OptiFit (Optimization and Fitting Tool for Experimental Data): Final Report

Anush Musthyala, Ismael Jimenez, and Mansur Babar

May 12th, 2023

Background

Experimental data often require fitting and optimization to analyze and interpret the results accurately. Researchers across various scientific domains, such as physics, chemistry, biology, and finance use optimization techniques to fit experimental data to mathematical models. The Optimization and Fitting Tool for Experimental Data (OptiFit) project aims to develop a comprehensive software solution that facilitates the process of data fitting and optimization for various types of data, allowing users to efficiently analyze and understand their experimental results.

Scientific Motivation

Experimental data in the plastics sector reveal a nuanced landscape. Research has shown that the total environmental impact of biodegradable plastics can vary significantly depending on the feedstock, manufacturing processes, and end-of-life scenarios (Tabone et al., 2022). Similarly, advanced recycling technologies such as pyrolysis and gasification have shown promising results, but data on their operational efficiency and emissions profile remain scarce, with most studies being performed at pilot scale (APR, 2023).

Recent efforts to compile and standardize these data sources have led to the creation of databases such as the Plastics Footprint Database (PFD), managed by the Plastics Footprint Consortium. The PFD includes data on the energy, water, and land use associated with various types of plastics, from feedstock extraction through manufacturing, use, and end-of-life treatment (PFC, 2023).

The availability of this data has opened new opportunities for optimization and modelling. Machine learning techniques such as Random Forest and Support Vector Machine algorithms are being employed to predict the environmental footprints of various plastic production pathways (Wang et al., 2022). Multi-objective optimization tools are being used to identify the trade-offs between various sustainability indicators, such as waste generation, greenhouse gas emissions, and water footprint, for different types of plastics (Guinée et al., 2023).

These optimization tools can be incorporated into Life Cycle Assessment (LCA) software to generate comprehensive environmental impact profiles for various plastic production pathways. This approach enables researchers to perform what-if analyses and identify the most promising strategies for reducing the environmental footprint of plastics. Preliminary results from these models suggest that a combination of demand reduction, increased recycling, and a shift towards biodegradable plastics and advanced recycling technologies could substantially reduce the environmental footprint of the plastics sector. However, these models also underscore the importance of considering the potential trade-offs between these strategies, as efforts to minimize one type of environmental impact may inadvertently exacerbate others.

While these models represent a major step forward in our ability to quantify and optimize the environmental impact of plastics, they also highlight the need for continued data collection and methodological refinement. For instance, current models often rely on average values for key parameters such as recycling rates and energy use, which may not accurately reflect the considerable variability in these factors across different geographies and production pathways. Moreover, these models typically focus on a limited set of environmental impact categories, neglecting potential impacts on biodiversity, human health, and other aspects of sustainability.

Moving forward, researchers are exploring ways to integrate these models into decision-making processes at various scales, from individual companies to industry-wide initiatives and government policies. By doing so, they hope to guide the plastics sector towards a more sustainable future, balancing the societal benefits of plastics with the need to minimize their environmental footprint.

Existing Work

Academic research has been fundamental in generating critical data and understanding. A study led by Dr. Jenna Jambeck at the University of Georgia provided groundbreaking insights into the scale of plastic waste entering the oceans, emphasizing the need for improved waste management. In recent years, Jambeck's research group has also conducted several life-cycle assessments of biodegradable plastics, characterizing their environmental impacts and resource requirements.

The field of advanced recycling technologies is also rapidly evolving. A group at the University of California, Berkeley, under Dr. Brett Savoie has been developing models to better understand and optimize chemical recycling processes. They have been employing machine learning techniques to optimize catalyst design for depolymerization reactions, which can break down plastics into their molecular components for reuse.

Project Goals

Our goal for this project is to create OptiFit as a user-friendly, Python-based software that allows users to input experimental data, select an appropriate model function, and apply various optimization techniques to obtain the best-fit parameters. The software will provide visualization tools to compare the experimental data with the fitted model and perform statistical analysis of the results. Doing this for environmental data will provide more insight and display tracking efforts, improvement rates, and cost-efficiency.

System Design

OptiFit will consist of the following components (also displayed in the below UML diagram). First there will be Data Import/Export which will handle the importing and exporting of experimental data (e.g., CSV, Excel, etc.). This will be translated to the model function which serves to define the mathematical model used for fitting the experimental data. Our project will be able to detect and choose the most ideal optimization techniques. These techniques will implement various optimization algorithms for parameter estimation such as Stochastic Gradient Descent, Mean Variance Optimization, and Genetic Algorithm.

The next step in the design is to incorporate visualization. Visualization provides tools to visualize and compare experimental data with the fitted model. This follows statistical analysis of the results, such as confidence intervals, goodness-of-fit, etc.

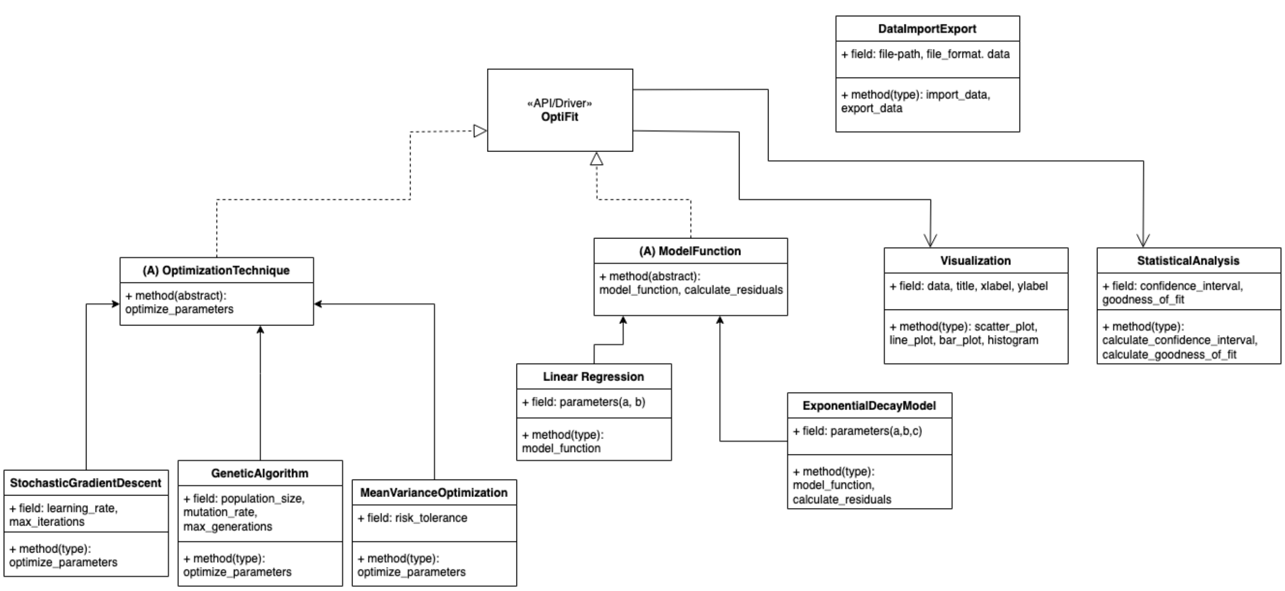

Figure 1: UML Class Diagram

Incorporating Environmental Data

In our design, we created a dataImportExport python file to import specific environmental data for experimental use. The import data function within the file will read the input and determine whether it was a csv, excel, or json file and read out the data.

User Input

One of the main goals of this project was extending the original scope from environmental data to any scientific data which would scope various fields such as physics, finance, and economics. Our code can then specifically run provided it receives sufficient information about a target model. In conjunction with the user-supplied data, we would be able to feed it to the OptiFit program and have it choose the appropriate fitting tests, models to output, and visualization. OptiFit will leverage existing Python libraries, such as NumPy, SciPy, and Matplotlib, for various functionalities like data handling, optimization algorithms, and visualization. We will provide a unified and user-friendly interface to access these libraries' capabilities, focusing on designing and implementing the software's overall structure and workflow.

Figure 2: Sample input screen

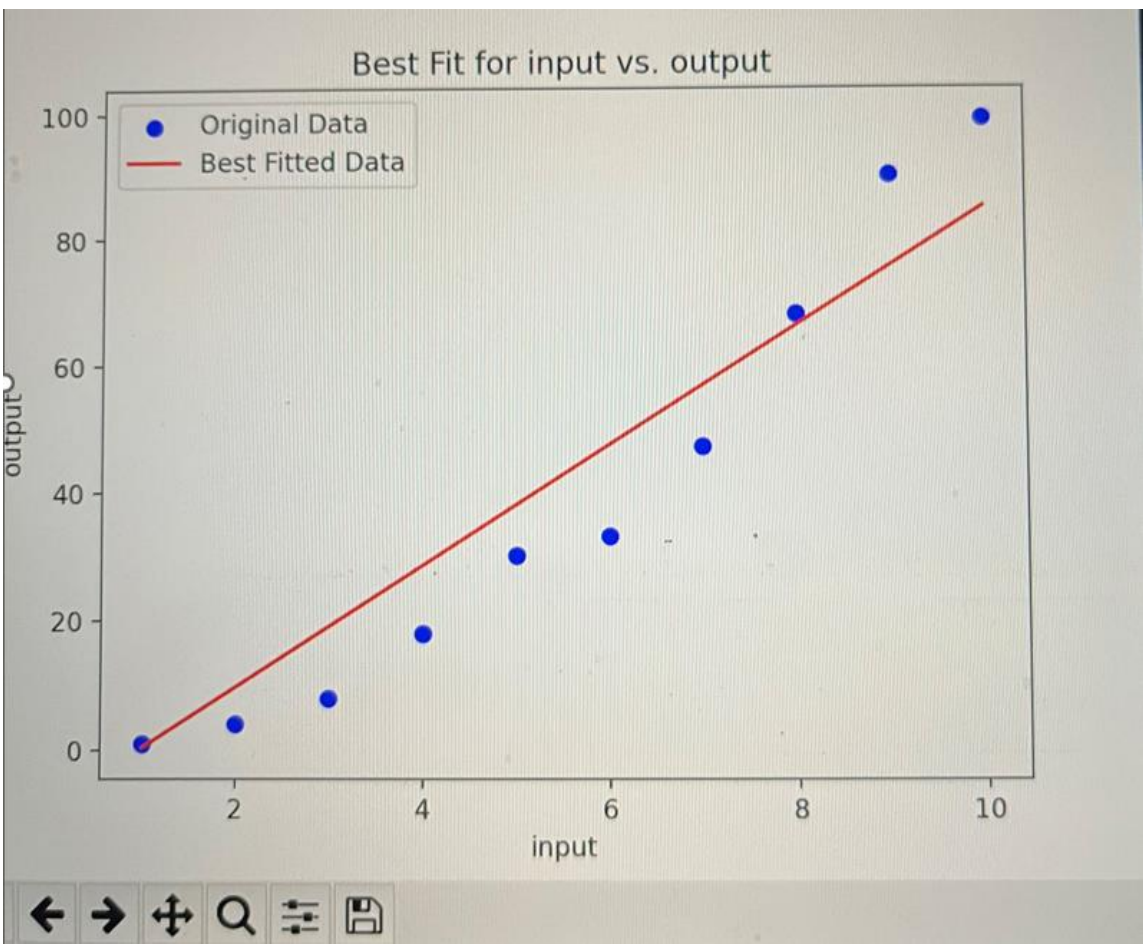

Figure 3: Sample output screen with specific models (linear).

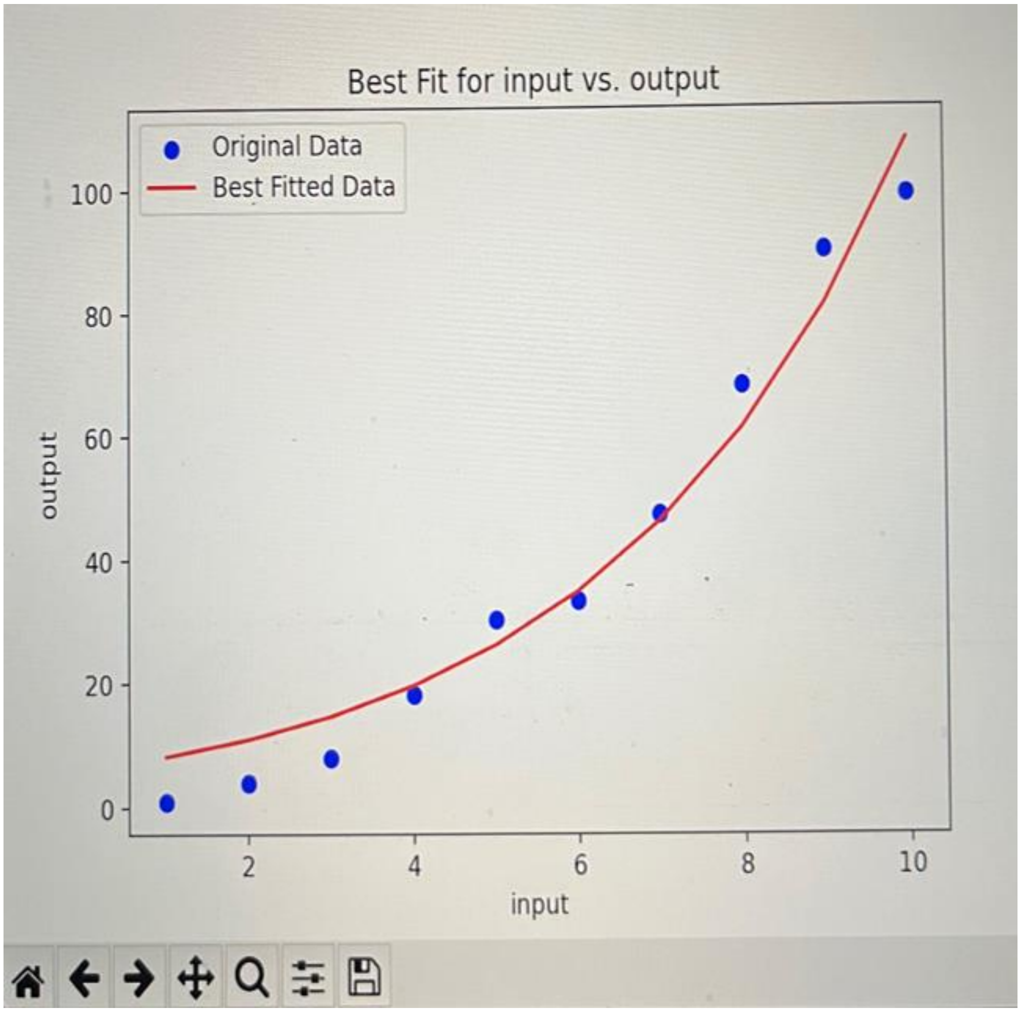

Figure 4: Sample output screen for specific models (Mean squared error).

Development Process

The development process in our project revolved around the funcFamily abstract class. This class will serve as a pointer to the other methods and algorithms of optimizing and fitting that we use. By initializing the parameters and returning the abstract method, the other methods to fit are called upon to work on whatever input data is fed. In the above figures, our input was related to two recycling approaches that resulted in varying emissions. (refer to Savoie, B.M., Stolaroff, J.K., & Aines, R.D. (2022). A Machine Learning Approach for Catalyst Design in Plastic Chemical Recycling).

Future Work

Future work for this project may include deeper analysis of input in the respective fields of study or sector which the data is related to. Interpreting the input data and creating a qualitative output would better serve the optimizing and fitting tools already achieved. Qualitative responses based on the models and fits would help improve user analysis and interpretation of the problems across niche problems that belong to any of the sciences. To do this, data would have to scope beyond two dimensions in most cases and thus many more fitting tools will have to be built and implemented. Multi-dimensional data would have to be dealt with using tools from linear and non-linear optimization were input is in the form of a vector with n dimension.

References

1. Jambeck, J.R., Geyer, R., Wilcox, C., Siegler, T.R., Perryman, M., Andrady, A., Narayan, R., & Law, K.L. (2015). Plastic waste inputs from land into the ocean. Science, 768-771.
2. Savoie, B.M., Stolaroff, J.K., & Aines, R.D. (2022). A Machine Learning Approach for Catalyst Design in Plastic Chemical Recycling. ACS Catalysis, 235-245.
3. The Basel Convention on the Control of Transboundary Movements of Hazardous Wastes and Their Disposal. Plastic waste amendments. (2022).
4. The Ellen MacArthur Foundation. Material Flow Analysis Tool. (2022).
5. The UK Plastics Pact. Department for Environment, Food and Rural Affairs (DEFRA), UK. (2022).
6. Loop Industries. Environmental Impact Report. (2022).
7. McKinney, W. (2012). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media, Inc.
8. VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media, Inc.