# Amortized inference of Gaussian process hyperparameters for improved concrete strength trajectory prediction


Kristen Severson[1*]
Olivia Pfeiffer[2]
Jie Chen[1]
Kai Gong[2]
Jeremy Gregory[2]
Richard Goodwin[1]
Elsa Olivetti[2]

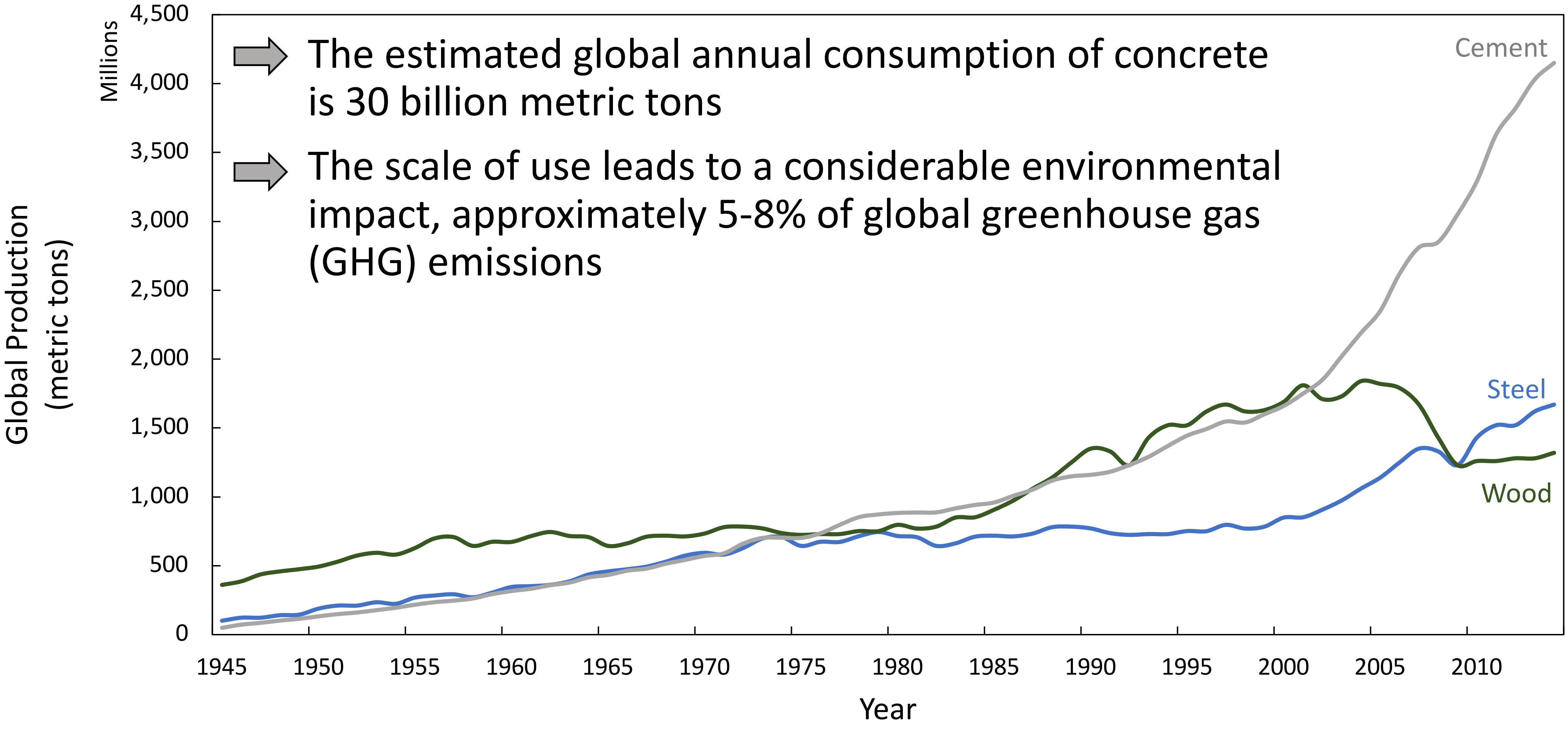
[1]MIT-IBM Watson AI Lab, IBM Research
[2]Massachusetts Institute of Technology

Olivetti Group
MIT | Dept. of Materials Science & Engineering
MIT CSHub
IBM | MIT

*kseverson@microsoft.com


# The concrete industry has a sizable GHG footprint

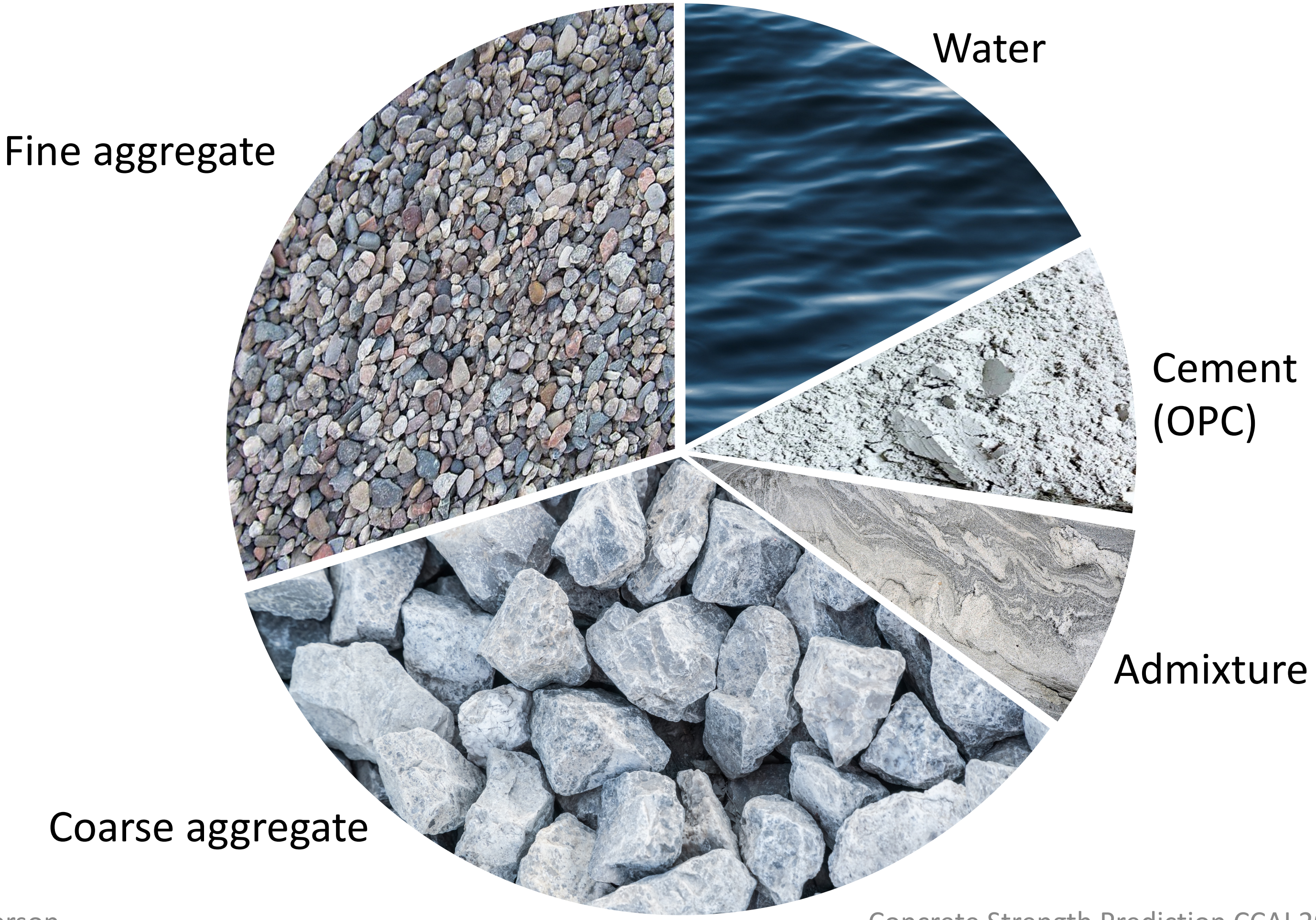
- The estimated global annual consumption of concrete is 30 billion metric tons
- The scale of use leads to a considerable environmental impact, approximately 5-8% of global greenhouse gas (GHG) emissions




Historical statistics for mineral and material commodities in the United States. US Geological Survey http://minerals.usgs.gov/minerals/pubs/historical-statistics/ (2014).

# Cement is an outsized contributor to the GHG footprint

Typical concrete mixture




Cement is approximately **10%** of the concrete mass but accounts for approximately **80%** of the GHG footprint

Partially replacing OPC with alternative materials from waste and by-products can reduce the GHG footprint

# Design of novel concrete formulations can decrease the GHG impact of the industry

Replacement materials often have higher compositional and mineralogical variability; their availability typically varies by location

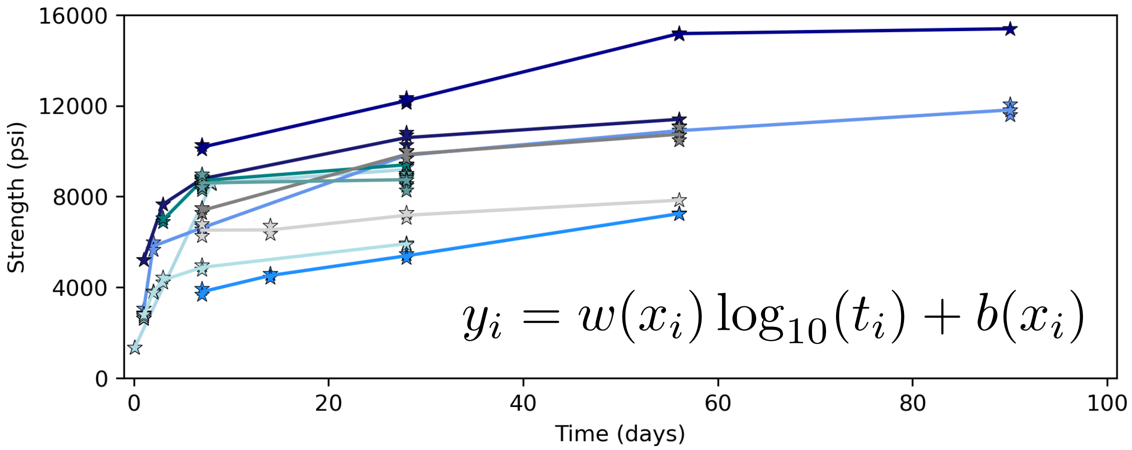
Trial and error approaches are used to design formulations with alternatives; this is very labor intensive and does not scale

Our work seeks to improve upon this paradigm by using optimal experimental design to formulate concrete mixes which minimize GHG while achieving the desired quality criteria

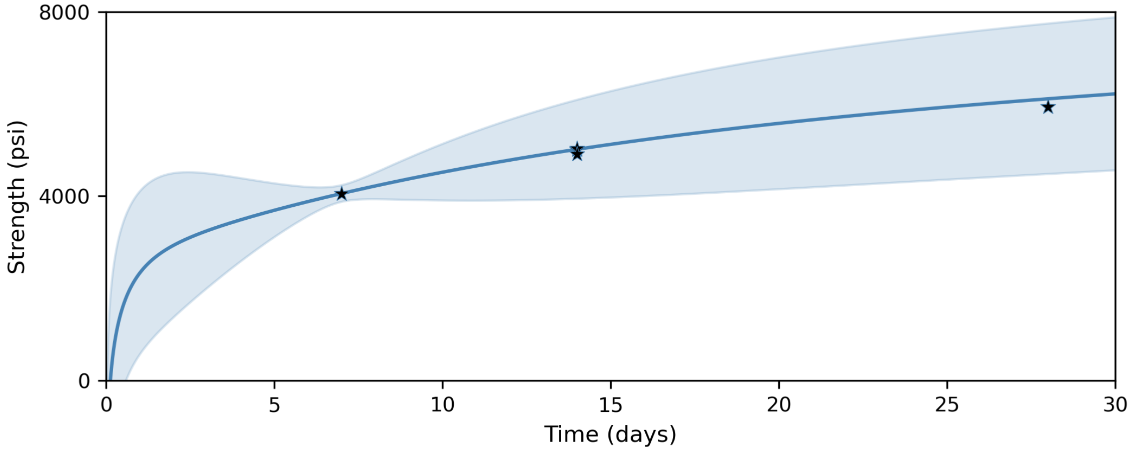
# Designing a model of concrete strength

Goal: Given a formulation $(\mathrm{x})$, estimate **strength** $(\mathrm{y})$ and predictive **uncertainty**

Concrete strength is not a single value, but a time-trajectory as the concrete sets




Uncertainty estimates are used to guide an exploration / exploitation trade-off



# Designing a model of concrete strength

Goal: Given a formulation $(\mathrm{x})$, estimate **strength** $(\mathrm{y})$ and predictive **uncertainty**

Approach: Model each formulation $(\mathrm{i})$ using a Gaussian process (GP)

$$y_i = f(t_i;\theta_i) + \epsilon_i \qquad f(t_i;\theta_i) \sim \mathcal{N}\left(\boxed{m(t_i;\theta_i)}, \boxed{k(t_i,t_i';\theta_i)}\right)$$

$$\boxed{m(t_i;\theta_i) = (\theta_i)_1 \log_{10}(t_i) + (\theta_i)_2}$$

$$\boxed{k(t_i,t_i';\theta_i) = (\theta_i)_3 \exp\left(-\frac{\|t_i - t_i'\|^2}{2(\theta_i)_4^2}\right)}$$

Share an inference network among the population to predict the GP hyperparameters $(\theta)$

$$\theta_i = \mathrm{MLP}(x_i)$$

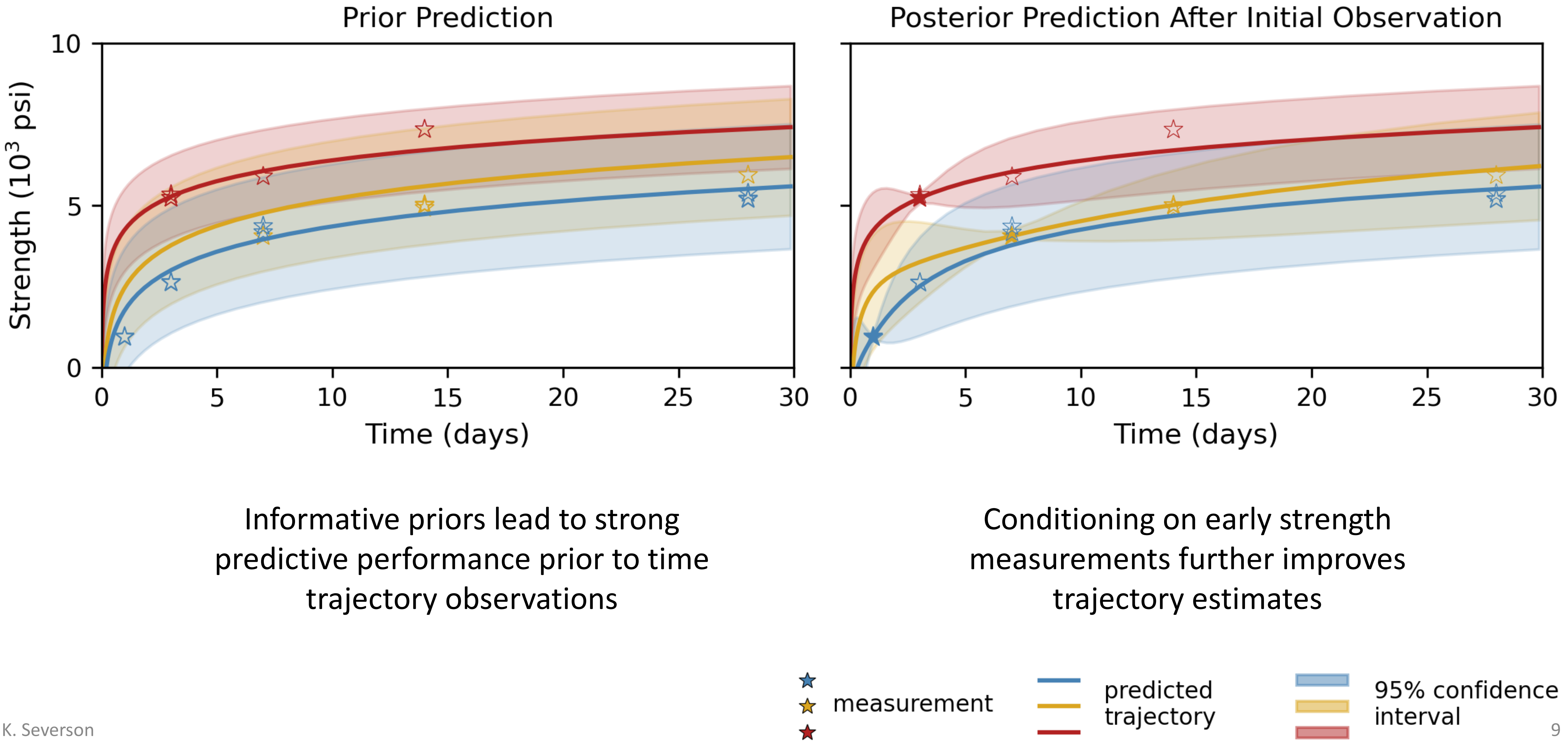
# Advantages of the approach

- The proposed model predicts strength time-trajectory and predictive uncertainty
- It models time continuously and can therefore handle irregularly sampling
- Predictions can be updated for a formulation if and when measurements become available
- Domain knowledge is incorporated into modeling decisions

# Data for evaluation

- Data was provided by an industrial concrete producer
- Strength and formulation information was available for 10,796 mixes with an average of 4.1 strength measurements per mix (range 1 – 26)
- Each formulation was characterized by 13 measures: 12 constituents and 1 derived variable, the water to cementitious material ratio

# Example of prediction results on test data




Informative priors lead to strong predictive performance prior to time trajectory observations

Conditioning on early strength measurements further improves trajectory estimates

# Comparison to random forest model

| | RMSE *before* sampling (psi) | RMV (psi) | RMSE *after* sampling (psi) |
|---|---|---|---|
| Proposed model | 950 | **871** | **831** |
| Random forest | **905** | 929 | 898 |

RMSE: root mean square error
RMV: root mean variance

# Conclusions: Moving towards a cleaner concrete industry

- An optimal experimental design approach fits the needs for designing concrete formulations with alternatives to reduce the GHG footprint
- Our approach better suits the modeling requirements to achieve this goal
- We estimate that increasing the fraction of alternatives from 5-15% to 15-25% would be associated with an emissions savings totaling over 200 billion kg $CO_2$eq per year

# Thank you!

Funding for this work was provided by the MIT-IBM Watson AI Lab

MIT-IBM
Watson
AI Lab