Sustainable Concrete via Bayesian Optimization

Sebastian Ament[1]   ament@meta.com
Andrew Witte[2]   dreww2@illinois.edu
Nishant Garg[2]   nishantg@illinois.edu
Julius Kusuma[1]   jkusuma@meta.com
[1]Meta and [2]University of Illinois at Urbana-Champaign.

Abstract

Eight percent of global carbon dioxide emissions can be attributed to the production of cement, the main component of concrete, which is also the dominant source of CO_2 emissions in the construction of data centers. The discovery of lower-carbon concrete formulae is therefore of high significance for sustainability. However, experimenting with new concrete formulae is time consuming and labor intensive, as one usually has to wait to record the concrete's 28-day compressive strength, a quantity whose measurement can by its definition not be accelerated. This provides an opportunity for experimental design methodology like Bayesian Optimization (BO) to accelerate the search for strong and sustainable concrete formulae. Herein, we 1) propose modeling steps that make concrete strength amenable to be predicted accurately by a Gaussian process model with relatively few measurements, 2) formulate the search for sustainable concrete as a multi-objective optimization problem, and 3) leverage the proposed model to carry out multi-objective BO with real-world strength measurements of the algorithmically proposed mixes. Our experimental results show improved trade-offs between the mixtures' global warming potential (GWP) and their associated compressive strengths, compared to mixes based on current industry practices. Our methods are open-sourced at github.com/facebookresearch/SustainableConcrete.

1. Introduction

Eight percent of global carbon dioxide emissions can be attributed to the production of cement (Lehne and Preston, 2018), the main reactive component of concrete, contributing significantly to anthropogenic climate change (Solomon et al., 2009). By comparison, the global annual emission from commercial aviation was estimated at 2.4% in 2019 (Graver et al., 2019). Concrete is also the leading source of CO_2 emissions in data center construction, accounting for 20-30% of the associated emissions, and making the reduction of the carbon footprint of concrete necessary to de-carbonize the operations of modern technology companies. Further, concrete mixtures that simultaneously exhibit a small carbon footprint and safe strength levels could become a critical piece in achieving societal de-carbonization goals and the mitigation of climate change.

However, conventional concrete is mainly optimized for cost, availability, and compressive strength at the 28-day mark. To meet construction and sustainability goals, concrete needs to be optimized for additional, often opposing objectives: curing speed and low environmental impact, where the latter is commonly expressed as the global warming potential (GWP), typically in kilo-gram of CO_2 per cubic meter. The optimization of these opposing objectives

is the primary goal of this work, and is part of a program to develop low-carbon concrete for data center construction which includes model development, lab testing, pilot projects, and at-scale application at Meta's data centers (Ge et al., 2022; Sudhalkar et al., 2022).

Herein, we give an overview of our methodology and validated experimental results that enable reliable strength predictions and the optimization of the trade-offs that are inherent to the design of low-carbon concrete. In particular, we 1) propose a probabilistic model that maps concrete formulae to compressive strength curves, 2) formulate the search for sustainable concrete as a multi-objective optimization problem, and 3) employ Bayesian optimization (BO) in conjunction with the proposed model to accelerate the optimization of the strength-GWP trade-offs using *real-world* compressive strength measurements.

2. Background

2.1 Gaussian Processes

Gaussian Processes (GP) constitute a general class of probabilistic models that permit exact posterior inference using linear algebraic computations alone (Rasmussen, 2004), which includes the quantification of the uncertainty of the model predictions. For this reason, most BO approaches use GPs as a model for the objective function f that is to be optimized and presumed to be too expensive to evaluate frequently. GPs can be defined as a class of distributions over functions whose finite-dimensional marginals are multi-variate Normal distributions. Formally, a real-valued f is distributed according to a GP if for any set of points $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$ in the domain $\mathbb{X}$,

$$f(\mathbf{X}) = [f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)] \sim \mathcal{N}(\mu(\mathbf{X}), k(\mathbf{X}, \mathbf{X})), \tag{1}$$

where $\mu : \mathbb{X} \to \mathbb{R}$ is called the mean function and $k : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ the covariance function, or simply "kernel", and we overloaded the notation of f, μ and k applied to the *set* of inputs $\mathbf{X}$ with a "broadcasting" over the elements of $\mathbf{X}$: $\mu(\mathbf{X}) = [\mu(\mathbf{x}_1), \ldots, \mu(\mathbf{x}_n)]$ and $k(\mathbf{X}, \mathbf{X}')_{ij} = k(\mathbf{x}_i, \mathbf{x}'_j)$ is a matrix which is positive semi-definite if $\mathbf{X} = \mathbf{X}'$.

2.2 Multi-Objective Bayesian Optimization

Multi-objective optimization (MOO) problems generally exhibit trade-offs between $m > 1$ objectives $\mathbf{f} = [f_1, \cdots, f_m]$ that make it impossible to find a single input that jointly maximizes the objectives. Instead, one is usually interested in finding a set of optimal *trade-offs*, also called the Pareto frontier (PF) between multiple competing objectives, usually under the constraint that each objective is above a minimum acceptable value $f_i(\mathbf{x}) > r_i$. Collectively, the set of lower bounds $\mathbf{r} = [r_1, \cdots, r_m]$ is referred to as the *reference point*.

The *hypervolume* of a discrete Pareto frontier $\mathcal{P} = \{\mathbf{y}_i\}_i$ bounded by a reference point $\mathbf{r}$ is a common measure of the quality, formally $\mathrm{HV}(\mathcal{P}, \mathbf{r}) := \lambda\left(\bigcup_{\mathbf{y}_i \in \mathcal{P}}[\mathbf{r}, \mathbf{y}_i]\right)$, where $[\mathbf{r}, \mathbf{y}_i]$ denotes the hyper-rectangle bounded by vertices $\mathbf{r}$ and $\mathbf{y}_i$, and λ is the Lebesgue measure. Thus, a natural acquisition function for MOO problems is the expected hypervolume improvement

$$\mathrm{EHVI}(\mathbf{x}) = \mathbb{E}\left[[\mathrm{HV}(\mathcal{P} \cup \mathbf{f}(\mathbf{x}), \mathbf{r}) - \mathrm{HV}(\mathcal{P}, \mathbf{r})]_+\right], \tag{2}$$

from obtaining a set $\mathcal{Y} \sim \mathbf{f}(\mathbf{X}) := [\mathbf{f}(\mathbf{x}_1), \cdots, \mathbf{f}(\mathbf{x}_q)]$ of q new observations. If $q = 1$ and the objectives are modeled with independent GPs, EHVI can be expressed in closed form (Yang et al., 2019), otherwise Monte Carlo approximations are necessary (Daulton et al., 2020).

Notably, DeRousseau et al. (2021) proposed a simulation-optimization framework for low-carbon concrete which jointly optimizes deterministic models of 28-day strength, cost, and embodied carbon using using an evolutionary algorithm (EA). The approach is promising but unlikely to be competitive for real-world experimentation without modifications, as EAs tend to be significantly less sample efficient than BO.

3. A Probabilistic Model of Compressive Strength

In prior work, Carino and Lew (2001) proposed analytical forms for the evolution of concrete's compressive strength as a function of time, Luo and Paal (2021) proposed a non-temporal, non-probabilistic kernel regressor to predict concrete's lateral strength, and Ge et al. (2022) used conditional variational autoencoders to predict the compressive strength at discrete t-day intervals. Herein, we propose a probabilistic model $\texttt{strength}(\mathbf{x}, t)$ that jointly models dependencies on composition $\mathbf{x}$ and time t. The proposed model is accurate even in the low data regime due to data transformations, augmentations, and a customized kernel function.

Zero-day zero-strength conditioning A simple but key characteristic of concrete strength is that it is zero at the time of pouring. As no actual measurement is made upon pouring, strength data sets generally do not contain records of the zero-day behavior. Direct applications of generic machine learning models to these data sets can consequently predict non-physical, non-sensical values close to time zero.

We propose a simple data augmentation that generalizes to any model type to fix this: For any concrete mixture in a training data set, we add an artificial observation at time zero with corresponding strength zero. We add additional artificial observations at time zero for a randomly chosen subset of compositions $\mathbf{x}$ to encourage the model to conform to the behavior for compositions that are dissimilar to any observed ones.

Non-stationarity in time The evolution of compressive strength is non-stationary in time, as the strength increases quickly and markedly early in the curing process, but converges monotonically to a terminal strength value. Therefore, we transform the time dimension to be logarithmic before passing the inputs to the model: $t \to \log t$. This transformation enables the application of stationary covariance kernels on the log-time dimension, while leading to good empirical predictive performance.

Kernel design Aside from data augmentation and transformation, a GP's kernel chiefly determines the behavior of the model and permits the incorporation of problem-specific, physical structure (Ament et al., 2021). Notably, the $\texttt{strength}(\mathbf{x}, \cdot)$ curves generally share a similar shape for any composition $\mathbf{x}$. For this reason, we introduce an additive composition-independent time-dependent component k_{time} to a generic kernel over all variables k_{joint}:

$$k((\mathbf{x}, t), (\mathbf{x}', t')) = \alpha k_{\text{time}}(\log t, \log t') + \beta k_{\text{joint}}((\mathbf{x}, \log t), (\mathbf{x}', \log t')), \tag{3}$$

where $\alpha, \beta > 0$ are variance parameters which are inferred with the other kernel hyper-parameters using marginal likelihood optimization (Rasmussen, 2004). For our experiments, we chose k_{time} to be an exponentiated quadratic kernel and k_{joint} to be a Matérn-5/2 kernel with automatic relevance determination (ARD). This structure allows the model to infer the general smooth shape of strength curves independent of the particular composition $\mathbf{x}$, leading the model's predictions to be physically sensible and accurate for long time horizons.
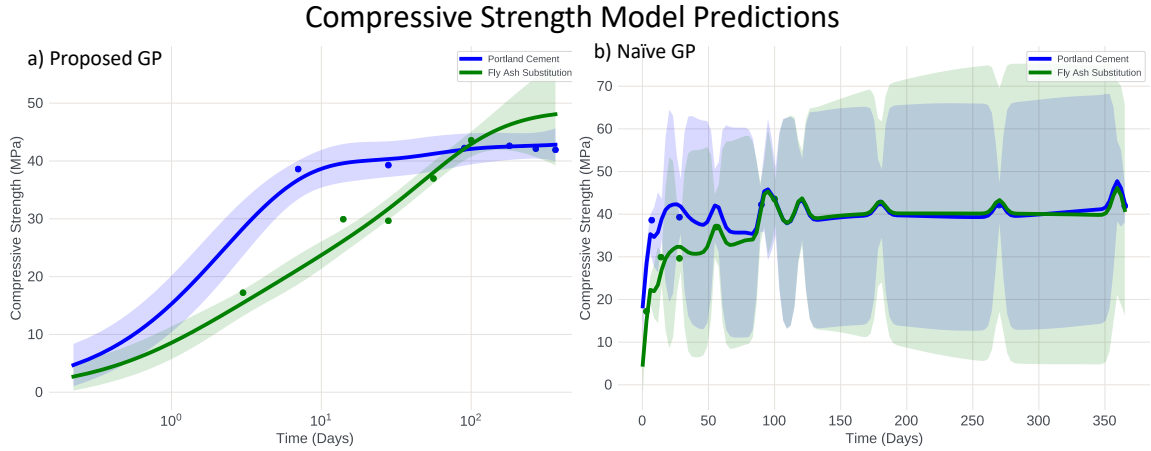
Figure 1: Strength curve predictions by our proposed model (left) and a standard GP (right) for two different compositions $\mathbf{x}$: 1) 100% cement (blue) and 2) 20% fly-ash (green).

Model Evaluation Figure 1 shows the predicted strength curves $\texttt{strength}(\mathbf{x}, \cdot)$ for two fixed compositions $\mathbf{x}$, but ranging over time t, of our model (left) and a naïve GP (right). Both models were trained on the UCI concrete strength dataset (Yeh, 2007), which we used to develop our model before starting our own experiments. While a naïve application of a GP to the data (dots) is unsuccessful, our model is accurate and well calibrated.

4. Sustainable Concrete as a Multi-Objective Optimization Problem

Increasing sustainability is the motivating factor of this work, though it is simultaneously critical to maintain concrete's compressive strength above application-specific thresholds.

Specifically, our primary objective is making concrete *sustainable*, a multi-faceted notion that includes, but is not limited to the carbon impact of production. Here, we focus on the global warming potential (GWP) to quantify concrete mixtures' sustainability, but the methods are general and can be extended to metrics that quantify complementary aspects of sustainability. Importantly, any decrease in GWP would be rendered meaningless by the inability of the associated concrete formula to be used for construction with usually tight deadlines. We therefore add compressive strength at short 1-day and long 28-day curing durations to our list of objectives.

The strength objectives are a-priori unknown and are estimated using the model proposed in Section 3 by evaluating $\texttt{strength}(\mathbf{x}, 1)$ and $\texttt{strength}(\mathbf{x}, 28)$. We model GWP as a deterministic linear function $\texttt{GWP}(\mathbf{x}) = \boldsymbol{\alpha}^\top \mathbf{x}$ where each α_i quantifies the GWP for each unit of the ith mixture ingredient. A more precise quantification of GWP would also dependent on location and transportation (Kim et al., 2022), but the linear model is a reasonable approximation for our purposes. Further, given an accurate strength model, we can infer Pareto-optimal mixes for a post-hoc change in the GWP model, similar to the post-hoc changes in the constraints explored in Sec. 5.3. Formally, the associated MOO problem is

$$\max_{\mathbf{x} \in \mathbb{X}} \left(\texttt{strength}(\mathbf{x}, 1), \ \texttt{strength}(\mathbf{x}, 28), \ -\texttt{GWP}(\mathbf{x}) \right). \tag{4}$$
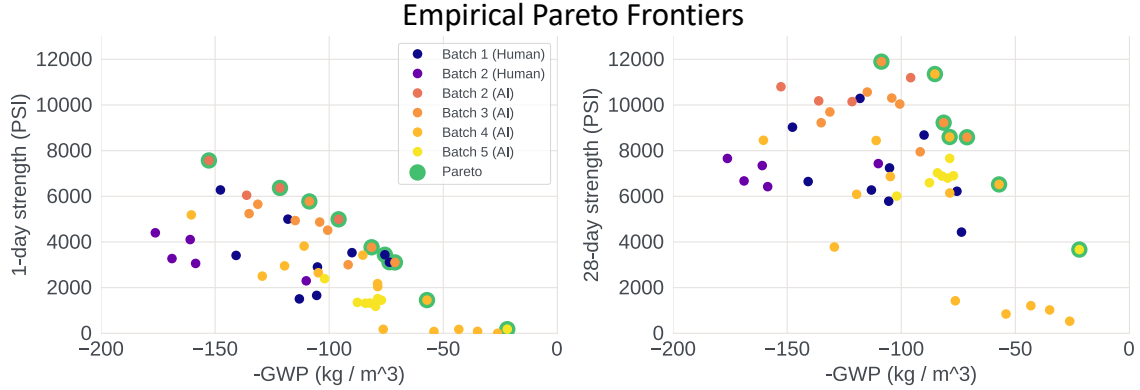
4

Figure 2: Empirical PFs for 1-day (left) and 28-day (right) compressive strength and GWP. The Pareto-optimal measurements are circled green. The darker colors correspond to human-generated mixes, while colors on the spectrum from dark orange to yellow correspond to AI-proposed mixes.

We then employ BO with Ament et al. (2023)'s qLogNEHVI acquisition function – the LogEI variant of Daulton et al. (2020)'s qNEHVI – to design batches of compositions $\mathbf{X}$, optimizing for the PF of the MOO problem of Eq. (4) using real-world compressive strength experiments.

5. Experiments

5.1 Real-World Experimental Setup

All experimental testing of mortar specimens was performed in accordance with ASTM C109, as outlined below. Two-inch mortar cubes were mixed and cured at 22 °C, first mixing fine aggregate with half the water and then adding all cementitious material with the remaining water. Superplasticizer was added during the second stage of mixing, as needed. After mixing and tamping, plastic lining was applied to each mold to prevent significant moisture loss. After 24 hours of curing inside the molds, mortar cube specimens were removed and submerged in a lime-saturated water bath at room temperature (22 °C). Three specimens each were prepared for curing ages of one, three, five, and twenty-eight (1, 3, 5, and 28) days. All specimens were subjected to a compressive load at a constant loading rate of 400 lb/s using a Forney compressive testing machine.

5.2 Empirical Optimization Results

Figure 2 shows the experimentally achieved trade-offs between 1-day (left) and 28-day (right) compressive strength, and GWP. The proposed mixtures quickly improve on both a random initial set (Batch 1, human) and a set of mixtures inspired by industry practices (Batch 2, human). The AI-proposed batches push the empirical Pareto frontier outward, providing an increasingly fine set of trade-offs between sustainability and strength. Fortunately, the highest-GWP Pareto-optimal composition shown in Figure 2 dominates the GWP-strength trade-off of a pure-cement mix (not shown in Fig. 2), implying that a GWP-reduction can –
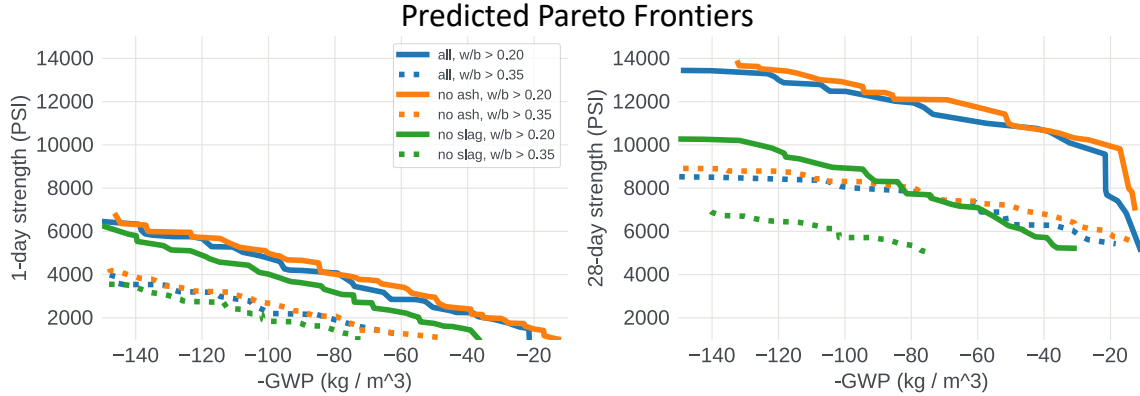
Figure 3: Predicted PFs for 1-day (left) and 28-day (right) compressive strength and GWP.

to a non-negligible extent – be achieved without sacrificing strength, though greater GWP reductions do require such trade-offs.

5.3 Inferred Pareto Frontier

In addition to using the proposed concrete strength model for BO, we can use it to compute inferred PFs conditioned on application-specific constraints. Querying the model in this way is useful both to gain scientific insight and because ingredients like fly ash or slag – waste-products of coal and steel plants, respectively – might only be effectively sourced in specific regions. By computing the inferred PF based on location-specific constraints, we can generate composition recommendations that are *customized to specific construction projects*.

Figure 3 shows a numerical approximation of the inferred PF conditioned on different water-to-binder (w/b) ratios, which affect the workability of the concrete, with higher w/b usually being more workable, as well as constraining the mixtures not to include fly ash (orange) or slag (green). The inferred PFs yield several insights: 1) the water-to-binder ratio has a significant effect on compressive strength, a phenomenon that has been remarked by in the literature on concrete (Nagaraj and Banu, 1996; Rao, 2001), 2) removing ash from the composition space has a negligible effect on the achievable trade-offs up to 28 days, and 3) removing slag has a significant negative effect. We stress that these results are based on predictions, but qualitatively match expert consensus on the variables' effects.

6. Conclusion

We introduced a probabilistic model for the temporal evolution of concrete's compressive strength as a function of its composition, formalized the problem of finding strong yet sustainable concrete formulae as a multi-objective optimization problem, and leveraged BO to propose new concrete mixtures for real-world testing. We seek to accelerate the development of sustainable concrete by open-sourcing our methods at `github.com/facebookresearch/SustainableConcrete`. The work has the potential to decrease the carbon footprint of data center construction and the construction industry at large, with possibly global impact.

References

Sebastian Ament, Maximilian Amsler, Duncan R. Sutherland, Ming-Chiang Chang, Dan Guevarra, Aine B. Connolly, John M. Gregoire, Michael O. Thompson, Carla P. Gomes, and R. Bruce van Dover. Autonomous materials synthesis via hierarchical active learning of nonequilibrium phase diagrams. *Science Advances*, 7(51):eabg4930, 2021. doi: 10.1126/sciadv.abg4930. URL https://www.science.org/doi/abs/10.1126/sciadv.abg4930.

Sebastian Ament, Samuel Daulton, David Eriksson, Maximilian Balandat, and Eytan Bakshy. Unexpected improvements to expected improvement for bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Nicholas J. Carino and H. S. Lew. *The Maturity Method: From Theory to Application*, pages 1–19. 2001. doi: 10.1061/40558(2001)17. URL https://ascelibrary.org/doi/abs/10.1061/40558%282001%2917.

Samuel Daulton, Maximilian Balandat, and Eytan Bakshy. Differentiable expected hypervolume improvement for parallel multi-objective bayesian optimization. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 9851–9864. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/6fec24eac8f18ed793f5eaad3dd7977c-Paper.pdf.

M.A. DeRousseau, J.R. Kasprzyk, and W.V. Srubar. Multi-objective optimization methods for designing low-carbon concrete mixtures. *Frontiers in Materials*, 8, 2021. ISSN 2296-8016. doi: 10.3389/fmats.2021.680895. URL https://www.frontiersin.org/articles/10.3389/fmats.2021.680895.

Xiou Ge, Richard T Goodwin, Haizi Yu, Pablo Romero, Omar Abdelrahman, Amruta Sudhalkar, Julius Kusuma, Ryan Cialdella, Nishant Garg, and Lav R Varshney. Accelerated design and deployment of low-carbon concrete for data centers. In *ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS)*, pages 340–352, 2022.

Brandon Graver, Kevin Zhang, and Dan Rutherford. emissions from commercial aviation, 2018. In *International Council on Clean Transportation*. 2019.

Alyson Kim, Patrick R Cunningham, Kanotha Kamau-Devers, and Sabbie A Miller. Open-concrete: a tool for estimating the environmental impacts from concrete production. *Environmental Research: Infrastructure and Sustainability*, 2(4):041001, oct 2022. doi: 10.1088/2634-4505/ac8a6d. URL https://dx.doi.org/10.1088/2634-4505/ac8a6d.

Johanna Lehne and Felix Preston. Making concrete change. *Innovation in Low-carbon Cement and Concrete*, 2018. URL https://www.chathamhouse.org/sites/default/files/publications/2018-06-13-making-concrete-change-cement-lehne-preston-final.pdf.

Huan Luo and Stephanie German Paal. Metaheuristic least squares support vector machine-based lateral strength modelling of reinforced concrete columns subjected to earthquake loads. *Structures*, 33:748–758, 2021. ISSN 2352-0124. doi: https:

//doi.org/10.1016/j.istruc.2021.04.048. URL `https://www.sciencedirect.com/science/article/pii/S2352012421003477`.

T.S. Nagaraj and Zahida Banu. Generalization of abrams' law. *Cement and Concrete Research*, 26(6):933–942, 1996. ISSN 0008-8846. doi: https://doi.org/10.1016/0008-8846(96)00065-8. URL `https://www.sciencedirect.com/science/article/pii/0008884696000658`.

G.Appa Rao. Generalization of abrams' law for cement mortars. *Cement and Concrete Research*, 31(3):495–502, 2001. ISSN 0008-8846. doi: https://doi.org/10.1016/S0008-8846(00)00473-7. URL `https://www.sciencedirect.com/science/article/pii/S0008884600004737`.

Carl Edward Rasmussen. *Gaussian Processes in Machine Learning*, pages 63–71. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

Susan Solomon, Gian-Kasper Plattner, Reto Knutti, and Pierre Friedlingstein. Irreversible climate change due to carbon dioxide emissions. *Proceedings of the national academy of sciences*, 106(6):1704–1709, 2009.

Amruta Sudhalkar, Julius Kusuma, Ryan Cialdella, Andy Gibbon, and Katie Sanford. Meta's road to net zero: How artificial intelligence can be leveraged by project teams to scale use of low carbon concrete. In *GREENBUILD Conference*. 2022.

Kaifeng Yang, Michael Emmerich, André Deutz, and Thomas Bäck. Multi-objective bayesian global optimization using expected hypervolume improvement gradient. *Swarm and Evolutionary Computation*, 44:945 – 956, 2019.

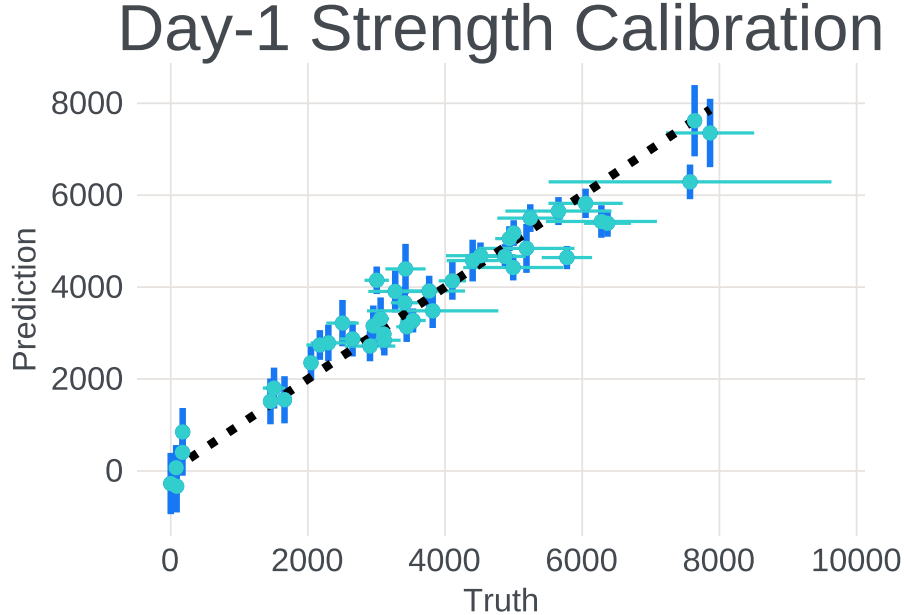I-Cheng Yeh. Concrete Compressive Strength. UCI Machine Learning Repository, 2007. DOI: https://doi.org/10.24432/C5PK67.

Figure 4: Cross-validation results for day-1 compressive strength predictions.

Appendix A. Background

Additional Information on Gaussian Processes Given a set of input and output pairs $(\mathbf{X}, \mathbf{y})$ and a Gaussian likelihood with zero mean and variance σ^2, the posterior distribution of a GP is also a GP defined by the posterior mean μ_p and kernel Σ_p:

$$
\begin{aligned}
\mu_p(\mathbf{X}^*) &= \mu(\mathbf{X}^*) + \Sigma(\mathbf{X}^*, \mathbf{X})(\Sigma(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I})^{-1}(\mathbf{y} - \mu(\mathbf{X})), \\
\Sigma_p(\mathbf{X}^*, \mathbf{X}'^*) &= \Sigma(\mathbf{X}^*, \mathbf{X}'^*) - \Sigma(\mathbf{X}^*, \mathbf{X})(\Sigma(\mathbf{X}, \mathbf{X}) + \sigma^2 \mathbf{I})^{-1}\Sigma(\mathbf{X}, \mathbf{X}'^*).
\end{aligned}
\tag{5}
$$

These equations show that, without a-priori knowledge about the specific values of the target function, the kernel function Σ is the primary factor in determining the behavior of the model. In the literature, Σ is usually also a function of hyper-parameters like length-scales which control how quickly the function is expected to vary with respect to specific input dimensions. In this work, we use well-established marginal likelihood optimization to optimize our model's hyper-parameters.

In the single-outcome ($M = 1$) setting, $\mathbf{f}(\mathbf{x}) \sim \mathcal{N}(\boldsymbol{\mu}(\mathbf{x}), \boldsymbol{\Sigma}(\mathbf{x}))$ with $\boldsymbol{\mu} : \mathbb{X}^q \to \mathbb{R}^q$ and $\boldsymbol{\Sigma} : \mathbb{X}^q \to \mathcal{S}_+^q$.

In the sequential ($q = 1$) case, this further reduces to a univariate Normal distribution: $f(\mathbf{x}) \sim \mathcal{N}(\mu(\mathbf{x}), \sigma^2(\mathbf{x}))$ with $\mu : \mathbb{X} \to \mathbb{R}$ and $\sigma : \mathbb{X} \to \mathbb{R}_+$.

Appendix B. Probabilistic Compressive Strength Model

Strength Model Cross-Validation Figure 4 shows cross validation results across all day-1 strength values in our data set, highlighting generally good predictive accuracies and well-calibrated uncertainties.

On monotonicity Another characteristic of concrete strength is its monotonic increase to a terminal value. While techniques for including this constraint on the derivatives of the GP have been proposed, they introduce additional complexities that we have – so far – not found to be worth the potential increase in model performance. The model's predictive mean is already monotone in our empirical observations, due to the other model components, like the log-time transform, the additive time-dependent component, and a range of different measurement times in the training data. Including this constraint could be lead to an increase in trust and uptake of the model by practitioners long term.