Data Science Capstone 2 WRF

Abhinit Sundar; aks94@njit.edu, Ansel Obergfell; ao383@njit.edu, Kamil Arif; kma56@njit.edu, Rafael Rodriguez; rar22@njit.edu

***Abstract*—This technical report outlines the research and development undertaken to advance optimization work with the Weather Research and Forecasting (WRF) model, specifically focusing on the WRF-Solar-BNL variant. The primary objective was to devise a computational steering framework to optimize model parameters and improve forecasting accuracy. The report details the challenges encountered, methodologies explored, and the ultimate success in implementing a customized solution. The investigation extends to alternative optimization techniques, ensuring a comprehensive understanding of the possibilities in refining weather forecasting models.**

I. Introduction

Ensuring precise weather predictions holds significant importance across diverse sectors, spanning from agriculture to energy generation. At the forefront of this project lies the WRF model, a prominent tool for atmospheric simulation. This project aimed to reproduce the findings of Wuji Liu's 2021 research[1]. Following the compilation of WRF-Solar and exploration of alternative methodologies, our subsequent approach involved engaging with the original researchers at Brookhaven National Laboratory who contributed to the development of the WRF model. Through ongoing discussions and meetings, we refined and enhanced our results.

II. Testing of other optimization methods

Bayesian Optimization (BO) is a robust method for global optimization of black-box functions, particularly useful in cases where evaluations are costly. For our use case, WRF runtimes can range from 3 to 6 hours, making BO a logical choice. Additionally, since the parameters of the WRF model are related to physical constants, we can reliably find a starting point that is near the true optimum. BO is able to capitalize on such situations and tune the starting parameters further.

Due to time constraints we were not able to thoroughly test every algorithm planned, however none of the proposed algorithms- with the exception of Stochastic Approximation- are likely to outperform BO for this use-case. Stochastic Approximation may be able to find the optimum faster, but there is much more variance in the amount of iterations required to do so compared to BO.

III. New Work

Our new step to work on optimizing our results from the previous endeavors was to collaborate and communicate with the BNL collaborators to further advance our success. After some questions about how the model functions and any way to further tune the model for better results they were able to distinguish that the main cause of our limited progress was the input data we were provided as that data was not completely compatible with the WRF model. It was missing proper data to help calculate CLDFRA(cloud fraction). Another suggestion given was on adjusting the domain size of the model since being too small would not result in accurate simulations. The BNL collaborators helped prepare some new sets of inputs for us to advance in our project. After receiving the appropriate data and making modifications to the model we were able to run some successful simulations and start on some BO simulated runs.

A. Acquisition Function Testing

We tested a variety of acquisition functions on this problem- due to time constraints the amount of iterations was limited, but we were still able to observe some differences in the optimization path between these different methods.

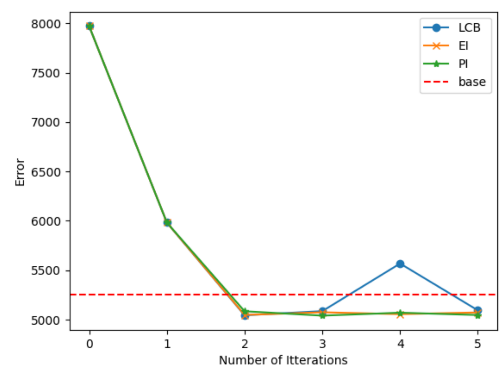

Fig. 1. Optimization paths for 3 acquisition functions compared against a baseline MSE.

IV. Future Work

Due to the effectiveness of Bayesian Optimization for this problem, future work should focus on variations of the algorithm suited to the project requirements, or potentially combining these methods.

A. Multi-Objective BO

The WRF model provides two relevant channels of output in direct and diffused irradiance. From these total irradiance can be derived. While current work has focused on optimizing based on total irradiance, better results may be found by optimizing on direct & diffused irradiance directly. Diversity-Guided Efficient Multi-Objective Optimization (DGEMO) is particularly suited to this task, as it is designed explicitly to optimize on multiple objective functions at once. [2]

B. Mixed-Variable BO

The WRF model has a variety of discrete parameters used to vary how it simulates certain effects, some of which can be changed without radically altering the model configuration. These were tuned manually for purpose of this project, but these parameters could instead be tuned alongside other continuous parameters using Mixed-Variable BO. The MIVABO algorithm [3] would be a good starting point for this approach.

C. Parallel BO

Evaluating multiple configurations of WRF in parallel would mitigate many of the issues caused by its long runtimes. BO typically is a sequential process that evaluates one point at a time- however work has been done to develop Batch BO, which tests multiple points at once. [4] From this, each evaluation would require each WRF instance to be set up in its own container, returning the simulation results at the end. WRF typically makes use of multiple cores for one evaluation, so some work would be required to find a balance between resources allocated to individual WRF instances, and the overall runtime of each evaluation batch.

V. Individual Contributions / Roles

- **Abhinit Sundar:** Spearheaded background research for this project, providing necessary context and aiding in the development process
- **Ansel Obergfell:** Continued to work on setup and debugging of the WRF model, as well as primarily managing its dedicated server. Also served as the point of contact with researchers at BNL.
- **Kamil Arif:** Handled the initial set-up of the WRF dedicated server, and explored a variety of alternatives to he WRF model throughout the course of the project.
- **Rafael Rodriguez:** Assisted model testing, planning future efforts, and collating project information / results.

Acknowledgment

We would like to extend our gratitude to professor, Dr.Wu, for his expert guidance, our BNL collaborators, Xin Zhou and Tao Zhang, for providing advice and knowledge on troubleshooting and tuning the WRF model, as well as Koli, Vrushali and Zhang, Yijie for their continued assistance on the project.

References

[1] W. Liu, Q. Ye, C. Q. Wu, Y. Liu, X. Zhou, and Y. Shan, "Machine learning-assisted computational steering of large-scale scientific simulations," in *2021 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)*, 2021, pp. 984–992. DOI: 10.1109/ISPA-BDCloud-SocialCom-SustainCom52081.2021.00138.

[2] M. Konakovic Lukovic, Y. Tian, and W. Matusik, "Diversity-guided multi-objective bayesian optimization with batch evaluations," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[3] E. Daxberger, A. Makarova, M. Turchetta, and A. Krause, "Mixed-variable bayesian optimization," in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, ser. IJCAI-PRICAI-2020, International Joint Conferences on Artificial Intelligence Organization, Jul. 2020. DOI: 10.24963/ijcai.2020/365. [Online]. Available: http://dx.doi.org/10.24963/ijcai.2020/365.

[4] N. Hunt, "Batch bayesian optimization," Available at https://hdl.handle.net/1721.1/128591, S.M. Thesis, Massachusetts Institute of Technology, Cambridge, MA, Feb. 2020.