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

Kristen A. Severson[1,2], **Olivia P. Pfeiffer**[2,3], **Jie Chen**[1,2], **Kai Gong**[2,3], **Jeremy Gregory**[2,3], **Richard Goodwin**[1,2], **Elsa A. Olivetti**[2,3]

[1]IBM Research [2]MIT-IBM Watson AI Lab
kseverson@microsoft.com*, {chenjie, rgoodwin}@us.ibm.com
[3]Department of Materials Science and Engineering, Massachusetts Institute of Technology
{oliviap, kaig, jgregory, elsao}@mit.edu

Abstract

Designing and utilizing alternative concrete formulations which supplant the use of ordinary portland cement with alternative binders have been identified as central goals in reducing the greenhouse gas impact of the concrete industry. Given the variability in availability and quality of alternatives, these goals call for an optimal design of experiment approach to designing formulations, which can be adapted to local needs. The realization of this goal hinges on an ability to predict key properties. Here, we present and benchmark a Gaussian process (GP) model for predicting the trajectory of concrete strength, an essential performance measure. GPs are a desirable model class for the application because of their ability to estimate uncertainty and update predictions given additional data. In this work, rather than manually tuning hyperparameters for different concrete mix models, we propose a new method based on amortized inference leveraging mixture attributes, leading to models which are better fit for use in Bayesian optimization of concrete formulation. We demonstrate the success of the approach using a large, industrial concrete dataset.

1 Introduction

Concrete is the most used building material in the world, with an estimated global annual consumption of 30 billion metric tons [1]. The global scale of concrete use leads to a significant environmental footprint: the production of cement, a key ingredient in concrete, represents 5-8% of global greenhouse gas (GHG) emissions [2]. To meet GHG reduction targets, there is a need to reduce the environmental footprint of concrete while still achieving performance and production volume requirements.

A concrete mix is made of a binder (usually ordinary portland cement, OPC), aggregates (gravel and sand), water (which mixes with the binder to form a hardened paste), and admixtures (chemicals that can change the properties of the concrete). OPC is an attractive binder because it can be produced uniformly and in large quantities to create concrete with highly tailored performance, such as strength, stiffness, or durability. However, while it typically represents about 10% of the concrete mass, it accounts for approximately 80% of the GHG footprint [3]. Many alternative binders exist that can supplement or partially replace OPC, including those made from industrial by-products and waste materials. These alternative binders have lower carbon footprints than OPC and thus, their increased use will lower the carbon footprint of the concrete industry. However, there are several

*Work completed while at IBM Research, updated contact provided

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2021.

challenges associated with using these alternative binders, particularly the significant compositional and mineralogical variability among sources, leading to higher uncertainty in the performance of the resulting concrete products. To partially replace OPC with these alternatives, companies use trial-and-error techniques to find formulas that work for a given industrial by-product or waste available in a given location. This trial-and-error approach is very labor-intensive and does not scale well because of the above-mentioned variability in these alternative sources. This leads to significant under-utilization of alternatives.

In our work, we seek to design concrete with alternative binders from waste materials to reduce the overall GHG impact of the industry. To achieve this goal, we envision an optimal experimental design paradigm which seeks to minimize the GHG impact of a concrete formulation while achieving desired quality criteria (see Appendix Figure 2). We emphasize the importance of this framework, as opposed to a single formulation, due to variability, geographically and over time, of available feedstocks. As GHG impact can be estimated using standard factors given a concrete mix, we focus our work on learning a surrogate model of strength which can be used to accelerate the design of experiments by using Bayesian optimization. Specifically, we aim to predict the strength time-trajectory given the mix formulation. Our interest in predicting strength evolution as opposed to the more typical point estimate of strength is two-fold: (1) in a variety of applications, the evolution of the strength is necessary to evaluate if a concrete is fit-for-purpose and (2) as we begin to validate the newly proposed formulations, access to the full trajectory prediction provides earlier time points for comparison, possibly shortening experiment times using approaches such as [4]. In this paper, we present our progress on the development of the surrogate model.

2 Gaussian process models to predict concrete strength time-trajectory

Past work has demonstrated the success of machine learning models for the concrete strength prediction task, e.g., [5], however, for predictive models to be fit for purpose in the context of novel formulation design, well-calibrated uncertainties as well as estimates of strength as a function of time are required. To achieve this, we propose Gaussian process (GP) models where each formulation's strength time trajectory, $\mathbf{y}_i$, is modeled as a noisy observation of a GP $\mathbf{f}_i$ at time points $\mathbf{x}_i$, i.e., $\mathbf{y}_i = \mathbf{f}_i(\mathbf{x}_i; \boldsymbol{\theta}_i) + \epsilon_i$. The GP is specified using a linear mean function in logarithmic time (i.e., $\mathbf{x} = \log_{10}(\text{time})$) and squared exponential covariance function

$$m(\mathbf{x}_i; \boldsymbol{\theta}_i) = (\boldsymbol{\theta}_i)_1 \mathbf{x}_i + (\boldsymbol{\theta}_i)_2, \quad k(\mathbf{x}_i, \mathbf{x}_i'; \boldsymbol{\theta}_i) = (\boldsymbol{\theta}_i)_3 \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_i'\|^2}{2(\boldsymbol{\theta}_i)_4^2}\right). \tag{1}$$

The choice of a linear mean function with a logarithmic time scale is based on domain knowledge of typical concrete trajectories. Instead of learning point estimates for the GP hyperparameters $\boldsymbol{\theta}_i$, we propose to learn a function which maps mixture attributes $\mathbf{z}_i$ to hyperparameter values by using multi-layer perceptrons; i.e., $\boldsymbol{\theta}_i = \text{MLP}(\mathbf{z}_i)$. We design the MLP to encode domain knowledge. Specifically, the output $(\boldsymbol{\theta})_1$ is constrained to be positive by using a softplus activation, encoding the prior knowledge that concrete strength increases over the time interval of interest. The outputs $(\boldsymbol{\theta})_3$ and $(\boldsymbol{\theta})_4$ are similarly constrained to be positive to ensure that the covariance function is positive-definite. The observation noise ϵ is assumed to be distributed according to a zero-mean Gaussian distribution, the covariance of which is specified based on an analysis of repeated strength measures for a given mix and time point in the training dataset.

A more typical application of a GP model for this task would concatenate $\mathbf{x}_i$ and $\mathbf{z}_i$, by repeating $\mathbf{z}_i$ for each time point, and use that as input to the GP $\mathbf{f}_i$. Then the GP hyperparameters $\boldsymbol{\theta}_i$ would be optimized after some number of observations from mix i using the log marginal likelihood of the data as the objective, or the hyperparameters would be fit using the same criteria for the entire population, i.e. $\boldsymbol{\theta}_i = \boldsymbol{\theta}$. We instead treat the static and dynamic data as different types of model inputs and learn a function that maps data concerning stationary aspects of the mix, i.e. the formulation, to the parameters of the time-series model. The approach has several advantages. First, we immediately achieve our goals of modeling strength over time and generating predictive uncertainties by using a GP as the base model. Second, we gain the ability to incorporate domain understanding, namely the relationship between strength and time, into the network and GP design. Lastly, the posterior predictions of the GP can be updated as data become available, enabling more accurate predictions and possibly shortening experiment times. While this is possible in the aforementioned standard scenarios, we note that the proposed individualized models are more sensitive to new data points

Table 1: RMSE of concrete strength prediction, in units of psi, for the training and testing dataset.

	Initial Prediction				Updated Prediction	
	RF	MLP	GP	Oracle	GP	Oracle
Training	278	858	887	194	775	216
Testing	905	953	950	221	831	245

as the GP is fit only to a particular mix, and, we maintain the ability to make predictions prior to observing time series data due to informative priors on m_i. We demonstrate the main advantages of the model below and provide additional detail on the model formulation in Appendix A.

2.1 Related Work

Many studies have considered data-driven predictions of concrete properties (see [6] for a recent review). To the best of our knowledge, none have focused on predicting the trajectory of strength measurements and instead focus on point measurements. However, the need to predict the evolution of strength is intuitive: for instance, in vertical construction, long-term strength is needed to support the building, and rapid early strength development is needed to remove forms and continue construction on the floors above.

Our proposed Gaussian process model is similar to [7], who propose a "task-agnostic" GP with amortized inference. In this model, the neural network which models the GP hyperparameters is trained on synthetic data and is shown to generalize to unseen use cases. The authors propose a hierarchical self-attention-based neural network and do not address the multi-modal data scenario which is the focus of our work.

3 Application

3.1 Experimental setup

We compare the performance of the proposed GP to random forest (RF) and multi-layer perceptron (MLP) models. We additionally present the results of an "oracle" model. The oracle has the ability to observe the data prior to making a prediction and therefore either predicts exactly the observed data, or in the case of multiple measurements, predicts the average. We present this result to provide context to the variability in strength measurement, even at fixed formulation and time points.

Random Forest The random forest regressor includes 100 decision trees and uses bootstrapping with replacement to create samples that are the same size as the original data. The model considers a maximum of 3 features at each split, a value determined via hyperparameter tuning.

Multi-layer Perceptron We use a multi-layer perceptron regressor with hidden layer size (64, 64, 64), logistic activation function, and mini-batch size of 200. The model weights are optimized using the Adam solver [8]. Using a constant learning rate of 0.001, the model converges after 155 iterations based on the early stopping condition (loss improvement less than 1e-4 for 10 consecutive iterations).

3.2 Dataset

We demonstrate the model using field data from an industrial concrete producer. Information is available for 10,796 mixes (44,490 measurements) each with a varying number of strength measurements at different ages (range 1-26, average 4.1 measurements per mix). Each mix has corresponding formulation information for the 12 constituent quantities (cement, water, coarse aggregate, fine aggregate, fly ash, silica fume, slag, five types of chemical admixtures) and one derived quantity: the ratio of water to cementitious material (cement, fly ash, silica fume and slag); these 13 measures constitute $\mathbf{z}$. We randomly partition the data based on mix and use 80% for training the models and the remaining 20% for evaluation.

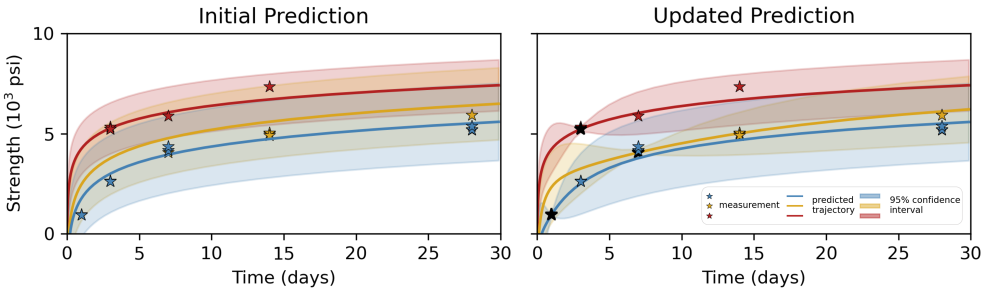

Figure 1: Strength trajectory predictions for three typical mixes from the test dataset using the GP model. *Left*: predictions using only the 13 mix descriptors. *Right*: predictions after observing one early strength measurement, denoted using a black star, for the same three mixes.

3.3 Strength Prediction

We consider two prediction settings: using only the mix formulation to predict all time points (initial prediction) and using the mix formulation along with conditioning on the first observed sample to predict all future time points (updated prediction). The results of the three models for these tasks are summarized in Table 1. Additional results are in Appendix B Table 2 and Figure 3. Although the random forest has the lowest error for the initial prediction task, the GP model provides additional advantages. For example, in Figure 1, the initial predicted trajectory (e.g. see yellow line) does not closely match the observed strength measurements. However, after seeing one early point, the updated trajectory is significantly improved. The value of this in the context of formulation optimization is described below.

3.4 Optimal experimental design

In addition to an accurate predictive model, we desire a model with well-calibrated uncertainties. As a comparison to the GP, we use the variance of the RF base estimators as an uncertainty estimate. We find that both models have 95% confidence intervals which span 94% of the testing data points. However, we find that the GP has narrower predictive intervals, with a root mean variance (akin to the average standard deviation [9]) of 871 psi, vs. 929 psi in the RF model. In effect, this implies that the GP has more accurate uncertainty, which is desirable in a Bayesian optimization where we hope to test novel formulations with particular strengths but also explore regions of high uncertainty.

The additional advantage of the GP model is its ability to adapt its predictions given additional data. The typical quality criteria for concrete formulation in industry is 28-day strength, leading to considerable data-lag during formulation optimization. We propose that the *predicted* 28-day strength estimated from the posterior estimate of the model conditioned on an earlier time point (e.g., 14-day strength measurement) can be used as surrogate data to shorten experiment times. As a proof-of-principal, we consider the 105 mixes from the testing dataset which have measurements at 14 and 28 days. We propose an early stopping criteria for the experiment based on the prediction of 14 day strength, i.e., if our prediction at baseline for 14-day strength is close to the observation at 14-days, accept the posterior prediction for 28-day strength and terminate the experiment, else continue the experiment until 28-days. Setting a threshold of 15% error, we achieve 28-day error of 575 psi (6%) for the mixes that meet the criteria and are able to terminate 79% of experiments early, reducing experiment time by half for those points. In comparison, the points that don't meet the early stopping criteria have a 28-day error of 1077 psi (20%).

4 Moving towards a cleaner concrete industry

This work highlights the potential of using amortized inference for the hyperparameters of a GP model for accurate prediction of the strength evolution for concrete containing industrial by-products, which is crucial to increasing their use in concrete production and lowering the overall global warming potential (GWP) of the concrete industry. The further enhanced predictive performance with updated early-age measurements will reduce the need for long-term measurements and hence save time and

cost for concrete producers during formulation optimization. The urgency of this problem is clear. The US alone uses 240 million cubic yards of concrete each year [10]. Using the industrial dataset, we estimate that mixes containing 5-15% alternatives have an average global warming potential of 367 kg CO_2eq/cubic yard of concrete (see Appendix C). Increasing this proportion to 15-25% would be associated with emissions savings totaling over 20 billion kg CO_2eq per year.

References

[1] P. Monteiro, S. A. Miller, , and A. Horvath, "Towards sustainable concrete," *Nature Materials*, vol. 16, p. 698, 2017.

[2] J. Oliver, G. Janssens-Maenhout, and J. P. M. Muntean, "Trends in global CO_2 emissions," The Hague: PBL Netherlands Environmental Assessment Agency; Ispra: European Comission, Joint Research Centre, Tech. Rep., 2016.

[3] S. Miller, A. Horvath, and P. Monteiro, "Readily implementable techniques can cut annual CO_2 emissions from the production of concrete by over 20%," *Environmental Research Letters*, vol. 11, p. 074029, 2016.

[4] A. Grover, T. Markov, P. Attia, N. Jin, N. Perkins, B. Cheong, M. Chen, Z. Yang, S. Harris, W. Chueh, and S. Ermon, "Best arm identification in multi-armed bandits with delayed feedback," in *AISTATS*, 2018.

[5] B. A. Young, A. Hall, L. Pilon, P. Gupta, and G. Sant, "Can the compressive strength of concrete be estimated from knowledge of the mixture proportions?: New insights from statistical analysis and machine learning methods," *Cement and Concrete Research*, vol. 115, pp. 379–388, 2019.

[6] W. B. Chaabene, M. Flah, and M. Nehdi, "Machine learning prediction of mechanical properties of concrete: Critical review," *Construction and Building Materials*, vol. 260, p. 119889, 2020.

[7] S. Liu, X. Sun, P. Ramadge, and R. Adams, "Task-agnostic amortized inference of gaussian process hyperparameters," in *NeurIPS*, 2020.

[8] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[9] K. Tran, W. Neiswanger, J. Yoon, Q. Zhang, E. Xing, and Z. W. Ulissi, "Methods for comparing uncertainty quantifications for material property predictions," *Machine Learning: Science and Technology*, vol. 1, no. 2, p. 025006, 2020.

[10] S. H. Kosmatka and M. L. Wilson, *Design and control of concrete mixtures: the guide to applications, methods, and materials*, 15th ed., ser. Engineering bulletin. Portland Cement Association, 2011, no. 1, OCLC: 798212309.

[11] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[12] J. Gardner, G. Pleiss, D. Bindel, K. Weinberger, and A. Wilson, "Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration," in *NeurIPS*, 2018.

[13] P. Dayan, G. Hinton, R. Neal, and R. Zemel, "The Helmholtz machine," *Neural Computation*, vol. 7, pp. 889–904, 1995.

[14] S. Gershman and N. Goodman, "Amortized inference in probabilistic reasoning," in *Annual Meeting of the Cognitive Science Soceity*, 2014.

[15] D. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv prepring arXiv:1312.6114*, 2013.

[16] A. O'Hagan, "Curve fitting and optimal design for prediction," *Journal of the Royal Statistical Society B*, vol. 40, pp. 1–42, 1978.

[17] R. Raganath, S. Gerrish, and D. Blei, "Black box variational inference," in *AISTATS*, 2014.

[18] D. Kingma and M. Welling, "Stochastic gradient vb and the variational auto-encoder," in *ICLR*, 2014.

[19] D. W. D.J. Rezende, S. Mohamed, "Stochastic propagation and approximate inference in deep generative models," in *ICML*, 2014.

[20] M. Titsias and M. Lázaro-Gredilla, "Doubly stochastic variational bayes for non-conjugate inference," in *ICML*, 2014.

[21] The Athena Sustainable Materials Institute, "A cradle-to-gate life cycle assessment of ready-mixed concrete manufactured by NRMCA members – version 3," 2020. [Online]. Available: https://www.nrmca.org/wp-content/uploads/2020/02/NRMCA_LCA_ReportV3_20200416.pdf

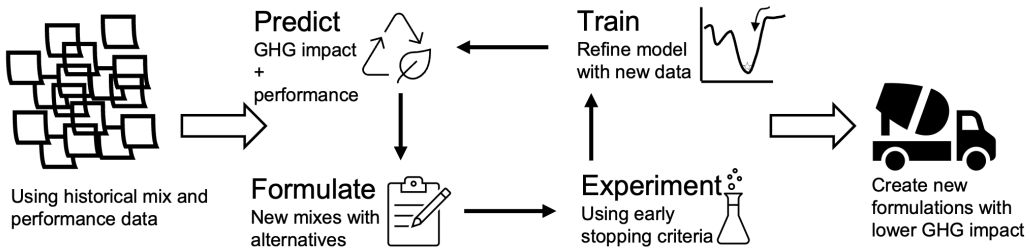

Figure 2: An overview of the optimal experimental design framework. Using green house gas (GHG) estimates and performance predictions, new concrete formulas can be proposed via optimization. These new formulas can be tested experimentally, possibly using the proposed early stopping criteria to speed up iterative rounds of experimentation. New data can be used to refine the model and the process can be repeated until concrete formulations with the desired properties and carbon footprints are achieved.

A Gaussian processes models with shared hyperparameter inference networks

We present a brief overview of Gaussian processes to introduce the central background for the proposed model. For a more detailed introduction to Gaussian processes, please refer to [11]. A GP models distributions over functions of the form $f : \mathcal{X} \to \mathbb{R}$. Any finite collection of function values, $\mathbf{f} = [f(\mathbf{x}_1), \ldots, f(\mathbf{x}_N)]$, is assumed to have a joint Gaussian distribution with mean $[\mu]_i = m(\mathbf{x}_i)$ and covariance $[\Sigma]_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, where m is a mean function and k is a positive-definite covariance function. Given a dataset $\mathcal{D} = \{\mathbf{x}_i, y_i\}_{i=1}^N$, we model the output as a noisy observation of a Gaussian process

$$y_i = f(\mathbf{x}_i) + \epsilon_i \tag{2}$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma^2 I)$. This formulation leads to closed-form estimates of the predictive distribution $f_*|\mathbf{x}_*, \mathcal{D}$ at an unobserved point $\mathbf{x}_*$. Although a GP is a non-parametric model, its success often critically depends on the specification of the mean and covariance functions and their corresponding hyperparameters. Typically hyperparameter values are determined via maximum likelihood estimation using the log marginal likelihood of $\mathcal{D}$.

In the proposed model, we consider a dataset of N samples where each sample i has two types of available data: a set of non-time-varying measures, denoted $\mathbf{z}_i \in \mathbb{R}^d$, and time-varying measures $\{x_{it}, y_{it}\}_{t=1}^{T_i}$.Given this dataset, we seek to make predictions y_{*t} given a new sample $\mathbf{z}_*$ and possibly time-series measures $\{x_{*j}, y_{*j}\}_{j=1}^{J_*}$ where typically $j < t$. Note that x_{it} and y_{it} need not be scalar, however are in all of the presented analysis. It is straightforward to extend the model to higher dimensional settings. In the context of the concrete formulation problem, $\mathbf{z}_i$ are the constituent amounts, $\mathbf{x}_i$ are the time points, and $\mathbf{y}_i$ are the corresponding strength measurements.

Each sample i is modeled as a GP

$$\mathbf{y}_i = \mathbf{f}_i(\mathbf{x}_i; \mathbf{z}_i) + \epsilon_i \tag{3}$$

where

$$\mathbf{f}_i(\mathbf{x}_i; \mathbf{z}_i) \sim \mathcal{N}(m(\mathbf{x}_i; \theta_i), k(\mathbf{x}_i, \mathbf{x}_i'; \theta_i) \tag{4}$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2 I) \tag{5}$$

where θ_i is the result a function, specifically a multi-layer perceptron (MLP), that models the corresponding parameters of the mean and covariance functions given the static attributes $\mathbf{z}_i$, i.e. $\boldsymbol{\theta}_i = \text{MLP}(\mathbf{z}_i)$. The choice of mean and covariance functions are problem dependent and are described in the main text for the concrete strength prediction problem.

We refer to the collective parameters of the θ functions as ϕ, which are learned by maximizing the log marginal likelihood of the training data

$$\mathcal{L}(\phi) = \sum_{i=1}^{N} \log p(\mathbf{y}_i|\mathbf{x}_i, \mathbf{z}_i, \sigma^2, \phi) \tag{6}$$

which is the result of marginalizing over the function values $\mathbf{f}_i$. Note that σ^2 may be learned jointly with ϕ or specified *a priori* based on domain knowledge concerning observation noise, e.g. instrument noise.

Eqn. 6 is optimized using ADAM [8] and the model is implemented in GPyTorch [12].

For a new formulation $\mathbf{z}_*$, we make predictions *prior* to observing strength measurements via our informative prior on $\mathbf{f}_*$ which makes use of the learned MLP, $\boldsymbol{\theta}_* = \text{MLP}(\mathbf{z}_*)$. As observations y_{*t} are made at time x_{*t}, the model can be updated using the closed-form of the posterior predictive distribution.

A.1 Model discussion

Our model leverages auxiliary sample information to improve time-trajectory estimates by learning a shared inference network to estimate Gaussian process model hyperparameters. This procedure can be seen as sort of amortization of the model selection process and is inspired by ideas in amortized variational inference [13, 14, 15]. The approach affords several advantages as compared to the more standard approaches of either learning one global Gaussian process model for the entire dataset or learning hyperparameters for each Gaussian process model via maximum likelihood estimates. Specifically our formulation has lower computational cost, is better able to incorporate prior knowledge, and has greater improvements to predictive performance as additional data becomes available. Each of these advantages is discussed below.

GPs are well know to have cubic complexity in the number of samples because of the required matrix inversion. In our setting, if we were to propose one global GP, the computational complexity would be $\mathcal{O}((\sum_{i=1}^{N} T_i)^3)$ however our formulation has complexity $\mathcal{O}(\sum_{i=1}^{N} T_i^3)$.

Often in applications we have ready access to prior knowledge concerning the pararameterization of $\mathbf{y}$ as a function of time ($\mathbf{x}$) and do not have such prior knowledge for the impact of other measurements. For instance, in the concrete application, we expect strength ($\mathbf{y}_i$) to increase over time, however parametric assumptions around the effect the constituent amounts ($\mathbf{z}_i$) on strength are much less clear. Modeling each mix using a unique GP enables the incorporation of this expected behavior. Without the shared inference network, *a priori* all estimates of time trajectory would be equivalent, limiting the model's applicability for optimal experiment design.

Lastly, by modeling each mix using a unique GP, the posterior estimates are more sensitive to updates from additional data, whereas when the entire population is modeled using one GP, the marginal data point has limited influence.

A.2 Future model development

Instead of point estimates for the GP parameters, we might be interested in full posterior inference and the proposed model extends naturally to this setting. To illustrate this, consider that the mean function is a linear function, possibly using a set of basis functions $h(\mathbf{x})$, with coefficients $\beta \in \mathbb{R}^m$. Let $g(\mathbf{x}) = f(\mathbf{x}) + \beta^{\mathrm{T}} h(\mathbf{x})$ where $f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$. Placing a Gaussian prior on the coefficients, $\beta \sim \mathcal{N}(b, B)$, then $g(\mathbf{x})$ is the GP

$$g(\mathbf{x}) \sim \mathcal{N}(b^{\mathrm{T}} h(\mathbf{x}), k(\mathbf{x}, \mathbf{x}') + h(\mathbf{x})^{\mathrm{T}} B h(\mathbf{x})) \tag{7}$$

[16]. Point estimates could be learned as above or Bayesian inference can be performed to learn the posterior distribution. For many choices of prior distributions, posterior inference is intractable. Therefore, we instead resort to variational inference and approximate the true posterior distribution $p(\theta|\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ with an approximate distribution $q_\phi(\theta)$. Note that it is possible to consider point estimates for some parameters and posterior distributions for others. Here, for notational simplicity, we group all parameters as θ and assume we use a fully variational approach. In this setting, instead of using neural networks to parameterize the values of the GP parameters, we use neural networks to parameterize the distributions of these parameters. The variational parameters can then be learned by optimizing the evidence lower bound

$$\begin{aligned} \mathcal{L}(\phi) = \sum_{i=1}^{N} \mathbb{E}_{q_\phi(\theta|z)} \log p(\mathbf{y}_i | g_i(\mathbf{x}_i), \theta) + \\ \log p(g_i(\mathbf{x}_i)|\mathbf{x}_i, \theta) - KL(q_\phi(\theta|\mathbf{z}_i)||p(\theta)). \end{aligned} \tag{8}$$

Table 2: MAPE of concrete strength prediction, in units of %, for the training and testing dataset.

	Initial Prediction				Updated Prediction	
	RF	MLP	GP	Oracle	GP	Oracle
Training	2.8	10	13	1.6	8.4	1.8
Testing	11	11	13	1.5	7.7	1.9

We propose Monte Carlo approximations [17, 18] combined with reparameterized gradients [19, 20] to calculate the intractable expectations. We only present results for point estimates of the GP parameters, hence why we refer to this aspect of the model as future work.

B Additional experiment details and results

The Gaussian process model is trained for 500 iterations and uses a batch size of 875. 10 random initializations are used and the final model is selected using the negative log marginal likelihood of the training data. Table 2 reports the mean absolute percent error (MAPE) for the various models, analogously to the RMSE results. Figure 3 shows the predicted and observed strength results for the testing dataset for all four models.

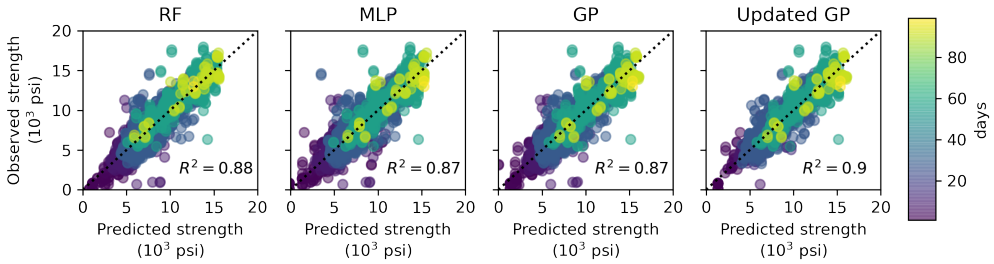

Figure 3: Predicted and observed strength for the testing dataset; the title indicates the corresponding model. The color of the point corresponds to the timestamp of the test in days. The random forest has the best predictive performance using only constituent information, but the GP has the ability to adapt to new strength data, without requirements on the timepoint of the sample.

C Global warming impact calculations

Global warming potential (GWP) is an index which expresses the total warming effect of all GHGs in terms of CO2 equivalents (CO2-eq). For example, the GWP of methane is about 30, therefore 1 kg of methane has about 30 times the warming potential of 1 kg of CO2 over a fixed time period (e.g., a 100 years). We calculate the GWP for a concrete mix on a cradle-to-gate basis, which considers the GHG emissions from three product stages: 1) raw materials supply, 2) transportation, and 3) manufacturing. We consider two primary data sources for the environmental impact analysis: the ecoinvent inventory database for raw materials and the life cycle assessment (LCA) report commissioned by the National Ready Mixed Concrete Association (NRMCA) [21] for transportation and manufacturing. The resulting function produces the global warming impact of a concrete mixture in units of kg CO2-eq per cubic yard of concrete given a concrete formulation (i.e., constituent quantities). Specific parameter values can be found in Table 3.

We qualitatively evaluate the GWP calculation by applying it to all mixes in the dataset. Figure 4 shows the GWP of concrete mixtures as a function of the percent of supplementary cementitious material (SCM; slag, fly ash and silica fume) used in the concrete, i.e., the ratio of SCM/(SCM+cement). As expected, we observe a general trend that the total concrete GWP decreases with increasing SCM content.

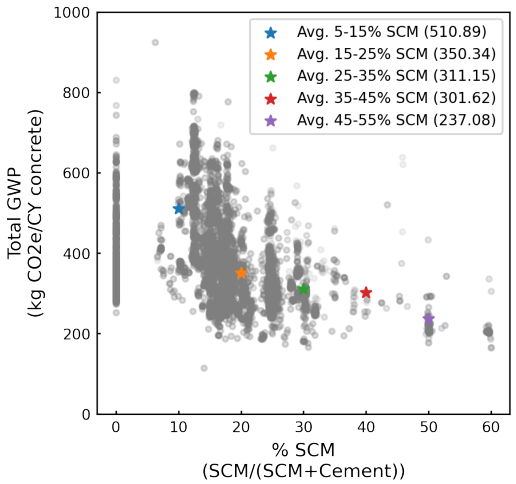

Figure 4: Percent supplementary cementitious material (% SCM) vs. total global warming potential (GWP) for all mixes in the RMC dataset. The starts represent the average GWP values for the indicated ranges of % SCM.

Table 3: Parameters used in calculating the global warming potential of concrete formulas. The impacts associated with production are calculated per cubic yard of concrete, and the relevant factor is 5.79 kg CO2-eq/CY concrete.

Constituent	**Raw Material Impact Factors (kg CO2-eq per mass [1])**	**Transportation Impact Factors (kg CO2-eq per mass [1] material shipped)**
Coarse aggregate	0.002	0.003
Cement	0.473	0.013
Slag	0.067	0.009
Other or Unknown admixture	0.063	0
Silica fume	0	0
Fly ash	0	0.009
High water reducing admixture	0.053	0
Water reducing admixture	0.053	0
Fine aggregate	0.002	0.003
Accelerating admixture	0.063	0
Air entraining admixture	0.015	0
Water	0.001	0

[1] Mass of admixtures are measured in oz, and mass of all other constituents are measured in lbs; the factors are scaled accordingly.