ORIGINAL ARTICLE

Prediction of heat of hydration of cementitious systems using Gaussian process regression enables mass concrete thermal modeling

Luna E. Al-Hasani · Greisi Perez · Hana N. Herndon · Jisoo Park · Jonathan L. Poole · Iris Tien · Newell R. Washburn · Yong K. Cho · T. Russell Gentry · Kimberly E. Kurtis

Abstract The temperature and time-dependent heat of hydration of cementitious pastes is a fundamental property to understand concrete performance. Cement hydration can be depicted using kinetic models by considering the effect of the chemical and physical properties of the cementitious pastes and the curing conditions. Supplementary cementitious materials and fillers are used as partial replacements for Portland cement to advance concrete performance and durability, which can exert varying effects on hydration kinetics. This adds a level of complexity that is difficult to capture with existing modeling methods. Here, the time and temperature-dependent heat of hydration of cementitious pastes was predicted using the machine learning Gaussian process regression (GPR) with information on the chemical and physical characteristics of the cementitious systems. Results show that high-fidelity heat of hydration predictions can be achieved using the GPR model when compared with isothermal calorimetry experiments. Moreover, the predicted heat of hydration was successfully used to perform mass concrete thermal modeling, which demonstrates the applicability of the model when upscaled to depict concrete performance.

Keywords Heat of hydration · Calorimetry · Machine learning · Gaussian process regression · Mass concrete

L. E. Al-Hasani · G. Perez · H. N. Herndon ·
I. Tien · Y. K. Cho · K. E. Kurtis (✉)
School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
e-mail: kkurtis@ce.gatech.edu

J. Park
Department of Built Environment, Indiana State University, Terre Haute, IN 47809, USA

J. L. Poole
MJ2 Consulting, Cedar Park, TX 78630, USA

N. R. Washburn
Department of Chemistry, Carnegie Mellon University, Pittsburg, PA 15289, USA

T. R. Gentry
School of Architecture, Georgia Institute of Technology, Atlanta, GA 30332, USA

1 Introduction

Cement hydration is a highly exothermic process. Determining the rate of heat release and the cumulative heat of hydration is essential in understanding cement behavior and useful in predicting the property evolution of the hardening material. In massive structures for instance, the exothermic nature of cement hydration leads to a rise in heat energy in concrete elements and a subsequent temperature rise. Information on the heat of hydration of cementitious mixes is required to predict the heat of hydration of

concrete and the adiabatic and internal temperature rises in mass elements. Adiabatic refers to the maximum theoretical temperature that concrete can attain. Internal temperatures are calculated from information on the adiabatic temperature rise, the initial placement temperature, and any losses due to heat transfer with the surroundings [1].

For a certain mix design, finding the adiabatic temperature rise requires that heat of hydration time-histories are measured at a range of isothermal temperatures thought to be representative of mass concrete temperatures. Isothermal calorimetry is a widely used, robust, and standardized laboratory technique, in accordance with ASTM C1679 [2], to measure the rate of heat release of cementitious mixes at isothermal test temperatures. The rate of heat release, also known as power, is related to the rates of reaction of a given system. Cumulative heat of hydration is calculated as the area under the rate of heat release curve. The procedure to find adiabatic temperature rise from isothermal calorimetry heat of hydration can be found in the literature [3–5].

As an alternative to the experimental isothermal calorimetry, semi-empirical kinetic-based models can be used to model the rate of heat release and cumulative heat of hydration of cement when information about the chemical and physical characteristics of the cement is available. The mechanisms controlling the hydration of cement can be explained by dividing the rate of heat release curve into three main periods: (1) up to the induction period, (2) the main hydration peak, and 3) the period after the hydration peak [6]. In a recent review on C_3S hydration, which is the most abundant phase in the cement composition and controls early age hydration, it was demonstrated that different mechanisms control the different hydration phases. The induction phase is best modeled by dissolution controlled by undersaturation. The main hydration peak is explained by C–S–H precipitation on cement grains in the form of needles. And finally, a combination of space-filling, precipitation, and dissolution mechanisms needs to be considered for the third and last phase [7]. Therefore, each period requires a separate model in its simulation. Moreover, accurately modeling the heat of hydration requires including the contribution of other phases, like C_3A, which complicates the modeling process.

In addition to cement, the concrete mix design can also include supplementary cementitious materials (SCMs) and fillers, which impact the heat of hydration development [8]. SCMs and fillers are used to improve the performance of concrete to be more suitable for specific use (e.g., enhancing a concrete's durability performance) and promote more sustainable concrete mix designs [9]. In mass concrete, SCMs are used as a partial low-heat replacement for Portland cement to reduce the maximum concrete temperatures. SCMs include fly ash and slag. Limestone is an example of a filler. The addition of these materials means that an appropriate model needs to be able to simulate all the phases and materials in a specific mixture, which further complicates the modeling procedure. Finally, additional considerations would need to be made to consider the effect of different curing temperatures on the heat of hydration development [10].

Since a single kinetic model that describes the entirety of the heat evolution process does not capture the complexity of the reactions, a combination of models which describe different phases can be used. The latter approach renders the modeling efforts complicated and labor-intensive [11]. Alternative to kinetic-based models, empirical models exist in the literature to find the rate of heat release and cumulative heat of hydration. Schindler and Folliard proposed a three-parameter exponential model to fit the cumulative heat of hydration curves [12]. The three parameters can be found either by performing calorimetry experiments to find the best fit of the model or using regression models that have been trained on the characteristics of a few hundred cementitious systems [13]. There are a few drawbacks to using the three-parameter exponential model, which include the inability of the regression models for the three parameters to explain more modern cements and the inability of the form of the three-parameter exponential model to fit heat of hydration curves of more complicated systems, such as those containing high amounts of fly ash or slag [14].

There is a critical need for a modeling approach that is capable of accurately predicting the heat of hydration of more complex blended cementitious systems and at various curing temperatures to be used for mass concrete analysis. A promising approach to predict the properties of cement and concrete using artificial intelligence machine learning approaches [15], which would overcome some of the complexities associated with more analytical models. For example, machine learning has been used to predict cement phase

mechanical properties [16] and constituent phases [17] from microstructural maps. Mechanical and thermal properties of concrete have also been predicted, such as compressive strength [18], thermal conductivity [19], the coefficient of thermal expansion [20], and the adiabatic temperature rise [21, 22], among many other applications. More specific to the heat of hydration, Wang et al. have modeled early age hydration kinetics using a flexible neural tree algorithm for Portland cement (PC) [23]. Similarly, early age hydration ($\sim$ 24 h) predictions for blended cements containing limestone and metakaolin have been performed by Cook et al. with high-fidelity using random forests [11]. The previous examples demonstrate the promise of using machine learning techniques to predict materials performance by providing knowledge of cementitious systems' physical and chemical characteristics.

Prior machine learning efforts to predict heat of hydration have explored PC systems [23] or blended systems. However, these are not appropriate for mass concrete; i.e., they contain high heat metakaolin [11]. Moreover, the existing models predict hydration kinetics at room temperature only. Here, a Gaussian process regression (GPR) model is proposed to predict the cumulative heat of hydration used in the thermal simulation of mass concrete structures. GPR is a non-parametric regression method, which makes it generalizable even when trained on small datasets, such as that developed here. Moreover, GPR is a Bayesian approach [24], and therefore provides an uncertainty estimate for its predictions which is useful for decision-making processes [25], including the design problem addressed here. In this work, the predictions are performed for blended cements containing fly ash, blast furnace slag, limestone, or a combination of the above, at curing temperatures ranging between 5 °C and 60 °C. A database of mixes with different cementitious physical and chemical characteristics, SCM and filler substitution, water/solids ratio (w/s), and curing temperatures was considered. The proposed model was validated against an instrumented concrete structure to assess the prediction quality of maximum temperatures in two mass concrete mid-scale experiments.

2 Materials and methods

2.1 Data collection

2.1.1 Isothermal calorimetry

Isothermal calorimetry provides an experimentally measured rate of heat release in watts, also denoted as thermal power (P), as a function of time at a given isothermal temperature. The thermal power per gram of solids can be found by normalizing the experimental curves with respect to the mass content of solids in the overall test specimen. The cumulative heat of hydration (H) in Joules per gram is the area under the normalized thermal power curve, as by definition:

$$H(t) = \int_0^t P(t)\mathrm{d}t \qquad (1)$$

Isothermal calorimetry tests were conducted in accordance with ASTM C1679 [26] on 23 different PC and blended PC-SCM mixes. The materials included in the tests were two different ASTM C150 Type I/II Portland cements (PC1, PC2), ASTM C150 Type II (MH) cement, ASTM C618 Class C fly ash (FC), ASTM C618 Class F fly ash (FF), Class CF fly ash which is a blend of Class C and Class F fly ashes (FCF), and ASTM C989 Grade 100 blast furnace slag. The w/s ratios ranged between 0.35 and 0.55. The tests were performed at temperatures 10, 23, 30, 40, 50 and 60 °C. This results in a total of 133 isothermal calorimetry tests. Table 1 shows a summary of the mix designs of the performed tests.

The tests were performed using an eight-channel (TAM Air) microcalorimeter, which has a precision of ±2mW and an accuracy greater than 95%. The temperature of the specimens was kept as close as possible to the temperature of the calorimeter to avoid condensation. Specimens were heated using an electric oven to maintain the temperature of water and cementitious materials between 30 and 60 °C before mixing. An environmental chamber was used to maintain the materials at 10 °C.

2.1.2 Data supplementation

In addition to the performed isothermal calorimetry experiments (denoted source 1), the database was supplemented using the heat of hydration histories collected from the literature, as follows: a dataset of 40

Table 1 Mixes used in isothermal calorimetry tests

No.	Cement type	SCM type	SCM (mass %)	w/s ratio	Test temperature (°C)
1	PC1	–	–	0.35	10, 23, 30, 40, 50, 60
2	PC1	–	–	0.45	10, 23, 30, 40, 50, 60
3	PC1	–	–	0.55	10, 23, 30, 40, 50, 60
4	PC1	FF	20	0.35	10, 23, 30, 40, 50, 60
5	PC1	FF	40	0.35	10, 23, 30, 40, 50, 60
6	PC1	FF	20	0.45	10, 23, 30, 40, 50, 60
7	PC1	FF	40	0.45	10, 23, 30, 40, 50, 60
8	PC1	FC	40	0.35	10, 23, 30, 40, 50
9	PC1	FC	20	0.45	10, 23, 30, 40, 50, 60
10	PC1	FC	40	0.45	10, 23, 30, 40, 50, 60
11	PC1	FCF	20	0.35	10, 23, 30, 40, 50, 60
12	PC1	FCF	40	0.35	10, 23, 30, 40, 50, 60
13	PC1	FCF	20	0.45	10, 23, 30, 40, , 50, 60
14	PC1	FCF	40	0.45	10, 23, 30, 40, 50, 60
15	PC1	Slag	40	0.45	10, 23, 30, 40, 50, 60
16	PC1	Slag	60	0.45	10, 23, 30, 40, 50, 60
17	PC2	–	–	0.45	10, 23, 30, 40, 50, 60
18	PC2	FF	25	0.45	23, 50, 60
19	PC2	FF	45	0.45	10, 23, 30, 40, 50, 60
20	PC2	FF/Slag	25/20	0.45	10, 23, 30, 40, 50, 60
21	MH	–	–	0.35	10, 23, 30, 40, 50, 60
22	MH	–	–	0.45	10, 23, 30, 40, 50, 60
23	MH	–	–	0.55	10, 23, 30, 40, 50

different PC mixes and blended mixes containing fly ash and blast furnace slag, conducted at 5 different temperatures (5, 15, 23, 38 and 60 °C), resulting in a total of 200 observations (source 2) [27], 40 heat of hydration curves of PC and limestone cement pastes with fly ash and slag substitutions at room temperature (source 3) [28], 29 curves of PC pastes, limestone cement pastes, and PC-limestone blends at room temperature (source 4) [29], and 5 tests of PC paste at different temperatures (source 5) [30]. The total number of tests is therefore 407, where 133 are unique cementitious systems and the remaining include variations of the same systems at varying test temperatures. The data from the different sources is different due to the varying temperatures and different SCMs used. However, the data has been grouped together since all isothermal calorimetry tests were performed following the same methodology.

2.2 Machine learning approach

A machine learning approach was utilized to create a model for the prediction of the cumulative heat of

hydration. The model is applicable and generalizable to mix designs whose properties are similar to those used for training the model, i.e. those containing ASTM C150 Type I/II cements, ASTM C595 Type IL cements, and the following supplementary materials: low-lime fly ash, high-lime fly ash, blast furnace slag, and limestone. Many of the mix designs were selected based on their common use for mass concrete construction.

2.2.1 The selection of features

The output of the machine learning model is the cumulative heat of hydration in Joules per gram of paste solids. The cumulative heat of hydration evolves with time, and therefore its values were retained every around 30 min up to 72 h. This culminates in a total of 106,814 total recorded instances. The predictors, or input features, were selected based on their known exothermic effects on the heat of hydration during the first 72 h. Since the heat of hydration is time-dependent, time (hrs) is included as a feature. A change in the curing temperature (or the isothermal

temperature at which the tests were conducted) greatly influences the heat of hydration, where a higher curing temperature increases both the rate of heat release and the ultimate heat value [31]. Finding heat of hydration curves using isothermal calorimetry at different temperatures is necessary for understanding the effect of evolving internal temperatures—expected in massive elements—on the subsequent in-situ rate of heat release [4]. Curing temperatures are provided in °C. The water to solids ratio and fineness of the cement also accelerate the rate of heat release and increase the cumulative heat of hydration with time [31]. For the cement, the Blaine fineness in m^2/kg was included. The w/s ratio is unitless.

The phase composition of the cement should also be considered. Hydration kinetics in the first few days are mainly driven by the C_3S and C_3A contents. C_3S, for instance, is known to significantly affect the rate at which heat is released [32]. Similarly, the chemical composition of cement was found to affect early age hydration. A higher equivalent alkali content accelerates early age hydration and reduces the induction period [33], and a higher SO_3 content accelerates the C_3S reaction, which increases the rate of heat release [34]. The phases of the cement and the chemical oxides were included as features as mass percent of the cement content.

Additional features include the contents and characteristics of the SCMs and the fillers used. Because it is used as a diluent for cement, fly ash (FA) retards hydration in both the induction and acceleration periods [35], and the impact was found to be affected by the fly ash substitution ratio and its lime (CaO) content [12]. The slag content, similarly, retards the rate of heat release and reduces the cumulative heat of hydration [36]. Cement substitution by limestone (LS) accelerates cement hydration and increases cumulative heat if the substitution ratio is less than 10% [37]. Moreover, a larger specific surface area of the materials results in a short induction period and a higher rate of acceleration [38]. Therefore, seven features which include the percent content of the materials, their specific surface areas in kg/m^2, and the fly ash CaO content were added.

One final feature; a rate factor, is used to depict the temperature sensitivity of the hydration reaction [39], which is the Arrhenius calculation for every test entry normalized with respect to the Arrhenius calculation at a reference temperature, as follows:

$$\text{Rate factor} = \frac{\exp\left(-\frac{E_a}{R.T}\right)}{\exp\left(-\frac{E_a}{R.T_{\text{ref}}}\right)} \tag{2}$$

where E_a is the apparent activation energy of the paste in J/mol, R is the gas constant and equals 8.314 J/mol.K, T is the temperature in Kelvin, and T_{ref} is the reference temperature in Kelvin equivalent to 23°C [40]. Equation (2) requires a calculation of the activation energy, which in this study was either found using the single linear approximation method [41] for the group of data where tests at different isothermal temperatures are available or approximated using regression models [42] when the paste was only tested at a single temperature.

The dataset is preprocessed to account for any missing values, which were encountered for the equivalent alkali content of three cements and the specific surface area (SSA) of five SCMs; more specifically, four fly ashes and one slag. Since values were missing at random, regression imputation was used to fit statistical models to the variables with the missing data [43]. Data imputation is performed since the dataset is relatively small, and therefore it is preferred over leaving out incomplete data. Pearson's correlation was first performed to determine the correlation between the different variables. It was determined that the equivalent alkali content had good correlation with cement's composition phases (C_3S, C_2S, C_3A, and C_4AF), in addition to the MgO and SO_3 contents. The R^2 value of the analysis was 0.69. The SSA of fly ash was correlated with the maximum value for heat of hydration, the fly ash content, and the fly ash CaO content, with an R^2 value of 0.95. The SSA of slag was correlated with the maximum value for heat of hydration and the slag content, with an R^2 value of 0.89. The smaller R^2 value could be due to lack of information on compositional content for slag. However, the value is still considered satisfactory for the purposes of this work. In the condition where different variable values were determined for the same material, the average value was taken. The total count of input features is 16. The statistical distribution of features of the individual data sources is summarized in Table 2.

2.2.2 Gaussian process regression

GPR is a supervised, non-parametric, probabilistic algorithm that defines a distribution over functions given certain data. Some of the advantages of GPR

Table 2 Statistical analysis of data from different sources

Variable	Unit	Data source 1				Data source 2				Data source 3				Data source 4				Data source 5			
		Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max
Cumulative heat	J/g	173.9	87.3	0.0	365.2	151.3	90.2	0.0	341.2	179.0	72.7	0.0	296.1	194.9	71.4	0.0	306.4	272.3	103.1	0.0	391.1
Time	Hour	34.8	20.6	0.00	71.9	34.5	20.4	0.0	72.0	35.9	20.8	0.0	72.0	35.9	20.8	0.0	72.0	33.8	20.2	0.0	72.0
w/s	Unitless	0.42	0.05	0.35	0.55	0.44	0.00	0.44	0.44	0.40	0.00	0.40	0.40	0.38	0.00	0.38	0.40	0.41	0.00	0.41	0.41
Temperature	°C	34.2	16.1	10.0	60.0	27.4	19.0	5.0	60.0	25.0	0.0	25.0	25.0	25.0	00.0	25.0	25.0	33.5	13.7	15.0	55.0
Blaine	m²/kg	407.3	19.6	303.0	411.0	379.7	12.5	365.4	390.9	524.2	66.3	428.0	606.0	487.5	64.1	395.0	597.0	373.0	0.0	373.0	373.0
C₃S	%	58.1	4.3	49.0	62.0	63.1	3.1	57.2	66.5	56.8	4.7	50.8	64.3	58.0	3.6	50.8	62.0	63.6	0.0	63.6	63.6
C₃A	%	7.4	1.4	0.0	7.8	6.8	2.7	3.9	9.6	3.4	1.1	1.5	5.0	5.4	2.3	2.0	7.0	2.3	0.0	2.3	2.3
SO₃	%	2.9	0.2	2.8	3.4	2.9	0.5	2.4	4.2	3.1	0.1	2.9	3.3	3.3	0.1	3.0	3.5	2.6	0.0	2.6	2.6
Na₂Oₑq	%	0.50	0.10	0.25	0.56	0.67	0.09	0.57	0.85	0.44	0.07	0.30	0.50	0.50	0.09	0.37	0.60	0.69	0.00	0.69	0.69
FA	%	24.7	15.4	0.0	45.0	20.1	15.3	0.0	40.0	9.4	7.3	0.0	15.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
FA CaO	%	10.4	7.7	0.0	22.1	10.6	17.8	0.0	28.9	9.1	11.7	0.0	26.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
FA SSA	m²/kg	349.4	206.3	0.0	594.0	320.9	290.1	0.0	744.0	301.5	259.9	0.0	627.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Slag	%	7.4	16.9	0.0	60.0	12.6	23.3	0.0	70.0	15.6	23.1	0.0	50.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Slag SSA	m²/kg	109.2	228.4	0.0	587.1	150.1	262.7	0.0	610.0	176.6	262.2	0.0	566.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Limestone	%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	18.3	7.6	0.0	30.0	0.0	0.0	0.0	0.0
LS SSA	m²/kg	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	492.4	307.3	0.0	917.7	0.0	0.0	0.0	0.0
Rate factor	Unitless	2.4	1.9	0.4	8.2	1.9	1.9	0.3	8.7	1.1	0.0	1.1	1.1	1.1	0.0	1.1	1.1	1.9	1.2	0.7	4.4

Std refers to standard deviation

include the generalizability of the model for datasets that are not large, and the uncertainty estimates it provides for its predictions. The approach is based on defining a prior over functions and translating it into a posterior over functions after observations are introduced into the training process. The fact that the performed predictions are probabilistic is advantageous since GPR is able to compute a confidence interval for the response. The approach is also kernel-based, which enables it to handle nonlinear data, as is the case with the heat of hydration histories, with great efficiency.

For a training dataset $D = \{x_i, y_i\}, i = 1, \ldots, n$, it can be defined that $x \in R^{dxn}$ is the input belonging to a d-dimensional vector space R^{dxn}, and $y \in R^n$ is the output belonging to a one-dimensional vector space R^n. n is the number of data points. Using GPR, the output y can be calculated as:

$$y = f(x) + \varepsilon \tag{3}$$

The error ε follows a normal distribution with a mean value of zero and a variance of σ^2, such that $\varepsilon \sim N(0, \sigma_n^2) \in R^n$. The distribution of y is assumed to be jointly Gaussian in GPR, which is a satisfactory assumption for many applications and has a mean vector $\mu(x)$ and a covariance matrix $\sum(x)$. The covariance matrix is a function of a kernel (κ) which determines the similarity between adjacent observations and the information they provide about each other. It is expected that training observations with similar input vectors will have a similar response. Likewise, similar testing and observed training samples will have similar responses. The covariance matrix takes the following form: $\Sigma_{ij} = \kappa(x_i, x_j)$. To perform predictions, the joint distribution can be written as follows:

$$\begin{pmatrix} f \\ f_* \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu \\ \mu_* \end{pmatrix}, \begin{pmatrix} K & K_* \\ K_*^T & K_{**} \end{pmatrix} \right) \tag{4}$$

where X, f, X_* and f_* are the training input features and response and testing input features and response at a specific instance, respectively. K_* is the covariance matrix for the training and testing dataset and takes the form of $K_* = \kappa(X_*, X)$. Similarly, we find $K = \kappa(X, X)$ and $K_{**} = \kappa(X_*, X_*)$. The posterior can therefore be expressed as:

$$p(f_* | X_*, X, f) = \mathcal{N}(f_* | \mu_*, \Sigma_*)$$

$$\mu_* = \mu(X_*) + K_*^T K^{-1}(f - \mu(X))$$

$$\Sigma_* = K_{**} - K_*^T K^{-1} K_* \tag{5}$$

A suitable kernel function needs to be selected to perform the training process. The choice of kernel decides the hyperparameters that one needs to optimize, which is performed using a Bayesian approach, i.e. maximizing the marginal likelihood. A sample kernel is the Matérn 5/2 kernel and is given by:

$$\kappa\left(x, x^{'}\right) = \sigma_f^2 \left(1 + \sqrt{5}r + \frac{5}{3}r^2 \right) \exp\left(-5\sqrt{r}\right) \tag{6}$$

$$r = \frac{\|x - x^{'}\|_2}{l}$$

The hyperparameters in the above equation are σ_f^2 which controls the vertical variation, and the length scale l which specifies the width of the kernel and implies variation along feature dimensions in the modeled function [44]. Terms x and x' are points in the dataset. It is possible to specify a different length scale for different feature dimensions to determine their relevance by using a product of kernels over the dimensions. For example, the product of Matérn 5/2 kernels is known as Matérn 5/2—ARD, where ARD is the automatic relevance determination [45]. Using this approach, a larger length scale denotes a smaller variation along that dimension and therefore a less relevant feature. Examples of other commonly used kernels are Matérn 3/2—ARD, squared exponential—ARD and rational quadratic—ARD.

2.2.3 Evaluation of model fit

The performance of the model on the testing dataset was quantitatively described using three different statistical measures to allow for a more comprehensive evaluation. The calculations are the coefficient of determination (R^2), the root mean square error (RMSE), and the mean absolute error (MAE). The measures are calculated by comparing predictions ($y^{'}$) and actual measurements (y) through the formulations shown in Eqs. (7) to (9). $\overline{y}$ is the average of the actual measurements in the testing set, and n is the number of observations.

$$R^2 = 1 - \frac{\sum_i \left(y_i - y_i'\right)^2}{\sum_i \left(y_i - \overline{y}_i\right)^2} \qquad (7)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_i \left|y_i - y_i'\right|^2} \qquad (8)$$

$$\text{MAE} = \frac{1}{n}\sum_i \left|y_i - y_i'\right| \qquad (9)$$

2.2.4 GPR model

The collected data was randomly divided into training and testing sets with an 80/20 split, where each isothermal calorimetry test was treated as an individual entry. As mentioned earlier, GPR was selected for its ability to train on small datasets. Here, the number of unique data points in the dataset is small. However, each unique data point is a time series with short time intervals, which makes up a relatively large number of data entries. For this reason, the fit and prediction processes were performed using the Subset of Dataset (SoD) approximation, where the GPR model is applied to a subset of the entire dataset of size $m < n$ to reduce computational complexity [46]. Matérn 5/2—ARD was selected as a suitable kernel function due to its ability to match physical patterns realistically as a result of its differentiability [47]. Separate length scales were used for the predictors using ARD. k-fold cross-validation with the number of folds taken as 10 was performed during model training to assess the measure of performance over different unseen training and testing datasets [48].

It is also important to understand the stability of the model and how it would be affected by changes in the training and testing datasets. Resampling using the bootstrap method was performed to further assess and validate the performance and robustness of the model [49]. The method involves dividing the data into training and testing datasets at random with replacement at a fixed split ratio, and the process is repeated for a determined number of iterations. The performance of the model is evaluated at each iteration. Here, 250 iterations were performed. The performance measures were calculated at every iteration, and their statistical distributions were obtained at the end of the analysis.

3 Results and discussion

The cumulative heat of hydration of different cementitious pastes at different isothermal temperatures was predicted using a GPR model. The dataset was compiled by performing isothermal calorimetry tests to obtain heat of hydration histories, in addition to histories collected from the literature. The model was trained on a randomly selected testing dataset, and the robustness of the model was evaluated using bootstrapping. The applicability of the model to real-life structural elements was validated by upscaling the predicted heat of hydration curves to compare the predicted internal temperatures with those measured from two mid-scale experiments. Finally, the use of the uncertainty estimates that the model provides as part of the decision-making process is demonstrated.

### 3.1 Isothermal calorimetry heat of hydration curves

Representative heat of hydration curves obtained using isothermal calorimetry are illustrated in Fig. 1. Figure 1a shows the effect of w/s on the heat of hydration for a PC mix, where the effect of w/s (within the range examined here) on the rate of heat development is minimal during the first 24 h, and afterward, cumulative heat increases with increasing water content. Cement fineness has a more pronounced effect on the heat of hydration, where a coarser cement releases lower cumulative heat, as shown in Fig. 1b. Figure 1c demonstrates that cement replacement with SCMs leads to lower heat of hydration release with time, as well as a slower rate of heat development as in the case of fly ash, as expected [12, 35]. Moreover, low lime ashes (Class F) experience a lower cumulative heat at 72 h. Figure 1d compares the effect of different limestone fineness $(40, 25, 3\,\mu\text{m})$ on the heat of hydration at a 20% replacement ratio. The extent of reduction on heat of hydration is greater with a coarser limestone [37]. Finally, higher temperatures increase both the rate of heat release and cumulative heat of hydration, as shown in Fig. 1e.

3.2 Prediction of heat of hydration

The heat of hydration was predicted using two approaches: (1) having one model for the entire heat of hydration curve, and (2) having separate models to depict different hydration processes. The GPR model

Fig. 1 Factors affecting cumulative heat of hydration
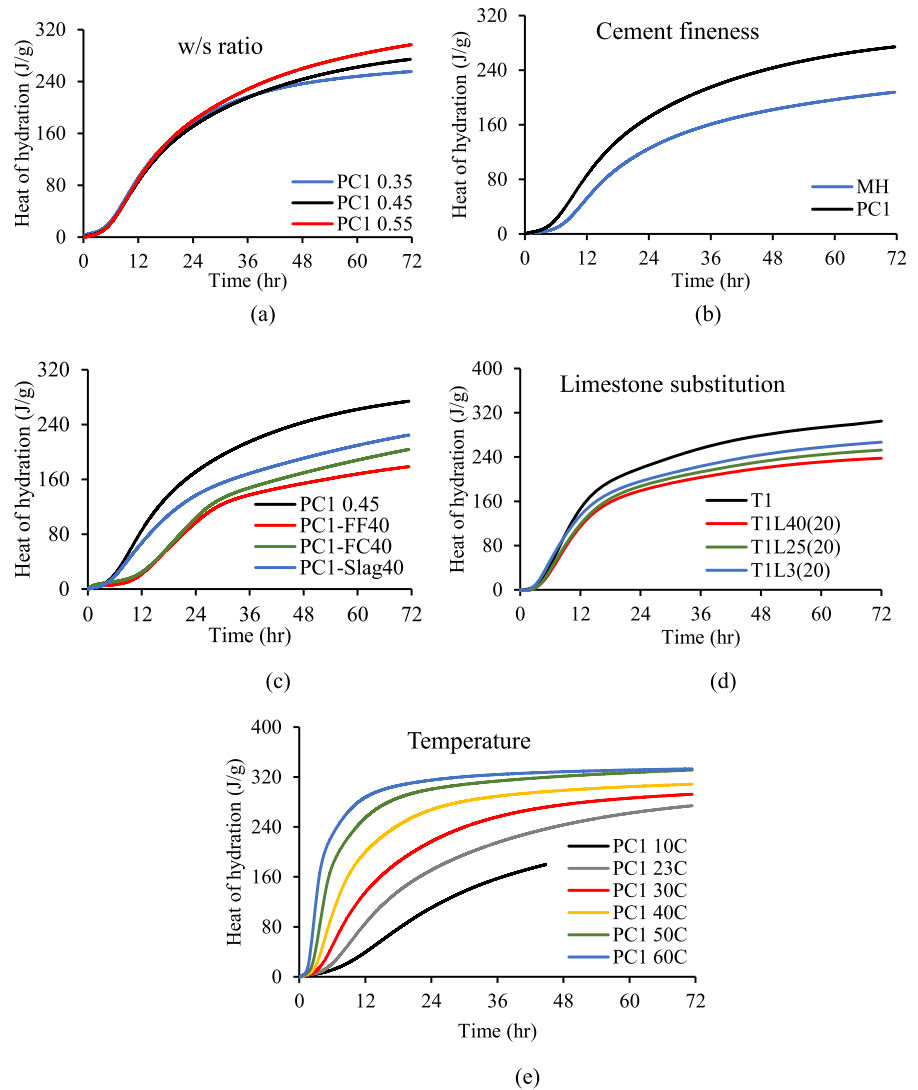
was trained using the training dataset, and the prediction accuracy was evaluated using the testing dataset. First, the heat of hydration was predicted for 72 h, although shorter periods were also considered based on the experiment details. Tenfold cross-validation was conducted in order to assess model performance over unseen data. The average loss of the validation process was found in terms of the root mean square error, and is equal to 8.4 J/g. The standard deviation of the losses is 0.35 J/g. The model's hyperparameters were automatically optimized during training using automatic relevance determination. The log of length scale for all features is shown in Fig. 2 to demonstrate feature relevance,

which is greater for smaller length scale values. It should be noted that the feature relevance might slightly differ based on the randomization of the training and testing datasets. The figure indicates that the SSA of fly ash is among the least relevant features.

Figure 3 illustrates heat of hydration prediction fits and accompanying RMSE for a representative group of cementitious mixes. The predicted heat of hydration matches very well with the observed heat of hydration for all four mixes. All predictions fall within the 95% confidence interval limit. The confidence interval was narrowest for the first mix, which was PC and cured at 23 °C, which indicates that the certainty of prediction is higher for this mix. Performance measures for both
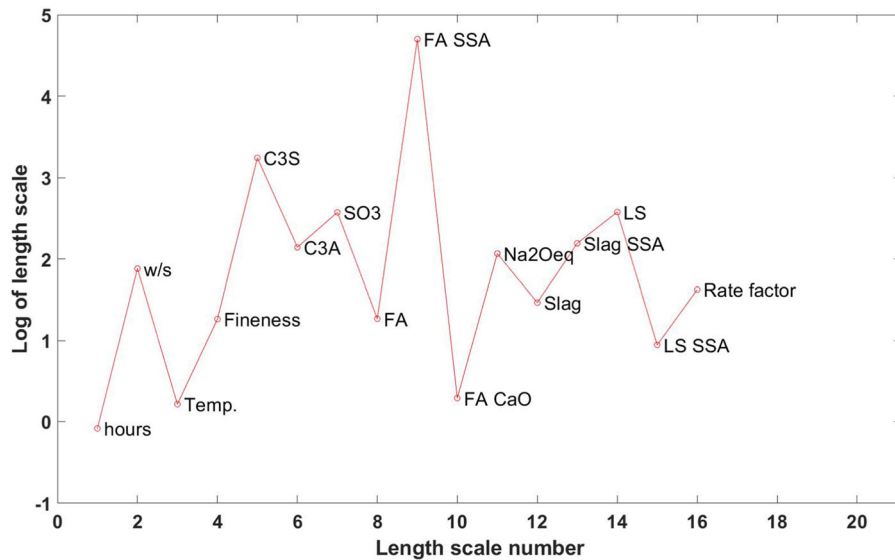
Fig. 2 Log of length scales of features, where a shorter length scale corresponds to greater relevance. FA and LS refer to fly ash and limestone, respectively

Fig. 3 Representative examples of heat of hydration prediction using GPR. PC is Portland cement, FF is Class F fly ash
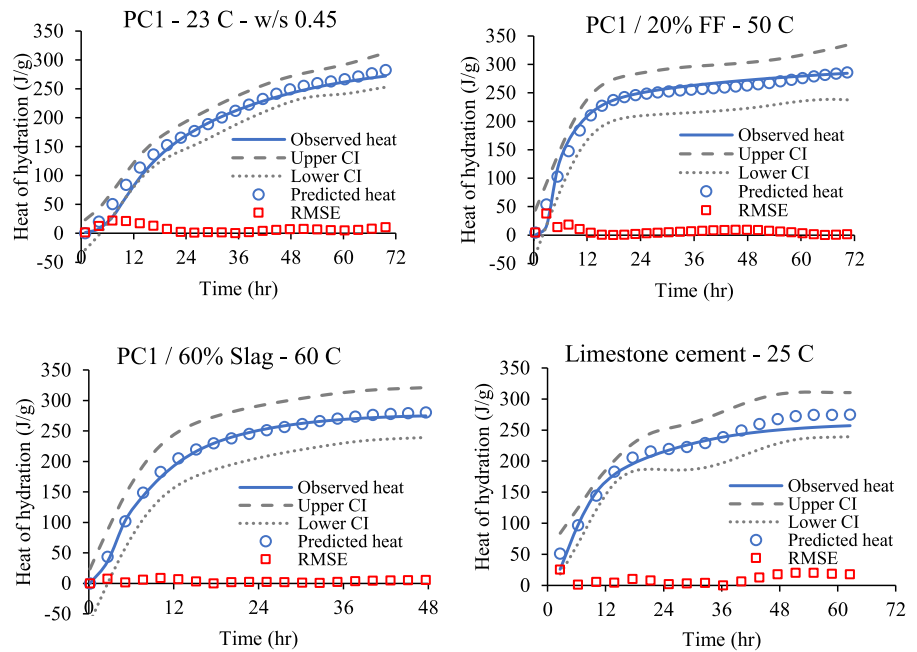
the testing and training sets are summarized in Table 3. It can be observed that some of the predictions in Fig. 3 are subject to changes in the slope with time, which is not a true attribute of the heat of hydration curves. The GPR algorithm is non-parametric, and therefore some flexibility in predictions is anticipated. The predictions will be used to find the internal temperatures of mass concrete elements and the extent

to which prediction errors influence the final results will be explored in the following section.

It is shown in Table 3 that the RMSE error for the training dataset is 7.077 J/g in comparison with 15.35 J/g for the testing dataset. The R^2 of the testing set is 97.2%. Figure 4 gives an example of the distribution of residuals for all mixes in the testing dataset and how it progresses with time. As

Table 3 Evaluation measures of GPR fit

Criteria	R^2	RMSE (J/g)	MAE (J/g)
With 10-fold cross validation			
Training	0.993	7.077	3.509
Testing	0.972	15.35	10.48
After bootstrapping			
Mean	0.974	14.05	9.752
Standard deviation	0.005	1.418	0.781
Maximum	0.984	18.98	12.11
Minimum	0.952	10.91	7.881
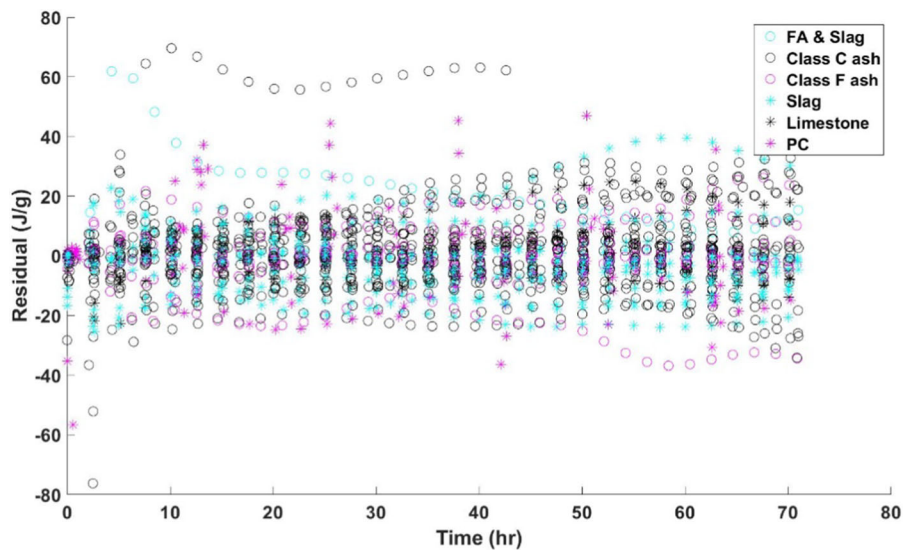
Fig. 4 Distribution of model residuals with time

demonstrated in the figure, the majority of the mixes have similar residuals, which are concentrated around zero. A few mixes containing high-lime fly ash (Class C) are showing slight deviations, which is a result of the random selection of the training and testing datasets. This could also indicate that more data points are needed for the better prediction of cementitious systems with SCMs. It can also be observed that the fluctuation of residuals does not change significantly with time. For the majority of the mixes, the accuracy of prediction was found to be slightly lower in the first 10 h and was relatively stable for the remainder of the period.

As stated earlier, the rate of hydration curve can be divided into three phases: up to the end of the induction period (approximately < 3 h), the main hydration peak (approximately between 3 and 24 h), and after the hydration peak (approximately > 24 h) [7]. It is possible to predict the behavior of each of the phases separately, which can be useful in determining where the largest errors reside, and how features influence the behavior of each of the phases. Next, the results obtained by dividing the rate of hydration curve into two segments are demonstrated. The input features were kept the same for both models. The first model, which trains the heat of hydration data up to 24 h, has training and testing R^2 values of 0.99 and 0.97, respectively. The second model, which trains the heat of hydration data after 24 h, has training and testing R^2 values of 0.99 and 0.95, respectively. Residual development with time is shown in Fig. 5 for both models.

It can be seen from the figure that most residuals are very close to zero when the first 24 h are considered. After the first 24 h, most residuals fall within $\mp 20 J/g$, and the distribution is similar over time. A few mixes
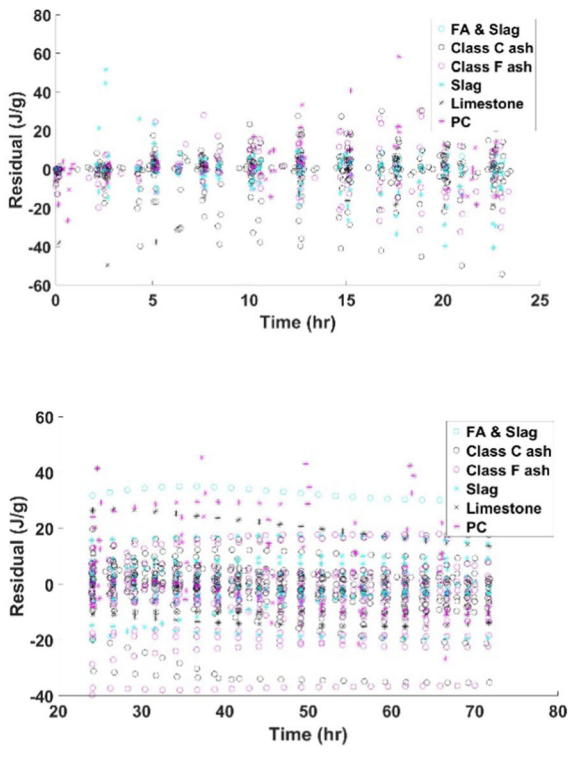
Fig. 5 Distribution of model residuals with time (left) and feature relevance (right) for heat of hydration modeling **a** up to 24 h and **b** later than 24 h. FA and LS refer to fly ash and limestone, respectively

have greater residuals for both models. The results correspond to those obtained earlier in Fig. 4 where the hydration curve was treated in entirety. Length scales for the features of both models are also shown in Fig. 5 to demonstrate feature relevance. The SSA of fly ash and slag are among the least relevant features. Here, it is shown that despite some differences in feature importance and prediction accuracy for both hydration phases, the results do not deviate much from when one model for the entire 72 h is considered, and deviations can be overlooked for the application of interest in this work, which is the thermal modeling of mass concrete structures.

3.3 Evaluating model stability using bootstrapping

The bootstrap method was used to evaluate the stability of the model by resampling the dataset. Here, 250 iterations were performed where the training and testing datasets were selected at random at an 80/20 split ratio. The statistical distributions of the
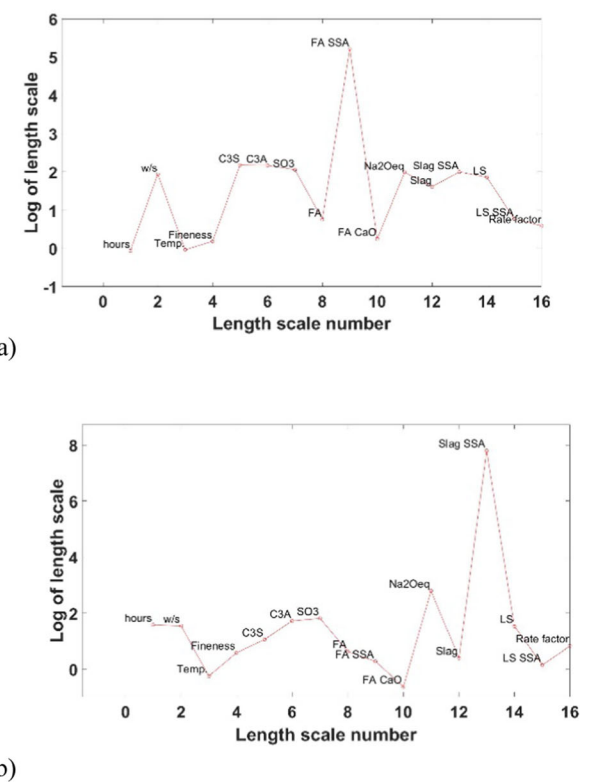
evaluation metrics R^2, RMSE, and MAE from all iterations are shown in Table 3 for the testing dataset. The average performance criteria values are 0.974, 14.05 J/g and 9.752 J/g for R^2, RMSE, and MAE, respectively. The average values demonstrate that the model is able to predict heat of hydration time histories with good accuracy. The standard deviation values further show the stability of the results.

Figure 6 illustrates the statistical distribution using histograms of the performance metrics obtained from the 250 iterations of bootstrapping. The bin width of each histogram was determined using the Freedman-Diaconis rule [50]. It can be concluded from the histograms that the model is stable and robust with the majority of R^2 values greater than 95%. The MAE outcomes are concentrated between around 7.8 J/g and 12 J/g, whereas the RMSE values demonstrate a wider variability. This is expected due to the greater sensitivity of the RMSE method to severe outliers [51]. Overall, the majority of the iterations give satisfactory results.
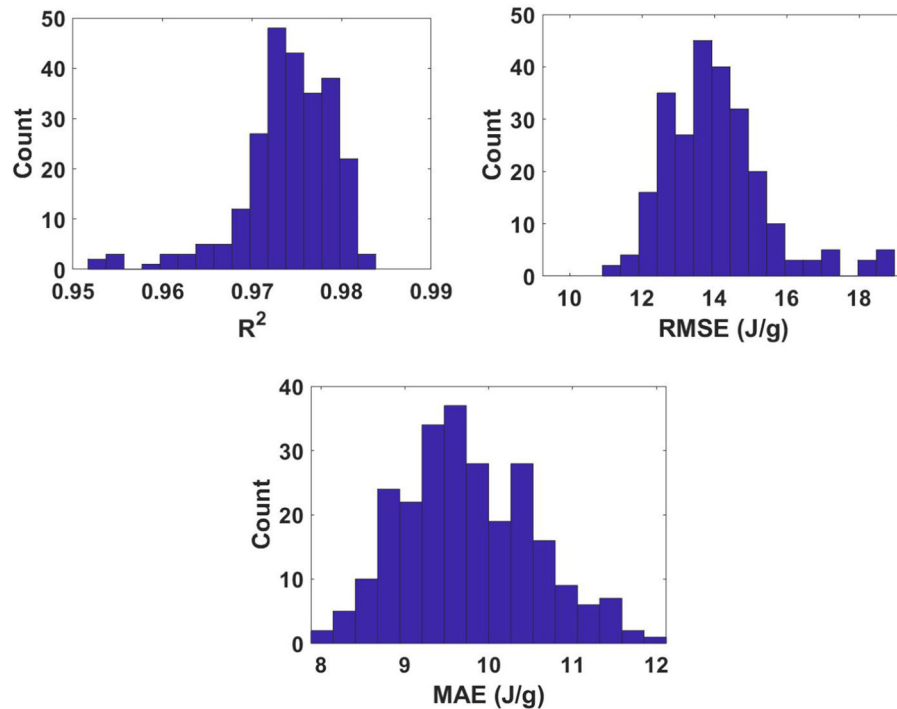
Fig. 6 Histograms showing the distribution of performance metrics obtained from bootstrapping iterations

3.4 Validation with experimental data: adiabatic temperature-rise in mass concrete

This section demonstrates how the developed GPR predictions can be used to predict the internal temperatures of massive concrete structures, which can eventually be used to formulate thermal control plans. The process is illustrated in Fig. 7. Here, the GPR model was used to find the heat of hydration curves required for thermal modeling of two mass concrete mid-scale experiments (MSE1 and MSE2) constructed using instrumented concretes [5]. The experiments involved reinforced concrete prisms with dimensions of $1.21 \times 1.21 \times 1.83$m. For the first experiment (MSE1), a PC concrete mix was used with a water/cement ratio of 0.444. The relevant characteristics of the ASTM C150 PC are shown in Table 4. Construction took place during January in Atlanta, Georgia, USA, and the concrete placement temperature was 13.3 °C. Internal temperature data collection was conducted using thirteen sensors designated with letters L, S and T, situated as shown in Fig. 8.

The second experiment (MSE2) involved a PC concrete mix with 25% Class F fly ash replacement. The water/binder ratio was 0.444. The relevant characteristics of the ASTM C618 fly ash are shown in Table 4. Construction took place during May in Atlanta, Georgia, USA, and the concrete placement temperature was 22.2 °C. The second mid-scale experiment involved an embedded cooling pipe system used as an active thermal control measure [52]. Four open circuit cooling loops using 3/8" PEX cooling pipes, which have an outside diameter of 12.7 mm and an inside diameter of 9.1 mm, were used. Cooling lasted between hours 14 and 22 after concrete placement. The cooling water temperature inside the pipes was 13.3 °C flowing at 0.4 m³/hr. Internal temperature data collection was conducted using eight temperature sensors designated with the letters TS, placed as shown in Fig. 8.

Isothermal calorimetry tests were conducted for both mixture designs at 23, 30, 40, 50 and 60 °C curing temperatures to serve as comparison to the GPR model predictions [5]. For the machine learning model, two additional data points were created, which contain information on the chemical and physical characteristics of the materials in MSE1 and MSE2. The database collected from isothermal calorimetry tests and the literature was used for training the model. Predictions were performed using the GPR model to
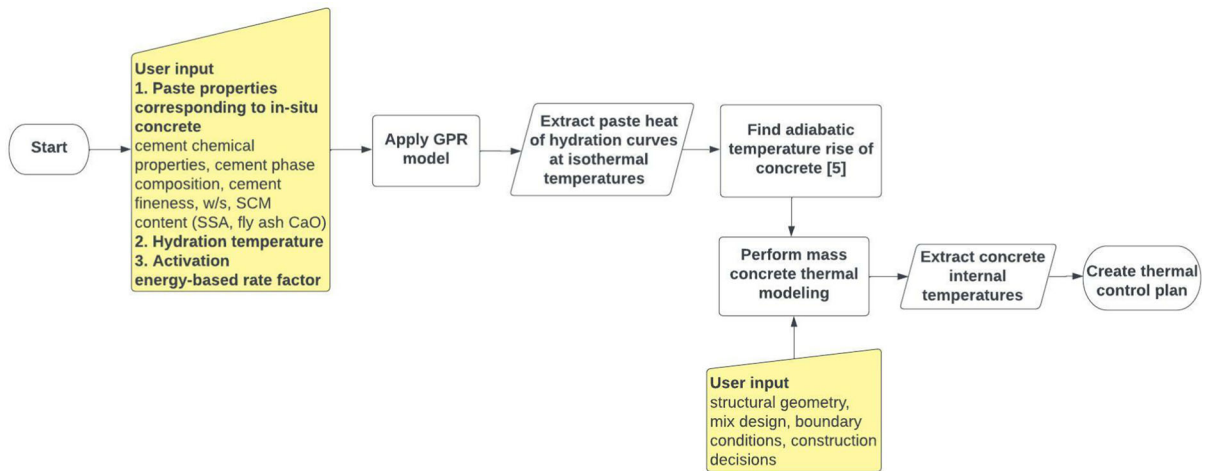
Fig. 7 Diagram of the machine learning-based mass concrete modeling process

Table 4 Chemical and physical characteristics of the materials of the mid-scale experiment

	PC	Class F fly ash
Oxide analysis—mass %		
MgO	1.70	–
SO$_3$	3.30	–
Na$_2$O$_{eq}$	0.48	–
CaO	–	6.99
Phase composition—mass %		
C$_3$S	59	
C$_2$S	12	–
C$_3$A	7	–
C$_4$AF	10	–
Fineness (m2/kg)		
SSA	–	338.9
Blaine	391	–

find the corresponding cumulative heat of hydration for both cementitious pastes at different isothermal temperatures. Figure 9 shows how the GPR model predictions compare to observed data using isothermal calorimetry. The RMSE for the training dataset was found to be 12 J/g, in comparison to 29.9 J/g and 21.1 J/g for MSE1 and MSE2 data, respectively. Similarly, the training dataset had an R^2 value of 0.98, compared to 0.92 and 0.95 for the data of MSE1 and MSE2, respectively.

The results show that the model accurately predicts the magnitude of the heat of hydration, whereas it is underperforming when it comes to predicting the early

rate of heat release. The mixture designs of both prisms included quantities of ASTM C494 high-range water-reducing admixture (PCHRWR) and ASTM C494 Type B&D low-range water-reducing/retarding admixture (WRRET), which are expected to increase the induction period and the rate of heat release [13]. Since the GPR model does not include features on admixtures, their effect on the heat of hydration development was not considered, which may have caused the deviation at times earlier than 12 h.

To predict the internal temperature rise of the concrete prisms, the heat of hydration curves predicted at the various temperatures using GPR was used to find the adiabatic temperature rise of the mixture designs and the corresponding in-situ heat of hydration of concrete. The procedure is explained in the literature [4, 5, 53]. The concrete's heat of hydration was then used in the finite element software b4cast [54] to find internal temperatures at all sensor locations. Figure 10 shows the internal temperature histories found using results from the machine learning model and compares them with measured temperatures. The comparison is conducted for sensors T4 and TS2 for MSE1 and MSE2, respectively. Sensor locations are illustrated in Fig. 8.

Figure 10 shows that the predicted heat of hydration curves was able to simulate the internal temperatures of the mid-scale experiment with good accuracy. For MSE1, results tend to be on the conservative side in comparison to the sensor data with a 4.7% error in the maximum temperature prediction. In contrast, the maximum temperature for MSE2 is underpredicted, with an error of 5.3%. The shape of the temperature rise
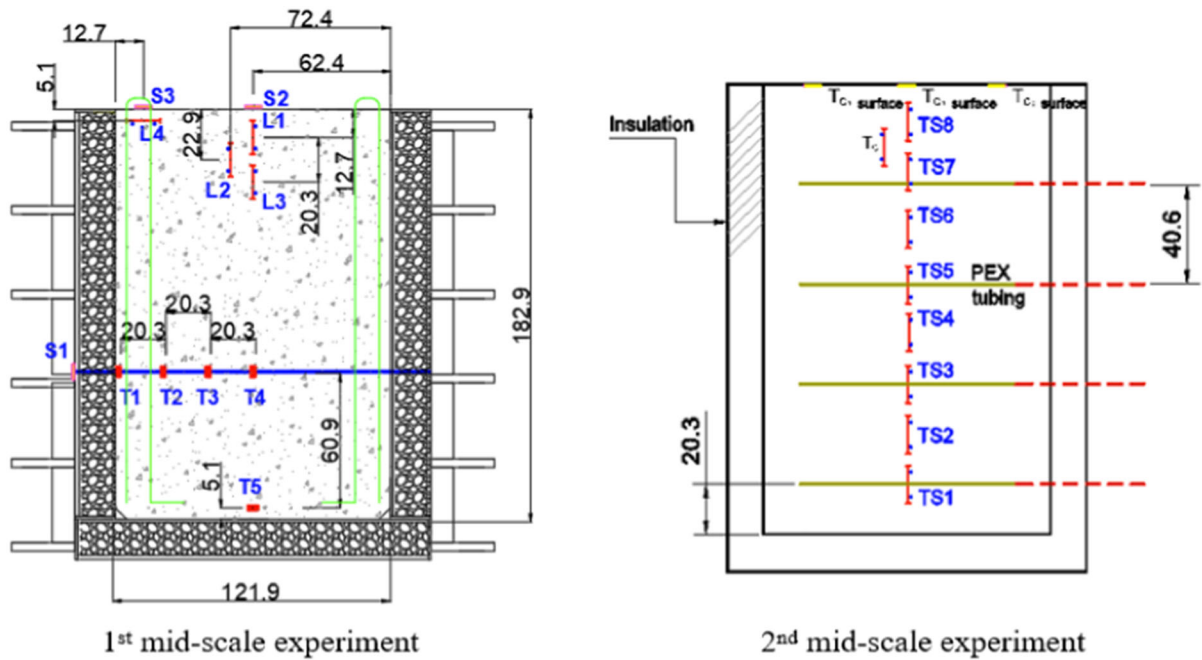
Fig. 8 Mid-scale experiments. All dimensions are in cm
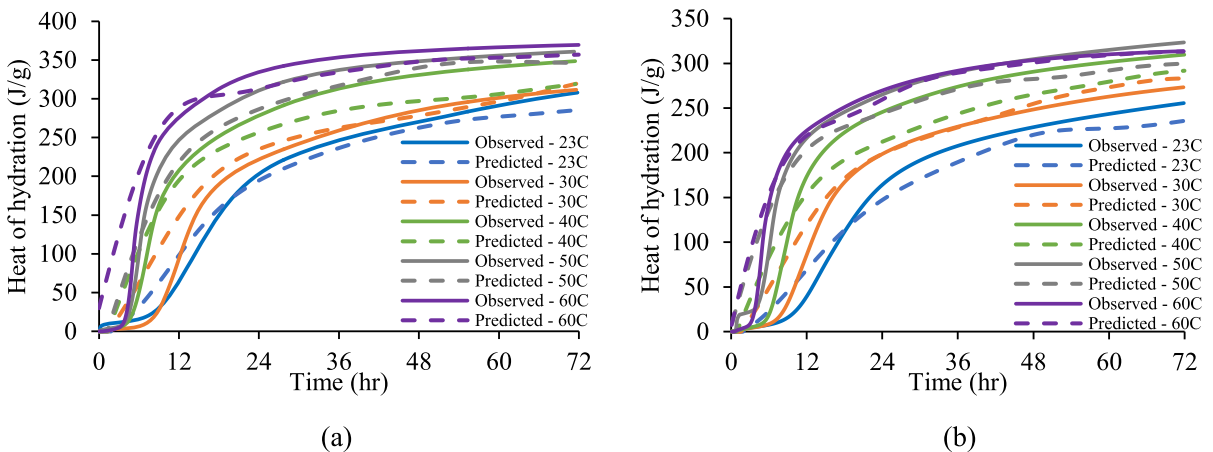

Fig. 9 Heat of hydration curves at different temperatures using isothermal calorimetry (observed) versus GPR model (predicted) for cementitious pastes used in MSE1 **a** and MSE2 **b** concretes

is well-depicted for both experiments, with some deviation in the early rate of heat release due to the uncaptured effect of admixtures.

3.5 Decision-making using uncertainty estimates

In previous attempts, random forest was used to find the heat of hydration of cementitious systems containing metakaolin and limestone with model performance similar to what is reported in Table 3 [11], though the previous model was trained on cementitious pastes of high heat of hydration, which are not suitable for use in mass concrete. Moreover, only room-level curing temperatures were considered, which is not representative of mass concrete cores. One additional advantage of using GPR is that it provides uncertainty estimates for the prediction, which is useful for decision-making scenarios. Here,
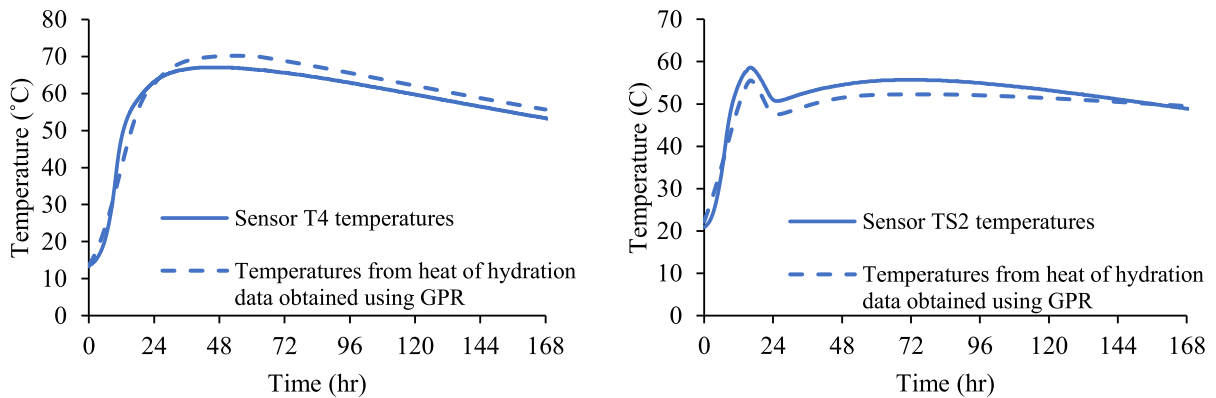
Fig. 10 Measured versus predicted temperature histories of MSE1 at T4 and MSE2 at TS2

a 95% confidence interval was obtained for each cumulative heat of hydration prediction, as previously shown in Fig. 3. The upper and lower confidence limits in the heat of hydration for any cementitious system and at various curing temperatures can be used to find the adiabatic temperature rise and corresponding in-situ heat of hydration, which enables the simulation of an upper and lower limit of possible internal temperature rises in the mass concrete element.

Figure 11 illustrates the uncertainty interval for the internal temperature rise calculated for the second mid-scale experiment. It can be concluded that the model provides more confidence in the predictions and that the maximum temperature is better captured in comparison to the history in Fig. 10, which was calculated using the mean heat of hydration prediction. Deviation from the internal temperatures is still observed after around 24 h, which can be attributed to the difference between the in-situ and predicted boundary condition properties (i.e. formwork and insulation) necessary for the finite element simulation, but the overall shape of the predicted curves matches the measured one well.

3.6 Comparison to empirical and kinetic models

Physics or chemistry-based models for the heat of hydration could be preferred for some applications. Such models were considered at the onset of this investigation but were deemed inappropriate for scaling up to mass concrete use. For empirical and kinetic semi-empirical models, capturing the nonlinear relationship between the chemical and physical characteristics of cementitious blends and the heat of
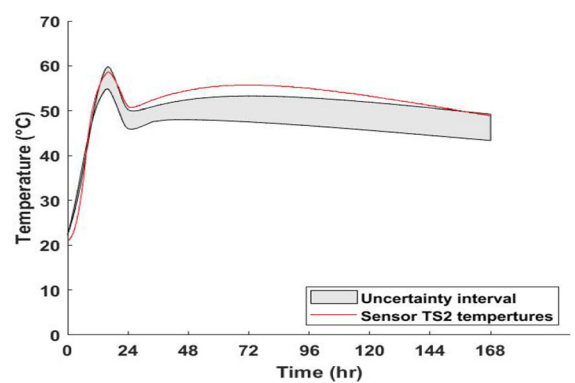
Fig. 11 Uncertainty interval of MSE2 predicted internal temperatures

hydration property is currently not feasible, especially with the incorporation of SCMs, fillers and admixtures, which adds a level of complexity to predictions. Semi-empirical kinetic models which depict the rate of heat release considering the several unique phases of the hydration reaction (initial phase, induction phase, acceleration phase, and deceleration phase) are available. Simple kinetic models, such as diffusion-based models, attempt to depict the heat of the hydration curve in its entirety based on assumptions to reduce the complexity of the reactions [10]. These semi-empirical models do not agree with general theoretical principles [7] and rely heavily on mathematical fitting of each dataset to explain the effects of the different cementitious characteristics, which limits their generalizability and predictive capabilities. Many other approaches have also been theoretically refuted, including nucleation and growth models and confined growth models. All of these models use a certain degree of fitting while relying on visual trial and error

model evaluation instead of more rigorous optimization techniques based on training, testing, and validation [55]. It should also be recalled that heat of hydration predictions of blended systems is currently too complicated using semi-empirical kinetic models, which might not be appropriate for mass concrete simulations.

Alternatively, empirical models which rely on the physical and chemical characteristics of cementitious systems to depict the shape of hydration can be used. One such example is the three-parameter exponential model, which relies on an exponential formulation for the cumulative heat of hydration [12]. The model, however, is not able to describe the shape of more complicated blended cementitious systems [14].

The challenges experienced with empirical and kinetic models can be overcome with the use of machine learning, which has proven to offer a practical approach for the heat of hydration prediction. The utility of machine learning approaches was demonstrated here with GPR, which was able to capture the nonlinear nature of the time and temperature-dependent heat of hydration of cementitious systems. The high-fidelity predictions demonstrate the ability of machine learning models to consider the complex relationship between composition and property while also including the correlations and interactions between different composition variables.

4 Conclusions

Accurate prediction of the heat of hydration of cementitious mixes is necessary to determine many of concrete's mechanical and thermal properties. In mass concrete, the heat of hydration curves at different temperatures are required to find the adiabatic temperature rise of a specific concrete. There is a need in the literature for unified models that are able to capture the heat of hydration of more complex systems, which include supplementary cementitious materials and fillers. For this reason, artificial intelligence machine learning offers a data-driven technique that enables the prediction of the heat of hydration by training on information related to the chemical and physical characteristics of the cementitious materials used, curing temperatures, and mix designs. Here, Gaussian process regression was utilized to model and predict heat of hydration for 407 PC and blended cementitious systems.

Heat of hydration histories were predicted up to 72 h at different isothermal temperatures using GPR. The results were demonstrated for cementitious mixes containing fly ash, blast furnace slag, and limestone at different replacement ratios and different characteristics. During training, model performance was assessed using tenfold cross-validation, and model stability was demonstrated using the bootstrap resampling method. The results have proven the ability of the GPR model to perform the predictions with good accuracy. The average R^2 from all the bootstrapping iterations is around 97.4%. The model was also used to predict the heat of hydration curves for cement from two mid-scale mass concrete experiments, which were subsequently used in finite element modeling to simulate internal concrete temperatures. The ability to use the predictions of the machine learning model and upscale them for application in real-life engineering and decision-making systems has been proven. The model can be expanded in future work to include more variations and combinations of SCMs, include admixtures, and improve the quality of compositional data for SCMs.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. ACI PRC-207.2–07 Report on thermal and volume change effects on cracking of mass concrete. https://www.concrete.org/store/productdetail.aspx?ItemID=207207&Format=DOWNLOAD&Language=English&Units=US_AND_METRIC. Accessed 3 Dec 2021
2. ASTM International (2017) ASTM C1679 Standard practice for measuring hydration kinetics of hydraulic

cementitious mixtures using isothermal calorimetry. ASTM International, West Conshohocken, PA

3. Wadso L (2003) An experimental comparison between isothermal calorimetry, semi-adiabatic calorimetry and solution calorimetry for the study of cement hydration (NT TR 522). NORDTEST, Finland

4. Xu Q, Wang K, Medina C, Engquist B (2015) A mathematical model to predict adiabatic temperatures from isothermal heat evolutions with validation for cementitious materials. Int J Heat Mass Transf 89:333–338

5. Al-Hasani L, Park J, Perez G et al (2022) Quantifying concrete adiabatic temperature rise based on temperature-dependent isothermal calorimetry; modeling and validation. Mater Struct 55:1–20

6. Bullard JW, Jennings HM, Livingston RA et al (2011) Mechanisms of cement hydration. Cem Concr Res 41:1208–1223. https://doi.org/10.1016/j.cemconres.2010.09.011

7. Scrivener K, Ouzia A, Juilland P, Mohamed AK (2019) Advances in understanding cement hydration mechanisms. Cem Concr Res 124:105823

8. Lothenbach B, Scrivener K, Hooton RD (2011) Supplementary cementitious materials. Cem Concr Res 41:1244–1256

9. Juenger MC, Siddique R (2015) Recent advances in understanding the role of supplementary cementitious materials in concrete. Cem Concr Res 78:71–80

10. Lin F, Meyer C (2009) Hydration kinetics modeling of Portland cement considering the effects of curing temperature and applied pressure. Cem Concr Res 39:255–265

11. Cook R, Han T, Childers A, et al (2021) Machine learning for high-fidelity prediction of cement hydration kinetics in blended systems. Mater Des 109920

12. Schindler AK, Folliard KJ (2005) Heat of hydration models for cementitious materials. ACI Mater J 102:24

13. Riding KA, Poole JL, Folliard KJ et al (2012) Modeling hydration of cementitious systems. ACI Mater J 109:225–234

14. Riding KA, Vosahlik J, Bartojay K, et al (2019) Methodology comparison for concrete adiabatic temperature rise. Mater J 116:45–53. https://doi.org/10.14359/51714451

15. Rios R, Childs C, Smith S et al (2021) Advancing cement-based materials design through data science approaches. RILEM Tech Lett; 6:140–149. https://doi.org/10.21809/rilemtechlett.2021.147

16. Ford E, Kailas S, Maneparambil K, Neithalath N (2020) Machine learning approaches to predict the micromechanical properties of cementitious hydration phases from microstructural chemical maps. Constr Build Mater 265:120647

17. Ford E, Maneparambil K, Neithalath N (2021) Machine learning on microstructural chemical maps to classify component phases in cement pastes. J Soft Comput Civ Eng 5:1–20

18. Oey T, Jones S, Bullard JW, Sant G (2020) Machine learning can predict setting behavior and strength evolution of hydrating cement systems. J Am Ceram Soc 103:480–490

19. Sargam Y, Wang K, Cho IH (2021) Machine learning based prediction model for thermal conductivity of concrete. J Build Eng 34:101956

20. Nilsen V, Pham LT, Hibbard M et al (2019) Prediction of concrete coefficient of thermal expansion and other properties using machine learning. Constr Build Mater 220:587–595

21. Trtnik G, Kavčič F, Turk G (2008) The use of artificial neural networks in adiabatic curves modeling. Autom Constr 18:10–15

22. Evsukoff AG, Fairbairn EM, Faria ÉF et al (2006) Modeling adiabatic temperature rise during concrete hydration: a data mining approach. Comput Struct 84:2351–2362

23. Wang L, Yang B, Chen Y et al (2012) Modeling early-age hydration kinetics of Portland cement using flexible neural tree. Neural Comput Appl 21:877–889

24. Rasmussen CE (2003) Gaussian processes in machine learning. In: Summer school on machine learning. Springer, pp 63–71

25. Tien I, Pozzi M, Der Kiureghian A (2016) Probabilistic framework for assessing maximum structural response based on sensor measurements. Struct Saf 61:43–56. https://doi.org/10.1016/j.strusafe.2016.03.003

26. C09 Committee Practice for Measuring Hydration Kinetics of Hydraulic Cementitious Mixtures Using Isothermal Calorimetry. ASTM International

27. Poole JL, Riding KA, Juenger MCG et al (2010) Effects of supplementary cementitious materials on apparent activation energy. J ASTM Int 7:1–16

28. Nadelman EI (2016) Hydration and microstructural development of portland limestone cement-based materials. PhD Thesis, Georgia Institute of Technology

29. Cardelino NH (2018) Design of self-consolidating precast concrete using powdered limestone. PhD Thesis, Georgia Institute of Technology

30. Dolphyn BP (2016) Laminar cracking in post-tensioned concrete nuclear containment buildings. PhD Thesis, Georgia Institute of Technology

31. Van Breugel K (1998) Prediction of temperature development in hardening concrete. Prev Therm Crack Concr Early Ages 15:51–75

32. Bogue RH (1955) The chemistry of Portland cement. LWW, Philadelphia

33. Huang L, Yan P (2019) Effect of alkali content in cement on its hydration kinetics and mechanical properties. Constr Build Mater 228:116833

34. Quennoz A, Scrivener KL (2013) Interactions between alite and C3A-gypsum hydrations in model cements. Cem Concr Res 44:46–54

35. Langan BW, Weng K, Ward MA (2002) Effect of silica fume and fly ash on heat of hydration of Portland cement. Cem Concr Res 32:1045–1051

36. Kolani B, Buffo-Lacarrière L, Sellier A et al (2012) Hydration of slag-blended cements. Cem Concr Compos 34:1009–1018

37. Wang D, Shi C, Farzadnia N et al (2018) A review on use of limestone powder in cement-based materials: mechanism, hydration and microstructures. Constr Build Mater 181:659–672

38. Costoya Fernández MM (2008) Effect of particle size on the hydration kinetics and microstructural development of tricalcium silicate. EPFL, Lausanne

39. Kada-Benameur H, Wirquin E, Duthoit B (2000) Determination of apparent activation energy of concrete by isothermal calorimetry. Cem Concr Res 30:301–305

40. Carino NJ (1984) The maturity method: theory and application. Cem Concr Aggreg 6:61–73

41. Poole JL, Riding KA, Folliard KJ et al (2007) Methods for calculating activation energy for Portland cement. ACI Mater J 104:303–311

42. Riding KA, Poole JL, Folliard KJ et al (2011) New model for estimating apparent activation energy of cementitious systems. ACI Mater J 108:550–557

43. Musil CM, Warner CB, Yobas PK, Jones SL (2002) A comparison of imputation techniques for handling missing data. West J Nurs Res 24:815–829

44. Murphy KP (2012) Machine learning: a probabilistic perspective. MIT Press, Cambridge

45. Duvenaud D (2014) Automatic model construction with Gaussian processes. PhD Thesis, University of Cambridge

46. Chalupka K, Williams CKI, Murray I (2013) A framework for evaluating approximation methods for gaussian process regression. J Mach Learn Res 14:333–350

47. Stein ML (1999) Interpolation of spatial data: some theory for kriging. Science, Berlin

48. The 'K' in K-fold Cross Validation. https://arpi.unipi.it/handle/11568/962587. Accessed 6 Dec 2021

49. De Bin R, Janitza S, Sauerbrei W, Boulesteix A-L (2016) Subsampling versus bootstrapping in resampling-based model selection for multivariable regression. Biometrics 72:272–280. https://doi.org/10.1111/biom.12381

50. Freedman D, Diaconis P (1981) On the histogram as a density estimator: L 2 theory. Z Für Wahrscheinlichkeitstheorie Verwandte Geb 57:453–476

51. Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)?: Arguments against avoiding RMSE in the literature. Geosci Model Dev 7:1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

52. Maggenti R (2007) From passive to active thermal control. Concr Int 29:24–30

53. Wadso, L. (2003) An experimental comparison between isothermal calorimetry, semi-adiabatic calorimetry and solution calorimetry for the study of cement hydration (NT TR 522). In: NORDTEST. http://www.nordtest.info/wp/2003/03/28/an-experimental-comparison-between-isothermal-calorimetry-semi-adiabatic-calorimetry-and-solution-calorimetry-for-the-study-of-cement-hydration-nt-tr-522/. Accessed 5 Dec 2021

54. b4cast - Simulation of Hardening Concrete. http://www.b4cast.com/b4cast/b4cast.html. Accessed 1 Apr 14AD

55. Ouzia A, Scrivener K (2019) The needle model: a new model for the main hydration peak of alite. Cem Concr Res 115:339–360