Cement and Concrete Composites 136 (2023) 104903

Contents lists available at ScienceDirect

Cement and Concrete Composites

journal homepage: www.elsevier.com/locate/cemconcomp

A combined data-driven, experimental and modelling approach for assessing the optimal composition of impregnation products for cementitious materials

Janez Perko [a,*], Eric Laloy [a], Rafael Zarzuela [b], Ivo Couckuyt [c], Ramiro Garcia Navarro [d], Maria J. Mosquera [b]

[a] Belgian Nuclear Research Centre SCK CEN, Boeretang 200, B-2400 Mol, Belgium
[b] TEP-243 Nanomaterials Group. Department of Physical-Chemistry, Faculty of Sciences, University of Cadiz, 11510, Puerto Real, Spain
[c] IDLab, Department of Information Technology, Ghent University-imec, Gent, Belgium
[d] SIKA, R&D Department, Crta. de Fuencarral, 72, 28108, Alcobendas, Spain

ARTICLE INFO

Keywords:
Impregnation products
Optimization
Machine learning
Gaussian processes
Penetration depth
Pore size distribution
Mortars

ABSTRACT

The effectiveness of sol-gel based treatments for the protection of concrete depends on their capacity to penetrate into the material pores. Optimization of sol formulation to achieve maximum penetration depth is not a straightforward process, as the influence of different physical properties of the sol varies with the pore size distribution of each concrete. Thus, a comprehensive experimental programme to evaluate this large number of materials would require a significant number of experiments. This manuscript describes an approach, using combined computational and experimental approach to design tailor-made impregnation products with optimized penetration depth on concrete or cementitious materials with different pore size distributions. First, a process-based numerical model, calibrated experimentally for one sol composition and several cementitious material samples with different pore structures is developed. The model calculates the penetration depth for a specific pore structure. The optimization process utilizes the probabilistic and non-parametric Gaussian Processes regression method Gaussian Processes at two steps; first to make the choice of the optimal experimental design, and second to make predictions of physical properties based on the obtained training points. In the final step, the penetration depth is calculated for each mix combination in defined parameter range. The effectiveness of this approach is demonstrated on three cases. In the first instance, we optimized the impregnation product for the maximum penetration depth without any restrictions. With another two cases, we impose the restrictions on the gelation time, i.e. the time in which the sol reacts to gel. The validation of the procedure has been made by the use of a blind validation and shows promising results. The impregnation product penetrated significantly deeper with a product selected by using the described procedure compared to the considered best product before this optimization. The proposed procedure can be applied to a wide range of cementitious materials based on their pore size distribution data. This offers significant advantage compared to purely experimental approaches, where a set of experiments is required for each considered material.

1. Introduction

Concrete structures undergo various chemical and physical changes when exposed to environmental conditions. Many of these chemical changes affect mechanical properties and so affect the durability and service life of concrete structures. Typical processes are carbonation, internal swelling reactions, calcium leaching, thermal fatigue or ice damage [1–4]. In case of the preservation of cultural heritage, the deterioration processes include also aesthetical changes, such as black crust formation. Almost all of these processes are based on—or accentuated by—the water availability within the structure. Hence, one of the paths for the preservation of concrete structures is via treatment by impregnation by hydrophobic impregnation products [5–7] that increase the material cohesiveness and/or prevent water ingress without completely sealing the pores. This is important for the durability of the impregnation products as sealing the pores may result in delamination

* Corresponding author.
E-mail address: janez.perko@sckcen.be (J. Perko).

https://doi.org/10.1016/j.cemconcomp.2022.104903
Received 16 May 2022; Received in revised form 31 October 2022; Accepted 16 December 2022
Available online 17 December 2022

of the impregnation layer due to insufficient water vapour transport within the concrete matrix [8]. According to the internal Sika Market Analysis, the market of protection and preservation approaches as a whole represented about $950 MM in 2019, with a Compound Annual Growth Rate (CAGR) of 4% through 2026. Around 10–15% of these for dedicated product for construction chemicals solutions. Hydrophobic impregnation of concrete elements are, along with the corrosion inhibition and the modification of wetting ability of different solvent-substrate systems are the most represented solutions. The technology of paints and coatings technology has a significant energy demand during its manufacturing. Optimization of the products in terms of consumption and durability can have a large impact on CO_2 footprint, water and energy consumption.

Alkoxysilanes and their derivatives are typically used as impregnation treatments due to their ability to form amorphous silica gels that interact with silicate/aluminosilicate phases of the aggregate and cement matrix by creation of Si–O–Si or Si–O–Al bonds [9] and even react with the cement hydration phases yielding reaction products with a similar composition. The sol can react with Portlandite ($Ca(OH)_2$) yielding a calcium silica hydrate gel (C–S–H) or it can be incorporated into the structure of the existing C–S–H, which results in an increase of its Si/Ca ratio [5]. The sol may also react with Aluminium containing cement phases (e.g. Ettringite, monocarboaluminate, Katoite) resulting in amorphous aluminosilicate gels [10]. The use of impregnation products can be combined with the integrated microtexture of concrete surfaces [11]. The durability of the impregnation product depends, along with its chemical/mechanical stability and interaction with the substrate, on its penetration depth inside the pore structure. As such, it is of great importance to develop products that can penetrate as deep as possible for a given material and at given environmental conditions. This is especially important for high performance concrete with a high content of fines (e.g. limestone, silica fume, metakaolin) and consequently small pores, where a sufficient penetration depth is difficult to achieve [12]. In general, the penetration rate of any liquid—a sol in the particular case of alkoxysilane-based products—inside a porous material is governed by capillarity, viscous drag forces and gravity [13,14]. The intensity of these forces depends on different factors related to the pore structure of the material [15] (i.e. size distribution, total porosity, geometry) and the physical properties of the liquid itself (i.e. surface tension, viscosity, density). In the case of alkoxysilane sols, however, simplified models such as Washburn's [13] are not sufficient, since the sol properties evolve with time because of gelification and evaporation processes. Sol properties during gelification are normally obtained experimentally, though this process can be time- and resource-consuming when multiple composition variables are considered. In principle, some of the properties could be obtained by theoretical approaches, such as patchy Brownian Cluster Dynamics to predict the structure and gelation processes of sols with different $\%H_2O$ [16]. In a previous work, gelation time and initial viscosity were measured considering a rigidity percolation threshold and friction coefficient. Such approaches, however, pose a challenge, as correlating the model parameters with the experimental properties is not trivial. Furthermore, the relative influence of each physical property of the sol towards penetration rate varies with pore size. In a previous work, we calibrated a model to predict penetration of alkoxysilane-based sols on mortars with varying pore size distributions, and determined, by Monte Carlo combined with the Partial Rank Correlation Coefficient sensitivity analysis, its correlation with different sol properties for each considered pore size range [17]. By knowing the effect of composition variations on the sol properties, these relations could be used to predict the optimal formulation for a given material.

The prediction of a penetration depth from the mixture proportions for a given material and number of components in a product's mixture is a two-step process. First, the relevant physical properties of the considered mixture need to be estimated from the proportions of the mixture's ingredients. This requires some form of statistical or "data-driven" modelling since it cannot be feasibly done by physically-based numerical modelling. Second, penetration depth can be modelled from physical principles, knowing the physical properties of the mixture and the pore size distribution of the considered material.

In practical terms, a general difficulty in obtaining experimental data from the processes related to fine pore materials is that the experiments are time-consuming, and with a large number of independent predictor variables also the number of experiments increases. This makes it quite challenging as only a limited number of experiments, typically less than a few tens, can be performed to construct or "*train*" a data-driven model of the product's physical properties. Most of the recent work related to alkoxysilane systems focus on experimental observations and consider the composition variables as independent as for example in Ref. [18]. In Ref. [19] the characteristics of similar sols with different proportions of water and n-octylamine (n-8) are studied. The gel time showed an almost linear relation with n-8 content, and water had less influence. Viscosity only increased slightly with $\%H_2O$, except at higher %n-8, where its effect becomes more evident. In this work, we use the well-established machine learning (ML) technique model called Gaussian processes (GP) [20,21] to model the relationship between mixture proportions and physical properties. There has been rapid growth of the use of ML methods related to cementitious materials in the last 10 years. A very comprehensive overview has been given in Ref. [20] and the references therein. Most often, ML is used for regression tasks e.g. Refs. [21–23]. However, to the best of our knowledge, this is the first use of ML for the optimization of the impregnation products. The GP regression method is probabilistic and non-parametric and is known to be better suited than other ML methods for small datasets. Furthermore, GP regression not only provides point predictions but also their associated uncertainty estimates in a straightforward way) [20]. To make the most of the affordable small number of laboratory experiments, we select our training points iteratively using an adaptive design of experiments (DOE). The latter builds the training set iteratively, adding new design points at each iteration based on information provided by the intermediate data-driven GP model, trained with the design points collected so far [24].

Overall, the goal of this work is to answer the following two questions:

- What is the optimal mixture with respect to the penetration depth?
- How does the optimal composition change for different materials (pore structures)?

This methodology enables us not only to obtain the mixture with the maximum penetration depth, but also to select the most optimal mixture based on different criteria. In this work, we show an example of optimal mixture under restrictions on the maximum gelation time.

The remainder of this paper is organized as follows. Section 2 details our proposed approach for assessing the optimal composition of a 3-component mixture product. Section 3 then presents our application results and provides some discussion, before section 4 offers a short conclusion and future perspectives.

2. An overview of the optimization approach to find the optimal composition of the impregnation product

The optimization of the final impregnation product is based on several consecutive steps, graphically presented also in the graphical abstract, which are described below. The first step of the process consists of linking the pore structure of a material and the physical properties of the impregnation product or "*sol*" with the penetration depth and uptake of the impregnation product. The relationship is described by a physical model, which describes the interaction between fluid momentum, viscous, capillary and gravitational forces. The model and the results are described in detail in Ref. [17]. Especially in fine porous media such as cementitious materials, laboratory experiments are significantly

time-consuming. It is therefore of great importance to make a design of experiments in such a way that the predictive capabilities are maximized and hence the number of experiments needed are minimized. Hence, in the second step we optimize the experimental design based on the minimisation of the largest variance of the estimated physical parameters, weighted according to the parameter importance with respect to the penetration depth (obtained in the first step). The methods of optimum design were developed in the early 20th century to describe the choice of the values of the explanatory variables or predictors in a regression model at which observations should be taken [25]. However, the most common experimental approach in design of experiments is either a one-factor-at-a-time method, where only one independent variable is changed, or a factorial design, where the effects and possible interactions of several factors, i.e. independent variables, can be studied. However, the number of experiments in a full factorial design grows quickly with the number of independent variables. Instead, we resort to a GP-based adaptive design [24] in order to define the experiment set with both the best predictive capabilities and the least number of experimental points. Initially, we start with an arbitrarily chosen set of first experiments. With this initial set we train a GP. With four physical parameters, we need to train four separate GP models. The GP predictions are made of both, a mean and its associated variance. The predictive uncertainty from the current GP models is used to select new points where the predictive uncertainty is largest before a new set of four GP models are trained with all the points collected so far. This process continues until a given criterion is met (e.g., maximum number of experiments, predictive accuracy threshold, etc.). In this study, we started with an initial set of 18 3-dimensional input vectors for which each dimension includes 3 regularly spaced data points in (dimensions 1 and 2) and 2 data points (dimension 3). In addition, we included 5 experimental points from the initial experimental campaign which resulted in an initial 23 experimental points. After training the first set of GP models, 6 new experimental points were defined based on the predictive variance of each predicted output (i.e. physical property), weighted by the importance. Section 4.2 elaborates in detail how this has been performed. This point selection and re-training process was repeated 2 times, sequentially leading from 23 to 29 and then 35 training points in total. For practical reasons, including the fact that the time incurred by a single experiment varies from 0.5 h to more than 24 h, no more than 35 experiments were performed.

In the third step, the physical parameters that define the penetration of the impregnation product are predicted for different possible mixtures. As detailed in section 4.2, each of the 35 training points consists of 3 mixture proportions (the input vector): the mass percentage of deionized water, the mass percentage of n-octylamine, and the mass percentage of an admixture of aminoalkyl alkoxysilane. For each mixture four measured physical properties of the resulting sol (the output vector): gel time, surface tension, density, and viscosity are predicted. A fifth physical property, namely the contact angle between the sol and mortar, was only measured for the initial 18 training points and was therefore taken as a mean value in the subsequent penetration depth calculations made with the model developed in Ref. [17].

As the final step of the optimization, the penetration depth has been modelled along an hypergrid of input values such that one can choose the input combination(s) that either give the largest penetration depth or additionally satisfy more criteria. For example, the largest penetration depth for a maximum gelation time. The current analysis does not include aesthetic criteria such as colour and transparency, but these are straightforward to implement providing that they can be described by numerical values.

3. Experimental methods and results

3.1. Properties of the impregnation product

The penetration depth of sol depends on five relevant physical properties, contact angle θ [°], viscosity μ [mPa•s], surface tension σ [mN/m], density ρ [kg/m3] and gelation time [min]. The values of these properties are time-dependent due to the gelation process or evaporation of ethanol (gelation by-product) from the sol, and are controlled by the amount and type of sol composition. The considered sols contain the following main components:

(i) De-ionized water, which acts as a reagent and allows the formation of inverse micelles in the reaction media (also named %H_2O in the text below);
(ii) n-octylamine, which acts as a catalyst of the sol-gel process and as a surfactant (also named %n-8 in the text below);
(iii) An oligomeric silica precursor (TES40, from Wacker): admixture of monomeric and oligomeric ethoxysilanes with an average chaing length of 5 Si–O units;
(iv) A monomeric alkylalkoxysilane (3-aminipropyl trimethoxysilane, APTMS) containing terminal –NHx groups of high polarity, to modify surface tension (also named %APTMS in the text below).

In this work, the components (i), (ii), and (iv) were varied, while component (iii) represents the vast majority of the component mass (96%–99% depending on the amount of other components). The initial "reference" sol composition is taken the same as developed in previous works to study the impregnation treatment for concrete and cement mortars [7]. The composition of this sol was initially defined according to practical restrictions for their application, namely:

- Min %H_2O and %n-8: lower proportions led to excessively slow gelation (unfeasible for practical use)
- Max %H_2O: higher proportions lead to unstable sols in closed vessels (short storage time).
- Max % n-8: higher proportions lead to excessively fast gelation (impractical for real applications)
- Max %APTMS: higher proportions lead to phase separation or excessively fast gelation.

For the reference sol, the contact angle, viscosity, surface tension, and density have been measured with time as described in Ref. [17]. The reference sol has been applied to six mortars with different pore structures and pore size distribution. Based on the measured penetration depth of the reference sol into each mortar, the importance of different sol physical properties with respect to penetration depth can be estimated. In Ref. [17] it has been shown that for pores with diameter below 10^{-5} m, the most important sol properties influencing the penetration depth are gelation time (correlation coefficient $r_t = 0.69$) and surface tension ($r_\sigma = 0.53$). Viscosity and contact angle are less important parameters with respectively $r_\mu = -0.29$ and $r_\varphi = -0.28$. Density plays a negligible role in the penetration depth for pore diameters typical for cement paste ($r_\rho \approx 0$). Correlation coefficients between considered output and the penetration depth are based on the modified Washburn equation [13] and Monte-Carlo analysis combined with partial rank coefficient determination. The analysis in Ref. [17] concludes that the influence of each physical in property to the penetration depth varies with the pore size following a different trend. This is especially true for pores >10 μm. The results in Fig. 1 show relatively constant values for the pores with diameters below 10 μm for all physical parameters. Since most of the pores in cementitious systems are below 10 μm diameter size, the parameter importance (or correlation coefficient) is taken from this range.

3.2. Properties of the tested samples

Five sets of mortar samples were prepared for the analysis of the penetration depth. One reference sol has been applied to each sample

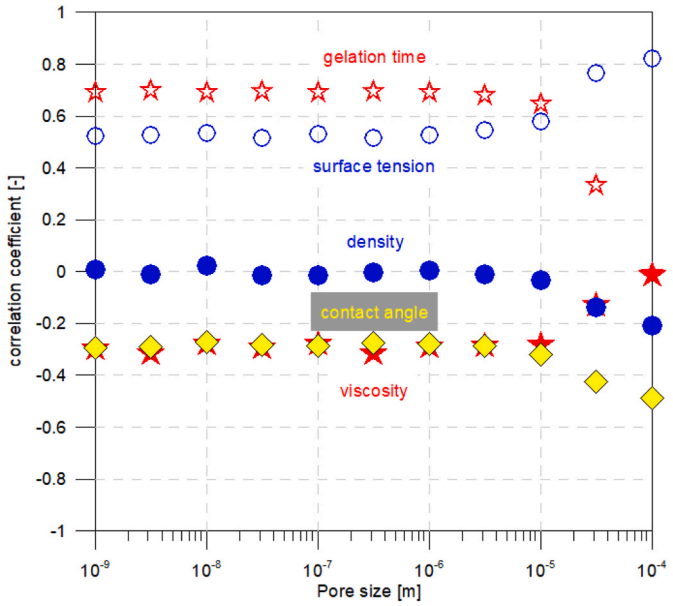

Fig. 1. Ranked correlation coefficients of the penetration depth for sol physical properties and different pore sizes.

and the penetration depth has been measured, together with the sol uptake. Prismatic OPC mortar specimens (4 × 4 × 16 cm^3) were prepared using CEM I 42.5R cement and CEN-compliant sand. In order to obtain specimens with different porosities and pore size distributions, three compositional parameters were modified: (1) the water-to-cement (w/c) ratios, (2) the sand-to-cement (s/c) ratios, and (3) the presence/absence of fine sand aggregates as described in Ref. [17]. The case without fine sand, the sand with size below 0.5 mm was removed by sieving. Analysed samples are distinguished based on w/c, s/c and fine sand fraction included or not (FS) and are shown in Table 1.

For all prepared samples, the characterization has been performed using MIP (Mercury Intrusion Porosimetry) on a Micrometrics AutoPore IV 9500 V[1].09 pore sizer. Samples were soaked in isopropanol for two days and then dried in a desiccator prior to analysis. MIP provides information on the intrusion volume of mercury as a function of pressure. The resulting intrusion curve is transformed into an equivalent pore frequency or equivalent pore size distribution assuming a cylindrical pore geometry and a contact angle of mercury of 138°. Integration over all pore size frequencies results in the material's porosity and averaging provides information on the equivalent pore size. Pore size distributions for the six mortars are given in Fig. 2.

3.3. Application on the mortars and determination of penetration depth

Prior to application, the mortar specimens were cut to 4 × 4x2.5 cm^3 prisms. The impregnation product was applied by brushing on a single face until apparent saturation. Successive layers were applied leaving the sample 10 min to absorb the sol between each. The mortar was considered saturated once the applied layer was not absorbed after 10 min. After the last application, any excess that was not absorbed for 20 min was removed with a cloth. The uptake was determined by the weight difference before and after the treatment (i.e. last layer).

Table 1
Identification and composition of the mortar samples used for the analysis.

Sample name	w/c	s/c	Fine sand (FS)
s/c 3/1 w/c 0.5	0.5	3/1	Yes
s/c 3/1 w/c 0.5 FS	0.5	3/1	No
s/c/3/1 w/c 0.7	0.7	3/1	Yes
s/c 5/1 w/c 0.7	0.7	5/1	Yes
s/c 5/1 w/c 0.7 FS	0.7	5/1	No
s/c 5/1 w/c 0.5 FS	0.7	5/1	No

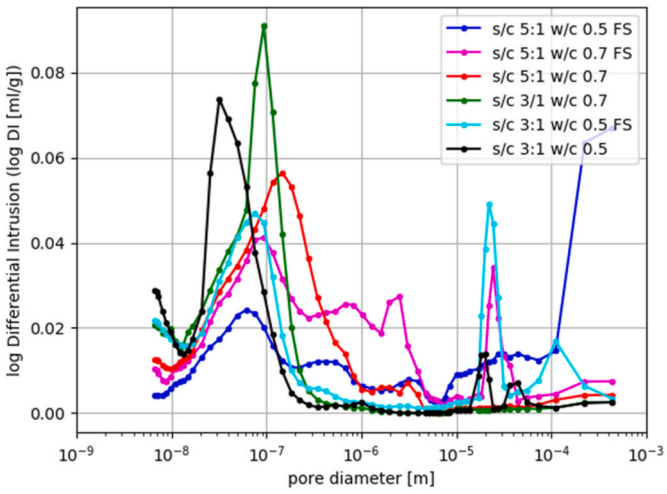

Fig. 2. Pore size distributions of examined mortar samples.

The penetration depth was experimentally determined via acid etching. The treated samples were cured for 24 h at atmospheric conditions (20 °C, 40% RH). Afterwards, the samples were cut in half and the cross-section was placed in a Petri dish with a 0.1 M HNO_3 solution during 20 s. The regions where the treatment is present are more resistant to the acid and, therefore, absorb less water than the regions where the impregnation did not take place. Penetration depth was calculated by image analysis (using ImageJ software) as the average of a minimum of 15 equidistant measurement points across the penetration front per replicate.

4. Modelling approaches

4.1. Nonlinear regression with Gaussian Processes

Gaussian processes (GP) can be thought of as a Gaussian probability distribution of functions [26]. The core idea is that the original process or function to be approximated, $y = F(\mathbf{x})$, resembles a Gaussian stochastic process $G(\cdot)$, which is completely determined by its mean function, $m(\mathbf{x})$, and covariance function, $k(\mathbf{x}, \mathbf{x}')$, and, if applicable, homoscedastic and uncorrelated Gaussian observational noise, ε, with standard deviation σ_ε [7]

$$F(\mathbf{x}) \approx G(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'), \sigma_\varepsilon) \tag{1}$$

Let us represent the ensembles of training and test data points by the $n \times d_X$ $\mathbf{X}$ and $n^* \times d_X$ $\mathbf{X}^*$ arrays, respectively, with n the number of training instances, n^* the number of test instances to be predicted, and d_X the dimensionality of $\mathbf{x}$. Similarly, the corresponding ensembles of training and yet to be predicted test outputs are denoted by the $n \times 1$ $\mathbf{y}$ and $n^* \times 1$ $\mathbf{y}^*$ vectors, respectively. Specifically, we use separate single-output GPs for the different quantities of interest to be predicted. Assuming $m(\mathbf{x}) = 0$, it can be shown that the predictive distribution of $\mathbf{y}^*$ is given by:

$$p(\mathbf{y}^*|\mathbf{X}, \mathbf{y}, \mathbf{X}^*) \propto N(\bar{\mathbf{y}}^*, cov(\mathbf{y}^*)) \tag{2}$$

With

$$\bar{\mathbf{y}}^* \triangleq \mathbb{E}[\mathbf{y}^*|\mathbf{X}, \mathbf{y}, \mathbf{X}^*] = k(\mathbf{X}, \mathbf{X}^*)\left[k(\mathbf{X}, \mathbf{X}) + \sigma_\varepsilon^2 \mathbf{I}\right]^{-1}\mathbf{y} \tag{3}$$

and

$$cov(\mathbf{y}^*) = k(\mathbf{X}^*, \mathbf{X}^*) - k(\mathbf{X}^*, \mathbf{X})\left[k(\mathbf{X}, \mathbf{X}) + \sigma_\varepsilon^2 \mathbf{I}\right]^{-1} k(\mathbf{X}, \mathbf{X}^*) \tag{4}$$

where $\mathbb{E}[\cdot]$ means the expectation, $k(\mathbf{X}^*,\mathbf{X}^*)$ is the $n^*\times n^*$ matrix of covariances between test input points, $k(\mathbf{X}^*,\mathbf{X})$ is the $n^*\times n$ matrix of covariances between the test and training input points and $k(\mathbf{X},\mathbf{X})$ is the $n\times n$ matrix of covariances between the training input points. For the prediction of each scalar output (physical quantity), y^*, equations (3) and (4) then reduce to

$$\mathbb{E}[y^*|\mathbf{X},\mathbf{y},\mathbf{x}^*] = k(\mathbf{x}^*,\mathbf{X})^T\left[k(\mathbf{X},\mathbf{X})+\sigma_\varepsilon^2\mathbf{I}\right]^{-1}\mathbf{y} \tag{5}$$

$$\mathbb{V}[y^*|\mathbf{X},\mathbf{y},\mathbf{x}^*] = k(\mathbf{x}^*,\mathbf{x}^*) - k(\mathbf{x}^*,\mathbf{X})^T\left[k(\mathbf{X},\mathbf{X})+\sigma_\varepsilon^2\mathbf{I}\right]^{-1}k(\mathbf{x}^*,\mathbf{X}) \tag{6}$$

where $\mathbb{V}[\cdot]$ denotes the variance and $k(\mathbf{x}^*,\mathbf{X})$ signifies the $n\times 1$ vector of covariances between the single test input point, $\mathbf{x}^*$, and the n training input points in $\mathbf{X}$.

There are many possibilities for the choice of the covariance kernel $k(\cdot,\cdot)$. Here we used the very popular squared exponential or Gaussian or radial basis function (RBF) kernel, which is the de-facto kernel for GPs. Preliminary results with other kernels such as the Matérn and linear kernels (and combinations thereof) did not show any improvement over the standard RBF kernel (see Ref. [27] for details about GPs' kernels). The RBF kernel is given by

$$k(\mathbf{x},\mathbf{x}') = \sigma_k^2\exp\left[-\frac{1}{2}\sum_{i=1}^{d_x}\left(\frac{(x_i-x_i')}{l_i}\right)^2\right] \tag{7}$$

where the kernel variance, σ_k^2, and correlation lengths (or lengthscales), $\mathbf{l}=[l_1,\ldots,l_{d_X}]$ are fitting parameters. Since $d_X=3$ in this study, the total number hyperparameters for each GP is 5: $[\sigma_k^2,l_1,l_2,l_3,\sigma_\varepsilon^2]$. The σ_ε^2 variable is jointly optimized with the other GP parameters rather than set to an estimate, as leave-one-out cross validation (LOOCV, see, e.g. Ref. [7], for details) showed that optimizing σ_ε^2 leads to better predictions. We construct and train our GPs using the GPFlow2 python-based toolbox [28,29]. Training is done by the classical approach of marginal likelihood maximization with a local, deterministic gradient-based search method [7], which we combine with a multi-start strategy to mitigate the risk of being trapped in a local optimum. Hence, each time a GP is trained we launch 20 local optimization runs with random initialization and select the one leading to the maximum marginal likelihood. Furthermore, the training data, i.e., the input and output vectors, are standardized

$$\mathbf{x} = \frac{\mathbf{x}^{\text{raw}}-\boldsymbol{\mu}_{\text{x}}^{\text{raw}}}{\boldsymbol{\sigma}_{\text{x}}^{\text{raw}}} \tag{8}$$

and

$$\mathbf{y} = \frac{\mathbf{y}^{\text{raw}}-\boldsymbol{\mu}_{\text{y}}^{\text{raw}}}{\boldsymbol{\sigma}_{\text{y}}^{\text{raw}}} \tag{9}$$

where $\mathbf{x}^{\text{raw}}$ ($\mathbf{y}^{\text{raw}}$) represents the untransformed input (output) vectors, $\boldsymbol{\mu}_{\text{x}}^{\text{raw}}$ ($\boldsymbol{\mu}_{\text{y}}^{\text{raw}}$) contains the means of the untransformed inputs (outputs) and $\boldsymbol{\sigma}_{\text{x}}^{\text{raw}}$ ($\boldsymbol{\sigma}_{\text{y}}^{\text{raw}}$) contains the standard deviations of the untransformed inputs (outputs).

4.2. Description of the adaptive design of experiments

At each step of the adaptive design process, the predictive performance is evaluated by leave-one-out cross validation (LOOCV) process. The latter consists of training with all available samples but one, and using this left-out sample to evaluate predictive performance. By cycling through the available data, one thus gets one prediction per data point. LOOCV can be performed with or without re-optimization of the GP model parameters (kernel variance, 3 kernel length scales and output noise variance) at each iteration. To avoid erratic variations in the optimized model parameters that can arise in such a small data regime as ours, we used LOOCV with no re-optimization. To quantitatively assess the LOOCV performance we compute for each of the 4 considered output types the coefficient of determination, Q_2, and root-mean-square error, RMSE, between the true and predicted left-out values

$$Q_2 = 1-\frac{\sum_{j=1}^{n}\left(y_j^t-y_j^p\right)^2}{\sum_{j=1}^{n}\left(y_j^t-\overline{\mathbf{Y}^t}\right)^2} \tag{10}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{j=1}^{n}\left(y_j^t-y_j^p\right)^2}{n}} \tag{11}$$

$n=35$ is the number of training points that are cycled over throughout the LOOCV, the $j=1,\ldots,n$ y_j^t and y_j^p are the true (observed) and predicted left-out output points in the original space ($y_j^t=y_j^{raw}$ in equation (9)), respectively, and $\overline{\mathbf{Y}^t}$ denotes the mean of the n true output points.

As written above, an initial dataset of 23 measurements (see section 3.1) is used for training. Afterwards, the 4 trained GPs are used to predict the $i=1,\ldots,4$ means, $\mu_{i,j}$ and standard deviations, $\sigma_{i,j}$ corresponding to $j=1,\ldots,40\times40\times40=64{,}000$ training input points regularly spaced along the 3 dimensions of the input space.

We propose to select new points in areas where the predictive standard deviations are large (see Fig. 3 for the case of the gelation time output). In order to obtain a relevant representation of the importance of the 64,000 potential candidates, the $\sigma_{i,j}$ values are weighted with the correlation coefficient (see section 3.1) between the considered output and penetration depth, r_i. More specifically, the $\sigma_{i,j}$ values are first normalized to lie in the [0,1] interval for each predicted quantity, i, before being multiplied by the associated correlation coefficients and summed up over i to get a score, χ, as shown in Eq. (12)

$$\chi_j = \sum_{i=1}^{4}\widetilde{\sigma}_{i,j} \tag{12}$$

where the $\widetilde{\sigma}_{i,j}$ are the normalized $\sigma_{i,j}$ values. Based on these scores, additional experiments are iteratively selected by computing the 0.5, 0.6 0.7 0.8, 0.9, and 1.0 percentiles of the χ distribution and retaining the associated input data. As mentioned earlier, doing this two times led to a total of 35 experiments.

As mentioned, the sols used in this work are based on an oligomeric tetra-ethoxysilane precursor containing de-ionized water (named % H2O), n-octylamine (named %n-8) and a monomeric –NHx terminated alkylalkoxysilane (named %APTMS).

The mix of these three factors changes the physical properties that, in turn, influence the penetration depth in a non-linear and non-monotonic way. The levels for each factor are defined as given in Table 2.

First, we performed a single factor experiment. Only one factor (% H_2O, %n-8 or %APTMS) changes from the reference point (0.5% H_2O, 0.16% n-8 and 0% APTMS) in these experiments. Some combinations of components resulted in very long gelation times. The details can be consulted in Table S1 in the supplementary material. Because it is difficult to estimate when the sol uptake stopped (i.e., sol becomes gel) for these cases, we did not consider these experiments as our input points due to their large uncertainty. Similarly, we excluded the case where gelation is too fast and was consequently not measured with sufficient accuracy. Contact angle was not measured for these experiments. The reason is that it was very difficult to accurately measure contact angle on the material because of its suction properties. For 18 experiments, the contact angle was indirectly measured by the angle of a liquid at the pore wall in a glass tube with a 1.15 mm diameter using Washburn's equation. This approach gives a good indication of true contact angle because the surface energy of glass, which determines the contact angle with the sol, varies in the 60–70 erg/cm^2. This is similar to the cementitious matrix, which typically has value of 40–65 erg/cm^2. Hence, for contact angle, we have fewer experimental points and they could not be predicted by GP satisfactorily. Taking the mean value of

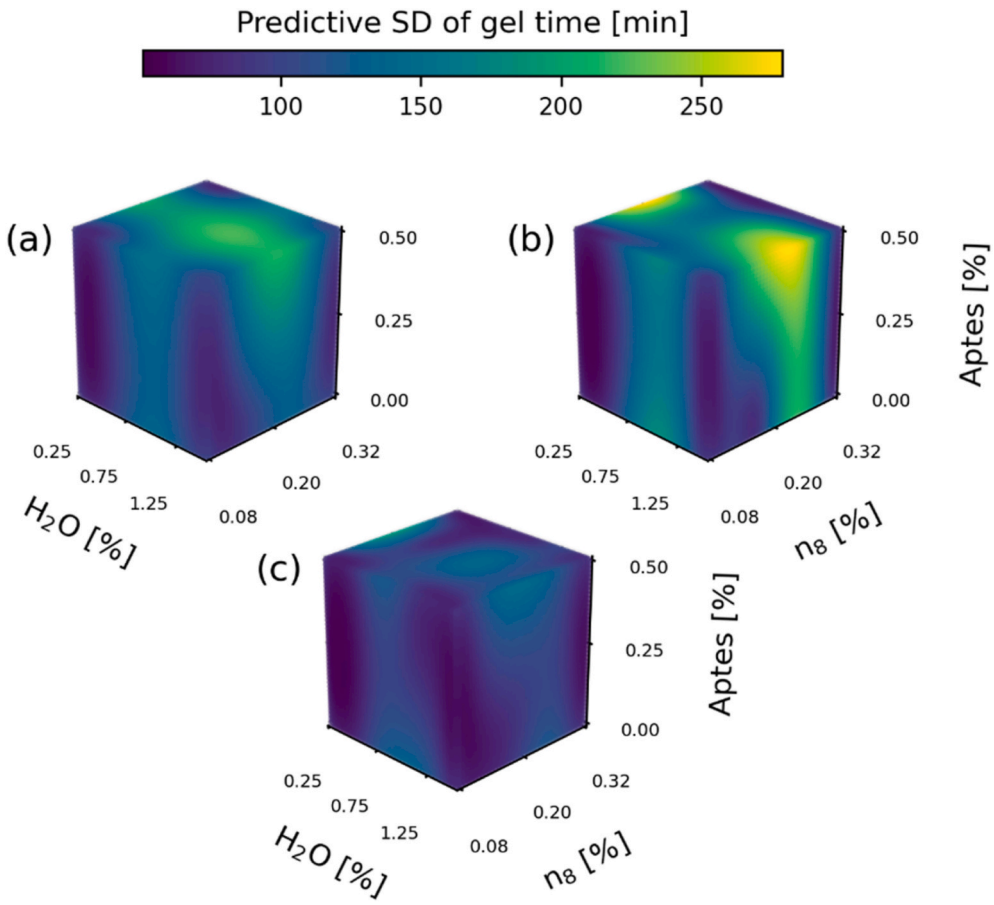

Fig. 3. Predictive standard deviation (SD) of the gelation time variable obtained after training the corresponding GP with (a) the initial set of 23 points, (b) the second training of 29 points (initial set plus the next 6 selected points), and (3) the third and final training set. Of 35 points (second training set plus the next 6 selected points).

Table 2
Factor levels for the sol mix.

Factor	Min level	Max level
$\%H_2O$	0.25	1.50
%n-8	0.08	0.32
%APTMS	0.00	0.50

contact angle was indeed found to be a better prediction than using the trained GP with 18 training points only.

We performed additional experiments in the corners (bound levels) of each factor dimension to obtain more representative sampling space. These results in 8 more experimental points (see Table S2 in the supplementary material for details).

For the basic experimental set, we included midpoints for $\%H_2O$ and %APTMS. This resulted in additional 10 experimental points. The reason to reduce the number of factor values in %n-8 is that it was less sensitive and to reduce the number of experiments. More details on these experimental results can be found in Table S3 in the supplementary material.

Beyond this set of 23 valid experiments, the additional experiments is defined by the use of GP as explained in section 4.2 for 6 additional points each time. The details on the selection of these experimental points are given in the Supplementary Material Table S4 and Table S5.

4.3. Performance of the trained GPs

Fig. 4 display the LOOCV results associated with the 4 trained GPs using the final 35-point dataset. With Q_2 of about 0.83 and 0.74, it is observed that the prediction quality is the greatest for the two most important output types, gelation time and surface tension (see Fig. 4.), respectively. Prediction quality is substantially lower for density but this is not an issue given its very low importance or influence with respect to penetration depth (see the near-zero correlation coefficient of density in Fig. 1). Overall, the results presented in Fig. 4 are deemed sufficiently accurate to use the corresponding GPs to optimize the impregnation product composition (see section 4).

5. Results from the optimization of the impregnation product

The final step in the optimization of the impregnation product is to find which input combination leads to the largest penetration depth of the resulting sol into cementitious materials. In this process, the GP-predicted physical properties associated with a given set of input combinations are used to simulate penetration depth and uptake of sol. In this study, we selected the input combinations to evaluate using a grid search approach. The input space was divided into 6 equal divisions in each direction, by %H2O, %n-8, and %APTMS, and we calculated the penetration depth for each of the resulting 216 sampling points.

By ranking the penetration depth results we can derive the optimal sol mixture. Since in this example we have three factors, we can present the results graphically in 3D plots (see Figs. 6–8). The size of the spheres in Figs. 6–8 denotes the penetration depth for a given mixture of %H2O, %n-8, and %APTMS.

We can explore different dependences of penetration depth. First, we investigated the influence of the number of experimental points. The derived 50 largest penetration depths are compared for the cases where

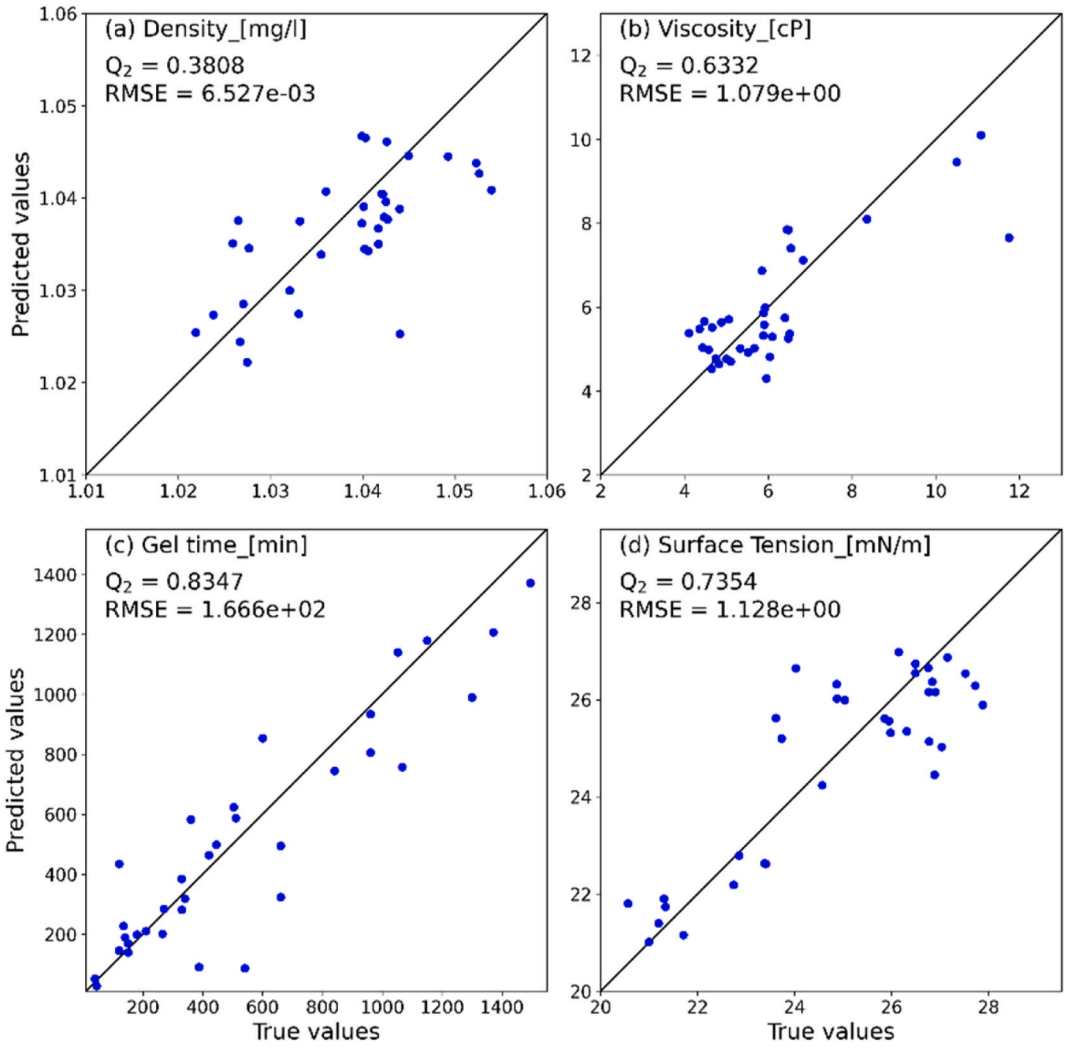

Fig. 4. 1-1 plots of LOOCV performance for each of the four GPs, obtained when using the 35-point dataset. The Q_2 coefficient and RMSE denote the coefficient of determination and root-mean-square-error, respectively, between the 35 true and predicted left-out points.

the GPs are trained with 18 experimental points (3 variations for %H2O and %APTMS and two variations for %n-8), 23 points, 29 points, and 35 experimental points.

5.1. Dependence of the derived penetration depth on the number of experimental points

Due to the significant time that is needed to realise one experiment, we want to keep the number of experiments as low as possible. The purpose of this section is to determine how robust the penetration depth estimates are with respect to the number of used experimental points (that is, GP training points). To do so we compare the input combinations associated with the derived 10 highest penetration depths for a different number of experimental points.

For the standard cementitious material sample (water/cement ratio of 0.5 and cement/sand weight ratio of 1/3) Fig. 5 shows that the highest penetration depth estimates can be very different for different number of GP training points. The average value of the penetration depth of the 50 most penetrating mixtures decreases with the number of experimental points, and the difference between the averaged values of the penetration depth related to number of training points tends to converge. The largest differences are for the highest amounts of n-8 and H_2O, especially between 18 and 23 experimental points. The details are provided in the supplementary materials Table S6 and Table S7.

With more training points also the derived maximum penetration depth tend to decrease, which indicates the presence of over- (and under-) estimations, especially at the corners of the experimental domain. We observe a convergence of results, since the absolute difference in the estimated penetration depth between 29 and 35 is lower compared to the difference between 18 and 23, and 23 and 29 experimental points.

5.2. Dependence of the sol composition for different materials

Because the importance of sol's physical parameters for the penetration depth depends on the pore size distribution of the cementitious material, the optimal mix might not be the same for all pore structures. As shown in Fig. 2, the studied mortars are characterized by a wide variation of pore size distributions for testing this hypothesis. The 6 available samples are compared. For all samples, the results are obtained based on 35 training points.

Results show that the general trend of largest penetration is always with little catalyst, but with either more water or less water combined with different amounts of APTMS precursor.

A more detailed look at the results in the supplementary material reveals that the largest penetration depths are obtained for different mixtures for different pore structures. Table S8 gives the details on the ranking of the largest penetration depths. Although there are several

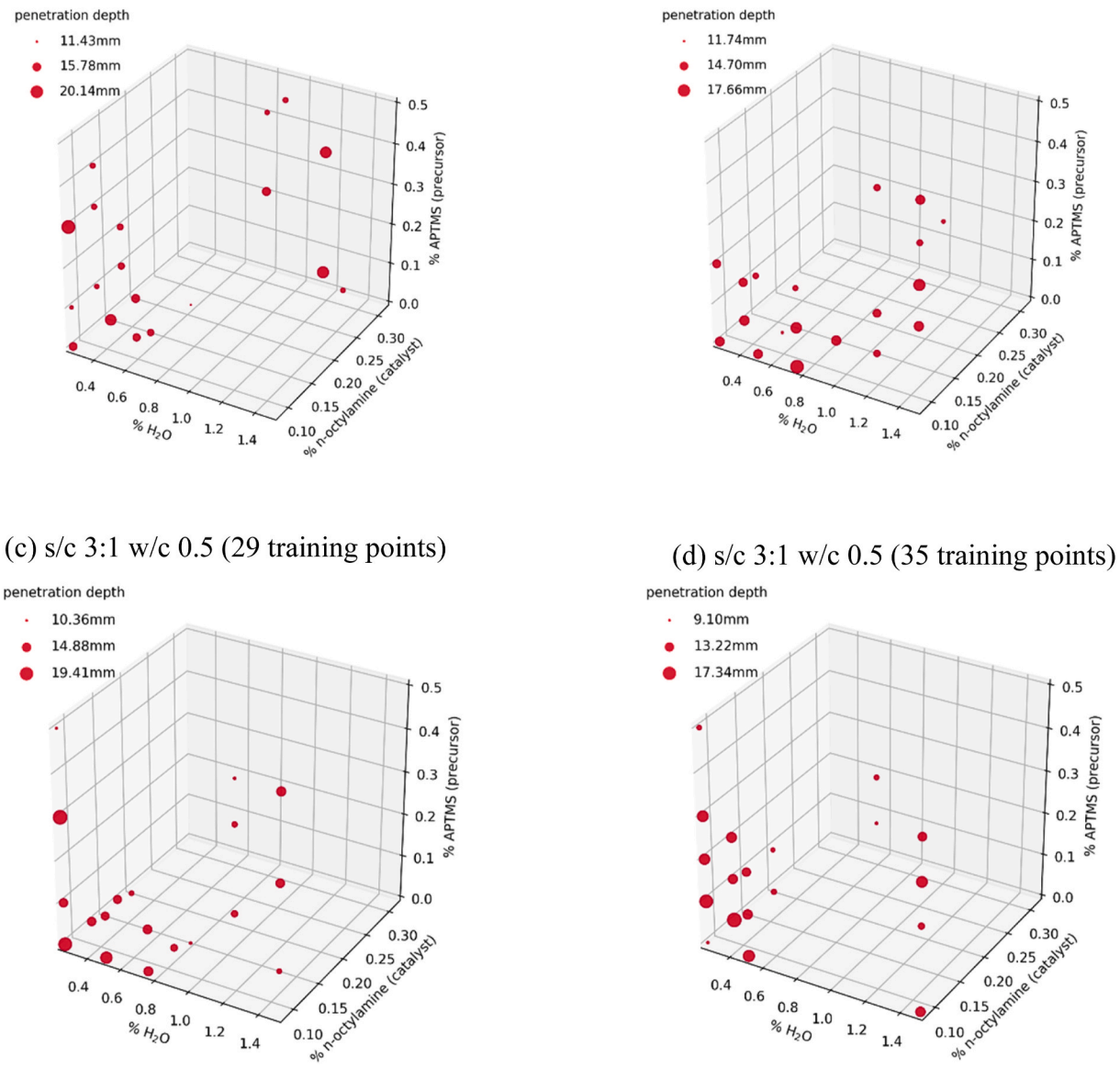

Fig. 5. Visualisation of the penetration depth as a function of input factors for the 50 cases with the largest penetration depth for: (a) 18 training points, (b) 23 training points, (c) 29 training points and (d) 35 training points.

matching results, it becomes clear that the optimal mixtures are not the same for all pore size distributions. The mix, which appears at least in four different pore structures, is marked in Table S8. From the matching compositions (colour-coded in the tables), it is obvious that different mixtures are better suited for specific groups of pore structures. For example, the mixture with 1.5%H_2O, 0.08%n-8 and 0.3%APTMS provides good penetration depth for s/c 5:1 w/c 0.7, s/c 3:1 w/c 0.7 and s/c 5:1 w/c 0.5 FS but it is not within the best choices for other samples. Hence, this mixture is better suited for materials with larger pores. A more wide range mixture would be with 0.25%H_2O, 0.08%n-8 and 0.2% APTMS that gives good results for five out of six samples. Just a little different APTMS concentration would also give good results for the first sample.

In general, one can conclude that for materials with an increasing amount of larger pore sizes, the importance of surface tension increases and the importance of gelation time decreases.

5.3. Optimization of the product imposing restrictions

5.3.1. Restriction on curing time

The results for the largest penetration depth presented above consider long gelation times, up to one day. For practical purposes, it may be convenient to limit the gelation time to shorter periods. With the optimization procedure described in this work, it is possible to find the best mixture(s) for any required maximum time. For the demonstration of optimization capabilities, the results for two maximum curing times 12 h and 6 h, respectively, are shown. In order to determine the sensitivity of the results with respect to pore structure, two samples, s/c 3:1 w/c 0.5 (predominantly small pores) and s/c 5:1 w/c 0.5 FS (predominantly large pores) are considered.

Fig. 7 shows the predicted penetration depths when applying restrictions on the gelation time. The gelation time is predominantly controlled by a catalyst, which is reflected in the results shown in Fig. 7. To have complete gelation in 6 h, the amount of catalyst should be between 0.2 and 0.3%. For 12 h maximum gelation, this amount varies between 0.1% and 0.15% for the best penetration depth. Within this range of catalysts, there is a trend that more precursor (APTMS) is

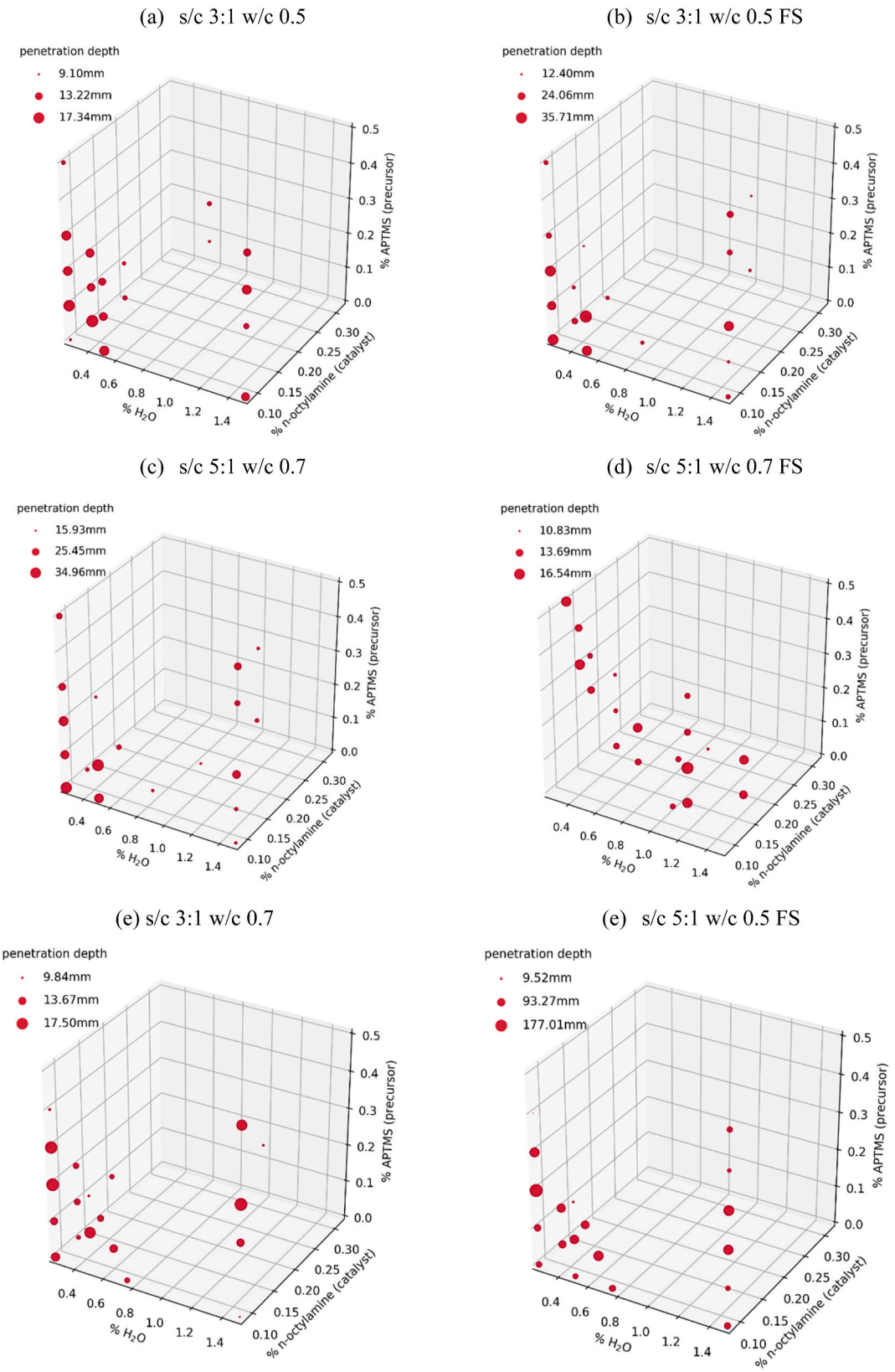

Fig. 6. Visualisation of the penetration depth as a function of input factors for 6 pore structures and for 35 training points.

needed if less water is added.

Comparison between finer and coarser pore structure reveals, that the penetration depth is very similar for both materials if the time of gelation is limited. While the estimated maximum penetration depth for 12 h gelation is 9.41 mm for fine pore materials, the corresponding estimated penetration depth for coarser materials is only slightly lower, with a value of 8.78 mm. This is because the capillary force in smaller pores is dominant and the transport due to capillary forces is faster than

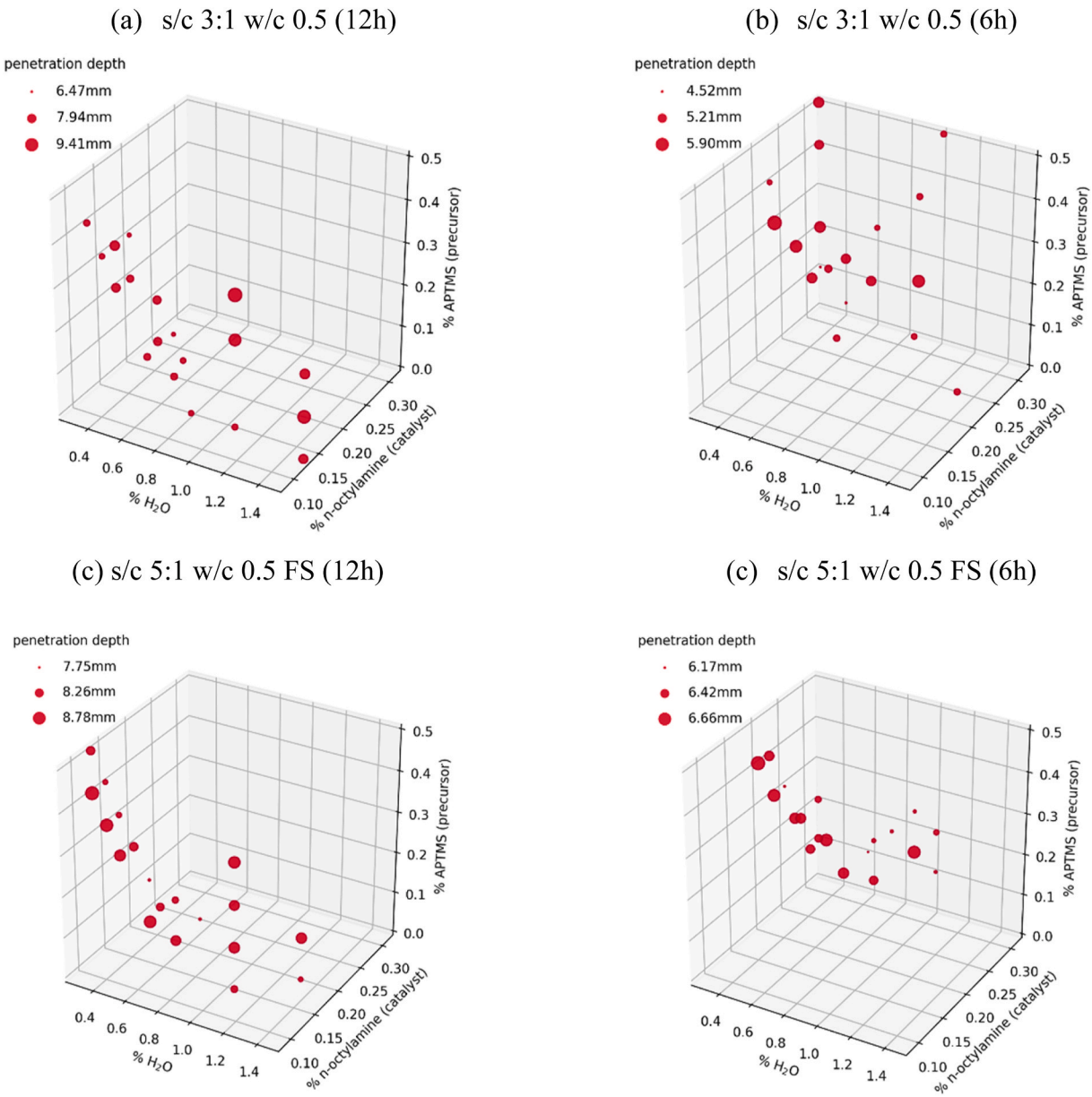

Fig. 7. Visualisation of the penetration depth (50 largest penetration depths) with the restriction on time as a function of input factors for 2 pore structures and 35 training points.

the gravitationally driven transport. Please note that our penetration depth model considers gravity oriented in the opposite direction to suction. This corresponds to the way that experiments are performed. In practical applications, this would represent the application of the impregnation product on the ceiling of the structure (i.e. applied from below). If the product is applied at the top, then large pores and large fractures would be filled very quickly.

For 6 h gelation time, the estimated penetration depth for fine pore materials is 5.9 mm and 6.66 mm for coarse materials. The mixtures that give the best penetration depth are similar between the two-pore structures. For example, 4 out of the 5 mixtures that give the best penetration depth in 12 h are the same. For 6 h curing time, this is the case for two mixtures. For 12 h gelation, only one common mixture appears to provide the optimal penetration depth. For maximum allowed 6 h gelation time, three such mixtures exist as seen in Table S9 in the supplementary material.

5.4. Validation of the numerical results

The complete process of optimization presented above is based on limited amount of data with six calibration points for the physical model, using one “reference” sol applied to six materials. On the other hand, for the training of the GP model, 35 experiments on one material and different mixtures of the impregnation product were used. Our aim is to predict the most optimal mixture for any material. To validate our proposed optimization process, two samples with known pore structure were chosen. For each of these materials a blind prediction has been made for the optimal (a) optimal mixture, (b) physical properties and (c) resulting penetration depth. For the optimal mixes, we selected a limitation on curing time of 12 h (720 min). As a comparison the reference mixture named UCA-T [5], optimized in non-structured way (i.e. based on practical qualitative criteria) was used. Corresponding mixes are presented in Table 3.

Using our trained GP models the physical values for these mixtures are calculated. Note that, as mentioned above, the mix composition of

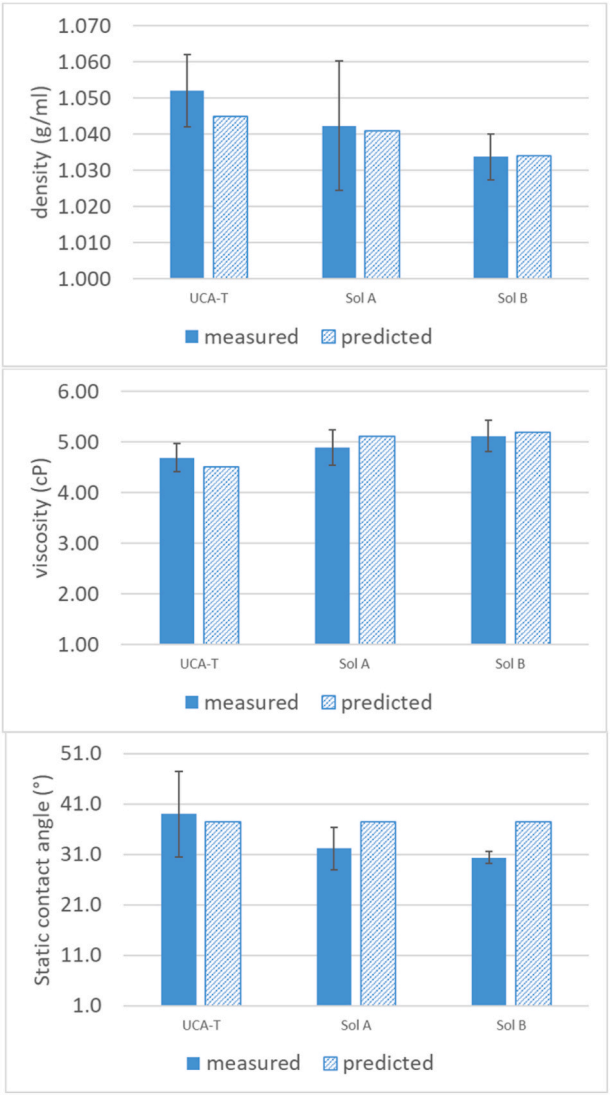

Fig. 8. Comparison between measured and predicted physical values: density (top), viscosity (mid) and contact angle (bottom).

Table 3
Addition of mix components in [%]. . UCA-T is a reference mixture (not optimized through the proposed method), Sol A and Sol B are optimized mixtures for two different materials.

Product	%H2O	%n-8	%APTMS
UCA-T	0.50	0.160	0.00
Sol A	0.25	0.176	0.20
Sol B	1.25	0.080	0.20

UCA-T is not obtained by the numerical optimization. From Table 4 can be seen that the predicted gelation time for UCA-T is higher than 12 h, while the other sols gave the gelation time below 12 h, which was the imposed optimization criterion.

Based on these predicted physical properties the penetration depth can be calculated. The penetration depths are given in Table 5. Sol A was optimized for the s/c 5:1 w/c 0.7 sample and Sol B for s/c 5:1 w/c 0.7 FS sample. In both cases, the predicted penetration depths are estimated to be larger for the optimized formulation compared to the non-optimized UCA-T formulation.

The intention of this validation is to test the predictive capabilities of the developed procedure. Experiments were performed for each mixture. Comparison for density and viscosity are very good as seen in Fig. 8 and Table 5.

The contact angle was approximated only by the mean of measured values, as the number of experiments for this property was too low as explained above. Indeed, the averaged values fall within the experimental uncertainty. In addition, the static contact angle measurements are subject to uncertainty because the sol droplet is quickly absorbed in the mortar (less than 1 s for a 5 μl drop) and thus varies fast with time, which is limited by the framerate of the camera and may not be fully representative of an equilibrium situation. However, the predictions of surface tension are always higher than the experimental values as shown in Fig. 9 and in Table 6 where the predictions for gelation time are consistently lower. This difference might be due to experimental variations between reagent batches that the model was trained on. In the case of gel time, fluctuations in atmospheric moisture and temperature (i.e. the experiments were made at room conditions), can interfere in the process. Standard deviation for gelation time is not given, as the measurement have to be “destructive” in the sense that the sol completely changes its state. Replicates within the same experiment give the same values. Gelation times, however, depend heavily on the ambient conditions. Especially humidity has a large impact and the validation experiments were performed at different time of year than the calibration experiments.

Uncertainties in gelation time affect also the predicted penetration depth because of its high correlation (see Fig. 1). The penetration depths in Fig. 10 were calculated as the average from a minimum of 15 equidistant points (for each of the 2 replicates) through the penetration front. Experimental results are shown in Fig. 11. Aside from the uncertainties in sol properties, a possible source of differences between predictions and experiments could be in the mortar pore structure. For instance, surface carbonation may affect chemical composition and pore structure over the timespan between the calibration and validation experiments because the samples were not stored in a sealed environment. However, this may only be validated if further studies in controlled environment were performed.

Despite all described uncertainties, the trend of the penetration depth of different samples and different mixes is well captured. The prediction matches the trend of penetration depth for different materials.

The results also demonstrate that the penetration depth is better with optimized mix compared to the original UCA-T product. Fig. 11 presents of the cross-section images of the impregnated materials, cleaved 24 h after application are presented in Fig. 11. For the visualisation purposes, the section was etched with 0.1 M HNO_3 and red lines are added as visual guide to better distinguish the penetration front.

Table 4
Predicted physical parameters for the selected mixes. UCA-T is a reference mixture (not optimized through the proposed method), Sol A and Sol B are optimized mixtures for two different materials.

Product	ρ (g/ml)	η (cP)	SCA (°)	Gelation time (min)	γ (mN/m)
UCA-T	1.045	4.53	37.4	862	21.7
Sol A	1.041	5.11	37.4	678	26.4
Sol B	1.034	5.19	37.4	663	26.5

Table 5
Predicted penetration depths for the selected mixes. UCA-T is a reference mixture (not optimized through the proposed method), Sol A and Sol B are optimized mixtures for two different materials.

	Penetration depth in mm	
Product	s/c 5:1 w/c 0.7	s/c 5:1 w/c 0.7 FS
UCA-T	9.8	8.8
Sol A	13.0	
Sol B		16.5

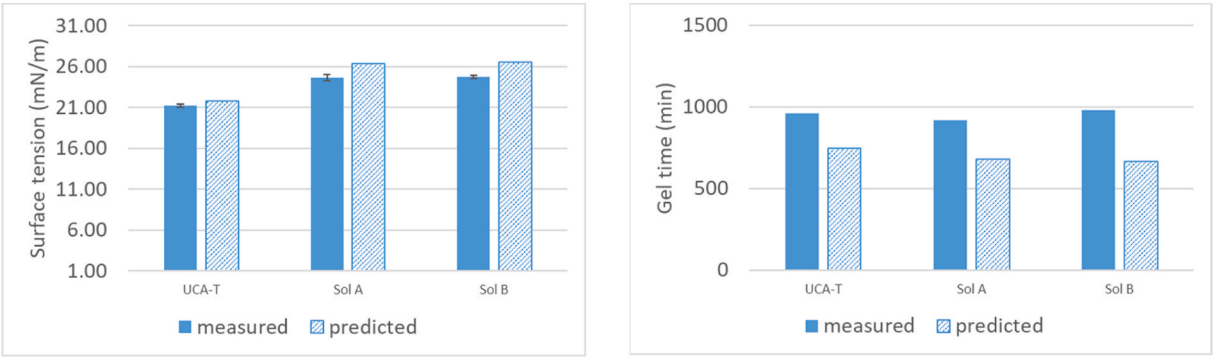

Fig. 9. Comparison between measured and predicted physical values: surface tension (left), and gelation time (right).

Table 6
Measured properties for the validation mixes.

Product	ρ (g/ml)	η (cP)	SCA (°)	Gelation time (m)	γ (mN/m)
UCA-T	1.052	4.69	39.0	960	21.20
Sol A	1.042	4.89	32.2	919	24.64
Sol B	1.034	5.12	30.4	980	24.71

6. Conclusions

The purpose of the presented analysis is to determine the optimal composition of the impregnation product. The focus of our optimization approach is twofold. On one hand, we want to define the mixture's composition that leads to a maximal impregnation depth for a specific pore size distribution of a material and given constraints related to other factors, such as maximum gelation time. On the other hand, our optimization methodology is designed to reduce the required number of time-consuming laboratory experiments while maintaining a good predictive capability.

The first part of the analysis consists of the development of a physically-based model that can predict the penetration depth and the uptake of the impregnation product from five physical parameters of the product and the material pore size distribution. The second part is concerned with the design of experiments where, from the initial factorial design, additional laboratory experiments are chosen stepwise to form a training set for training a (data-driven) Gaussian process (GP) model that predicts the product's physical parameters from its composition. New training points are selected iteratively, based on the largest predictive variance of the estimated physical parameters for the current GP, weighted according to parameter importance with respect to penetration depth. We increased the number of experiments (training points) incrementally by 6 in each step.

In the third step, our trained data-driven GP models were used to predict the physical parameters that are needed by our physics-based model to simulate penetration of the impregnation product. For our case, these physical parameters were determined for the mixture of three impregnation product components (precursor, catalyst and water). In many cases, nanoparticles are added to the impregnation products, but for the purpose of this work, we didn't consider them as this would introduce additional uncertainties. Namely, the presence of nanoparticles can have a significant effect on the sol properties, especially at larger concentration of nanoparticles. The nanoparticles change physical properties, such a viscosity and gelation time. Viscosity increase depends on particle content and their size and shape. Gel time is affected by acting as nucleation sites, modifying pH or adsorbing the catalyst. Moreover, because of nanoparticles size the penetration is not determined only by the changes in physical properties of the sol, but also by physical restriction because smaller may clog by the particles during the process. In the presented cases, we divided the parameter space of each component into 6 equally spaced values. This resulted in 216 observation points for each physical parameter. The physical parameters were used by the model to calculate the penetration depth.

Our results show we can see that the optimal mixture depends on the material pore structure. Hence, with a single product we cannot obtain the best penetration depth for all materials. However, we can find the mixture that makes a better compromise for all materials. Additionally, we demonstrated the ability of the proposed procedure to find the optimal product under specific user defined constraints. We considered constraining gelation time, which is relevant for practical applications. The results for the gelation time-limited cases highlight again that the optimal mixtures are different for different materials, but the differences in the penetration depths vary less compared to the cases with no limitation on the gelation time. For these non-limited gelation time cases, there is indeed a one order of magnitude difference in penetration depth between the considered fine pore and large pore systems. When the gelation time is restricted to a maximum of 6 or 12 h, the penetration depth in the fine and large pore matrix systems is almost the same.

The procedure has been validated by the blind prediction of the

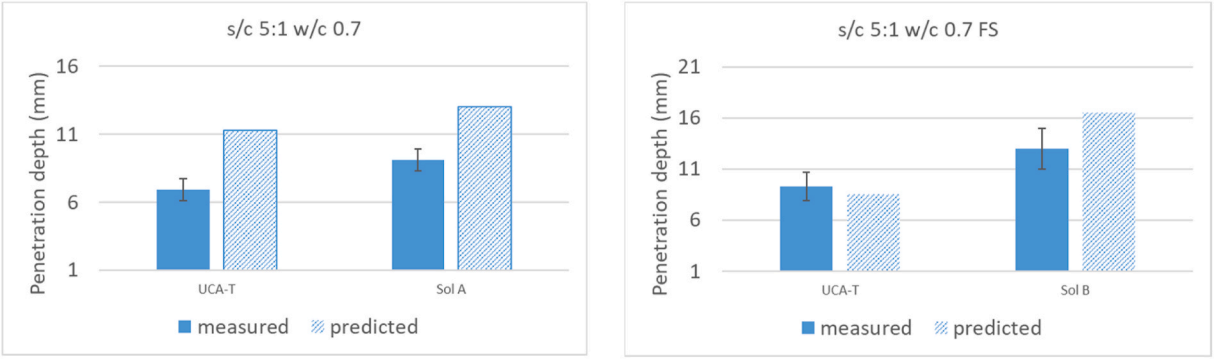

Fig. 10. Comparison between measured and predicted penetration depths: s/c 5:1 w/c 0.7 (left), and s/c 5:1 w/c 0.7 FS (right).

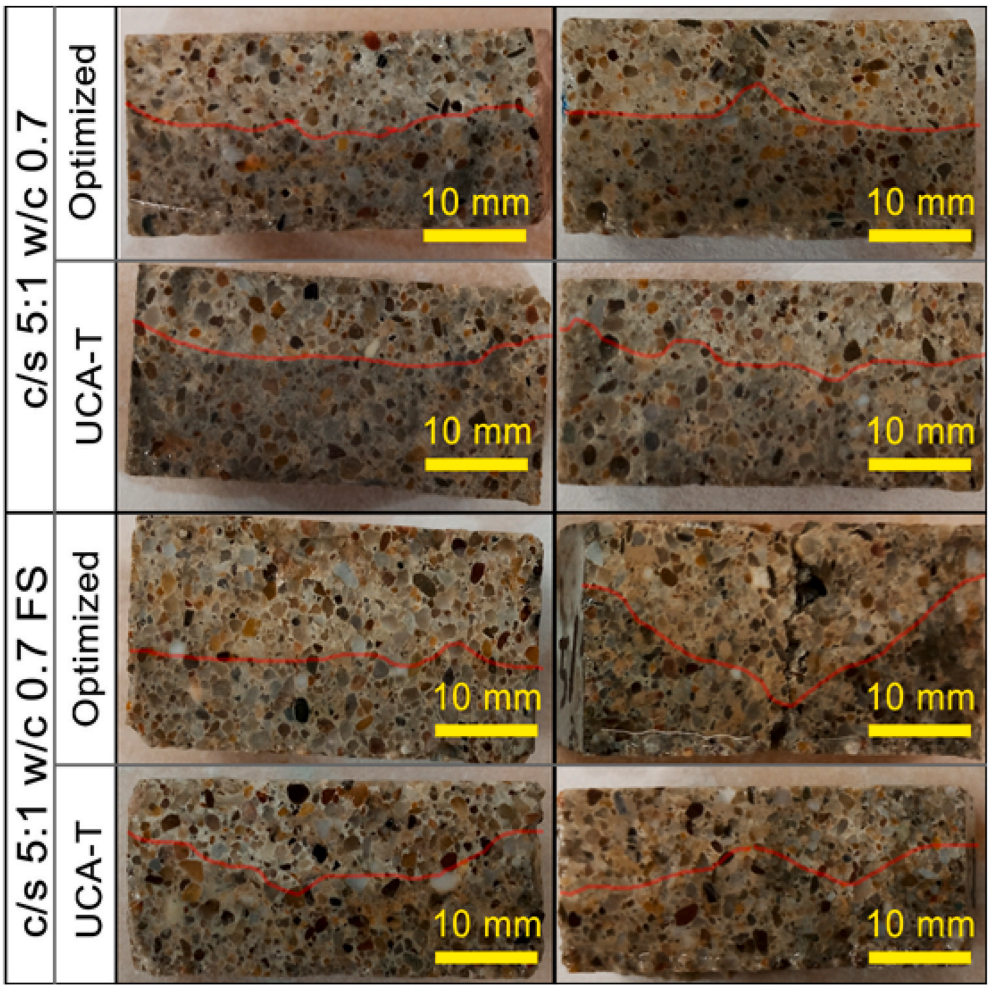

Fig. 11. Comparison of penetration depths for optimized and not optimized (UCA-T) product for two different materials. The red line denotes the penetration depth (the product is applied at the top). Left and right figures correspond to different replicates. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

optimal mix. The prediction of sol properties shows a very good agreement with the measured values for the density and viscosity. For surface tension and gelation time, the predicted values have quantitative differences, but the trend is consistent. In addition, the penetration depth is a bit overestimated, however, the trend is correctly captured and the optimized product performs considerably better.

Future research can be extended to account for other environmental variables that play an important role in the penetration depth, such as temperature and relative humidity. Then the proposed optimization workflow would apply to a wide range of climate conditions for which protection is required. In addition, the process can be improved to account for different types of impregnation treatments or substrates. One of the limitation of the current research is that for materials with a very specific pore shape, such as slit-shaped pores that are common in clay-based materials, the physical model to predict penetration depth, and even more the sol uptake, should be calibrated and additional variables may need to be considered.

The optimization in this work is done on Ordinary Portland Cement (OPC) only. This choice is based on the fact, that most of the civil structures and monuments are made by OPC based concrete. The purpose of this work is to propose a methodology and use an available set of experiments to prove the concept. The methodology described in this manuscript is not restricted to OPC and can be used, in principle, for blended cement, stones, brick or other building materials. For different materials, however, the interaction between sol and material can change sol-gel kinetics (gelation time) and/or contact angle. These differences may affect the exact penetration depth values, but the trends predicted by the methodology are expected to be similar because the trend depends on the pore size distribution of the material.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The work described in this manuscript has been performed under InnovaConcrete EC project, supported by funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant AgreementNº 760858. This work has been supported by the Flemish Government under the 'Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen'.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.

org/10.1016/j.cemconcomp.2022.104903.

References

[1] F. Glasser, J. Marchand, E. Samson, Durability of concrete Degradation phenomena involving detrimental chemical reactions, Cement Concr. Res. 38 (2) (2008) 226–246.

[2] K. Kovler, V. Chernov, Failure, distress and repair of concrete structures, in: Types of Damage in Concrete Structures, Woodhead publishing limited, Oxford, Cambridge, New Delhi, 2009, pp. 32–56.

[3] B. Zoran, G. Topličić Ćurčić, D. Nebojša, J. Savić, Damage of concrete and reinforcement of reinforced-concrete foundations caused by environmental effects, Procedia Eng. 117 (2015) 411–418.

[4] Z.J. Shiping Wei, L. Hao, D. Zhou, M. Sanchez-Silva, Microbiologically induced deterioration of concrete - a review, Braz. J. Microbiol. 44 (4) (2013) 1001–1014.

[5] R. Zarzuela, M. Luna, L. Carrascosa, M. Yeste, I. Garcia-Lodeiro, M. Blanco-Varela, M. Cauqui, J. Rodríguez-Izquierdo, M. Mosquera, Producing C-S-H gel by reaction between silica oligomers and portlandite: a promising approach to repair cementitious materials, Cement Concr. Res. 130 (2020), 106008.

[6] P. Maravelaki-Kalaitzaki, N. Kallithrakas-Kontos, D. Korakaki, Z. Agioutantis, S. Maurigiannakis, Evaluation of silicon-based strengthening agents on porous limestones, Prog. Org. Coating 57 (2006) 140–148.

[7] I. García-Lodeiro, R. Zarzuela, M. Mosquera, M. Blanco-Varela, Consolidation of artificial decayed portland cement mortars with an alkoxysilane-based impregnation treatment and its influence on mineralogy and pore structure, Construct. Build. Mater. 304 (2021), 124532.

[8] S.-S. Park, Y.Y. Kim, B.-J. Lee, B.-J. Lee, S.-J. Kwon, S.-J. Kwon, Evaluation of concrete durability performance with sodium silicate impregnants, Adv. Mater. Sci. Eng. 1 (2014) 1–11.

[9] X. Pan, Z. Shi, C. Shi, T.-C. Ling, N. Li, A review on concrete surface treatment Part I: types and mechanisms, Construct. Build. Mater. 132 (2017) 578–590.

[10] I. Garcia-Lodeiro, P. Carmona-Quiroga, R. Zarzuela, M.J. Mosquera, M. Blanco-Varela, Chemistry of the interaction between an alkoxysilane-based impregnation treatment and cementitious phases, Cement Concr. Res. 142 (2021), 106351.

[11] M. Horgnies, J.J. Chen, Superhydrophobic concrete surfaces with integrated microtexture, Cement Concr. Compos. 52 (2014) 81–90.

[12] T. Bader, B.J. Waldner, S.H. Unterberger, R. Lackner, On the performance of film formers versus penetrants as water-repellent treatment of High-Performance Concrete (HPC) surface, Construct. Build. Mater. 203 (2019) 481–490.

[13] E. Washburn, The Dynamics of capillary flow, Phys. Rev. 18 (1921) 273–283.

[14] S. Babaei, S. Seetharam, U. Muehlich, A. Dizier, G. Steenackers, B. Craeye, A multiscale framework to estimate water sorption isotherms for OPC-based materials, Cement Concr. Compos. 105 (2020), 103415.

[15] M. Bofeldt, B. Nyman, Penetration depth of hydrophobic impregnating agents for concrete/Eindringtiefe von Hydrophobierungsmitteln in Beton, Restor. Build. Monuments 8 (2–3) (2014) 217–232.

[16] A. Prabhu, J.S. Dolado, E.A.B. Koenders, R. Zarzuela, M.J. Mosquera, I. Garcia-Lodeiro, M.T. Blanco-Varela, A patchy particle model for C-S-H formation, Cement Concr. Res. 152 (2022), 106658.

[17] J. Perko, R. Zarzuela, I. Garcia-Lodeiro, M.T. Blanco-Varela, M.J. Mosquera, T. Seemann, L. Yu, The importance of physical parameters for the penetration depth of impregnation products into cementitious materials: modelling and experimental study, Construct. Build. Mater. 257 (2020), 119595.

[18] A.A. Issa, A.S. Luyt, Kinetics of alkoxysilanes and organoalkoxysilanes, MDPI Polymers 11 (537) (2019).

[19] D.S. Facio, M. Luna, M.J. Mosquera, Facile preparation of mesoporous silica monoliths by an inverse micelle mechanism, Microporous Mesoporous Mater. 247 (2017) 166–176.

[20] Z. Li, J. Yoon, R. Zhang, F. Rajabipour, W.V. Srubar Iii, I. Dabo, A. Radlińska, Machine learning in concrete science: applications, challenges, and best practices, npj Computational Materials 8 (1) (2022) 127.

[21] M. Ashraf, M.F. Iqbal, M. Rauf, M.U. Ashraf, A. Ulhaq, H. Muhammad, Q.-f. Liu, Developing a sustainable concrete incorporating bentonite clay and silica fume: mechanical and durability performance, J. Clean. Prod. 337 (2022), 130315.

[22] Q.-f. Liu, M.F. Iqbal, J. Yang, X.-y. Lu, P. Zhang, M. Rauf, Prediction of chloride diffusivity in concrete using artificial neural network: modelling and performance evaluation, Construct. Build. Mater. 268 (2021), 121082.

[23] A. Ahmad, F. Farooq, K.A. Ostrowski, K. Sliwa-Wieczorek, S. Czarnecki, Application of novel machine learning techniques for predicting the surface chloride concentration in concrete containing waste material, MDPI Materials 14 (2021) 2297.

[24] T. Santner, B. Williams, W. Notz, The Design and Analysis of Computer Experiments, Springer, 2003.

[25] K. Smith, On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations, Biometrika 12 (1918) 1–85.

[26] C. Rasmussen, C. Williams, Gaussian Processes for Machine Learning, MIT Press, 2006.

[27] D. Kristjanson Duvenaud, Automatic Model Construction, University of Cambridge, Cambridge, 2014.

[28] A.G. Matthews, M. van der Wilk, T. Nickson, K. Fujii, A. Boukouvalas, P.G.Z. Leòn-Villagrà, J. Hensman, GPflow: a Gaussian process library using TensorFlow, J. Mach. Learn. Res. 18 (40) (2017) 1–6.

[29] M. van der Wilk, V. Dutordoir, S. John, A. Artemev, V. Adam, J. Hensman, A Framework for Interdomain and Multioutput Gaussian Processes, 2020 *arXiv: 2003.01115*.