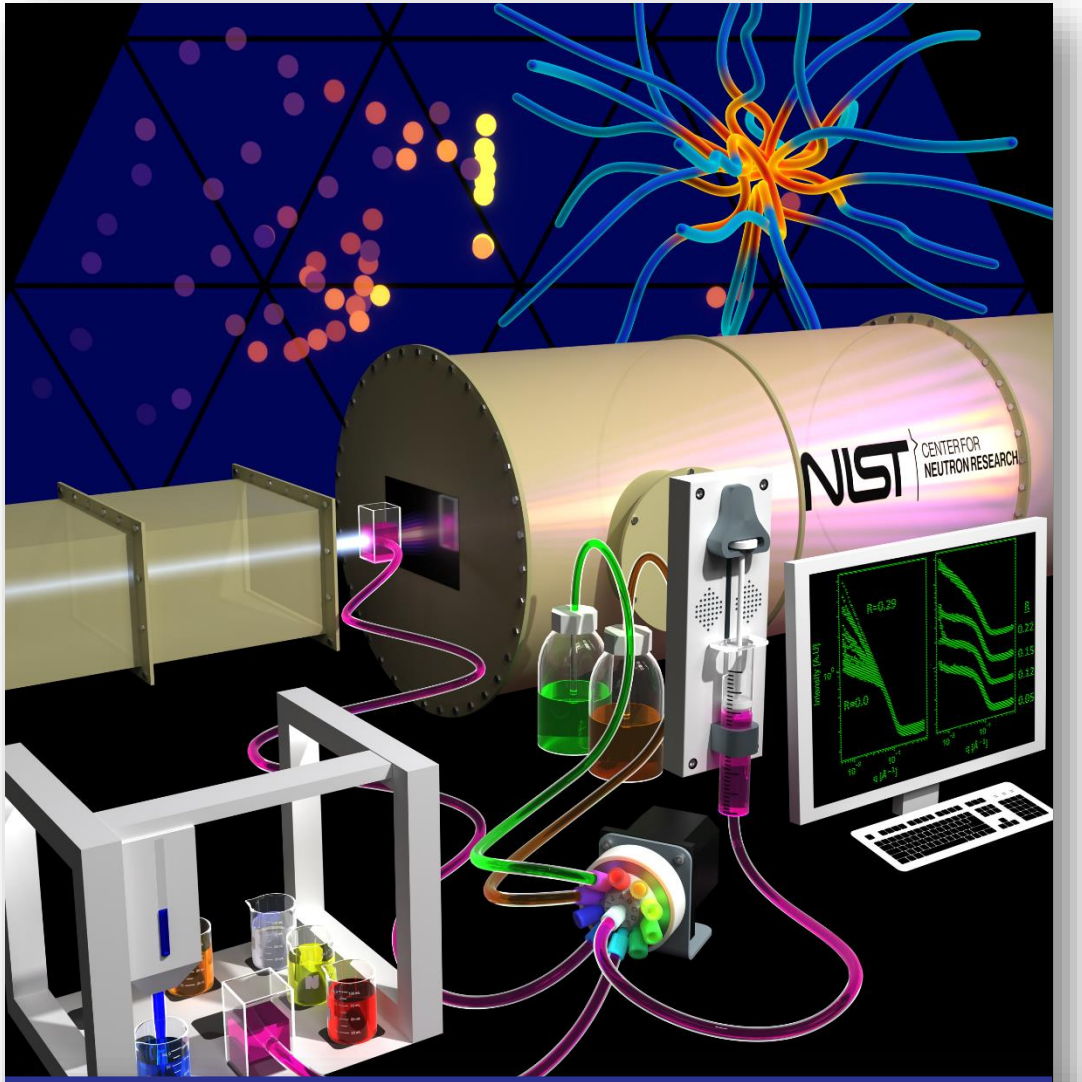


Tuning an AI Agent for Autonomous Phase Discovery

Sophia Liu (Poolesville High School)
Mentor: Dr. Tyler B. Martin (MML; NIST)

August 8th 2024

Formulations are Complicated

Ingredients of dish soap (~22 components):

Water, Sodium lauryl sulfate, Sodium laureth sulfate, C10-16 Alkydime-Thylamine oxide, Alcohol denat., PPG-26, Sodium chloride, Sodium hydroxide, PEI-14 PEG-24/PPG-16 copolymer, Phenoxyethanol, Methylisothiazolinone, Colorants, Fragrances, C9-11 Pareth-8, Tetrasodium glutamate, diacetate, Chloroxylenol, Phenoxy-isopropanol, Glycerin, Sodium cumene-sulfonate, Propylene glycol, Terpineol

- **The performance of formulations is dependent on their composition**
- **Small changes to composition can greatly change the formulation properties**

dawn-dish.com/en-us/how-to/what-dawn-is-made-of-ingredients/

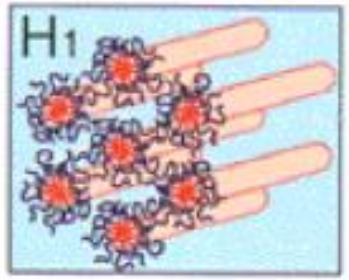
Formulation Composition Controls Nanostructure

Poloxamers!

Hydrophilic Hydrophobic Hydrophilic

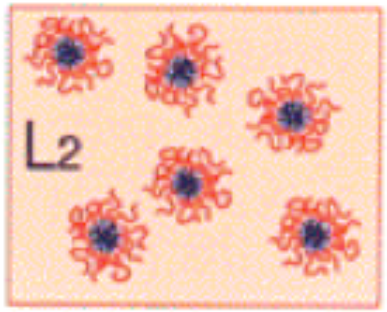
0% oil
50% water
50% poloxamer

Cylinders

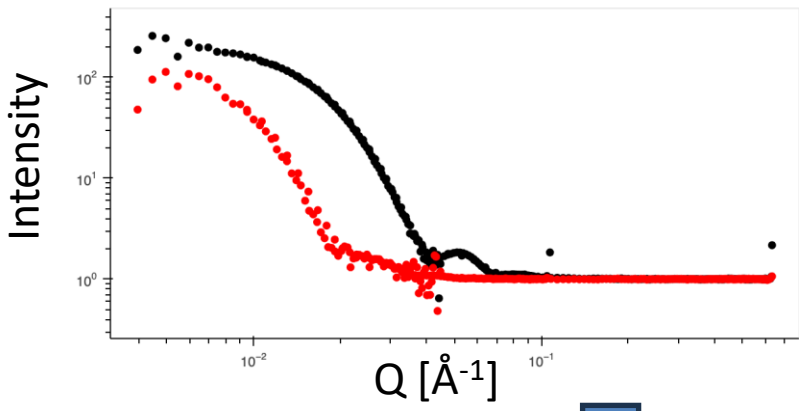
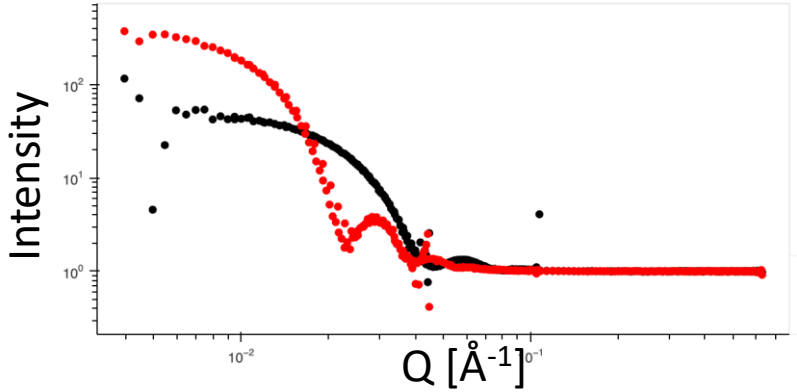
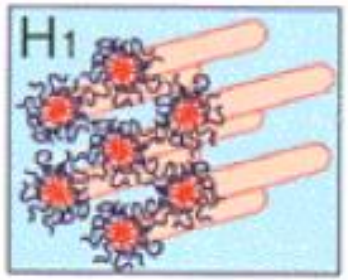
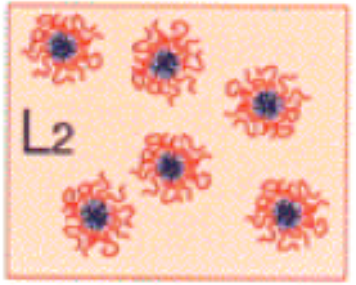


50% oil
0% water
50% poloxamer

Spherical Micelles



Measuring Formulation Microstructure

Composition	0% oil 50% water 50% poloxamer	50% oil 0% water 50% poloxamer
Scattering Data		
Analysis	 **Cylinders**	 **Spherical Micelles**

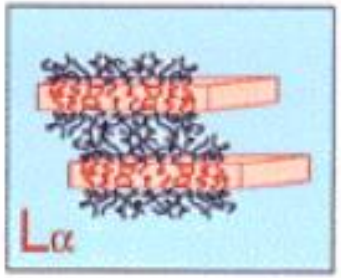
*The two scattering patterns are just different sizes/concentrations of samples with the same composition.

Scattering can be used to identify the structures of the sample.

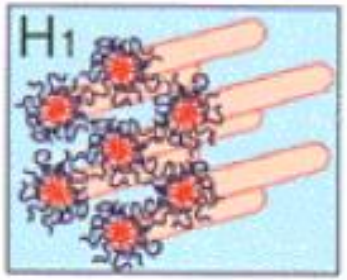
Phase Diagrams – Visualizing Phase Boundaries

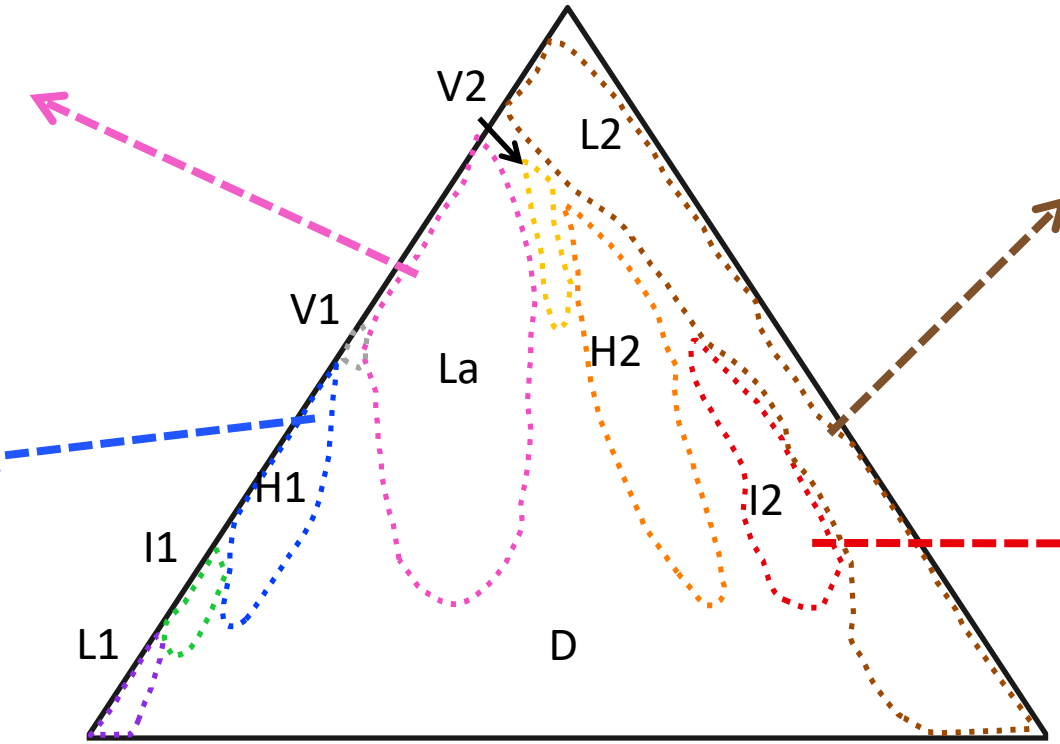
Phase diagrams show how formulation microstructures changes with composition

Lamellae

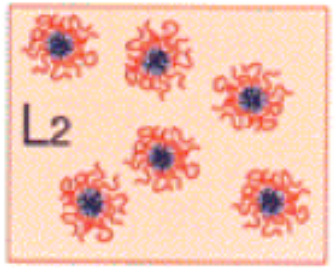


Cylinders

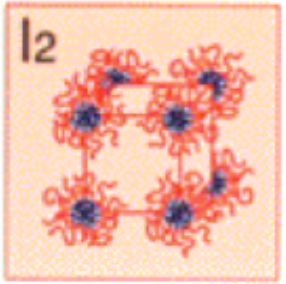




Spherical Micelles



Ordered Micelles



Langmuir 1998, 14, 2627-2368

Why should we care about formulations?

Ingredients of dish soap (~22 components):

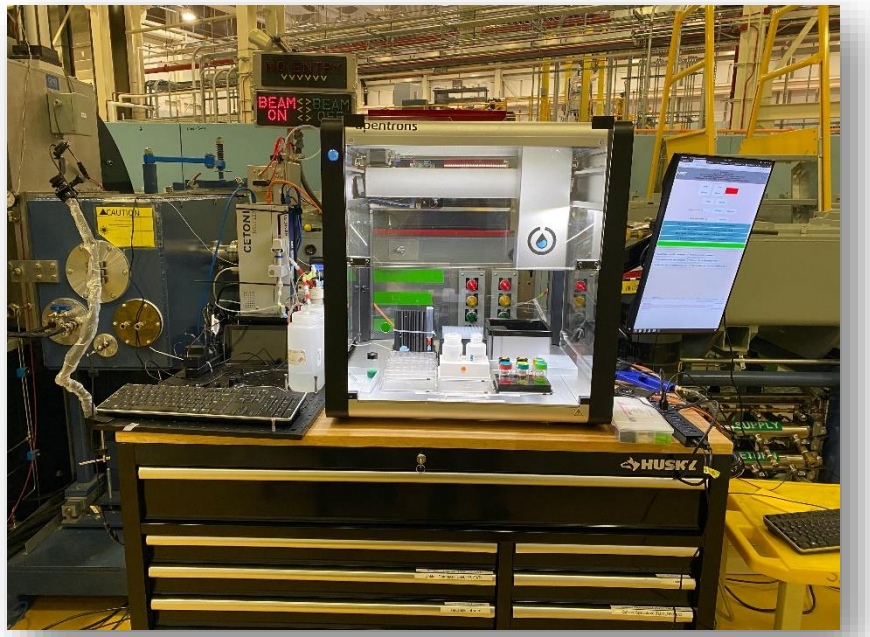
Water, Sodium lauryl sulfate, Sodium laureth sulfate, C10-16 Alkydime-Thylamine oxide, Alcohol denat., PPG-26, Sodium chloride, Sodium hydroxide, PEI-14 PEG-24/PPG-16 copolymer, Phenoxyethanol, Methylisothiazolinone, Colorants, Fragrances, C9-11 Pareth-8, Tetrasodium glutamate, diacetate, Chloroxylenol, Phenoxy-isopropanol, Glycerin, Sodium cumene-sulfonate, Propylene glycol, Terpineol

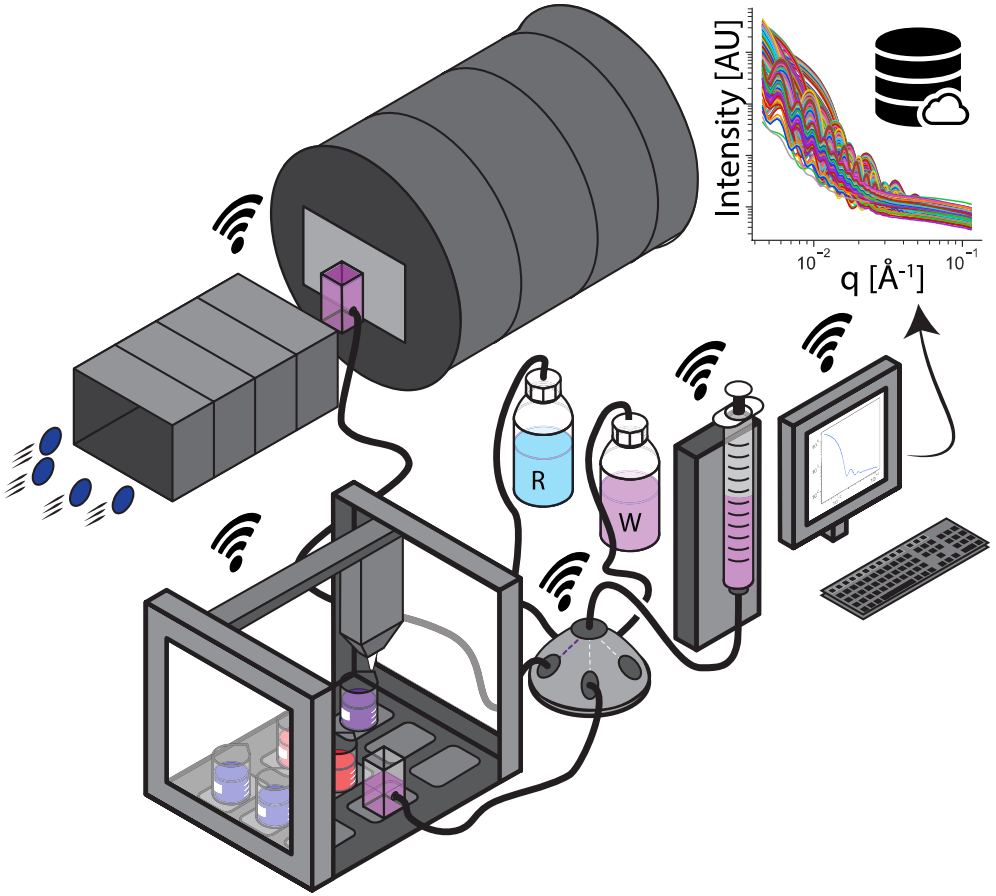
- **Changing regulations and consumer desire are forcing companies to reformulate**
- **Reformulation is expensive and tedious**

dawn-dish.com/en-us/how-to/what-dawn-is-made-of-ingredients/

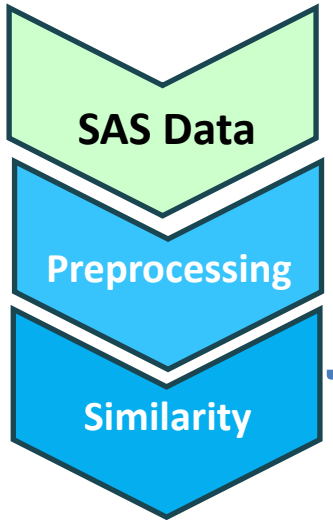
The Autonomous Formulation Lab!

Autonomously conducts x-ray/neutron scattering experiments on liquid compositions.

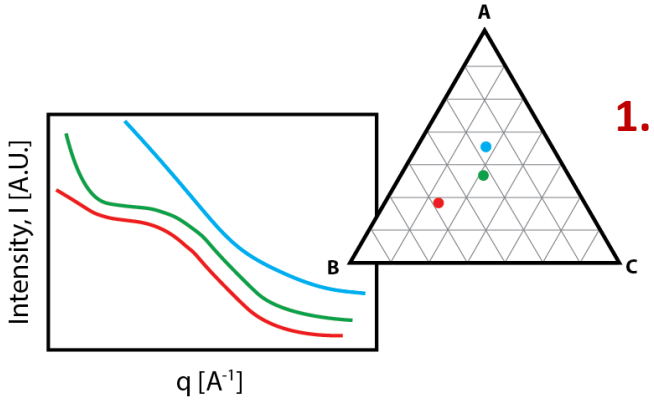




The AI Agent's Pipeline



sample1
sample2
sample3
...



1. Input scattering data

sample1 → new_and_improved_sample1

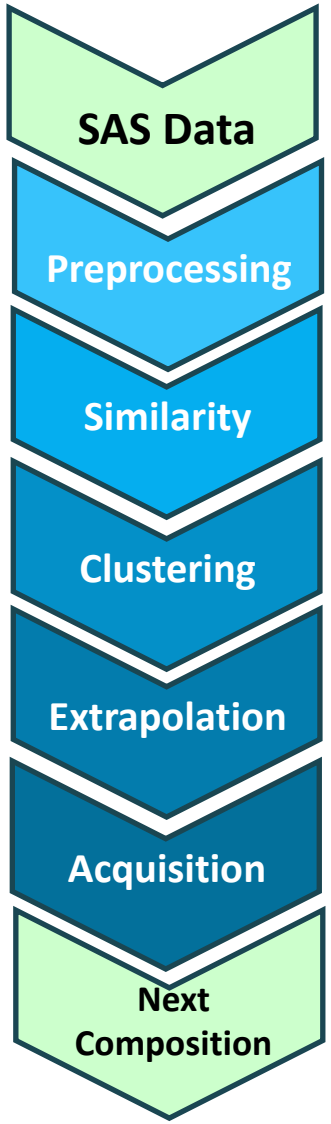
2. Change the data to be optimal for calculations

Example	Sample 1	Sample 2	Sample 3
Sample 1	1.0	0.2	0.3
Sample 2	0.2	1.0	0.8
Sample 3	0.3	0.8	1.0

3. Calculate the similarity matrix

Similarity Scale
Larger number = more similar
(1.0 is the max)

The AI Agent's Pipeline

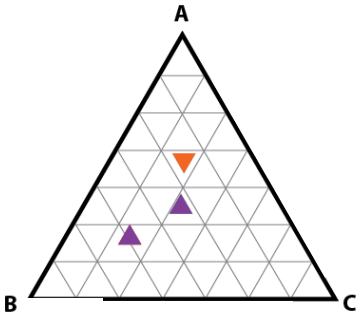


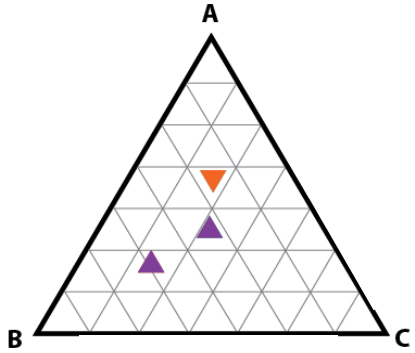
sample1
sample2
sample3
...

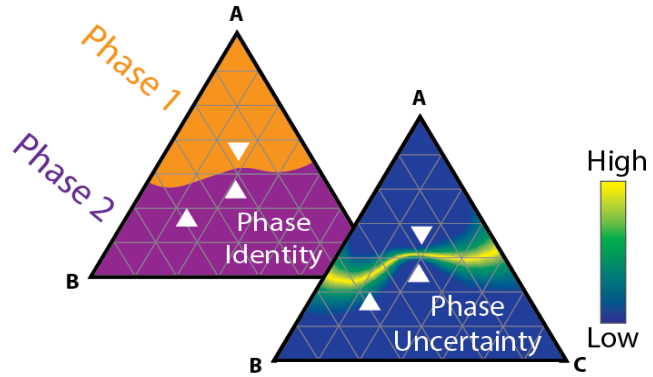
4. Assign the samples to a phase

Phase 1
sample1, ...

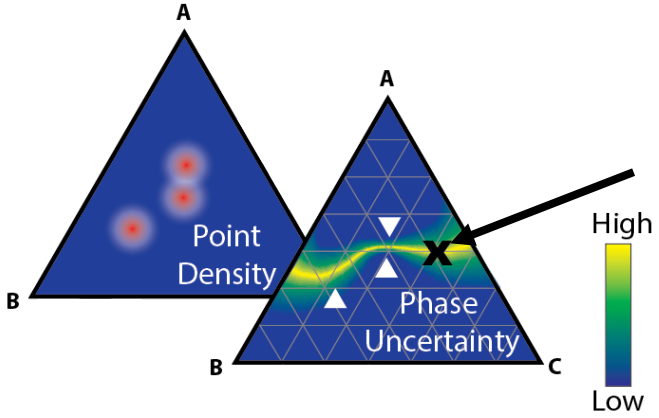
Phase 2
sample2, sample3, ...







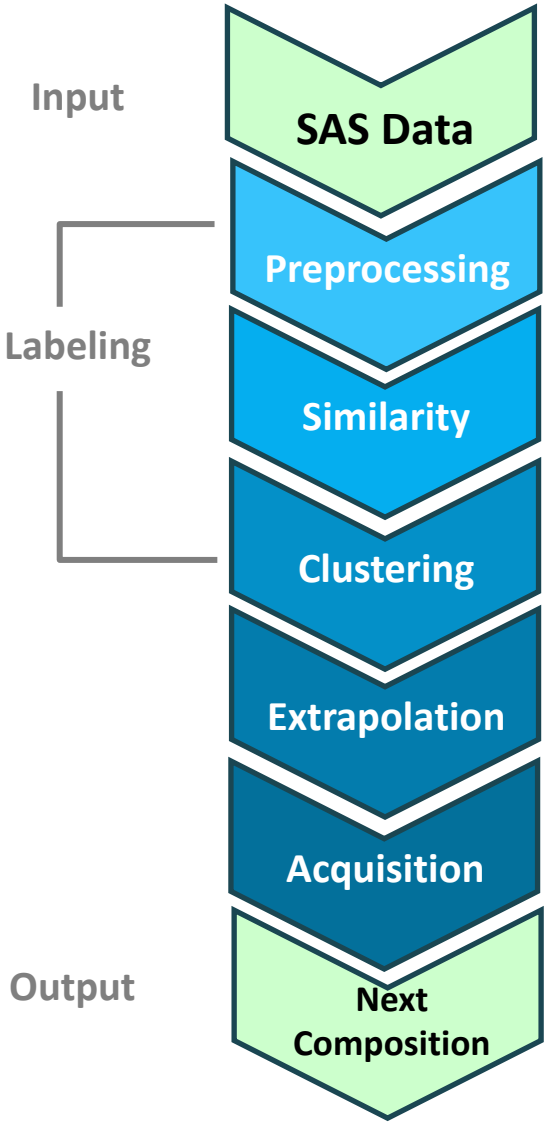
5. Predict the entire phase diagram



Next sample to measure.

6. Acquire the next sample to measure

The AI Agent's Parameters



Coordinate system?

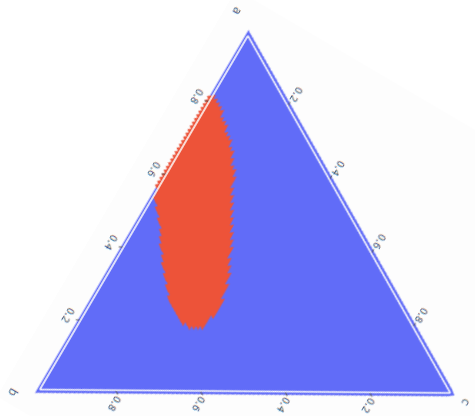
Similarity Input?
Similarity Calculation?

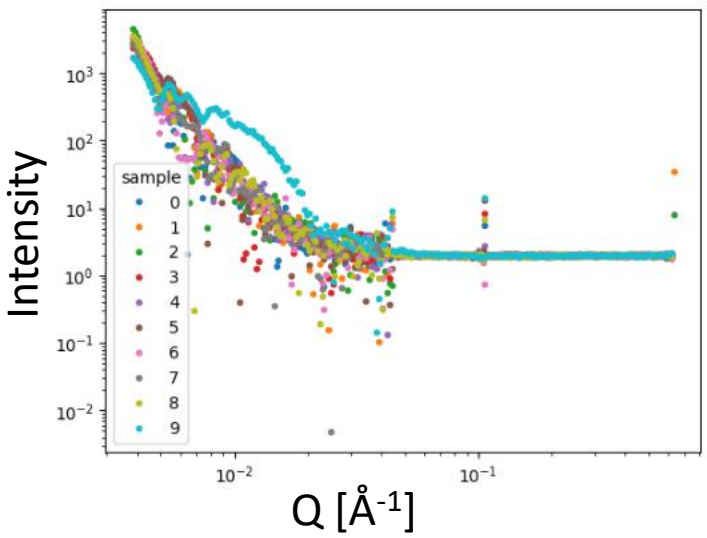
Manipulating uncertainties?

How do we know what the optimal value/algorithm is?

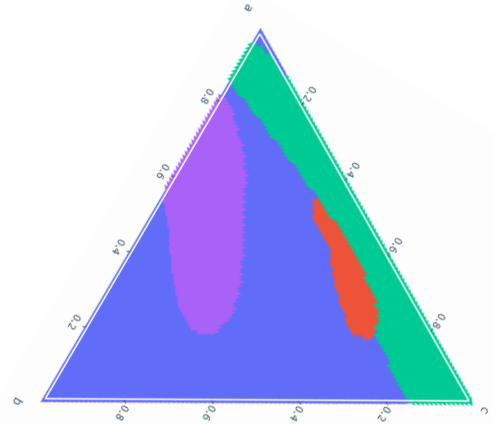
The Challenges

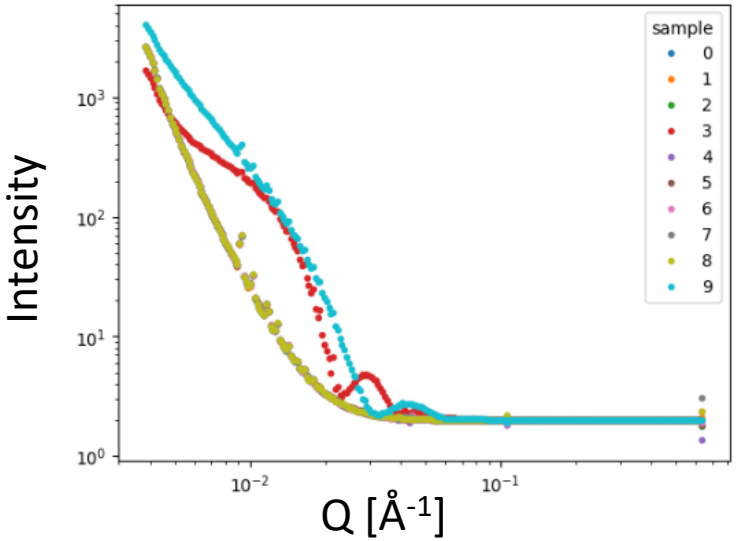
Challenge 1 - Noise



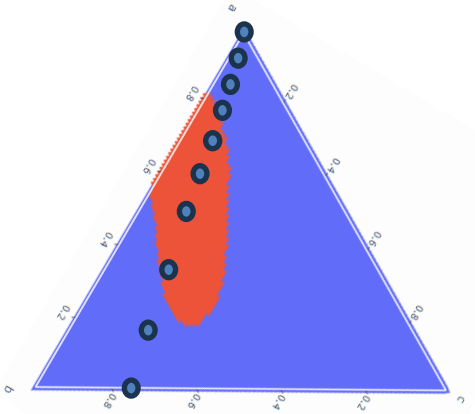


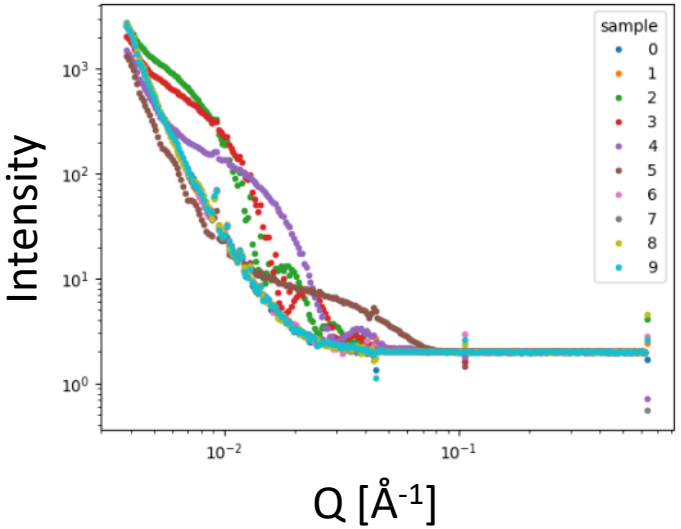
Challenge 2 – Multiple Phases





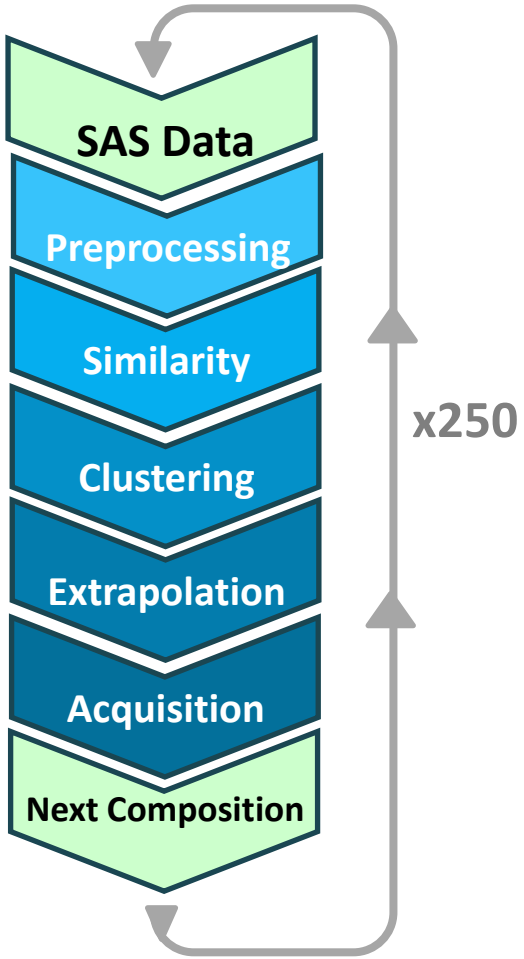
Challenge 3 – Variation within the phase

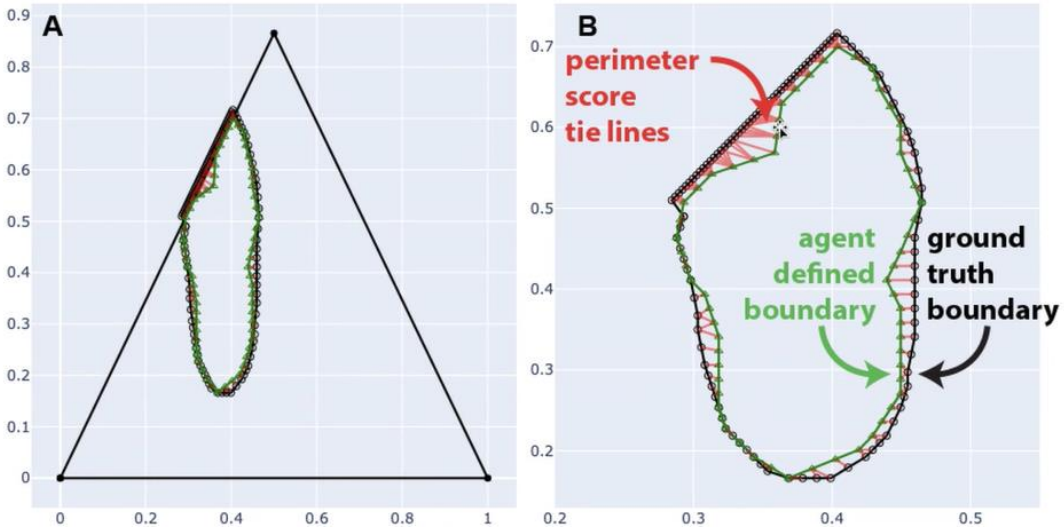




Determining Accuracy

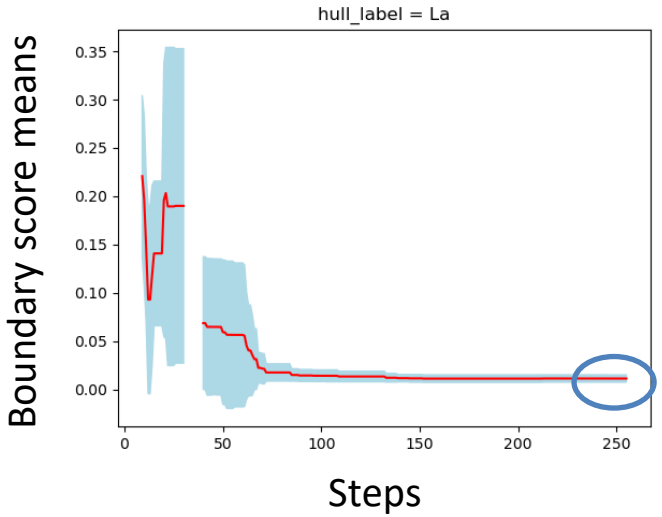
Campaign – 250 steps of the pipeline





Boundary score mean – Average distance between the ground truth and agent defined boundary

Boundary standard deviation – deviation of scores



End Campaign Accuracy – Average of last 10 score means

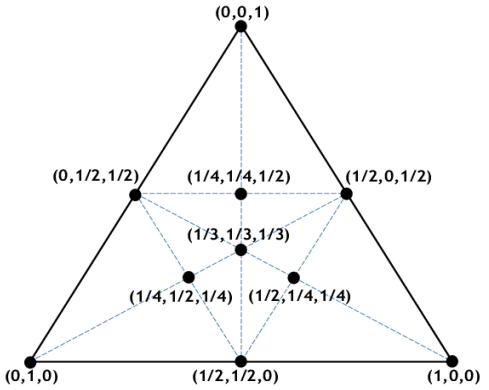
End Campaign Error – Last 10 standard deviations propagated

*In the bar graphs, the values are the average of around 3 campaigns

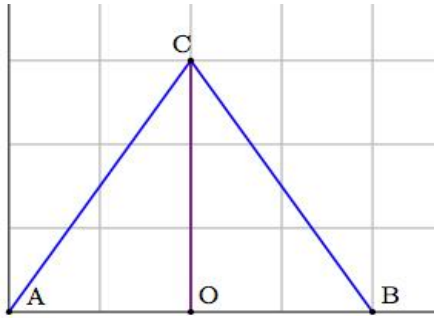
Barycentric vs. Cartesian - Deciding a Coordinate System

Preprocessing

Composition coordinates can be converted into cartesian coordinates.

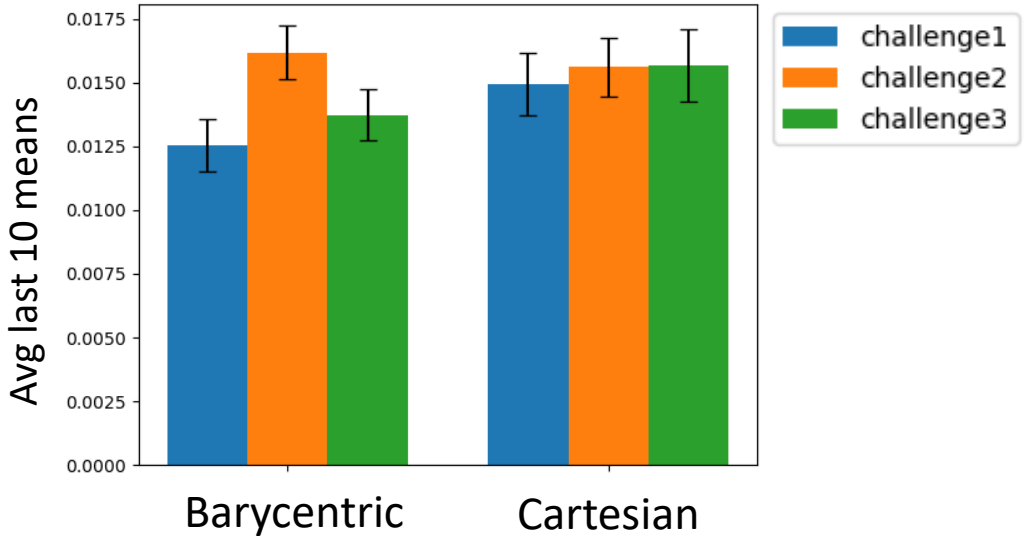


(0, 0.866)
(0, 0.577)
(0, 0.288)
(0, 0)

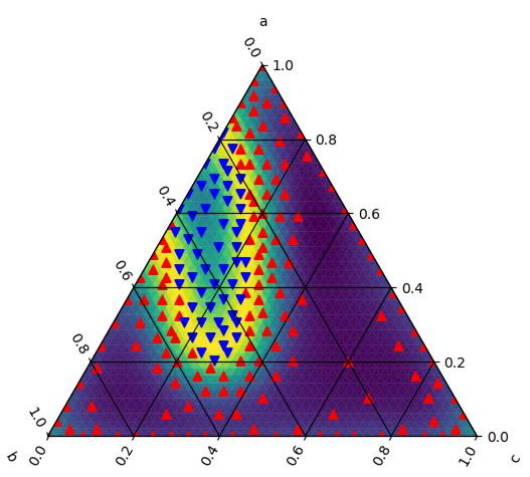


(0.5, 0)
(1, 0)

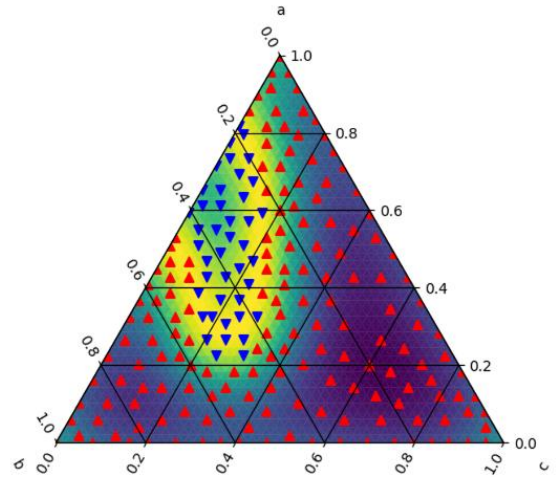
Barycentric is generally better.



Cartesian is more grid-like.



Barycentric



Cartesian

Challenge 1, Step 200

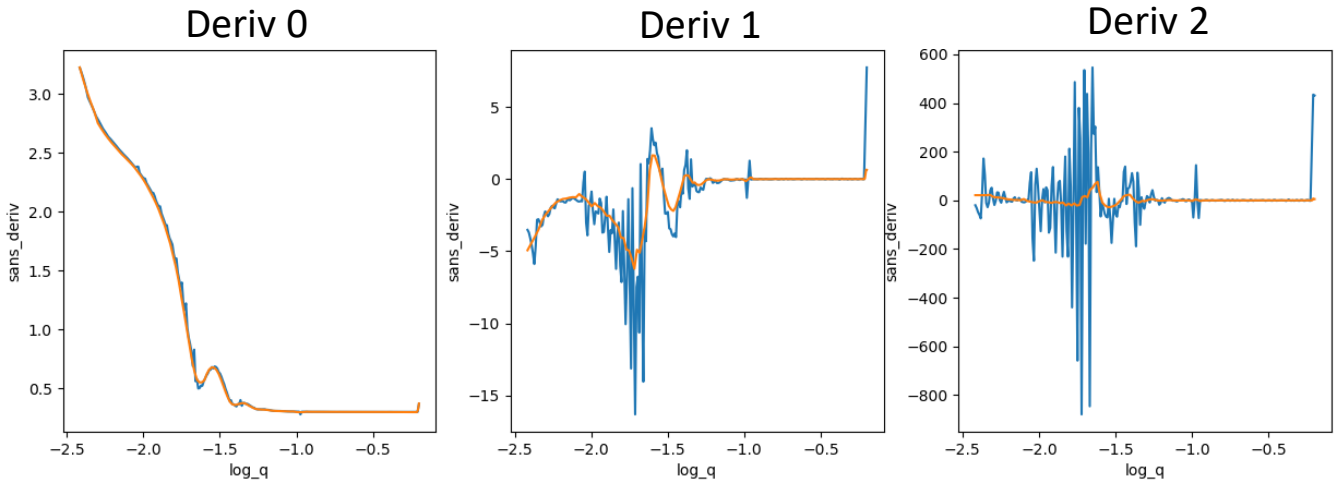
tasks.illustrativemathematics.org/content-standards/HSG/GPE/B/tasks/1687
www.researchgate.net/figure/Barycentric-coordinates-on-an-equilateral-triangle_fig5_264825595

Derivative - Similarity in Shapes

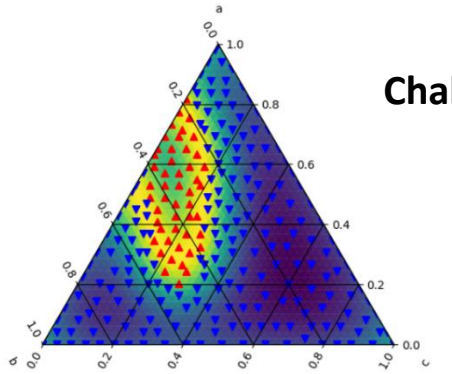
Similarity

The derivative can tell us about the shape of the scattering data.

Derivatives do not work well with noise.

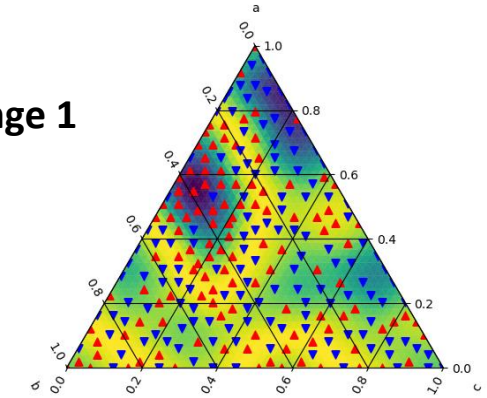


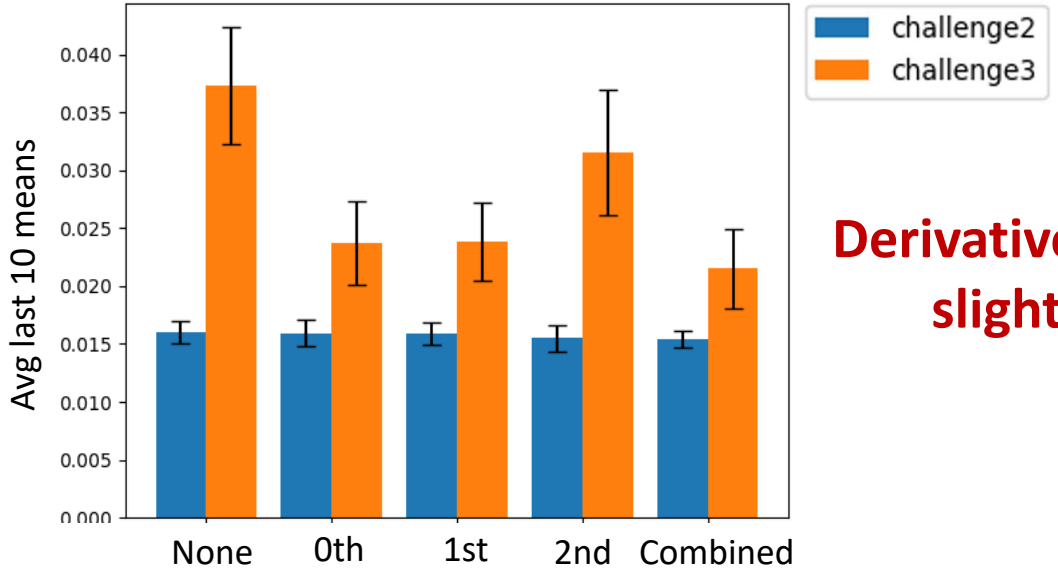
No Derivative



Challenge 1

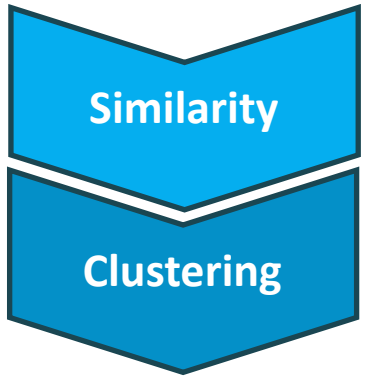
0th Derivative





Derivatives performed slightly better.

Gamma - Tuning the Similarity



Laplacian kernel – Method to calculate similarity

Uses the formula: $\exp(-\gamma\|x-y\|_1)$

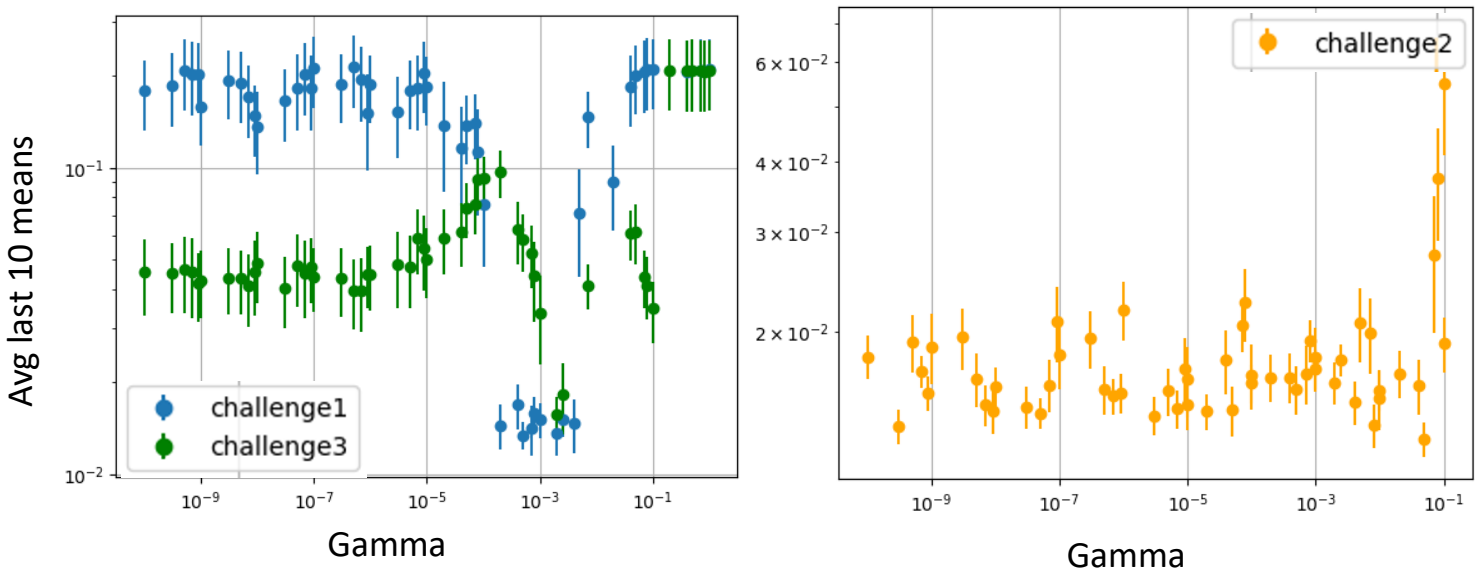
To the power of e

Gamma (a constant)

Difference between samples

Similarity matrix

	s1	s2	s3
s1	1.0	0.2	0.3
s2	0.2	1.0	0.8
s3	0.3	0.8	1.0



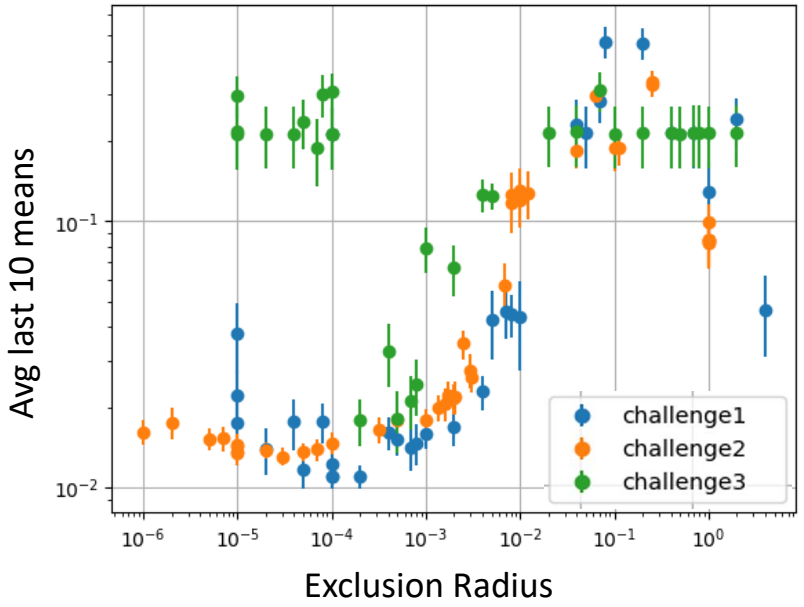
The optimal gamma is around 1e-3.

Exclusion Radius – Distance between points

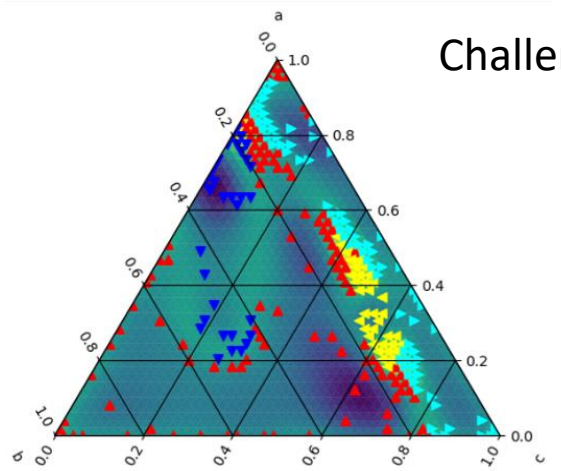
Acquisition

The certainty around each point is artificially increased.

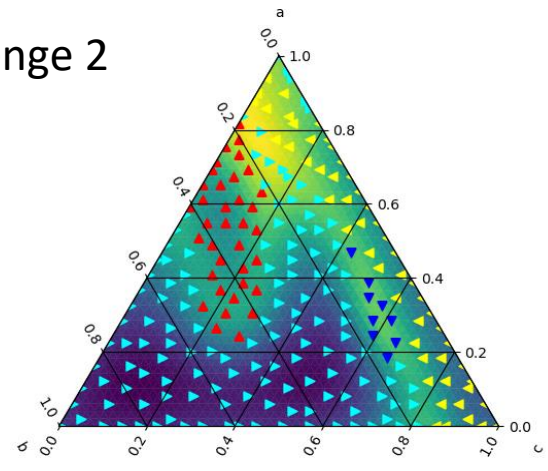
The **exclusion radius** (ER) determines how large that area is.



Challenge 2



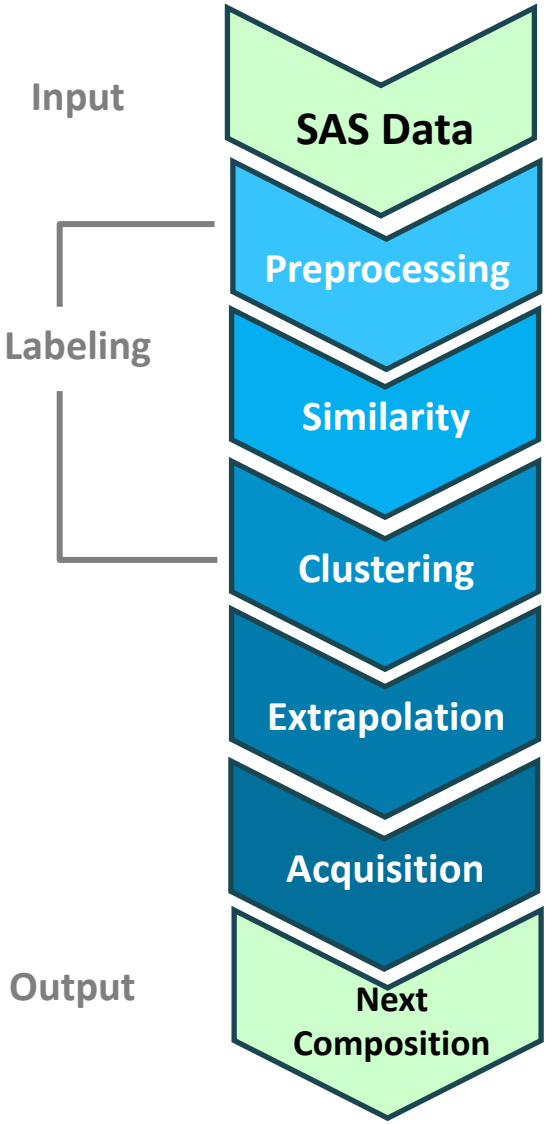
ER = 1e-5



ER = 1e-3

The ideal exclusion radius is between 1e-4 and 8e-2

Summary



Through approx. 1000 virtual campaigns, we found..

Coordinate system? **Barycentric is generally better**

Similarity Input? **Use derivatives**

Similarity Calculation? **Gamma between 1e-4 to 1e-2**

Distance between points? **Exclusion radius between 1e-4 to 8e-2**

Thank you to Dr. Tyler Martin for mentoring me,
thank you to CHRNS for making the SHIP program possible, and
thank you to Dr.'s Julie Borchers, William Kao, and Jessica Staymates for letting me participate in the SHIP program!