Composites Part B: Engineering
Available online 31 March 2021, 108801

Coupling machine learning with thermodynamic modelling to develop a composition-property model for alkali-activated materials

Xinyuan Ke[1*] Yu Duan[2]

1. Department of Architecture and Civil Engineering, the University of Bath, Bath BA2 7AY, United Kingdom
2. Department of Mechanical Engineering, Imperial College London, London, UK

Abstract

Alkali-activation is one of the most promising routes for utilisation of versatile aluminosilicate resources. However, the variations of chemical compositions in these resources have increased the challenge of designing alkali-activated materials (AAMs) with multiple sources, posing the demand for establishing composition-property correlations that can represent a wide range of AAMs. This study proposes a data-driven approach to develop such composition-property correlations combining machine learning with global sensitivity analysis and thermodynamic modelling. The strength performance of alkali-activated concretes was investigated for a benchmark study (196 data inputs). The impact of the five key chemical compositions, CaO-SiO_2-Al_2O_3-MgO-Na_2O, has been assessed. The results show that despite the use of different aluminosilicate precursors, there appear to be coherent connections between bulk binder chemical compositions, phase assemblages, and the performance of AAMs. The composition-property correlations established via machine learning can be used to facilitate the on-demand design of AAMs utilising varying aluminosilicate resources.

Keywords:

Alkali-activated materials, weighted Gaussian processes, Sobol global sensitivity indices, thermodynamic modelling, mechanical performances.

*Corresponding author. E-mail: x.ke@bath.ac.uk
Both authors contributed equally.

1. Introduction

The production of clinker cement currently takes around 25% of the global industry CO_2 emission, which is

expected to grow by further 7 % in 2050 under the as-is scenario [1], posing significant pressure on achieving the 2050 net-zero target set out by the Paris Agreement. Alkali-activated materials (AAMs) are non-clinker cementitious materials that have demonstrated great potential as sustainable and durable alternatives to Portland cement [2-4]. Despite some standing engineering challenges [5], the AAMs have demonstrated superior performances in specialised engineering applications, such as marine structures [6], sewage pipes [7], and even hazardous waste immobilisations [8]. The choice of AAMs over traditional clinker cement for these specialised areas can result in improved performances and longevity, and thus improving the sustainability and resource efficiency of infrastructures.

Alkali-activation is one of the most promising and efficient chemical routes that can drastically improve the resource efficiency by utilisation of versatile aluminosilicate resources, including natural minerals [9-11], waste and recycled materials [12-14]. The reaction kinetics, phase evolution, mechanical performances, and durability of alkali-activated materials are primarily controlled by the chemistry of the aluminosilicate precursors and the dosage and type of activators [2, 15, 16]. A wide range of aluminosilicate rich precursors can be used to prepare AAM, from high calcium content precursors (e.g. blast furnace slag), to intermedia and/or low calcium content precursors (e.g. fly ash, rice husk ashes, and metakaolin) [17-19]. Sodium based alkalis, including siliceous sodium silicate and non-siliceous sodium hydroxide and sodium carbonate, have been widely used as chemical activators due to their effectiveness of reaching desirable engineering performances and affordable costs [20-22].

However, the variations of chemical compositions in these aluminosilicate precursors have increased the challenge of designing AAMs with desirable performances. This has posed great demand for establishing composition-property correlations that can: 1) reflect the physical and chemical nature of the AAM binder system and relative importance of different chemical compositions; 2) estimate the performance of AAMs with known binder chemical compositions. Such understanding of the AAM system from the composition-performance

perspectives, independent of specific types of precursors used, is particularly important for providing essential guidance for performance-based on-demand materials design and optimisation of a wide range of AAMs.

Despite a large number of existing studies on the effects of the chemistry of precursors and/or activators on the performances of alkali-activated materials [15, 16, 19, 23-28], it is still very difficult to establish a general composition-property correlation for AAMs in a single model. This is because the reaction kinetics and microstructure developments in AAMs are extremely complex and vary when different precursors are been used [15, 29-31]. Kinetic-controlled reactions, such as the dissolution of the precursors, are determined by much more complex factors, such as surface areas and mineralogy, and should not be overlooked [32-34]. However, from the thermodynamic point of view, the bulk chemical compositions of the AAM binders play dominant roles in determining the phase assemblage and microstructure (e.g. density) of the AAM [35-37], which further affect the mechanical and durability performances of AAMs [28, 38, 39]. The thermodynamic modelling of the AAM binder systems can provide a reliable semi-quantified prediction of the phase assemblages, density, as well as durability performances in the AAM system for matured samples [35-37]. Similar thermodynamic modelling approaches have been proven effective for OPC binder systems [40]. This implies that there might be coherent connections between bulk chemical compositions, phase assemblages, and performance of AAMs, serving as the theoretical support of the composition-performance correlations. However, the number of experimental tests and materials characterisations required for verifying these hypotheses would increase exponentially with the increasing number of chemical compounds investigated. It is sometimes impractical to obtain sufficient data from single laboratory-based research to perform rigid statistical parametric analysis.

In order to tackle these challenges, this paper proposes the use of a data-driven approach to developing such composition-property model for alkali-activated materials. The data-driven approach has gained increasing interest in recent years to establish composition-performance correlations for understanding materials performances and developing new materials [41-45]. Machine learning models (such as support vector machines, artificial neural networks, and Gaussian processes models) , trained and validated using experimental-based observations, can

serve as the surrogate models for describing the relationships between inputs (e.g. concrete ingredients) and outputs (e.g. mechanical or durability performances) [46-50]. Then, sensitivity analysis of the model inputs can be performed using the machine learning model to identify the significance of input variables [50-52].

Sensitivity analysis includes both local sensitivity analysis (around fixed data point) and global sensitivity analysis (full data input ranges). The variance-based sensitivity analysis (Sobol indices)[53] is particularly useful for global sensitivity analysis, in order to identify and quantify input(s) whose uncertainty has/have the largest impact on the variability of the model output. The impact of interactions of two or more inputs on the model output can also be quantitively evaluated through second or higher-order Sobol indices. By identifying the high impact (sensitive) input variables, the sensitivity analysis would also enable further reduction of the model complexity. The machine learning assisted sensitivity analysis can also be used to assess whether the data-driven model follows the physical or chemical roles of the simulated system [50, 54]. Details of the mathematical background of the Sobol indices can be found in section 2.2.

In this study, a composition-strength model for AAMs is investigated as a benchmark. A Gaussian processes (GP) model is trained using organised dataset from existing literature to predict the compressive strength of alkali-activated concrete (AAC) from binder composition and concrete mix design (section 4.1.). Global sensitivity analysis is performed on the trained GP model using Monte Carlo sampling methods for analysing the impact of chemical compositions and mix design on the strength performance of AAC, suggesting that the trained machine learning model follows the physical and chemical law of the studied binder system (section 4.2.). The proposed composition-property model is further assessed via linking the machine learning model predicted strength performance with binder phase assemblages, evaluated through thermodynamic modelling from bulk binder composition (sections 4.3.). The application and possible future improvement of this proposed composition-property framework coupling machine learning with thermodynamic modelling are also discussed.

2. Methodology

In this section, we will give a brief introduction to the Gaussian processes, variance-based global sensitivity analysis (Sobol indices), and the thermodynamic modelling.

2.1.Gaussian processes (GP)

Gaussian processes (GP) is a nonparametric Bayesian machining learning method. which can be viewed as a one-layer feed-forward Bayesian neural networks with an infinite number of hidden units [55]. One unique feature of GP is that there is no predefined model structure, such as classes of parametric functions, for describing the relationships between the input and output datasets, which has the benefit of achieving more flexible predictive models [56]. The mathematical background of the GP is briefly summarised as following. For readers who are interested in more details about Gaussian processes, please refers to [57].

2.1.1. Mathematical definition

GP is designed to estimate the y^* (unknown value) at a new input $\vec{x}^*$ (known value), given the observations (training datasets) $\vec{y} = \{y_i\}_{i=1}^n$ at n sets of the input $X = \{\vec{x}_i\}_{i=1}^n$, where $y_i \in R$ is an observed function output due to the i[th] input $\vec{x}_i \in \mathrm{R}^{\mathrm{D}}$. In particular, X is a $n \times D$ matrix and y_i is subject to the observational $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$. Predicting y^* can also be expressed as finding the most plausible $y^*$with the knowledge of $\vec{y}$. Therefore, the problem becomes

$$\underset{\vec{x} \in A}{\operatorname{argmax}} P(y^* | \vec{y}) \qquad \text{Eq. 1}$$

According to the Bayes' theorem, $p(y^*|\vec{y})$ in Eq. *1* can be written as

$$P(y^* | \vec{y}) = \frac{P(\vec{y}|y^*)P(y^*)}{P(\vec{y})} \qquad \text{Eq. 2}$$

GP assumes that any finite number of a collection of outputs of a latent function follow a multivariate Gaussian distribution [57-59]. Hence:

$$\vec{y} \sim \mathcal{N}(\vec{m}, K) \qquad \text{Eq. 3}$$

in which $\vec{m}$ represents the prior mean of $\vec{y}$. K is the covariance matrix of $\vec{y}$. The covariance matrix essentially describes the data structure and allows the data to determine the form of the latent function. The element in K are calculated using the kernel function $k(\vec{x_i}, \vec{x_j})$. In this study, the $\vec{y}$ is the composite strength with respect to the different composite design. Meanwhile $\vec{x_i}$ stands for the composite design which lead to the strength y_i. For notational

simplicity, $\vec{m}$ values are normally set to zeros. The GP with the prior mean values as zeros is also called as the simple kriging in some contexts. Since the mean of the posterior process is not confined to zero, this simplification is not necessarily a limitation [57]. The measured data $\vec{y}$ with an unknown y^* at a corresponding location $\vec{x}^*$, also forms a multivariate Gaussian distribution, which can be written as Eq. *4*.

$$\begin{bmatrix} \vec{y} \\ y^* \end{bmatrix} \sim \mathcal{N}\left(\vec{0}, \begin{bmatrix} K & (K^*)^T \\ K^* & K^{**} \end{bmatrix}\right) \quad \text{Eq. 4}$$

In Eq. 4, ^{T} denotes the transpose of the vector or matrix. K^* is a $1 \times n$ matrix containing the covariance between y^* and y_i, and K^{**} is the self-covariance of y^*. The element in K^* and K^{**} are also calculated using the kernel function. Subsequently, $P(y^* | \vec{y})$ is:

$$P(y^*|\vec{y}) = \mathcal{N}(K^*K^{-1}\vec{y}, K^{**} - K^*K^{-1}(K^*)^T) \quad \text{Eq. 5}$$

2.1.2. Hyperparameters determination

As discussed in 2.1.1, the reliability of a GP model depends on the form and hyperparameters of the selected covariance function. The design of the covariance function is an active area of research in the GP research community. More about covariance functions can be found in [57]. In this study, the squared exponential (SE) covariance function (kernel function) is chosen as it generally performances well [60].

With the form of $k(x_i, x_j)$ determined, the problem then becomes finding the most plausible $\vec{\theta}$ with the knowledge of $\vec{y}$ and X. According to the Bayes' theorem, the posterior over the hyperparameters is defined as

$$p(\vec{\theta}|\vec{y}, X) = \frac{p(\vec{y}|X, \vec{\theta})p(\vec{\theta})}{\int p(\vec{y}|X, \vec{\theta})p(\vec{\theta})d\vec{\theta}} \quad \text{Eq. 6}$$

, where $p(\vec{\theta})$ is the prior of the hyperparameters and $p(\vec{y}|X, \vec{\theta})$ is the marginal likelihood. The fraction $\int p(\vec{y}|X, \vec{\theta})p(\vec{\theta})d\vec{\theta}$ is a normalisation constant. To avoid strong bias, $p(\vec{\theta})$ has to be a flat function over the domain. The maximum of $p(\vec{\theta}|\vec{y}, X)$, is effectively determined by $p(\vec{y}|X, \vec{\theta})$ when $p(\vec{\theta})$ is sufficiently flat [60]. Hereby, determination of the robust $\vec{\theta}$ becomes $argmax\left(p(\vec{y}|X, \vec{\theta})\right)$. In practice, the logarithm of $p(\vec{y}|X, \vec{\theta})$, Eq. 7, is usually maximised.

$$\log p(\vec{y}|X,\vec{\theta}) = -\frac{1}{2}\vec{y}^T K^{-1}\vec{y} - \frac{1}{2}\ln|K| - \frac{n}{2}\ln 2\pi \quad \text{Eq. 7}$$

, where n is the number of observations. Optimisation algorithm is then applied to determine the robust set of $\vec{\theta}$. In this study, the adaptive Nelder–Mead algorithm [61] is applied to identify the optimal hyperparameters.

2.1.3. Weighted GP model

The accuracy of a GP model is purely relying on the hyperparameters of $\vec{y}$ once the form of the kernel function is determined. In other words, a set of hyperparameters defines a unique GP model. It is well known that most of the optimisation algorithms may end at a local maximum due to the improper initial points and may cause overfitting or under-fitting. To mitigate this effect, we randomise the initial points and run the adaptive Nelder-Mead search for N times. We only select the N' sets of hyperparameters corresponding to the N' largest $\ln p(\vec{y}|\vec{x},\vec{\varphi})$ values. Based on the marginal likelihood calculated using the N' largest sets of hyperparameters, the weight of the i[th] GP model, defined by $\vec{\theta}_i$, can be calculated using

$$w_i' = \frac{p(\vec{y}|X,\vec{\theta}_i)p(\vec{\theta}_i)}{\sum_{i=0}^{N'} p(\vec{y}|X,\vec{\theta}_i)p(\vec{\theta}_i)} = \frac{p(\vec{y}|X,\vec{\theta}_i)}{\sum_{i=0}^{N'} p(\vec{y}|X,\vec{\theta}_i)} \quad \text{Eq. 8}$$

predicted by the GP model with $\vec{\theta}_i$.

$$E(y^*) = \sum_{i=0}^{N'} w_i'\alpha_i \quad \text{Eq. 9}$$

In this study, a thousand adaptive Nelder-Mead searches with random initialisation points are performed. In each search, it stops when the difference between the maximum and minimum in the search is smaller than 10^{-3}. The hyperparameters contributing to the first 25 largest marginal likelihoods have been selected, contributing to 99.9% of the total weights.

2.1.4. Leave-one-out-cross-validation (LOO-CV)

Since the simplifications adopted in the concept, such adjoint gaussian distribution assumption and improper hyperparameters may lead to a strongly biased machine learning model. Before applying the trained model in other usage, it is vital to check the validity of the model. In this study, the leave-one-out cross-validation (LOO-CV) test is adopted to check the

validity of the weighted GP model. In LOO-CV, the training dataset is divided into two groups, namely the LOO-CV training data $\vec{y}_{-i} = \{y_1, \cdots y_{i-1}, y_{i+1}, \cdots y_n\}$ and LOO-CV validation data y_i. The weighted GP model is then applied to predict the y_i. The integrity of the emulator is then qualitatively checked using graphical comparison (quantity-to-quantity plot and error histogram) and quantitatively assessed using the following metrics:

Root Mean Square Error (RMSE):

$$RMSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}_i^*)^2 \quad \text{Eq. 10}$$

Pearson Correlation Coefficient (ρ)

$$\rho = \frac{cov(\vec{y}, \vec{y}^*)}{\sigma_y \sigma_{y^*}} \quad \text{Eq. 11}$$

, where $cov(\cdot,\cdot)$ is the covariance of two vectors, σ_y and σ_{y^*} stand for the standard deviations of $\vec{y}$ and $\vec{y}^*$, respectively.

2.2. Global sensitivity analysis

In this study, a variance based Sobol global sensitivity analysis methods [53] is used to 1) measure the overall importance of each variables in the predictive surrogate model (total effect); 2) measure the relative importance of the five key chemical compositions and their interactions. The mathematical background can be summarised as follows.

Any given integrable model can be described as $f(\vec{x})$, where $\vec{x} = (x_1, \cdots x_n)$ is the n-dimensional input variable. $f(\vec{x})$ can be decomposed into 2^n orthogonal functional terms as:

$$f(\vec{x}) = f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{1 \le i \le j \le d} f_{ij}(x_i, x_j) + \cdots + f_{1,2,\cdots,n}(x_1, \cdots x_d) \quad \text{Eq. 12}$$

Due to the orthogonality of the components in Eq. 12, the variance of $f(\vec{x})$ can be written as follows:

$$V = \sum_{i=1}^{n} V_i + \sum_{1 \le i \le j \le n} V_{i,j} + \cdots + V_{1,2,\cdots,d} \quad \text{Eq. 13}$$

, where V_i is the partial variance due to x_i whilst others are interactive variance contributions due to different combination of inputs.

The Sobol indices (s_u) are defined as the relative variance to the total variance of the model

$$s_i = \frac{V_i}{V}, s_{ij} = \frac{V_{ij}}{V}, \cdots, s_{1,2,\ldots,n} = \frac{V_{1,2,\ldots,n}}{V} \quad \text{Eq. 14}$$

, where s_i is the first-order Sobol sensitivity index (the main effect sensitivity index), s_{ij} is the second-order Sobol sensitivity index, and $s_{1,2,\ldots,n}$ is the n-th order Sobol sensitivity index. Higher Sobol index value means that the corresponded variance(s) has/have higher contribution to the total variance. For each individual input assessed, the total effect sensitivity index is the total sum of the first to the n-th order Sobol indices involving this variable,

$$s_i^t = \frac{T_i}{V} = \frac{\sum_{\{u: i \in u\}} V_u}{V} \quad \text{Eq. 15}$$

, where $\{u: i \in u\}$ means any subset of $\{1, 2, \cdots, d\}$, which includes i in it. As suggested in the definition, the total effect index measures the contribution of variance x_i to the output, including all variance caused by its interactions with any other variables. Both the main and the total effective indices are portions of the output variance that can be attributed to the influence of a single input alone.

According to the definition, calculation of the exact Sobol indices require computing several integrals, which can sometimes be impractical. For this reason, the Monte Carlo estimation is used in this framework, and the Sobol indices defied in Eq. 14 and Eq. 15 are approximated using the estimation proposed by Owen [62]. Details of the mathematical deduction has been explained in [50]. To estimate the Sobol indices of each inputs of the GP surrogate model, two sets of inputs each containing 100,000 pairs of inputs are generated. The element in each set of inputs is sampled uniformly using the latin hypercube sampling (LHS) method within the related data range (Table 1).

2.3. Thermodynamic modelling

A thermodynamic modelling programme based on Gibbs energy minimisation was used to predict the stable phases that may be expected to form as reaction products in the alkali-activated binder system. The open source software GEM-Selektor v.3 (http://gems.web.psi.ch/GEMS3/) is used to perform the modelling. An extended Debye–Hückel equation is used for calculating the activity coefficients of the aqueous species, Eq. 16:

$$\log_{10}\gamma_i = \frac{-Az_i^2\sqrt{I}}{1+Ba\sqrt{I}} + bI + \log_{10}\frac{x_{iw}}{X_w} \quad \text{Eq. 16}$$

where γ_i is the activity coefficient, z_i is the charge of the i^{th} aqueous species, and A ($kg^{0.5}mol^{-0.5}$) and B ($kg^{0.5}mol^{-0.5}cm^{-1}$) are the temperature- and pressure-dependent electrostatic parameters. It was assumed that the aqueous phase is dominated by NaOH, so the average ion size and parameter for common short-range interactions of charged species (a and b as shown in Eq. 16) are 3.31 Å and 0.098 $kg \cdot mol^{-1}$ [63]. The term x_{iw} represents the molar quantity of water, and X_w is the total molar amount of the aqueous phase. I is the total ionic strength of the aqueous solution, which is calculated using Eq. 17.

$$I = \frac{1}{2}\sum c_i \cdot z_i^2 \quad \text{Eq. 17}$$

where c_i is the concentration of the i^{th} ionic species and z_i is its charge.

The recently published Cemdata18 [64] was used as the main chemical thermodynamic database. The solid solution models for alkali-substituted calcium aluminate silicate hydrate gels (CNASH_ss) and for hydroxylated hydrotalcite (MgAl-OH-LDH_ss) were used, the thermodynamic properties of these end members are described in detail in [35, 65]. The formation of quartz was excluded in the calculations due to their slow formation kinetics under ambient conditions [66]. In addition to the Cemdata18 database, the recently published zeolite20 database [67, 68] was also included in this study in order to represent the alkali aluminosilicate gel (N-A-S-H) formed in AAM with different Si/Al ratios. Crystallised zeolitic phases are not observed experimentally in most of the AAMs. However, instead of crystallinity and framework structures, the bulk chemical compositions (mainly Si/Al ratios and extra-framework cations) play primary roles in determining the Gibbs energy of formation of framework aluminosilicate minerals [69, 70]. Therefore, it is reasonable to interpret the entirety of zeolites as representing the N-A-S-H gel of different compositions. The same approximation treatment has been adapted in a recent study [37], where the exclusion of mordenite (both Na

and Ca endmembers) zeolite was also justified. A full degree of reaction model is used in all simulations for simplification, where 1.0 g of dry air is added in all simulations as the gas atmosphere. All simulations were performed at 25 °C under standard atmospheric pressure.

3. Data structure

The original data used to train the composition-strength predictive model are summarised from 18 published research [71-88]. The alkali-activated concrete mix designs and the corresponded compressive strength involving the three most extensively studied aluminosilicate precursors, BFS, fly ash, and metakaolin are included in this study. Five major elements in their oxide forms, CaO, SiO_2, MgO, Al_2O_3, and Na_2O, are been investigated in this study, which representing the bulk chemical compositions without distinguishing the crystalline and amorphous phases (such information is largely unavailable from the literature). Figure 1 shows the distribution of these five main chemical compositions in the precursors included in this study. It suggests that BFS is the main CaO and MgO source while all precursors contain a significant amount of SiO_2 and a considerable amount of Al_2O_3. The precursors themselves contain a negligible amount of Na_2O.

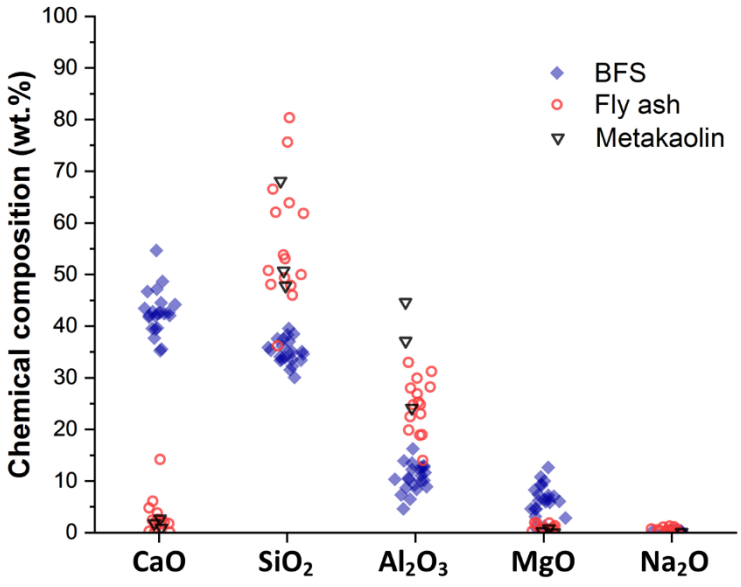


Figure 1 Distribution of CaO-SiO_2-Al_2O_3-MgO-Na_2O chemical compositions of BFS, fly ash and metakaolin precursors assessed in this study.

The chemical compositions of the binder in each mix design, which includes the precursor and the anhydrous activator, are recalculated from the bulk chemical composition of the precursor and dosage of the activator. It should be noted that only samples prepared using the sodium hydroxide and/or sodium silicate activators are included in this study. In order to investigate if the complex alkali-activated materials system can be reasonably represented by the simplified composition model, the silicate content from both the precursors and the activator are been added up as the total SiO_2 content. Since the precursors included in this study contain only a negligible amount of alkalis, the Na_2O content is solely contributed by the activators. The total water content (free water plus water from activators), fine aggregate, and coarse aggregate content are also included. Table 1 shows the data ranges of concrete design summarised in this study by binder chemical compositions in the unit of density (kg/m^3). A total of 196 data points is used in this study for training the composition-strength predictive model in section 4.1, each contains 9 variables. Since the GP model assumes that input variables are independent, the density of each chemical composition is used as the input variable rather than the mass percentage (of the total binder). The distribution of chemical compositions as the mass percentage of the total binder of the training data set can be found in the Supporting Information (Figure S-1) document. Sample curing temperature was not taking into consideration as training parameters as 175 out of the total 196 data points were cured under ambient conditions (20-25 °C).

Table 1 Data structures and ranges of the alkali-activated concrete mix design assessed in this study as summarised by bulk density. The natural logarithm values of testing age are used as training data.

	Chemical compositions		
	Unit	**Min**	**Max**
CaO	kg/m^3	1.2	227.0
SiO_2	kg/m^3	121.6	410.3
MgO	kg/m^3	0	59.6
Al_2O_3	kg/m^3	27.6	130.0
Na_2O	kg/m^3	11.4	63.8
Water	kg/m^3	83.5	263.1
Fine aggregate	kg/m^3	500.0	1247.0

Coarse aggregate	kg/m^3	415.0	1221.2
Testing age	ln(day)	3.0	365.0
Comp. Str.	MPa	4.9	94.0

4. Results and discussion

4.1. Training and validation of GP model

The LOO-CV predictions of the weighted GP model is plotted against the training data in Figure 2A. The results show that the predicted values are close to the equilibrium line (the solid line with slop equals to 1.0), where the Pearson correlation (ρ) between the LOO-CV prediction and the training data is 0.965. The root mean square error (RMSE) of the prediction is 5.80 MPa, representing the general accuracy of the trained GP model. The distribution of error percentage of the LOO-CV results, defined as {LOO-CV prediction – training data}×100/training data, is shown in Figure 2B. The results indicate that 181 out of 196 of the LOO-CV prediction results are within the ±25% error range, accounting for 92.3% of LOO-CV prediction results. Figure 2C presents the histograms and the fitted probability distribution functions (PDFs) of the training data and the LOO-CV prediction. The close alignment of the fitted PDFs suggests that the statistical characters of the trained GP model are akin to the training data set.

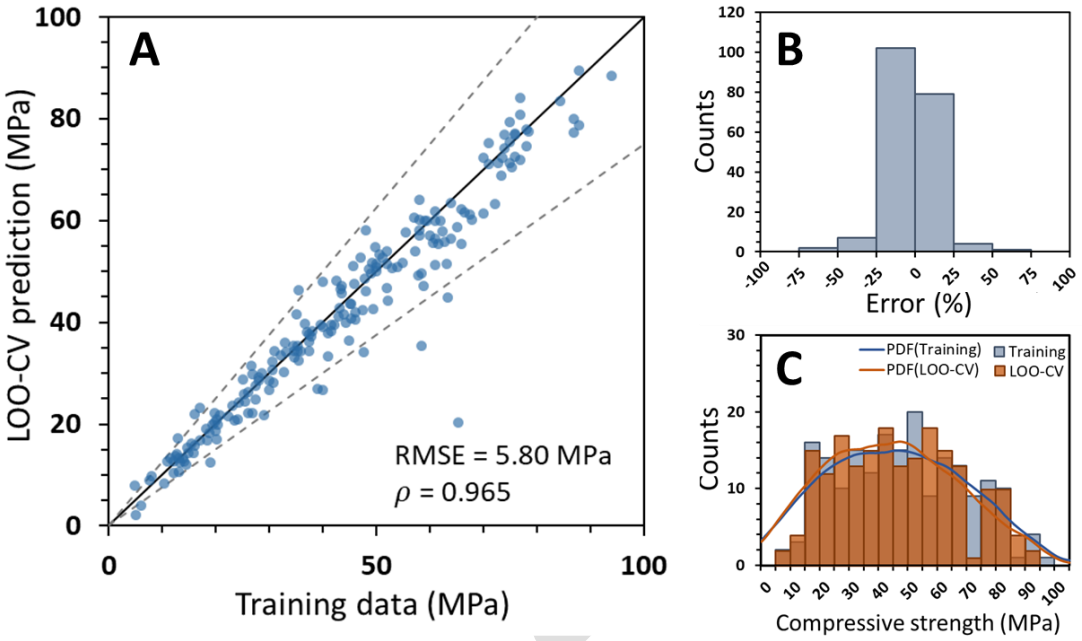


Figure 2 (A) Leave-one-out-cross-validation (LOO-CV) results; (B) distribution of errors (%) of the LOO-CV results; (C) probability density distribution (PDF) comparison of the training data compressive strength and the LOO-CV predicted values.

4.2. Global sensitivity analysis: Sobol indices

Figure 3 presents the total effect, first-order and second-order Sobol indices of the bulk binder chemical compositions, water content, fine and coarse aggregate, and age (in natural logarithm values) with regards to the compressive strength of the AAC. Figure 3(top) shows that SiO_2 has the highest total effect, followed by the coarse aggregate, CaO, water, Na_2O, fine aggregate, MgO, age, and Al_2O_3 content. The total effect Sobol indices of coarse aggregate, CaO and water are very close to each other, while the total effect Sobol indices of the fine aggregate, MgO, age, and Al_2O_3 content are much lower in comparison with the highest five factors. However, the first-order and second-order Sobol indices, Figure 3(bottom), show different trends comparing to the total effect values. SiO_2 has the highest first-order Sobol index, followed by the coarse aggregate, water, and Na_2O content. The CaO content has none first-order effect, instead, the second-order interactive effect between CaO and SiO_2 is very significant.

Since the total effect index is the total sum of the first to the n-th order Sobol indices involving the assessed variable, the results suggest that for the trained GP predictive model, SiO_2 content, out of the five major chemical compositions investigated, has the most dominant effect on the predicted strength performances, followed by the CaO and the Na_2O content. The relative impacts of different chemical compositions have not been assessed quantitatively from experiments. However, the relative importance of variables assessed here is consistent with experimental observations. It has been observed from various experimental studies that the CaO content in the precursor is crucial for determining the strength given phase sodium and aluminium-substituted calcium silicate hydrate (C-(N)-A-S-H) gel [30, 73, 79, 89], corresponding to the high CaO total effect and significant CaO and SiO_2 interactive effect. The SiO_2 content in the training data set includes both SiO_2 content from the precursor and the activator. In general, the use of sodium silicate activators can result in a denser microstructure and higher strength performances in AAMs [24, 90], corresponding to the high first-order Sobol index of SiO_2 and the relatively significant interactive effect between Na_2O and SiO_2. Apart from the five major binder chemical compositions, the significant total effect of water content corresponds to its importance as reaction media and affecting the overall porosity [91]. Between fine aggregate and coarse aggregate, the coarse aggregate has a much higher total and first-order effect. This is consistent with the average paste thickness (APT) theory, where the largest diameter size of the aggregate has the most dominant effect on the strength performance of concrete [92, 93]. In additions, it is important to note that since the natural logarithm values of sample ages are used in the training data, the Sobol indices relating to sample ages reflect the impact of age in logarithm values instead of actual sample ages. Nevertheless, the global sensitivity analysis of the trained GP model for predicting the strength performances of AAC from binder chemical compositions suggests that the trained surrogate model follows the basic physical and chemical rules in the AAC system.

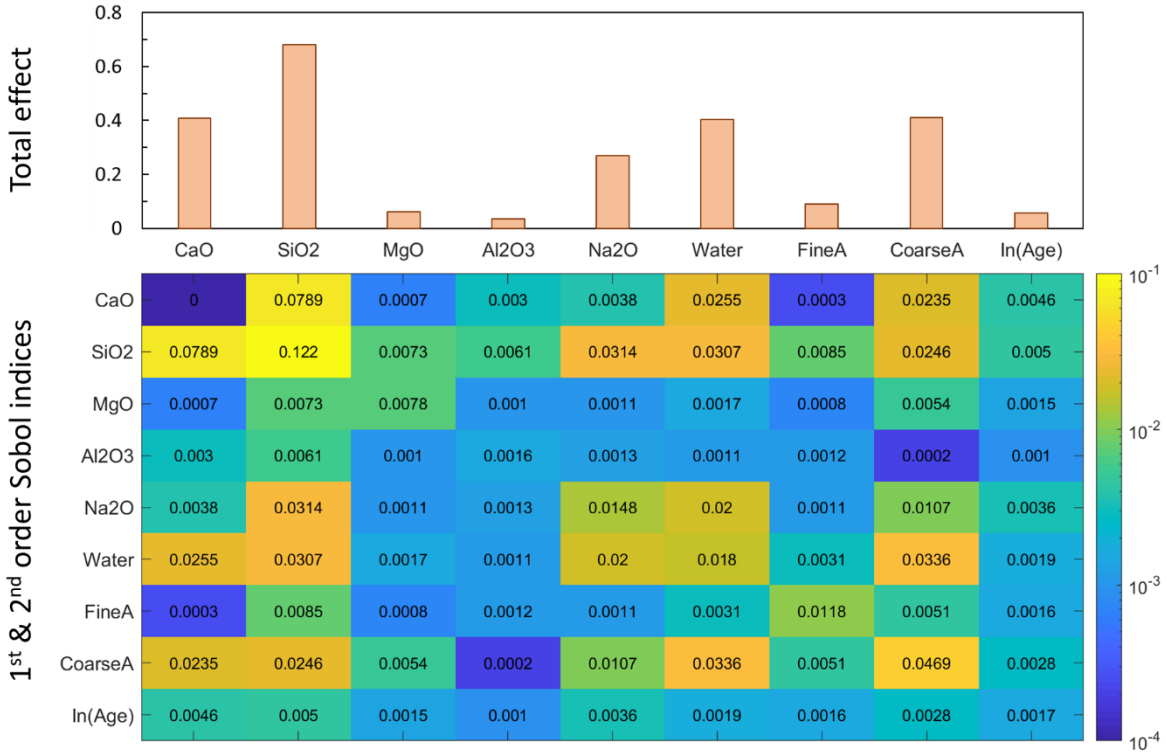


Figure 3 The total effect (top), and 1st and 2nd order Sobol indices of the predictors in the machine learning surrogate model for predicting the compressive strength of alkali-activated concretes (Note: $ln(Age)$ is the natural logarithm of 'Age').

4.3. Coupling GP model with thermodynamic modelling

In order to further assess the validity of the proposed composition-strength model, parametric studies of the five major chemical compositions are performed. A Monte Carlo sampling process is used to generate AAC binder compositions within the training data ranges while fixing the water to binder ratio, aggregate contents, and the sample age as constants. A total binder content of 450 kg/m^3 is assumed, which is approximately the average binder content of the training data set. A total of 3803 AAC binder chemical compositions are generated. A typical water to binder ratio of 0.4 is used, resulting in 180 kg/m^3 total water content. The fine and coarse aggregate contents are chosen as fixed values, 800 kg/m^3 and 1000 kg/m^3. The compressive strength of these simulated AAC mix design (by binder chemical compositions) are predicted using the validated GP model from section 4.1. The results of the 28 days strength performance are shown in Figure 4. The results of AAC with different water to binder ratios and different fixed aggregate content are included in the Supporting

Information (Figure S-2), demonstrating that the fixed factors do not cause qualitative changes to the trend of results discussed in this session.

Figure 4 suggests that out of the 3803 simulated AAC mix design by chemical composition, the highest AAC strengths are been predicted within two binder composition ranges:

(I) CaO content 36-44 wt.%, SiO_2 content 38-44 wt.%, MgO content 0-3 wt.%, Al_2O_3 content 9-17 wt.% and Na_2O content 3-6 wt.%. The binder compositions in this range have the bulk Ca/Si elemental ratio around 1.0, and Si/Al ratio between 2.0 to 4.5.

(II) CaO content 1-3 wt.%, SiO_2 content 67-71 wt.%, MgO content 0-2 wt.%, Al_2O_3 content 17-23 wt.% and Na_2O content 5-11 wt.%. The binder compositions in this range have the bulk Si/Al ratio between 2.5 to 3.5.

For binder compositions within the range (I), the C-(N)-A-S-H phase is observed experimentally as the dominate reaction product [34, 38, 94]; while for binder compositions within range (II), the N-A-S-H phase is observed experimentally as the dominate reaction product [95-97]. Both C-(N)-A-S-H and N-A-S-H phases are recognised as the main strength given phases in alkali-activated materials [2]. As suggested by Figure 4A, it appears that AACs containing a mixed of C-(N)-A-S-H and N-A-S-H phases result in lower strength performance when comparing with AAC dominated by either of these two phases. Figure 4B shows that higher Na_2O content does not seem to favour the high-Ca binder system, while moderate amount of Na_2O content is required to achieve higher strength in the low-Ca binder system. In addition, higher strength performance is predicted in low MgO content mix designs. However, under the given results, it is still unclear whether the composition-strength correlation is also consistent with the detailed physiochemical properties of the binder system, such as the formation of secondary phases, changes in the bulk compositions in the C-(N)-A-S-H and N-A-S-H phases, and the bulk porosity. Further assessment of the composition-strength correlation is conducted via coupling with thermodynamic models.

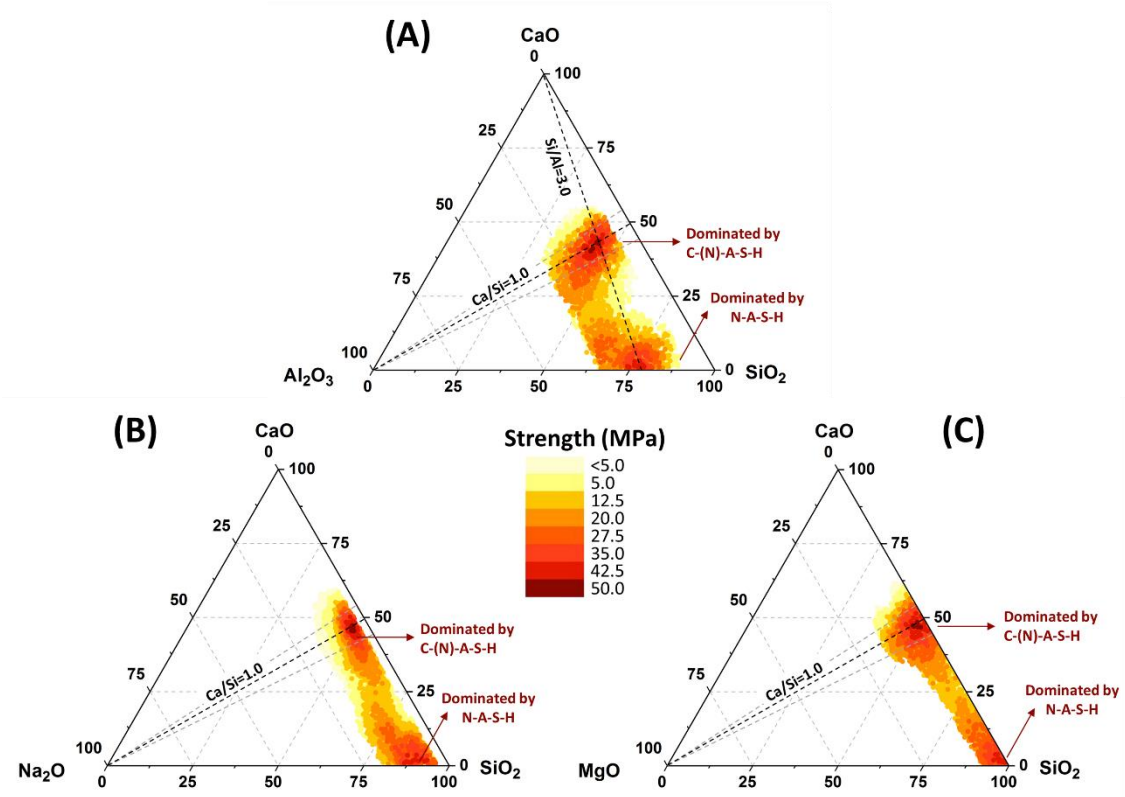


Figure 4 Distribution of 3803 data points (chemical composition and corresponded compressive strength) presented in ternary diagrams, (A) CaO-Al_2O_3-SiO_2, (B) CaO-Na_2O-SiO_2, (C) CaO-MgO-SiO_2. The grey dash line on both sides of the Ca/Si=1.0 (molar ratio) indication lines representing the bulk Ca/Si ratio of 0.8 and 1.5.

4.3.3. The role of CaO-SiO_2-Al_2O_3 binder composition

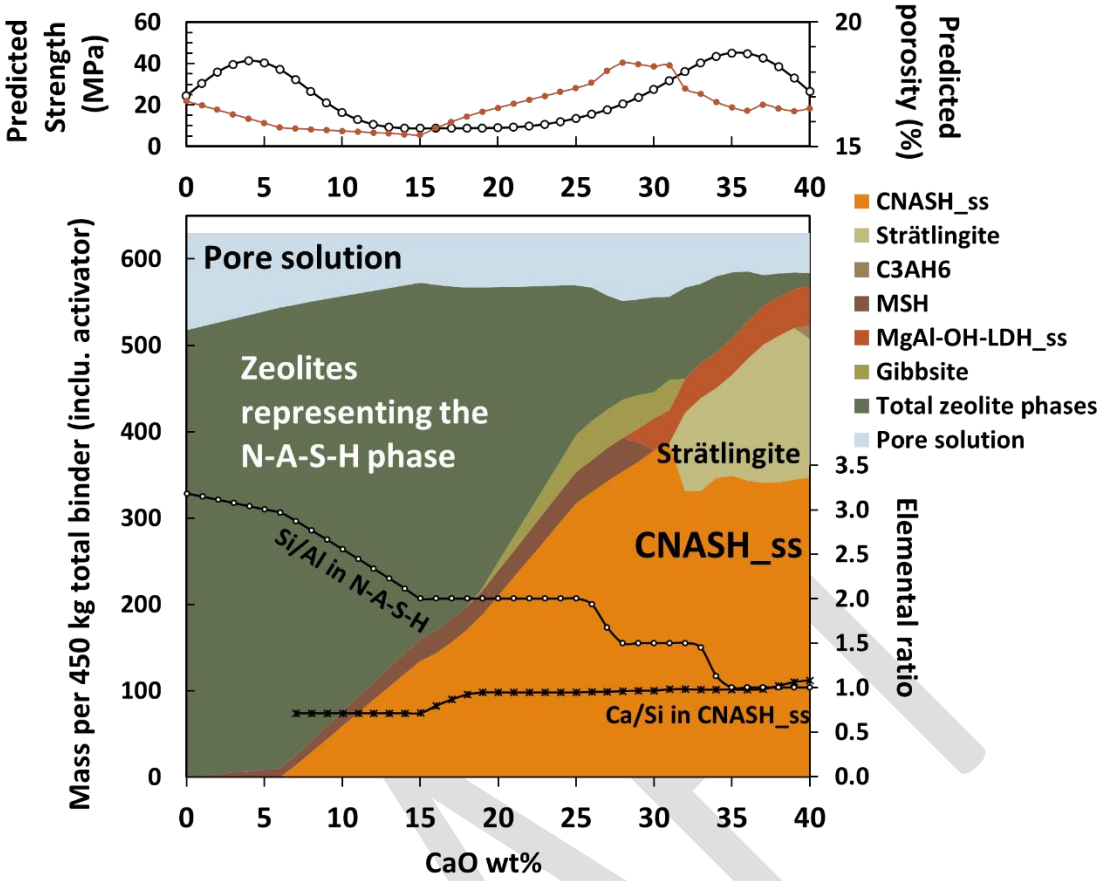
Figure 5 shows the phase assemblages of AAM binder system predicted using thermodynamic modelling via GEMs, where CaO wt.% in the total binder ranges from 0 wt.% to 40 wt.%, SiO_2 wt.% and Al_2O_3 wt.% decreases uniformly from 75 wt.% to 35 wt.% and 20.0 wt.% to 16.0 wt.% respectively. The MgO wt.% is set to be 10% of corresponded CaO content, while Na_2O wt.% is fixed at 5 wt.%. These composition ranges are been setup to reflect the role of CaO-SiO_2-Al_2O_3 compositions on the phase assemblages, corresponding to Figure 4A. The predicted phase assemblages show that N-A-S-H gels (represented by the zeolite phases) and a very small fraction of magnesian silicate hydrate (M-S-H) are been predicted at low CaO content (high SiO_2 and Al_2O_3 content) region. The formation of M-S-H is due to the presence of a trace of MgO setup by the model. However, since MgO content is barely observed from metakaoalin and fly ash, the M-S-H gel is not identified experimentally. The C-(N)-A-S-H gel

is not been predicted to form until the CaO content is above 6 wt%. Instead, zeolites with Ca as extra framework cations are been predicted at low CaO binder content. This is consistent with experimental observations that Ca can be accommodated in the amorphous aluminosilicate phase through ion-exchange [98, 99]. The total mass and bulk Si/Al ratio of the predicted N-A-S-H phase start to decline as C-(N)-A-S-H gel starts to form and consumes Si. Mg-Al type layered double hydroxide (hydrotalcite) and Ca-Al type layered double hydroxide (strätlingite) are been predicted when the binder CaO content is above 30 wt%, together with the C-(N)-A-S-H gel (bulk Ca/Si ratio around 1.0) as dominate reaction product and a small fraction of N-A-S-H gel (bulk Si/Al around 1.0 to 1.5), corresponding to the phase assemblage of typical alkali-activated slags [100, 101] and alkali-activated slag with a small amount of metakaolin [38, 102]. Figure 5B illustrates the predicted zeolites, chabazite (Na-endmember: $Na_2(Al_2Si_4)O_{12}(H_2O)_6$; Ca-endmember: $Ca(Al_2Si_4)O_{12}(H_2O)_6$), natrolite ($Na_2(Al_2Si_3)O_{10}(H_2O)_2$), zeolite 4A ($Na_2(Al_2Si_2)O_8(H_2O)_{4.5}$) and hydroxysodalite ($Na_8(Al_6Si_6)O_{24}(OH)_2(H_2O)_2$), representing the N-A-S-H phase as shown in Figure 5A. The chabazite type zeolite and hydroxysodalite both contains 6 member silica ring structure (6-R), consistent with the experimental observations suggesting that 6-R and 4-R structures might be the main secondary building units of the amorphous N-A-S-H gel formed in low CaO content AAMs [103-105]. The natrolite and zeolite 4A predicted from high CaO content AAMs consist of tetrahedral silica with 4 alumina cordination $Q^4(4Al)$ and 4-R structure. This is also consistent with the deconvoluted ^{29}Si solid-state NMR results of aged sodium silicate-activated slag samples [101] and the identification of a small fraction of crystalline gismondine (containing 4-R structure) from alkali-activated slag with a small amount of metakaolin replacement (~10 wt.%) [38].

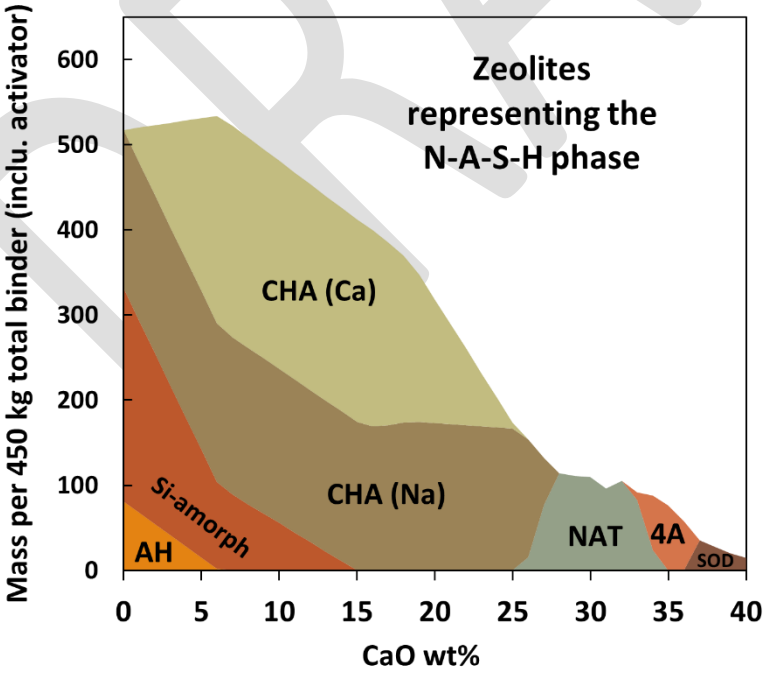
Compressive strength of the AAMs with the same bulk chemical compositions as the corresponding GEMs inputs are predicted using the trained GP model, as shown in the top plot in Figure 5A. The results show that the optimal strengths are predicted from either the N-A-S-H gel only region, or the C-(N)-A-S-H dominent region with hydrotalcite and strätlingite as secondary reaction products and a small fraction of N-A-S-H gel. Much lower compressive strengths are predicted from AAMs binders containing intermedia level of CaO content, reflecting experimental observations where lower strength performances are often observed from hybrid AAMs, where high calcium precursors (e.g. BFS) are placed by low calcium precursors (e.g. fly ash, metakaolin) by approximately 50% while keeping the total alkali dose

constent[79, 106]. This might relate to the fact that low calcium precursors often require high alkali dose to achieve sufficient reactivity, which is not always favouring the reactivity of high calcium precursors [106, 107].

The micromechanical performances, such as the elastic modulus, of individual reaction products formed in AAMs can affect the macroscale mechanical performances of the AAMs [108-110]. The elastic modulus of calcium silicate hydrate is around 10-15 GPa with a Ca/Si ratio of 1.0 and gel porosity around 20 % [111]. The elastic modulus of N-A-S-H with 7% to 10% alkali content ranges from 13 to 15 GPa [112]. The elastic modulus of hydrotalcite, the Mg-Al type LDH has not been reported in the literature. However, the elastic modulus of carbonated Ca-Al type LDH, monocarbonate is around 35 to 45 GPa [113]. It is possible that hydrotalcite would possess a similar elastic modulus due to the similarity in crystal structures between these two LDH type minerals. In addition, the formation of strätlingite has been identified as a critical strength given phase in Imperial Roman architectural mortar [114], suggesting that strätlingite might also significantly contribute to the overall strength performances of AAMs with high binder CaO content. For AAMs containing mixed reaction products, the elastic modulus values are around 18 GP for low calcium content binders [110, 115], and around 25 to 30 GP for high calcium content binders [113, 116].



(A)



(B)

Figure 5 (A) Comparison of phase assemblages (bottom) of alkali-activated materials predicted from thermodynamic modelling and compressive strength (top) of the

corresponding concrete mix design by chemical composition as predicted from the predictive surrogate model. While the CaO wt.% in the total binder ranges from 0 wt.% to 40 wt.%, SiO_2 wt.% and Al_2O_3 wt.% decreases uniformly from 75 wt.% to 35 wt.% and 20.0 wt.% to 16.0 wt.% respectively. The MgO wt.% is set to be 10% of corresponded CaO content, while Na_2O wt.% is fixed as 5 wt.%. A total water to binder ratio of 0.4 is used. (B) Zeolites representing the N-A-S-H phase as predicted corresponding to the result shown in (A).

Besides, a higher total mass of the reaction products is been predicted when the binder content of CaO increases to above 32 wt.%, resulting in higher total volumes of reaction products and thus lower overall porosity. The corresponded volume phase plot of Figure 5A can be found in Supporting Information (Figure S-3). The total porosities of the predicted AAMs are estimated using the method described in [117] from the thermodynamic modelling results. Comparing the predicted strength with the predicted total pore volume (plotted together in Figure 5A), in the high CaO content region, the predicted high strength binder compositions correspond to the lower predicted total porosity, consistent with the experimental observations [24, 118]. A similar trend is observed from the N-A-S-H gel only regions (CaO content below 5%), where a decrease of predicted porosity corresponds to increased predicted strength. However, there shows no clear correlation between the predicted strength and the predicted total porosity in the intermedia CaO content range.

4.3.4. The role of Na_2O content

Figure 6 shows the predicted phase assemblage of AAMs with varying Na_2O content from 2 wt.% to 10 wt.%, the proportions of CaO wt.%, SiO_2 wt.% and Al_2O_3 wt.% are fixed to match the bulk Ca/Si molar ratio of 1.0 and the bulk Si/Al molar ratio of 2.0. The MgO wt.% is set to be 10% of corresponded CaO content. These composition ranges are been setup to reflect the role of Na_2O content on the phase assemblages at the optimal CaO, SiO_2 and Al_2O_3 proportions, corresponding to the high CaO content region in Figure 4B. Since in this study, the Na_2O contents in the binder are contributed solely by the activators, the increase of Na_2O content is similar to the increase of total alkali content in the activator. The results show that, the formation of strätlingite increases as Na_2O content increases from 2 wt.% to 6 wt.% and decreases to form C3AH6 as Na_2O content continuously increasing to 10 wt.%. The formation of C3AH6 at high alkalinity binder system matches with experimental observations [94, 119].

The strength performances predicted from the GP model of the same binder compositions show the optimal strength at around 4 wt.% Na_2O content, where a significant increase in total alkali content results in lower strength performances. Comparing the predicted strength with the phase assemblages, it appears that with constant total binder content, the formation of C3AH6 at high Na_2O content leads to a lower total mass of reaction product, as well as a lower total volume of the reaction products (Supporting Information, Figure S-4), and thus a higher predicted porosity. This suggests that extremly high alkali dosage will lead to coarsen of the pore structure, impairing the strength performances. This might also be used to explain the efflorescence phenomena that are often observed in AAMs with higher Na_2O content [120]. In addition, the results suggest that this trained composition-property model can be used to estimate the optimal activator dosage for precursors of known chemical compositions.

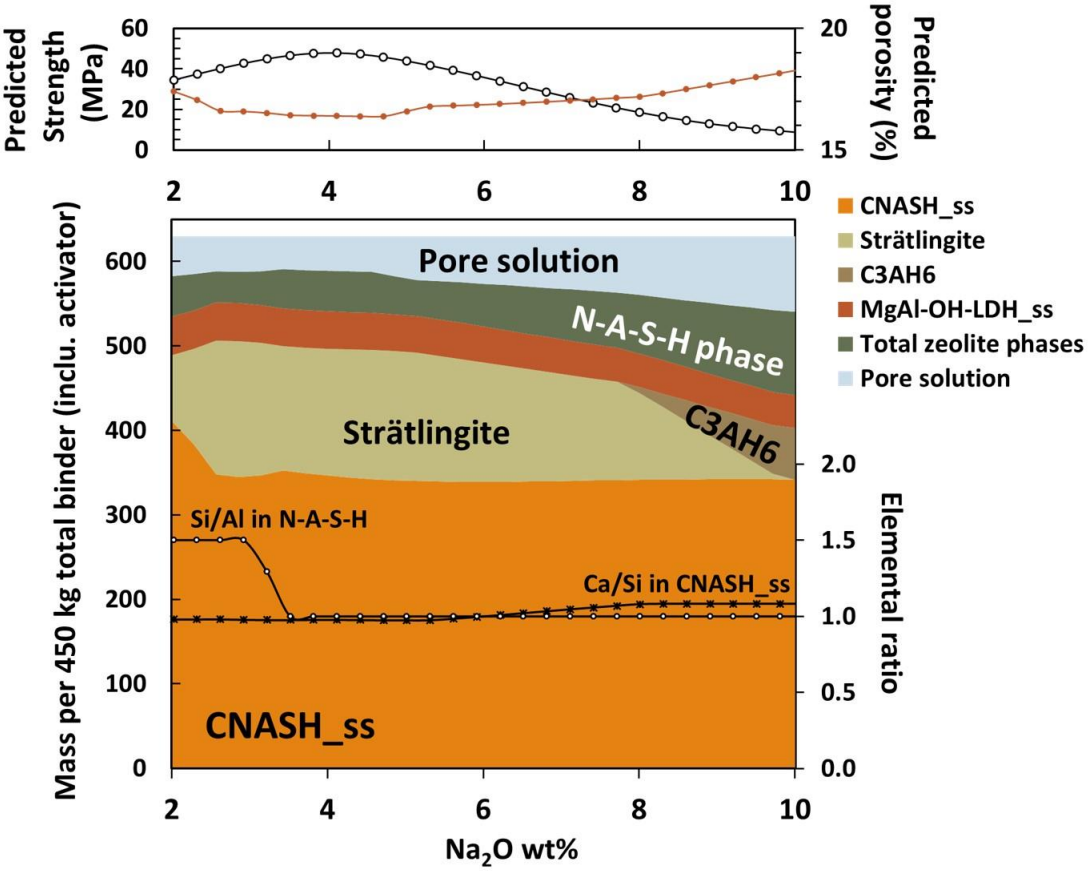


Figure 6 Comparison of phase assemblages (bottom) of alkali-activated materials predicted from thermodynamic modelling and compressive strength (top) of the corresponding concrete

mix design by chemical composition as predicted from the predictive surrogate model. While the Na_2O wt.% in the total binder ranges from 2 wt.% to 10 wt.%, the proportions of CaO wt.%, SiO_2 wt.% and Al_2O_3 wt.% are fixed to match Ca/Si molar ratio of 1.0 and Si/Al molar ratio of 2.0. The MgO wt.% is set to be 10% of the corresponded CaO content.

4.3.5. The role of MgO content

Figure 7 shows the predicted phase assemblages of AAMs with varying MgO content from 0 wt.% to 12 wt.%. Similar to the previous session, the proportions of CaO wt.%, SiO_2 wt.% and Al_2O_3 wt.% are fixed to match the bulk Ca/Si molar ratio of 1.0 and Si/Al molar ratio of 2.0. The Na_2O wt.% is fixed at 5 wt.%. The results show that as MgO content in the total AAM binder increases, the amount of strätlingite decreases while the amount of hydrotalcite increases. This is consistent with experimental observations where increased quantity of hydrotalcite-like phase together with decreased Ca-Al LDHs are identified from alkali-activated slag with higher MgO content [34]. The formation of hydrotalcite consumes Al and thus destabilising strätlingite. Comparing with the compressive strengths of the same binder compositions predicted from the GP model, when the MgO content is below 3wt.%, the predicted strength increases as the MgO content increases; while for MgO content above 3 wt.%, the predicted strength decreases at higher MgO content. Similar trends have also been identified when adding reactive MgO into alkali-activated slag where very high MgO content resulted in lower compressive strength [121, 122]. However, this contradicts with the results reported by Ben Haha et al. [118], where slightly higher mortar compressive strength was achieved by sample with higher MgO content (around 6-8 wt.%) when sodium silicate activator was used. But the same study also reported that when sodium hydroxide was used as activator, slightly lower mortar compressive strength was achieved by higher MgO content samples. Different from previous sections, there appear to be no clear correlations between the strength performances, phase assemblages, and the predicted porosity when the MgO content is the main variance. According to the data source (Figure 1) and data structure (Table 1) used in this study for training the predictive model, MgO processes the lowest mass percentage among the five major elements and is mainly contributed by the BFS. It has been reported by various studies that MgO play critical roles in affecting the early age reaction kinetics, as well as determining the degree of reaction of alkali-activated slag [19, 23, 34, 118, 122, 123]. The main effect of MgO on the strength performance of AAM is relating to the densification of the microstructure [81,

118, 122], as a result of increased degree of reaction and increased total reaction products. Therefore, the effect of MgO content might be dominated by the reaction kinetics rather than the phase assemblages. Further studies are required to investigate the role of MgO on micromechanical performances of AAMs in addition to its effect on microstructure development.

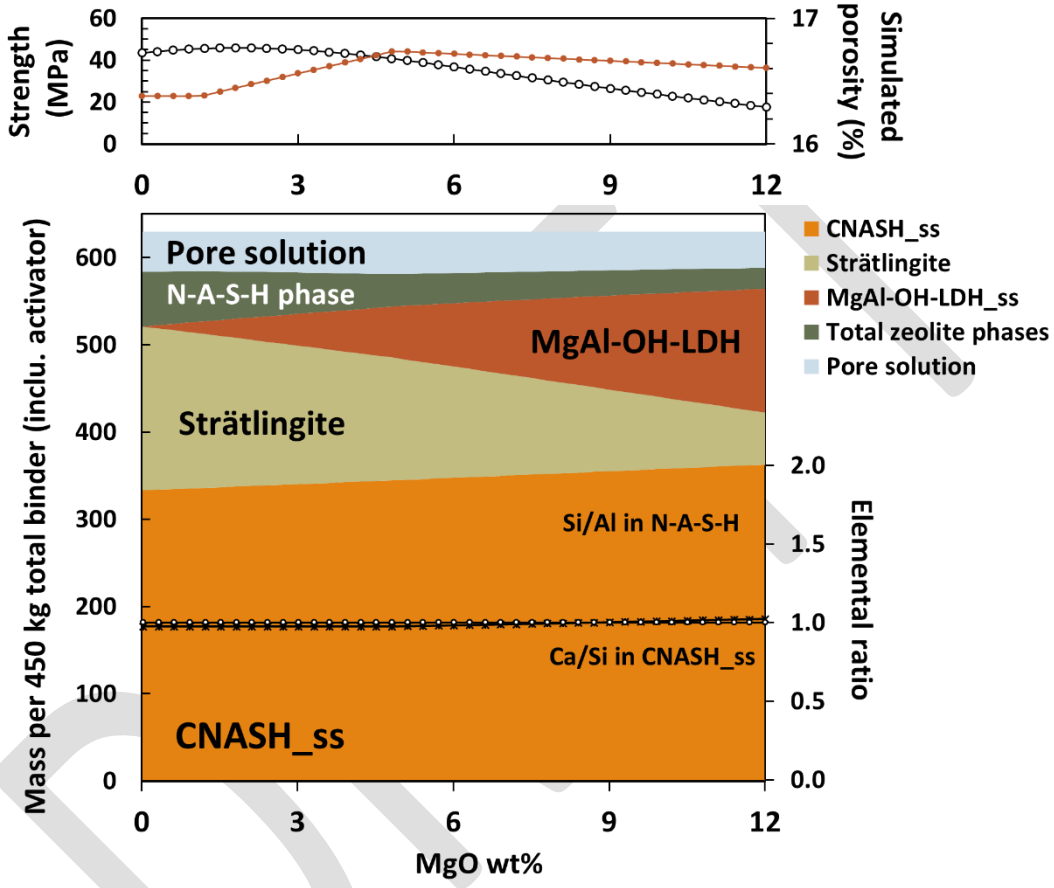


Figure 7 Comparison of phase assemblages (bottom) of alkali-activated materials predicted from thermodynamic modelling and compressive strength (top) of the corresponding concrete mix design by chemical composition as predicted from the predictive surrogate model. While the MgO wt.% in the total binder ranges from 0 wt.% to 12 wt.%, the proportions of CaO wt.%, SiO_2 wt.% and Al_2O_3 wt.% are fixed to match Ca/Si molar ratio of 1.0 and Si/Al molar ratio of 2.0. The Na_2O wt.% is fixed at 5 wt.%.

5. Perspectives and future improvement

The outcomes of this study prove that the composition-property correlations of alkali-activated materials developed using the data-driven approach can reflect the chemistry of the binder system. Since the composition-property model is independent of the types of precursors been used, it can be used to conduct performance-based materials design, as well as selection of potential precursor materials for their best uses. The present study showcased the composition-strength correlations of AAMs. However, similar composition-property correlations can also be developed for assessing a wide range of properties, including chloride permeability, carbonation resistance and environmental impact. The variance-based Sobol indices and thermodynamic models can be used to evaluate whether the trained machine learning model can reflect the true physical and chemical nature of the studied system. Then the mathematically and chemically verified machine learning surrogate model can be used to perform high-throughput "experiments" to achieve fit-for-purpose design and optimisations. Nevertheless, a purely data-driven model puts emphasis on the quality and completeness of well-organised data sources, which are not often readily available for interpretations. The proposed model can be further improved if additional factors, such as particle sizes, surface areas, degree of reactions, and amorphous content can also be included (if the completed dataset is available) in the training model. The inclusion of additional physical factors can enable the model to take into consideration the effects of microstructure and kinetic-driven reactions.

For coupling the composition-property model with thermodynamic modelling, one important factor that will need to be considered to further improve the robustness of the model is the degree of reaction (DoR). The DoR will determine the percentage of precursors participating in the reaction, and thus affecting the predicted total porosity. Different methods have been adapted for determining the degree of reaction in AAM, including selective chemical dissolution [124, 125] and deconvolution of solid-state ^{29}Si NMR [126-128]. Figure 8 summarises the DoR of AAMs reported in the literature at 28 days sample age, plotted versus the binder chemical contents. The limited existing information suggests that there is no significant correlation among bulk CaO, SiO_2 and MgO content in the AAM; however, higher Al_2O_3 content and higher Na_2O content appear to result in higher DoR. Besides, the lack of a universal testing protocol for determining the DoR in AAMs might also cause uncertainties in existing observations. Further studies need to be carried out in order to understand the

correlation between DoR and bulk binder compositions, as well as additional factors such as particle sizes and surface areas.

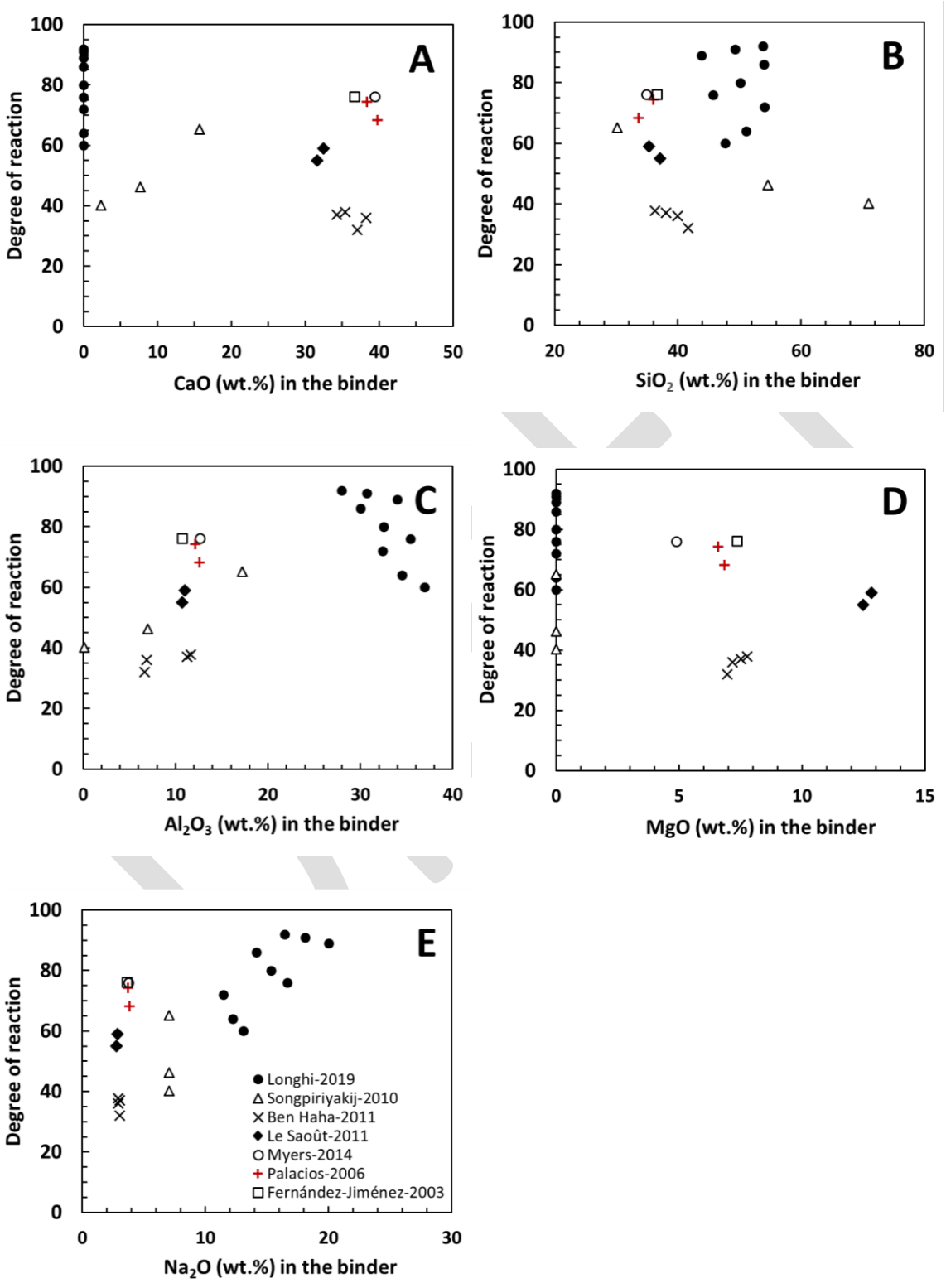


Figure 8 Degree of reaction reported in literature of alkali-activated materials at 28 days with different binder compositions. The reported degree of reaction is plotted versus (A) CaO wt.%, (B) SiO_2 wt.%, (C) Al_2O_3 wt.%, (D) MgO wt.%, (E) Na_2O wt.%. For all the data included in this figure, sodium silicate and/or sodium hydroxide were used as activators. Longhi et al.-2019 [124], Songpiriyakij et al.-2010 [125], Ben Haha et al.-2011 [24], Le Saout et al.-2011 [126], Myers et al.-2014 [127], Palacios et al.-2006 [128], Fernandez-Jimenez et al.-2003 [27].

6. Conclusions

This study proposes a novel material analysis framework coupling machine learning with thermodynamic modelling for studying the composition-property correlations of alkali-activated materials. The results suggest that the trained and validated GP predictive model can reflect the statistical features of the training data, with a general model accuracy of 5.80 MPa. The global sensitivity analysis further proves that the trained GP model can reflect the relative importance of the model inputs (binder chemical compositions and mix designs) with regards to the known chemistry of the AAMs system. The strength performances of 3803 random AAM mix designs with varying chemical compositions (generated using the Monte Carlo process) are predicted by the trained GP model. The results suggest that higher AAM strength performances can be expected from either C-(N)-A-S-H dominant binder system or N-A-S-H gel dominant binder system, corresponding to the thermodynamic modelling results. The predicted strengths within these regions are also consistent with porosity predicted from the thermodynamic modelling, where lower porosities are predicted from binder compositions with higher strength.

Acknowledgement

X. Ke acknowledges University of Bath for sponsoring her Prize Fellowship.

Conflict of interest statement

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.

Replication of Results

The data required to reproduce these findings will be available upon request.

References

[1] Habert G, Miller SA, John VM, Provis JL, Favier A, Horvath A, et al. Environmental impacts and decarbonization strategies in the cement and concrete industries. Nature Reviews Earth & Environment. 2020;1(11):559-73.
[2] Provis JL, Bernal SA. Geopolymers and related alkali-activated materials. Annual Review of Materials Research. 2014;44:299-327.
[3] Grant Norton M, Provis JL. 1000 at 1000: Geopolymer technology—the current state of the art. J Mater Sci. 2020;55(28):13487-9.
[4] Shi C, Qu B, Provis JL. Recent progress in low-carbon binders. Cement and Concrete Research. 2019;122:227-50.
[5] Pacheco-Torgal F, Abdollahnejad Z, Miraldo S, Kheradmand M. Alkali-Activated Cement-Based Binders (AACB) as Durable and Cost-Competitive Low-CO2 Binder Materials: Some Shortcomings That Need to be Addressed. Handbook of Low Carbon Concrete2017. p. 195-216.
[6] Criado M, Provis JL. Alkali Activated Slag Mortars Provide High Resistance to Chloride-Induced Corrosion of Steel. Frontiers in Materials. 2018;5(34).
[7] Bakharev T, Sanjayan JG, Cheng YB. Resistance of alkali-activated slag concrete to acid attack. Cement and Concrete Research. 2003;33(10):1607-11.
[8] Ke X, Bernal SA, Sato T, Provis JL. Alkali aluminosilicate geopolymers as binders to encapsulate strontium-selective titanate ion-exchangers. Dalton Transactions. 2019;48:12116-26.
[9] Khalifa AZ, Cizer Ö, Pontikes Y, Heath A, Patureau P, Bernal SA, et al. Advances in alkali-activation of clay minerals. Cement and Concrete Research. 2020;132:106050.
[10] Slaty F, Khoury H, Wastiels J, Rahier H. Characterization of alkali activated kaolinitic clay. Applied Clay Science. 2013;75-76:120-5.
[11] Tchakoute Kouamo H, Mbey JA, Elimbi A, Kenne Diffo BB, Njopwouo D. Synthesis of volcanic ash-based geopolymer mortars by fusion method: Effects of adding metakaolin to fused volcanic ash. Ceramics International. 2013;39(2):1613-21.
[12] Reig L, Tashima MM, Borrachero MV, Monzó J, Cheeseman CR, Payá J. Properties and microstructure of alkali-activated red clay brick waste. Construction and Building Materials. 2013;43:98-106.
[13] Ye N, Yang J, Ke X, Zhu J, Li Y, Xiang C, et al. Synthesis and Characterization of Geopolymer from Bayer Red Mud with Thermal Pretreatment. Journal of the American Ceramic Society. 2014;97(5):1652-60.
[14] Cristelo N, Fernández-Jiménez A, Vieira C, Miranda T, Palomo Á. Stabilisation of construction and demolition waste with a high fines content using alkali activated fly ash. Construction and Building Materials. 2018;170:26-39.
[15] Winnefeld F, Ben Haha M, Le Saout G, Costoya M, Ko S-C, Lothenbach B. Influence of slag composition on the hydration of alkali-activated slags. Journal of Sustainable Cement-Based Materials. 2015;4:85-100.
[16] Ben Haha M, Lothenbach B, Le Saout G, Winnefeld F. Influence of slag chemistry on the hydration of alkali-activated blast-furnace slag — part II: effect of Al_2O_3. Cement and Concrete Research. 2012;42(1):74-83.

[17] Juenger MCG, Winnefeld F, Provis JL, Ideker JH. Advances in alternative cementitious binders. Cement and Concrete Research. 2011;41(12):1232-43.
[18] He J, Jie Y, Zhang J, Yu Y, Zhang G. Synthesis and characterization of red mud and rice husk ash-based geopolymer composites. Cement and Concrete Composites. 2013;37:108-18.
[19] Bernal SA, San Nicolas R, Myers RJ, Mejía de Gutiérrez R, Puertas F, van Deventer JSJ, et al. MgO content of slag controls phase evolution and structural changes induced by accelerated carbonation in alkali-activated binders. Cement and Concrete Research. 2014;57(0):33-43.
[20] Li N, Shi C, Zhang Z, Wang H, Liu Y. A review on mixture design methods for geopolymer concrete. Composites Part B: Engineering. 2019;178:107490.
[21] Adesanya E, Perumal P, Luukkonen T, Yliniemi J, Ohenoja K, Kinnunen P, et al. Opportunities to improve sustainability of alkali-activated materials: A review of side-stream based activators. Journal of Cleaner Production. 2021;286:125558.
[22] Passuello A, Rodríguez ED, Hirt E, Longhi M, Bernal SA, Provis JL, et al. Evaluation of the potential improvement in the environmental footprint of geopolymers using waste-derived activators. Journal of Cleaner Production. 2017;166:680-9.
[23] Criado M, Walkley B, Ke X, Provis JL, Bernal SA. Slag and Activator Chemistry Control the Reaction Kinetics of Sodium Metasilicate-Activated Slag Cements. Sustainability. 2018;10(12):4709.
[24] Ben Haha M, Le Saout G, Winnefeld F, Lothenbach B. Influence of activator type on hydration kinetics, hydrate assemblage and microstructural development of alkali activated blast-furnace slags. Cement and Concrete Research. 2011;41(3):301-10.
[25] Reddy MS, Dinakar P, Rao BH. A review of the influence of source material's oxide composition on the compressive strength of geopolymer concrete. Microporous Mesoporous Mater. 2016;234:12-23.
[26] Li N, Shi C, Zhang Z. Understanding the roles of activators towards setting and hardening control of alkali-activated slag cement. Composites Part B: Engineering. 2019;171:34-45.
[27] Fernández-Jiménez A, Puertas F. Effect of activator mix on the hydration and strength behaviour of alkali-activated slag cements. Advances in Cement Research. 2003;15:129-36.
[28] Duxson P, Mallicoat SW, Lukey GC, Kriven WM, van Deventer JSJ. The effect of alkali and Si/Al ratio on the development of mechanical properties of metakaolin-based geopolymers. Colloids and Surfaces A: Physicochemical and Engineering Aspects. 2007;292(1):8-20.
[29] Bernal SA. Advances in near-neutral salts activation of blast furnace slags. RILEM Technical Letters. 2016;1:39-44.
[30] Yip CK, Lukey GC, van Deventer JSJ. The coexistence of geopolymeric gel and calcium silicate hydrate at the early stage of alkaline activation. Cement and Concrete Research. 2005;35(9):1688-97.
[31] Criado M, Palomo A, Fernández-Jiménez A. Alkali activation of fly ashes. Part 1: Effect of curing conditions on the carbonation of the reaction products. Fuel. 2005;84(16):2048-54.
[32] Wilińska I, Pacewska B. Comparative investigation of reactivity of different kinds of fly ash in alkaline media. Journal of Thermal Analysis and Calorimetry. 2019;138(6):3857-72.
[33] Kumar S, García-Triñanes P, Teixeira-Pinto A, Bao M. Development of alkali activated cement from mechanically activated silico-manganese (SiMn) slag. Cement and Concrete Composites. 2013;40:7-13.
[34] Ke X, Bernal SA, Provis JL. Controlling the reaction kinetics of sodium carbonate-activated slag cements using calcined layered double hydroxides. Cement and Concrete Research. 2016;81:24-37.

[35] Myers RJ, Lothenbach B, Bernal SA, Provis JL. Thermodynamic modelling of alkali-activated slag cements. Applied Geochemistry. 2015;61:233-47.
[36] Lothenbach B, Gruskovnjak A. Hydration of alkali-activated slag: thermodynamic modelling. Advances in Cement Research. 2007;19(2):81-92.
[37] Ke X, Bernal SA, Provis JL, Lothenbach B. Thermodynamic modelling of phase evolution in alkali-activated slag cements exposed to carbon dioxide. Cement and Concrete Research. 2020;136:106158.
[38] Bernal SA, Provis JL, Rose V, Mejía de Gutierrez R. Evolution of binder structure in sodium silicate-activated slag-metakaolin blends. Cement and Concrete Composites. 2011;33(1):46-54.
[39] Wan Q, Rao F, Song S, García RE, Estrella RM, Patiño CL, et al. Geopolymerization reaction, microstructure and simulation of metakaolin-based geopolymers at extended Si/Al ratios. Cem Concr Compos. 2017;79:45-52.
[40] Lothenbach B, Zajac M. Application of thermodynamic modelling to hydrated cements. Cement and Concrete Research. 2019;123:105779.
[41] Himanen L, Geurts A, Foster AS, Rinke P. Data-Driven Materials Science: Status, Challenges, and Perspectives. Advanced Science. 2019;6(21):1900808.
[42] Nyshadham C, Rupp M, Bekker B, Shapeev AV, Mueller T, Rosenbrock CW, et al. Machine-learned multi-system surrogate models for materials prediction. npj Computational Materials. 2019;5(1):51.
[43] Liu H, Fu Z, Yang K, Xu X, Bauchy M. Machine learning for glass science and engineering: A review. Journal of Non-Crystalline Solids. 2019:119419.
[44] Hu Y-J, Zhao G, Zhang M, Bin B, Del Rose T, Zhao Q, et al. Predicting densities and elastic moduli of SiO2-based glasses by machine learning. npj Computational Materials. 2020;6(1):25.
[45] Ward L, Agrawal A, Choudhary A, Wolverton C. A general-purpose machine learning framework for predicting properties of inorganic materials. npj Computational Materials. 2016;2(1):16028.
[46] Oey T, Jones S, Bullard JW, Sant G. Machine learning can predict setting behavior and strength evolution of hydrating cement systems. Journal of the American Ceramic Society. 2020;103(1):480-90.
[47] Jalal M, Grasley Z, Gurganus C, Bullard JW. Experimental investigation and comparative machine-learning prediction of strength behavior of optimized recycled rubber concrete. Construction and Building Materials. 2020;256:119478.
[48] Song H-W, Kwon S-J. Evaluation of chloride penetration in high performance concrete using neural network algorithm and micro pore structure. Cement and Concrete Research. 2009;39(9):814-24.
[49] Cai R, Han T, Liao W, Huang J, Li D, Kumar A, et al. Prediction of surface chloride concentration of marine concrete using ensemble machine learning. Cement and Concrete Research. 2020;136:106164.
[50] Ke X, Duan Y. A Bayesian machine learning approach for inverse prediction of high-performance concrete ingredients with targeted performance. Construction and Building Materials. 2021;270:121424.
[51] Nossent J, Elsen P, Bauwens W. Sobol' sensitivity analysis of a complex environmental model. Environmental Modelling & Software. 2011;26(12):1515-25.
[52] Prieto D, Asensio MI, Ferragut L, Cascón JM. Sensitivity analysis and parameter adjustment in a simplified physical wildland fire model. Advances in Engineering Software. 2015;90:98-106.
[53] Sobol′ IM. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation. 2001;55(1):271-80.

[54] Cheng K, Lu Z, Ling C, Zhou S. Surrogate-assisted global sensitivity analysis: an overview. Structural and Multidisciplinary Optimization. 2020;61(3):1187-213.
[55] Neal RM. Priors for Infinite Networks. In: Neal RM, editor. Bayesian Learning for Neural Networks. New York, NY: Springer New York; 1996. p. 29-53.
[56] Svendsen DH, Morales-Álvarez P, Ruescas AB, Molina R, Camps-Valls G. Deep Gaussian processes for biogeophysical parameter retrieval and model inversion. ISPRS Journal of Photogrammetry and Remote Sensing. 2020;166:68-81.
[57] Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning: MIT Press; 2006.
[58] Joseph VR, Hung Y, Sudjianto A. Blind Kriging: A New Method for Developing Metamodels. Journal of Mechanical Design. 2008;130(3).
[59] Couckuyt I, Forrester A, Gorissen D, De Turck F, Dhaene T. Blind Kriging: Implementation and performance analysis. Advances in Engineering Software. 2012;49:1-13.
[60] Roberts S, Osborne M, Ebden M, Reece S, Gibson N, Aigrain S. Gaussian processes for time-series modelling. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2013;371(1984):20110550.
[61] Gao F, Han L. Implementing the Nelder-Mead simplex algorithm with adaptive parameters. Computational Optimization and Applications. 2012;51(1):259-77.
[62] Owen AB. Variance Components and Generalized Sobol' Indices. SIAM/ASA Journal on Uncertainty Quantification. 2013;1(1):19-41.
[63] Helgeson HC, Kirkham DH, Flowers GC. Theoretical prediction of the thermodynamic behavior of aqueous electrolytes by high pressures and temperatures; IV, Calculation of activity coefficients, osmotic coefficients, and apparent molal and standard and relative partial molal properties to 600 °C and 5kb. American Journal of Science. 1981;281:1249-516.
[64] Lothenbach B, Kulik DA, Matschei T, Balonis M, Baquerizo L, Dilnesa B, et al. Cemdata18: A chemical thermodynamic database for hydrated Portland cements and alkali-activated materials. Cement and Concrete Research. 2019;115:472-506.
[65] Myers RJ, Bernal SA, Provis JL. A thermodynamic model for C-(N-)A-S-H gel: CNASH_ss. Derivation and validation. Cement and Concrete Research. 2014;66(0):27-47.
[66] De Weerdt K, Plusquellec G, Belda Revert A, Geiker MR, Lothenbach B. Effect of carbonation on the pore solution of mortar. Cement and Concrete Research. 2019;118:38-56.
[67] Ma B, Lothenbach B. Thermodynamic study of cement/rock interactions using experimentally generated solubility data of zeolites. Cement and Concrete Research. 2020;135:106149.
[68] Ma B, Lothenbach B. Synthesis, characterization, and thermodynamic study of selected Na-based zeolites. Cement and Concrete Research. 2020;135:106111.
[69] Chermak JA, Rimstidt JD. Estimating the thermodynamic properties (ΔG_f^0 and ΔH_f^0) of silicate minerals at 298 K from the sum of polyhedral contributions. American Mineralogist. 1989;74(9-10):1023-31.
[70] Navrotsky A, Tian ZR. Systematics in the Enthalpies of Formation of Anhydrous Aluminosilicate Zeolites, Glasses, and Dense Phases. Chemistry – A European Journal. 2001;7(4):769-74.
[71] Bakharev T, Sanjayan JG, Cheng YB. Resistance of alkali-activated slag concrete to carbonation. Cement and Concrete Research. 2001;31(9):1277-83.
[72] Bernal SA, Mejía de Gutiérrez R, Pedraza AL, Provis JL, Rodriguez ED, Delvasto S. Effect of binder content on the performance of alkali-activated slag concretes. Cement and Concrete Research. 2011;41(1):1-8.
[73] Bernal SA, Mejía de Gutiérrez R, Provis JL. Engineering and durability properties of concretes based on alkali-activated granulated blast furnace slag/metakaolin blends. Construction and Building Materials. 2012;33:99-108.

[74] Bernal SA, Provis JL, Mejía de Gutiérrez R, van Deventer JSJ. Accelerated carbonation testing of alkali-activated slag/metakaolin blended concretes: effect of exposure conditions. Materials and Structures. 2015;48(3):653-69.
[75] Bernal SA, San Nicolas R, Provis JL, Mejía de Gutiérrez R, van Deventer JSJ. Natural carbonation of aged alkali-activated slag concretes. Materials and Structures. 2014;47(4):693-707.
[76] Ana M. Fernandez-Jimenez AP, Cecilio L-H. Engineering Properties of Alkali-Activated Fly Ash Concrete. ACI Materials Journal. 2006;103(2).
[77] Gunasekara C, Law DW, Setunge S. Long term permeation properties of different fly ash geopolymer concretes. Construction and Building Materials. 2016;124:352-62.
[78] Häkkinen T. The influence of slag content on the microstructure, permeability and mechanical properties of concrete: Part 2 technical properties and theoretical examinations. Cement and Concrete Research. 1993;23(3):518-30.
[79] Humad AM, Kothari A, Provis JL, Cwirzen A. The Effect of Blast Furnace Slag/Fly Ash Ratio on Setting, Strength, and Shrinkage of Alkali-Activated Pastes and Concretes. Frontiers in Materials. 2019;6(9).
[80] Ismail I, Bernal SA, Provis JL, San Nicolas R, Brice DG, Kilcullen AR, et al. Influence of fly ash on the water and chloride permeability of alkali-activated slag mortars and concretes. Construction and Building Materials. 2013;48:1187-201.
[81] Khan MSH, Castel A. Effect of MgO and Na_2SiO_3 on the carbonation resistance of alkali activated slag concrete. Magazine of Concrete Research. 2018;70(13):685-92.
[82] Law DW, Adam AA, Molyneaux TK, Patnaikuni I. Durability assessment of alkali activated slag (AAS) concrete. Materials and Structures. 2012;45(9):1425-37.
[83] Lee NK, Lee HK. Setting and mechanical properties of alkali-activated fly ash/slag concrete manufactured at room temperature. Construction and Building Materials. 2013;47:1201-9.
[84] Li Z, Li S. Carbonation resistance of fly ash and blast furnace slag based geopolymer concrete. Constr Build Mater. 2018;163:668-80.
[85] Nath P, Sarker PK. Effect of GGBFS on setting, workability and early strength properties of fly ash geopolymer concrete cured in ambient condition. Construction and Building Materials. 2014;66:163-71.
[86] Nuaklong P, Sata V, Chindaprasirt P. Properties of metakaolin-high calcium fly ash geopolymer concrete containing recycled aggregate from crushed concrete specimens. Construction and Building Materials. 2018;161:365-73.
[87] Pouhet R, Cyr M. Formulation and performance of flash metakaolin geopolymer concretes. Construction and Building Materials. 2016;120:150-60.
[88] Sarker PK, Kelly S, Yao Z. Effect of fire exposure on cracking, spalling and residual strength of fly ash geopolymer concrete. Materials & Design. 2014;63:584-92.
[89] Li C, Sun H, Li L. A review: The comparison between alkali-activated slag (Si+Ca) and metakaolin (Si+Al) cements. Cement and Concrete Research. 2010;40(9):1341-9.
[90] Krizan D, Zivanovic B. Effects of dosage and modulus of water glass on early hydration of alkali–slag cements. Cement and Concrete Research. 2002;32(8):1181-8.
[91] Bernal SA, San Nicolas R, van Deventer JSJ, Provis JL. Water content modifies the structural development of sodium metasilicate-activated slag binders. ALCONPAT J. 2015;5(1):29-40.
[92] Chidiac SE, Moutassem F, Mahmoodzadeh F. Compressive strength model for concrete. Magazine of Concrete Research. 2013;65(9):557-72.
[93] Moutassem F, Chidiac SE. Assessment of concrete compressive strength prediction models. KSCE Journal of Civil Engineering. 2016;20(1):343-58.

[94] Shi Z, Shi C, Wan S, Li N, Zhang Z. Effect of alkali dosage and silicate modulus on carbonation of alkali-activated slag mortars. Cement and Concrete Research. 2018;113:55-64.
[95] Walkley B, Ke X, Hussein OH, Bernal SA, Provis JL. Incorporation of strontium and calcium in geopolymer gels. Journal of Hazardous Materials. 2019:121015.
[96] Guo X, Shi H, Dick WA. Compressive strength and microstructural characteristics of class C fly ash geopolymer. Cement and Concrete Composites. 2010;32(2):142-7.
[97] Ismail I, Bernal SA, Provis JL, San Nicolas R, Hamdan S, van Deventer JSJ. Modification of phase evolution in alkali-activated blast furnace slag by the incorporation of fly ash. Cement and Concrete Composites. 2014;45(0):125-35.
[98] Garcia-Lodeiro I, Palomo A, Fernández-Jiménez A, Macphee DE. Compatibility studies between N-A-S-H and C-A-S-H gels. Study in the ternary diagram Na2O–CaO–Al2O3–SiO2–H2O. Cement and Concrete Research. 2011;41(9):923-31.
[99] García-Lodeiro I, Fernández-Jiménez A, Palomo A, Macphee DE. Effect of Calcium Additions on N–A–S–H Cementitious Gels. Journal of the American Ceramic Society. 2010;93(7):1934-40.
[100] Bernal SA, Provis JL, Walkley B, San Nicolas R, Gehman JD, Brice DG, et al. Gel nanostructure in alkali-activated binders based on slag and fly ash, and effects of accelerated carbonation. Cement and Concrete Research. 2013;53:127-44.
[101] Walkley B, Ke X, Provis JL, Bernal SA. Influence of the activator anion type on the nanostructure of aged alkali-activated slag cements. Inorganic Chemistry. 2021;Submitted.
[102] Bernal SA, Mejía de Gutiérrez R, Provis JL, Rose V. Effect of silicate modulus and metakaolin incorporation on the carbonation of alkali silicate-activated slags. Cement and Concrete Research. 2010;40(6):898-907.
[103] Oh JE, Monteiro PJM, Jun SS, Choi S, Clark SM. The evolution of strength and crystalline phases for alkali-activated ground blast furnace slag and fly ash-based geopolymers. Cement and Concrete Research. 2010;40(2):189-96.
[104] Rees CA, Provis JL, Lukey GC, van Deventer JSJ. The mechanism of geopolymer gel formation investigated through seeded nucleation. Colloids and Surfaces A: Physicochemical and Engineering Aspects. 2008;318(1–3):97-105.
[105] White CE, Page K, Henson NJ, Provis JL. In situ synchrotron X-ray pair distribution function analysis of the early stages of gel formation in metakaolin-based geopolymers. Applied Clay Science. 2013;73(Supplement C):17-25.
[106] Puertas F, Martínez-Ramírez S, Alonso S, Vázquez T. Alkali-activated fly ash/slag cements: Strength behaviour and hydration products. Cement and Concrete Research. 2000;30(10):1625-32.
[107] Williamson T, Juenger MCG. The role of activating solution concentration on alkali–silica reaction in alkali-activated fly ash concrete. Cement and Concrete Research. 2016;83:124-30.
[108] Haecker CJ, Garboczi EJ, Bullard JW, Bohn RB, Sun Z, Shah SP, et al. Modeling the linear elastic properties of Portland cement paste. Cement and Concrete Research. 2005;35(10):1948-60.
[109] Hu C, Li Z. A review on the mechanical properties of cement-based materials measured by nanoindentation. Construction and Building Materials. 2015;90:80-90.
[110] Němeček J, Šmilauer V, Kopecký L. Nanoindentation characteristics of alkali-activated aluminosilicate materials. Cement and Concrete Composites. 2011;33(2):163-70.
[111] Hu Z, Wyrzykowski M, Griffa M, Scrivener K, Lura P. Young's modulus and creep of calcium-silicate-hydrate compacts measured by microindentation. Cement and Concrete Research. 2020;134:106104.

[112] Luo Z, Li W, Gan Y, Mendu K, Shah SP. Applying grid nanoindentation and maximum likelihood estimation for N-A-S-H gel in geopolymer paste: Investigation and discussion. Cement and Concrete Research. 2020;135:106112.
[113] Wilson W, Sorelli L, Tagnit-Hamou A. Unveiling micro-chemo-mechanical properties of C–(A)–S–H and other phases in blended-cement pastes. Cement and Concrete Research. 2018;107:317-36.
[114] Jackson MD, Landis EN, Brune PF, Vitti M, Chen H, Li Q, et al. Mechanical resilience and cementitious processes in Imperial Roman architectural mortar. Proceedings of the National Academy of Sciences. 2014;111(52):18484-9.
[115] Khedmati M, Kim Y-R, Turner JA, Alanazi H, Nguyen C. An integrated microstructural-nanomechanical-chemical approach to examine material-specific characteristics of cementitious interphase regions. Materials Characterization. 2018;138:154-64.
[116] Thomas RJ, Gebregziabiher BS, Giffin A, Peethamparan S. Micromechanical properties of alkali-activated slag cement binders. Cement and Concrete Composites. 2018;90:241-56.
[117] Lothenbach B, Matschei T, Möschner G, Glasser FP. Thermodynamic modelling of the effect of temperature on the hydration and porosity of Portland cement. Cement and Concrete Research. 2008;38(1):1-18.
[118] Ben Haha M, Lothenbach B, Le Saout G, Winnefeld F. Influence of slag chemistry on the hydration of alkali-activated blast-furnace slag — part I: effect of MgO. Cement and Concrete Research. 2011;41(9):955-63.
[119] Ke X, Bernal SA, Ye N, Provis JL, Yang J. One-Part Geopolymers Based on Thermally Treated Red Mud/NaOH Blends. Journal of the American Ceramic Society. 2015;98(1):5-11.
[120] Zhang Z, Provis JL, Ma X, Reid A, Wang H. Efflorescence and subflorescence induced microstructural and mechanical evolution in fly ash-based geopolymers. Cement and Concrete Composites. 2018;92:165-77.
[121] Jin F, Al-Tabbaa A. Strength and drying shrinkage of slag paste activated by sodium carbonate and reactive MgO. Construction and Building Materials. 2015;81:58-65.
[122] Park SM, Jang JG, Lee HK. Unlocking the role of MgO in the carbonation of alkali-activated slag cement. Inorganic Chemistry Frontiers. 2018;5(7):1661-70.
[123] Snellings R, Paulhiac T, Scrivener K. The effect of mg on slag reactivity in blended cements. Waste Biomass Valoriz. 2014;5:369-83.
[124] Longhi MA, Walkley B, Rodríguez ED, Kirchheim AP, Zhang Z, Wang H. New selective dissolution process to quantify reaction extent and product stability in metakaolin-based geopolymers. Composites Part B: Engineering. 2019;176:107172.
[125] Songpiriyakij S, Kubprasit T, Jaturapitakkul C, Chindaprasirt P. Compressive strength and degree of reaction of biomass- and fly ash-based geopolymer. Construction and Building Materials. 2010;24(3):236-40.
[126] Le Saoût G, Ben Haha M, Winnefeld F, Lothenbach B. Hydration Degree of Alkali-Activated Slags: A 29Si NMR Study. Journal of the American Ceramic Society. 2011;94(12):4541-7.
[127] Myers RJ, Bernal SA, Gehman JD, van Deventer JSJ, Provis JL. The role of Al in cross-linking of alkali-activated slag cements. Journal of the American Ceramic Society. 2014;98(3):996-1004.
[128] Palacios M, Puertas F. Effect of carbonation on alkali-activated slag paste. Journal of the American Ceramic Society. 2006;89(10):3211-21.