

Accelerating materials discovery is a critical challenge in both academic and industrial research. Traditional trial-and-error approaches are often slow and resource-intensive, especially when exploring vast compositional and processing spaces. Active learning revolutionizes materials discovery by strategically guiding experimental synthesis through machine learning algorithms. In this work, we demonstrate how we used machine learning, powered by active learning algorithms, can strategically guide experimental synthesis to maximize information gain and efficiency. In this study, we applied a Bayesian optimization framework to select the most informative next experiments in the development of two targets properties. This approach enabled us to significantly reduce the number of experiments required to achieve target properties, compared to conventional grid search methods. To facilitate the adoption of these techniques by researchers and industry partners, we developed MADGUI, an intuitive, open-source graphical user interface built with Streamlit. MADGUI streamlines the active learning workflow, allowing users to visualize predictions, monitor uncertainty, and interactively select new candidates for synthesis. Our collaborative results showcase the practical impact of integrating active learning into real-world materials research, highlighting substantial gains in efficiency and discovery speed. This work underscores the transformative potential of machine learning-driven experimentation for accelerating innovation in materials science.

## Active learning

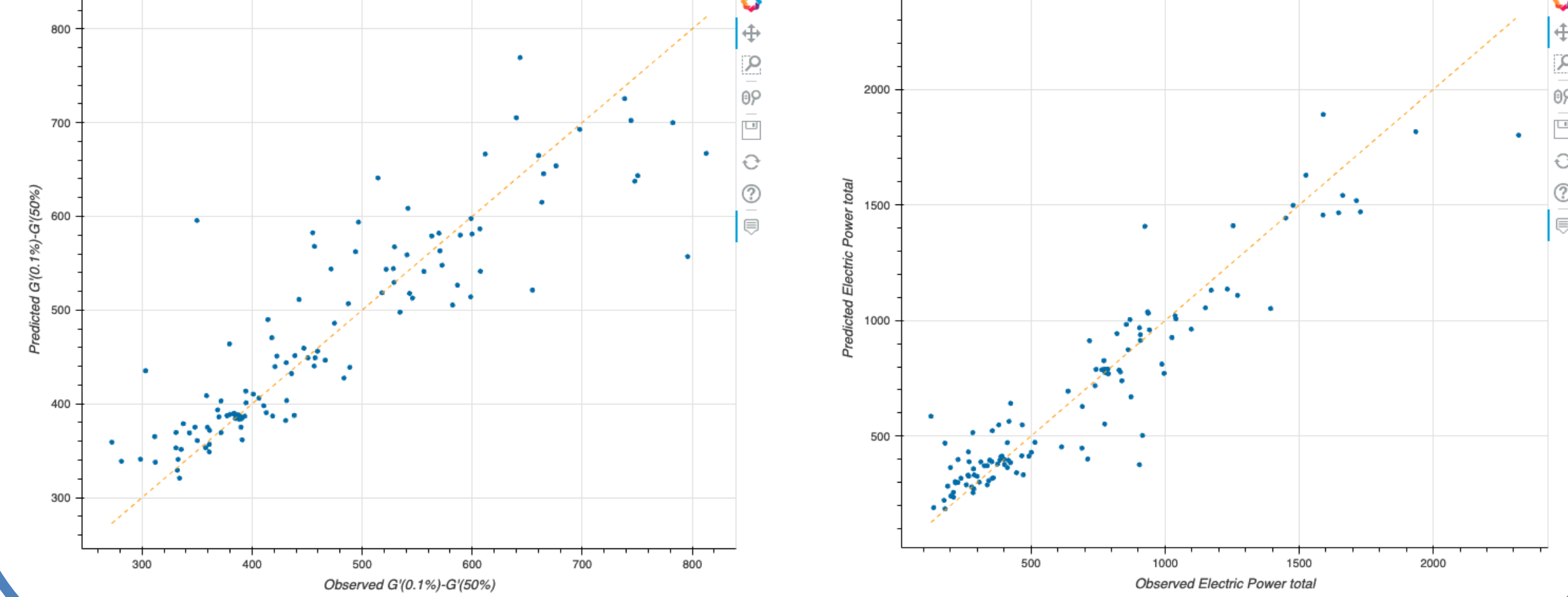
In this study, active learning was implemented across 11 experimental parameters, creating a vast search space of approximately  $10^{18}$  possible combinations, far beyond what could be exhaustively explored. The workflow operates in iterative cycles: after each round of experiments, the experimentalists provide newly measured data along with updated parameter boundaries. This data is then pre-processed to meet the requirements of the machine learning algorithm. Some linear correlation and prediction model are applied to analyze the data and Bayesian Optimization propose the most informative next experiments. These suggested parameter sets are then returned to the experimentalists for synthesis and measurement, and the cycle repeats, enabling efficient exploration and accelerated materials discovery.

## Prediction model

After several iterations, the active learning process yields a comprehensive understanding of the parameter space, allowing us to efficiently map regions associated with optimal material properties. By leveraging the algorithm's balance between exploration (sampling uncertain regions) and exploitation (focusing on promising candidates), we progressively refine our predictive models. This approach enables accurate identification of parameter sets that meet or exceed target performance criteria, greatly enhancing our ability to predict and achieve desired material outcomes with fewer experiments.

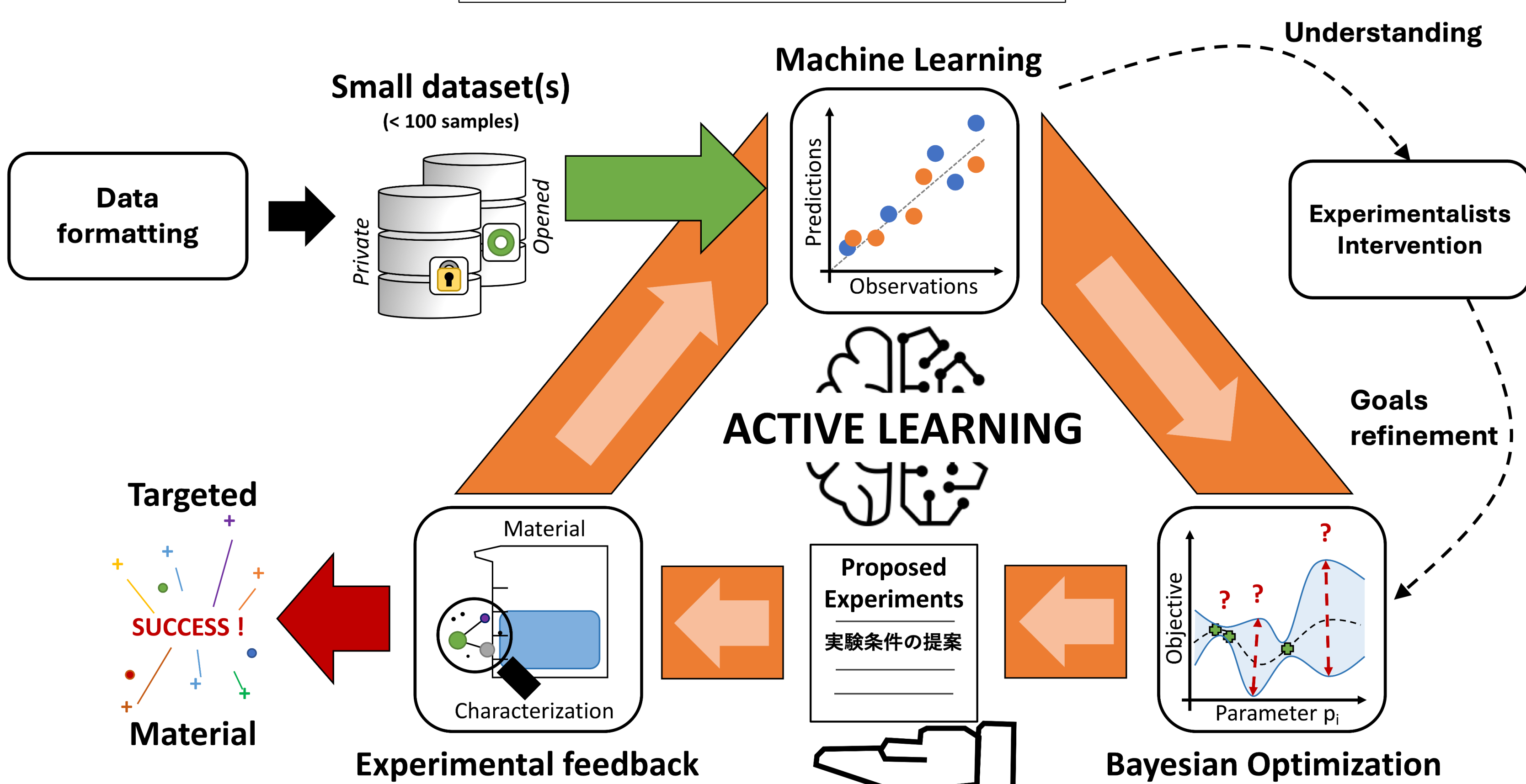
Non-linear regression on  $G'(0.1\%)-G'(50\%)$  with RandomForestRegressor and K-Fold = 10 : MAE: 38.116 +/- 10.789 RMSE: 55.431 +/- 17.979

Non-linear regression on Electric Power total with RandomForestRegressor and K-Fold = 13 : MAE: 103.799 +/- 39.021 RMSE: 139.946 +/- 56.885



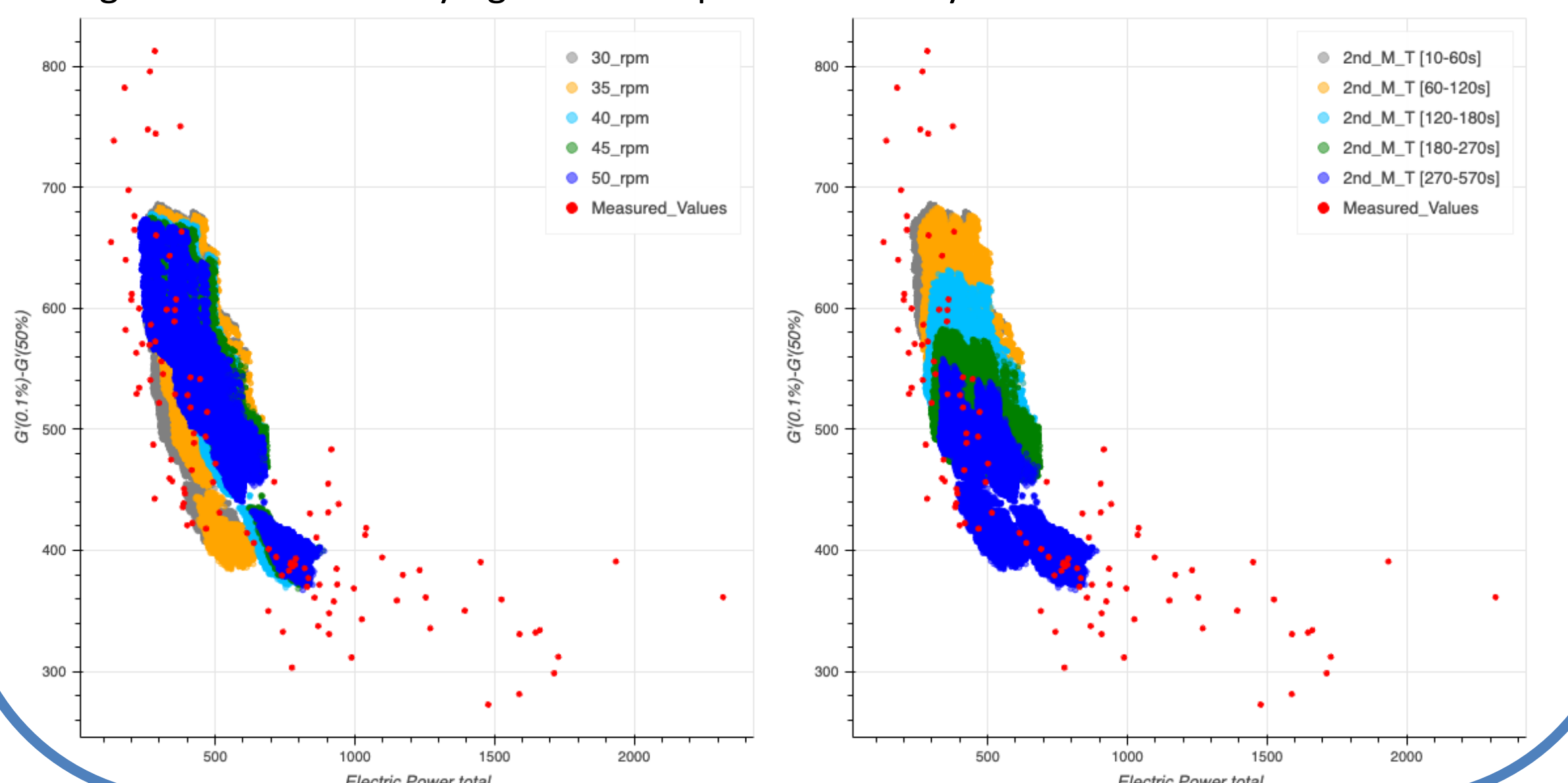
Prediction model for the two targets of this study

## Active learning general principle



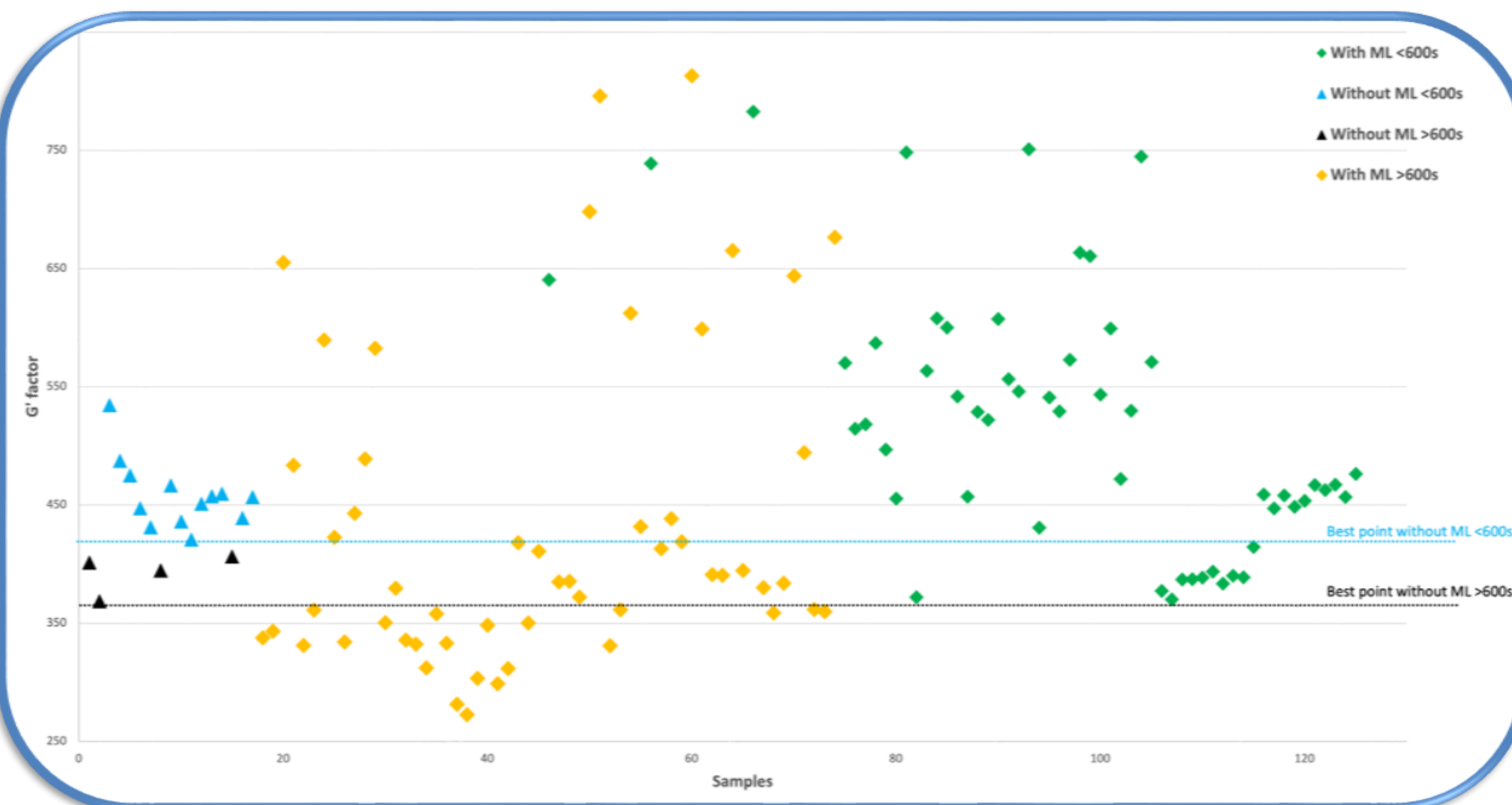
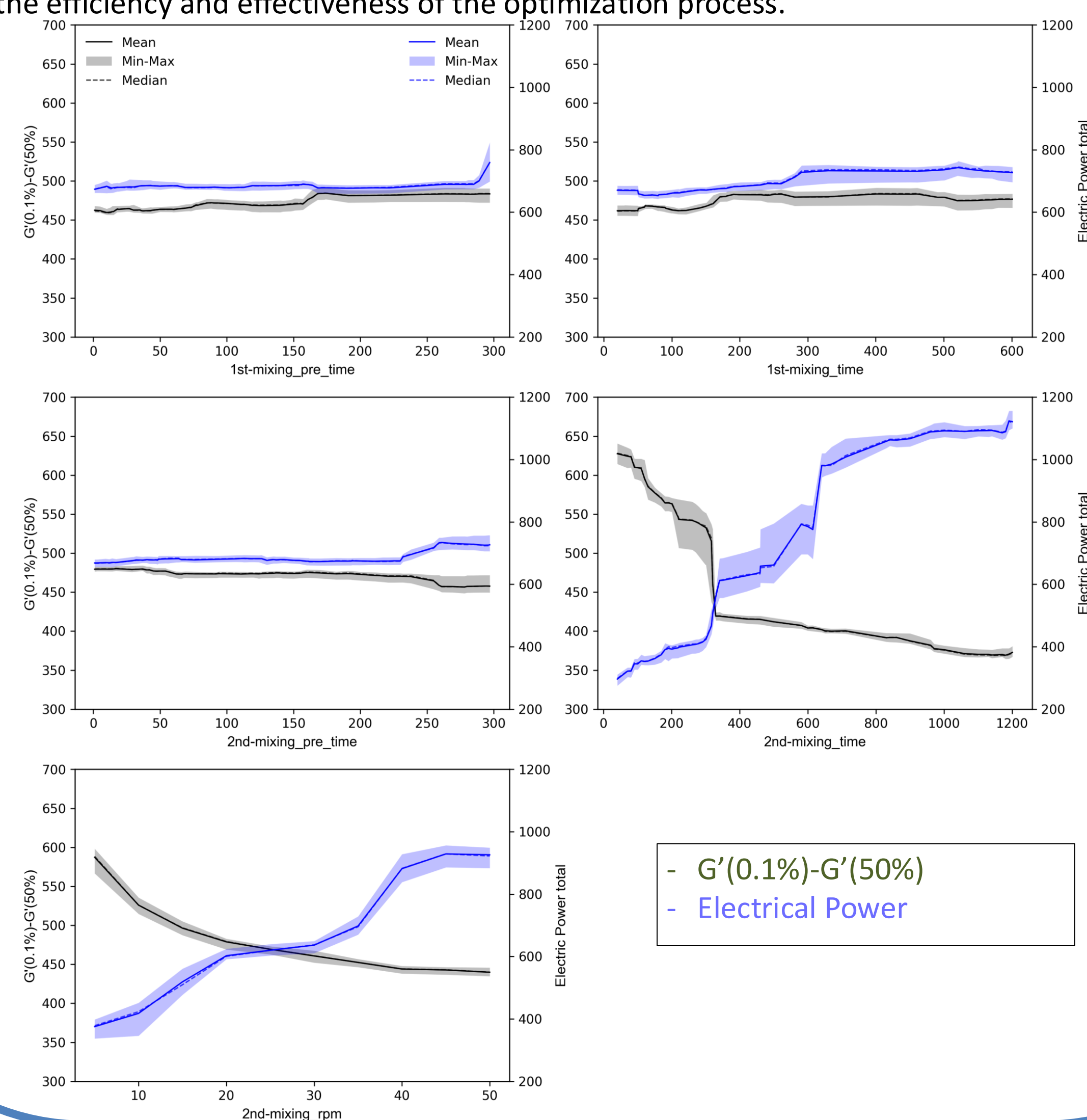
## Data Analysis

For the latest iterations, we reduced the number of parameters from 11 to 5, focusing on those most relevant to our target properties. Using a prediction model trained on the experimental dataset, we predicted target values based on these selected parameters. This analysis revealed the trade-offs between the two key target properties and allowed us to identify the two most influential parameters for each target. Notably, we observed a clear linear correlation between rpm and electrical power, while mixing time exhibited a negative linear correlation with the  $G'$  factor, providing valuable insights into the underlying relationships within the system.



## Features importance

A well-performing prediction model enables us to quantify the influence of each parameter on the optimization of target properties. By analyzing feature importance, we can identify parameters that have negligible impact on the target values, allowing us to reduce the dimensionality of the problem and streamline the experimental search space. Conversely, parameters with significant influence are prioritized, enabling us to focus subsequent studies on the variables that most strongly affect material performance. This targeted approach enhances both the efficiency and effectiveness of the optimization process.



Overall, in this study, the machine learning focus on the diminution of both targets at the same time resulted in an improvement of 46.3% with only 10 iterations, 11 parameters and a better understanding of the two main properties impact on the targets properties.