis and presented in Chapter 3 pointed out the high importance of the moisture (RH) of the clay (both after and before a physico-chemical modification) in affecting the final properties of morphology and surface activity. The latter were correlated to catalytic application and are key features in adsorption application. The unexpectedly high importance of RH stimulated the curiosity to amplify the investigation. Thus, during my international stay at the Department of Physics and Chemistry, University of Palermo, I aimed to understand the role of moisture in natural clays and tried to correlate this information to the adsorptive application. Moreover, since a drying process is a common step during the industrial production of various clay-based materials, the effect of the thermal treatment at 150° C was also investigated. This temperature provokes the release of weakly adsorbed water molecules while maintaining the crystal structures. The investigation was conducted by modulated thermogravimetric analysis TGA measurement (MTGA), which provides fast calculation of the activation energy (Ea) associated with the removal of the confined water. In particular, we studied the behavior of the adsorbed water in five diverse clay structures recording the MTGA profiles up to 260ºC. Thermal treatment at 150ºC provokes the release of weakly adsorbed moisture. It is expected to modify the moisture capture capacity when the dried samples are reconditioned under controlled humidity with 75%wt saturated NaCl salt solution. Parallelly, as a proof of concept, the Ea and other physico-chemical properties were statistically correlated to the adsorption activity of natural clays in two diverse applications, i.e., mycotoxin aflatoxin B1 (AFB1) and β-carotene mitigation. We designed the adsorption experiments according to typical application requirements. Specifically, the adsorption test of AFB1 is designed by simulating the animal's gastrointestinal

tract, while the β-carotene is designed by employing a vacuum pump and high temperature. Details of the employed methods and experiments are presented in Chapter 6.

The mechanism of clays' dehydration depends on the morphology and structural characteristics, e.g., the type of exchangeable cations and swelling properties.[4–7] Figure 14 contains the MTGA profiles for the five classes of 2:1 phyllosilicates.

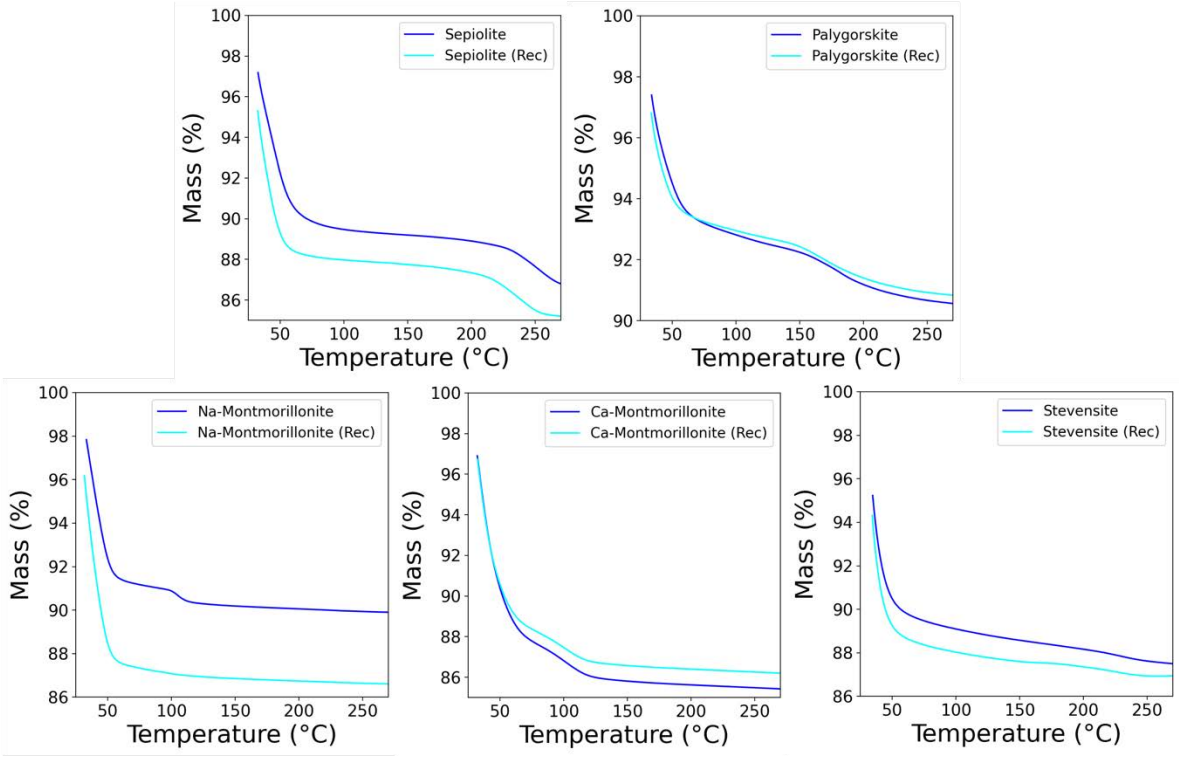


Figure 14. MTGA of the conditioned sample under constant humidity (blue lines) and reconditioned samples after drying overnight and 150ºC (cyan lines).

The interlayer space contains weakly adsorbed water and water molecules strongly adsorbed to interlayer cations. The first loss at ca. 50°C is associated with the partial removal of the moisture content and weakly/physisorbed water from the micro and mesoporous surfaces of clays. The second loss, occurring at a higher temperature, is associated with removing water in the microporosity of tunnels in fibrous clay and interlayer space of stevensite and montmorillonites,

firmly bound to exchangeable cations. Table 4 summarizes the main parameters extracted by the MTGA profiles.

Table 4. Weight losses and activation energy calculated by MTGA for the conditioned sample under constant humidity and reconditioned (Rec) samples after drying overnight and 150ºC.

Clay mineral	**Temperature (°C)**	**First loss (%)**	**Ea (kJ/mol)**	**ΔE (kJ/mol)**
Na-Montmorillonite	45	8.9	100	-13
Na-Montmorillonite (Rec)		12.7	87	
Ca-Montmorillonite	39	12.3	82	8
Ca-Montmorillonite (Rec)		12	90	
Palygorskite	39	7.2	164	72
Palygorskite (Rec)		7	236	
Stevensite	38	11.7	85	-3
Stevensite (Rec)		12.3	82	
Sepiolite	47	10.9	187	-53
Sepiolite (Rec)		12.2	133	

Although no difference in temperature of the first loss was observed after the reconditioning procedure, the drying process modified the moisture capture capacity. The second loss remains relatively stable for non-swelling clay minerals being not affected by thermal treatment at 150ºC, meaning that the water uptake in the microporosity is a reversible reaction that does not affect structural and morphological properties. On the contrary, the second loss disappears from the MTGA profile of swelling Na-montmorillonite when treated at 150ºC and reconditioned. The

latter is probably correlated to the closing of its interlayer space when the micro-confined water is removed, making the process irreversible.[5,7] The Ea of conditioned fibrous clays was higher than lamellar minerals suggesting that the morphologically different clay particles influence the interaction of water and clay surface in terms of the energy required to activate the process. Indeed, removing weakly bound water from the mesoporosity of fibrous clays is a more energetically costly process than the same process occurring in lamellar clay minerals. The dehydration mechanisms are controlled by nucleation and growth, followed by a diffusion-controlled reaction.[8] The differences between the two morphologies could be most likely related to the tortuous path created by the arrangement of sepiolite fibers or palygorskite hindering the water molecules' removal.[9–13] After the thermal treatment, the moisture uptake behavior changed significantly. For example, it is enhanced for Na-montmorillonite, stevensite, and sepiolite, as observed by the increase of first loss percentages in Table 4. A complete discussion of the results observed by recording the MTGA is available in Chapter 6.

We correlated the Ea and a complete set of physico-chemical features with the adsorption of two target molecules: β-carotene and AFB1 mycotoxin. The property-activity correlation was estimated by calculating Spearman's coefficient (ϱ), shown in Figure 15 (see Method section of Chapter 6). The low ϱ of 0.3 with an associated p of 0.62 is shown between AFB1 mycotoxin and β-carotene. The lack of noticeable correlation (low values of ϱ) reflects the different mechanisms involved, including the different experimental setups. The adsorption experiment of β-carotene is performed at high temperatures and under vacuum conditions while AFB1 is at room temperature and pressure.

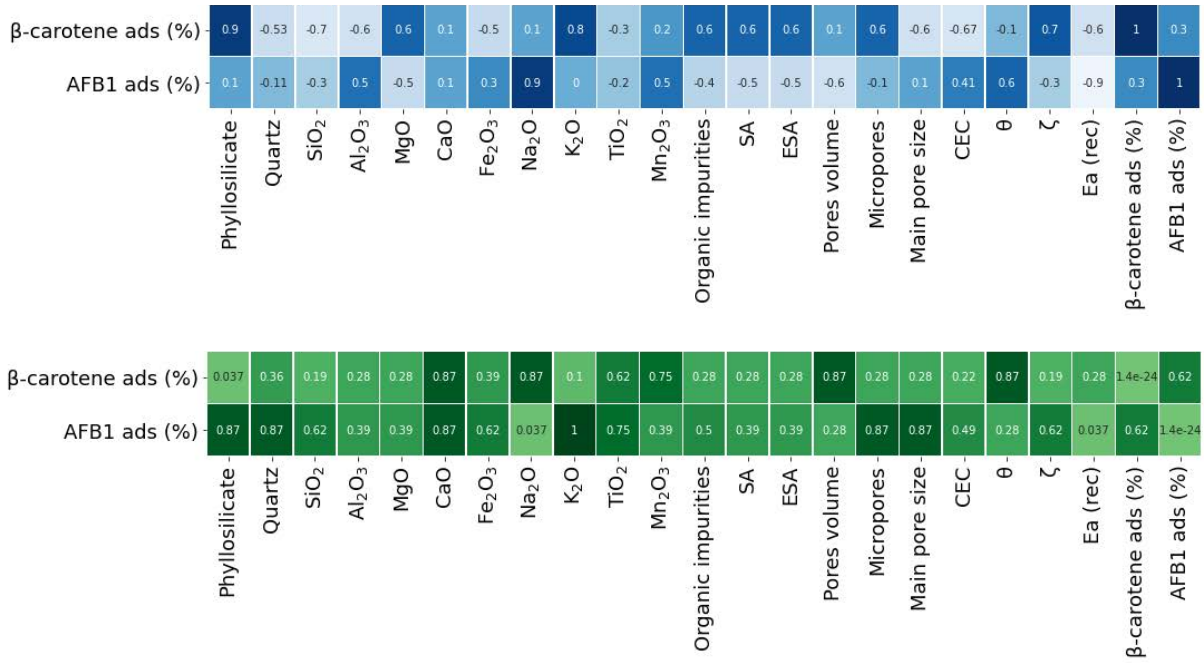


Figure 15. Spearman correlation coefficients (blue) and their associated p-values (green) between the adsorptive outcomes and the physico-chemical properties.

As expected, the adsorption of AFB1 suffers from the presence of the confined water because the temperature and pressure do not remove it. The most correlated feature with AFB1 is the activation energy required for the desorption of the confined water in the reconditioned samples, with a Spearman's score of -0.9 (p=0.037). The negative value means that the AFB1 adsorption increases when the Ea(rec) decreases, remarking the intense competition between AFB1 and water, independent of the proper clay structure. The latter also suggests that the thermal treatment deteriorates the capability of palygorskite in capturing AFB1 since it increases its Ea(rec) of 72 kJ/mol, while being practical for sepiolite structures providing a decrease of 53 kJ/mol, hence favoring the AFB1 adsorption for the conditioned no-treated sepiolite.

Additionally, we conducted the AFB1 adsorption experiments on the untreated samples of the trioctahedral group. We found an AFB1 adsorption of 93% for the raw stevensite and 50% for sepiolite. The results remark that clays in which Ea remains almost unmodified by the thermal treatment show the same adsorption capacity. In contrast, clays like sepiolite benefit from the

thermal treatment, which facilitates the interaction between AFB1 and the clay surface by reducing the Ea.

On the other hand, during the β-carotene mitigation experiments, the confined water is forcefully removed, favoring the entrapping of the pigment on the clay surfaces. In this case, we observed a decrease in the correlation score (ρ = -0.6, p=0.28), suggesting a faint competition between the guest molecule and water, being removed by the process. The correlation between clay structure (amount of phyllosilicate) and β-carotene shows the highest Spearman score of 0.9 (p=0.037). This observation can be interpreted considering the vacuum pumping and the high temperature promoting a good interaction between β-carotene and the clay by exposing the clay surfaces to the water. Hence the adsorption, in this case, is mainly guided by the content of active minerals.

2.2. Thesis's impact

Facilitating the design of clay-based materials for specific applications is the core of this Industrial PhD thesis. The conversion from raw minerals to clay products involves tuning clay characteristics (e.g., surface area, porosity, and physical-chemical surface properties) by modifying and controlling the preparation process. The conventional approach is time-consuming and costly due to the extensive experimental work required. The constant development of improved products is crucial for the growth of companies such as Tolsa to enter new markets and maintain "the competitive edge".

Under the main goal established in collaboration between IMDEA Materials and Tolsa company, this PhD thesis explores the application of data-driven tools to accelerate the experimentations of clay-based materials. The collaboration with Tolsa company defined the direction of the outcomes of the investigation progressed during this thesis. Three crucial application fields were selected in which Tolsa wanted to improve its market position and test the

effectiveness of machine learning design of saleable competitive products. The dedication of the research was centered into:

1) Design of acid nano-catalyzers
2) Design mycotoxin detoxifiers and understand their mode of action
3) Design of rheological additives

Improved materials were rapidly pinpointed through machine learning, and the chemistry was captured by waiting for the importance of the involved material features.

The investigation progressed during this PhD thesis is fully presented in Chapters 3-5. The materials preparation and characterization of their features and properties were carried out on a lab scale while reproducing later the real industrial processes. The success in the correspondence between predictions and experiments moved the implementation of the approach to the large-scale development of promising novel products for Tolsa. The machine learning models, and datasets of Chapters 3-5 were readapted to control the industrial technologies and propose prototypes to be scaled-up. Nowadays, the benefits of this industrial PhD thesis cloud also be quantified in sales as proof of the fruitful cooperation between research and industry.

The application's lines in which this PhD thesis was centered cover the three leading brands ranges of Tolsa's products:

1) Acid nano-catalyzers, corresponding to Minclear brand.
2) Mycotoxin detoxifiers, corresponding to Atox brand.
3) Rheological additives, corresponding to Berkbent brand.

Minclear brand incorporates products for bleaching earth application, meanly in edible oils. Two prototypes have been designed by machine learning during this PhD thesis. These prototypes were scaled-up, validated in the market by selected customers, and launched by Tolsa SA under the commercial names Minclear S296 and Minclear S285. Both are sepiolite-based acid nanocatalyzers with high performance as bleaching agents employed in oil refining. In particular:

- Minclear S296 is specially designed for rapeseed oil treatment, possessing remarkable filtration rates and low oil retentions. Minclear S296 is recommended for difficult-to-treat vegetable oils, competing even against some high-performance bleaching earth grades. The market opportunity behind this product is expected to reach the 2 M€ of turnover. The sales in 2022, the launching year, for this product, have been 120 k€, and one big oil refinery plant in the UK has been achieved, standing for 2023 1500 Tn/year, so 1.125 M€.
- Minclear S285 is designed for refining companies of multiple oil types of varying qualities with an excellent filtration rate. This product is recommended for customers with drastic variations of oil quality throughout the year, representing a versatile addition to Tolsa's product portfolio, essential for Minclear's positioning in the sector. The market opportunity expected for its first year will be 500 k€, with an opportunity for the third year three of 2 M€. The sales in 2022 have been during the validation phase, while the launch to the market is expected during the year's first quarter.

Atox brand corresponds with the mycotoxin detoxifiers products for animal feed. The investigation carried out during this PhD thesis was provided to confirm it as the top performed MDT of Tolsa's portfolio and discover its mode of action. The latter is continuously supporting the consolidation of Atox Nature Silver in the market, allowing it to achieve 950 k€ turnover during 2022, with a growth potential of 30% each year.

Berkent brand comprises products for drilling and civil engineering applications. The on-working model presented in Chapter 4 is exploited to design unique materials achieving particular properties while covering various product grades. Three prototypes were designed and scaled-up:

- Berkbent BTG is the product proposed for tunnel operations based on an optimized mixture of montmorillonites. The validation was completed successfully at the beginning of 2020. 300 Tn of the product was offered for tunnel work in Portugal in September '22, corresponding with 100 k€. During 2023 is expected to reach 500 k€ of the potential, opening opportunities outside of Spain.

- Berkbent HF40 is designed for the deep foundations market for hydrofraise and under challenging soil conditions. The opportunity represents 1 M€ for the third quarter of the 2023 year.
- Berkbent HY+ is a developed product for the horizontal directional drilling market that has evolved to a higher yield point. The potential behind this product for the first year is 1,2 M€ in Spain and more than 1 M€ in the UK. The launch is expected for the second quarter of 2023.

Another rheological product for the construction sector (Cimsil brand) is being scaled up, expecting validation during the rest of 2023, with a launching during the first of 2024. It is a global product with a significant market opportunity of 1,5M€ for the third year.

The success of this industrial PhD thesis highlights the benefits of a a profound collaboration between scientists and industries in focusing and exploiting the research to concrete applications. Moreover, the published contents are expected to assist the researcher in the context of natural porous materials. The main contributions can be grouped into three points:

1) Define a proper descriptor space for clays.
2) Rationalize the differences between the modification of clay materials and final physico-chemical properties.
3) Identify the key features affecting specific applications.

The challenge accepted in point one provides a material representation considering the nature of clay minerals (i.e., considering the impurities) rather than computing the clay features. Chapter 3 presents an example of how the natural contaminations of five samples of stevensites facilitate or prevent clay modification. The comprehensive information that the implementation of machine learning models can extract was exploited as tools to study the differences between the modification of clay materials and final physico-chemical properties (point two). For instance, the rationalization of the results of Chapter 4 provided the spots to understand the mode of action of

clay-based mycotoxin detoxifier, which would require lengthy and costly experimentation. Moreover, Chapter 5 presents an exciting play of diversity algorithms coupled with machine learning prediction, highlighting that material modification and properties may be correlated by a non-linear function and may be hard to capture by a simple human intuition. Machine learning allowed observing the difference in the gel structure created by diverse formulations of rheological additives. Visualizing the importance score assigned to the whole set of descriptors highlights the stand-out feature to be tuned when we design an optimal material. The latter, incorporated in point three, suggests to other researchers what should be the most interesting variables to be explored and studied, which may be hard to identify with pure experimentation. For example, the unexpectedly high importance of the clay moisture content, which we found during the development of the work presented in Chapter 2, stimulated the curiosity to conduct a more profound investigation into the role of confined water in clay minerals. Thus, in agreement with Tolsa and IMDEA, I invested an international stay in the Department of Physics and Chemistry of the University of Palermo, running experiments of modulated thermogravimetric analysis (MTGA). We found differences in the value of the calculated activation energy for removing the adsorbed water (Ea). The Ea was higher in fibrous clays than lamellar structures and highly correlated to adsorption applications. The investigation is detailed in Chapter 5.

2.3. General conclusions and future actions

The development of this Industrial PhD thesis is evidence of the collaboration between scientific research (IMDEA Materials) and industry (Tolsa company). The principal objective is to adopt machine learning technics to the material design, projecting companies such as Tolsa to the new paradigms of Industry 4.0. The thesis application's preliminary stage was oriented to selective clay-based adsorbents for organic compounds, e.g., emerging contaminants and mycotoxins. The reliable results that the project provided prompted the decision to re-plan the focus and extend the implementation of the approach to catalytic and rheological applications. Along this line, this PhD thesis touched the most critical markets of Tolsa products with great success, which, nowadays, can be counted in terms of sales. The company's benefits have been so promising that, at the end of the three years of the Industrial PhD fellowship, Tolsa created a new job position hiring me with the role of Machine learning Engineer in the Research & Technological Innovation department. During this time (March 2022 – January 2023) I was involved in most of Tolsa's industrial projects applying the expertise acquired during this thesis and coordinating various actions in developing novel products. The continous achieved improvement, has promoted my fast profesional growth into the Company. Nowadays, I rised the Laboraty's Manager of the Research & Technological Innovation department job position. My tasks and efforts are the coordination of all the department's projects, designing and organizing the required experimental work of the technicians and applying machine learning when it is needed.

Parallelly, with IMDEA Materials, we intend to keep open the investigation of the synergic effect and the design function, creating a smart design platform that may serve as a strategy in other similar areas of interest in materials science.

2.4. References

(1) Rudin, C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nat. Mach. Intell.* **2019**, *1* (5), 206–215. https://doi.org/10.1038/s42256-019-0048-x.

(2) Terayama, K.; Sumita, M.; Tamura, R.; Tsuda, K. Black-Box Optimization for Automated Discovery. *Acc. Chem. Res.* **2021**, *54* (6), 1334–1346. https://doi.org/10.1021/acs.accounts.0c00713.

(3) Zhong, X.; Gallagher, B.; Liu, S.; Kailkhura, B.; Hiszpanski, A.; Han, T. Y. J. Explainable Machine Learning in Materials Science. *npj Comput. Mater.* **2022**, *8* (1), 1–19. https://doi.org/10.1038/s41524-022-00884-7.

(4) Rinnert, E.; Carteret, C.; Humbert, B.; Fragneto-Cusani, G.; Ramsay, J. D. F.; Delville, A.; Robert, J. L.; Bihannic, I.; Pelletier, M.; Michot, L. J. Hydration of a Synthetic Clay with Tetrahedral Charges: A Multidisciplinary Experimental and Numerical Study. *J. Phys. Chem. B* **2005**, *109* (49), 23745–23759. https://doi.org/10.1021/jp050957u.

(5) Cases, J. M.; Bérend, I.; Besson, G.; Francois, M.; Uriot, J. P.; Thomas, F.; Poirier, J. E. Mechanism of Adsorption and Desorption of Water Vapor by Homoionic Montmorillonite. 1. The Sodium-Exchanged Form. *Langmuir* **1992**, *8* (11), 2730–2739. https://doi.org/10.1021/la00047a025.

(6) Derkowski, A.; Drits, V. A.; McCarty, D. K. Rehydration of Dehydrated-Dehydroxylated Smectite in a Low Water Vapor Environment. *Am. Mineral.* **2012**, *97* (1982), 110–127.

(7) Bérend, I.; Cases, J. M.; François, M.; Uriot, J. P.; Michot, L.; Masion, A.; Thomas, F. Mechanism of Adsorption and Desorption of Water Vapor by Homoionic Montmorillonites: 2. The Li+, Na+, K+, Rb+, and Cs+-Exchanged Forms. *Clays Clay Miner.* **1995**, *43* (3), 324–336. https://doi.org/10.1346/CCMN.1995.0430307.

(8) Zhang, X. hui; He, C.; Wang, L.; Li, Z. quan; Deng, M.; Liu, J.; Li, H. kui; Feng, Q. Non-Isothermal Kinetic Analysis of Thermal Decomposition of the Ca-Bentonite from Santai,

China. *Mineral. Petrol.* **2015**, *109* (3), 319–327. https://doi.org/10.1007/s00710-014-0331-9.

(9) Kuligiewicz, A.; Derkowski, A. Tightly Bound Water in Smectites. *Am. Mineral.* **2017**, *102* (5), 1073–1090. https://doi.org/10.2138/am-2017-5918.

(10) Kuang, W.; Facey, G. A.; Detellier, C. Dehydration and Rehydration of Palygorskite and the Influence of Water on the Nanopores. *Clays Clay Miner.* **2004**, *52* (5), 635–642. https://doi.org/10.1346/CCMN.2004.0520509.

(11) Bahranowski, K.; Klimek, A.; Gaweł, A.; Serwicka, E. M. Rehydration Driven Na-Activation of Bentonite—Evolution of the Clay Structure and Composition. *Materials (Basel).* **2021**, *14* (24), 1–13. https://doi.org/10.3390/ma14247622.

(12) Stackhouse, S.; Coveney, P. V.; Benoit, D. M. Density-Functional-Theory-Based Study of the Dehydroxylation Behavior of Aluminous Dioctahedral 2:1 Layer-Type Clay Minerals. *J. Phys. Chem. B* **2004**, *108* (28), 9685–9694. https://doi.org/10.1021/jp037608p.

(13) Ogorodova, L. P.; Kiseleva, I. A.; Vigasina, M. F.; Kabalov, Y. K.; Grishchenko, R. O.; Mel'Chakova, L. V. Natural Sepiolite: Enthalpies of Dehydration, Dehydroxylation, and Formation Derived from Thermochemical Studies. *Am. Mineral.* **2014**, *99* (11–12), 2369–2373. https://doi.org/10.2138/am-2014-4804.

Chapter 3 - Machine-learning-accelerated multimodal characterization and multiobjective design optimization of natural porous materials

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] and Maciej Haranczyk[a]

[a]IMDEA Materials Institute, C/Eric Kandel 2, 28906 Getafe, Madrid, Spain.

[b]Department of Materials Science and Engineering, Universidad Carlos III de Madrid, Avda. de la Universidad, 30. 28911 Leganés, Madrid, Spain.

[c]Tolsa Group, Carretera de Madrid a Rivas Jarama, 35, 28041, Madrid, Spain.

Chemical Science, 2021, 12 (27), 9309-9317.

Abstract. 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.

3.1. 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 and 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 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]

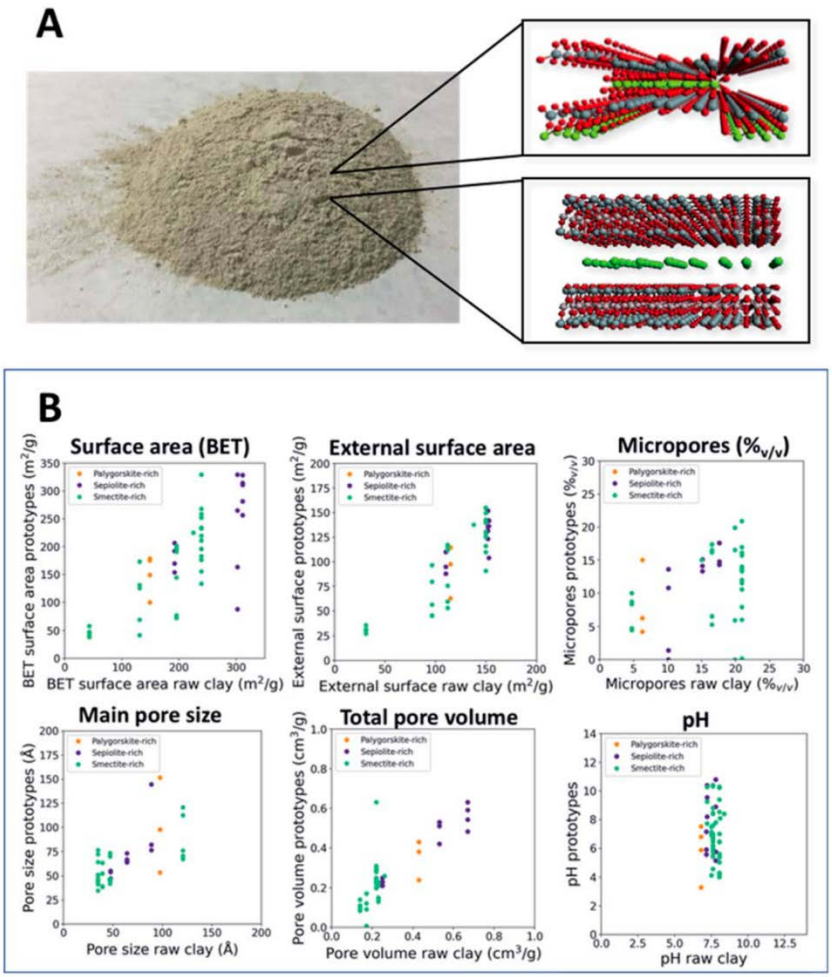


Fig. 1 Photo of a powdered clay together with two examples of fibrous sepiolite (top) and lamellar smectite (bottom) (A). Morphological properties and pH of processed materials (prototypes) compared to the corresponding features of the pristine, natural clays (B). Each raw material represents a discrete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

3.2. Methods

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 (pKa), 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 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 pKa, 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 mm 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.

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.

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):

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

where i indicates the target of Table 2 while x_{raw_clay}, $x_{additive}$, and $x_{modification\_process}$ 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]

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

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 ˚A 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% data- points (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 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.

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

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:

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

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 mm was fixed to 75%. The DF improvements with respect to the starting values were calculated as (DFproc - DFraw)/DFraw x 100, where DFproc and DFraw 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†).

3.3. Results and discussion

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 consid- ered, enhancing the errors.

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 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 behavior 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]

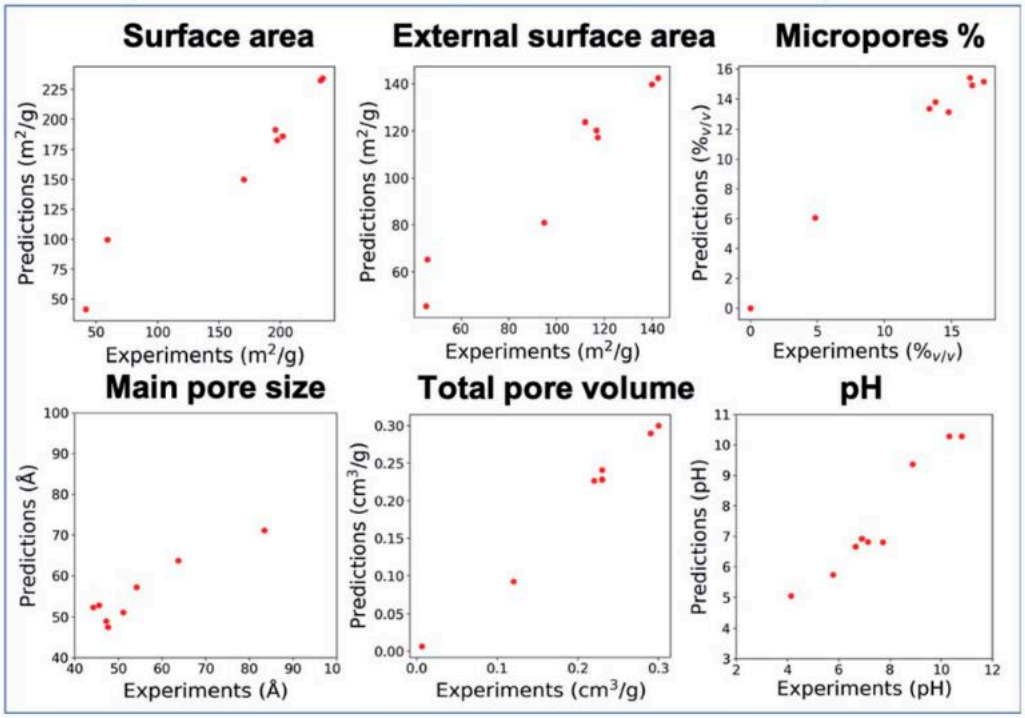


Fig. 2 Graphical model assessment results for the prediction of the labeled data of the test set.

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

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 char- acter 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.

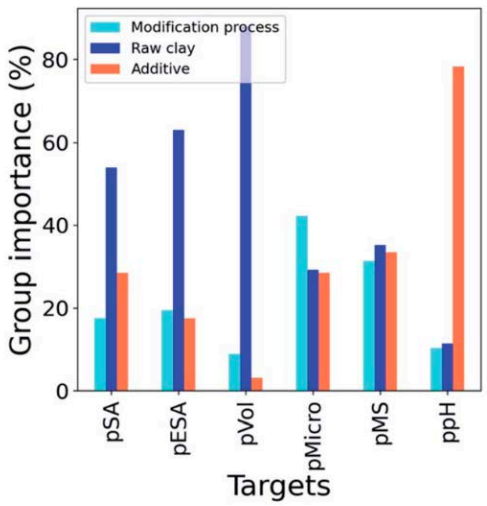


Fig. 3 Summed up importance scores of the features which were grouped according to Table 1.

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. Doble 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 steven- sites. 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.

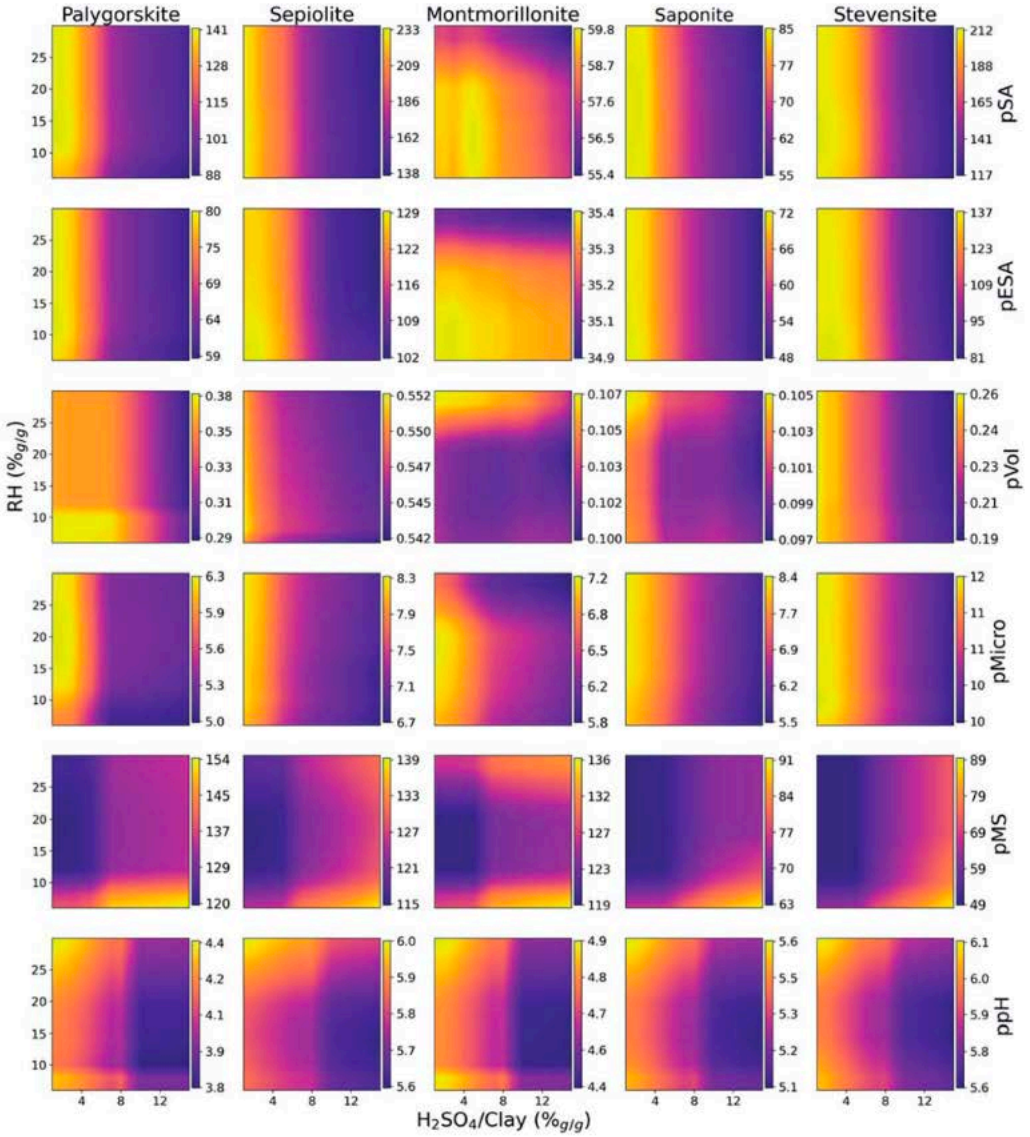


Fig. 4 Effect of the starting moisture content and H_2SO_4/clays% on the predicted target pSA, pESA, pVol, pMicro, pMS and ppH.

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 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 high-lighted 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.

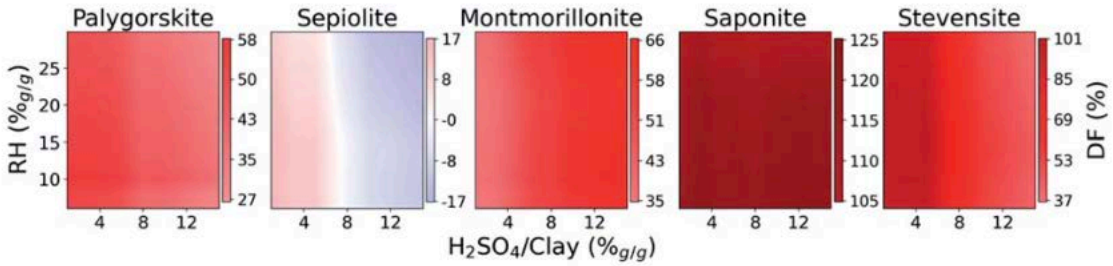


Fig. 5 Trend of design function improvements (%) with different additives/clays (%g/g) and starting moisture contents (RH%g/g).

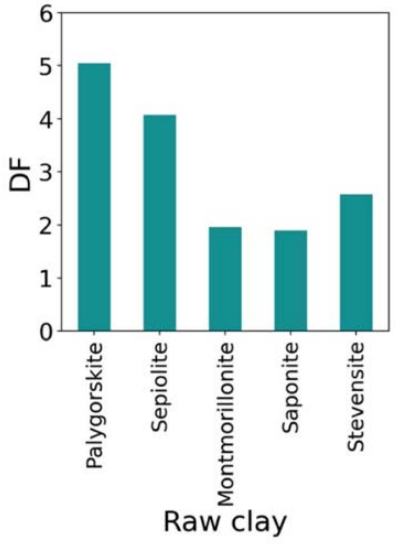


Fig. 6 Design function prediction for promising nano-catalyzers derived from different starting materials modified with sulfuric acid at its optimal amount (8% for palygorskite, 5% sepiolite, 12% montmorillonite, 3% for saponite and 6% for stevensite).

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

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 mm was set to 75%. To check if the above 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 a er 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.

3.4. 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 morpho- logical 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 mate- rial. 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.

3.5. Additional information

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 Alvaro Peña Nuñez performed the experiments and analysed the data with assistance from Veronica Carcelen. 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.

3.6. 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. Rodr´ıguez-Castell´on 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-Valdes, 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.Seoand 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. Petrovic, L. Rozic, Z. Vukovic, T. Novakovic 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. 49–81.

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. Rom´an-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.Heand 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. Madejov´a, 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.

3.7. Supporting information

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. N2 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 N2 isotherm.[6] Figure S1 shows examples of N2 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]

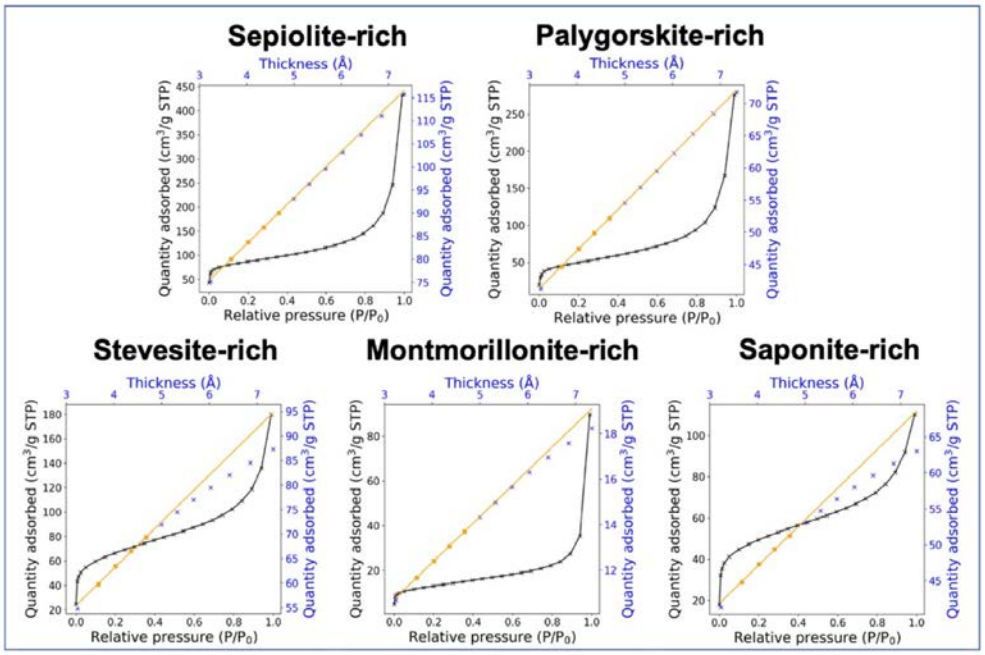


Fig. S1. Nitrogen adsorption isotherm linear plot and t-plots of natural clay minerals rich in sepiolite, saponite, palygorskite, stevensite and montmorillonite.

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.

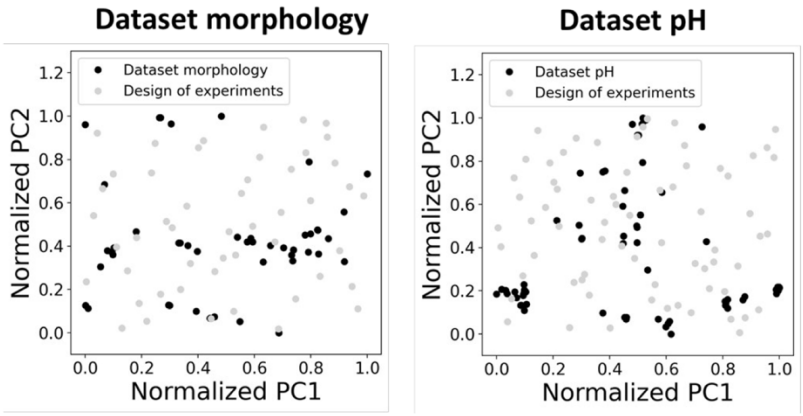


Figure S2. Graphical visualization of data point distribution into normalized PC1 and PC2 descriptors.

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.

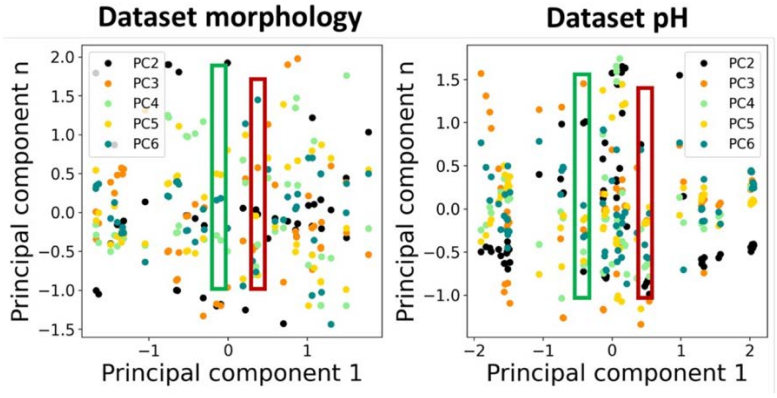


Figure S3. The plots of five principle components PC2-PC6 with respect to the PC1. The ranges of PCn values of prototype P1 is given in green box and P2 in red box.

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.

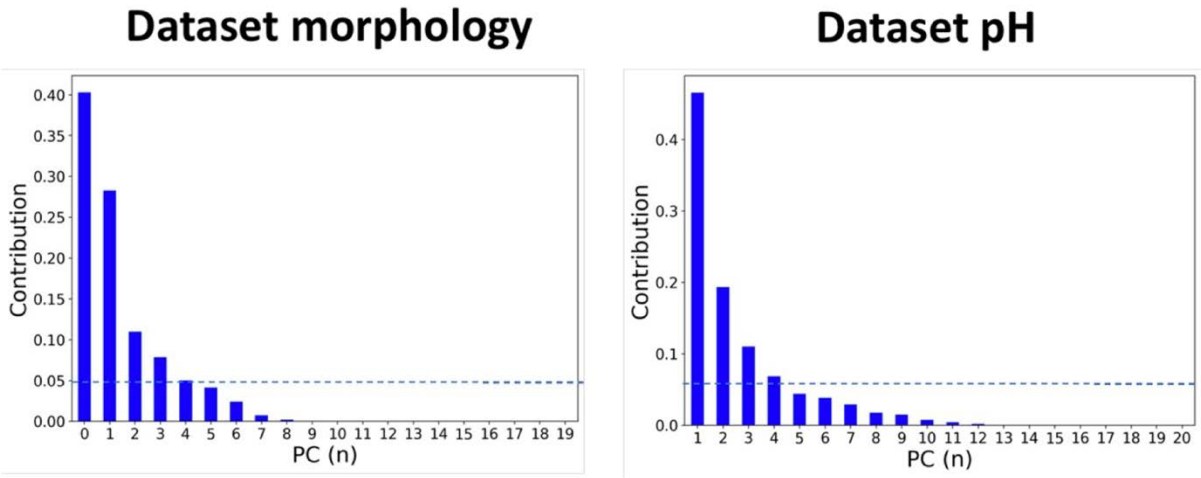


Figure S4. Individual contributions of the principal component in explain the total variance of the employed dataset (morphology and pH). The contribution score of 0.05 was chosen as threshold allowing the feature reduction to 4 main principal components.

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.

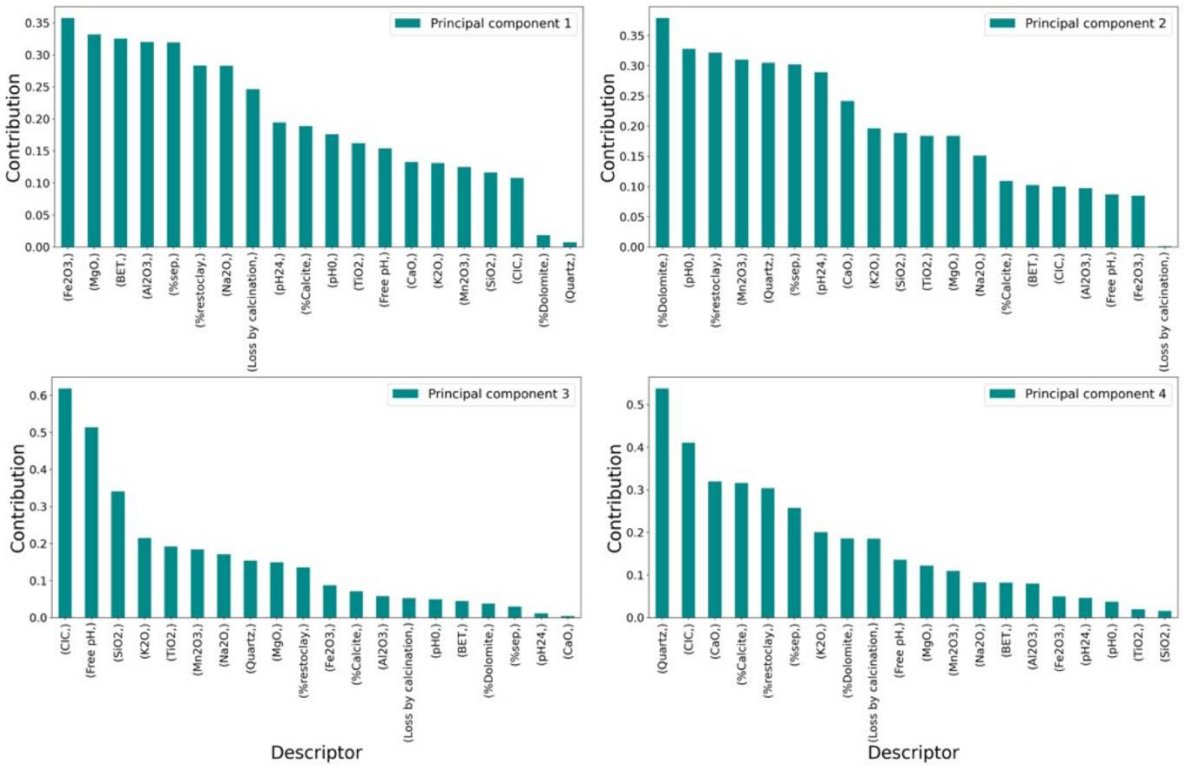


Figure S5. Contribution of the pristine features describing the properties of raw clay materials in the morphological dataset reduced by PCA. The 4 main PC are shown.

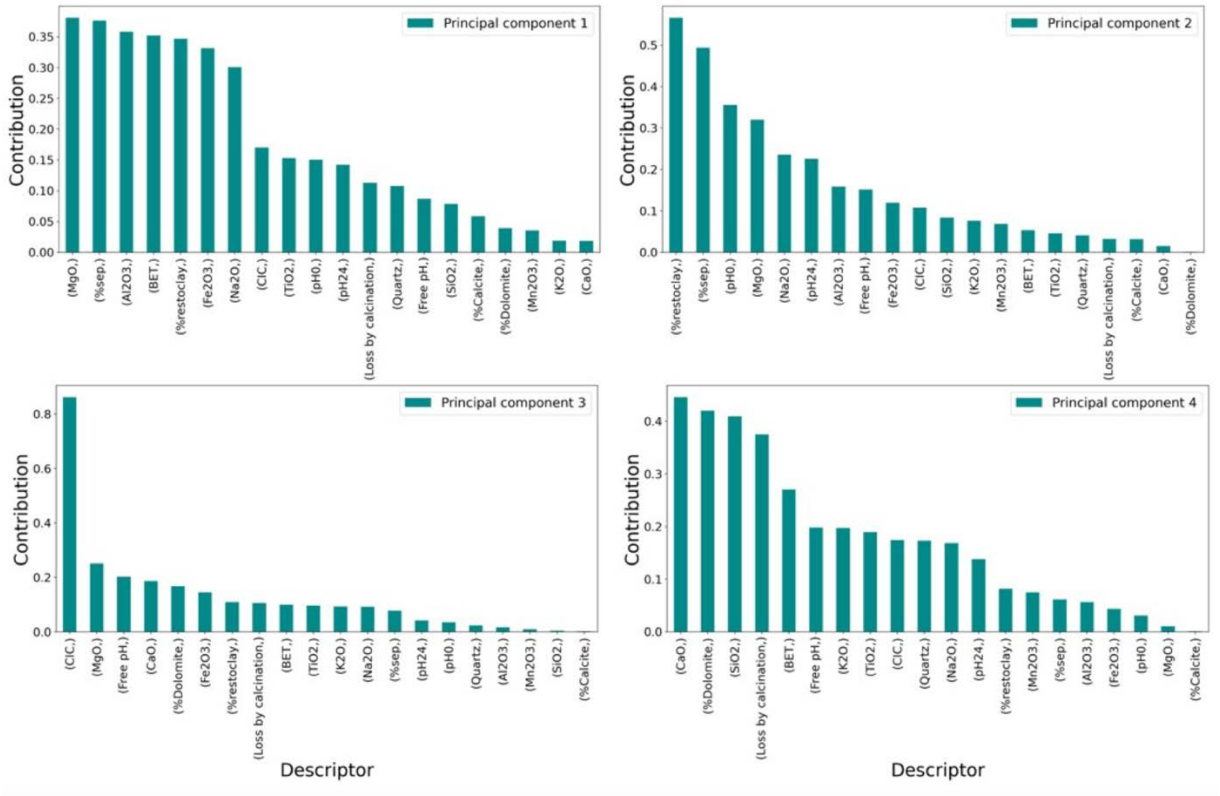


Figure S6. Contribution of the pristine features describing the properties of raw clay materials in the surface activity dataset reduced by PCA. The 4 main PC are shown.

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 H2SO4/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.

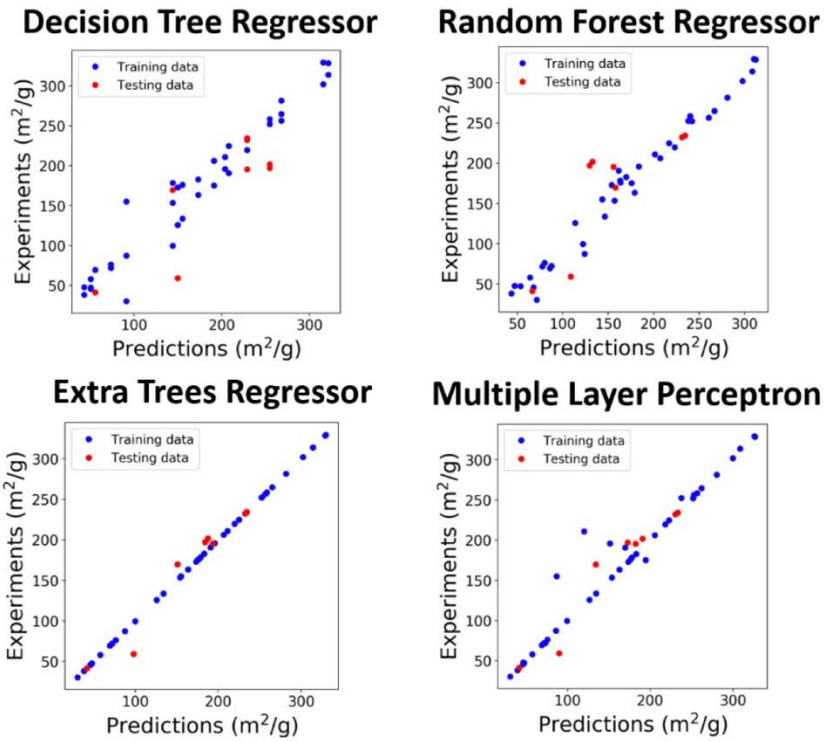


Fig S7. Visual assessment of BET surface area prediction models for modified clay. The dataset includes 49 data points, which were split 85-15% respectively into the training and test sets.

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

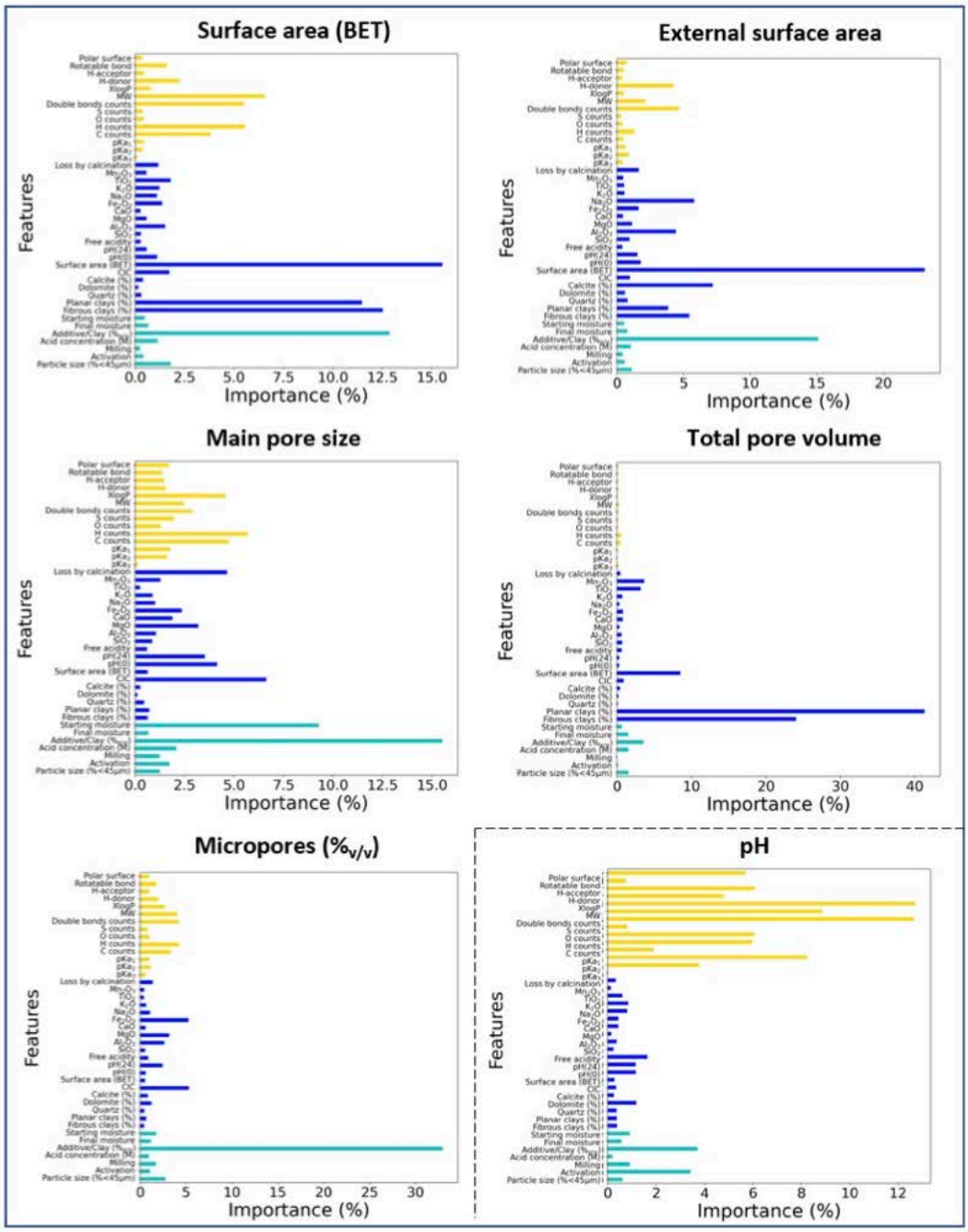


Fig S8. Feature importance scores for the Extra Trees Regressor-based models of properties of modified clay-based materials.

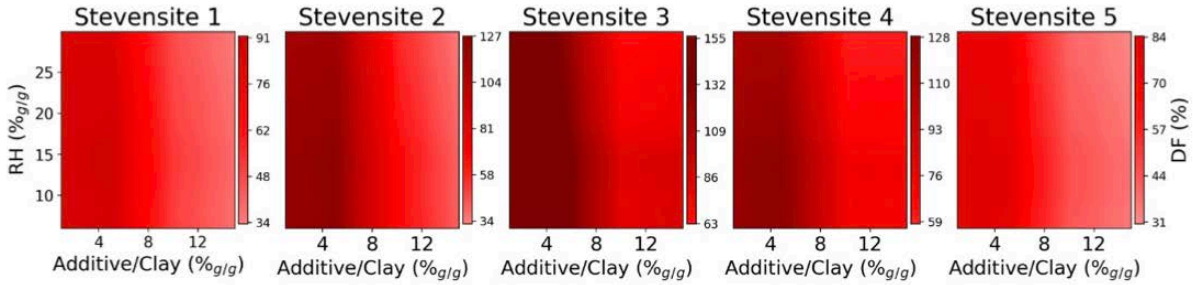


Fig S9. The plots of the design function improvements (in %) in the selected ranges of the additive/clay ratio and starting moisture content (RH %g/g) for five stevensites with different grade of purity (from 80% to 62% of phyllosilicate content).

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

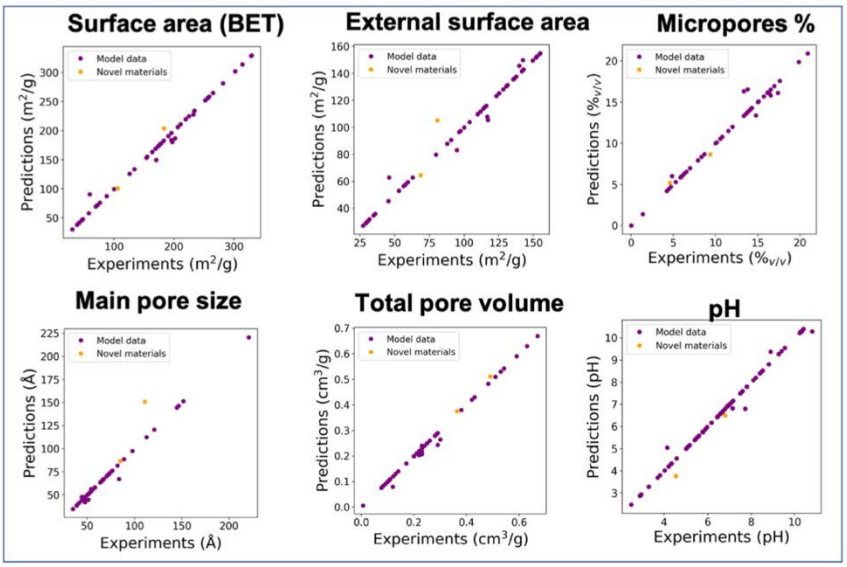


Fig S10. Visual assessment of the quality of Extra Tree model predictions for the novel materials synthesized in this work (P1 and P2).

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.

Chapter 4 - Machine earning-aided design of composite mycotoxin detoxifier material for animal feed.

Machine earning-aided design of composite mycotoxin detoxifier material for animal feed.

Giulia Lo Dico[a,b,c,+], Siska Croubels,[d,+] Verónica Carcelén,[c,+] and Maciej Haranczyk[a,*,+]

[a]IMDEA Materials Institute, C/Eric Kandel 2, 28906 Getafe, Madrid, Spain.

[b]Department of Materials Science and Engineering, Universidad Carlos III de Madrid, Avda. de la Universidad, 30. 28911 Leganés, Madrid, Spain.

[c]Tolsa Group, Carretera de Madrid a Rivas Jarama, 35, 28041, Madrid, Spain.

[d]Department of Pathobiology, Pharmacology and Zoological Medicine, Faculty of Veterinary Medicine, Ghent University, Salisburylaan 133, 9820, Merelbeke, Belgium.

[+]these authors contributed equally to this work

Scientific Reports, 2022, 12 (1), 1-11

Abstract. The development of food and feed additives involves the design of materials with specific properties that enable the desired function while minimizing the adverse effects related with their interference with the concurrent complex biochemistry of the living organisms. Often, the development process is heavily dependent on costly and time-consuming in vitro and in vivo experiments. Herein, we present an approach to design clay-based composite materials for mycotoxin removal from animal feed. The approach can accommodate various material compositions and different toxin molecules. With application of machine learning trained on in vitro results of mycotoxin adsorption–desorption in the gastrointestinal tract, we have searched the space of possible composite material compositions to identify formulations with high removal capacity and gaining insights into their mode of action. An in vivo toxicokinetic study, based on the detection of biomarkers for mycotoxin-exposure in broilers, validated our findings by observing a significant reduction in systemic exposure to the challenging to be removed mycotoxin, i.e., deoxynivalenol (DON), when the optimal detoxifier is administrated to the animals. A mean reduction of 32% in the area under the plasma concentration–time curve of DON- sulphate was seen in the DON + detoxifier group compared to the DON group ($P = 0.010$).

4.1. Introduction

The growing world population and its impact on the environment creates a need for advanced agriculture technologies and products, for example, highly target-specific animal feed additives to improve the condition and growth of livestock[1-3]. Mycotoxin detoxifiers (MDTs) are an example of such additives[4,5]. Mycotoxins are toxic secondary metabolites typically produced by fungi growing in food and animal feed commodities[6,7], and exposure to these contaminants may result in death or disease[8,9]. The maximum levels of some of the most highly prevalent mycotoxins in animal feed are already controlled guidance values or recommendations, rendering the need for mitigation strategies, one of which is mycotoxin capture by detoxifier feed additives[10]. The global market of the latter is projected to grow at the rate of 3.1% from 2020, and reaching a value of USD 3.1 billion by 2027[11-14].

A typical mode of action of MDTs involves either forming bulky non-absorbable complexes with mycotoxins in the gastrointestinal tract, hence reducing their oral bioavailability, or promoting their degradation into non-toxic metabolites by bio-transforming agents, such as bacteria or enzymes[1,15,16]. Adsorbents capture mycotoxins which are delivered through the gastrointestinal tract of the animals with their feed, and the mycotoxin-MDT complex is eliminated with feces thus minimizing the absorption in the blood stream[17]. Among the mycotoxin- binding agents (binders), inorganic porous materials such as clays minerals are recognized effective especially for sequestering of aflatoxin B1 (AFB1)[18,19]. However, the specific mode of action strongly depends on the affinity between clay and mycotoxins, which is a function of multiple variables representing the structure and properties of the mycotoxins, the adsorbing materials as well as the process, e.g. inclusion rate of MDTs[18]. For example, montmorillonites are commonly used for aflatoxins binding by creating complexes with the exchangeable cations, while stevensites exploit their larger surface area for efficient entrapping of ochratoxin A (OTA) and zearalenone (ZEN)[20,21]. Material processing strategies such as the preparation of organo-aluminosilicates have made improvement in the uptake capability of ZEN, OTA and T-2 toxin (T2)[22]. Similarly, other porous materials such

as activated charcoal (AC) have been recognized as strong adsorbents for several mycotoxins, including deoxynivalenol (DON)[23]. However, the required AC doses necessary for a significant detoxification leads to sequestrating of essential micronutrients such as vitamins or minerals[24]. Composite materials based on a mixture of the above may offer to capture and remove multiple large number of mycotoxins while minimizing the interference effect. Besides the regulated mycotoxins, yet-unregulated and emerging mycotoxins occur frequently in agricultural products[25], thus it is desired that the detoxifier design should address a wide variety of toxic fungal metabolites without compromising the animal health[26].

The development of advanced MDTs additives represents a multidisciplinary endeavor. It involves the material design of the additive to optimize selective toxin adsorption while relying on the fields of biochemistry, biology and veterinary medicine to understand the mechanisms of action and properly assess the impact, efficacy and potential side effects of the additive. In vitro experiments imitating the gastrointestinal environment may be used for the preliminary assessment of promising MDT, and are recommended by, for example, the European Food Safety Authority (EFSA)[27]. However, low fidelity of reproduction of the complexity of the true animal's digestive system by such models requires in vivo experiments for definitive performance evaluation, one of which are strategies based on biomarker for exposure detection[28-31].

Interestingly, if the interdisciplinary MDTs development were to rely on the state of the art of the respective disciplines, it would have to combine seemingly conflicting research and development strategies. In the case of material science, (semi)automated high-throughput synthesis and characterization, as well as computational screening approaches are allowing to generate and assess the performance of an unprecedented number of high performing materials, including porous materials[32-34]. On the contrary, the research involving living animals is guided by the principle of the "3Rs", i.e. replacement, reduction and refinement when animals are used for scientific purposes to improve animal welfare and to minimize the environmental impact[35,36]. Herein, we present an approach that addresses this issue and allows for screening of vast number of material formulations

while incorporating the animal-derived models for efficacy testing. Specifically, we have built machine learning (ML) models that incorporate three distinctive factors underlying the detoxifier performance, i.e. the material formulation, the chemical structure of the targeted toxin and the process in which a MDT is applied. The models being trained using an extensive set of in vitro experiments are used as surrogates of real experiments for the exploration of promising MDTs. In this study, we aim to demonstrate three-fold applications of our approach (1) in the identification of high performing formulations for the regulated toxins, (2) in predicting detoxification of a wide set of yet-unregulated mycotoxins, and (3) in gaining insights into the in vitro detoxification mode of action through model feature importance analysis. Finally, biomarker detection-based in vivo validation of the material identified in (1) is demonstrated in a challenging DON detoxification trial in broiler chickens.

4.2. Results and discussion

Predictive models and approach overview. The performance of MDTs could be most definitely determined on the basis of long-term studies assessing the growth and health of living animals[37,38]. It is, however, impractical and costly to implement such studies in the context of MDT material design. Instead, the performance is measured by either short-term in vivo determination of toxin biomarkers for exposure after single bolus administration, or by much faster in vitro experiments[27], which are set up to mimic the toxin adsorption–desorption cycle experienced by the detoxifier passing through the gastrointestinal tract[39] (see Supplementary Sect. S1 for details). The performance metrics corresponding to the latter approach, Ads(%) and Eff(%), are the percentages of the initial toxin level actively adsorbed on the MDT in the modeled gut experiment and removed from the digestive tract with the MDT, respectively. The sequestering capability of MDT is the function of the composition of the MDT material itself, the chemical structure of the targeted toxin as well as a number of parameters defining the application process such as dose[40]. In the applications thus far, the identification of the optimum of this function has been tedious and costly

due to a large parameters space and heavy reliance on experimentation determination of Ads(%) and Eff(%). Herein, we introduce an approach, which workflow is summarized in Fig. 1, that provided to build surrogate models of Ads(%) and Eff(%) allowing for a quick identification of powerful MDTs.

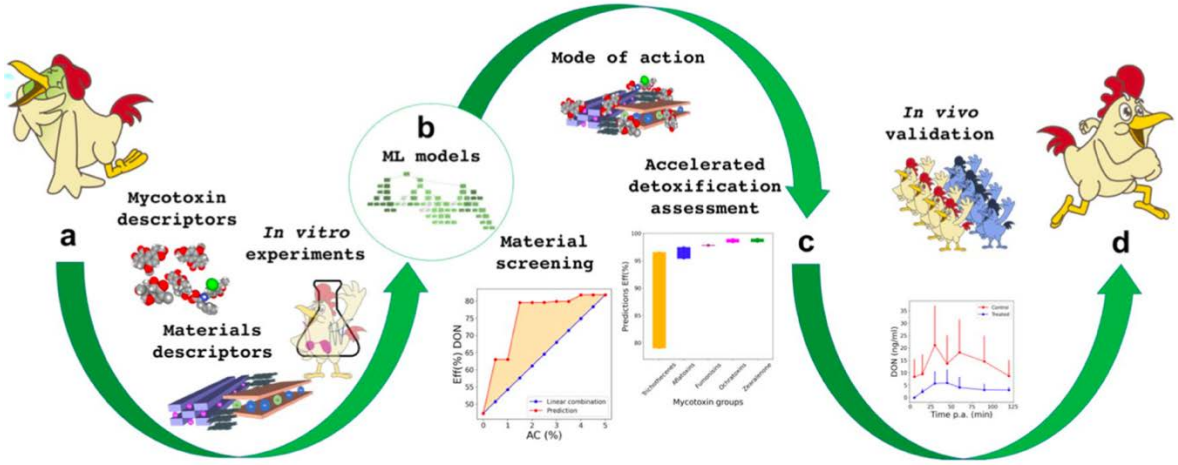


Figure 1. Workflow representation starting from selection of feature vector space (**a**) which describes the in vitro adsorption and efficiency (model targets). Machine learning models (**b**) trained on in vitro dataset providing tools for material screening, wide in vitro detoxification assessment, and mode of action capturing (**c**). In vivo validation of the findings extracted by our approach (**d**).

In our approach, the information on MDT performance is a function of descriptor space which encodes clay materials with experimentally-determined physico-chemical features, mycotoxins with molecular descriptors, and the modeled in vitro experiments with five setting parameters (Fig. 1a). Then, we develop random forest regressor models (RF) predicting both Ads(%) (denoted RFads) and Eff(%) (denoted RFeff) using 84 experimental data points (Fig. 1b). The data distribution shown in Supplementary Fig. S1, demonstrates the low correlation between in vitro adsorption, desorption and efficiency outcomes contained in the dataset. The dataset comprises 15 diverse toxin-detoxifier materials (MDTs), i.e., 10 natural clays and 5 clay-based composites in which

formulation the organic component is activated charcoal (AC) (see method section for details). The two independent machine learning models exhibit high predictive performances with R^2 score of 0.92 and 0.96 for RFads and RFeff, respectively (details in method section). Upon the positive assessment of the models, three main contributions were investigated, i.e., screening of optimal solutions, wide in vitro detoxification assessment, and mode of action capturing (Fig. 1c). Finally, biomarker detection-based in vivo trial in broiler chickens was performed as proof-of-concept to validate our approach, hence, the sequestration power of the top powerful MDT towards the sequestration of a mycotoxin (DON) selected among the most challenging to be mitigated (Fig. 1d).

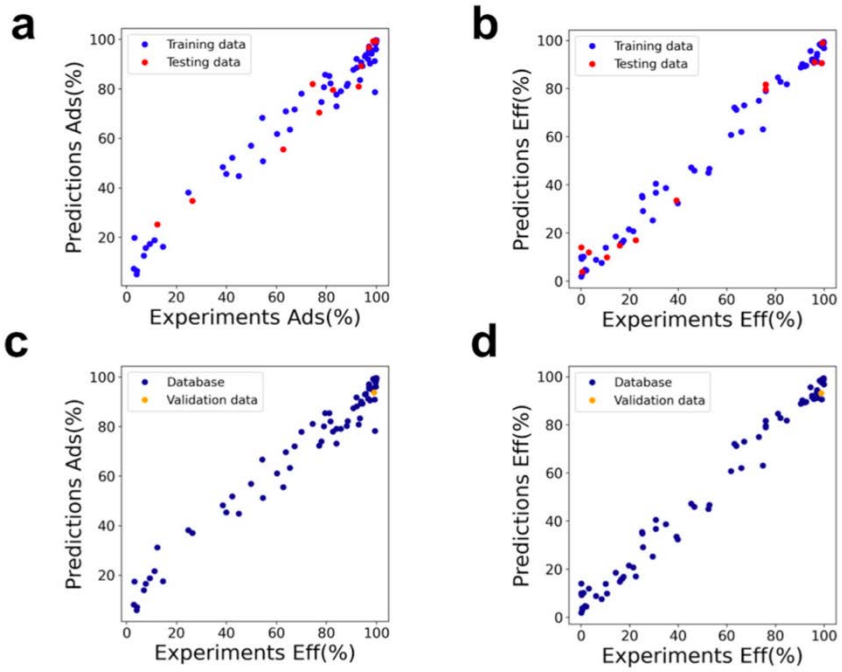


Figure 2. Graphical assessment of RFads (**a**) and RFeff (**b**) of 6 mycotoxins by 15 MDTs under different experimental conditions. Validation of RFads (**c**) and RFeff (**d**) toward one yet-unregulated toxin (DAS).

Figure 2a, b provide the graphical models assessment and Supplementary Fig. S2 summarizes the associated mean absolute error distributions given by the model, which exhibit apparent absence of systematic errors. The models were further validated by predicting and testing the Ads(%) and

Eff(%) of a composite material, which incorporates 5% AC, toward an out-of-trainset trichothecene toxin, i.e. diacetoxyscirpenol (DAS). Our models are able to capture and quantify the correlation between the differences among mycotoxin chemical structures and the physico-chemical features of MDT providing extremely accurate predictions (Fig. 2c, d) as well as reproducing the experimentally obtained ranking: DON > DAS > T2. Specifically, the chemical structures of two trichothecenes contained into the dataset (DON, T2) were compared with the chemical structure of DAS and shown in Supplementary Fig. S3. T2 joins two acetoxy and one methylbutanoate group in the DON structure, while DAS only shares with T2 the acetoxy groups. The corresponding slight variations of mycotoxin molecular properties complicate the human interpretability, while the appreciable model accuracy allows screening MDT materials toward the removal of yet-unregulated mycotoxins enabling fast future detoxification assessment in case of emerging contaminants.

Mode of action analysis. Feature importance analysis of the ML models provides insights into the main descriptors affecting the detoxification and, indirectly, the mode of action of MDT in in vitro scenarios. Figure S4 depicted the feature importance scores extracted by RFads and RFeff. We analyzed separately the impact of each group of descriptors by summing the corresponding scores assigned to the descriptors belonging to its group (Fig. 3a). We then normalized the individual contributions from 0 to 100% where the 100% is assigned to the summed importance score of the experimental conditions (Fig. 3b), materials (Fig. 3c) and mycotoxin (Fig. 3d) group of features. The main contribution to Ads(%) comes from material descriptors while Eff(%) is mostly determined by the conditions of the in vitro experiments (Fig. 3a), in which the inclusion rate and the MDT/toxin ratio possess the highest score (Fig. 3b). The high contribution of the latter enables our models to screen possible dosages identifying the recommended amount of MDT administration for specific contamination levels. The mode of action is dominated by the

unequivocal heavy contribution of Mg^{2+} (Fig. 3c), which suggests that the adsorption mechanism involves substitutions of the exchangeable cations and chelate formations.

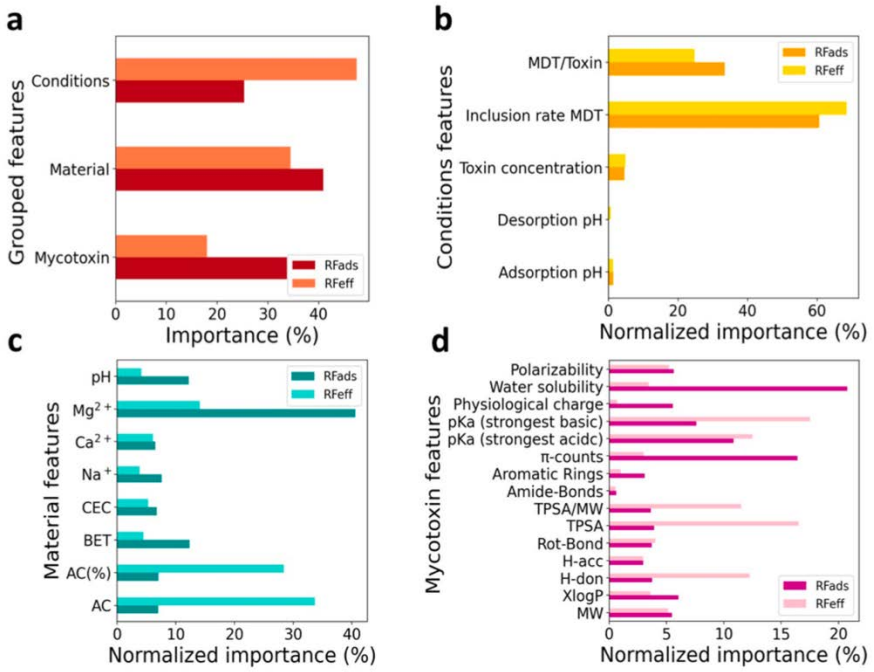


Figure 3. Summed importance score of the groups of features outlined in Table 2, extracted by RFads and RFeff (**a**). Normalized individual contribution corresponding to in vitro experimental conditions (**b**) materials (**c**) and mycotoxin (**d**) group of features.

However, the importance score distribution indicates that a multifunction of diverse interactions e.g. hydrogen bonds, Van der Waals, electrostatic and hydrophobic interaction are involved. The type of AC, i.e. high or low surface area form, and its content in composite formulation (AC(%)) are dominating the RFeff. Water solubility and π-counts are the main mycotoxin descriptors (Fig. 3d) involved during the adsorption step, suggesting that the primary interaction is entropically promoted by hydrophobic and π effects which involves replacing of solute–solvent and surface-solvent bonds by solute-surface and solvent–solvent bonds. Those are weak interactions which prevent a robust binding with the detoxifier along the whole digestive tract and lose importance when the desorption step is considered. The high contribution in RFeff of mycotoxin's pKa, the

topological polar surface area, and amount of hydrogen-bond donors indicates that electrostatic interactions with MDT surface are responsible for ensuring an effective sequestration during the modelled gastrointestinal tract.

Towards optimal MDT designs. The experimental in vitro performance of three different pure clays against the uptake of the six main mycotoxins included in the dataset (Supplementary Fig. S5) reveals the clay's selectivity towards specific mycotoxins, e.g. sepiolite and montmorillonite exhibit a stronger affinity towards DON than smectite. An effective uptake of multiple mycotoxins can be achieved by incorporating various components in the MDT design. The latter can be achieved by the optimal formulation in which the components positively cooperate increasing the simultaneous binding of diverse mycotoxins.

In order to identify MDTs with high performance, expressed by RFeff, we screened a number of hybrid formulations which incorporate clays with activate carbon. The latter, being highly porous may cause the sequestration of micronutrients prejudicing the animal health. Together with the highest efficiency we addressed the optimization looking for those formulations in which each component positively cooperates with the others in order to reduce the amount of activated carbon need to reach the desired performance, hence limiting the interferences with micronutrients. Thus, we proposed a function to estimate the synergic effect between components and it is expressed by Eq. (4). The synergy was estimated as the difference between the predicted Eff(%) of hypothetical formulations and the linear combination of Eff(%) associated with a mixture of pure components (details in "Methods" section). The descriptors of hypothetical materials were calculated as a linear combination of the pristine descriptors (see "Methods" section for details). The capability of RFeff in estimating the synergistic effect was assessed predicting the Eff(%) of a number of sepiolite-montmorillonite mixtures. Supplementary Fig. S6 demonstrates the positive cooperation between the two explored clays (sepiolite and montmorillonite). One of the formulations that expressed the highest synergy, composed of sepiolite/montmorillonite in ratio 1:4, was prepared and tested for

an in vitro detoxification of ZEN. The experiments are close to the prediction of Eff(%) within the corresponding error bars, i.e., the model predicted an Eff(%) of 58.3 ± 4.89 while the experimented Eff(%) was 67.5 ± 5.7. The following exploration of possible composite material compositions have additionally incorporated high surface AC as a third component. In this case, the synergy was defined as the difference between the predicted outcomes of hypothetical formulations and the linear combination of the outcomes of composite formulation with the highest amount of AC (fixed at 5%) (details in "Methods" section).

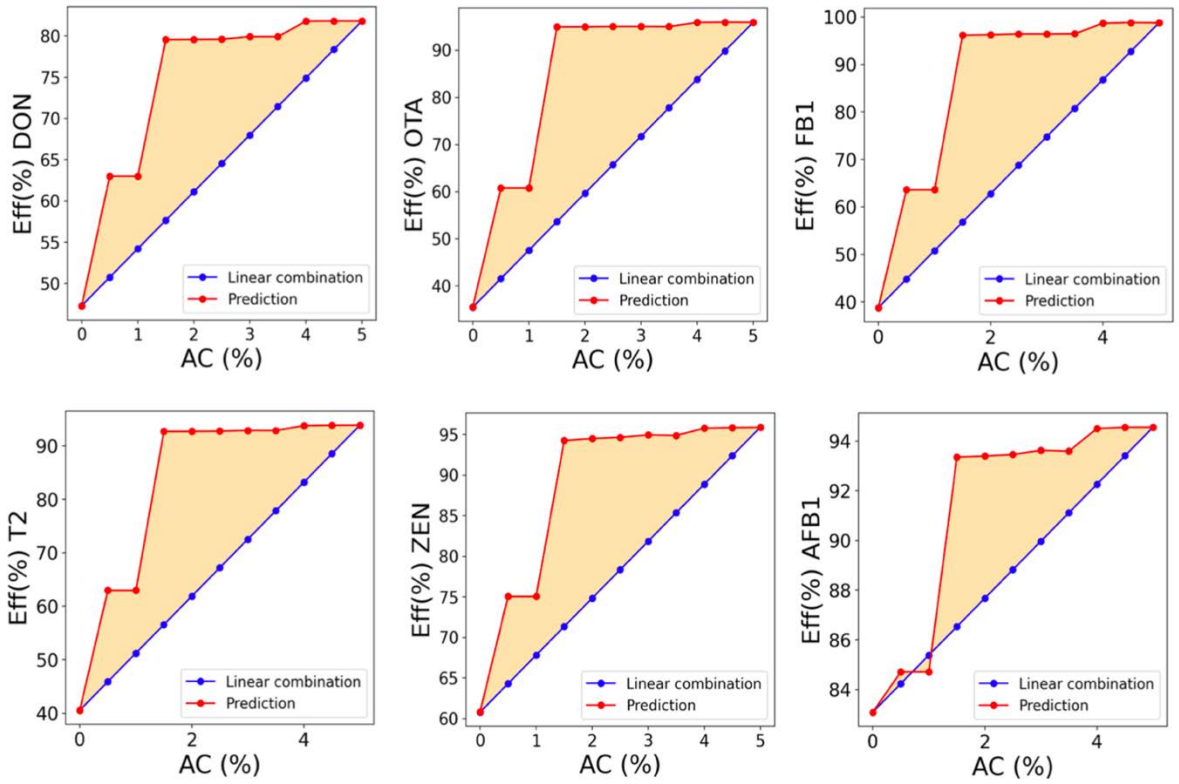


Figure 4. Synergy capturing by RF predicting efficiency for a series of hybrids of sepiolite-montmorillonite-charcoal in which the sepiolite-montmorillonite ratio was fixed to 1/4. The orange area is assigned to the positive synergistic effect. The experimental settings for the uptake of DON, OTA, T2, FB1 and ZEN were fixed to 2 kg/t of inclusion rate of MDT, 2 µg/ml of toxin concentration. The pH during the adsorption experiment was fixed to 3 while the desorption pH was 6.5.

All the possible sepiolite-montmorillonite-charcoal combinations were screened by RFeff and the optimal formulations were identified in those incorporating 20–40% of sepiolite and at least AC 1.5%. Figure 4 enables synergy exploration of optimal formulations during the successful removal of the main regulated mycotoxins. We observed a plateau starting from those hybrids which incorporates sepiolite-montmorillonite in ratios of 1/4 and 1.5% of AC. The latter formulation also express the highest synergy. However, the composite sepiolite-montmorillonite in ratio 1:4 and 2% of AC (SEP/MONT/AC) was selected as the top performing MDT due to easier material preparation scale-up.

Machine learning-aided assessment of the identified SEP/MONT/AC composite. Our computer-aided approach enables fast screening of emerging contaminants limiting the need for animal tests, hence following the 3Rs principles. Moreover, the practical implementation of RFeff in screening diverse inclusion rates and toxin concentration values recommends the optimal to be administrated MDT dosage under specific contamination levels. The high accuracy of RFeff model in predicting the detoxification of out-of-dataset molecules (Fig. 2d) was employed to assess SEP/MONT/AC against the removal of a large set of yet-to-be-regulated mycotoxins (Supplementary Table S1). Supplementary Figure S7 summarizes the individual predicted efficiencies. Figure 5a reveals the reliable detoxification of SEP/MONT/AC achieving an efficiency of at least 94% in all the explored toxin groups, except for the trichothecenes, which are the most heterogeneous group of molecules. DON was found to be the most challenging-to-be-removed toxin by SEP/MONT/AC, thus it was selected to validate our findings through an in vivo test.

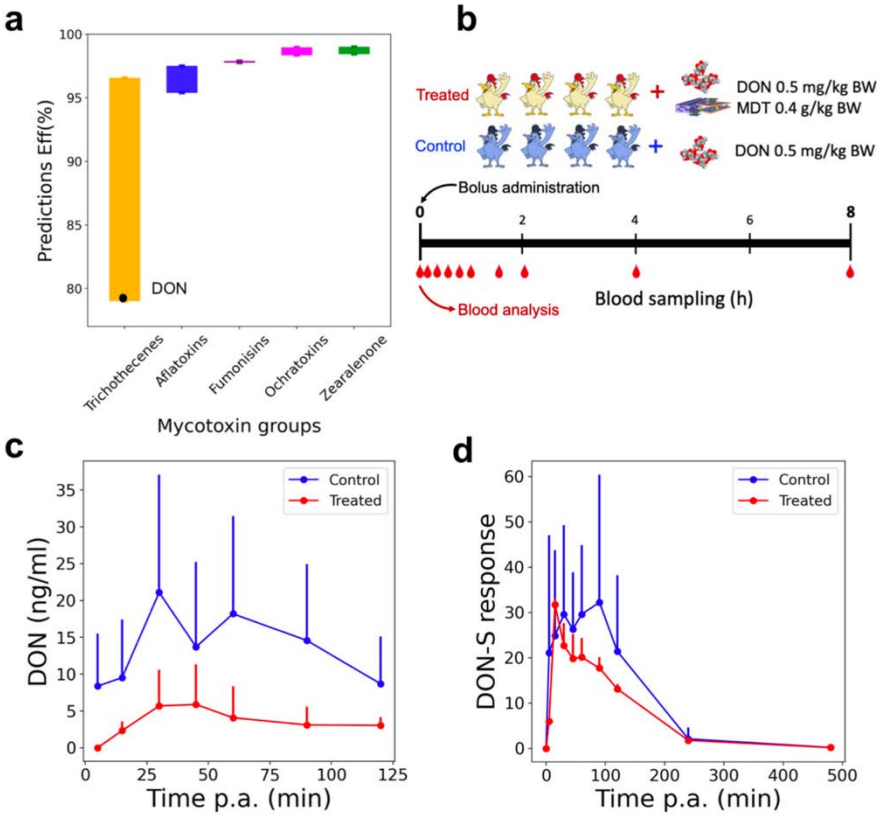


Figure 5. Response of predicted in vitro efficiency of SEP/MONT/AC towards the removal of the explored mycotoxin groups (**a**). The predictions were obtained by RF fixing the experimental setting to 2 kg/t of inclusion rate of SEP/MONT/AC, 2 μg/ml of toxin concentration, adsorption and desorption pH to 3 and 6.5, respectively. Graphical representation of in vivo design of experiment validating the detoxification of DON by SEP/MONT/AC detoxifier (**b**). Mean plasma concentration and standard deviation of DON after single oral bolus administration of DON alone (0.5 mg/kg BW) and DON in combination with SEP/MONT/AC (0.4 g/ kg BW) to 8 broiler chickens (**c**). Mean response and standard deviation of deoxynivalenol-sulphate (DON-S) in plasma after single oral bolus administration of DON alone (0.5 mg/kg BW) and DON in combination with SEP/MONT/AC (0.4 g/kg BW) to 8 broiler chickens (**d**).

In vivo validation assessment. We further evaluated the sequestration power of the identified SEP/MONT/AC composite through in vivo experiments carried out by a biomarker detection of DON[27], albeit after a single bolus toxin exposure. We performed the in vivo trial to validate our

findings by observing the DON relative oral bioavailability when the detoxifier is administrated to broiler chickens. The design of in vivo trial is schematized in Fig. 5b and detailed in method section and Supplementary Sect. S1. The broiler chickens were equally divided into treated (0.5 mg/kg BW of DON and 0.4 g/kg BW of SEP/MONT/AC, inclusion rate of 4 kg/t) and control (0.5 mg/kg BW of DON) groups, their repetitive blood analysis was performed recording the mean plasma concentration–time curve of DON. Figure 5c displays a significant reduction in DON oral availability after oral bolus administration of DON, whether combined with the detoxifier (treated group) compared to control group. Since in many samples DON could not be detected, or the concentration was below the limit of quantification of 1 ng/ml, no accurate toxicokinetic modeling could be performed for DON. Therefore, the metabolite DON-sulphate (DON-S) was used to perform the modeling. This metabolite is a major phase II biotransformation product of DON and hence a suitable biomarker for DON exposure in broiler chickens[41-43]. Figure 5d demonstrates the mean plasma response-time curve of the metabolite DON-S after oral bolus administration of DON, whether or not combined with the mycotoxin detoxifier to 8 broiler chickens. Table 1 summarizes the results of the most important toxicokinetic parameters of the metabolite DON-S after oral bolus administration of DON to 8 broiler chickens, whether or not combined with the detoxifier. A significantly lower systemic exposure of DON-S, expressed as $AUC_{0-\infty}$, was seen for the DON + detoxifier group compared to the DON group ($P = 0.010$). Also, a statistically significant lower Cmax for DON-S was seen for the DON + detoxifier group compared to the DON group ($P = 0.013$). Furthermore, the relative oral bioavailability of DON-S in the DON + detoxifier group compared to the DON group was only 67.8%. This shows a relevant reduction in in vivo systemic exposure to DON, since more than 20% reduction was seen. Moreover, the 90% CI of the ratio of $\log(AUC_{0-\infty})$ was 54–88% and therefore almost completely fell outside the range 80–125%, showing a significant effect of the detoxifier in reduction of systemic exposure to DON.

TK parameter (units)	DON	DON + SEP/MONT/AC	*P* value
$AUC_{0-\infty}$(min.response)	4305.39 ± 1684.29	2919.59 ± 1032.94	**0.010**
C_{max} (response)	51.93 ± 21.04	34.39 ± 17.06	**0.013**
T_{max} (min)	64.38 ± 42.71	56.25 ± 38.24	0.637
k_e (1/min)	0.013 ± 0.001	0.013 ± 0.002	0.976
$T_{1/2e}$ (min)	53.09 ± 6.09	53.77 ± 9.40	0.877
Relative F $AUC_{0-\infty}$ (%)	/	67.81	/
90% CI for log($AUC_{0-\infty}$)	/	[0.54; 0.88]	/

Table 1. Toxicokinetic (TK) parameters of the metabolite DON-S after oral administration of DON alone (0.5 mg/kg BW) and DON in combination with the detoxifier (0.4 g/kg BW, inclusion rate of 4 kg/t) to 8 broiler chickens. The mean ± standard deviation (SD) is shown. Significant values are in bold. $AUC_{0-\infty}$ area under the DON-S response-time curve from time 0 to infinity, *Cmax* maximum DON-S response in plasma, *Tmax* time at maximum DON-S plasma response, *Relative F* relative oral bioavailability, *CI* confidence interval, /: not applicable.

4.3. Conclusions

We have demonstrated a machine learning-aided approach to the design of mycotoxin detoxifiers (MDTs). We used a dataset of experimental in vitro assessment of the adsorption and efficiency of various MDTs against regulated mycotoxins (DON, T2, ZEN, OTA, FB1, AFB1) to build two random forest (RF) models which predict the adsorption (RFads) and efficiency (RFeff). The model feature space was defined by descriptors representing MDT and mycotoxin physico-chemical properties and in vitro experimental setting which modelled the animal gastrointestinal tract. The model achieved satisfactory performance with R^2 of 0.92 and 0.96 and MAE of 5.5% and 4.8% for RFads and RFeff, respectively. By means of feature importance analysis of the models, we gain insights into the MDT mode of action i.e., the unequivocal heavy contribution of Mg^{2+} suggests that the main mechanism processes through substitutions of the exchangeable cations and chelate formations. Our RFeff being skilled to capture the synergy between components MDT formulations was employed in the recognizing of specific material preparation

which avoids interferences originated by micronutrient adsorption. The top promising formulation incorporates sepiolite, montomorillonite and activated carbon in ratio 1:4:0.1 (SEP/MONT/AC). Our RFeff targets DON as the most challenging to-be-removed molecule which was trained for the in vivo validation of the real performance of SEP/MONT/AC on the oral bioavailability of DON in broiler chickens. The specifically designed in vivo trial avoids the repetition of pointless use of animals following the 3Rs principles. Both DON and the corresponding phase II metabolite (DON-S) detection demonstrated the relevant reduction in in vivo systemic exposure to DON in broiler chickens, confirming the findings of our approach, i.e., the capability of the identified material to mitigate the presence of the challenging to be removed DON. Our computer-aided approach enables versatile applications, e.g. assisting the recommendation of the precise administration dosage of MDTs for specific contamination levels, and providing the estimation of adsorption efficiency of MDTs against yet-to-be-regulated mycotoxins projecting future assessment of emerging contaminants.

4.4. Methods

MDT in vitro assessment. The in vitro adsorption experiments employ a specific amount of MDT with the appropriate amount of mycotoxin (ranging between 0.5 and 5 μg/ml) in phosphate buffer adjusted to the required pH (see details in Supplementary Sect. S1). The adsorption is then estimated against the amount of toxin in the standard by the following Eq. (1):

$$\mathrm{Ads}(\%) = 100 - (A_{ft}/A_{st} * 100) \quad (1)$$

where A_{ft} and A_{st} are the amount of free toxin, after separating the complexed MDT-mycotoxin and the toxin amount in the standard, respectively. The MDT-mycotoxin complex is then dispersed in buffered aqueous solution of neutral pH, followed by the incubation and centrifugations steps. The desorption is calculated as follow (Eq. (2)):

$$Des(\%) = A_{rt}/A_{st} * 100 \quad (2)$$

where A_{rt} is the amount of toxin released by the MDT. The efficiency is then calculated as (Eq. (3)):

$$Eff(\%) = Ads(\%) - Des(\%) \quad (3)$$

The dataset of in vitro adsorption and desorption experiments contains 51 data points related to 10 natural clays and 33 data points of 5 clay-based composites which incorporate activated charcoal (AC) as organic component. The natural clays represent four morphological diverse classes and were, i.e., 2 sepiolites, 2 swelling montmorillonites, 2 non-swelling montmorillonites, and 4Mg-rich smectites, with high purity (> 80% of phyllosilicate content, the remaining amount are impurity such as carbonates, quartz, and calcites, not considered as factor of interest in our application). The descriptor space was defined by selecting the main features which can efficiently represent MDT materials, mycotoxins and in vitro experimental settings (summarized in Table 2). Clays are identified by experimentally obtained properties (details in Supplementary Sect. S1). Two additional descriptors were included completing the vector space for the recognition of composite materials, i.e., type of AC (0 for inorganic material, 1 for CA with 900 m^2/g and 2 for AC with 1200 m^2/g), and its content with respect to dried inorganic component (AC (%)). Mycotoxins were featured by the most common molecular descriptors supplied in PubChem repository and Toxin and Toxin Target Database (T3DB). In vitro experiments were modeled by adsorption and desorption pH, inclusion rate, toxin concentration and the relative amount of MDT with respect to the toxin concentration (MDT/Toxin) details in Supplementary Sect. S1). The targets were the previously defined Ads(%) and Eff(%) outcomes.

Table 2. Input vector space.

Mycotoxin molecular descriptor	Detoxifier material descriptors	Experimental settings
Molecular weight (MW)	Specific surface area (BET)	Adsorption pH
Octanol/water partition coefficient (XlogP)	Cation exchange capacity (CEC)	Desorption pH
Hydrogen bond donor (H-don)	pH	Inclusion rate MDT
Hydrogen bond acceptor (H-acc)	Exchangeable cations (Na^+, Ca^{2+}, Mg^{2+})	Toxin concentration
Rotatable bonds (Rot-Bond)	AC	MDT/toxin
Topological polar surface area (TPSA)	AC (%)	
Specific topological polar surface area (TPSA/MW)		
Amide bonds		
Aromatic rings		
π counts		
pKa (strongest acid)		
pKa (strongest basic)		
Physiological charge		
Water solubility		
Polarizability		

Model architectures. Random forest regressor (RF) was trained using sci-kit learn Python library. The performance, in terms of cross validation score, was assessed against Extra trees regressor (EXT) predicting the detoxification efficiency. Both models predicting Ads(%) and Eff(%) were hyperparameter optimized on the base of k-fold cross validation with K = 5. The optimized hyperparameters as well as the assessment scores are given in Supplementary Tables S2 and S3, respectively. Random forest was selected for its outperforming, low variability within the fold switching and clear model interpretability, exhibiting a R^2 of 0.84±0.08 in cross-validation of trainset (K = 5) compared to 0.76 ± 0.14 of EXT. The algorithms were trained on 85% of the available dataset and tested in the remaining 15%. Diverse predictive models, i.e., support vector machine, multiple layer perceptron and K-nearest neighbors were benchmarked. The algorithms were assessed by means of R^2, MEA, accuracy, and p-value, the latter being calculated with the 5×2cv paired t test (details available in Supporting Information). The computed p-values allowed to compare RF to the other models to observe a significant difference ($P < 0.05$) in the performances of multiple layer perceptron and K-nearest neighbors and considerable statistical

difference between RF and support vector machine. Details are available in Supporting information, Sect. S2, Fig. S8, and Table S4. The models' codes are available in zenodo repository with the identifier https://doi.org/10.5281/ zenodo.5793956.

Capturing synergy. A number of hybrid formulations were screened for searching of optimal detoxifier material. RF predicted the efficiency allowing to identify the top promising formulation in composite materials composed by SEP, MONT and AC organic compound. A set of diverse formulations was proposed fixing the inorganic portion to SEP/MONT 1:4 and varying the AC from 0 to 5% with respect to the composite weight. Each material descriptor value of the proposed formulations was linearly calculated as Σ_i $a_i x_i$, where a_i is the amount (%) of the i-component in the hybrid formulations and x_i is the descriptor value corresponding to the pure i-component.

RF predicted the efficiency of the hypothetical composites (y_{pred}) against the removal of deoxynivalenol (DON), T2-toxin (T2), ochratoxin A (OTA), zearalenone (ZEN), fumonisin B1 (FB1), aflatoxin B1 (AFB1). We compared the prediction with the linear combination of efficiency of the pure components, being calculated as Σ_i $a_i y_i$, where y_i is the predicted target of the pure i-component. The gap between predictions and linear combination of pristine efficiency was assigned to the synergistic effect between components (Eq. (4)), being it positive or negative depending on the specific binder-toxin affinity.

$$\text{Synergy} = y_{pred} - \Sigma a_i y_i \quad (4)$$

In vivo validation. An in vivo trial was conducted with 8 healthy broiler chickens, females and males equally divided and about the same body weight (BW) at arrival (see supporting information, Supplementary Sect. S3 for details). The in vivo study was conducted at CER-Groupe, a GLP (Good Laboratory Practice) compliant test site (Marloie, Belgium). The animal study was approved

by the Ethical Committee of CER-Groupe (approval number CE/Sante/Residus-1). The animal experiment was conducted in compliance with Council Directive No. 2010/63/EU of 22 September 2010 on the protection of animals used for scientific purposes[44] as well as the ARRIVE guidelines. Randomization was performed at arrival, based on the sex and BW of the broiler chickens in such a way that 2 groups with 4 birds (2 males and 2 females) each were formed with about the same average BW/group. The feed was analyzed before administration by a multi-mycotoxin LC–MS/MS method (liquid chromatography-tandem mass spectrometry) and was found to contain low levels of DON (61.3 μg/kg) and OTA (1.6 μg/kg). These contamination levels were well within the acceptance criteria of the EU (2006/576/ EC)[2]. The treatment consisted of a single oral bolus administration with either DON or DON in combination with SEP/MONT/AC detoxifier (0.500 mg DON/kg BW, corresponding to the maximum EU guidance level of 5 mg/kg DON in feed, and 0.4 g detoxifier/kg BW, corresponding to an inclusion rate of 4 kg/t in the feed), administered as oral capsules directly in the crop and using a cross-over study design respecting a one-day wash-out period between treatments. Repetitive blood samples were taken from the *vena metatarsalis plantaris superficialis* (leg vein). The time points of blood sampling were 0 h (just before administration) and 0.08, 0.25, 0.5, 0.75, 1, 1.5, 2, 4, and 8 h (post administration, p.a.). The blood samples were centrifuged within 2 h after collection. Both DON and its major phase II metabolite DON-sulphate (DON-S) were analysed as appropriate biomarkers for exposure by UHPLC-MS/MS[45]. For DON-S, a qualitative UHPLC-MS/MS analysis was performed, i.e. chromatographic response or peak area ratio of DON-S/internal standard. Toxicokinetic modeling of the chromatographic response-time profiles of DON-S was done by non-compartmental toxicokinetic analysis (details in Supplementary Sect. S3). Following parameters were calculated: area under the response-time curve from time zero to infinite ($AUC_{0-\infty}$), maximal DON-S response in plasma (C_{max}), time at maximal plasma response (T_{max}), elimination half-life ($T_{1/2e}$) and elimination rate constant (k_e). The effect of the detoxifier on the oral absorption of DON was evaluated by comparing major toxicokinetic parameters between the

DON and DON+detoxifier treated broiler chickens, with special emphasis on $AUC_{0-\infty}$, Cmax and Tmax. Statistical analysis was performed with to evaluate possible significant differences between the DON and DON + detoxifier treatment. *P* values < 0.05 were considered significant. Moreover, the relative oral bioavailability (($AUC_{0-\infty}$ mycotoxin + detoxifier/$AUC_{0-\infty}$ mycotoxin)*100) was evaluated as marker for efficacy of the detoxifier. In general, two treatments are considered bioequivalent or thus not different from one another if the 90% confidence interval (CI) of the ratio of a log-transformed exposure measure (AUC) falls completely within the range 80–125%, as it is assumed that differences in exposure up to 20% are not relevant. If the CI falls completely out this specified range the treatments are considered not bioequivalent, and hence a significant effect of the detoxifier in reduction of systemic exposure can be concluded.

4.5. References

(1) Vila-Donat, P.; Marín, S.; Sanchis, V.; Ramos, A. J. A Review of the Mycotoxin Adsorbing Agents, with an Emphasis on Their Multi-Binding Capacity, for Animal Feed Decontamination. *Food Chem. Toxicol.* **2018**, *114*, 246–259. https://doi.org/10.1016/j.fct.2018.02.044.

(2) Aiko, V.; Mehta, A. Occurrence, Detection and Detoxification of Mycotoxins. *J. Biosci.* **2015**, *40* (5), 943–954. https://doi.org/10.1007/s12038-015-9569-6.

(3) Battilani, P.; Toscano, P.; Van Der Fels-Klerx, H. J.; Moretti, A.; Camardo Leggieri, M.; Brera, C.; Rortais, A.; Goumperis, T.; Robinson, T. Aflatoxin B 1 Contamination in Maize in Europe Increases Due to Climate Change. *Sci. Rep.* **2016**, *6* (April), 1–7. https://doi.org/10.1038/srep24328.

(4) Çelik, K. The Efficacy of Mycotoxin-Detoxifying and Biotransforming Agents in Animal Nutrition. *Nanomycotoxicology Treat. Mycotoxins Nano W.* **2019**, 271–284. https://doi.org/10.1016/B978-0-12-817998-7.00012-4.

(5) Pirouz, A. A.; Selamat, J.; Iqbal, S. Z.; Mirhosseini, H.; Karjiban, R. A.; Bakar, F. A. The Use of Innovative and Efficient Nanocomposite (Magnetic Graphene Oxide) for the Reduction on of Fusarium Mycotoxins in Palm Kernel Cake. *Sci. Rep.* **2017**, *7* (1), 1–9. https://doi.org/10.1038/s41598-017-12341-3.

(6) Marin, S.; Ramos, A. J.; Cano-Sancho, G.; Sanchis, V. Mycotoxins: Occurrence, Toxicology, and Exposure Assessment. *Food Chem. Toxicol.* **2013**, *60*, 218–237. https://doi.org/10.1016/j.fct.2013.07.047.

(7) Freire, L.; Guerreiro, T. M.; Pia, A. K. R.; Lima, E. O.; Oliveira, D. N.; Melo, C. F. O. R.; Catharino, R. R.; Sant'Ana, A. S. A Quantitative Study on Growth Variability and Production of Ochratoxin A and Its Derivatives by A. Carbonarius and A. Niger in Grape-Based Medium. *Sci. Rep.* **2018**, *8* (1), 1–11. https://doi.org/10.1038/s41598-018-32907-z.

(8) Antonissen, G.; Martel, A.; Pasmans, F.; Ducatelle, R.; Verbrugghe, E.; Vandenbroucke, V.;

Li, S.; Haesebrouck, F.; Van Immerseel, F.; Croubels, S. The Impact of Fusarium Mycotoxins on Human and Animal Host Susceptibility to Infectious Diseases. *Toxins (Basel).* **2014**, *6* (2), 430–452. https://doi.org/10.3390/toxins6020430.

(9) Alassane-Kpembi, I.; Gerez, J. R.; Cossalter, A. M.; Neves, M.; Laffitte, J.; Naylies, C.; Lippi, Y.; Kolf-Clauw, M.; Bracarense, A. P. L.; Pinton, P.; Oswald, I. P. Intestinal Toxicity of the Type B Trichothecene Mycotoxin Fusarenon-X: Whole Transcriptome Profiling Reveals New Signaling Pathways. *Sci. Rep.* **2017**, *7* (1), 1–14. https://doi.org/10.1038/s41598-017-07155-2.

(10) Wielogórska, E.; MacDonald, S.; Elliott, C. T. A Review of the Efficacy of Mycotoxin Detoxifying Agents Used in Feed in Light of Changing Global Environment and Legislation. *World Mycotoxin J.* **2016**, *9* (3), 419–433.

(11) European Union. Commission Recommendation 2006/579/EC of 17 August 2006 on the Presence of Deoxynivalenol, Zearalenone, Ochratoxin A, T-2 and HT-2 and Fumonisins in Products Intended for Animal Feeding. *Off. J. Eur. Union L 229/7* **2006**, *L 229/7*, 7–9.

(12) European Union. Commission Regulation (EC) No 1881/2006 of 19 December 2006 Setting Maximum Levels for Certain Contaminants in Foodstuffs. *Off. J. Eur. Union* **2006**, *L 364*, 5–24.

(13) Dohlman, E. Mycotoxin Hazards and Regulations. *Int. trade food Saf. Econ. theory case Stud.* **2003**, 97.

(14) MarketWatch. *Feed Mycotoxin Binders and Modifiers Market Size to Surpass US$ 3.21 Billion by 2027*; 2020.

(15) Li, P.; Su, R.; Yin, R.; Lai, D.; Wang, M.; Liu, Y.; Zhou, L. Detoxification of Mycotoxins through Biotransformation. *Toxins (Basel).* **2020**, *12* (2), 1–37. https://doi.org/10.3390/toxins12020121.

(16) Gzyl-Malcher, B.; Rudolphi-Skórska, E.; Sieprawska, A.; Filek, M. Manganese Protects Wheat from the Mycotoxin Zearalenone and Its Derivatives. *Sci. Rep.* **2019**, *9* (1), 1–12.

https://doi.org/10.1038/s41598-019-50664-5.

(17) Huwig, A.; Freimund, S.; Käppeli, O.; Dutler, H. Mycotoxin Detoxication of Animal Feed by Different Adsorbents. *Toxicol. Lett.* **2001**, *122* (2), 179–188. https://doi.org/10.1016/S0378-4274(01)00360-5.

(18) Di Gregorio, M. C.; Neeff, D. V. De; Jager, A. V.; Corassin, C. H.; Carão, Á. C. D. P.; Albuquerque, R. De; Azevedo, A. C. De; Oliveira, C. A. F. Mineral Adsorbents for Prevention of Mycotoxins in Animal Feeds. *Toxin Rev.* **2014**, *33* (3), 125–135. https://doi.org/10.3109/15569543.2014.905604.

(19) Mitchell, N. J.; Xue, K. S.; Lin, S.; Marroquin-Cardona, A.; Brown, K. A.; Elmore, S. E.; Tang, L.; Romoser, A.; Gelderblom, W. C. A.; Wang, J. S.; Phillips, T. D. Calcium Montmorillonite Clay Reduces AFB1 and FB1 Biomarkers in Rats Exposed to Single and Co-Exposures of Aflatoxin and Fumonisin. *J. Appl. Toxicol.* **2014**, *34* (7), 795–804. https://doi.org/10.1002/jat.2942.

(20) Pappas, A. C.; Tsiplakou, E.; Georgiadou, M.; Anagnostopoulos, C.; Markoglou, A. N.; Liapis, K.; Zervas, G. Bentonite Binders in the Presence of Mycotoxins: Results of in Vitro Preliminary Tests and an in Vivo Broiler Trial. *Appl. Clay Sci.* **2014**, *99*, 48–53. https://doi.org/10.1016/j.clay.2014.06.009.

(21) Sohling, U.; Haimerl, A. Use of Stevensite for Mycotoxin Adsorption. 2008.

(22) Schall, N.; Simmler-hibenthal, H.; Feldhaus, H. G. Mycotoxin Adsorbents. 2004.

(23) Devreese, M.; Antonissen, G.; De Backer, P.; Croubels, S. Efficacy of Active Carbon towards the Absorption of Deoxynivalenol in Pigs. *Toxins (Basel).* **2014**, *6* (10), 2998–3004. https://doi.org/10.3390/toxins6102998.

(24) Kihal, A.; Rodriguez-Prado, M.; Godoy, C.; Cristofol, C.; Calsamiglia, S. In Vitro Assessment of the Capacity of Certain Mycotoxin Binders to Adsorb Some Amino Acids and Water-Soluble Vitamins. *J. Dairy Sci.* **2020**, *103* (4), 3125–3132. https://doi.org/10.3168/jds.2019-17561.

(25) Fraeyman, S.; Croubels, S.; Devreese, M.; Antonissen, G. Emerging Fusarium and Alternaria Mycotoxins: Occurrence, Toxicity and Toxicokinetics. *Toxins (Basel).* **2017**, *9* (7), 1–26. https://doi.org/10.3390/toxins9070228.

(26) Gruber-Dorninger, C.; Novak, B.; Nagl, V.; Berthiller, F. Emerging Mycotoxins: Beyond Traditionally Determined Food Contaminants. *J. Agric. Food Chem.* **2017**, *65* (33), 7052–7070. https://doi.org/10.1021/acs.jafc.6b03413.

(27) Aquilina, G.; Bories, G.; Brantom, P.; Chesson, A.; Cocconcelli, P. S.; de Knecht, J.; Dierick, N. A.; Gralak, M. A.; Gropp, J.; Halle, I.; Kroker, R.; Leng, L.; Lindgren, S.; Lundebye Haldorsen, A.-K.; Mantovani, A.; Mézes, M.; Derek, R.; Saarela, M. Statement on the Establishment of Guidelines for the Assessment of Additives from the Functional Group 'substances for Reduction of the Contamination of Feed by Mycotoxins".' *EFSA J.* **2010**, *8* (7), 1–8. https://doi.org/10.2903/j.efsa.2010.1693.

(28) Lauwers, M.; Croubels, S.; Letor, B.; Gougoulias, C.; Devreese, M. Biomarkers for Exposure as a Tool for Efficacy Testing of a Mycotoxin Detoxifier in Broiler Chickens and Pigs. *Toxins (Basel).* **2019**, *11* (4). https://doi.org/10.3390/toxins11040187.

(29) Devreese, M.; Osselaere, A.; Goossens, J.; Vandenbroucke, V.; De Baere, S.; Eeckhout, M.; De Backer, P.; Croubels, S. New Bolus Models for in Vivo Efficacy Testing of Mycotoxin-Detoxifying Agents in Relation to EFSA Guidelines, Assessed Using Deoxynivalenol in Broiler Chickens. *Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess.* **2012**, *29* (7), 1101–1107. https://doi.org/10.1080/19440049.2012.671788.

(30) Osselaere, A.; Devreese, M.; Watteyn, A.; Vandenbroucke, V.; Goossens, J.; Hautekiet, V.; Eeckhout, M.; De Saeger, S.; De Baere, S.; De Backer, P.; Croubels, S. Efficacy and Safety Testing of Mycotoxin-Detoxifying Agents in Broilers Following the European Food Safety Authority Guidelines. *Poult. Sci.* **2012**, *91* (8), 2046–2054. https://doi.org/10.3382/ps.2012-02245.

(31) Grenier, B.; Hackl, M.; Skalicky, S.; Thamhesl, M.; Moll, W. D.; Berrios, R.; Schatzmayr, G.;

Nagl, V. MicroRNAs in Porcine Uterus and Serum Are Affected by Zearalenone and Represent a New Target for Mycotoxin Biomarker Discovery. *Sci. Rep.* **2019**, *9* (1), 1–14. https://doi.org/10.1038/s41598-019-45784-x.

(32) Simon, C. M.; Kim, J.; Gomez-Gualdron, D. A.; Camp, J. S.; Chung, Y. G.; Martin, R. L.; Mercado, R.; Deem, M. W.; Gunter, D.; Haranczyk, M.; Sholl, D. S.; Snurr, R. Q.; Smit, B. The Materials Genome in Action: Identifying the Performance Limits for Methane Storage. *Energy Environ. Sci.* **2015**, *8* (4), 1190–1199. https://doi.org/10.1039/c4ee03515a.

(33) Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q. Computation-Ready, Experimental Metal-Organic Frameworks: A Tool to Enable High-Throughput Screening of Nanoporous Crystals. *Chem. Mater.* **2014**, *26* (21), 6185–6192. https://doi.org/10.1021/cm502594j.

(34) Moosavi, S. M.; Chidambaram, A.; Talirz, L.; Haranczyk, M.; Stylianou, K. C.; Smit, B. Capturing Chemical Intuition in Synthesis of Metal-Organic Frameworks. *Nat. Commun.* **2019**, *10* (1), 1–7. https://doi.org/10.1038/s41467-019-08483-9.

(35) Hubrecht, R. C.; Carter, E. The 3Rs and Humane Experimental Technique: Implementing Change. *Animals* **2019**, *9* (10), 1–10. https://doi.org/10.3390/ani9100754.

(36) Russell, W. M. S.; Burch, R. L. The Principles of Humane Experimental Technique. *Methuen Co Ltd. London, UK* **1959**.

(37) Iheshiulor, O. O. M.; Esonu, B. O.; Chuwuka, O. K.; Omede, A. A.; Okoli, I. C.; Ogbuewu, I. P. Effects of Mycotoxins in Animal Nutrition: A Review. *Asian J. Anim. Sci.* **2011**, *5* (1), 19–33. https://doi.org/10.3923/ajas.2011.19.33.

(38) Holanda, D. M.; Kim, S. W. Efficacy of Mycotoxin Detoxifiers on Health and Growth of Newly-Weaned Pigs under Chronic Dietary Challenge of Deoxynivalenol. *Toxins (Basel).* **2020**, *12* (5). https://doi.org/10.3390/toxins12050311.

(39) González-Arias, C. A.; Marín, S.; Sanchis, V.; Ramos, A. J. Mycotoxin Bioaccessibility/Absorption Assessment Using in Vitro Digestion Models: A Review. *World*

Mycotoxin J. **2013**, *6* (2), 167–184. https://doi.org/10.3920/WMJ2012.1521.

(40) Boudergue, C.; Burel, C.; Dragacci, S.; Favrot, M.; Fremy, J.; Massimi, C.; Prigent, P.; Debongnie, P.; Pussemier, L.; Boudra, H.; Morgavi, D.; Oswald, I.; Perez, A.; Avantaggiato, G. Review of Mycotoxin-detoxifying Agents Used as Feed Additives: Mode of Action, Efficacy and Feed/Food Safety. *EFSA Support. Publ.* **2017**, *6* (9). https://doi.org/10.2903/sp.efsa.2009.en-22.

(41) Schwartz-Zimmermann, H. E.; Fruhmann, P.; Dänicke, S.; Wiesenberger, G.; Caha, S.; Weber, J.; Berthiller, F. Metabolism of Deoxynivalenol and Deepoxy-Deoxynivalenol in Broiler Chickens, Pullets, Roosters and Turkeys. *Toxins (Basel).* **2015**, *7* (11), 4706–4729. https://doi.org/10.3390/toxins7114706.

(42) Wan, D.; Huang, L.; Pan, Y.; Wu, Q.; Chen, D.; Tao, Y.; Wang, X.; Liu, Z.; Li, J.; Wang, L.; Yuan, Z. Metabolism, Distribution, and Excretion of Deoxynivalenol with Combined Techniques of Radiotracing, High-Performance Liquid Chromatography Ion Trap Time-of-Flight Mass Spectrometry, and Online Radiometric Detection. *J. Agric. Food Chem.* **2014**, *62* (1), 288–296. https://doi.org/https://doi.org/10.1021/jf4047946.

(43) Devreese, M.; Antonissen, G.; Broekaert, N.; De Mil, T.; De Baere, S.; Vanhaecke, L.; De Backer, P.; Croubels, S. Toxicokinetic Study and Oral Bioavailability of Deoxynivalenol in Turkey Poults, and Comparative Biotransformation between Broilers and Turkeys. *World Mycotoxin J.* **2015**, *8* (4), 533–539. https://doi.org/https://doi.org/10.3920/WMJ2014.1843.

(44) Anonymous. Council Directive No. 2010/63/EU of 22 September 2010 on the Protection of Animals Used for Scientific Purposes. Official Journal of the European Union, L276/33-79 2010.

(45) Lauwers, M.; De Baere, S.; Letor, B.; Rychlik, M.; Croubels, S.; Devreese, M. Multi LC-MS/MS and LC-HRMS Methods for Determination of 24 Mycotoxins Including Major Matrices from Pigs and Broiler Chickens. *Toxin* **2019**, *11* (3), 171–201.

https://doi.org/10.3390/toxins11030171.

4.6. Additional information

Acknowledgements

The authors acknowledge the Community of Madrid for its support of this work through the project IND2018/ IND-9819. The Laboratory of Pharmacology and Toxicology is member of the Ghent University expertise centre MSsmall.

Author contributions

G.L.D. and M.H. carried out the analysis of the in vitro experiments and the computational work with assistance from V.C. S.C. performed the in vivo trial. All authors provided input on the manuscript.

Competing interests

G.L.D. and V.C. are employed by Tolsa Group, which is commercializing one of the high performing composite materials identified in this study under the market name ANS. S.C. and M.H. declare no potential conflict of interest.

Data availability. The dataset and the model codes that support the findings of this study are available in zenodo repository with the identifier https://doi.org/10.5281/zenodo.5793956.

Received: 13 July 2021; Accepted: 23 February 2022

4.7. Supporting information

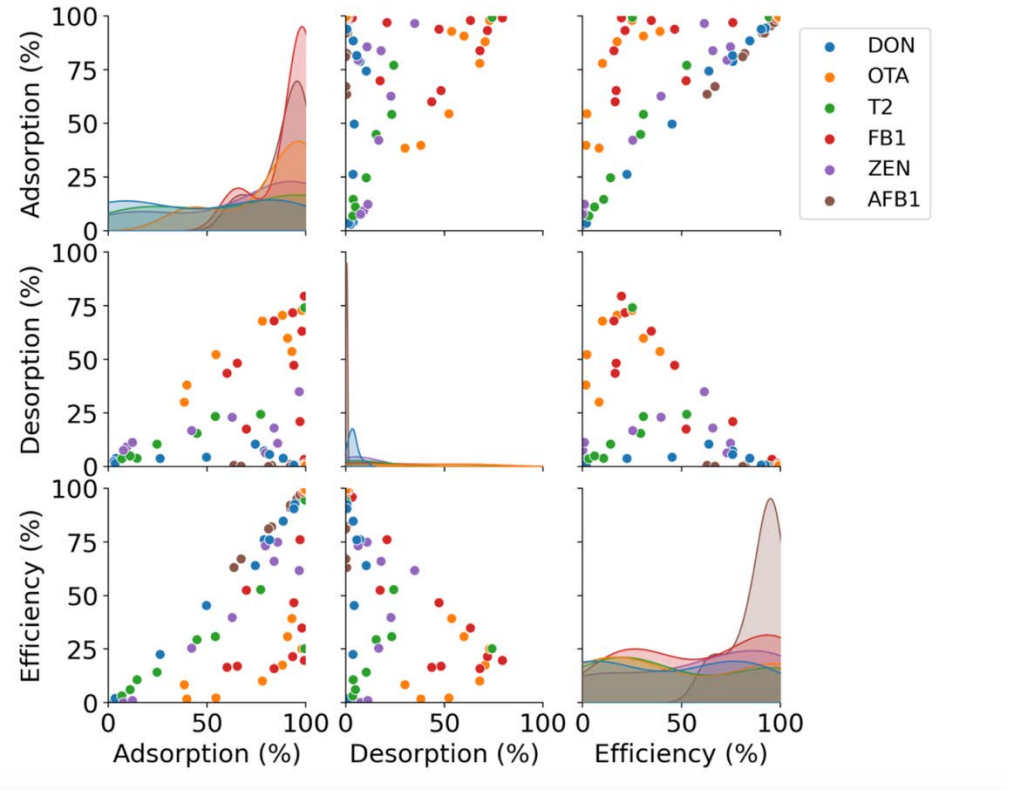


Figure S1. Pairwise relationships of adsorption, desorption, and efficiency in the dataset.

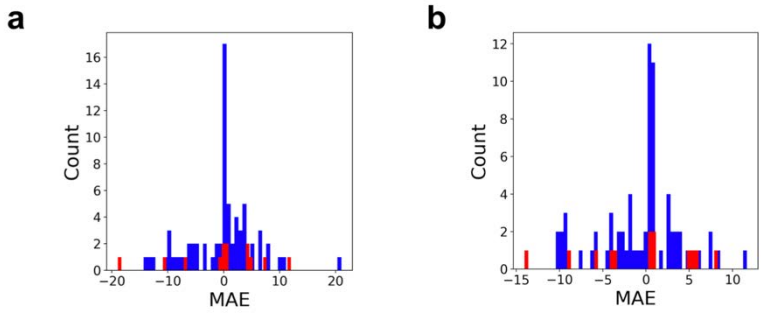


Figure S2. Mean absolute error distribution corresponding to the model predicting the adsorption (A) and efficiency (B).

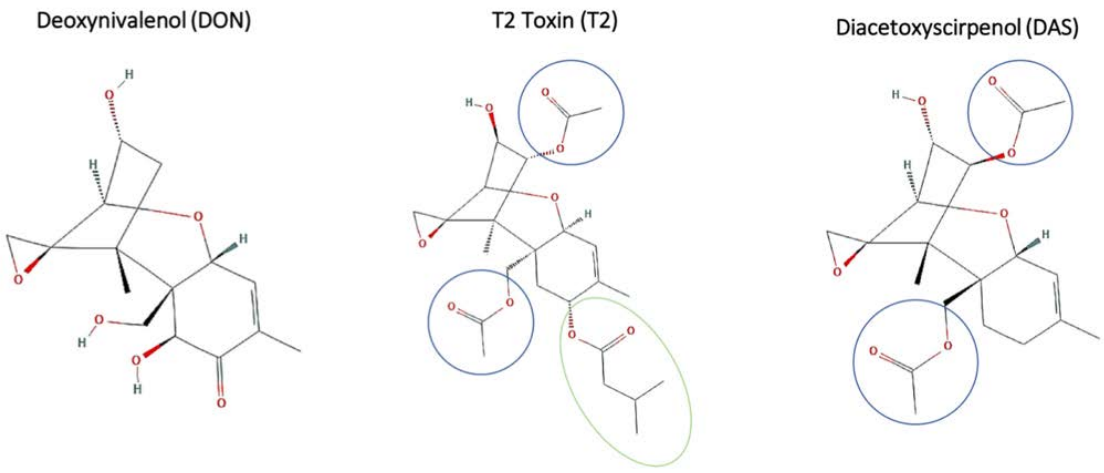


Figure S3. Chemical structures of three *Fusarium* toxins deoxynivalenol (DON), T-2 toxin (T2), and diacetoxyscirpenol (DAS). Blue circles highlight the carboxy groups, shared between T2 and DAS and absented in DON. Green circles highlight the methylbutanoate group present uniquely in T2 structure.

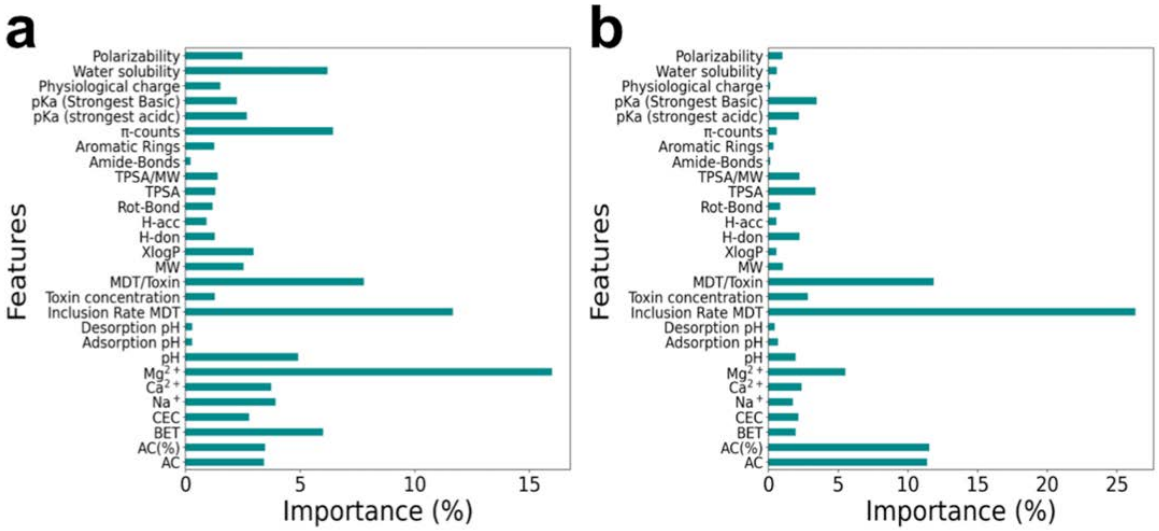


Figure S4. Feature importance score extracted by RFads (A) and RFeff (B).

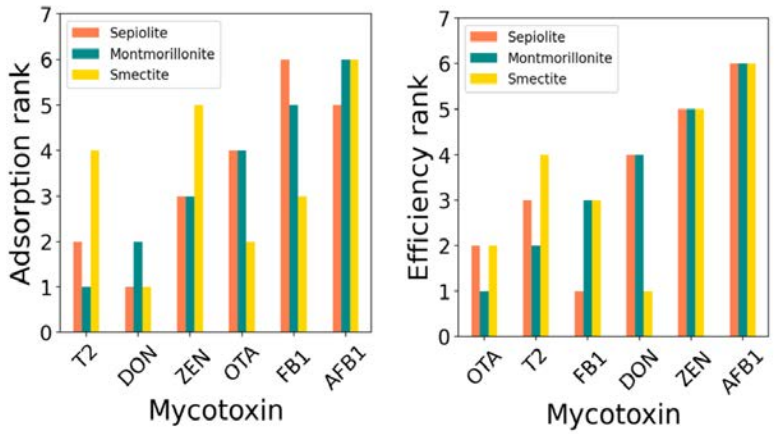


Figure S5. Performance ranking of sepiolite, Na-montmorillonite and Mg-rich smectite in terms of adsorption and efficiency. The experimental settings for the uptake of DON, OTA, T2, FUM and ZEN were fixed to 1kg/t of inclusion rate of MDT, 3 μg/ml of toxin concentration. The pH during the adsorption experiment was fixed to 3 while the desorption pH was 6.5. AFB1 uptake was assessed with 2kg/t of inclusion rate of MDT, 4 μg/ml of toxin concentration. The pH during both adsorption and desorption experiments was fixed to 5.

Table S1. The list of mycotoxins employed in the computer-aided towards multi-mycotoxin uptake.

Toxin	Cod.	Fungi species	Mycotoxin group
Deoxynivalenol	DON	*Fusarium* species	Trichothecenes
15-acetyl-deoxynivalenol	15ADON	*Fusarium* species	Trichothecenes
3-acetyl-deoxynivalenol	3ADON	*Fusarium* species	Trichothecenes
T2-toxin	T2	*Fusarium* species	Trichothecenes
T2-toxin tetraol	T2T	*Fusarium* species	Trichothecenes
HT2-toxin	HT2	*Fusarium* species	Trichothecenes
Diacetoxyscirpenol	DAS	*Fusarium* species	Trichothecenes
Aflatoxin B1	AFB1	*Aspergillus* species	Aflatoxins
Aflatoxin B2	AFB2	*Aspergillus* species	Aflatoxins
Aflatoxin G1	AFG1	*Aspergillus* species	Aflatoxins
Aflatoxin G2	AFG2	*Aspergillus* species	Aflatoxins
Aflatoxin M1	AFM1	*Aspergillus* species	Aflatoxins
Aflatoxin M2	AFM2	*Aspergillus* species	Aflatoxins
Fumonisin B1	FB1	*Fusarium* species	Fumonisins
Fumonisin B2	FB2	*Fusarium* species	Fumonisins
Fumonisin B3	FB3	*Fusarium* species	Fumonisins
Zearalenone	ZEN	*Fusarium* species	Zearalenone
alfa-zearalenol	AZEL	*Fusarium* species	Zearalenone
beta-zearalenol	BZEL	*Fusarium* species	Zearalenone
Ochratoxin A	OTA	*Aspergillus* species	Ochratoxins
Ochratoxin B	OTB	*Aspergillus* species	Ochratoxins
Ochratoxin C	OTC	*Aspergillus* species	Ochratoxins

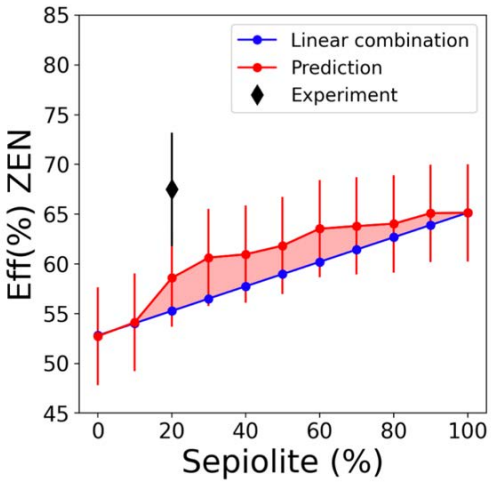


Figure S6. RF-based synergy-capturing assessment of efficiency in a series of hybrid SEP/MONT materials. The red area is assigned to the positive synergistic effect. The experimental settings for the uptake of DON, OTA, T2, FB1 and ZEN were fixed to 2 kg/t of inclusion rate of MDT, 2 µg/ml of toxin concentration. The pH during the adsorption experiment was fixed to 3 while the desorption pH was 6.5. The main values and the MAE are given.

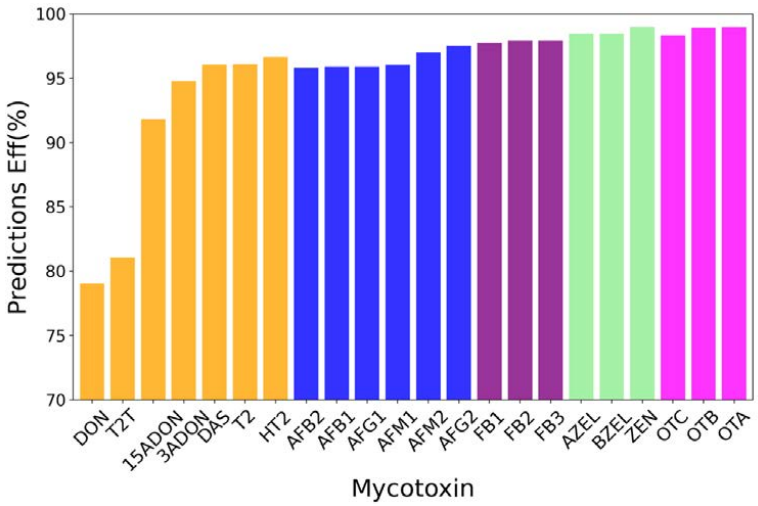


Figure S7. Efficiency prediction of a list of regulated and yet un-regulated mycotoxins. In orange the group of trichothecenes, aflatoxins in blue, fumonisins in purple, zearalenone in green and ochratoxins in magenta. The experimental setting was fixed to 2kg/t of inclusion rate of MDT, 2 µg/ml of toxin concentration. The pH during the adsorption experiment was fixed to 3 while the desorption pH was 6.5.

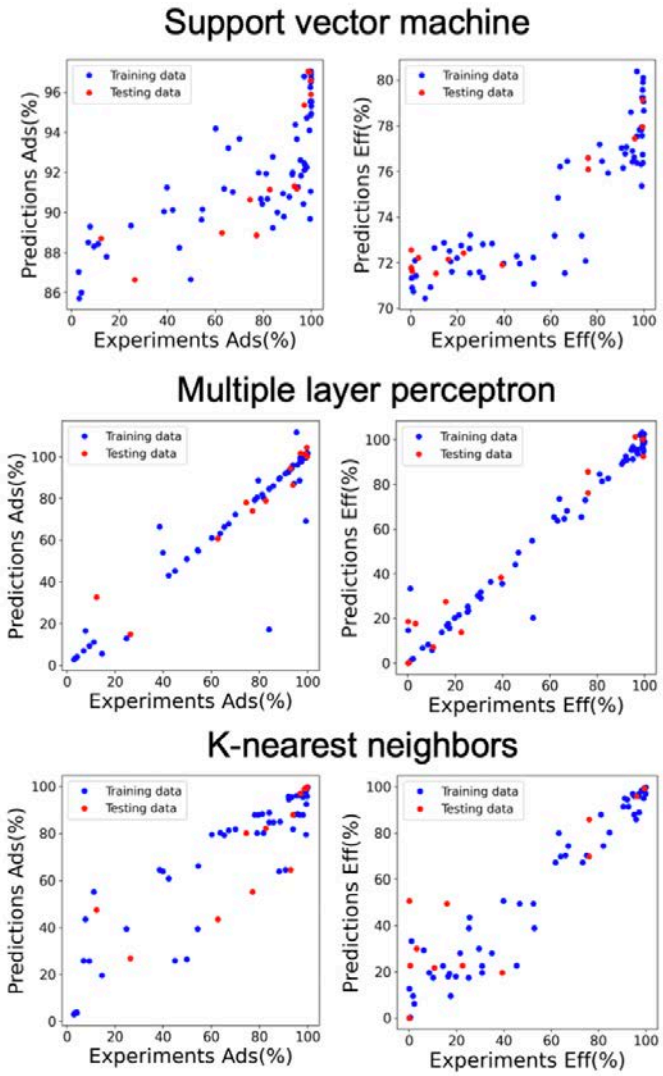


Figure S8. Graphical assessment of support vector machine, multiple layer perceptron and k-nearest neighbors predicting Ads(%) and Eff(%).

Table S2. Optimized hyperparameters of RFads and RFeff after cross-validation with K=5.

Hyperparameter	RFads	RFeff
n_estimators	2000	2000
max_depth	300	300
max_features	9	12
min_samples_leaf	1	1
min_samples_split	2	2

Table S3. Assessment scores of RFads and RFeff.

Assessment	RFads	RFeff
R^2 (testset)	0.91	0.97
R^2 (trainset, K=5)	0.71±0.054	0.84±0.08
MAE (testset)	5.5	4.8
MSE (testset)	60	38

Table S4. Model assessment of support vector machine, multiple layer perceptron y k-nearest neighbors and Random forest regressor predicting Ads(%) and Eff(%).

Score	Support vector machine	Multiple layer perceptron	K-nearest neighbors	Random forest regressor
Target	Ads(%)			
p-value	0.284	<0.0005	0.007	
R^2	0.66	0.87	0.7	0.91
MAE	12.6	7	9.1	5.5
Accuracy	0.39	0.46	0.54	0.69
Target	Eff(%)			
p-value	0.372	<0.0005	0.007	
R^2	0.84	0.95	0.73	0.97
MAE	10.34	6.2	13.9	4.8
Accuracy	0.46	0.69	0.38	0.69

[a] The p-value is calculated by 5x2cv paired t-test comparing each model-B (Support vector machine, multiple layer perceptron and k-nearest neighbors) against Random Forest regressor. The null hypothesis is that Random Forest and model-B have equal performance ($\alpha = 0.05$).

S1. Experimental material characterization

Specific surface area was estimated by Brunauer–Emmett–Teller (BET) theory on the adsorption branch of the physisorption isotherms. The isotherms were measured in a Micromeritics Gemini V surface area and pore size analyzer under N2 gas flow at 77K. The samples were previously degassed at 124°C for 18h. The pH of a 10 wt% suspension of raw clay in distilled water was

measured with Crison Basic 20 pH-meter, after stirring for 10 minutes at 25°C. The cation exchange capacity (CEC) was estimated quantifying the concentration of colored Cu^{2+}-triethylenetetramine complex by spectrophotometry. The exchange solution (1 l) was prepared dissolving in deionized water 1.45 ml of triethylenetetramine (97 %, Sigma-Aldrich) and 100 ml of copper sulfate standard solution 0.1 M (Sigma Aldrich). The exchange solution (30 ml) was added to 0.15g of clay sample stirring vigorously the suspension by vortex agitator until there are no more visible agglomerations detectable. The suspension was then centrifuged for 5 min at 4.500 min^{-1} and supernatant solution measured using a UV spectrophotometer (Agilent Technologies, Cary 60 UV-vis) at 580 nm. The cations leached by the clay structure (Na^{+}, Ca^{2+}, Mg^{+2}), in the supernatant solution, were evaluated by chemical analysis with ICP-OES Varian, Agilent 730.

The *in vitro* experiments are performed by Trilogy laboratory validated method modelling the adsorption and desorption through the gastrointestinal tract. Firstly, a specific amount of binder with the appropriate amount of mycotoxin were combined in phosphate buffers adjusted to the required pH, incubating the mixture at 37°C for 3 h under gently stirring. The solution was then centrifugated at 8 RPM separating the complexed binder-mycotoxin from the free toxin which is analyzed by HPLC-UV or HPLC-Fluorescence. The adsorption is calculated against the amount of toxin in the standard with the following equation 1:

$$Ads(\%) = 100 - (A_{ft}/A_{st} * 100) \qquad \text{Eq. 1}$$

Where A_{ft} and A_{st} are the amount of toxin in the supernatant solution and in the standard, respectively. The separated binder-mycotoxin complex was then taken through a desorption step. A buffered aqueous solution of neutral pH is added to the complex keeping the solution under stirring at 37°C for 3 hours. After incubation, the solution was centrifugated removing the remaining binder-mycotoxin complex from any toxin released from the binder. The supernatant

was the analyzed by HPLC-UV or HPLC-Fluorescence and the desorption calculated as follow (Eq.2):

$$Des(\%) = A_{rt}/A_{st}*100 \qquad \text{Eq. 2}$$

Where A_{rt} is the amount of toxin released by the binder. The efficiency is then calculated as:

$$Eff(\%) = Ads(\%) - Des(\%) \qquad \text{Eq. 3}$$

The starting amount of mycotoxin (toxin concentration) and binder (inclusion rate) used for filling the databases, as well as the adsorption and desorption pH are selected among the ranges shown in Table S5.

Table S5. Experimental setting implemented to test the *in vitro* adsorption and desorption of any mycotoxin-binder contained in the databases.

Parameter	Ranges
Adsorption pH	3-5
Desorption pH	6-7
Inclusion Rate (kg/t)	1-4
Toxin Concentration (µg/ml)	0.5-5

S2. Non-tree models architecture

Support vector machine and k-nearest neighbors were implemented using Scikit-learn library. Support vector machine was optimized on the base of k-fold cross validation finding the best hyperparameter, i.e., kernel='poly', gamma='scale', degree=3. Multiple layer perceptron (MLP) was implemented in TensorFlow and Keras framework. The network was built with 6 dense layers with ReLU activation function, 100 neurons in the hidden layers, Adaptive Moment Estimation (Adam)

as optimizer and 300 epochs. The performances were evaluated by R^2 score, mean absolute value (MAE), p-value and accuracy score. The latter is added to classify how often our regressors make predictions closed to the real value by considering the estimated experimental error as tolerance grade. The accuracy is defined as ratio of number of correct prediction and the total predictions corresponding to the testset size. The correct prediction was defined considering the estimated experimental error (5.7) and obtained when the absolute value of the difference between the predicted and real target is less the 5.7.

S3. *In vivo* trials.

An *in vivo* trial was conducted with 8 healthy broiler chickens, females and males equally divided, of the same breed (Ross 308, commercial supplier, Aye, Belgium), the same age (4 weeks at the beginning of the treatment) and about the same body weight (BW) at arrival. The *in vivo* study was conducted at CER-Groupe, a GLP (Good Laboratory Practice) compliant test site (Marloie, Belgium). The animal study was approved by the Ethical Committee of CER-Groupe. Randomization was performed at arrival, based on the sex and BW of the broiler chickens in such a way that 2 groups with 4 birds (2 males and 2 females) each were formed with about the same average BW/group. Both groups were housed in the same pen of 3 m2. The housing conditions were according to the EU guidelines and the Belgian guidelines[1,2]. The light schedule was a 18h/6h light/dark scheme. An acclimatization period of 8 days was respected (day 1-8). Throughout the acclimatization period, the chickens were allowed access to feed and water *ad libitum*. The chickens received commercial broiler chicken feed ACTI POUSSIN (batch 028590, SCAR Bullingen, Belgium). At least eight hours before the treatment (day 8), the feed was withdrawn, but water was available. Animals were fed again 4 h post administration. The feed was analysed by a multi-mycotoxin LC-MS/MS method (liquid chromatography-tandem mass spectrometry) and was found to contain low levels of DON (61.3 µg/kg) and OTA (1.6 µg/kg). These contamination

levels were well within the acceptance criteria of the EU (2006/576/EC)[3]. Analytical standard of DON was obtained from Fermentek (Jerusalem, Israel), and a stock solution was prepared in ethanol at a concentration of 10 mg/ml.

The treatment consisted of a single oral bolus administration with either DON or DON in combination with SEP/MONT/AC detoxifier (0.500 mg DON/kg BW, corresponding to the maximum EU guidance level of 5 mg/kg DON in feed, and 0.4 g detoxifier/kg BW, corresponding to an inclusion rate of 0.4% in the feed), administered as oral capsules directly in the crop and using a cross-over study design respecting a one-day wash-out period between treatments. The capsules contained cellulose and the appropriate volume of DON stock solution was added, whether or not combined with the detoxifier. Capsules were closed after 20 min in order to ensure evaporation of the solvent.

Repetitive blood samples (+/- 0.5 ml) were taken from the *vena metatarsalis plantaris superficialis* (leg vein). Intravenous (IV) catheters were placed in order to ensure continuous access to the blood vein. The time points of blood sampling were 0 h (just before administration) and 0.08, 0.25, 0.5, 0.75, 1, 1.5, 2, 4, and 8 h (post administration, p.a.). The blood samples were centrifuged within 2 hours after collection (±3,000 *g*, 10 min, 4°C). Plasma was stored at ≤-15°C until shipment to the Laboratory of Pharmacology and Toxicology for GLP-compliant analysis by UHPLC-MS/MS. Both DON and its major phase II metabolite DON-sulphate (DON-S) were analysed as appropriate biomarkers for exposure[4]. For DON-S, a qualitative UHPLC-MS/MS analysis was performed, i.e. chromatographic response or peak area ratio of DON-S/internal standard.

Toxicokinetic modeling of the chromatographic response-time profiles of DON-S was done by non- compartmental toxicokinetic analysis (WinNonlin 6.3, Pharsight Corporation, USA). Following parameters were calculated: area under the response-time curve from time zero to infinite (AUC0- ¥), maximal DON-S response in plasma (C_{max}), time at maximal plasma response (T_{max}), elimination half-life ($T_{1/2e}$) and elimination rate constant (k_e). The effect of the detoxifier on the oral absorption of DON was evaluated by comparing major toxicokinetic parameters

between the DON and DON+detoxifier treated broiler chickens, with special emphasis on AUC0-¥, Cmax and Tmax. Statistical analysis was performed with SPSS Statistics 26 to evaluate possible significant differences between the DON and DON+detoxifier treatment. P-values < 0.05 were considered significant. First, a Shapiro-Wilk test for normality was performed, including inspection of QQ-plots. If the data were not normally distributed, a log-transformation of the data was performed. If OK, a paired sample t-test was performed. Data for elimination half-life (T1/2e) were not normally distributed, even after log-transformation. Therefore, a nonparametric related-samples Wilcoxon signed rank test was performed.

Moreover, the relative oral bioavailability ((AUC0-¥ mycotoxin+detoxifier/ AUC0-¥ mycotoxin)*100) was evaluated as marker for efficacy of the detoxifier. In general, two treatment are considered bioequivalent or thus not different from one another if the 90% confidence interval (CI) of the ratio of a log-transformed exposure measure (AUC) falls completely within the range 80-125%, as it is assumed that differences in exposure up to 20% are not relevant. If the CI falls completely out this specified range the treatments are considered not bioequivalent, and hence a significant effect of the detoxifier in reduction of systemic exposure can be concluded.

References

1. (1) European Union. Commission Recommendation on Guidelines for the Accommodation and Care for Animals Used for Experimental and Other Scientific Purposes. *Off. J. Eur. Union* **2010**, *L276/33*.
2. (2) K.B. van 16 Januari 2006 Tot Vaststelling van de Nadere Regels van de Erkenningen, Toelatingen En Voorafgaande Registraties Afgeleverd Door Het Federaal Agentschap Voor de Veiligheid van de Voedselketen. Belgisch Staatsblad 2006.
3. (3) European Union. Commission Recommendation 2006/579/EC of 17 August 2006 on the Presence of Deoxynivalenol, Zearalenone, Ochratoxin A, T-2 and HT-2 and

Fumonisins in Products Intended for Animal Feeding. *Off. J. Eur. Union L 229/7* **2006**, *L 229/7*, 7–9.

4. (4) Lauwers, M.; De Baere, S.; Letor, B.; Rychlik, M.; Croubels, S.; Devreese, M. Multi LC-MS/MS and LC- HRMS Methods for Determination of 24 Mycotoxins Including Major Matrices from Pigs and Broiler Chickens. *Toxin* **2019**, *11* (3), 171–201. https://doi.org/10.3390/toxins11030171.

Chapter 5 - Data-driven experimental design of rheological clay-polymer composites

Data-driven experimental design of rheological clay-polymer composites

Giulia Lo Dico,[a,b,c] Borja Muñoz,[c] Verónica Carcelén,[c] and Maciej Haranczyk,[a]

[a]IMDEA Materials Institute, C/Eric Kandel 2, 28906 Getafe, Madrid, Spain.

[b]Department of Materials Science and Engineering, University of Carlos III of Madrid, Avda. de la Universidad, 30, 28911 Leganés, Madrid, Spain.

[c]Tolsa Group, Carretera de Madrid a Rivas Jarama, 35, 28041, Madrid, Spain.

KEYWORDS. Porous materials; Machine learning; Rheological agents.

Industrial Engineering and Chemical Research, 2022, 61, 31, 11455–11463

DOI: https://doi.org/10.1021/acs.iecr.2c00936

Chapter 6 - Toward the role of moisture in natural and thermally-treated clay materials

Toward the role of moisture in natural and thermally-treated clay materials

Giulia Lo Dico[a,b,c,d], Lorenzo Lisuzzo[d], Verónica Carcelén[c], Giuseppe Lazzara[d] and Maciej Haranczyk[a]

[a] IMDEA Materials Institute, C/Eric Kandel 2, 28906 Getafe, Madrid, Spain.

[b] Tolsa Group, Carretera de Madrid a Rivas Jarama, 35, Madrid, Spain.

[c] Department of Materials Science and Engineering, Universidad Carlos III de Madrid, Avda. de la Universidad, 30. 28911 Leganés, Madrid, Spain.

[d] Department of Physics and Chemistry, University of Palermo, Viale delle Scienze, pad. 17, 90128 Palermo, Italy.

Contribution of the international stay. Under revision by Microporous and Mesoporous Materials.

Highlights

- The Ea for the dehydration of micro and mesoporous clays was calculated by MTGA.
- The Ea is higher in fibrous clays compared to lamellar structures.
- The heating at 150° C decreases the Ea of Na-montmorillonite, stevensite, and sepiolite
- The heating at 150° C increases the Ea of palygorskite.
- The Ea is correlated to the adsorption of AFB1 and β-carotene with a Spearman score of -0.9 and -0.6.

Abstract. Clays are a class of porous materials which surfaces are naturally covered by moisture. Weak thermal treatment may be considered effective to remove the confined water molecules, changing the properties at the surfaces and make the micro and/or mesoporosities accessible to interact with other molecules. Herein, we report a modulated thermogravimetric analysis (MTGA) study of the behavior of the moisture on the structures of five, both fibrous and laminar, clays. The effect of the thermal treatment at 150° C, which provokes the release of the weakly adsorbed water molecules, was also investigated. From the single MTGA curve we calculated the activation energy for the removal of the adsorbed water (Ea) and found it higher in fibrous clays compared to lamellar structures. The thermal treatment enhances the rehydration in Na-montmorillonite, stevensite, and sepiolite structures with a decrease of the energy required to remove it, while Ea increases significantly in palygorskite. As a proof of concept, we statistically correlated the MTGA results, together with a full characterization of physico-chemical properties of the five clays to the adsorption of two molecules, i.e., aflatoxin B1 (AFB1) and β-carotene. The Ea was correlated with AFB1 adsorption with a Spearman score of -0.9. While, when the adsorbed water is forced to be removed, e.g., under vacuum conditions and high temperatures, the structure becomes the most important, decreasing the Spearman score between β-carotene and Ea to -0.6.

Keywords: Microporous clays; MTGA; Dehydration; Adsorption

Abbreviations: Activation energy=Ea; Activation energy of reconditioned samples=Ea (rec);zeta potential=ζ; Aflatoxin B1=AFB1; Inductively Coupled Plasma Optical Emission spectroscopy=ICP-OES; Specific surface area= SA; External surface area=ESA; Brunauer–Emmett–Teller theory=BET; Pore volume=Pore vol, Micropore content=Micro; Main pore size=MPS; Contact angle=θ; Scanning electron microscopy=SEM; High-performance liquid chromatography with ultraviolet detection=HPLC-UV; Amount of toxin in the supernatant solution= $A_{ft;}$ Amount of toxin in the standard solution=A_{st}; Spearman's coefficient=ϱ

6.1. Introduction.

Among porous materials, clays, being low-cost, biocompatible and environmentally friendly[1] are widely exploited as adsorbents of toxic molecules[2], organic pollutants[3] and pigment impurities[4–6], as well as nanocontainers for bioactive molecules[7,8] and nanocarriers for drug delivery systems[9]. The crystal structure of clays is formed by the repetition of one silicate tetrahedral layer and one octahedral layer (1:1) in materials such as kaolinite and halloysite, or by repetition of a layer consisting of an octahedral sheet sandwiched by two tetrahedral sheets (1:2). The latter gives rise to the 2:1 phyllosilicates which exhibit diverse morphologies, i.e., fibrous (such as sepiolites and palygorskites) or laminar (such as montmorillonites and stevensites).[1] The physico-chemical features are correlated to the size of clay particles, and their deposits. Fibrous clays, typically, possess higher surface area (ca. 180 m^2/g for palygorskites and ca. 300 m^2/g for sepiolites) compared to stevensites which exhibit ca. 70 m^2/g for the Al-rich and ca. 270 m^2/g for the Mg-rich.[10–13] Laminar clays also exhibit high cation exchange capacity (CEC) influencing their swelling properties.[14]

Natural moisture is present in clays as water is adsorbed on both external and internal surfaces during the crystal growth. The weakest-adsorbed water molecules are the externally adsorbed water molecules and the confined water in the micro and meso-porosities which are lost in a range of temperatures of 40-200ºC sepiolite and palygorskite or under vacuum treatment.[15,16] The presence of confined water molecules play a key role in the adsorption properties, determining the micro and/or mesoporosities accessibility to other molecules.[17,18] For instance, high moisture enhances the interlayer space of swelling montmorillonites changing their morphological properties and rendering the microporous geometrically exploitable to entrap organic molecules.[19] On the other hand, the presence of water layers on the clay micro and meso-porous may prevent fruitful interactions of the surfaces and the guest molecules.[20,21] In fact, the guest competes with water during the adsorption at the sorbent surface depending on its hydrophilic nature.[22] Weak thermal

treatment removes the moisture without significative structural modification, and it has been found to be effective to enhance the uptake of certain organic molecules.[21,23] The coordination water is responsible for the preservation of the crystal structure, and its loss at a temperature over 300ºC provokes a loss in crystallinity.[24,25] Thus, strong thermal treatment at high temperatures leads to profound changes in the clay morphological structure drastically reducing the accessible surface area, micro and mesoporous and, consequently, the adsorption properties.[23,26]

The implementation of vacuum pumping is a useful strategy to force the removal of confined water releasing the clay porosity and improving the loading rate[27]. Lisuzzo et al. investigated the water confinement within the cavity of halloysite by Knudsen thermogravimetry, describing the mechanism of halloysite loading.[28] Parallelly, modulated thermogravimetric analysis TGA measurement (MTGA) provides fast calculation of the activation energy (Ea) associated to the removal of the confined water.[29,30]

Beside to the role of the moisture, the absorption efficiency of clays strongly depends on the morphology, porosity and on the polar properties of the clays' surfaces. All of those properties tend to vary depending on natural environment and mines.[12,31] The main interactions between clay surface and target molecules are electrostatic interactions, strongly correlated with the clay zeta potential (ζ)[32]. In fact, the natural substitutions of the lattice structure and vacancies, charge the layer conferring to the clays the cation exchange capacity. Montmorillonite, sepiolite and stevensites switch their behavior along the pH with isoelectric points depending on the proper chemical structure and morphology[33–35]. Consequently, selective adsorption is possible by tuning the electrostatic forces between clay surface and the target molecule[36].

In this contribution we present a comprehensive study of the adsorbed water behavior for a set of 2:1 aluminosilicate structures of different morphology. We investigate the role of the moisture, in both natural and thermally-treated clay sources, by MTGA analysis calculating the kinetic

parameters associated to its removal from the clays' surfaces. As a proof of concept, the Ea as well as other physico-chemical properties were statistically correlated to the adsorption activity of natural clays in two diverse applications, i.e., mycotoxin aflatoxin B1 (AFB1) and β-carotene mitigation. We designed the adsorption experiments according to typical application requirements. Specifically, the adsorption test of AFB1 is designed by simulating the gastrointestinal tract of the animal, while the β-carotene is designed by employing a vacuum pump and high temperature.

6.2. Methods

Materials characterization. Five 2:1 aluminosilicates, i.e., fibrous sepiolite and palygorskite, and laminar Na-montmorillonites, Ca-montmorillonite and stevensites were provided by TOLSA S.A. and studied without previous treatment.

XRD diffractograms 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. The patterns were recorded between 2° and 70° (degree) at a goniometer speed of 0.5 s per step. Just in case to estimate the phyllosilicate content, the clay fraction (fine size material <45 μm) of the samples were separated by centrifugation and used as oriented aggregates for clay-mineral identification and semi-quantification. The calculate amount of each phyllosilicate were referred to the total content presents in each raw sample. Chemical analysis of metal oxides content was carried out with ICP-OES Varian, Agilent 730. The mass loss by calcination (referred as organic impurity) was estimated in a Heraeus M110 oven at 1000° C in air atmosphere. Morphological parameters were evaluated by physisorption isotherms, 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 and the morphological parameters were calculated in the adsorption branch of N_2 isotherm.[37–39] The latter are the specific surface area and external surface area (SA and ESA, respectively) by Brunauer–Emmett–Teller (BET) theory, the total pore volume (Pore vol), the micropore content (in %, Micro) and the main pore size (MPS) calculated

by the t-plot on the adsorption branch. The cation exchange capacity (CEC) was estimated by quantifying the concentration of colored Cu^{2+}-triethylenetetramine complex by spectrophotometry. In particular, the exchange solution (1 l) was prepared by dissolving in deionized water 1.45 ml of triethylenetetramine (97 %, Sigma-Aldrich) and 100 ml of copper sulfate standard solution 0.1 M (Sigma Aldrich). The exchange solution (30 ml) was added to 0.15g of clay sample stirring vigorously the suspension by vortex agitator until there were no more visible agglomerations detectable. The suspension was then centrifuged for 5 min at 4.500 rpm and supernatant solution measured using a UV spectrophotometer (Agilent Technologies, Cary 60 UV-vis) at 580 nm. The cations released by the clay structure (Na^{+}, Ca^{2+}, Mg^{+2}, K^{+}), in the supernatant solution, were evaluated by chemical analysis with inductively coupled plasma optical emission spectroscopy (ICP-OES) Varian, Agilent 730.

Zeta potential (ζ) experiments were carried out using a Zetasizer NANO-ZS (Malvern Instruments, London, UK) at 25 °C. Water contact angle tests were performed with an optical contact angle apparatus (OCA 20, Data Physics Instruments) equipped with a video measuring system having a high-resolution CCD camera and a high-performance digitizing adapter. Data acquisition was conducted by SCA 20 software (Data Physics Instruments). The contact angle (θ) of water in air was detected through the sessile drop method by placing a water droplet of 10 ± 0.5 mL onto the surface of the clay tablets. The measurements were conducted at 30.0 ± 0.1 °C. The morphology of the natural clays was evaluated by scanning electron microscopy (SEM) using a FE-SEM Apreo 2S LoVac- Field Emission Scanning Electron Microscope (FEG-SEM) without pretreating of the studied samples.

Thermal treatment and MTGA measurements. To investigate the role of the confined water at the micro and mesoporosities of the natural clays, the samples were treated by conditioning in desiccator under and the relative humidity was kept at 75% by saturated NaCl solution during 48h.

The samples were then dried at 150ºC during 24h to remove the adsorbed water and reconditioned under the same humidity conditions during 48h.

MTGA were performed with TA instrument, TGA 550. The modulate temperature setting was 5 °C for 200 s. The temperature ramp was 2 °C/min to 260°C. The MTGA is a quasi-isothermal TGA in which the temperature is rapidly modulated while its average within the time recording is kept constant.[29,30] The sinusoidal temperature oscillation of the MTGA enables to calculate the kinetics from a single experiment avoiding to run multiple TGA.[40]

The resulting curves were analyzed using the TA Instruments Universal Analysis 2000 software which implements discrete Fourier transformation to calculate the activation energy at the temperature of interest. The activation energy related to the evaporating process of the adsorbed water from the surfaces of both conditioned (Ea) and reconditioned (Ea (rec)) samples.

Adsorption experiments on AFB1 mycotoxin. Five re-conditioned clay after thermal treatment were tested in adsorption experiments of AFB1 mycotoxin. The latter were chosen to mimic the adsorption through the gastrointestinal tract.

In particular, 2 mg/g of clay were combined with 2 μg/ml of AFB1 mycotoxin dispersed in phosphate buffers adjusted to pH 5. The solution was incubated at 37°C for 3 h under gently stirring and centrifugated at 8 m/s for 10 minutes separating the complexed clay-mycotoxin from the free toxin. The amount of free toxin was quantified by high-performance liquid chromatography with ultraviolet detection (HPLC-UV). The adsorption is calculated against the amount of toxin in the standard according to Equation 1:

$$Ads(\%) = 100 - (A_{ft}/A_{st} * 100) \qquad \text{Eq. 1}$$

Where A_{ft} and A_{st} are the amount of toxin in the supernatant solution and in the standard, respectively.

Adsorption experiments of β-carotene. Five re-conditioned clay after thermal treatment and reconditioned as previously described, were tested in adsorption experiments of β-carotene. The clay (1%wt) is combined with 150 mg of crude palm oil in a glass batch (250 ml) and kept under stirring for 30 min with controlled pressure of 60 mbar and temperature of 95ºC. Then, the clay was removed by filtration over a Buchner funnel with Filterlab 1250 with Ø porous of 10-13 μm and the filtered oil was cooled at 70ºC and the amount of β-carotene was quantified using a Lovibond Tintometer Color Scale in a 0.25" (10 mm) glass cell. The amount of β-carotene was measured in both crude ($\beta\text{-carotene}_{st}$) and treated oil ($\beta\text{-carotene}_{free}$) samples, calculating the corresponding adsorbed amount according Eq. 2.

$$\beta\text{-carotene ads } (\%) = 100 - (\beta\text{-carotene}_{free} / \beta\text{-carotene}_{st} * 100) \qquad \text{Eq. 2}$$

Physico-chemical properties-adsorption correlation. To investigate the property-activity relationships, we calculate a correlation matrix based on Spearman's coefficient (ϱ). The latter returns a score which is related to a non-parametric statistical dependence between the ranking of two variables, avoiding assumptions about the type of correlation (e.g., linear, exponential). The score is calculated as follow:

a- Calculate the ranks for each property, i.e., ordering the property values corresponding to the five studied clays from the smallest to the greatest and assigning an increasing integer value.

b- Calculate the difference between ranks of each individual property (d).

c- Calculate the Spearman's score (ϱ) between each individual physico-chemical property (x) and the adsorption of AFB1 mycotoxin or β-carotene (y), reported in Table 3, by Eq. 3.

$$\varrho_{xy} = 1 - \frac{6\sum d_i^2}{5p6^2 - 1r} \qquad \text{Eq. 3}$$

The Spearman's coefficient ranges between -1 to 1. Values closed to 1 or -1 correspond to a positive or negative, respectively, monotonic correlation, while 0 relates to no-correlation between ranks. Together with ϱ, we considered the p-values to evaluate the statistical significance of the obtained correlation since our number of experiments is limited.

6.3. Results and discussion

Properties of natural clay minerals. The morphology of the fibrous (palygorskite and sepiolite) and planar (stevensite, Ca-montmorillonite, and Na-montmorillonite) clays were first observed by electron microscopy. Figure 1 includes the micrographs captured at 4 μm and the magnification at 500 nm of the corresponding white-selected area. The fibrous morphology of palygorskite and sepiolite was clearly observed (Fig 1A,B and 1C,D) with a fiber length of 0.5–4 μm and a thickness of ca. 30–50 nm. Figures 1E-J show the profiles of messily arranged lamellae whit a thickness of ca. 10 nm. Table 1 resumes the physico-chemical characterization of the five diverse clays. The purity grade was evaluated by semiquantitative analysis by XRD. The clays present a high purity grade with a phyllosilicate content at least of 80%. From chemical analysis montmorillonites possess at least of 17% of Al_2O_3 while stevensite and sepiolite the highest amount of MgO. The organic impurity was found be around 6-10% allowing use the natural mineral sources without initial purification treatment.

The BET surface areas of fibrous clays are higher than the laminar shaped minerals, while the natural montmorillonites exhibit the lowest surface area. Under the experimental condition in which the porosity of the clays is evaluated (see Methods section), the microporosity is that fraction of total microporosity that is mainly located at the edges of the crystallites[1,13]. While the presence

of mesopores and macropores, considered in the total pore volume and the main pore size, are assigned to the arrangement between fibers and plates themselves.[1,12] The cation exchange capacity (CEC) of the montmorillonites was higher that the Mg-rich clays. The layer charge and hence CEC and ζ potential are controlled by the vacancies in stevensite, that is trioctahedral, with Mg as the main octahedral cation. On the other hands, for montmorillonites, being them dioctahedral, both CEC and ζ potential are mainly controlled by the heterovalent substitutions in the structure.

The thermal treatment effect on the behavior of the adsorbed water. Natural moisture occupies the mesoporosity of clays, and the corresponding water molecules are lost with thermal treatment above 300ºC. We investigated the behavior of the adsorbed water recording the MTGA profiles up to 260ºC. Thermal treatment at 150ºC provokes the release of the weakly adsorbed moisture and is expected to modify the moisture capture capacity when the dried samples are reconditioned under controlled humidity with 75%wt saturated NaCl salt solution (details in Method section).

The mechanism of dehydration of clays depends on the on the morphology and structural characteristics, e.g., type of exchangeable cations, swelling properties[41–44]. Figure 2 contains the MTGA profiles for the five classes of 2:1 phyllosilicates. The interlayer space contains both weakly adsorbed water and water molecules strongly adsorbed to interlayer cations. The first loss at ca. 50°C is associated to partially removal of the moisture content and weakly/physisorbed water from the micro and mesoporous surfaces of clays. The second loss, occurring at higher temperature, is associated to the removal of water in the microporosity of tunnels in fibrous clay and interlayer space of stevensite and montmorillonites, strongly bound to exchangeable cations. Table 2 summarizes the main parameters extracted by the MTGA profiles. Despite no-difference in temperature of the first loss was observed after re-conditioning procedure, the moisture capture capacity was modified by the drying process. The second loss remains quite stable for non-swelling

clay minerals being them not affected by thermal treatment at 150ºC meaning that the water uptake in the microporosity is a reversible reaction which does not affect structural and morphological properties. On the contrary, the second loss disappears from the MTGA profile of swelling Na-montmorillonite when treated at 150ºC and reconditioned. The latter is probably correlated to the closing of its interlayer space when the micro-confined water is removed, making the process irreversible[42,44].

The Ea of conditioned fibrous clays was found higher than lamellar minerals suggesting that the morphological different clay particles influence the interaction of water and clay surface in terms of the energy required to activate the process. Indeed, the removal of weakly bound water from the mesoporosity of fibrous clays is a more energetically costly process compared to the same process occurring in lamellar clay minerals. The dehydration mechanisms is controlled by nucleation and growth, followed by a diffusion-controlled reaction.[50] The differences between the two morphologies could be most likely related to tortuous path created by the arrangement of fibers of sepiolite or palygorskite hinders the removal of the water molecules.[45–49]

After the thermal treatment is carried out, the moisture uptake is enhanced for Na-montmorillonite, stevensite and sepiolite as observed by the increase of first loss percentages in Table 2. More interestingly, the increase of moisture capture, after drying, is accompanied by a decrease of the energy required to remove the adsorbed water from the three clays and, more effectively, from sepiolite. As a consequence, these minerals can still adsorb more water molecules, but the interaction is weaker than pristine clays before thermal treatment.

On the other side, the activation energy values increased for Ca-montmorillonite and palygorskite after the drying process, despite no difference was observed in terms of moisture uptake efficiency, which remains quite stable. These results suggest that the thermal treatment promotes the adsorption of water molecules at the mineral surface making harder their removal, with particular emphasis for palygorskite.

Properties-activity correlation. One of the main applications of natural porous clays is the removal of certain organic molecules through adsorption on the clay surfaces. Here, the mitigation of β-carotene and AFB1 mycotoxin were chosen as example of practical application of clay-based adsorbent[51–53]. Two different experiments were performed reproducing the real adsorption process (details in Method section). The results are shown in Table 3. Palygorskite material removed the less amount of both molecules. Na-montmorillonite adsorbed more than 96% of AFB1, while stevensite was the most effective to mitigate the β-carotene content. We estimated the property-activity correlation by calculating the Spearman's coefficient (ϱ), shown in Figure 3 (see Method section for details). In particular, ϱ was calculated for each property listed in Table 1, together to the calculated Ea(rec), and the adsorption power of Table 3. Additionally, Figure 3 display the corresponding p-value for each pair of variables (green). Low ϱ of 0.3 with an associated p of 0.62 is shown between AFB1 mycotoxin and β-carotene. The no correlation reflects the different mechanisms involved including the different experimental setup. In fact, the adsorption experiment of β-carotene is performed at high temperature and under vacuum condition while AFB1 at room temperature and pressure. As expected, the adsorption of AFB1 suffers the presence of the confined water being it not removed by the temperature or the vacuum. In fact, the most correlated feature with AFB1 is the activation energy required for the desorption of the confined water in the reconditioned samples, with a Spearman's score of -0.9 ($p=0.037$). The negative value means that the AFB1 adsorption increases as Ea(rec) decreases, remarking the strong competition between AFB1 and water, independently on the proper clay structure. The latter also suggests that the thermal treatment deteriorates the capability of palygorskite in capturing AFB1, since increase its Ea(rec) of 72 kJ/mol, while is effective for sepiolite structures providing a decrease of 53 kJ/mol, hence favoring the AFB1 adsorption with respect to the conditioned no-treated sepiolite. Additionally, we conducted the AFB1 adsorption experiments on the untreated samples of trioctahedral group. We found an AFB1 adsorption of 93% for the raw stevensite and 50% for

sepiolite. The results remark that clays which Ea remains almost unmodified by the thermal treatment show the same adsorption capacity while clays like sepiolite take benefit from the thermal treatment which, reducing the Ea, facilitates the interaction between AFB1 and the clay surface.

On the other hand, during the β-carotene mitigation experiments the confined water is forced to be removed favoring the entrapping of the pigment on the clay surfaces. In this case we observed decreasing of the correlation score (ρ = -0.6, p=0.28) suggesting a faint competition between the guest molecule and water, being it removed by the process. The correlation between clay structure (amount of phyllosilicate) and β-carotene shows the highest Spearman score of 0.9 (p=0.037). This observation can be interpreted considering the vacuum pumping and the high temperature promoting a good interaction between β-carotene and the clay by exposing the clay surfaces from the water. Hence the adsorption in this case is mainly guided by the amount of active minerals. The same experiments were repeated using untreated samples and the results are show in Table S1. The observed differences with respect to the thermal treated materials remark that the clay moisture and the Ea does not influence significatively the adsorption of β-carotene due to the high temperature and the vacuum condition of the experimental process.

6.4. Conclusions.

We investigated the role of confined water in five diverse clay structures through modulated thermogravimetry analysis (MTGA). The calculated activation energy associated to the removal of the adsorbed water (Ea) was higher in fibrous clays compared to lamellar structures. The thermal treatment at 150ºC influenced the activation energy in different ways depending on the proper morphology and structure of the 2:1 phyllosilicates. In fact, water uptake of non-swelling clay was not significantly affected by the thermal treatment suggesting a reversible reaction which does not change the structural and morphological properties. On the contrary, swelling Na-montmorillonite closes its interlayer space when the micro-confined water is removed making the process

irreversible. The thermal treatment enhances the moisture uptake for Na-montmorillonite, stevensite and sepiolite with a decrease of the energy required to remove it, while Ea increases in palygorskite structures.

The structural, morphological and polar properties were correlated to the adsorption of two model organic molecules, aflatoxin B1 (AFB1) and β-carotene. The experiments were designed imitating the concrete application, i.e., additive for mycotoxin mitigation and oil refining. Ea is the most correlated feature with AFB1 adsorption with a Spearman score of -0.9 while showing a correlation of -0.6 with β-carotene adsorption which is mostly dependent on the phyllosilicate content. The presence of confined water on the clay surfaces prevents a good interaction with the target molecules, such as AFB1. When the adsorbed water is forced to be removed, for example under vacuum condition and high temperature, the structure and electrokinetical properties become the most correlated to the adsorption of organic molecule such as β-carotene.

Acknowledgements. The authors acknowledge the Community of Madrid for its support of this work through the project IND2018/IND-9819.

6.5. References

[1] F. Bergaya, B.K.G. Theng, G. Lagaly, Handbook of Clay Science, 2006. https://doi.org/10.1016/S1572-4352(05)01039-1.

[2] M.C. Di Gregorio, D.V. De Neeff, A.V. Jager, C.H. Corassin, Á.C.D.P. Carão, R. De Albuquerque, A.C. De Azevedo, C.A.F. Oliveira, Mineral adsorbents for prevention of mycotoxins in animal feeds, Toxin Rev. 33 (2014) 125–135. https://doi.org/10.3109/15569543.2014.905604.

[3] J.F. Lambert, Organic pollutant adsorption on clay minerals, in: Surf. Interface Chem. Clay Miner., 2018: pp. 195–253. https://doi.org/10.1016/B978-0-08-102432-4.00007-X.

[4] G. Lo Dico, F. Semilia, S. Milioto, F. Parisi, G. Cavallaro, G. Inguì, M. Makaremi, P.

Pasbakhsh, G. Lazzara, Microemulsion encapsulated into halloysite nanotubes and their applications for cleaning of a marble surface, Appl. Sci. 8 (2018) 1–10. https://doi.org/10.3390/app8091455.

[5] L. Rožić, S. Petrović, T. Novaković, β-Carotene removal from soybean oil with smectite clay using central composite design, Russ. J. Phys. Chem. A. 83 (2009) 1621–1624. https://doi.org/10.1134/S0036024409090374.

[6] G.K. Sarma, S. SenGupta, K.G. Bhattacharyya, Methylene blue adsorption on natural and modified clays, Sep. Sci. Technol. 46 (2011) 1602–1614. https://doi.org/10.1080/01496395.2011.565012.

[7] G. Lo Dico, B. Wicklein, L. Lisuzzo, G. Lazzara, P. Aranda, E. Ruiz-Hitzky, Multicomponent bionanocomposites based on clay nanoarchitectures for electrochemical devices, Beilstein J. Nanotechnol. 10 (2019) 1303–1315. https://doi.org/10.3762/bjnano.10.129.

[8] L. Lisuzzo, B. Wicklein, G. Lo Dico, G. Lazzara, G. Del Real, P. Aranda, E. Ruiz-Hitzky, Functional biohybrid materials based on halloysite, sepiolite and cellulose nanofibers for health applications, Dalt. Trans. 49 (2020) 3830–3840. https://doi.org/10.1039/c9dt03804c.

[9] M. Massaro, C.G. Colletti, G. Lazzara, S. Riela, The use of some clay minerals as natural resources for drug carrier applications, J. Funct. Biomater. 9 (2018) 1–22. https://doi.org/10.3390/jfb9040058.

[10] J.L. Martin de Vidales, M. Pozo, J.M. Alia, F. Garcia-Navarro, F. Rull, Kerolite-stevensite mixed-layers from the Madrid Basin, Central Spain, Clay Miner. 26 (1991) 329–342. https://doi.org/10.1180/claymin.1991.026.3.03.

[11] M. Suárez, J. García-Rivas, J. Morales, A. Lorenzo, A. García-Vicente, E. García-Romero, Review and new data on the surface properties of palygorskite: A comparative study, Appl. Clay Sci. 216 (2022). https://doi.org/10.1016/j.clay.2021.106311.

[12] M. Suárez, E. García-Romero, Variability of the surface properties of sepiolite, Appl. Clay Sci. 67–68 (2012) 72–82. https://doi.org/10.1016/j.clay.2012.06.003.

[13] S. Kaufhold, R. Dohrmann, M. Klinkenberg, S. Siegesmund, K. Ufer, N2-BET specific surface area of bentonites, J. Colloid Interface Sci. 349 (2010) 275–282. https://doi.org/10.1016/j.jcis.2010.05.018.

[14] S.L. Teich-McGoldrick, J.A. Greathouse, C.F. Jové-Colón, R.T. Cygan, Swelling Properties of Montmorillonite and Beidellite Clay Minerals from Molecular Simulation: Comparison of Temperature, Interlayer Cation, and Charge Location Effects, J. Phys. Chem. C. 119 (2015) 20880–20891. https://doi.org/10.1021/acs.jpcc.5b03253.

[15] F.X. Coudert, A. Boutin, A.H. Fuchs, Open questions on water confined in nanoporous materials, Commun. Chem. 4 (2021) 9–11. https://doi.org/10.1038/s42004-021-00544-9.

[16] J. Zhou, X. Lu, E.S. Boek, Confined water in tunnel nanopores of sepiolite: Insights from molecular simulations, Am. Mineral. 101 (2016) 713–718. https://doi.org/10.2138/am-2016-5430.

[17] R. Kecili, C.M. Hussain, Mechanism of adsorption on nanomaterials, Elsevier Inc., 2018. https://doi.org/10.1016/B978-0-12-812792-6.00004-2.

[18] H. Sadegh, G.A.M. Ali, V.K. Gupta, A.S.H. Makhlouf, R. Shahryari-ghoshekandi, M.N. Nadagouda, M. Sillanpää, E. Megiel, The role of nanomaterials as effective adsorbents and their applications in wastewater treatment, J. Nanostructure Chem. 7 (2017) 1–14. https://doi.org/10.1007/s40097-017-0219-4.

[19] J. Zhu, T. Wang, R. Zhu, F. Ge, P. Yuan, H. He, Expansion characteristics of organo montmorillonites during the intercalation, aging, drying and rehydration processes: Effect of surfactant/CEC ratio, Colloids Surfaces A Physicochem. Eng. Asp. 384 (2011) 401–407. https://doi.org/10.1016/j.colsurfa.2011.04.023.

[20] A. Ramírez, C. Sifuentes, F.S. Manciu, S. Komarneni, K.H. Pannell, R.R. Chianelli, The effect of Si/Al ratio and moisture on an organic/inorganic hybrid material:

Thioindigo/montmorillonite, Appl. Clay Sci. 51 (2011) 61–67. https://doi.org/10.1016/j.clay.2010.11.002.

[21] Ç. Çakany, H.C. Cabbar, Adsorption of p -xylene in dry and moist clay, 3 (2008) 29–36.

[22] E. Hunter-Sellars, J.J. Tee, I.P. Parkin, D.R. Williams, Adsorption of volatile organic compounds by industrial porous materials: Impact of relative humidity, Microporous Mesoporous Mater. 298 (2020) 110090. https://doi.org/10.1016/j.micromeso.2020.110090.

[23] S. Balci, Effect of heating and acid pre-treatment on pore size distribution of sepiolite, Clay Miner. 34 (1999) 647–655. https://doi.org/10.1180/000985599546406.

[24] N. Clauer, A.E. Fallick, E. Galán, M. Pozo, C. Taylor, Varied crystallization conditions for Neogene sepiolite and associated Mg-clays from Madrid Basin (Spain) traced by oxygen and hydrogen isotope geochemistry, Geochim. Cosmochim. Acta. 94 (2012) 181–198. https://doi.org/10.1016/j.gca.2012.07.016.

[25] Y. Sarıkaya, M. Önal, A.D. Pekdemir, Kinetic and thermodynamic approaches on thermal degradation of sepiolite crystal using XRD-analysis, J. Therm. Anal. Calorim. 140 (2020) 2667–2672. https://doi.org/10.1007/s10973-019-09053-3.

[26] M. Sarı Yılmaz, Y. Kalpaklı, S. Pişkin, Thermal behavior and dehydroxylation kinetics of naturally occurring sepiolite and bentonite, J. Therm. Anal. Calorim. 114 (2013) 1191–1199. https://doi.org/10.1007/s10973-013-3152-x.

[27] E. Abdullayev, Y. Lvov, Halloysite clay nanotubes for controlled release of protective agents, J. Nanosci. Nanotechnol. 11 (2011) 10007–10026. https://doi.org/10.1166/jnn.2011.5724.

[28] L. Lisuzzo, G. Cavallaro, P. Pasbakhsh, S. Milioto, G. Lazzara, Why does vacuum drive to the loading of halloysite nanotubes? The key role of water confinement, J. Colloid Interface Sci. 547 (2019) 361–369. https://doi.org/10.1016/j.jcis.2019.04.012.

[29] R.L.B.B.K. Hahn, Obtaining kinetic parameters from modulated thermogrvimetry, J.

Therm. Anal. (1998) 695–704.

[30] V. Mamleev, S. Bourbigot, Modulated thermogravimetry in analysis of decomposition kinetics, Chem. Eng. Sci. 60 (2005) 747–766. https://doi.org/10.1016/j.ces.2004.08.044.

[31] V. D'Ascanio, D. Greco, E. Menicagli, E. Santovito, L. Catucci, A.F. Logrieco, G. Avantaggiato, The role of geological origin of smectites and of their physico-chemical properties on aflatoxin adsorption, Appl. Clay Sci. 181 (2019) 105209. https://doi.org/10.1016/j.clay.2019.105209.

[32] D. Zadaka, A. Radian, Y.G. Mishael, Applying zeta potential measurements to characterize the adsorption on montmorillonite of organic cations as monomers, micelles, or polymers, J. Colloid Interface Sci. 352 (2010) 171–177. https://doi.org/10.1016/j.jcis.2010.08.010.

[33] P. Leroy, C. Tournassat, O. Bernard, N. Devau, M. Azaroual, The electrophoretic mobility of montmorillonite. Zeta potential and surface conductivity effects, J. Colloid Interface Sci. 451 (2015) 21–39. https://doi.org/10.1016/j.jcis.2015.03.047.

[34] E. Sabah, U. Mart, M. Çinar, M.S. Çelik, Zeta potentials of sepiolite suspensions in concentrated monovalent electrolytes, Sep. Sci. Technol. 42 (2007) 2275–2288. https://doi.org/10.1080/15275920701313616.

[35] E. Tombácz, M. Szekeres, Colloidal behavior of aqueous montmorillonite suspensions: The specific role of pH in the presence of indifferent electrolytes, Appl. Clay Sci. 27 (2004) 75–94. https://doi.org/10.1016/j.clay.2004.01.001.

[36] N. Tekin, A. Dinçer, Ö. Demirbaş, M. Alkan, Adsorption of cationic polyacrylamide onto sepiolite, J. Hazard. Mater. 134 (2006) 211–219. https://doi.org/10.1016/j.jhazmat.2005.11.005.

[37] M.F. De Lange, T.J.H. Vlugt, J. Gascon, F. Kapteijn, Adsorptive characterization of porous solids: Error analysis guides the way, Microporous Mesoporous Mater. 200 (2014) 199–215. https://doi.org/10.1016/j.micromeso.2014.08.048.

[38] M. Thommes, K. Kaneko, A. V Neimark, J.P. Olivier, F. Rodriguez-Reinoso, J. Rouquerol,

K.S.W. Sing, Physisorption of gases, with special reference to the evaluation of surface area and pore size distribution (IUPAC Technical Report), Pure Appl. Chem. 87 (2015) 1051–1069. https://doi.org/10.1515/pac-2014-1117.

[39] ISO9277:2010, Determination of the specific surface area of solids by gas adsorption — BET method, (2010).

[40] V. Mamleev, S. Bourbigot, Calculation of activation energies using the sinusoidally modulated temperature, J. Therm. Anal. Calorim. 70 (2002) 565–579. https://doi.org/10.1023/A.

[41] E. Rinnert, C. Carteret, B. Humbert, G. Fragneto-Cusani, J.D.F. Ramsay, A. Delville, J.L. Robert, I. Bihannic, M. Pelletier, L.J. Michot, Hydration of a synthetic clay with tetrahedral charges: A multidisciplinary experimental and numerical study, J. Phys. Chem. B. 109 (2005) 23745–23759. https://doi.org/10.1021/jp050957u.

[42] J.M. Cases, I. Bérend, G. Besson, M. Francois, J.P. Uriot, F. Thomas, J.E. Poirier, Mechanism of Adsorption and Desorption of Water Vapor by Homoionic Montmorillonite. 1. The Sodium-Exchanged Form, Langmuir. 8 (1992) 2730–2739. https://doi.org/10.1021/la00047a025.

[43] A. Derkowski, V.A. Drits, D.K. McCarty, Rehydration of dehydrated-dehydroxylated smectite in a low water vapor environment, Am. Mineral. 97 (2012) 110–127.

[44] I. Bérend, J.M. Cases, M. François, J.P. Uriot, L. Michot, A. Masion, F. Thomas, Mechanism of Adsorption and Desorption of Water Vapor by Homoionic Montmorillonites: 2. The Li+, Na+, K+, Rb+, and Cs+-exchanged Forms, Clays Clay Miner. 43 (1995) 324–336. https://doi.org/10.1346/CCMN.1995.0430307.

[45] A. Kuligiewicz, A. Derkowski, Tightly bound water in smectites, Am. Mineral. 102 (2017) 1073–1090. https://doi.org/10.2138/am-2017-5918.

[46] W. Kuang, G.A. Facey, C. Detellier, Dehydration and rehydration of palygorskite and the influence of water on the nanopores, Clays Clay Miner. 52 (2004) 635–642.

https://doi.org/10.1346/CCMN.2004.0520509.

[47] K. Bahranowski, A. Klimek, A. Gaweł, E.M. Serwicka, Rehydration driven na-activation of bentonite—evolution of the clay structure and composition, Materials (Basel). 14 (2021) 1–13. https://doi.org/10.3390/ma14247622.

[48] S. Stackhouse, P. V. Coveney, D.M. Benoit, Density-functional-theory-based study of the dehydroxylation behavior of aluminous dioctahedral 2:1 layer-type clay minerals, J. Phys. Chem. B. 108 (2004) 9685–9694. https://doi.org/10.1021/jp037608p.

[49] L.P. Ogorodova, I.A. Kiseleva, M.F. Vigasina, Y.K. Kabalov, R.O. Grishchenko, L. V. Mel'Chakova, Natural sepiolite: Enthalpies of dehydration, dehydroxylation, and formation derived from thermochemical studies, Am. Mineral. 99 (2014) 2369–2373. https://doi.org/10.2138/am-2014-4804.

[50] X. hui Zhang, C. He, L. Wang, Z. quan Li, M. Deng, J. Liu, H. kui Li, Q. Feng, Non-isothermal kinetic analysis of thermal decomposition of the Ca-bentonite from Santai, China, Mineral. Petrol. 109 (2015) 319–327. https://doi.org/10.1007/s00710-014-0331-9.

[51] Y. Li, G. Tian, L. Gong, B. Chen, L. Kong, J. Liang, Evaluation of natural sepiolite clay as adsorbents for aflatoxin B1: A comparative study, J. Environ. Chem. Eng. 8 (2020) 104052. https://doi.org/10.1016/j.jece.2020.104052.

[52] W.F. Jaynes, R.E. Zartman, W.H. Hudnall, Aflatoxin B1 adsorption by clays from water and corn meal, Appl. Clay Sci. 36 (2007) 197–205. https://doi.org/10.1016/j.clay.2006.06.012.

[53] K.C.M. Xavier, M.S.F. Santos, J.A. Osajima, A.B. Luz, M.G. Fonseca, E.C. Silva Filho, Thermally activated palygorskites as agents to clarify soybean oil, Appl. Clay Sci. 119 (2016) 338–347. https://doi.org/10.1016/j.clay.2015.10.037.

6.6. Figures and tables

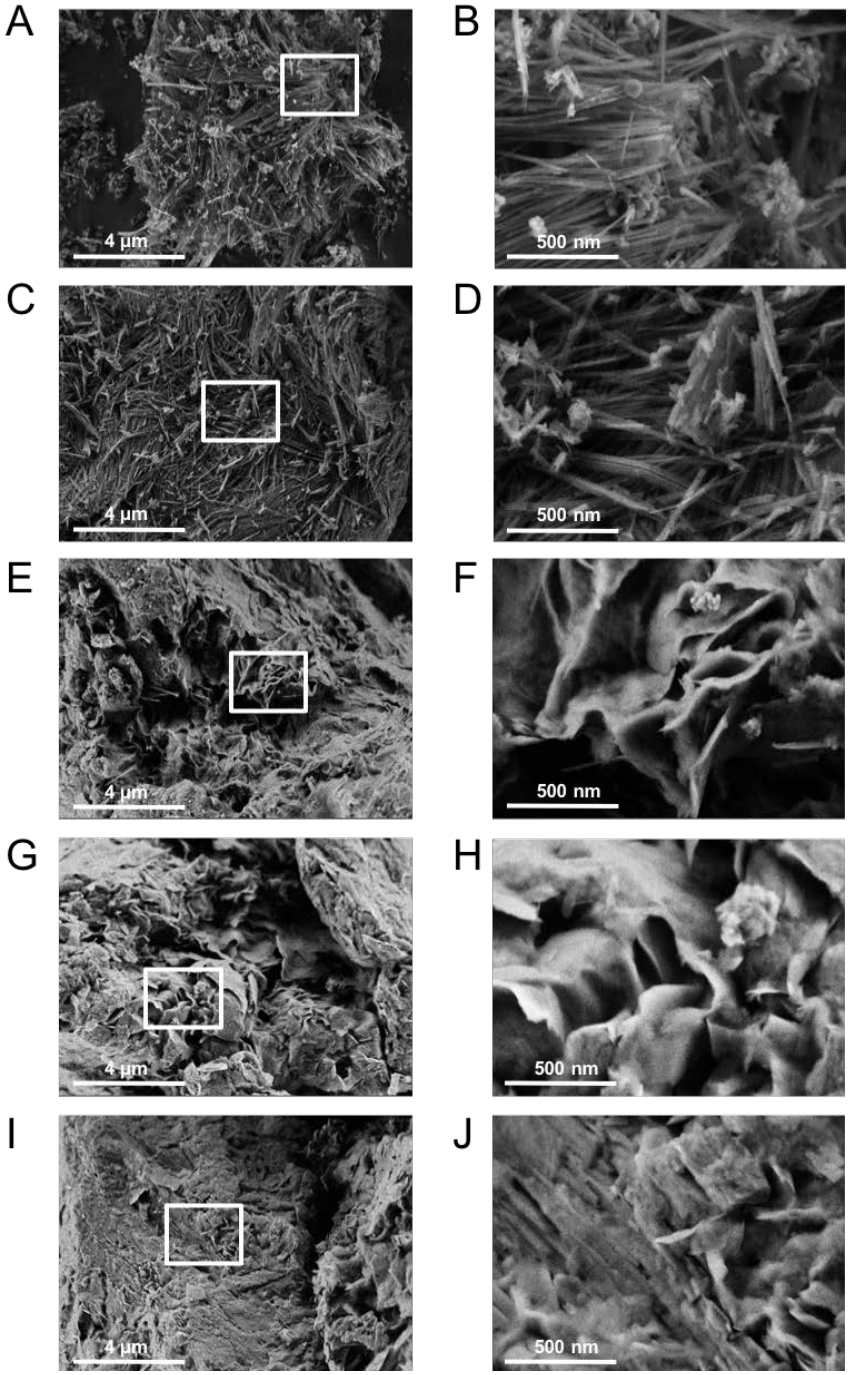


Figure 1. SEM micrographs of (A, B) natural palygorskite, (C, D) natural sepiolite, (E, F) natural stevensite, (G, H) Ca-montmorillonite (I, J) natural Na-montmorillonite.

Table 1. Physico-chemical characterization of the five selected natural clays.

Property	Palygorskite	Sepiolite	Stevensite	Na-montmorillonite	Ca-montmorillonite
Phyllosilicate [a]	81	96	82	88	90
Quartz[a]	3	0	1	1	0
SiO_2[a]	68.4	60.2	58.5	61.9	56.7
Al_2O_3[a]	8.2	2.0	5.4	19.3	17.3
MgO[a]	8.6	25.7	22.2	3.2	3.8
CaO[a]	1.5	1.0	1.5	1.4	4.7
Fe_2O_3[a]	3.7	0.5	1.5	4.3	5.2
Na_2O[a]	0.1	0.3	0.4	3.2	1.1
K_2O[a]	0.4	0.6	1.1	0.3	0.7
TiO_2[a]	0.3	0.1	0.2	0.2	0.5
Mn_2O_3[a]	0.0	0.0	0.0	0.0	0.1
Organic impurity [a]	8.7	10.1	9.0	6,1	9.3
SA (m^2/g)	150	310	248	45	63
ESA (m^2/g)	110	152	150	28	42
Pore vol (cm^3/g)	0.42	0.68	0.24	0.2	0.12
Micro (%)	6	12	20	5	4.7
Main pore size (Å)	97	87	37	102	113
CEC (meq/100g)	24	12	12	93	88
Na^+ (mmol/100 g)	1	1	1	81	3
Ca^{2+} (mmol/100 g)	9	7	7	17	67
Mg^{2+} (mmol/100 g)	8	4	4	13	10
K^+ (mmol/100 g)	1	1	1	1	1
Θ (degree)	18.4	15.6	38.1	41.8	8.3
ζ (mV)	-15	-13	-12	-26	-18

[a] The amount is calculated as %wt/wt with respect to the total (see Method section).

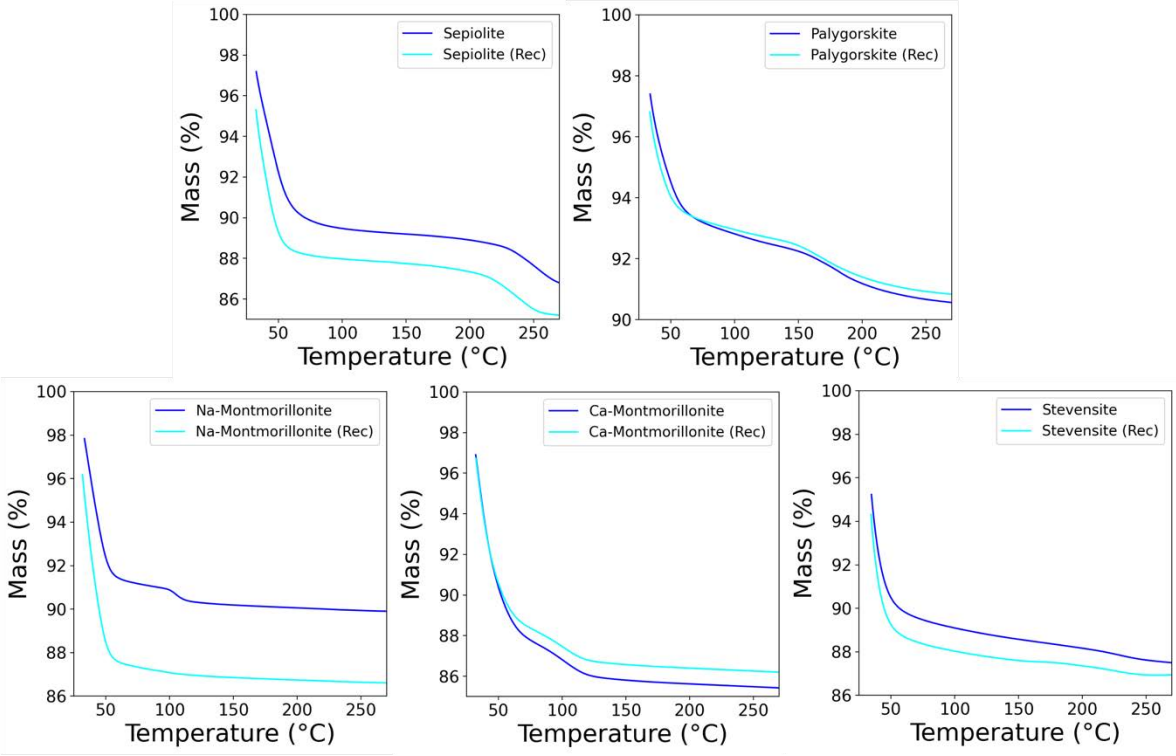


Figure 2. MTGA of conditioned sample under constant humidity (blue lines) and re-conditioned samples after drying overnight and 150ºC (cyan lines).

Table 2. Weight losses and activation energy calculated by MTGA for conditioned sample under constant humidity and re-conditioned (Rec) samples after drying overnight and 150ºC.

Clay mineral	Temperature (°C)	First loss (%)	Ea (kJ/mol)	ΔE (kJ/mol)
Na-Montmorillonite	45	8.9	100	-13
Na-Montmorillonite (Rec)		12.7	87	
Ca-Montmorillonite	39	12.3	82	8
Ca-Montmorillonite (Rec)		12	90	
Palygorskite	39	7.2	164	72
Palygorskite (Rec)		7	236	
Stevensite	38	11.7	85	-3
Stevensite (Rec)		12.3	82	
Sepiolite	47	10.9	187	-53
Sepiolite (Rec)		12.2	133	

Table 3. β-carotene and AFB1 adsorption results for the 5 class of minerals.

Clay minerals	β-carotene ads (%)[a]	AFB1 ads (%)[a]
Na-Montmorillonite (Rec)	31.5	97
Ca-Montmorillonite (Rec)	31.5	67
Palygorskite (Rec)	22	12
Stevensite (Rec)	35	91
Sepiolite (Rec)	32	63

[a] The amount is expressed as % of the adsorbed molecule with respect to the total starting amount (see Method section for details).

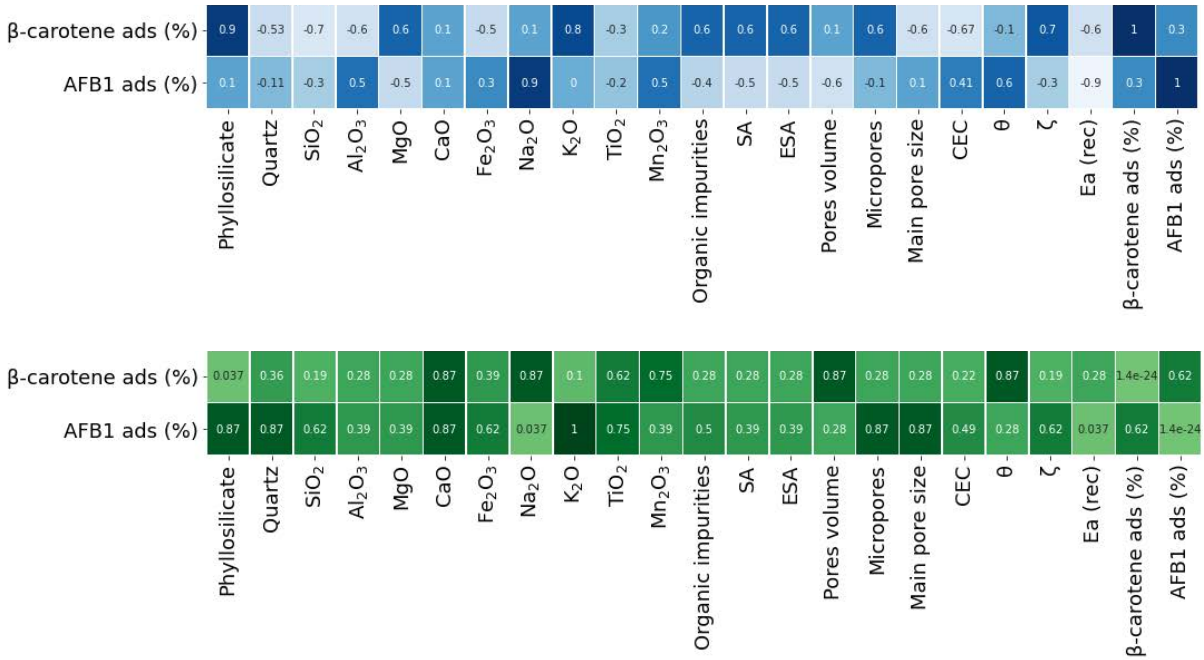


Figure 3. Spearman correlation coefficients (blue) and their associated p-values (green) between the adsorptive outcomes and the physico-chemical properties resumed in Table 1.

6.7. Supporting information

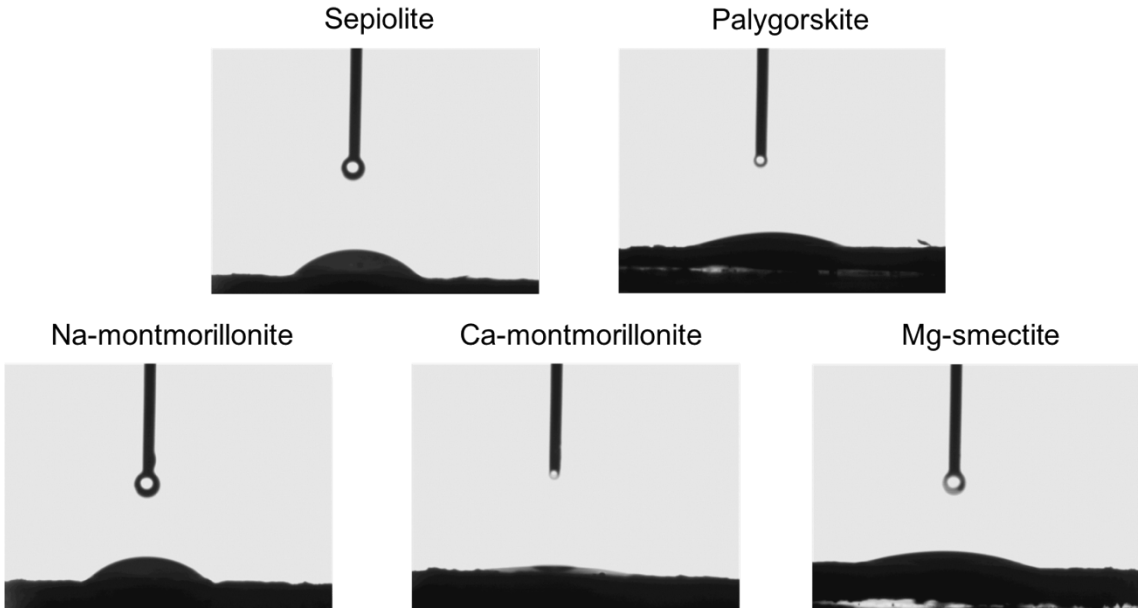


Figure S1. Images of the water droplets just after their deposition on the surface of the five clay samples.

Table S1. β-carotene and AFB1 adsorption results for the untreated 5 class of minerals.

Clay minerals	**β-carotene ads (%)[a]**
Na-Montmorillonite	30
Ca-Montmorillonite	33
Palygorskite	24
Smectite	35
Sepiolite	29

[a] The amount is expressed as % of the adsorbed molecule with respect to the total starting amount (see Method section for details).

Annex 1 - Synergistic effect in hybrid materials captured by machine learning

Synergistic effect in hybrid materials captured by machine learning

Giulia Lo Dico[a,b], and Maciej Haranczyk[a]

[a] IMDEA Materials Institute, C/Eric Kandel 2, 28906 Getafe, Madrid, Spain.

[b] Tolsa Group, Carretera de Madrid a Rivas Jarama, 35, Madrid, Spain.

In preparation manuscript

A.1. Introduction

Well-designed hybrid materials are known to be efficient materials exploiting a combination of the properties of single components. Commonly, the characteristics of an hybrids reside in between the original components and their final property, especially mechanical properties, can be predicted by the role-of-mixture (ROM).[1–4] The algorithm is a weighted mean which fit the physical behavior under certain assumptions. Despite ROM can approach a reasonable approximation of the hybrid properties, it fails when the components interact in such a way to create a new and unique physical, chemical, and/or biological properties not possessed by the original components.[5–7] The latter is associated to a positive or negative synergistic effect, defined as a positive or negative deviation of a certain mechanical property from the rule-of-mixtures behavior.[8–10] The synergistic effect is also exploited in different application such as adsorption and controlled release. For instance, active porous hybrid materials are known to be effective in the removal of a number of contaminants or combine the benefits of different components to synergize the outcome of drug delivery for therapeutic applications.[11–13]

The design of the effective hybrid materials is challenging due to the hard rationalization of the multi-parameters contributing to the final property. Additionally, the wide variability of physico-chemical features of natural materials, such as clay minerals, should be considered when those are incorporated in the hybrid formulation. In fact, depending on the morphology, chemical structure, cation exchange capacity (CEC), etc., clay minerals interact in different way with other compounds complicating the material design.[14,15]

Machine learning applied to material science is an emerging tool to replace classical approach, reduce the number of experiments needed in scientific investigation and extract physico-chemical information on the material behavior.[16–19] We have applied this methodology in two different field of applications of clay-based hybrid materials, i.e., mycotoxin detoxifier (MDT) and rheological

additive. [20,21] In particular, we have built and experimentally validated tree-based machine learning algorithms to predict the adsorption and viscosity of a high number of proposed formulations selecting the most interesting ones to be experimentally tested. During the development of the project of "Machine learning-aided design of composite mycotoxin detoxifier material for animal feed" we have started to explore the possibility of predicting the synergic effect between components and rapidly identify the optimal formulations. The approach allowed us to visualize the most interesting area of the descriptor space which have stimulated the curiosity to extend it to diverse set of data and application, such as the rheological additives.

Herein, we aimed to amplify the concept of synergy exploiting our on-working machine learning models while expanding the investigation with additional experimental data points.

A.2. Methods

Datasets, and model architecture. We exploited two different datasets, one for each property, i.e., adsorption of mycotoxins and viscosity of rheological additives. The formulations were defined by two different descriptor space reported in Table S1,2. Random forest was selected as machine learning algorithms were trained on 85% of each dataset (training set) and tested on the remaining 15% of the data (testing set). The models were hyperparameter optimized and validated with out-of-dataset experiments. RFeff predict the efficiency in mycotoxin adsorption while RFav2 apparent viscosity for rheological additives. The dataset and the model codes used to predict the adsorption and viscosity of hybrid materials are fully presented in Chapter 3 and 4, respectively.

Synergistic effect definition. The synergy between components in the hybrid formulations was calculated considering each material descriptor value as linear combination of the descriptors (Table S1) of the singular compound ($\sum_i a_i x_i$, where a_i is the amount (%) of the i-component in the hybrid formulations and x_i is the descriptor value corresponding to the pure i-component).

RFeff predicted the efficiency of the hypothetical composites (y_{pred}) against the removal of the mycotoxins contained into the database. The predictions were compared with the linear combination of efficiency of the pure components, being calculated as $\sum_i a_i y_i$, where y_i is the predicted target of the pure i-component. The difference between predictions and linear combination of pristine efficiency was assigned to the synergistic effect between components (Eq. 1).

For the exploration of the synergy in rheological composites the descriptors of the hypothetical formulations were again calculated as linear combination of the descriptor space of Table S2. In this case, the descriptor AlC is the one which encodes the information of the physico-chemical properties of the Al-Clay being based on Euclidean distances (see details in Supporting Information Section S1 and Figure S1). RFav2 predicted the apparent viscosity (y_{pred}) of the hypothetical hybrid and the pure components, and the synergy was then calculated with Eq. 1.

$$\text{Synergy} = y_{pred} - \sum_i a_i y_i \qquad \text{Eq. 1}$$

A.3. Results and discussion

Synergic effect in adsorption properties. We generated the linearly calculated descriptors of all the possible combinations of ternary formulations composed by sepiolite + montmorillonite + activated carbon (AC). The latter was explored from 0 to 5% since it rapidly reaches the plateau close to the 100% of efficiency. Figure 1 shown the prediction surface computed by RFeff in the 3D composition space of the hybrids. The vertices are the prediction associated to single or binary compositions reported in Table S3 and were used to calculate the equation plane in which each point is the linear combination of the prediction of the vertices. The resulting synergy contribution, calculated following the Eq. 1, was projected on the x,y plane.

The predictions displayed a noticeable displacement from the plane. Despite our models was trained by a dataset contained just information of pure MDTs and the descriptor space of the explored hypothetical MDTs was calculated by the rule of mixture, our approach was able to capture the hidden information about certain interaction between the compounds occurring when they are put working together. Thus, the predictions are out of the plane creating spots in which the performance is higher (painted in red) or lower (painted in blue) than the one estimated under the assumption that the components of the hybrid adsorb the mycotoxin as they do when they act alone. The maximum positive synergy is expressed for hybrid composed by ca. 30% of sepiolite and 1-2% of AC. However, the cooperation between clays can be visualized following the x,z plane.

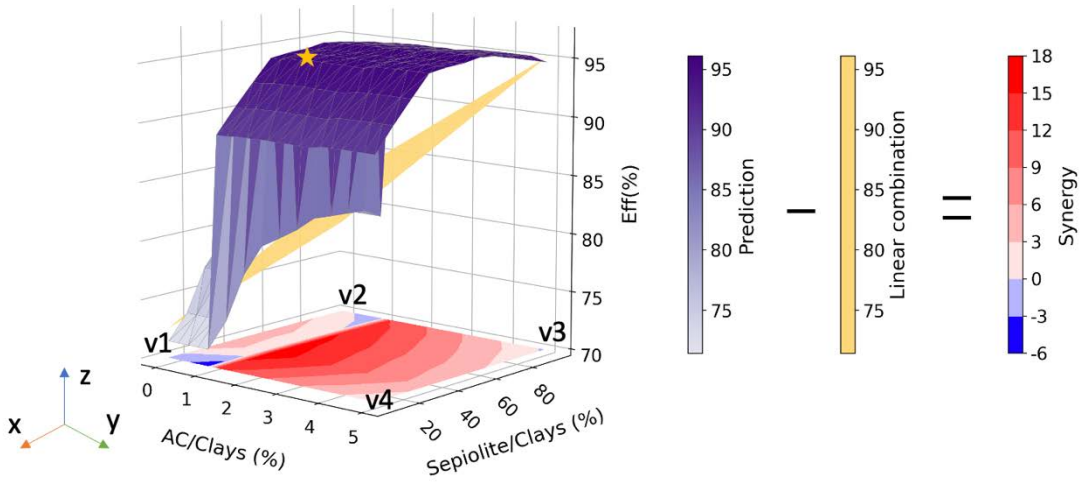


Figure 1. Prediction of efficiency in adsorption of zearalenone (ZEN) mycotoxin (purple surface) in the 3D space of ternary hybrid composed by sepiolite + montmorillonite + activated carbon (CA). The resulting synergic contribution was projected in the x,y plane. The composition of the vertices is displayed in TableS3 (B). The start represents the experimental datapoint. The experimental settings were fixed to 2 kg/t of inclusion rate, 2 μg/ml of toxin concentration pH 3 during the adsorption step and pH 6.5 during the desorption step.

The detoxification increases as the amount of sepiolite increases with a maximum of positive cooperation for those formulation which incorporated the sepiolite in ranges of 30-60%. Among

the possible formulations, we selected one placed in the reddest spot, which were prepared and tested obtaining an experimental value closed to the prediction.

To investigate the behavior of mixtures of different clays, we explore hybrid composition of Ca-montmorillonite and Na-montmorillonite (Na-AlC). RFeff predicted the efficiency of the selected mixtures in the adsorption of deoxynivalenol (DON), T2-toxin (T2), ochratoxin A (OTA), zearalenone (ZEN), fumonisin B1 (FB1), aflatoxin B1 (AFB1). The cooperation between clays changes depending on the to-be-adsorbed mycotoxin, suggesting that different modes of action are involved. The positive synergy starts from incorporating the 40% and 30% of Na-AlC for capturing T2 and AFB1 mycotoxins. On the other hands, for the uptaking of ZEA, OTA, DON and FB1, at least the 70% of Na-AlC should be incorporate in the MDT formulation to observe a positive cooperation. Those hybrids also revealed an enhanced efficiency with respect to the pure components and should be preferred to design suitable MDT. Additionally, it should be noted that when the components of the MDT cooperate negatively, the resulting predicted efficiency is lower than the one of the pure components and should be avoided to design MDTs. The latter interesting observation agrees with the concept that hybrid materials possess unique properties, far from the ones of the pure components and challenging to-be-estimated.

To validate the capacity of our approach to predict the sign of the synergy, we selected a formulation which exploits negative sign for the uptake of ZEN, DON, FB1, OTA, and T2 mycotoxins which positive synergy for AFB1. The experiments are shown with start shaped point. Even with certain displacements from the predictions, our approach was capable to make accurate prediction of the synergy sign.

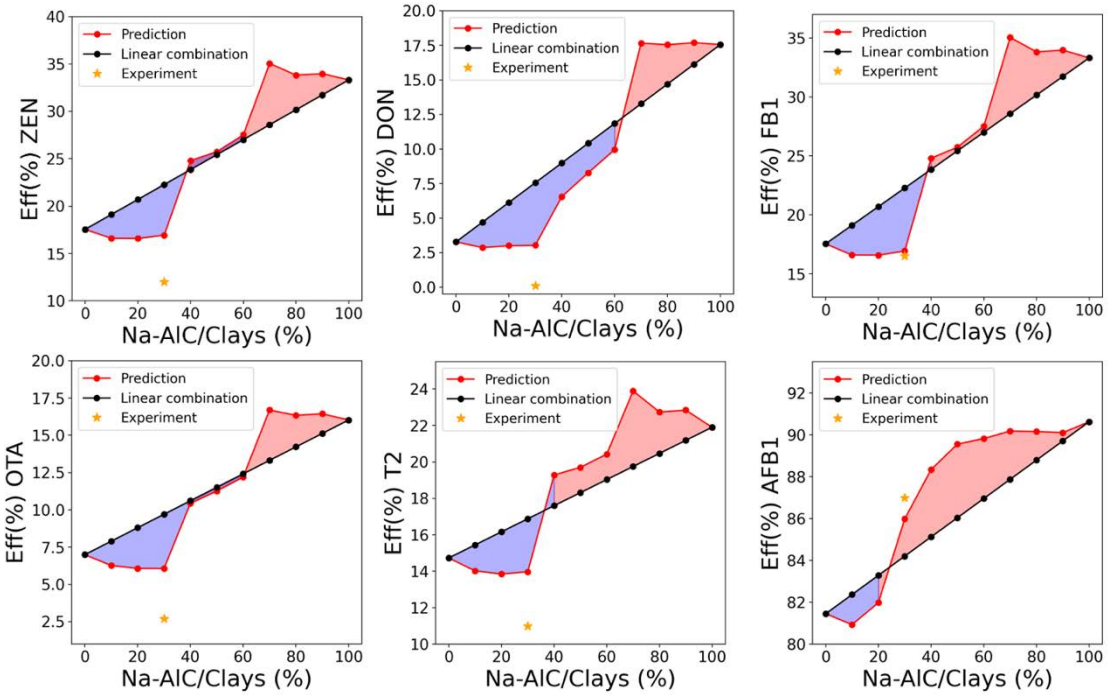


Figure 2. RFeff prediction of efficiency in adsorption (red points) of ZEN, DON, FB1, OTA, T2, AFB1 mycotoxin for mixtures of Ca-montmorillonite and Na-montmorillonite. The experimental settings were fixed to 1 kg/t of inclusion rate, 3 μg/ml of toxin concentration pH 3 during the adsorption step and pH 6.5 during the desorption step for all the toxins except AFB1 which was assessing at 2 kg/t of inclusion rate, 2 μg/ml.

A further validation was conducted on three formulations chosen among the possible combination of binary hybrids of smectite (MgC) and Na-montmorillonite. RFeff predicted the adsorption of ZEN mycotoxin. The incorporation of MgC up to 40% results in a negative cooperation with Na-montmorillonite. From 40% to 50% the hybrids behave following the rule of mixture, while increasing the MgC content a positive synergy can be observed. The three formulations were selected being showing negative, none, and positive synergy against the capture of ZEN mycotoxin. Again, our models correctly predicted the sign of the synergy contribution. Moreover, the experiment of the formulation incorporating the 65% of smectite revealed an enhanced efficiency with respect to the pure smectite. The latter remarks that well designed hybrids may be more effective of pure materials.

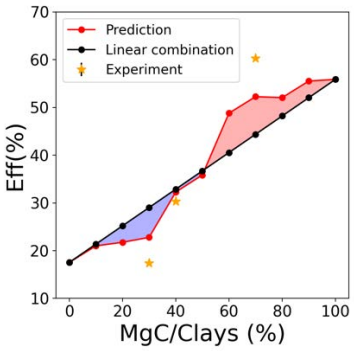


Figure 3. RFeff prediction of efficiency in adsorption (red points) of ZEN mycotoxin for mixtures of smectite (MgC) and Na-montmorillonite. The experimental settings were fixed to 1 kg/t of inclusion rate, 3 µg/ml of toxin concentration pH 3 during the adsorption step and pH 6.5 during the desorption step.

Synergic effect in rheological properties. We transferred the previously disclosed approach to investigate the synergy between the components of rheological hybrids. The validated RFav2 was used as surrogate model to predict the apparent viscosities of the water slurries. The hybrids are mixtures of smectite (MgC), montmorillonite (AlC), acrylic polymer (PolA), and cellulosic polymer (PolB) in various ratios. Figure 4 shows the predictions (purple surface), the calculated linear combination of the vertices p1-4 and d1-4 reported in TableS4, and the resulting synergic contribution. We explored changed the amount of MgC, AlC and PolA, while fixing PolB to the 1% (Figure 4A) and MgC, AlC and PolB, while fixing PolA to the 1% (Figure 4B). Similarly to the previous observations, the surface of predictions is displaced from the plane creating spots of higher viscosity (painted in red) or lower (painted in blue) with respect to the viscosity estimated under no-synergy assumption. The latter remarks the presence of the areas in which the conditions are promising to create a stable gel architecture and others in which the components flocculate or are not dispersed in a favorable environment. The components start to cooperate from the incorporation of 0.25% of acrylic polymer with a maximum of positive synergy in 0.75-1% (Fig. 4A). At least 1% of the cellulosic polymer should be added to observe cooperation and, consequently, the gel architecture. Both polymers optimize the electrostatic interparticle

interactions with high amount of Na-montmorillonite and can be appreciated following the x,z planes. On the contrary, the interaction between components in mixtures of smectite and Ca-montmorillonite is unfavorable when it is incorporated in high amount. The latter is an effect of the intrinsic property of the diverse clays, which is encoded in our AlC descriptor.

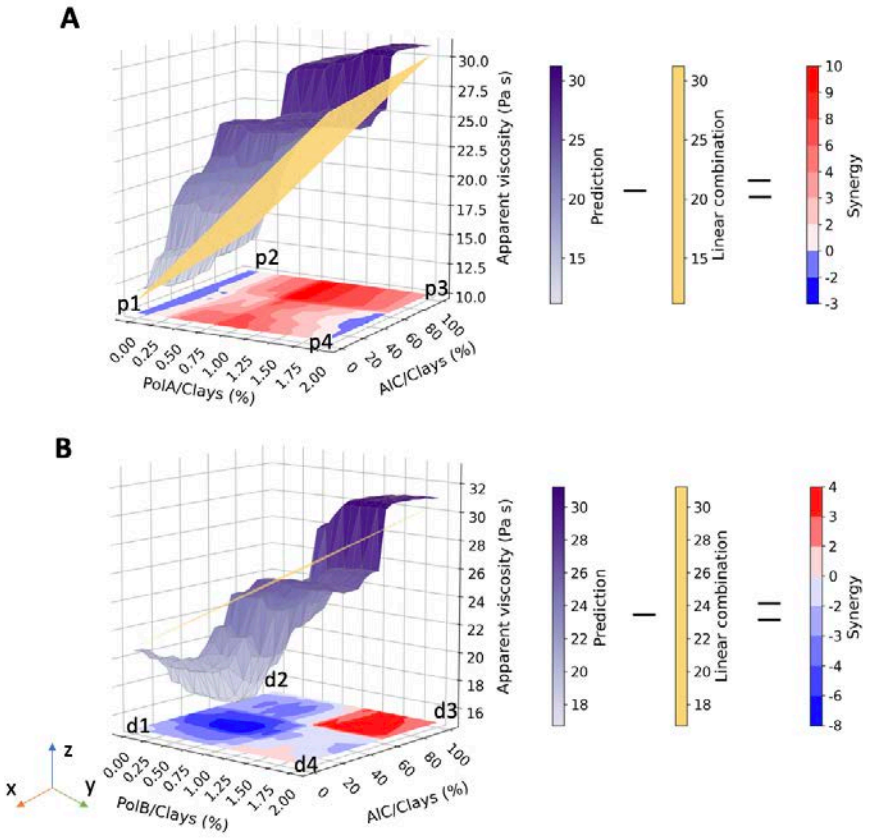


Figure 4. Prediction of the apparent viscosity of rheological hybrid composed by combination of smectite (MgC) + Na-montmorillonite (AlC) + acrylic polymer (PolA), fixing the cellulosic polymer (PolB) to 1% (A). Prediction of the apparent viscosity of rheological hybrid composed by combination of smectite (MgC) + montmorillonite (AlC) + cellulosic polymer (PolB), fixing the acrylic polymer (PolA) to 1% (B). The resulting synergic contribution was projected in the x,y plane. The compositions of the vertices are reported in TableS4.

To validate the capacity of capturing the synergy between smectite and a Ca-montmorillonite, named AlC6, RFav2 predicted the viscosity of several combinations varying the clay composition while fixing 0.7% PolA/Clays, 0.9% PolB/Clays. Two formulations were selected in the positive (red) and negative (blue) area of Figure 5. The experimental results further confirmed the capability of our approach to predict the sign and orientate the decisions in the material design process.

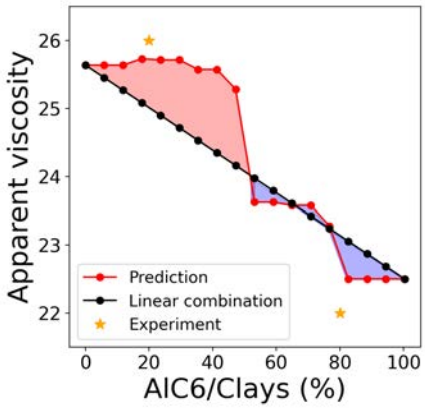


Figure 5. Prediction of the apparent viscosity of rheological hybrid composed by combination of smectite (MgC) + Ca-montmorillonite (AlC6) + 0.7% of acrylic polymer (PolA), and 0.9% of cellulosic polymer (PolB).

A.5. References

(1) Yerbolat, G.; Amangeldi, S.; Ali, M. H.; Badanova, N.; Ashirbeok, A.; Islam, G. Composite Materials Property Determination by Rule of Mixture and Monte Carlo Simulation. *Proc. 2018 IEEE Int. Conf. Adv. Manuf. ICAM 2018* **2019**, No. November, 384–387. https://doi.org/10.1109/AMCON.2018.8615034.

(2) Lee, D. N.; Kim, Y. K. On the Rule of Mixtures for Flow Stresses in Stainless-Steel-Clad Aluminium Sandwich Sheet Metals. *J. Mater. Sci.* **1988**, *23* (2), 558–564. https://doi.org/10.1007/BF01174685.

(3) Feng, B.; Xin, Y.; Sun, Z.; Yu, H.; Wang, J.; Liu, Q. On the Rule of Mixtures for Bimetal Composites. *Mater. Sci. Eng. A* **2017**, *704*, 173–180. https://doi.org/10.1016/j.msea.2017.08.005.

(4) Potluri, R.; Diwakar, V.; Venkatesh, K.; Srinivasa Reddy, B. Analytical Model Application for Prediction of Mechanical Properties of Natural Fiber Reinforced Composites. *Mater. Today Proc.* **2018**, *5* (2), 5809–5818. https://doi.org/10.1016/j.matpr.2017.12.178.

(5) Belachew, N.; Bekele, G. Synergy of Magnetite Intercalated Bentonite for Enhanced Adsorption of Congo Red Dye. *Silicon* **2020**, *12* (3), 603–612. https://doi.org/10.1007/s12633-019-00152-2.

(6) Klumpp, E.; Heitmann, H.; Schwuger, M. J. Synergistic Effects between Cationic Surfactants and Organic Pollutants on Clay Minerals. *Colloids Surfaces A Physicochem. Eng. Asp.* **1993**, *78* (C), 93–98. https://doi.org/10.1016/0927-7757(93)80314-5.

(7) Szeluga, U.; Kumanek, B.; Trzebicka, B. Synergy in Hybrid Polymer/Nanocarbon Composites. A Review. *Compos. Part A Appl. Sci. Manuf.* **2015**, *73*, 204–231. https://doi.org/10.1016/j.compositesa.2015.02.021.

(8) You, Y. J.; Kim, J. H. J.; Park, K. T.; Seo, D. W.; Lee, T. H. Modification of Rule of Mixtures for Tensile Strength Estimation of Circular GFRP Rebars. *Polymers (Basel).* **2017**, *9* (12). https://doi.org/10.3390/polym9120682.

(9) Prabhu, T. N.; Demappa, T.; Harish, V.; Prashantha, K. Synergistic Effect of Clay and Polypropylene Short Fibers in Epoxy Based Ternary Composite Hybrids. *Adv. Mater. Res.* **2015**, *4* (2), 97–111. https://doi.org/10.12989/amr.2015.4.2.097.

(10) Jung, B. N.; Jung, H. W.; Kang, D.; Kim, G. H.; Shim, J. K. Synergistic Effect of Cellulose Nanofiber and Nanoclay as Distributed Phase in a Polypropylene Based Nanocomposite System. *Polymers (Basel).* **2020**, *12* (10), 1–15. https://doi.org/10.3390/polym12102399.

(11) Natarajan, S.; Bajaj, H. C.; Tayade, R. J. Recent Advances Based on the Synergetic Effect of Adsorption for Removal of Dyes from Waste Water Using Photocatalytic Process. *J. Environ. Sci. (China)* **2018**, *65*, 201–222. https://doi.org/10.1016/j.jes.2017.03.011.

(12) Cuvitoglu, A.; Zhou, J. X.; Huang, S.; Isik, Z. Predicting Drug Synergy for Precision Medicine Using Network Biology and Machine Learning. *J. Bioinform. Comput. Biol.* **2019**, *17* (2). https://doi.org/10.1142/S0219720019500124.

(13) Cokol, M.; Kuru, N.; Bicak, E.; Larkins-Ford, J.; Aldridge, B. B. Efficient Measurement and Factorization of High-Order Drug Interactions in Mycobacterium Tuberculosis. *Sci. Adv.* **2017**, *3* (10). https://doi.org/10.1126/sciadv.1701881.

(14) Tournassat, C.; Davis, J. A.; Chiaberge, C.; Grangeon, S.; Bourg, I. C. Modeling the Acid-Base Properties of Montmorillonite Edge Surfaces. *Environ. Sci. Technol.* **2016**, *50* (24), 13436–13445. https://doi.org/10.1021/acs.est.6b04677.

(15) Teich-McGoldrick, S. L.; Greathouse, J. A.; Jové-Colón, C. F.; Cygan, R. T. Swelling Properties of Montmorillonite and Beidellite Clay Minerals from Molecular Simulation: Comparison of Temperature, Interlayer Cation, and Charge Location Effects. *J. Phys. Chem. C* **2015**, *119* (36), 20880–20891. https://doi.org/10.1021/acs.jpcc.5b03253.

(16) Zhang, P.; Yin, Z. Y.; Jin, Y. F.; Chan, T. H. T.; Gao, F. P. Intelligent Modelling of Clay Compressibility Using Hybrid Meta-Heuristic and Machine Learning Algorithms. *Geosci. Front.* **2021**, *12* (1), 441–452. https://doi.org/10.1016/j.gsf.2020.02.014.

(17) Lo Dico, G.; Nuñez, Á. P.; Carcelén, V.; Haranczyk, M. Machine-Learning-Accelerated

Multimodal Characterization and Multiobjective Design Optimization of Natural Porous Materials. *Chem. Sci.* **2021**, *12* (27), 9309–9317. https://doi.org/10.1039/d1sc00816a.

(18) Sharma, A.; Huang, R.; Malani, A.; Babarao, R. Computational Materials Chemistry for Carbon Capture Using Porous Materials. *J. Phys. D. Appl. Phys.* **2017**, *50* (1), 463002. https://doi.org/https://doi.org/10.1088/1361-6463/aa87e9.

(19) Mannodi-Kanakkithodi, A.; Pilania, G.; Ramprasad, R.; Lookman, T.; Gubernatis, J. E. Multi-Objective Optimization Techniques to Design the Pareto Front of Organic Dielectric Polymers. *Comput. Mater. Sci.* **2016**, *125*, 92–99. https://doi.org/10.1016/j.commatsci.2016.08.018.

(20) Lo Dico, G.; Croubels, S.; Carcelén, V.; Haranczyk, M. Machine Learning-Aided Design of Composite Mycotoxin Detoxifier Material for Animal Feed. *Sci. Rep.* **2022**, *12* (1), 1–11. https://doi.org/10.1038/s41598-022-08410-x.

(21) Lo Dico, G.; Muñoz, B.; Carcelén, V.; Haranczyk, M. Data-Driven Experimental Design of Rheological Clay–Polymer Composites. *Ind. Eng. Chem. Res.* **2022**, *61* (31), 11455–11463. https://doi.org/https://doi.org/10.1021/acs.iecr.2c00936.

A.6. Supporting information

Table S1. Input feature descriptors selected to address the design of hierarchical porous materials.

Raw clay	Additive	Modification process
Cation exchange capacity (CEC) Surface area (BET) Acid-base character (3 features)[a] Chemical composition (10 binary features)[b] Phyllosilicates composition (5 features)[c]	pKa_1 pKa_2 pKa_3 C, H, O, S counts #Double bonds Molecular weight XlogP H-donor H-acceptor Rotatable bond Polar surface	Activation Milling time Additive/Clay $\%_{g/g}$ Additive (M) RH $\%_{g/g}$ Final RH $\%_{g/g}$ Particle size

[a] pH measured at 0 and after 24h (pH0, pH24) 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.

Table S2. Feature vector space describing the datasets and the targets.

Descriptors
Maturation
Mixing rate
Polymers/Clays
PolA/Clays (%)
PolB/Clays (%)
AlC (%)
AlC[a]

[a] AlC descriptor is used only to train RFs2 with DS2.

Definition of the descriptor AlC.

Each Al-montmorillonites (AlC1-8) was characterized quantifying the metal oxide content with ICP-OES Varian, Agilent 730. The organic impurities contained in the clay samples were estimated as mass loss by calcination in a Heraeus M110 oven. The cation exchange capacity (CEC) was measured with NH_4 ion selective electrode, Metrohm 867-801 and Thermo Scientific Orion 960 titrator. 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 clay 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.

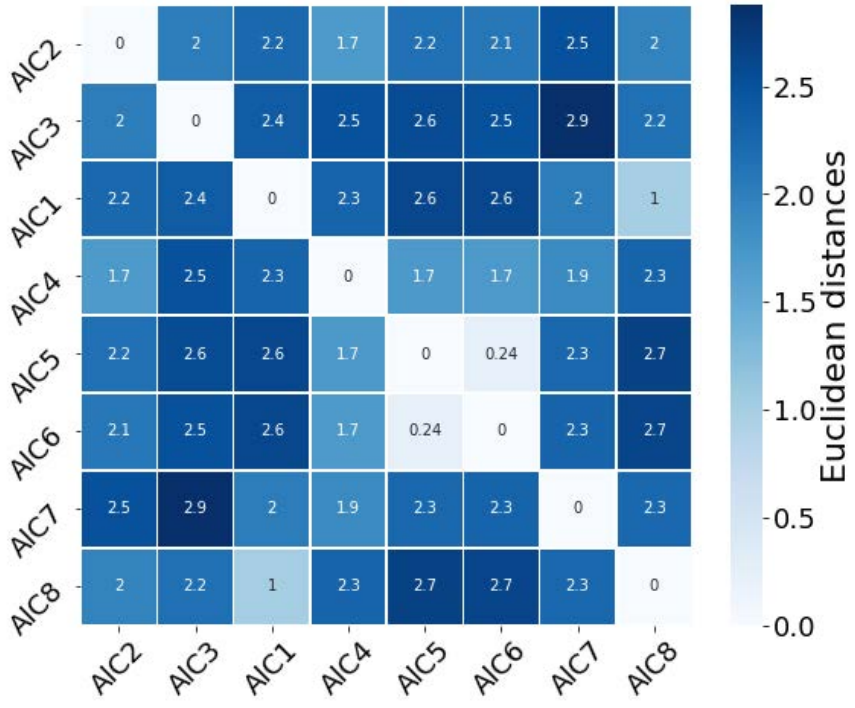


Figure S1. Distance matrix calculated on the base of Euclidean distances (Eq. 5) in the space of physico-chemical properties of the raw AlCs.

Table S3. Composition of the MDTs placed at the vertexes of Figure 1.

Point	AlC/Clays (%)	Sepiolite/Clays (%)	AC/Clays (%)
v1	100	0	0
v2	0	100	0
v3	0	100	5
v4	100	0	5

Table S4. Composition of the hybrids placed at the vertexes of Figure 4.

Point	MgC/Clays (%)	AlC/Clays (%)	PolA/Clays (%)	PolB/Clays (%)
p1	100	0	1	0
p2	0	100	1	0
p3	0	100	1	1
p4	100	0	1	1
d1	100	0	0	1
d2	0	100	0	1
d3	0	100	1	1
d4	100	0	1	1