Project: Data Processing Software for Field-Flow-Fractionation Instrument

**Executive Summary**

This report details the development of data processing software for a Field-Flow-Fractionation (**FFF**) instrument used by the Nanometrology team at the National Measurement Institute (**NMI**). The FFF instrument is designed to derive the size distributions of particles in liquid suspensions, a critical task in the characterisation of nanoparticles. This project involved developing a Python-based software template to automate data processing, including baselines correction, peak detection, despiking, and calculation of hydrodynamic radius and molar mass. The results demonstrate the capability of the software to accurately process complex datasets, providing reliable size distribution measurements. The implications of this work are significant for improving the efficiency and accuracy of nanometrology practices, aligning with NMI’s objectives to advance measurement standards for methods.

**Introduction**

FFF is a powerful technique widely used in nanometrology to separate and analyse particles based on their size, shape, and density. The method is highly effective for characterising nanoparticles in liquid suspensions, providing insights into particle size distribution and other critical properties. Accurate measurement and characterisation are essential for research and development in various fields, including material science, biotechnology, and pharmaceuticals.

At the NMI, the Nanometrology team uses FFF instruments to conduct research and provide services related to nanoparticle measurement and characterisation. However, processing the raw data generated by the FFF instrument can be labour intensive and prone to inconsistencies. The motivation behind this project is to develop and automated data processing pipeline that leverages programming techniques to enhance the accuracy, repeatability, and efficiency of data analysis, thereby adding value to the NMI’s capabilities.

**Aims and Objectives**

The primary aim of this project is to develop a robust software solution for processing data generated by the FFF instrument, allowing for accurate determination of particle size distributions in liquid suspensions. The specific objectives are as follows:

1. Automate baseline correction and peak detection in raw FFF data.
2. Implement despiking algorithms to remove noise and artifacts from the data.
3. Calculate hydrodynamic radii and molar masses of particles based on the corrected data.
4. Validate the software’s performance by comparing the processed results with known standards and experimental data.

**Methods and Approach**

The project employs a Python-based approach to develop the software pipeline for FFF data processing. The development was carries out using Jupyter Notebooks, allowing for interactive coding and visualisation of the data analysis process. The data from the FFF instrument consists of time-series measurements of light scattering and UV absorbance, which require several preprocessing steps to derive meaningful information.

**Data Preprocessing**

The raw data from the FFF instrument often contains spikes and baseline drifts that need to be corrected before further analysis. The software pipeline includes a despiking algorithm that identifies and removes outliers based on statistical threshold. Baseline correction is performed using a polynomial fitting approach, which subtracts the baseline signal to normalise the data.

**Peak Detection and Analysis**

Peaks corresponding to different particle populations are identified using a peak detection algorithm based on derivatives and local maxima criteria. Each detected peak is analysed to determine its area, height, and width, which are crucial for calculating particle sizes.

**Calculation of Hydrodynamic Radius and Molar Mass**

The software calculates the hydrodynamic radius (RH) and molar mass (M) of particles using established theoretical models. The Rh is derived from the Stokes-Einstein equation, while M is calculated based on UV absorbance data, using calibration curves obtained from standards.

**Tools and Libraries Used**

The software development utilised Python libraries such as NumPy, SciPi, pandas, and Matplotlib for numerical computation, data manipulation, and visualisation. These tools were selected for their efficiency and reliability in handling large datasets and performing complex mathematical operations.

**Results and Analysis**

The developed software was tested on multiple datasets generated by the FFF instrument to validate its accuracy and efficiency. The following are the key results obtained.

1. **Baseline Correction**: The polynomial fitting approach effectively corrected baseline drifts, providing a flat baseline that aligns with zero. This correction was crucial for accurate peak detection and subsequent size distribution analysis.
2. **Despiking Efficiency**: The despiking algorithm successfully removed noise without affecting the integrity of the underlying signal. This was validated by comparing the processed data with manually cleaned data, showing a high correlation between the two.
3. **Peak Detection**: The peak detection module accurately identifies peaks according to different particle sizes. The results were consistent with expected size distributions for standard samples, with deviations within acceptable limts.
4. **Hydroynamic Radius and Molar Mass Calculations**: Have not yet calculated…

Figures and tables summarising the results are provided in the appendices, including size distribution plots, baseline correction comparisons, and calculated parameters for vaiours samples.

**Discussion**

The software developed in this project provides a reliable and efficient method for processing FFF data, significantly reducing manual effort and time required for analysis. The automated pipeline ensures consistency in data processing, which is crucial for repeatability and comparability of results in scientific research and industrial applications.

The accuract of the results obtained using the software indicates that the methods implemented for baseline correction, despiking and peak detection are effective for FFF data. However, the software’s performance could be further enhanced by integrating machine learning algorithms for more sophisticated noise filtering and peak identification, especially for complex datasets with overlapping peaks.

The project faced several challenges, including optimising and despiking algorithm to handle varying noise levels across the different datasets and fine-tuning the peak detection parametrs for different particle size ranges. Future work could focus on refining these algorithms and incorporating more advanced data analysis techniques, such as deconvolution methods, to improve accuracy in complex samples.

**Conclusion**

This project successfully developed a Python-based software solution for processing data from the FFF instrument at NMI’s nanometrology team. The software automates critical steps such as baseline correction, despiking, and peak detection, enabling accurate determination of particle size distributions. The results demonstrate the software’s potential to enhance nanometrology practices at NMI, providing a valuable tool for research and development in nanoparticle characterisation. Future improvements could include incorporating more advanced data analysis techniques and expanding the software’s capabilities to handle a broader range of samples and experimental conditions

**Appendices**

Thow additional details, theory, code, extra data analysis, size dist. Plots

**References**

Add Mar-Dean’s references, the references I read, and the different scientific papers I looked at for layout, language & presentation ideas.