armccluskey.com/talks/m4colloids2019

Automating reflectometry reduction and analysis at Diamond Light Source

M4 Colloids — 2019/07/12
DOI: 10.6084/m9.figshare.8863589

Andrew R. McCluskey,[a,b] Tim Snow,[a,c] & Jos F. K. Cooper[d]

a. Diamond Light Source; b. University of Bath;
c. University of Bristol; d. ISIS Neutron and Muon Source

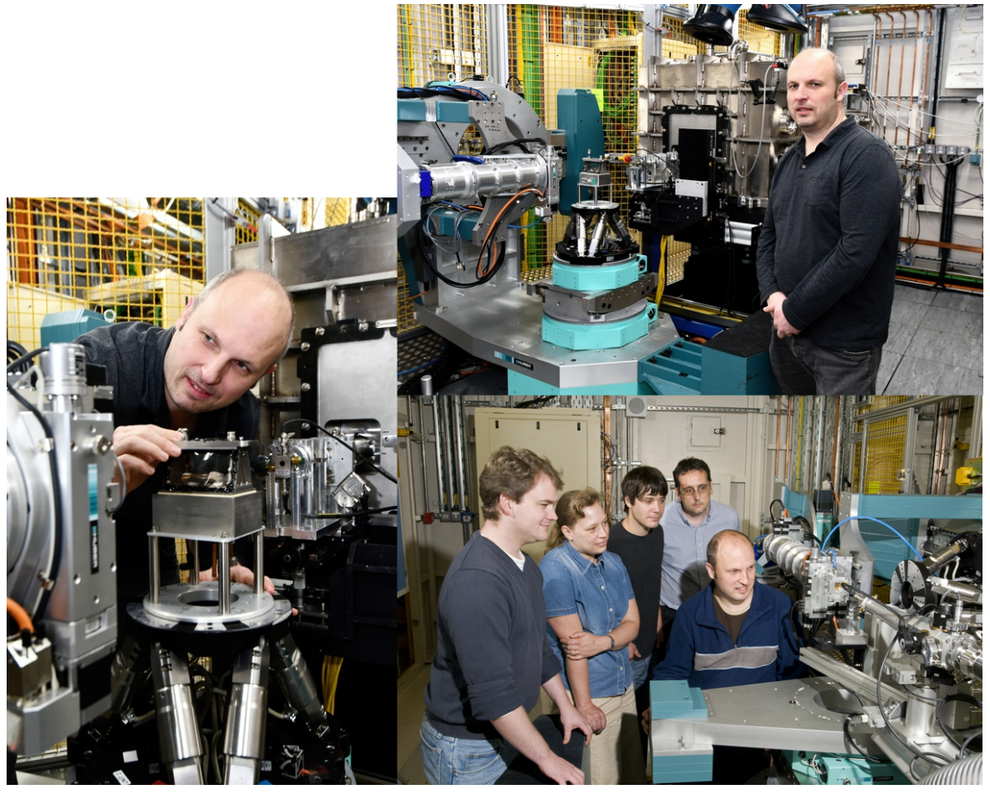
@ andrew.mccluskey@diamond.ac.uk
an_drewmcc
♂ (he/him)

@ andrew.mccluskey@diamond.ac.uk

reflectometry is a tool that uses x-rays or neutrons to probe interfaces

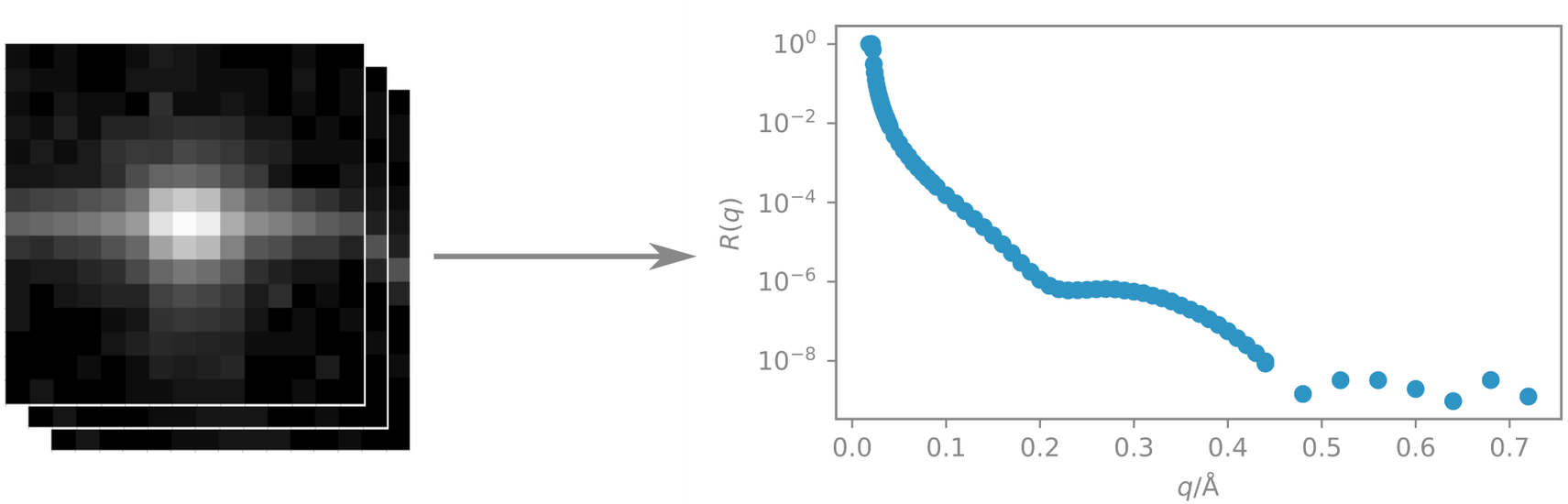
@ andrew.mccluskey@diamond.ac.uk

users bring cool science to the beamline



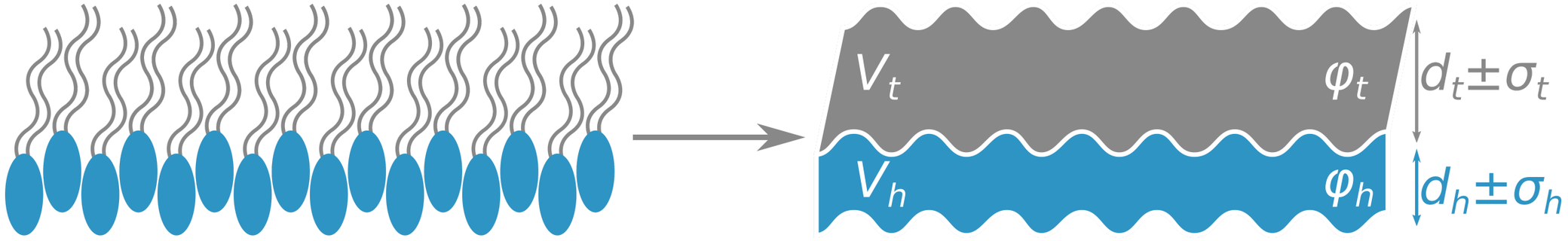
@ andrew.mccluskey@diamond.ac.uk

experimental data is reduced



@ andrew.mccluskey@diamond.ac.uk

data analysis is performed

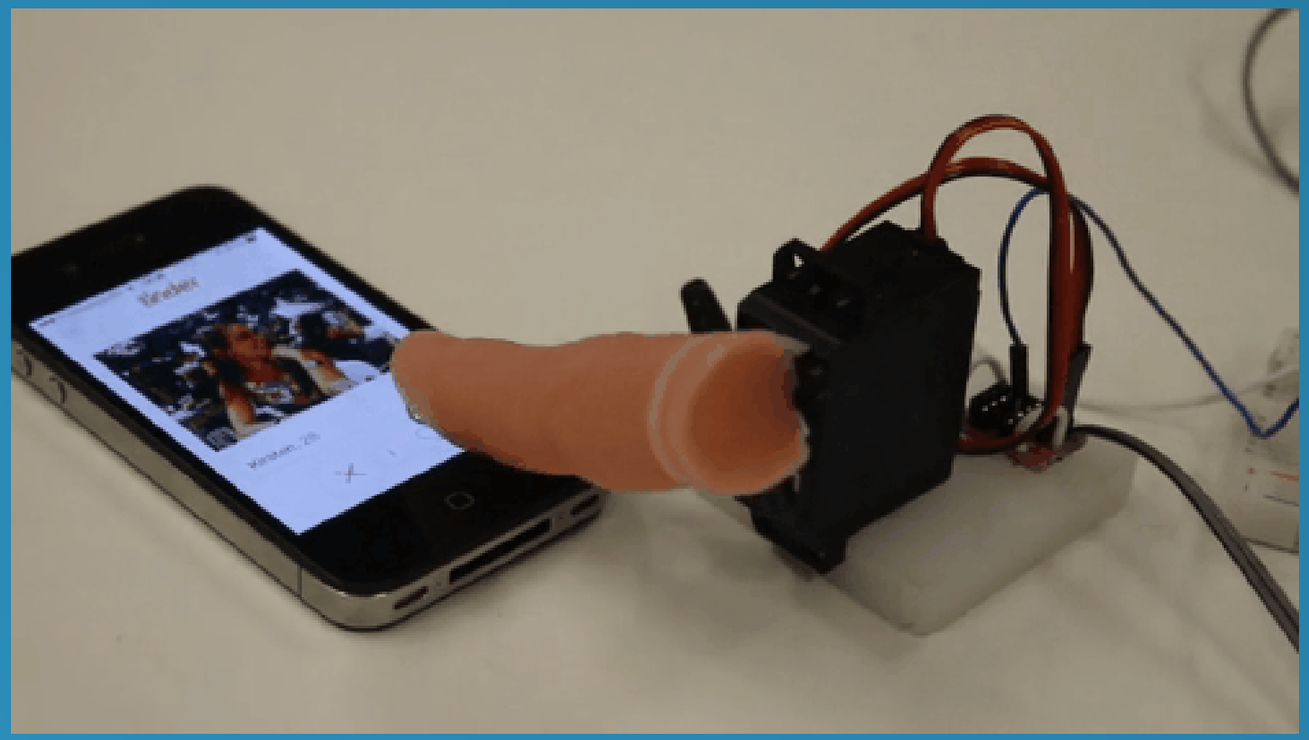


@ andrew.mccluskey@diamond.ac.uk

nature paper is written and published

@ andrew.mccluskey@diamond.ac.uk

greater automation → more science



hopefully good science too

@ andrew.mccluskey@diamond.ac.uk

what can we automate?

- Users bringing the cool science: 🤦
  - Data reduction: ✅
  - Data analysis: 🤞
  - Writing papers: 😔

@ andrew.mccluskey@diamond.ac.uk

let's automate data reduction

@ andrew.mccluskey@diamond.ac.uk

x-ray reflectometry reduction is well studied

Previous literature focuses on point detectors, while modern instruments utilise area detectors

A. Gibaud, G. Vignaud, & S. K. Sinha. *Acta Cryst.*, **A49**, 642-648, 1993
F. Salah, B. Harzallah, & A. van der Lee. *J. Appl. Crystallogr.*, **40**, 813-819, 2007.

@ andrew.mccluskey@diamond.ac.uk

not re-inventing the wheel, but re-building it

DAWN SCIENCE

Activities runtime-uk.ac.diamond.dawn.product - /scratch/lno76132/runtime-uk.ac.diamond.dawn.product - DAWN S... Mon Jul 8, 12:52:43

runtime-uk.ac.diamond.dawn.product - /scratch/lno76132/runtime-uk.ac.diamond.dawn.product - DAWN Science

File Tools Window Help

Quick Access | Resource | DataVis | Processing

Project Ex | File Viewe

data

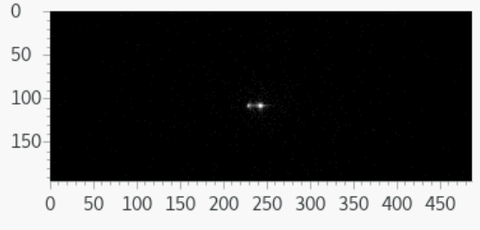
Data Slice View

312332.dat
312333.dat
312334.dat
312335.dat
312336.dat

Current slice of data: [0,:195,:487]

Input

file_image[0,:,:]



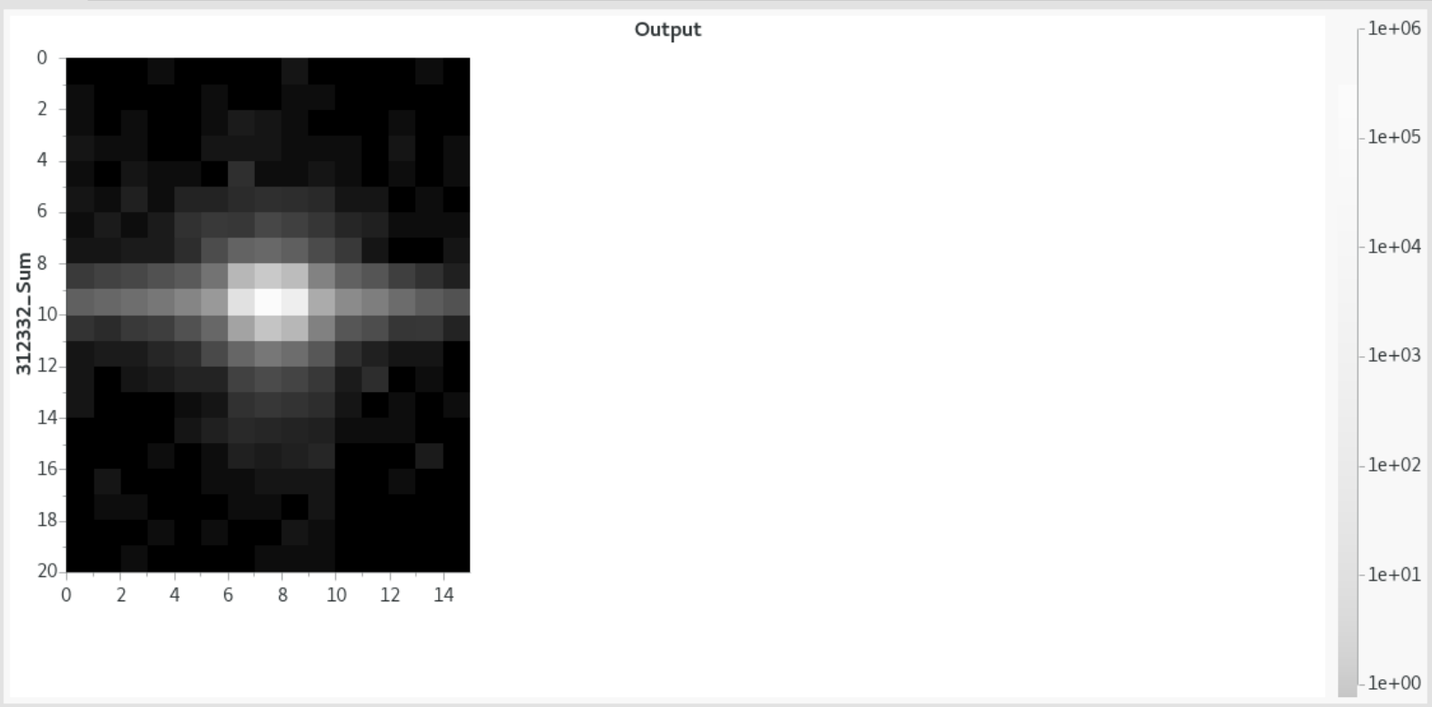
Processing

Name
Crop Around Peak [2D]
Image Integration
Gaussian Background Subtraction [Background Subtraction]
Set Poisson Error [Error]
Normalisation [metadata]
Q-Normalisation for DCD [Reflectivity]
Area Correction [Reflectivity]
Sum

Model 'Crop Around Peak [2D]'

Name	Value
x-Box Length	15.0
y-Box Length	20.0

Output



Activities runtime-uk.ac.diamond.dawn.product - /scratch/lno76132/runtime-uk.ac.diamond.dawn.product - DAWN S... Mon Jul 8, 12:52:59

runtime-uk.ac.diamond.dawn.product - /scratch/lno76132/runtime-uk.ac.diamond.dawn.product - DAWN Science

File Tools Window Help

Quick Access | Resource | DataVis | Processing

Project Ex | File Viewe

data

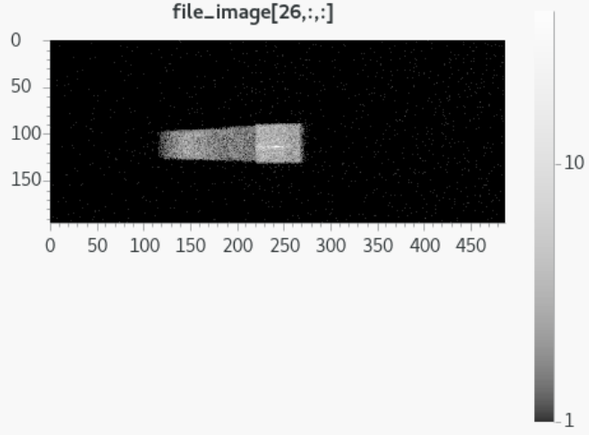
Data Slice View

312332.dat
312333.dat
312334.dat
312335.dat
312336.dat

Current slice of data: [26,:195,:487]

Input

file_image[26,:,:]



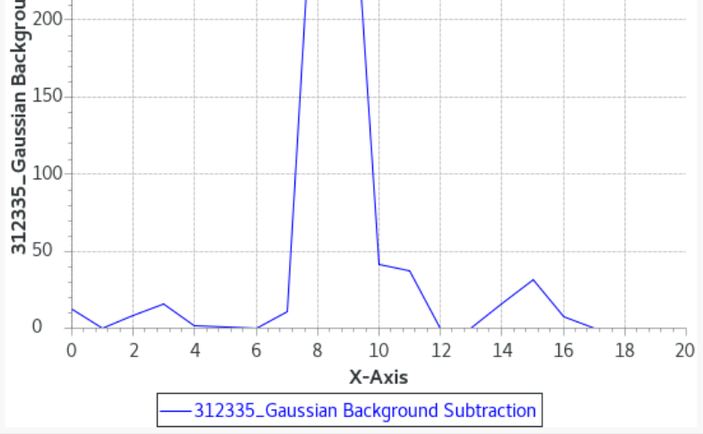
Processing

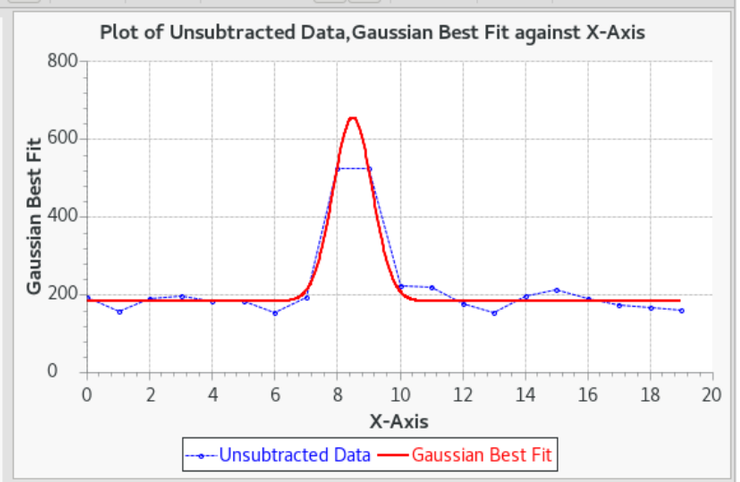
Name	
Crop Around Peak [2D]	
Image Integration	
Gaussian Background Subtraction	[Background Subtraction]
Set Poisson Error	[Error]
Normalisation [metadata]	
Q-Normalisation for DCD	[Reflectivity]
Area Correction	[Reflectivity]
Sum	

Model 'Gaussian Background Subtraction'

Name	Value

Output






```
A 1D Gaussian function with an x-axis offset
has been fit to the data.

The parameters are as follows:
Peak position = 8.506005E+00
Full-width half maximum = 1.473161E+00
Peak area = 7.395714E+02
x-Axis offset = 1.824224E+02
This offset has then been subtracted from the data.
```

Activities runtime-uk.ac.diamond.dawn.product - /scratch/lno76132/runtime-uk.ac.diamond.dawn.product - DAWN S... Mon Jul 8, 12:54:06

runtime-uk.ac.diamond.dawn.product - /scratch/lno76132/runtime-uk.ac.diamond.dawn.product - DAWN Science

File Plot Tools Transfer Window Help

6.6620E - 666.20E

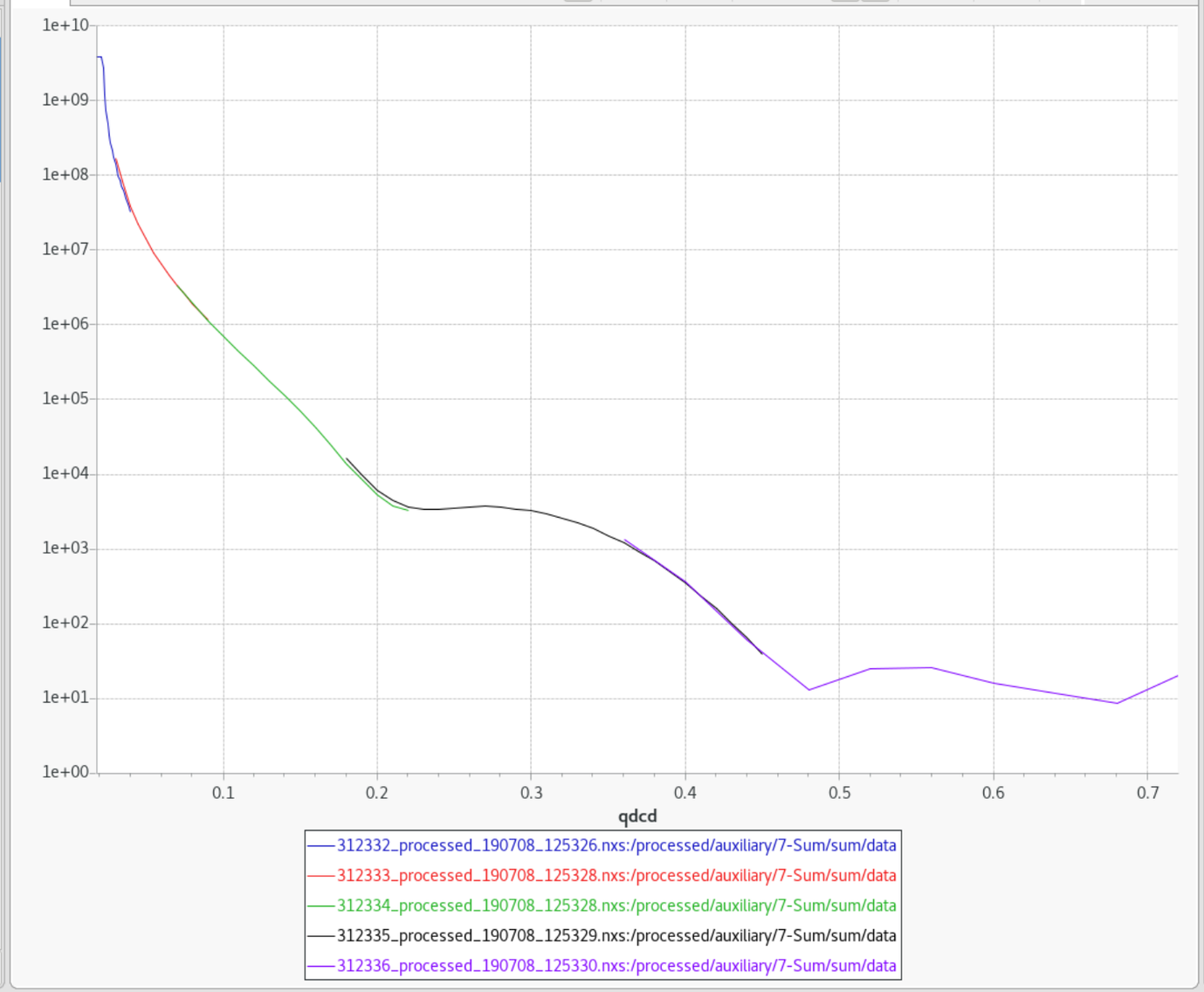
Quick Access | Resource | DataVis | Processing

Data Files

Filename
312332_processed_190708_125326.nxs
312333_processed_190708_125328.nxs
312334_processed_190708_125328.nxs
312335_processed_190708_125329.nxs
312336_processed_190708_125330.nxs

/home/lno76132/Desktop

Plot



Datasets

Dataset Name	Shape
/processed/result/data	[23, 20]
/processed/result/errors	[23, 20]
/processed/result/qdcd	[23]
/processed/process/origin/data_dimensions	[2]
/processed/auxiliary/2-Gaussian Background S	[23]
/processed/auxiliary/2-Gaussian Background S	[23]
/processed/auxiliary/2-Gaussian Background S	[23]
/processed/auxiliary/2-Gaussian Background S	[23]
/processed/auxiliary/2-Gaussian Background S	[23]
/processed/auxiliary/2-Gaussian Background S	[23]
/processed/auxiliary/2-Gaussian Background S	[23]
/processed/auxiliary/2-Gaussian Background S	[23]
/processed/auxiliary/7-Sum/sum/data	[23]
/processed/auxiliary/7-Sum/sum/errors	[23]
/processed/auxiliary/7-Sum/sum/qdcd	[23]
/processed/auxiliary/7-Sum/sum_of_sqrt/data	[23]
/processed/auxiliary/7-Sum/sum_of_sqrt/erro	[23]
/processed/auxiliary/7-Sum/sum_of_sqrt/qdc	[23]

Plot Type: Line

Dimensio	Display	Start:Stop:S	Axes
0 [23]	X	:	/processed/auxiliary/7-Sum/sum/q

Ready

0.642, 7.318

Activities runtime-uk.ac.diamond.dawn.product - /scratch/lno76132/runtime-uk.ac.diamond.dawn.product - DAWN S... Mon Jul 8, 12:55:28

runtime-uk.ac.diamond.dawn.product - /scratch/lno76132/runtime-uk.ac.diamond.dawn.product - DAWN Science

File Plot Tools Transfer Window Help

6.6620E - 666.20E

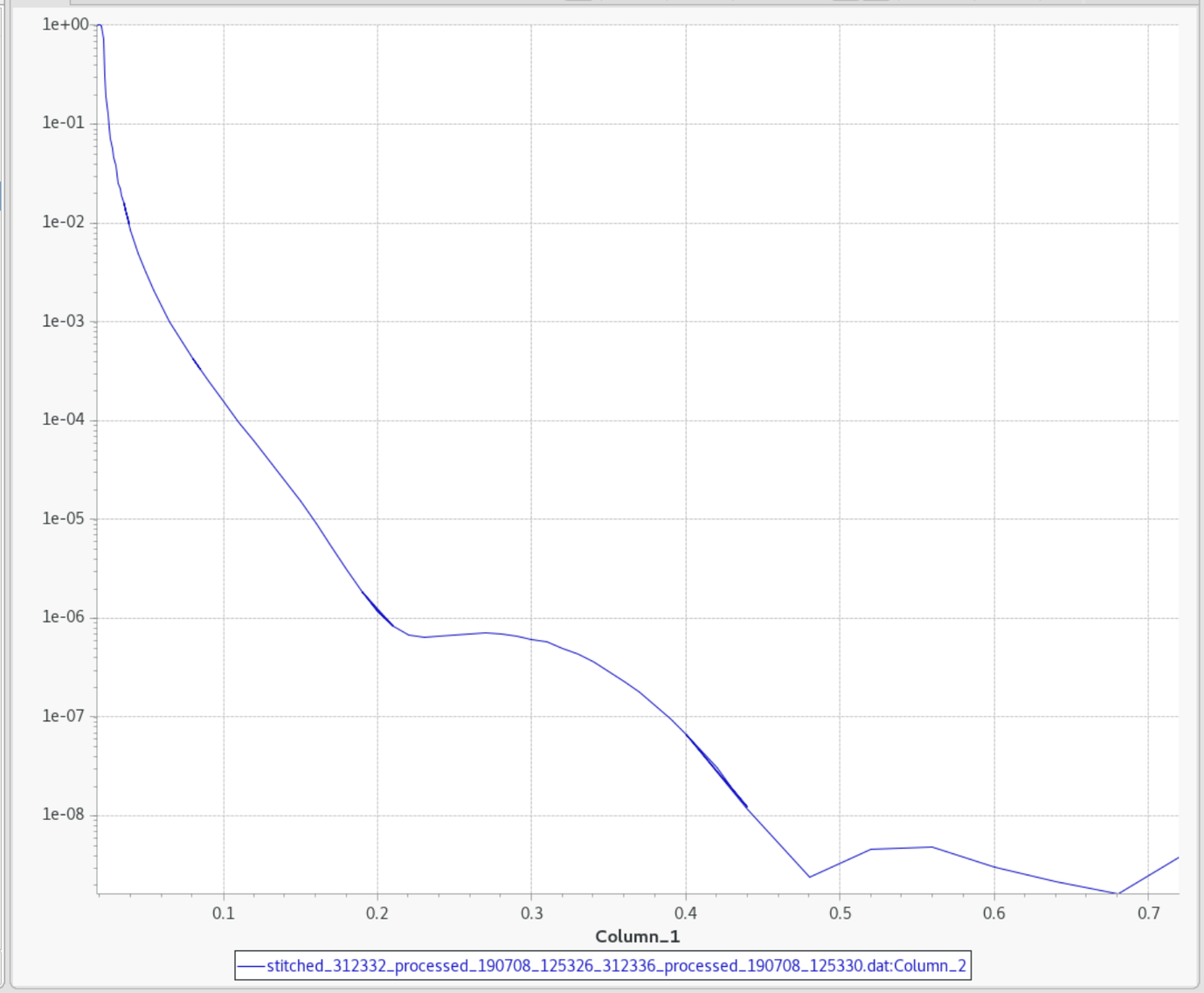
Quick Access | Resource | DataVis | Processing

Data Files

Filename
312332_processed_190708_125326.nxs
312333_processed_190708_125328.nxs
312334_processed_190708_125328.nxs
312335_processed_190708_125329.nxs
312336_processed_190708_125330.nxs
stitched_312332_processed_190708_125326_3123

/home/lno76132/Desktop

Plot



Datasets

Dataset Name	Shape
Column_1	[82]
Column_2	[82]
Column_3	[82]
Column_4	[82]

Plot Type: Line

Dimensio	Display	Start:Stop:S	Axes
0 [82]	X	:	Column_1

Ready

@ andrew.mccluskey@diamond.ac.uk

this will enable automated reduction

- Only a single processing pipeline for a given experiment
- The processing is modular
- Coming to i07 in October 2019

@ andrew.mccluskey@diamond.ac.uk

now for data analysis

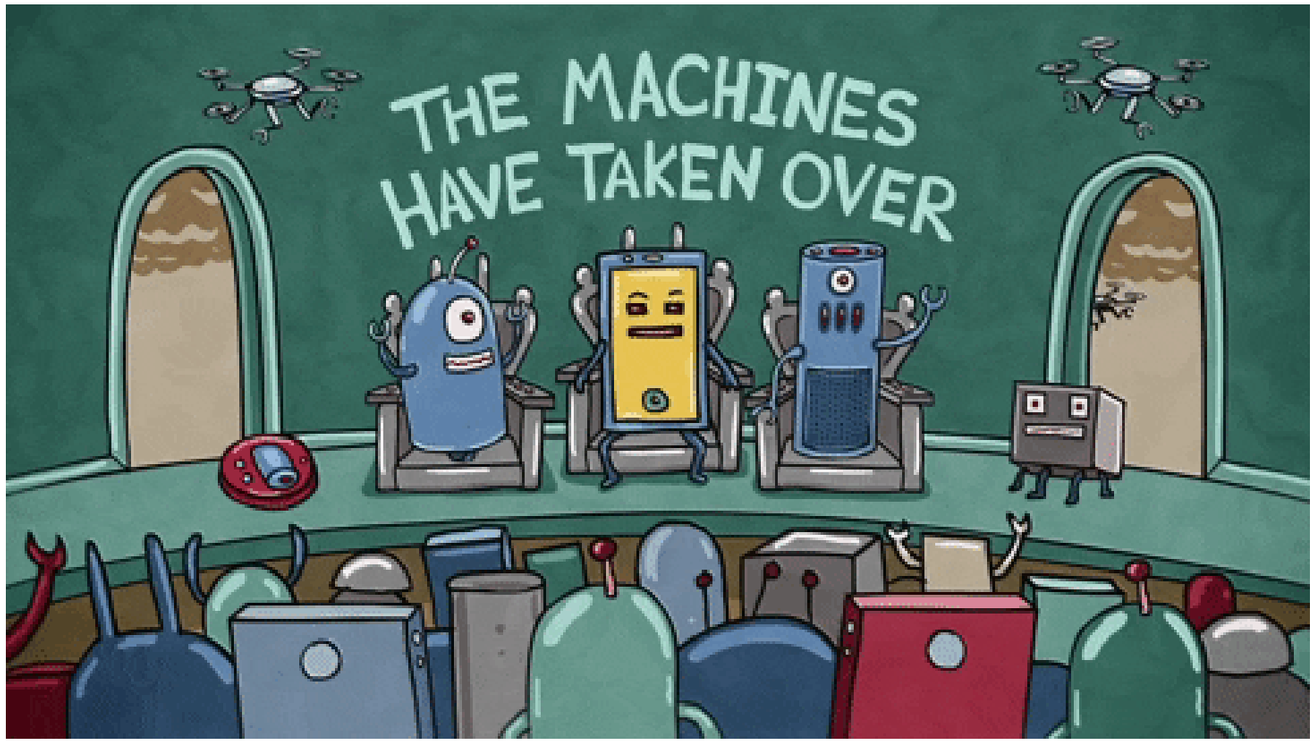
@ andrew.mccluskey@diamond.ac.uk

reflectometry analysis is model-dependent

Can we automate the model selection?

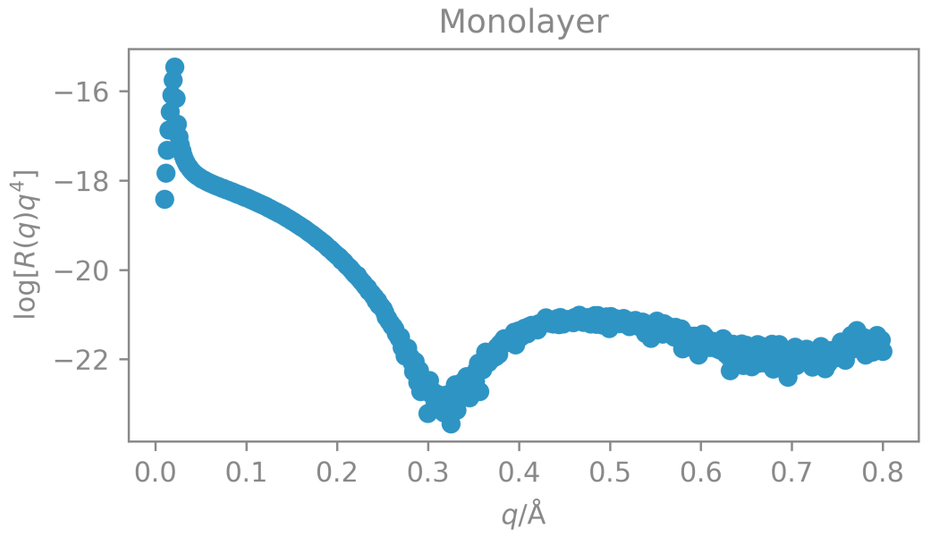
@ andrew.mccluskey@diamond.ac.uk

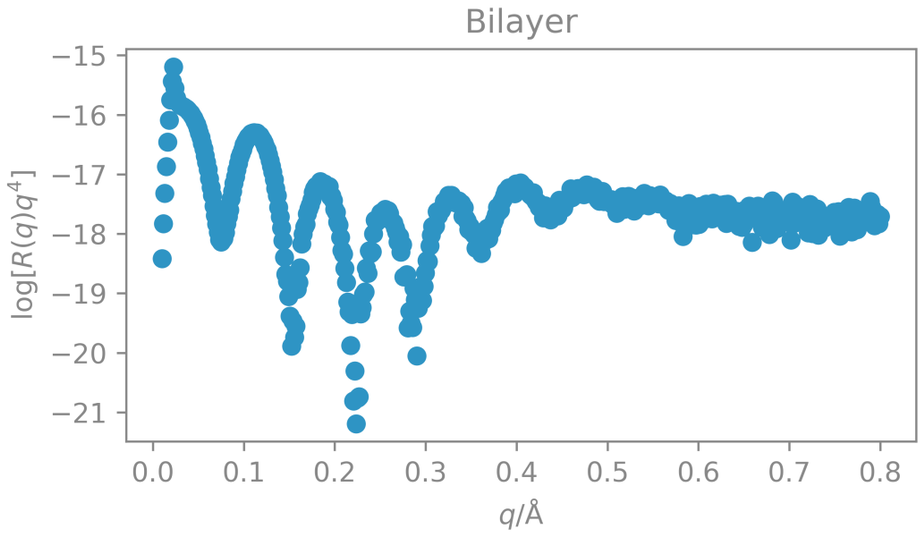
machine learning for model selection

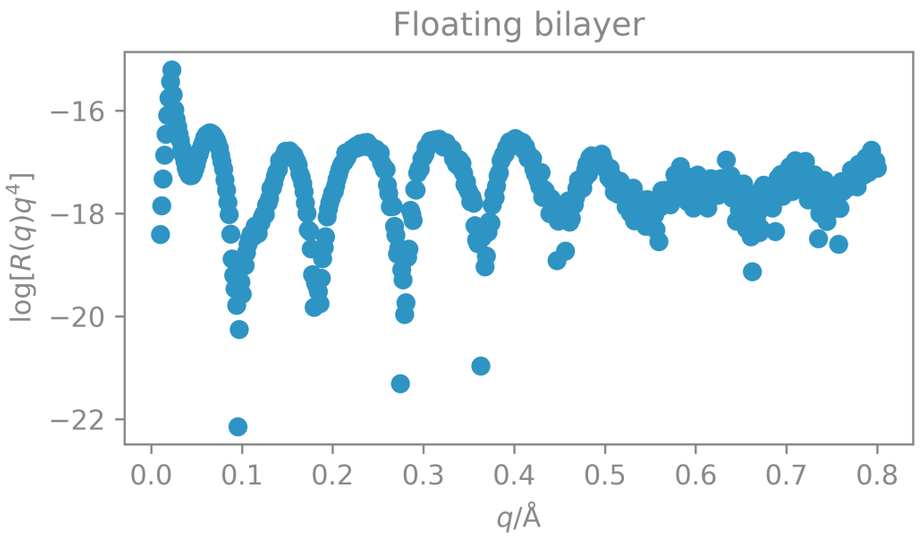


@ andrew.mccluskey@diamond.ac.uk

ml-driven classification

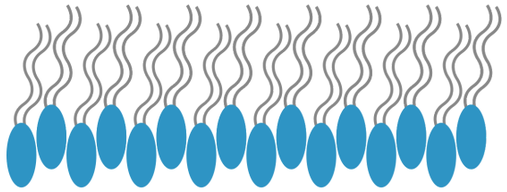




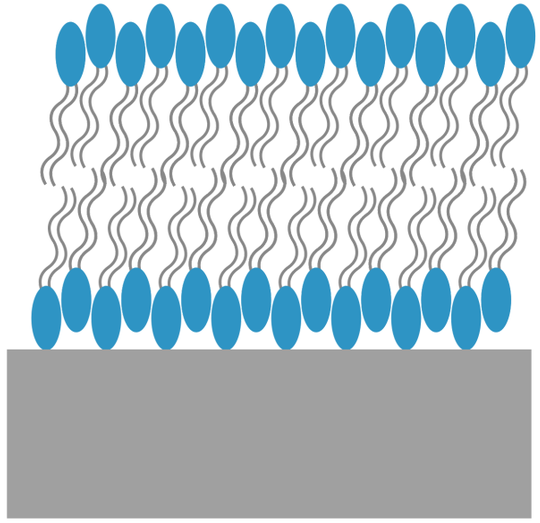


@ andrew.mccluskey@diamond.ac.uk

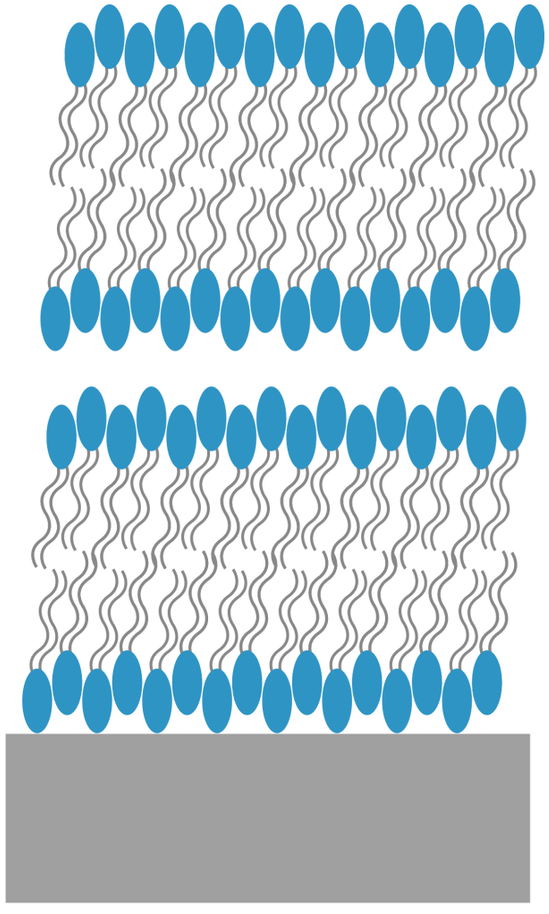
ml-driven classification



1



2



3

1. A. R. McCluskey, *et al.*. *Phys. Chem. Chem. Phys.*, **21**, 6133-6141, 2019.
2. Y. Gerelli. *Phys. Rev. Lett.*, **122**, 248101, 2019.
3. V. Rondelli, G. Fragneto, S. Motta, E. Del Favero, L. Cantù, *J. Phys. Conf. Ser.*, **340**, 012083, 2012.

@ andrew.mccluskey@diamond.ac.uk

convolutional neural network

Very quick and accurate for "model" data;
2 *epochs* → 100 % validation accuracy

@ andrew.mccluskey@diamond.ac.uk

how robust is our model?

	Monolayer	Bilayer	Floating Bilayer
DPPC Validation	100 %	100 %	100 %
DXPC Validation	100 %	100 %	100 %

@ andrew.mccluskey@diamond.ac.uk

why machine learning?

@ andrew.mccluskey@diamond.ac.uk

with ml we can advise analysis

We can create *generic* analysis interfaces which the machine learned classification would suggest

@ andrew.mccluskey@diamond.ac.uk

jupyter notebook

This could be in the form of a Jupyter Notebook that you could run from your office

jupyter **monolayer** Last Checkpoint: 3 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3

Run Markdown

Analysis of X-ray reflectometry from a surfactant monolayer

This Jupyter Notebook can facilite the analysis of your experimental X-ray reflectometry data from a surfactant/lipid monolayer. This will improve the reproduciblity and reliability of your analysis and results.

When writing your publication, please be sure to cite this Jupyter Notebook:

- A. R. McCluskey. (2019, July 12). Analysis of X-ray reflectometry from a surfactant monolayer (Version 1.0.0). Zenodo. DOI: 10.5281/zenodo.xxxxxxx

The publication that this model was taken from:

- A. R. McCluskey, A. Sanchez-Fernandez, K. J. Edler, S. C. Parker, A. J. Jackson, R. A. Campbell, and T. Arnold. (2019). *Phys. Chem. Chem. Phys.*, **21**, 6133-6141. DOI: 10.1039/C9CP00203K.

And the software that was used to perform the analysis:

- A. R. J. Nelson and S. W. Prescott. (2019). *J. Appl. Crystallogr.*, **52**, 193-200. DOI: 10.1107/S1600576718017296.

How to use

To use this analysis methodology, please follow through this notebook and fill in the required fields. Once all of the fields are completed, select "Cell" from the toolbar at the top of the page and choose "Run All". This will run the commands in all of the Python cells throughout the Notebook, generate an analysis output, which will describe the model system that best agrees with the experimental data, and publication quality figures for use in your paper/talk.

If you need any help along the way please contact andrew.mccluskey@diamond.ac.uk or scientificsoftware@diamond.ac.uk.

The model

The model is that of a surfactant or lipid monolayer that has self-assembled at an air-liquid interface. Shown below is a pictorial example of the expected structure, and the associated model that will be used in the analysis.

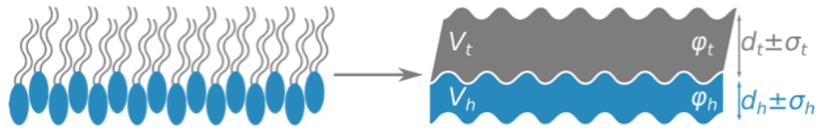


Figure 1. expected structure of the surfactant monolayer and a graphical description of the model.

@ andrew.mccluskey@diamond.ac.uk

automated script

Or an automated analysis script could be run on well tagged data

```
from __future__ import division
import refnx
from refnx.reflect import ReflectModel, SLD, Structure
from refnx.dataset import ReflectDataset
from refnx.analysis import (Transform, CurveFitter, Objective, GlobalObjective,

import numpy as np

import sys
sys.path.insert(0, "models/")
sys.path.insert(0, "utils/")

import two_layer as ol
import toolbox as tb

import json


surface_pressure = sys.argv[1]
seed = int(sys.argv[2])
contrasts = ['d13acmw', 'd13d2o', 'hd2o', 'd70acmw', 'd70d2o', 'd83acmw', 'd83d2
apm = float(sys.argv[3])
iden = sys.argv[4]
guide = json.load(open('scripts/guide.json', 'r'))
free_parameters = guide[iden]


data_dir = 'data/surf_pres_{}'.format(surface_pressure)

output_dir = 'output/all_data/{}/seed_{}'.format(iden, seed)

datasets = []
for i in range(len(contrasts)):
```

@ andrew.mccluskey@diamond.ac.uk

conclusions

- i07 will have automated data reduction by the year end
- We are developing ML-driven approaches to analytical model selection
- These will enable directed or automated down-stream analysis
- Stretch goal is users leave with analysed data

@ andrew.mccluskey@diamond.ac.uk

acknowledgements

- Tim Snow (DLS)
- Jos Cooper (ISIS)
- Keith Butler (SCD)

funding

ada lovelace centre

@ andrew.mccluskey@diamond.ac.uk

Penny and Sadie, good dogs