



Entwicklung eines automatisierten Systems zur PAT Modellierung mit mehreren Spektralsensoren

Bachelor- oder Masterarbeit, Start ist flexibel

Hintergrund

Spektroskopische Methoden können in der pharmazeutischen Industrie und Forschung eingesetzt werden, um **kritische Qualitätsmerkmale (CQAs)** – wie Produktaggregate oder Wirtszellproteine – während der Produktaufreinigung zu überwachen. Allerdings liefert die Spektroskopie selbst keine direkte Aussage über die CQAs. Dafür ist eine **chemometrische Modellierung** erforderlich. Mithilfe dieser können die relevanten Qualitätsattribute aus den Rohspektren extrahiert werden. Daraufhin werden **Machine Learning (ML) Modelle** trainiert die genaue Aussagen über das zu messende CQA geben. Häufig können mehrere spektroskopische Methoden zur Überwachung eines CQAs genutzt werden. Welcher Sensor dabei am besten funktioniert kann über bestimmte Metriken (wie Korrelationskoeffizient oder Modellfehler) bestimmt werden. Die Auswahl des passenden Sensors und des optimalen ML Modells ist dabei ausschlaggebend für eine Erfolgreiche Messung des zu Qualitätsattributs.

Ziel der Arbeit

Das Ziel der Arbeit ist die Entwicklung eines Systems zur Auswahl der passenden Sensoren für die Bestimmung von Proteinkonzentration und –aggregation von drei Modellproteinen. Dazu wird ein System verwendet, welches automatisiert Raman, IR, UV und DLS Daten aufnehmen kann. Anhand dieser Daten soll dann objektiv entschieden werden welcher Sensor für das entsprechende Protein verwendet werden soll. Dieser Entscheidungsfindungsprozess wird daraufhin in einem Aufreinigungsschritt validiert. Zuletzt soll getestet werden ob sich die Robustheit und Qualität des Modells erhöht, wenn zusätzliche Sensoren (Data Fusion) bei der Modellbildung berücksichtigt werden.

Angewandte Methoden

- Spektroskopie (Raman, Infrarot, UV/Vis, DLS)
- Off-line Analytik (SEC, RP-HPLC)
- ÄKTA-Chromatographie / Filtration (Tiefenfiltration, UF/DF)
- Machine Learning und Datenanalyse mit Python

Zeitplan

	Monat #					
	1	2	3	4	5	6
Literaturrecherche						
Aufnahme der Spektren						
Datenanalyse und Modellbildung						
Validierung des Verfahrens im Prozess						
Evaluierung zusätzlicher Sensoren						
Schreiben der Arbeit						

Falls wir dein Interesse geweckt haben sende uns eine kurze Nachricht an raphael.niess@kit.edu und julian.gentes@kit.edu

Development of an automated system for PAT modelling with multiple spectral sensors

Bachelor or Master Thesis, Starting from January 2026 (flexible)

Background

Spectroscopic methods can be used in the pharmaceutical industry and research to monitor **critical quality attributes (CQAs)** - such as product aggregates or host cell proteins - during product purification. However, spectroscopy itself does not provide a direct measure of the CQAs. For this purpose, **chemometric modelling** is required to extract the relevant CQAs from the spectral data. Subsequently, **machine learning (ML)** models are trained to provide accurate predictions of the CQA to be measured. Often, multiple spectroscopic methods can be used to monitor a given CQA. Which sensor performs best can be determined using specific metrics (such as correlation coefficient or model error). The selection of the appropriate sensor and the optimal ML model is therefore crucial for a successful measurement of the quality attribute.

Scope of the thesis

The scope of this thesis is the development of a workflow to decide on the most appropriate sensor for monitoring of protein concentration and aggregation of three model proteins. The data is acquired on a automatic system which measures Raman, UV, IR and DLS data. The data is then processed and used to train machine learning models. An objective decision, based on these datasets, should be made to decide which sensor is most appropriate for the given task. This workflow will then be validated in a purification unit operation, such as filtration or chromatography. Lastly, it will be investigated if the addition of multiple sensors ("data fusion") enhances the robustness and quality of the trained machine learning models.

Applied skills

- Spectroscopy (Raman, UV/Vis, IR, DLS)
- Äkta chromatography / filtration (depth filtration, UF/DF)
- Data analysis and machine learning using Python

Timeline

	Month #					
	1	2	3	4	5	6
Literature review						
Acquisition of spectral data						
Analysis and model building						
In-process validation						
Evaluation of data fusion techniques						
Reporting						

If we have sparked your interest, please send us a short message to raphael.niess@kit.edu and julian.gentes@kit.edu