Bayesian Optimization for Precision Agriculture with Scalable Probabilistic Models

Anonymous Authors
Anonymous Institution

Abstract

An overarching goal of precision agriculture is to select plants with desirable traits to boost crop yield and strengthen resilience to climate change and pest infestations. To select such crops, geneticists utilize predictive models to forecast plants' desirable traits from low-cost proxies. Although such data is abundant and easy to collect, identifying meaningful candidates is expensive. Bayesian optimization (BO) is a strong cost-effective solution to identify valuable proxies. In this work, we quantify the performance of BO with a collection of acquisition function and surrogate models for identifying good proxies, in a set of +4 million proxies. We find BO achieves comparable performance to random search while requiring significantly less computation. Despite traditional BO and random search techniques performing sufficiently well, both search techniques fail to leverage information from related tasks. To this end, we propose a pre-trained model as a transfer learning method. Using this benchmark, we conduct an extensive empirical study and demonstrate promising results on the transfer learning effect, paving a new way for an efficient optimization with a parsimonious sample size.

Keywords: Bayesian Optimization, Precision Agriculture

1. Introduction

While genetic data has become increasingly abundant and easy to collect, identifying meaningful traits (or their proxies) is expensive, with naive approaches requiring combinatorial search. To this end, Bayesian optimization (BO) is considered a strong cost-effective solution for faster search. BO (Frazier, 2018; Garnett, 2022; Shahriari et al., 2016) has consistently been proven to work well in a variety of benchmarks: hyperparameter tuning of machine learning models (Snoek et al., 2012), material design (Kiyohara et al., 2016; Balachandran et al., 2016), and drug discovery (Negoescu et al., 2011; Shields et al., 2021). BO requires a surrogate model and an acquisition function. Here we consider Gaussian Processes (GPs). Beyond being proven to be a reliable method, GPs support exact inference, are interpretable, and have straightforward uncertainty quantification, often making GPs with Matérn kernels the default choice of surrogate function in BO (Rasmussen and Williams, 2006).

This paper is focused on identifying wavelength ratios with the highest co-heritability with the desired trait, using BO as the algorithmic search procedure. By leveraging a sequential and adaptive search over wavelength ratios, the number of co-heritabilities calculated can be reduced. Further, we curate the first benchmark, which consists of more than four million proxies, to understand the performance of BO methods identifying proxies for genomic prediction. We observe that different traits can have similar co-heritability properties. To this end, we propose a pretrained model to transfer the knowledge learned from one set of traits and the corresponding wavelength spectrum to another set to reduce

the search cost. We present encouraging results showing that transfer learning of good wavelength ratios is possible for different traits.

Our contribution is as follows:

- We curate a dataset of more than four million proxy measurements that can be used to understand the performance of Bayesian Optimization (BO) algorithms in the context of precision agriculture. This unprecedented dataset offers a comprehensive foundation for evaluating and enhancing the efficacy of BO techniques for the genomic selection of plants with desirable traits to improve crop yield and resilience.

- We establish the first benchmark the performance of BO methods in the context of proxy search for genomic prediction. This benchmark, derived from our comprehensive dataset, provides a standard metric for assessing the efficacy of both traditional and novel BO approaches. It facilitates the comparative analysis of methods and drives the development of more efficient and effective optimization techniques tailored to the specific challenges of precision agriculture.

- We conduct an extensive empirical study demonstrating the significant impact of transfer learning on BO's performance in find proxy data for genomic prediction. Our findings demonstrate a promising transfer effect that paves the way for efficient optimization with minimal sample complexity.

- We propose a novel application of transfer learning to BO by introducing a pretrained model approach. This innovation addresses the scalability issues inherent in traditional Gaussian Process (GP) surrogates by effectively leveraging information from related tasks. Our method enhances the BO's capability to identify valuable proxies with reduced computational expense and improved optimization efficiency.

1.1. Background on the scientific motivation

Farmers are under immense pressure as the world's population, projected to hit 10 billion by 2050 (Searchinger et al., 2019), demands more food, while urbanization and industrialization decrease available farmland (Satterthwaite et al., 2010; Follmann et al., 2021). Precision farming aims to enhance crop yield, efficiency, and profitability while minimizing waste and environmental impact. This approach relies on data-driven decision-making systems, such as GPS and remote sensing (Delgado et al., 2019; Boursianis et al., 2022), to create a more sustainable, productive, and efficient agricultural system (Liaghat et al., 2010; Sishodia et al., 2020).

One method of harnessing technology, which can boost crop yield and strengthen resilience, is incorporating genomic prediction models for crop breeding. These models enable biologists to predict the genomic estimated breeding values (GEBVs) of targeted traits using genome-wide markers of plants. Genomics prediction enables biologists to identify plants with desirable traits early in the growing season, eliminating the need to wait for trait observation. Collecting data for these predictive models is costly, either requiring teams of scientists to go out into fields to collect data or to bring individual plants into labs (Gastal et al., 2015; Berger et al., 2020; Jurado et al., 2022). To reduce costs, high-throughput phenotyping data are used in place of directly measuring the trait of interest. For example,

scientists may favor plants with better water absorption. It is known that the degree of water absorption is correlated with the wavelength spectrum of crops (Roberts et al., 2018). Hence, an efficient way to collect data is to fly a drone mounted with a hyperspectral imaging sensor to measure the wavelengths emitted from each crop (Maimaitijiang et al., 2020; Jin et al., 2020).

Measuring the usefulness of wavelength spectra for training genomic prediction models hinges on two factors (Janssens, 1979; Fernandes et al., 2023): (i) the correlation between the wavelengths and the desired trait, $r(w, t)$, and (ii) the amount of variance in the wavelength(s) w and the trait t of the crop cohorts attributed to genetics, $h(w)$ and $h(t)$, respectively (Falconer and Mackay, 1983). These factors can be amalgamated into a metric called "co-heritability." Discovering the relationship between wavelength spectra to co-heritability is non-trivial because it is unique to the cohort of crops (e.g. genetics and species dependent). Moreover, a brute force search becomes increasingly impractical as the number of proxy traits increase (Fernandes et al., 2023; Azam et al., 2023). To address the expense associated with sorting through various wavelengths, we take the Bayesian optimization (BO) approach, which can effectively reduce the number of wavelengths (s) considered.

2. Preliminaries and Related Works

Co-heritability For a given trait $t \in \mathcal{T}$, heritability ($\mathbf{h}_t^2 : \mathcal{T} \to [0, 1]$) is a crucial metric that delineates the extent to which genetic factors account for the variability observed, as opposed to environmental influences (Falconer and Mackay, 1983). Mathematically, $\mathbf{h}_t^2 = \frac{\sigma_g}{\sigma_g + \sigma_e + \sigma_r}$, where $\sigma_g^2, \sigma_e^2, \sigma_r^2$ represents the genetic variance, environmental variance, and residual variance inherent in the phenotype t, respectively. These variances are typically modeled using a linear mixed-effect model (LMM) (Speed et al., 2012; Mrode, 2014). Genetic and environmental variances are accommodated within the covariance matrix $\mathbf{\Lambda}$, while residual variance corresponds to the model's residual error. Co-heritability ($\mathbf{f} : \mathcal{W} \times \mathcal{T} \to \mathbb{R}$) is a measure that combines the heritability of two traits and their Pearson correlation r between the two traits (Janssens, 1979; Fernandes et al., 2023): $\mathbf{f}(w, t) = \sqrt{\mathbf{h}_w^2} \times \sqrt{\mathbf{h}_t^2} \times r(w, t)$. By calculating the co-heritability of a desired trait with a proxy trait, we can quantify the goodness of utilizing the proxy for a task of genomic prediction.

Linear Mixed-effect Models By nature, genetic variability exists among individuals within a population, whereas environmental factors influence the entire population as a whole. Linear mixed models (LMMs) are among the most effective statistical tools for modeling the relationship between environmental and genetic factors, and their influence on the manifestation of a phenotype (Piepho and Mohring, 2007; Holland et al., 2003; Speed et al., 2012). LMMs are a two-level hierarchical Bayesian model consisting of two normal distributions with σ noise. The hierarchical structure allows us to distinguish group-level influences and crops' genetics from global environmental effects (Bates et al., 2014; Butler et al., 2017). The observed phenotype for N individuals, $\mathbf{Y} \in \mathbb{R}^N$, has the following data generating process: $\mathbf{Y}|u \sim \mathcal{N}(\mathbf{Z}\mathbf{\Lambda} u + \mathbf{X}\beta, \sigma^2 \mathbf{I}_N)$ where $u \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_Q)$. In other words, the observed dependent variable is modeled as the linear combination of the fixed and random independent effects. Each of these variables models a different type of relationship to the phenotype.

The fixed effect is an independent variable represented by $\mathbf{X} \in \mathbb{R}^{N \times P}$ whose regression coefficient is $\beta \in \mathbb{R}^P$, where P is the number of fixed effects. The fixed effect segment, $\mathbf{X}\beta$, represents population-level effects, which are factors shared among a large portion of the population (Bates, 2011). For example, it may represent the relationship between the location of the crops on the field and the nitrogen contents of their leaves. Unlike fixed effects, random effect, $\mathbf{Z} \in \mathbb{R}^{N \times Q}$, whose regression coefficient is $\mathbf{\Lambda}u$, is an independent variable that has variations depending on the groups they belong to. Random effects can represent smaller individual or group-level variations within a population (i.e., the genetic variations over a cohort of many crops). Specifically, $\mathbf{Z} = [z_1, ..., z_k]$ where z_k is a vector for each of the k grouping factors. Each individual or group-level effect will correspond to its effect, which is represented in a block diagonal covariance matrix $\mathbf{\Lambda} = \mathrm{diag}(\mathbf{\Lambda}_1, ..., \mathbf{\Lambda}_k)$.

LMM is fitted using the restricted maximum likelihood objective. We refer the reader to a tutorial of the lmer4 R package (Bates et al., 2014), which gives an in-depth understanding of the objectives and the numerical methods used to fit such models in practice. For this paper, LMMs are used to model the relationship between genetics and the environment to phenotype for a specific crop. By leveraging the hierarchical structure of the data and incorporating both fixed and random effects, LMMs offer a robust approach to quantifying genetic and environmental effects in plant populations while accounting for potential sources of variation and relatedness among individuals.

Bayesian Optimization BO (Frazier, 2018; Garnett, 2022; Shahriari et al., 2016) is a technique for the global optimization of expensive, black-box functions. Such methods aim to solve the following optimization problem: $x^* = \arg\max_{x \in X} f(x)$. The black-box function, f, is one where you don't have access to its derivative. BO utilizes a probabilistic surrogate model to estimate f, then adaptively selects the next data point, ensuring that we extract the maximum information from each experiment. In this work, we wish to find the spectra with the optimal co-heritability. The derivative over the co-heritability function is non-trivial. Furthermore, the computation of each co-heritability value demands ~ 0.2 seconds, a potentially sluggish process contingent upon the granularity and scale of the hyperspectral data. Given these challenges, treating co-heritability as a black box within a BO framework becomes a pragmatic approach. BO has consistently been proven to work well in a variety of benchmarks: hyperparameter tuning of machine learning models (Snoek et al., 2012), material design (Kiyohara et al., 2016; Balachandran et al., 2016), and drug discovery (Negoescu et al., 2011; Shields et al., 2021). BO requires a surrogate model and an acquisition function GP Beyond being proven to be a reliable method, GPs support exact inference, are interpretable, and have straightforward uncertainty quantification, often making GPs with Matérn kernels the default choice of surrogate function in BO (Rasmussen and Williams, 2006).

Probabilistic Surrogate Models The Gaussian process (GP) is a powerful tool for probabilistic modeling commonly used in BO, offering the benefits of flexibility and a principled approach to uncertainty quantification. However, its scalability to large datasets has been a significant limitation, primarily due to the $\mathcal{O}(n^3)$ computational complexity associated with inverting the covariance matrix, where n is the number of data points. This scalability issue has been a major barrier as the computational resources required to handle large-scale data are prohibitive Rasmussen and Williams (2006).

Recent advances in machine learning technology have begun to address these challenges. Stochastic Dual Descent (SDD) and Probabilistic Transformers present promising avenues for improving the scalability of GP regression (Lin et al., 2023). SDD offers a way to approximate the inverse of large matrices more efficiently, reducing the computational burden. On the other hand, Probabilistic Transformers leverage deep learning architectures to model distributions over functions, potentially bypassing the need for explicit covariance matrix inversions and enabling GP-like inference on much larger datasets (Wang and Ribeiro, 2020). These technologies herald a new era for GP regression, where its application to large-scale data and transfer learning becomes increasingly feasible. Transfer learning is a technique where a model trained on one task is repurposed or fine-tuned for another related task. It is beneficial if data is expensive or scarce, as it leverages knowledge gained from previous task(s) to improve performance on the target task. In our application of co-heritability search, as seen in Figure 6, we observe the similarities between the co-heritabilities spaces for four target phenotypes. We wish to leverage such similarities.

3. Experimental Setup

Our dataset comprises 869 Sorghum Lines from two growouts near the University of Illinois Urbana-Champaign (Dos Santos et al., 2020; Ferguson et al., 2021). In our experiment, we consider four target traits 1) Nitrogren Area (narea) 2) Specific Leaf area (sla) 3) PLSR Nitrogren Area (pn) 4) PLSR Specific Lead area (ps). For each crop, the data consists of spectrography ranging from 350nm–2500nm. In this work, the proxy trait consists of wavelength ratios (w_1/w_2) spectrography ranging from 350nm–2500nm. This is because wavelength ratios are commonly known to be a useful proxy (Roberts et al., 2018; Lu et al., 2018), and the space is small enough to illustrate. Note that proxies can be constructed with more complex linear combinations of functions, e.g., $\frac{w_1+w_2}{w_3+w_4}$).

In the experiments presented in Section 4.1 benchmark BO with GPs. We test Matérn kernels with $v = 1/2, 3/2, 5/2$ and radial basis kernels with adaptive length scales. The length scale is assigned a Gamma(3.0, 6.0) prior to the input and a Gamma(2.0, 0.15) prior to the output scale. Three acquisition functions were tested: 1) Expected Improvement (BO-EI) 2) Upper Confidence Bound (BO-UCB) and 3) Probability Improvement (PI). BO-UCB was executed with varying exploration levels $\beta = 0.1, 0.2, 0.5$. A uniform random search served as a baseline. Each experiment comprised five trials, beginning with 10 initial points. Queries were made for 300 iterations.

In Section 4.2, we present evidence that pretraining a GP model for multiple tasks improves the performance of co-heritability search on similar tasks. To illustrate this point, we perform (vanilla) GP regression on 500–3000 data points. At every step, points are uniformly sampled from each task. The validation set comprises 1000 uniformly sampled points, distinct from the training set. The GP employs a mixed kernel. This kernel combines Matèrn-kernel(s) for the continuous variable (w_i), i.e., wavelength pairs and a categorical kernel(s) for discrete variables (t_i), i.e., the trait:

$$K((\mathbf{w}_1, \mathbf{t}_1), (\mathbf{w}_2, \mathbf{t}_2)) = K_{\mathrm{mat}_1}(\mathbf{w}_1, \mathbf{w}_2) + K_{\mathrm{cat}_1}(\mathbf{t}_1, \mathbf{t}_2) + K_{\mathrm{mat}_2}(\mathbf{w}_1, \mathbf{w}_2) \cdot K_{\mathrm{cat}_2}(\mathbf{t}_1, \mathbf{t}_2). \quad (1)$$

Experiments conducted in Sections 4.1 and 4.2, both consist of GPs' with multiple hyperparameters (i.e. noise level, kernel's length scale). To select a good set of hyper-

parameters, we train a Vanilla GP on a smaller training set (i.e., 3000 points randomly sampled from all tasks) where the hyperparameters are tuned by minimizing the marginal log-likelihood using the default L-BFGS-B minimizer in sciPy. We observe that, regardless of the optimizer's initialization, the hyperparameters are shown to consistently converge to the same values.

4. Results and Discussion
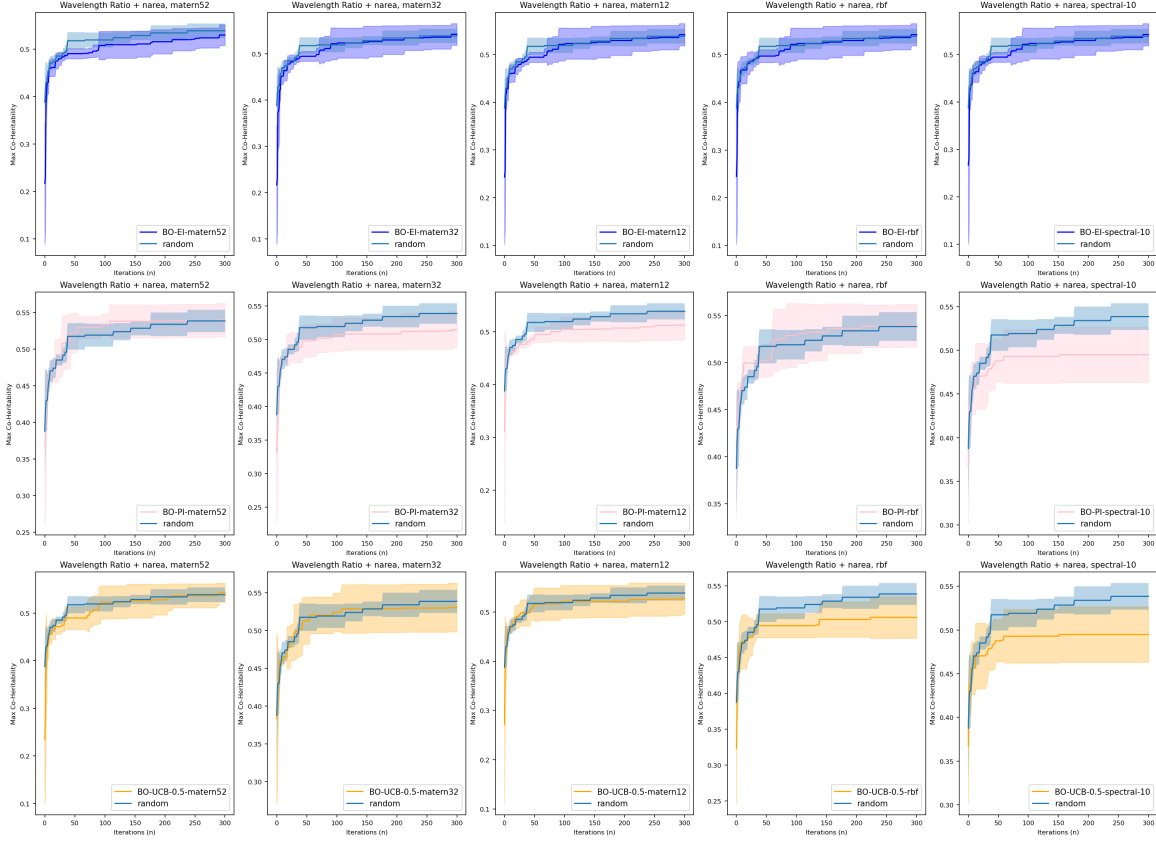
4.1. Bayesian Optimization without Transfered Learning

Figure 1: Performance (max co-heritability for narea) vs. number of iterations for BO for each co-heritability search. Each experiment was run for 300 iterations.

In this section, we present benchmarks for utilizing Bayesian Optimization with Gaussian Processes to identify high co-heritability wavelength ratios. As depicted in Figures 1, 3, 2 4, we benchmark three acquisition functions and 5 kernels over four different trait spaces. Our results indicate similar behavior over all four traits. With some exceptions, random search always performs within the standard deviation bound of BO after 300 iterations. In the other cases, like BO-UCB spectral-10 for ps, random search outperforms BO. This indicates that GP-based BO does not meaningfully outperform random search for
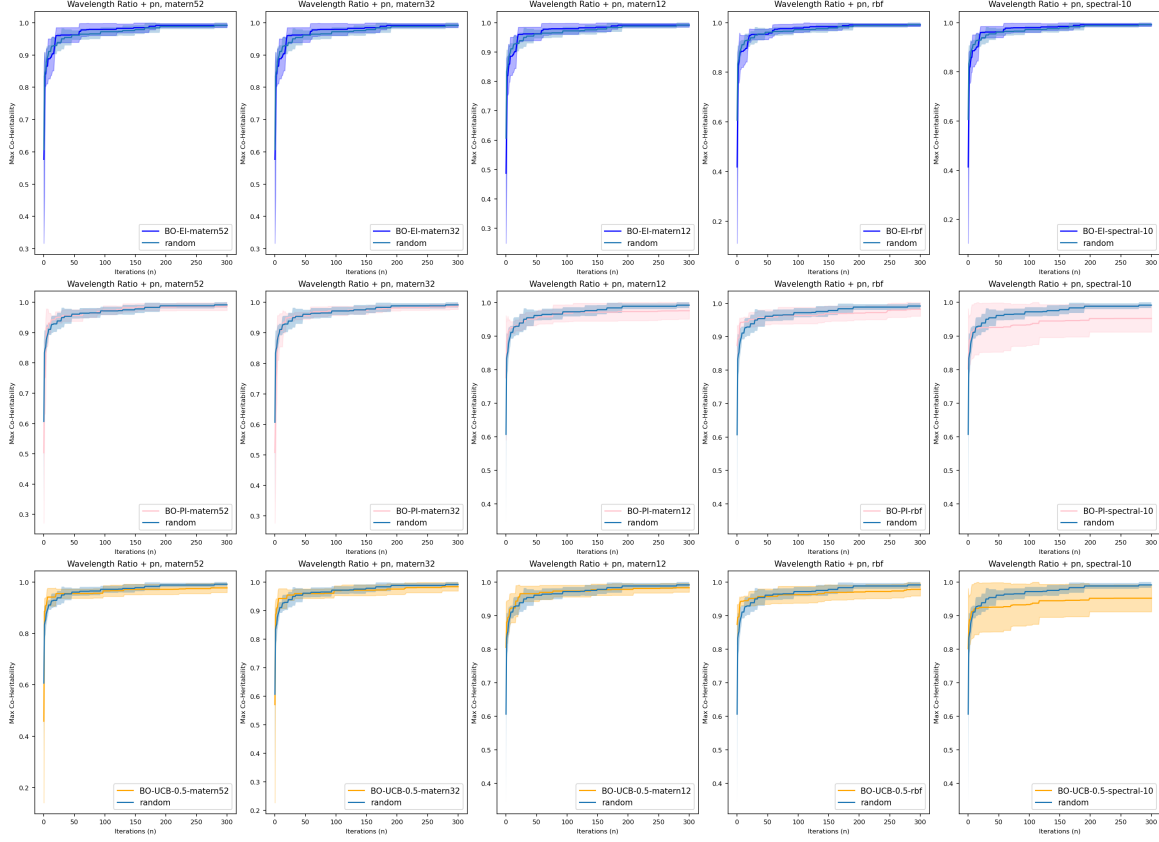
Figure 2: Performance (max co-heritability for pn) vs. number of iterations for BO for each co-heritability search. Each experiment was run for 300 iterations.

co-heritability search over wavelength ratio. Considering that the runtime of random search is $O(1)$ per iteration while fitting a Gaussian process takes $O(n^2)$, it becomes evident that the cost-performance trade-off does not warrant the use of GP-based Bayesian optimization despite its marginal performance advantage.

Our result is robust to the choice of hyperparameters of the acquisition function. As a demonstration, we look at the effects of changing the exploration-exploitation levels in the UCB objective on the performance of BO-GPs for search over co-heritabilities of wavelength ratios. To test the hyperparameter sensitivity, we execute the methods with different β values, $0.1, 0.2, 0.5$. The results, presented in Figure 5, indicate that random search performs better or only marginally worse than any BO-UCB, regardless of the hyperparameter. We conclude that BO-GP with UCB is stable to the tuning β for searching over hyperspectral co-heritability spaces. To the best of our knowledge, regardless of the acquisition function, BO with GPs for co-heritability search over wavelength ratios, surprisingly, achieves comparable performance to a simple random strategy while requiring significantly less computing cost.

Our experiments show that using Gaussian processes-based Bayesian optimization for finding high co-heritabilities may not perform much better than random search. We hy-
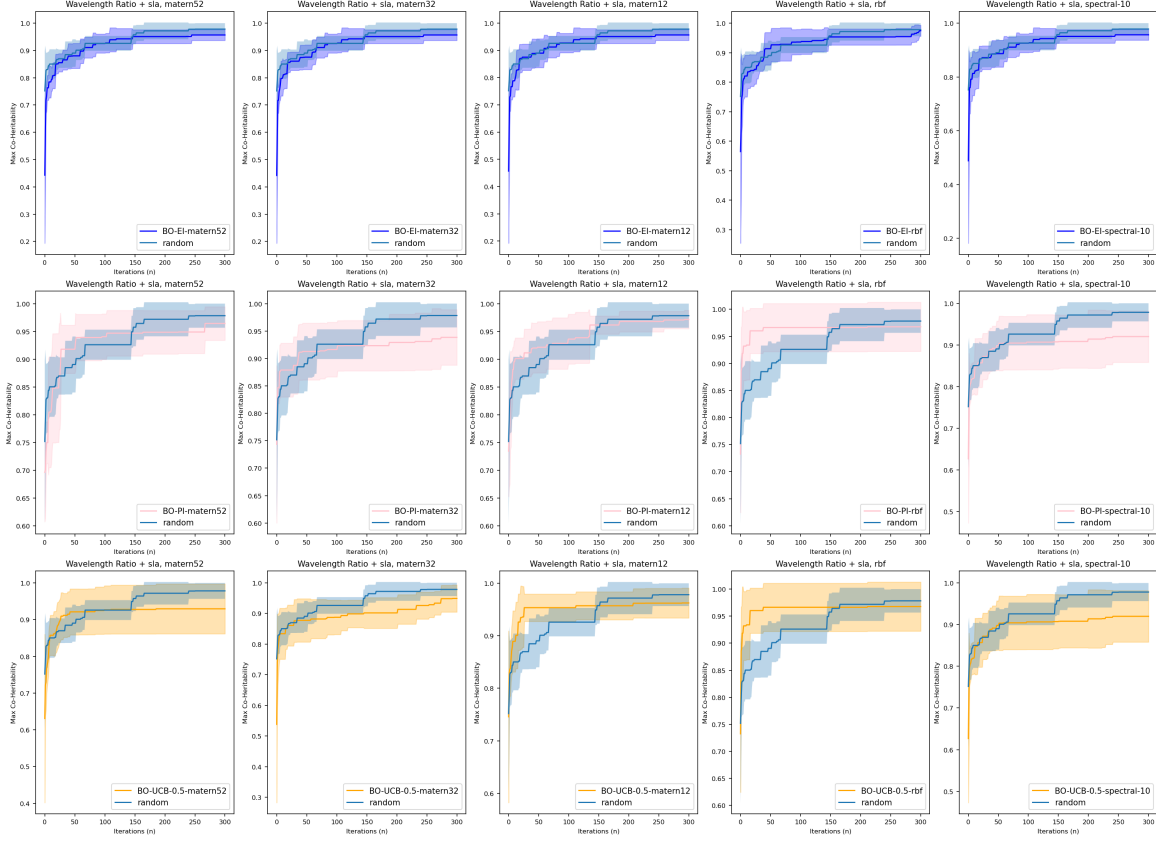
Figure 3: Performance (max co-heritability for sla) vs. number of iterations for BO for each co-heritability search. Each experiment was run for 300 iterations.

pothesize several characteristics leading to the small gap in performance between GP-based Bayesian optimization and random search. First is sharpness – there are large smooth regions with little to no signal towards the optimal co-heritability. Regarding Euclidean distance, as seen in Figure 6, the top 1% of points are extremely close to points with the lowest heritability values, making these regions extremely sharp. Since Gaussian processes with Matérn kernels are most effective for representing smooth functions, they are unlikely to represent extremely sharp regions accurately. Second is aperiodicity, the aforementioned patterns are repeated throughout the search space, yet in a non-periodic manner. Despite finding that GP-based BO performs similarly to random search, in the next section, we propose one potential direction for boosting the performance BO for co-heritability search.

4.2. Transfer Learning Improves Heritability Prediction

Transfer learning proves especially advantageous for tasks with similar characteristics, allowing the utilization of knowledge acquired from one task to enhance performance on related tasks. This approach has the potential to enhance predictive accuracy and efficiency in our context (Kendall et al., 2018; Standley et al., 2020; Bai et al., 2023). As illustrated
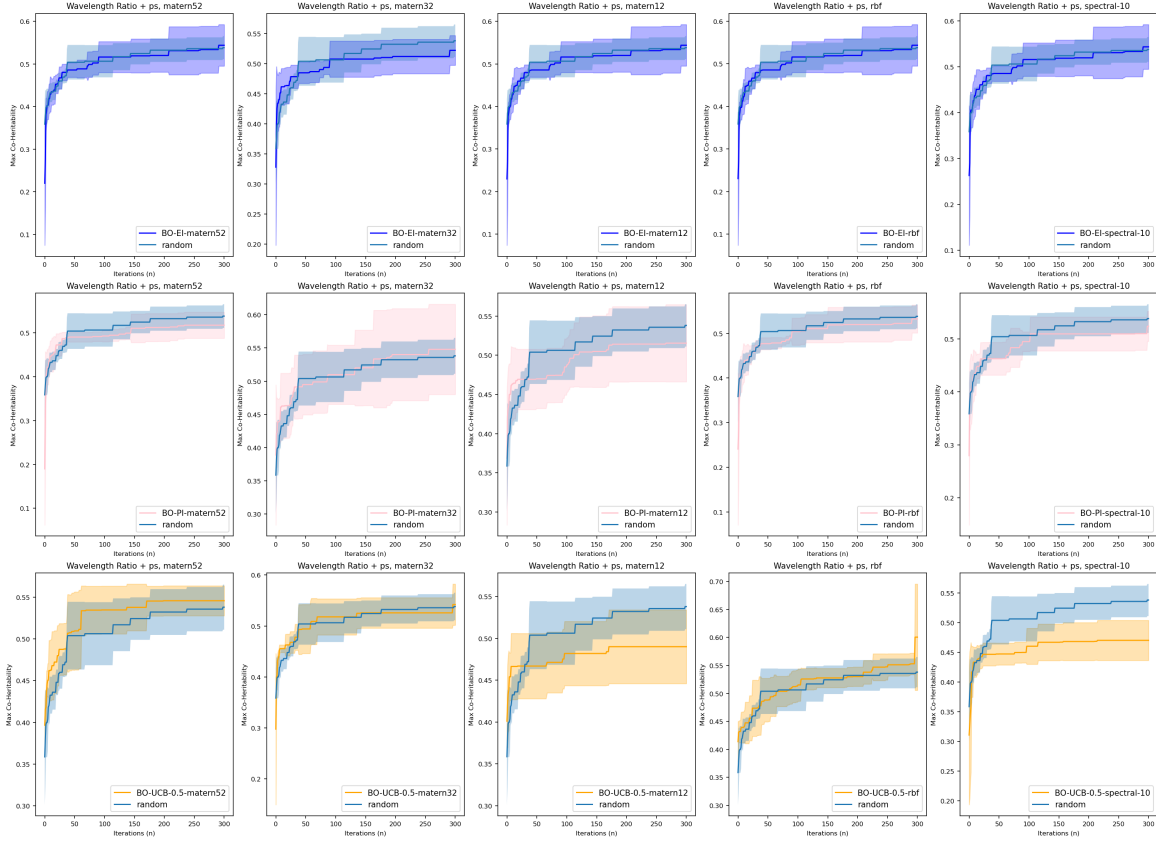
Figure 4: Performance (max co-heritability for ps) vs. number of iterations for BO for each co-heritability search. Each experiment was run for 300 iterations.

in Figure 6 the search space of co-heritabilities for various phenotypes, reveals multiple overlapping peaks and plateaus. In this section, we leverage the similarities between different co-heritability search spaces. Our objective is to develop a multi-task model capable of predicting the co-heritability of wavelength ratios across four distinct tasks simultaneously. Through this experiment, we aim to showcase the effectiveness of positive transfer in forecasting the co-heritability of wavelength ratios.

As illustrated in Figure 7, with 3000 data points, a consistent trend emerges: the validation loss steadily decreases across all tasks as the dataset size grows for each task. This reduction in mean squared error indicates that, on average, the training data from various tasks jointly improve the performance of all tasks. Furthermore, as depicted in Figure 6 (bottom), we present the GP model acquired after training on $N = 3000$ points. The resemblance of these plots to those in Figure 6 (top) assures that a robust GP model is being estimated. These findings affirm the potential of leveraging larger, diversified datasets to refine and strengthen GP models, thereby enhancing their applicability and effectiveness in complex multi-task scenarios.
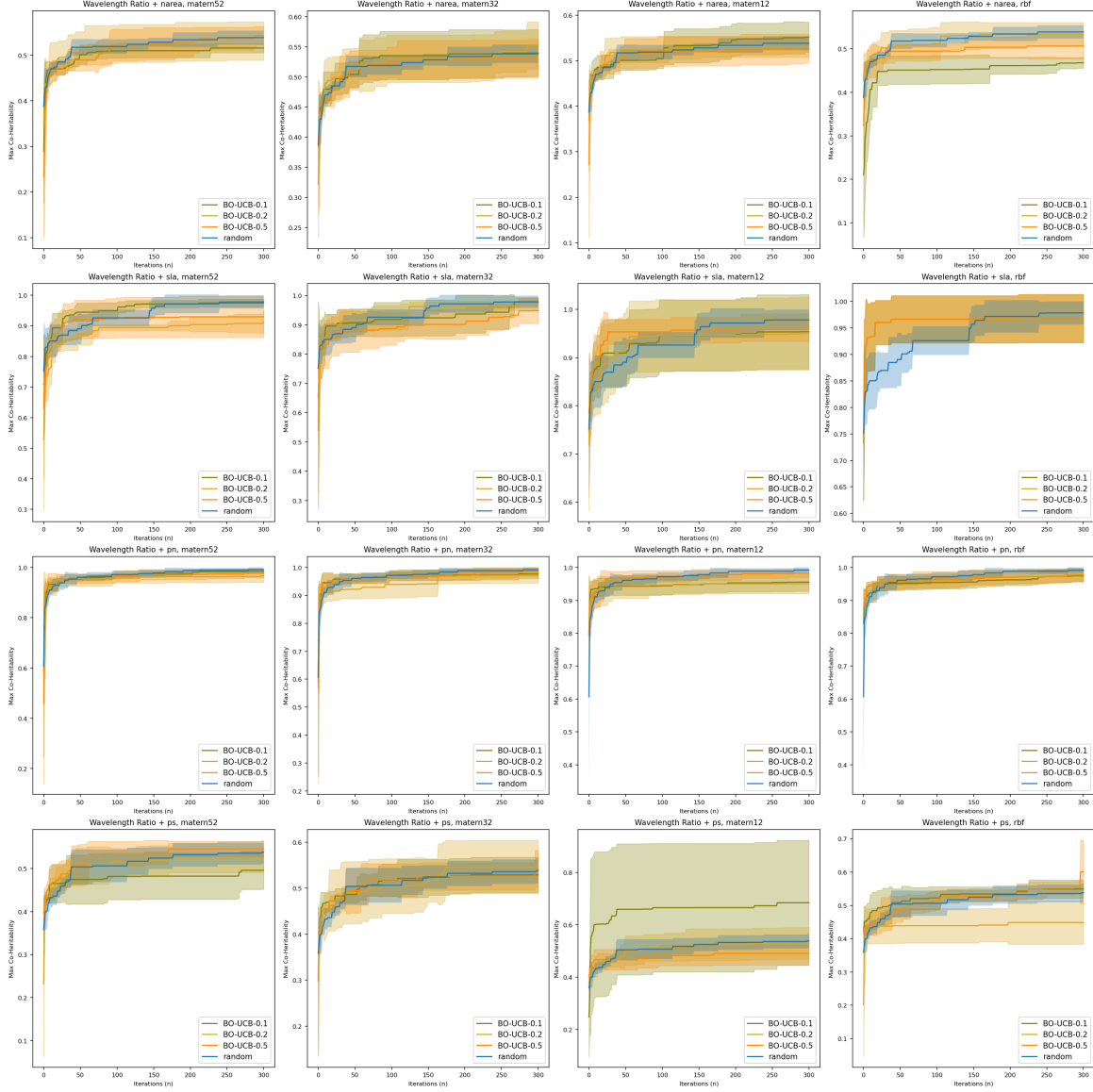
Figure 5: Performance (mean and standard deviation bands) vs. time to search over different co-heritability settings for BO-UCB. The method was run over different levels of exploration-exploitation, $\beta = 0.1, 0.2, 0.5$.

5. Conclusion & Future Work

We conduct an extensive empirical study in the effectiveness of using BO methods in the search of good proxies for genomic selection. Our results indicate that Bayesian optimization (BO) attains performance on par with random search, yet demands substantially fewer computational resources. In the second part of this study, we assess the posterior predictive distribution based on data from four distinct phenotypes. Our initial findings indicate a
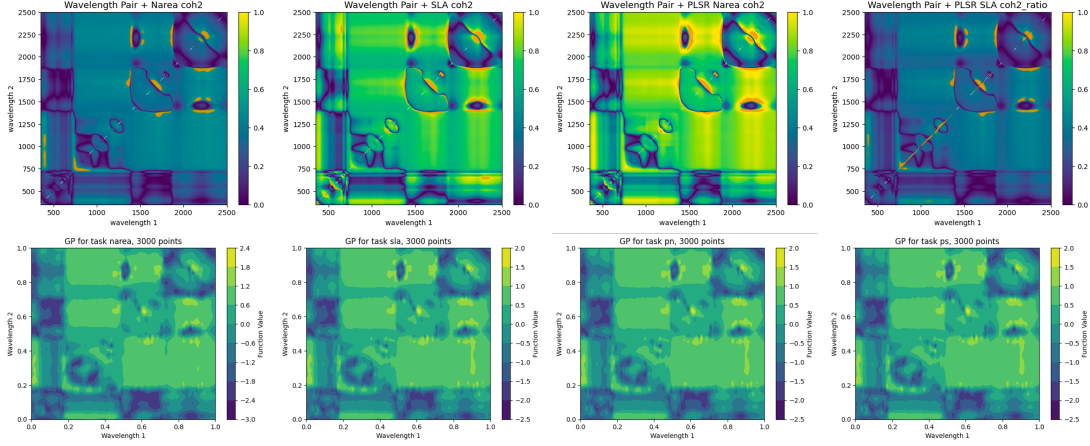
Figure 6: First row: An illustration of the search space of co-heritabilities of nitrogen area (upper left), specific leaf area (upper right), plsr nitrogren area (lower-left), and plsr specific leaf area (lower-right) for each wavelength ratio $\frac{w_1}{w_2}$. The orange highlights the top-1% of the search space in the crop Sorghum. Second row: Multitask GP posteriors after training on $N = 3000$ points.
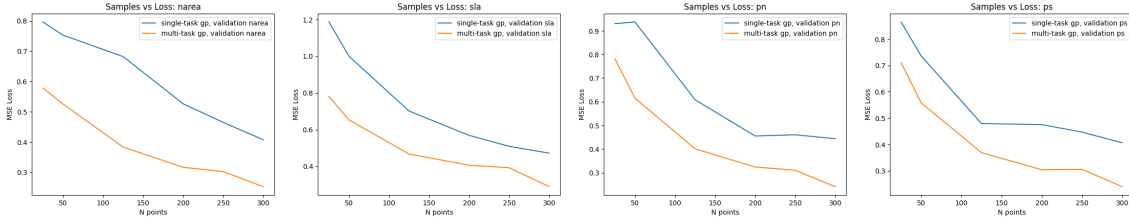

Figure 7: Vanilla GP trained on four target traits with increasing training points.

promising transfer effect in learning the co-heritabilities of various target traits for Sorghum, using wavelength ratios as indicators for genomic prediction models. This indicates that pretrained models are a promising direction for efficiently doing proxy search for a cohort of plants. However, a more thorough examination is required to fully confirm this effect and its transferability between specific phenotypes.

Our experiments involved $3,000$ data points. However, GPs suffer from high computational costs, with time complexity of $(O(n^3))$ and space complexity of $(O(n^2))$. The high computational costs of GPs can limit the number of points the pretrained model is able to capture. This becomes an issue as pretraining on 4 phenotypes can potentially be composed of up to 18.5 million wavelength ratios. As future work, to address the scalability issue, we intend to utilize a stochastic gradient method known as Stochastic Dual Descent (SDD) to streamline the fitting process. SDD has been proven to efficiently estimate kernel regression parameters through stochastic gradient descent by optimizing the dual objective. This approach necessitates domain expertise to select appropriate kernel hyperparameters. We hypothesize that an optimal hyperparameter, determined through Maximum Likelihood

Estimation, will remain effective as the volume of training data increases. If hyperparameters show to be stable, we will utilize these hyperparameters to decide on a kernel for the SDD method.

To the best of our knowledge, we are the first to study the effects of using Bayesian optimization as a method of proxy search for genomic selection. This paper proposes using pretrained models as a transfer learning approach for Bayesian optimization to identify valuable low-cost proxies for predicting desirable plant traits in precision agriculture, demonstrating promising results in leveraging information across related tasks. Our study addresses a way to reduce the costly data collection process of genomic prediction models, enabling more efficient and effective manners to bolster crop production.

References

Ruhana Azam, Sanmi Koyejo, Samuel B Fernandes, Mohammed Kebir, Andrew Leakey, and Alexander Lipka. Rethinking Bayesian optimization with Gaussian processes: Insights from hyperspectral trait search. In *NeurIPS 2023 AI for Science Workshop*, 2023.

Tianyi Bai, Yang Li, Yu Shen, Xinyi Zhang, Wentao Zhang, and Bin Cui. Transfer learning for Bayesian optimization: A survey. *arXiv preprint arXiv:2302.05927*, 2023.

Prasanna V. Balachandran, Dezhen Xue, James Theiler, John Hogden, and Turab Lookman. Adaptive Strategies for Materials Design using Uncertainties. *Scientific Reports*, 6:1–9, 2016. ISSN 20452322. doi: 10.1038/srep19660.

Douglas Bates. Computational methods for mixed models. *Vignette for lme4*, 1045:1046, 2011.

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*, 2014.

Katja Berger, Jochem Verrelst, Jean-Baptiste Féret, Zhihui Wang, Matthias Wocher, Markus Strathmann, Martin Danner, Wolfram Mauser, and Tobias Hank. Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. *Remote Sensing of Environment*, 242:111758, 2020.

Achilles D Boursianis, Maria S Papadopoulou, Panagiotis Diamantoulakis, Aglaia Liopa-Tsakalidi, Pantelis Barouchas, George Salahas, George Karagiannidis, Shaohua Wan, and Sotirios K Goudos. Internet of things (iot) and agricultural unmanned aerial vehicles (uavs) in smart farming: A comprehensive review. *Internet of Things*, 18:100187, 2022.

DG Butler, BR Cullis, AR Gilmour, BJ Gogel, and R Thompson. Asreml-r reference manual version 4. *VSN International Ltd, Hemel Hempstead, HP1 1ES, UK*, 2017.

Jorge A Delgado, Nicholas M Short Jr, Daniel P Roberts, and Bruce Vandenberg. Big data analysis for sustainable agriculture on a geospatial cloud framework. *Frontiers in Sustainable Food Systems*, 3:54, 2019.

Jhonathan PR Dos Santos, Samuel B Fernandes, Scott McCoy, Roberto Lozano, Patrick J Brown, Andrew DB Leakey, Edward S Buckler, Antonio AF Garcia, and Michael A Gore. Novel Bayesian networks for genomic prediction of developmental traits in biomass sorghum. *G3: Genes, Genomes, Genetics*, 10(2):769–781, 2020.

Douglas Scott Falconer and Trudy FC Mackay. *Quantitative genetics*. Longman London, 1983.

John N Ferguson, Samuel B Fernandes, Brandon Monier, Nathan D Miller, Dylan Allen, Anna Dmitrieva, Peter Schmuker, Roberto Lozano, Ravi Valluru, Edward S Buckler, et al. Machine learning-enabled phenotyping for gwas and twas of wue traits in 869 field-grown sorghum accessions. *Plant Physiology*, 187(3):1481–1500, 2021.

Samuel B. Fernandes, Ruhana Azam, Rachel E. Paul, Meilu Yuan, Mohammed El-Kebir, Sanmi Koyejo, Alexander E. Lipka, and Andrew Leakey. Including high-throughput phenotyping derived traits in multi-trait genomic analysis. Presented at the CSSA: Translational Genomics Workshop, Plant and Animal Genome XXI Conference., 2023.

Alexander Follmann, Maximilian Willkomm, and Peter Dannenberg. As the city grows, what do farmers do? a systematic review of urban and peri-urban agriculture under rapid urban growth across the global south. *Landscape and Urban Planning*, 215:104186, 2021.

Peter I. Frazier. A tutorial on Bayesian optimization. *INFORMS Tutorials in Operations Research*, 2018. URL https://pubsonline.informs.org/doi/10.1287/educ.2018.0188.

Roman Garnett. *Bayesian Optimization*. Cambridge University Press, 2022. URL https://bayesoptbook.com/.

François Gastal, Gilles Lemaire, Jean-Louis Durand, and Gaëtan Louarn. Quantifying crop responses to nitrogen and avenues to improve nitrogen-use efficiency. In *Crop physiology*, pages 161–206. Elsevier, 2015.

James B Holland, Wyman E Nyquist, Cuauhtemoc T Cervantes-Martínez, and J Janick. Estimating and interpreting heritability for plant breeding: an update. *Plant breeding reviews*, 22, 2003.

MJJ Janssens. Co-heritability: its relation to correlated response, linkage, and pleiotropy in cases of polygenic inheritance. *Euphytica*, 28(3):601–608, 1979.

Xiuliang Jin, Pablo J Zarco-Tejada, Urs Schmidhalter, Matthew P Reynolds, Malcolm J Hawkesford, Rajeev K Varshney, Tao Yang, Chengwei Nie, Zhenhai Li, Bo Ming, et al. High-throughput estimation of crop traits: A review of ground and aerial phenotyping platforms. *IEEE Geoscience and Remote Sensing Magazine*, 9(1):200–231, 2020.

Juan M Jurado, Alfonso López, Luís Pádua, and Joaquim J Sousa. Remote sensing image fusion on 3d scenarios: A review of applications for agriculture and forestry. *International Journal of Applied Earth Observation and Geoinformation*, 112:102856, 2022.

Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7482–7491, 2018.

Shin Kiyohara, Hiromi Oda, Koji Tsuda, and Teruyasu Mizoguchi. Acceleration of stable interface structure searching using a kriging approach. *Japanese Journal of Applied Physics*, 55(4):45502, 2016. ISSN 13474065. doi: 10.7567/JJAP.55.045502.

Shohreh Liaghat, Siva Kumar Balasundram, et al. A review: The role of remote sensing in precision agriculture. *American journal of agricultural and biological sciences*, 5(1): 50–55, 2010.

Jihao Andreas Lin, Shreyas Padhy, Javier Antorán, Austin Tripp, Alexander Terenin, Csaba Szepesvári, José Miguel Hernández-Lobato, and David Janz. Stochastic gradient descent for Gaussian processes done right, 2023.

Jinzhu Lu, Mingchuan Zhou, Yingwang Gao, and Huanyu Jiang. Using hyperspectral imaging to discriminate yellow leaf curl disease in tomato leaves. *Precision agriculture*, 19:379–394, 2018.

Maitiniyazi Maimaitijiang, Vasit Sagan, Paheding Sidike, Sean Hartling, Flavio Esposito, and Felix B Fritschi. Soybean yield prediction from uav using multimodal data fusion and deep learning. *Remote sensing of environment*, 237:111599, 2020.

Raphael A Mrode. *Linear models for the prediction of animal breeding values*. Cabi, 2014.

Diana M Negoescu, Peter I Frazier, and Warren B Powell. The knowledge-gradient algorithm for sequencing experiments in drug discovery. *INFORMS Journal on Computing*, 23(3): 346–363, 2011.

Hans-Peter Piepho and Jens Mohring. Computing heritability and selection response from unbalanced plant breeding trials. *Genetics*, 177(3):1881–1888, 2007.

Carl Edward Rasmussen and Christopher K. I. Williams. Gaussian processes for machine learning. *the MIT Press*, 2006.

Dar A Roberts, Keely L Roth, Erin B Wetherley, Susan K Meerdink, and Ryan L Perroy. Hyperspectral vegetation indices. In *Hyperspectral indices and image classifications for agriculture and vegetation*, pages 3–26. CRC press, 2018.

David Satterthwaite, Gordon McGranahan, and Cecilia Tacoli. Urbanization and its implications for food and farming. *Philosophical transactions of the royal society B: biological sciences*, 365(1554):2809–2820, 2010.

Tim Searchinger, Richard Waite, Craig Hanson, Janet Ranganathan, Patrice Dumas, Emily Matthews, and Carni Klirs. Creating a sustainable food future: A menu of solutions to feed nearly 10 billion people by 2050. final report, 2019.

Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2016. doi: 10.1109/JPROC.2015.2494218. URL https://ieeexplore.ieee.org/document/7352306.

Benjamin J. Shields, Jason Stevens, Jun Li, Marvin Parasram, Farhan Damani, Jesus I.Martinez Alvarado, Jacob M. Janey, Ryan P. Adams, and Abigail G. Doyle. Bayesian reaction optimization as a tool for chemical synthesis. *Nature*, 590(7844):89–96, 2021. ISSN 14764687. doi: 10.1038/s41586-021-03213-y. URL http://dx.doi.org/10.1038/s41586-021-03213-y.

Rajendra P Sishodia, Ram L Ray, and Sudhir K Singh. Applications of remote sensing in precision agriculture: A review. *Remote sensing*, 12(19):3136, 2020.

Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, 4: 2951–2959, 2012. ISSN 10495258.

Doug Speed, Gibran Hemani, Michael R Johnson, and David J Balding. Improved heritability estimation from genome-wide snps. *The American Journal of Human Genetics*, 91(6):1011–1021, 2012.

Trevor Standley, Amir Zamir, Dawn Chen, Leonidas Guibas, Jitendra Malik, and Silvio Savarese. Which tasks should be learned together in multi-task learning? In *International conference on machine learning*, pages 9120–9132. PMLR, 2020.

Yuxiang Wang and Bruno Ribeiro. Probabilistic transformers: Integrating deep learning with Bayesian inference for uncertainty modeling. *arXiv preprint arXiv:2006.12320*, 2020.