Design and Discovery of Sustainable Cementitious Binders Through Machine Learning Trained from a Small Database

Authors: Taihao Han, Aditya Kumar, Jie Huang, Gaurav Sant, Narayanan Neithalath

Presented by: **Taihao Han**, Ph.D.

Postdoctoral Researcher, Department of Materials Science & Engineering,

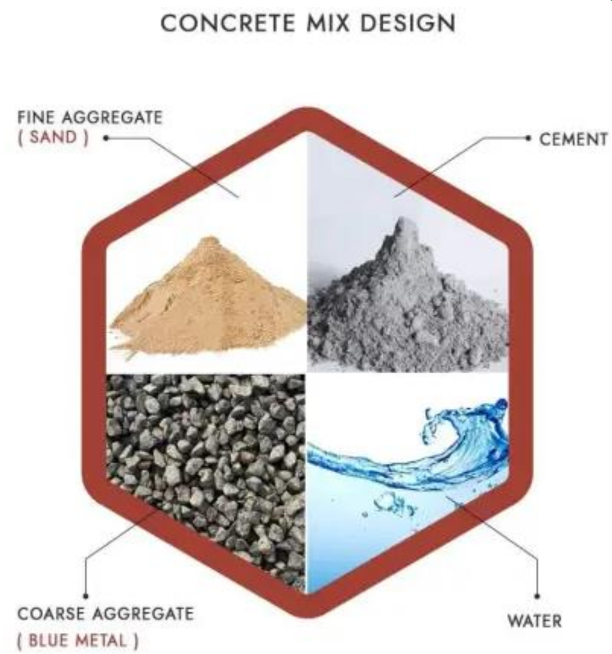
Missouri University of Science & Technology, Rolla, MO 65409

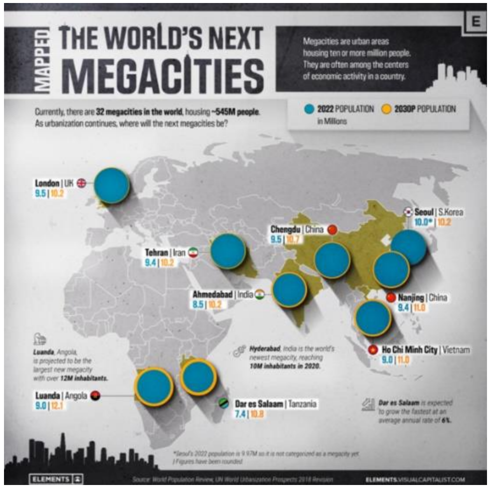
ACI Concrete Convention

San Francisco, CA, April 2nd, 2023

Concrete: Overview

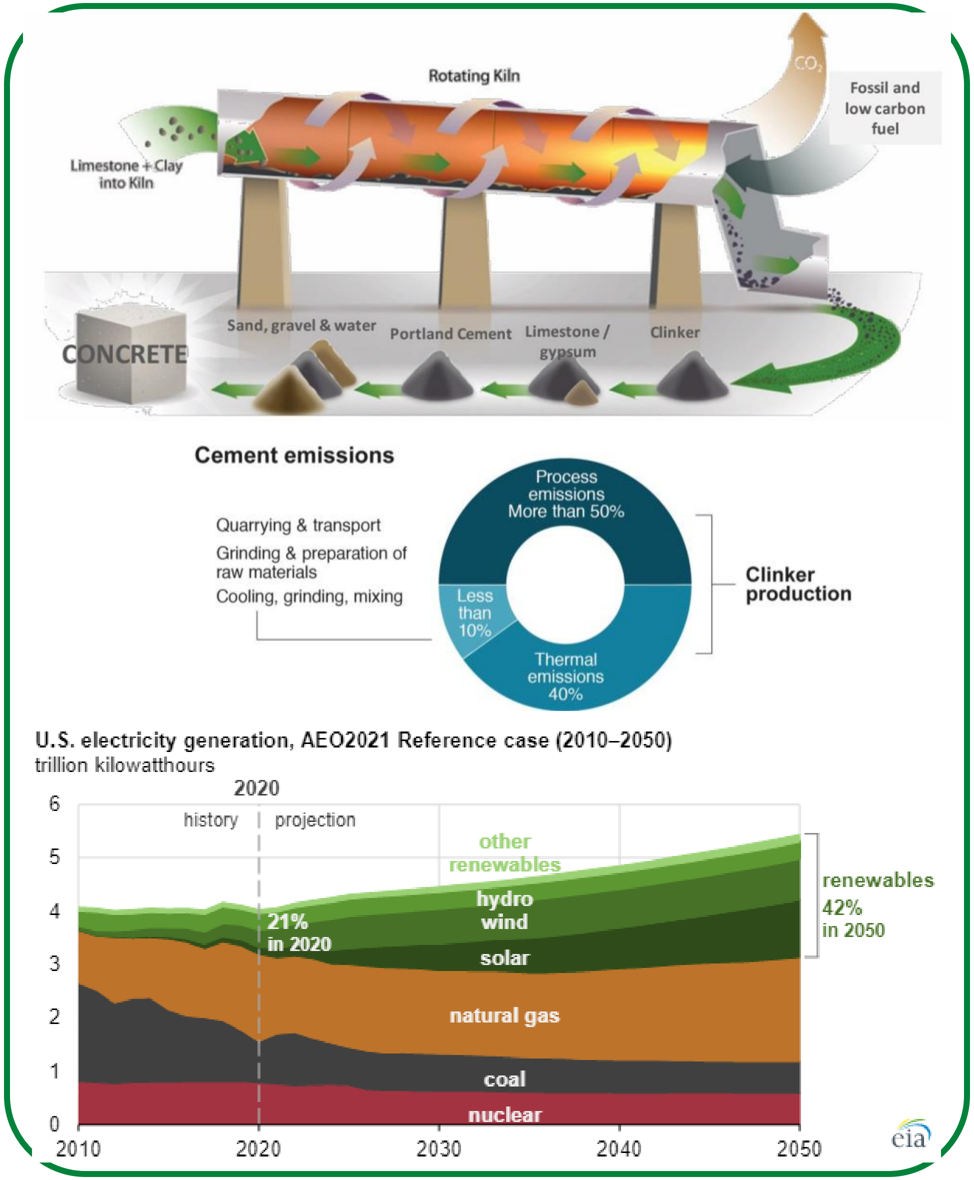
- Concrete: The principal material for construction of (nearly all) infrastructure
- Concrete = Portland cement + water + sand + stones + chemicals (to regulate properties)
- Production-and-use: 40 billion tons/year
  - Employs 10 million Americans
  - Creates $1.3 trillion worth of engineered systems
- 2050 projections
  - Global population: 10 billion
  - (Sub)urbanites: 6.6 billion
  - Rise of >50 megacities
  - Concrete production: ~60 billion tons/year





Concrete: The ugly

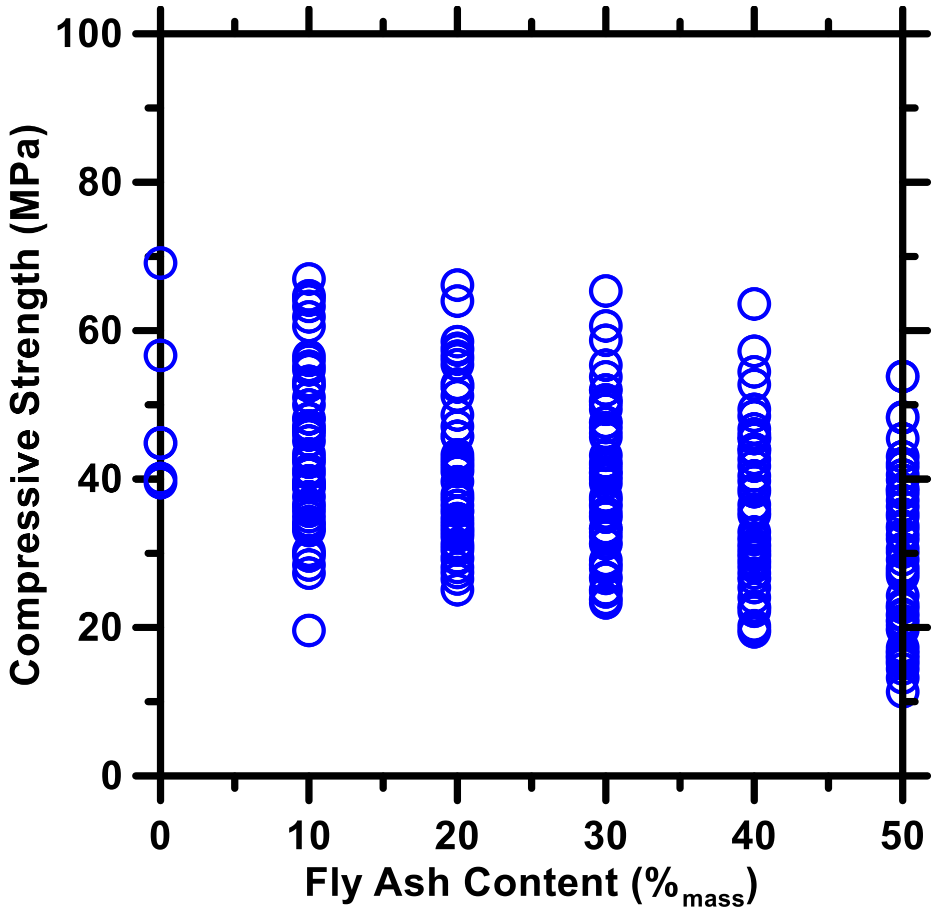
- The **carbon-footprint** problem
- Limestone ($CaCO_3$) is needed to produce cement: $70\%_{mass}$
  - CO_2 released at ~800 °C
- Clinkering temperature: 1450°C
  - Achieved using fossil fuels
  - Switching to electricity doesn't help (68% of electricity is generated from fossil fuels)
- 0.85 tons of CO_2 emitted for every ton of cement produced
- Cement production: 8% of all anthropogenic CO_2 emissions



Using supplementary cementitious materials

- Use supplementary cementitious materials (SCMs) to partially replace cement in concrete
  - Coal fly ash; slags; waste glass
  - Geological materials (e.g., clay)
- SCMs are not as reactive as cement: Cannot replace >50%
- Feature substantial batch-to-batch variations in composition
- Affect chemical reactions (cement hydration); microstructural evolution; property development in unpredictable ways.

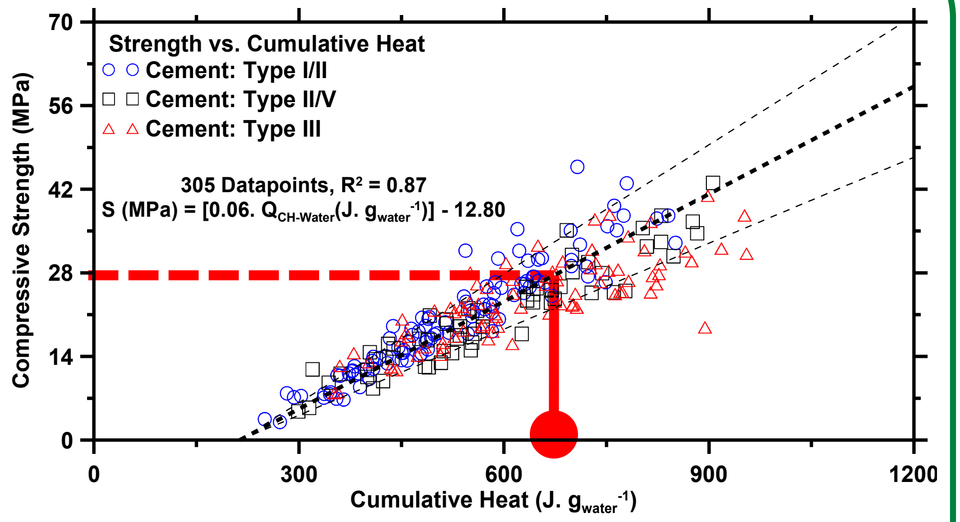
Snellings, RILEM Tech. Report 2016

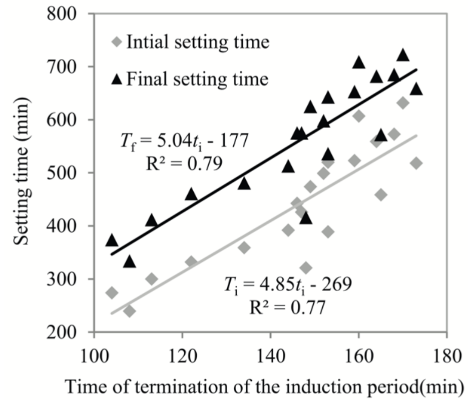


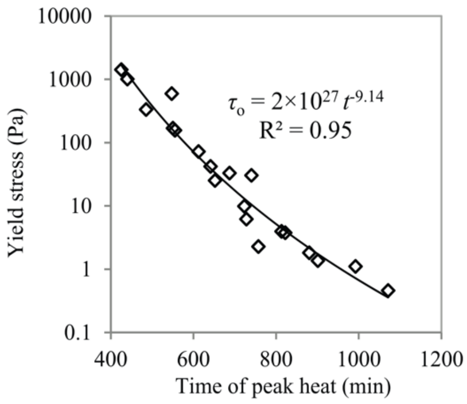
Lapeyre et al., Sci. Reports 2021
Cook et al. Mat. & Des. 2021
Han et al. Front. Mat. 2022

Why model hydration kinetics?

- *Apriori* prediction of hydration kinetics (calorimetry profile) is very useful
- If we know what % of cement has reacted, thermodynamic models can predict phase assemblage
- More heat = greater extent of cement hydration = more products = less porosity; more solid-to-solid phase connectivity; more strength
- Cumulative heat is linked to properties
  - strength; set time; rheology; etc.
- If hydration can be predicted, performance can be estimated
- Useful for cement design, mixture proportioning, etc.



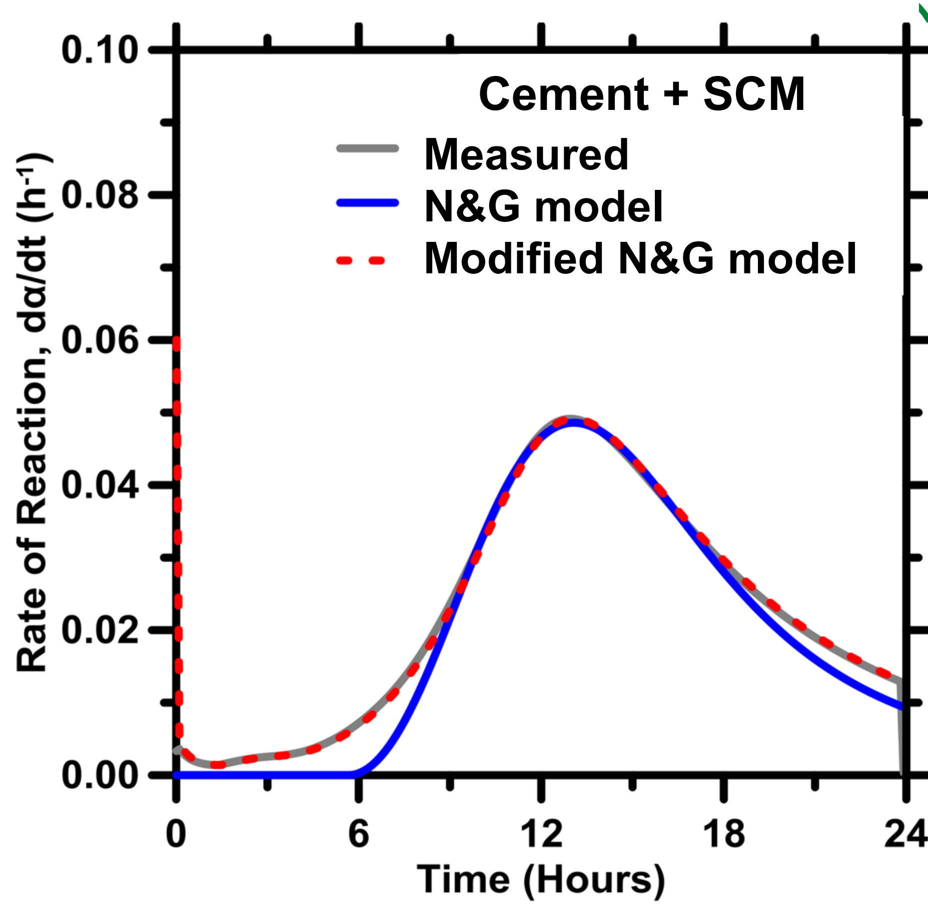




Kumar et al. Cem. Conc. Compos. 2013
Meng et al. Cem. Conc. Compos. 2019

Can we use theory-based models?

- Phase boundary nucleation & growth (N&G)
  - Equation captures the underlying mechanisms quite well
  - Allow product to nucleate and grow on SCM particle surfaces
- Excellent reproduction of experimental data
- Need to guess
  - Effective surface area of SCM
  - Nucleation density of the product
  - Rate of the growth of the product with respect to time
- *a priori* predictions are not possible

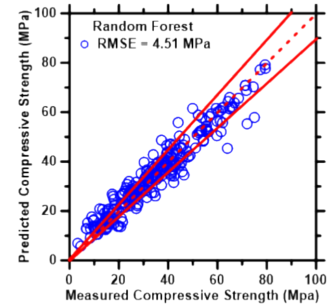


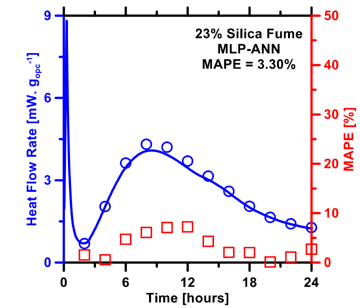
Lapeyre et al., J. Am. Cer. Soc. 2019 and 2018; Cook et al. J. Am. Cer. Soc. 2020; Meng et al. J. Phys. Chem. C. 2016; Hernandez et al. ACS Omega 2021; Kumar et al. J. Am. Cer. Soc. 2016

Supervised machine learning

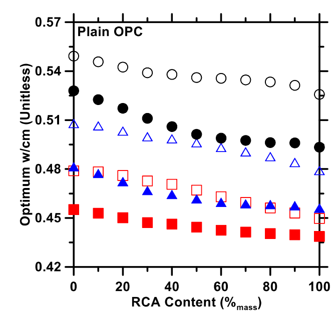
- Machine learning (ML) is a form of artificial intelligence: unsupervised and supervised
- **Supervised ML model** is first trained using a database
- The ML model develops patterns, input-output correlations in the data
  - Correlations may or may not be known from theory
- Once trained, the ML model can leverage input-output correlations to predict in new data-domains
- In the case of cement pastes/concretes
  - **Training**: ML model learns correlations between inputs (physiochemical properties of precursors) and output (heat evolution; elastic modulus; strength; etc.)
  - **Testing**: Predicts properties (heat evolution) of new pastes/concretes, using their mix design as input

Predict properties of new systems





Optimize mixture design of systems

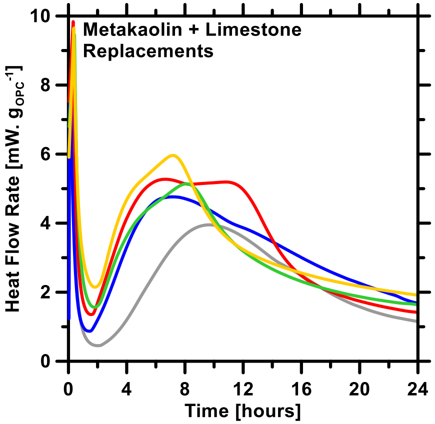


Database used for machine learning

- **Training:** hundreds of unique pastes
  - Plain, binary, and ternary pastes
- Cement: Commercial and synthetic
- SCMs
  - Different types: limestone; quartz; calcined clay; silica fume; fly ash; etc.
  - 0-to-60%$_{mass}$ cement replacement
- **Inputs:** Physiochemical properties (composition, PSD) of pastes at t = 0 h
- **Outputs:** heat flow rate with respect to time
- **Blind-testing:** Pastes, with new PSDs and replacement levels of SCMs; different cement compositions

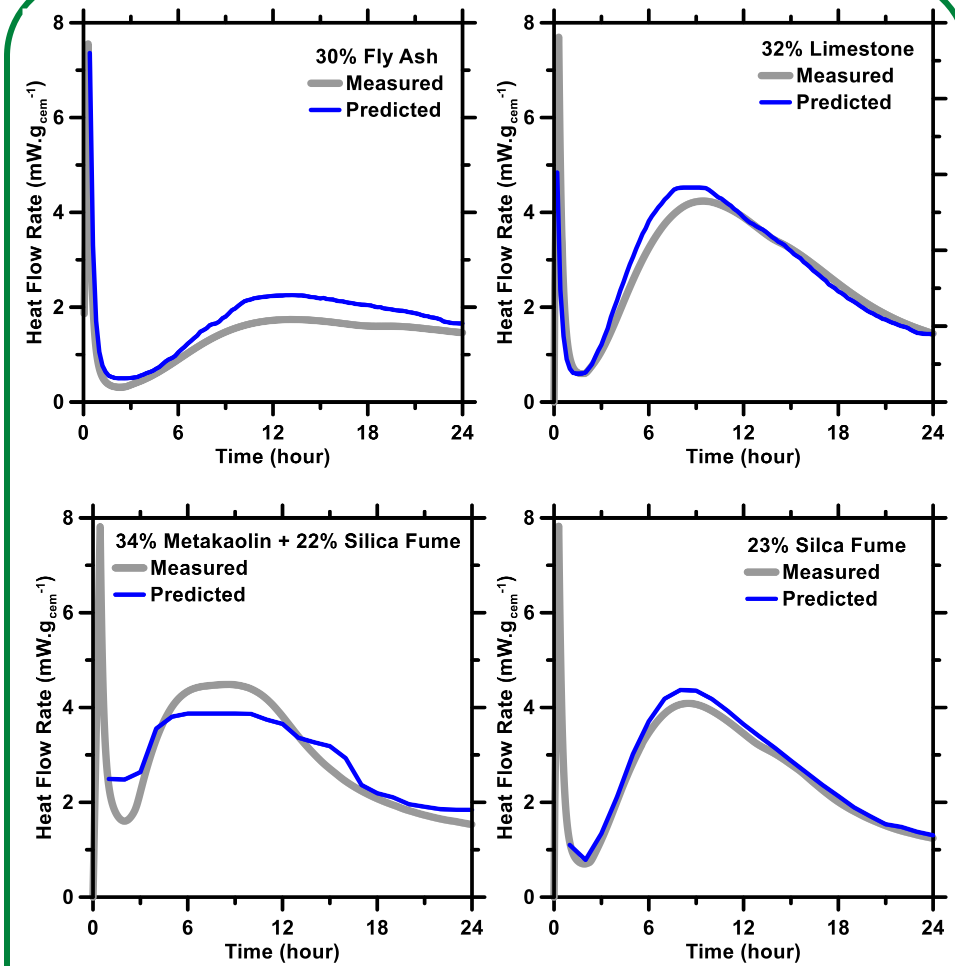
Input: PSDs, composition, and contents of cement and SCMs; & water content

Output: Time-dependent heat evolution



Machine learning: Results

- Good prediction performance in blind tests
- *a priori* predictions of hydration kinetics in complex systems, e.g., pastes with 2 SCMs
- No free parameters; just need simple inputs
  - Composition and fineness of materials; mix proportion
- Some predictions were not great
  - Reason: Limited database; inadequate training
  - Fix: Expand the database



Lapeyre et al., Sci. Reports 2020; Cook et al. Mat. & Des. 2021; Han et al. Front. Mat. 2022; Cook et al. J. Mat. Civ. Eng. 2019; Han et al. Const. Build. Mat. 2020; Han et al. Cem. Concr. Res. 2023; Bhat et al. Const. Build. Mat. 2022

Improving prediction performance

- Best approach: train the model with a larger database
- Second-best approach: simplify the database
- We use data-distillation approaches
  - Segmentation
  - Fast Fourier transformation (FFT)
  - ML guided theory-based model
- Significant improvement in prediction performance
- Faster computations

ImageNet Classification with Deep Convolutional Neural Networks

By Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com
Noam Shazeer* Google Brain noam@google.com
Niki Parmar* Google Research nikip@google.com
Jakob Uszkoreit* Google Research usz@google.com
Llion Jones* Google Research llion@google.com
Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu
Łukasz Kaiser* Google Brain lukaszkaiser@google.com
Illia Polosukhin* ‡ illia.polosukhin@gmail.com

ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION

Diederik P. Kingma* University of Amsterdam, OpenAI dpkingma@openai.com
Jimmy Lei Ba* University of Toronto jimmy@psi.utoronto.ca

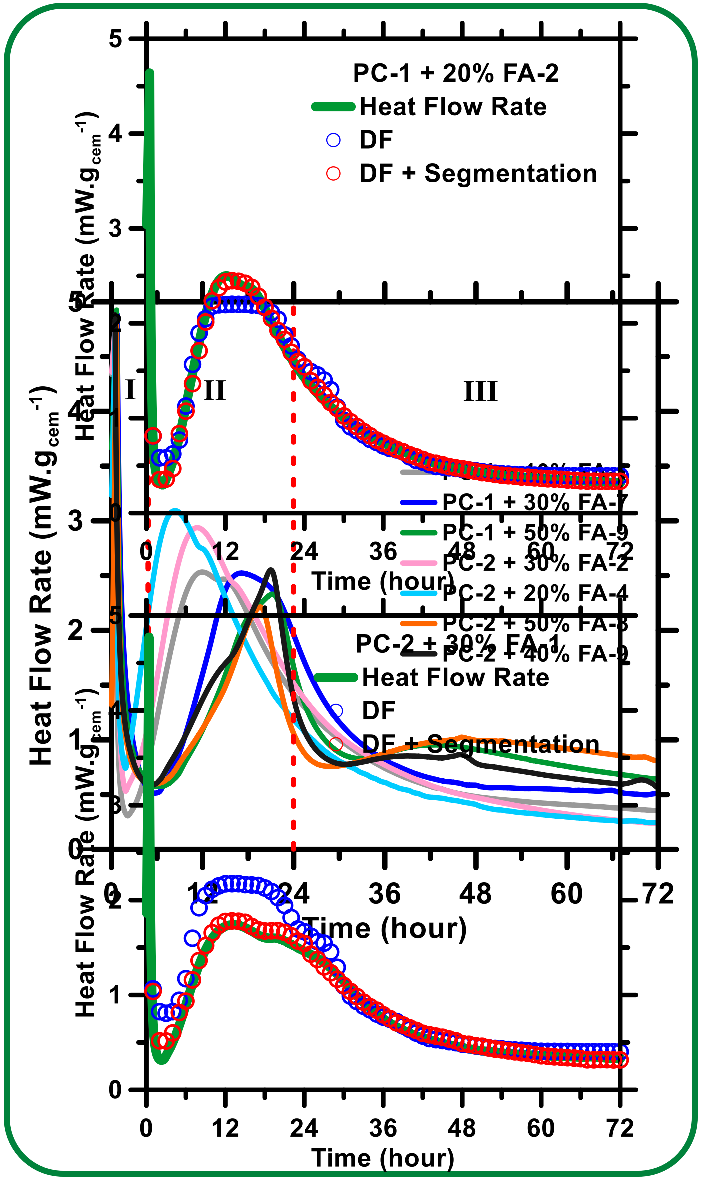
Using machine learning and feature engineering to characterize limited material datasets of high-entropy alloys

Dongbo Dai [a], Tao Xu [a], Xiao Wei [a, b], Guangtai Ding [a, b], Yan Xu [c], Jincang Zhang [b], Huiran Zhang [a, b, c]

Han et al. Front. Mat. 2022
Bhat et al. Const. Build. Mat. 2022

Segmentation

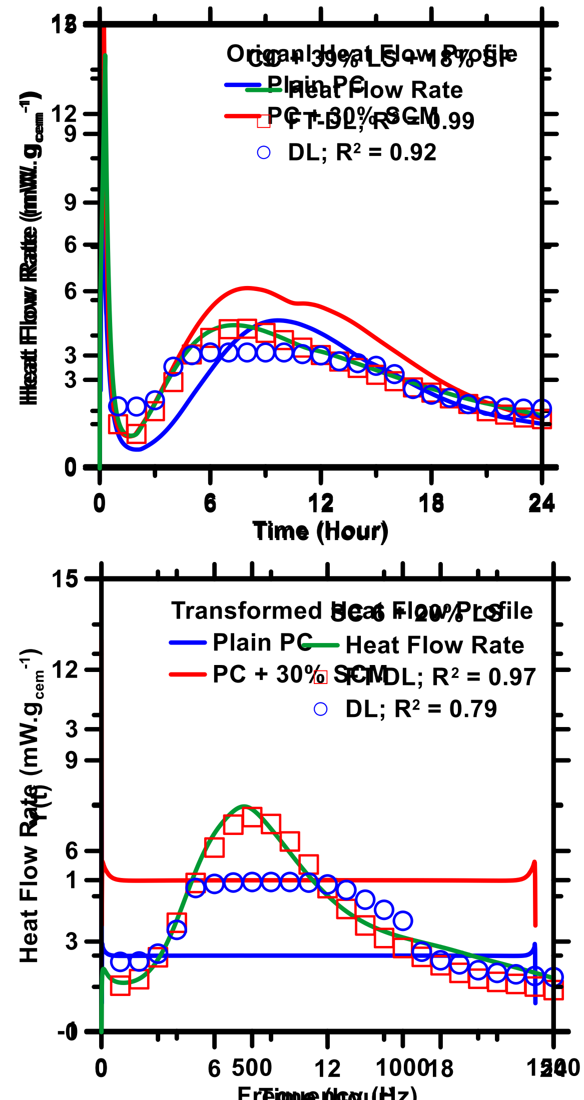
- Segmentation technique divides the heat evolution profiles into three segments.
- Hydration behaviors in the same segments are similar.
- Three segments: 0-4; 5-24; and 25+ hours.
- ML model integrated segmentation technique produce superior predictions to standalone ML models.
- Limitations: Hard to predict high sulfur cement



Han et al. Cem. Concr. Res. 2023

Fast Fourier transform

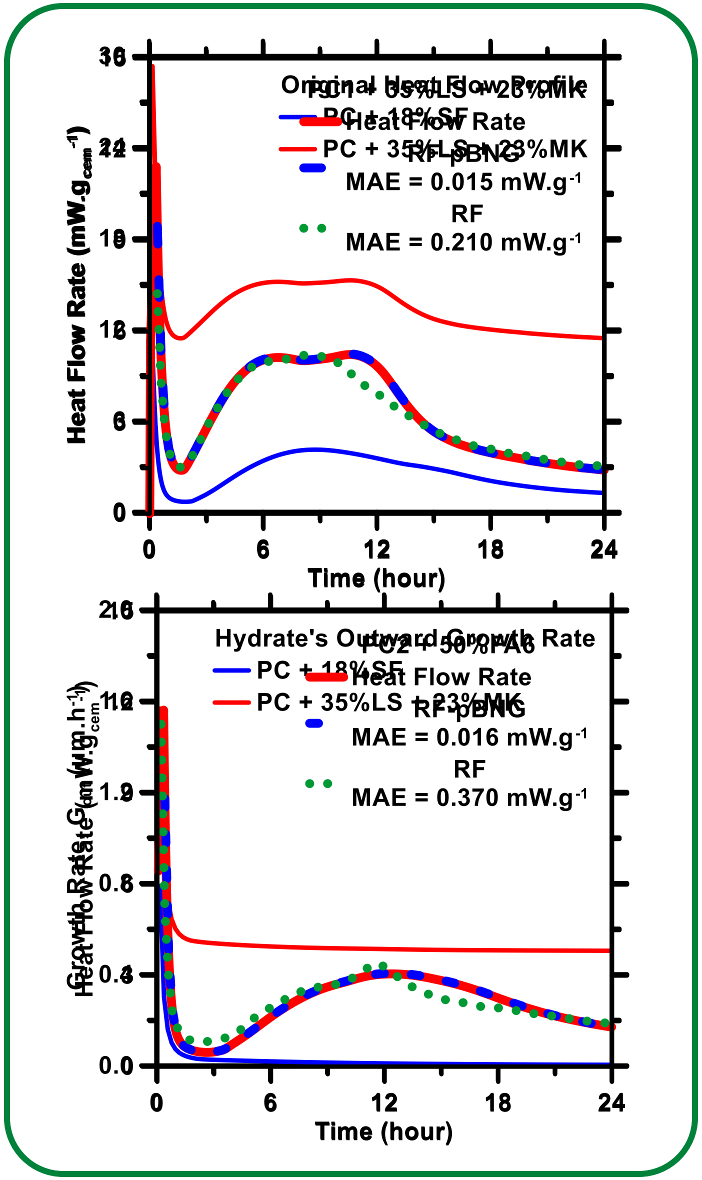
- Fourier transform converts a complex waveform to a simpler format
- Fourier transform reduced the degree of freedom of heat evolution profiles
- ML model predicted fewer data points, which save computational resources
- ML integrated with Fourier transform produced superior predictions to standalone ML
- Limitations: May violate material laws



Han et al. Frontiers 2022

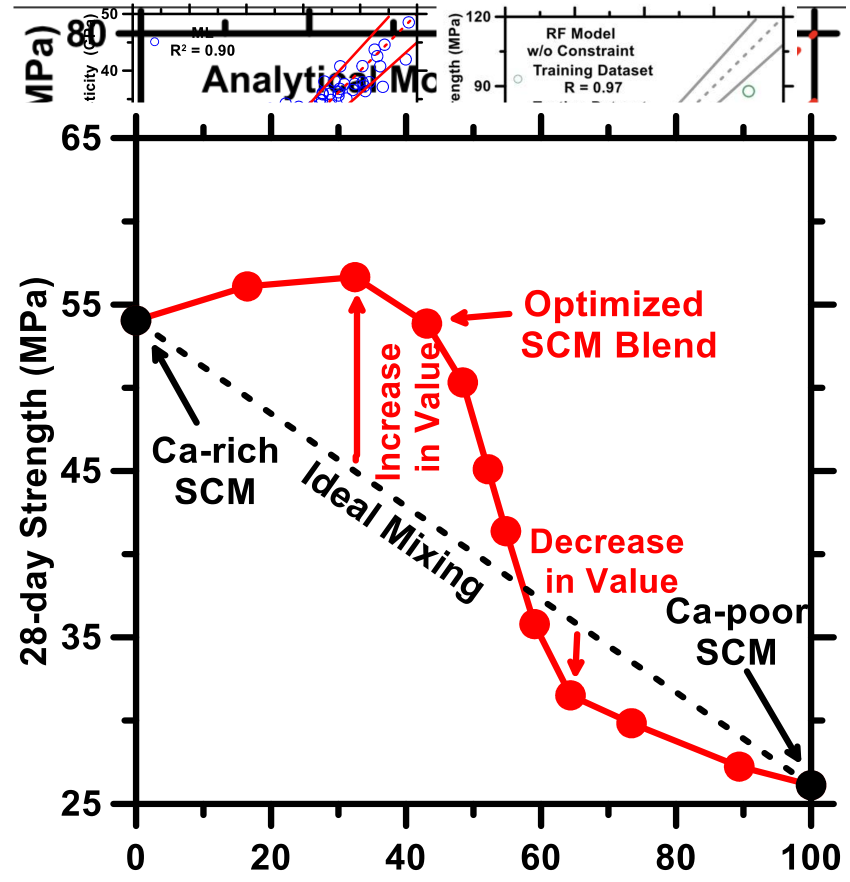
Nucleation and growth model

- Phase boundary nucleation and growth models require hydrate's growth rate (G_{out}) to reproduce heat flow profiles
- G_{out} profile showed simpler structure and trend
- ML optimized parameters for pBNG model, which allowed it to reproduce the heat flow rate profiles
- ML integrated with pBNG model produced superior predictions to standalone ML
- Nucleation and growth model regulate final outputs to avoid violation of material laws.

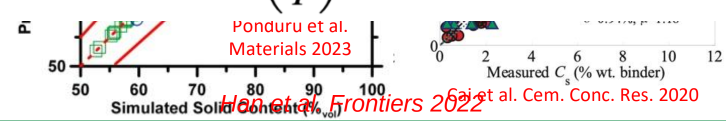


Other applications of machine learning

- Combine with ***game theory*** to develop simple, closed-form analytical models
  - (1) Quantify “importance” of each input parameter
  - (2) Construct a function using only “consequential” parameters
  - (3) Optimize the coefficients
- Predict constructability and compliance metrics of concrete
  - set time; rheology; steel rebar corrosion potential; strength
- Material design to **optimize performance**



$$CS = C_1X_{C3S}^2 + C_2X_{C2S} + C_3X_{C3A} + C_4X_{C4AF}^2 + C_5\left(X_{gypsum} + X_{FA}x_{SO3}\right) + C_6X_{FA}N_c + C_7exp\left(\frac{C_8}{T}\right) + C_9$$



Conclusions

- Machine learning is powerful tool
  - **Prediction**: To predict the behavior of materials (concrete) even—and especially—when theory is not well understood
  - **Optimization**: To design new sustainable formulations that satisfy user-defined criteria
  - **Simplification**: To develop simple, analytical models
- Database volume puts a ceiling on accuracy. Prediction performance can be improved
  - Using feature selection, data-distillation, and other techniques
  - Using theory-based models in tandem with machine learning models
- Bad at identifying functions/mechanisms. Theory based models will always be needed to reveal mechanisms.
  - Theory-based models can be used side-by-side; but that's not good enough
  - Deep integration—though difficult—is crucial

Questions?