3.36pt

Funded by NSF EXTREEMS-QED

Parameterization and Analysis of Viscous Fluid Conduit Edges for Dispersive Hydrodynamics

University of Colorado, Boulder
Advisors: Mark Hoefer, Michelle Maiden

March 5, 2016

Zoe Farmer
www.dataleek.io

Funded by NSF EXTREEMS-QED

CU Boulder Applied Math

Dispersive Hydrodynamics Lab

Our lab has four main facets

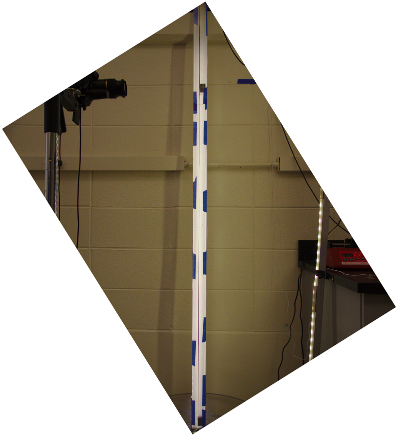
1. Experimental Observations
2. Analytical Solutions
3. Numerical Solutions
4. Data Analysis (part of the NSF EXTREEMS-QED program)

In order to have accurate results, we need to make sure we have accurate data.

Poor quality data yields poor quality results.

High Level Overview

- In our lab we examine the dynamics of a viscous fluid conduit.
- One large issue is that sometimes the conduits that are formed curve, and spiral around a central axis.
- We determine a method to measure the degree of spiraling, and to reject conduits based on this degree.

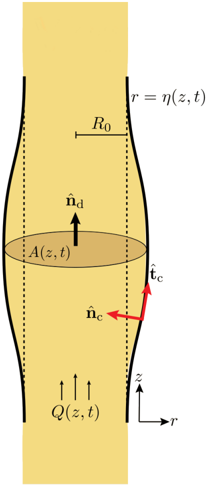


This is a **Data Quality Problem**.

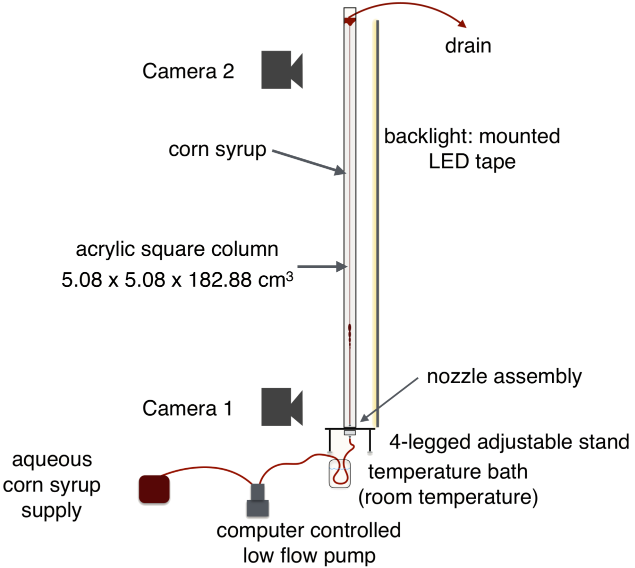
What's our Environment?

Viscous Fluid Conduits

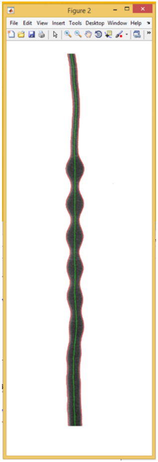
- Two viscous fluids, with inner forming axisymmetric conduit.
- Exterior Fluid: $\rho^{(e)}$ density and $\mu^{(e)}$ viscosity
- Interior Fluid: $\rho^{(i)}$ density and $\mu^{(i)}$ viscosity
- $\rho^{(i)} < \rho^{(e)} \Rightarrow$ buoyant flow
- $\mu^{(i)} << \mu^{(e)} \Rightarrow$ minimal drag
- Re $<< 1 \Rightarrow$ low Reynold's number

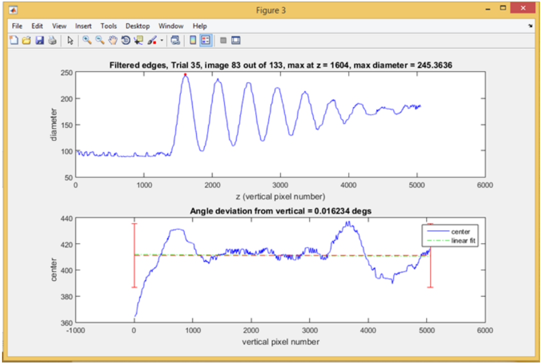


Experimental Setup



Data Collection



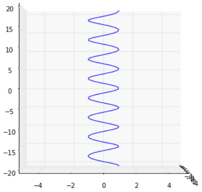


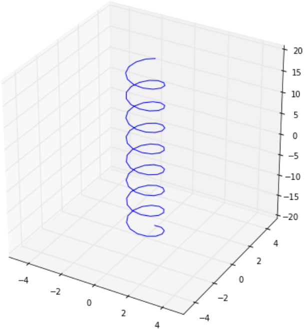
Numeric data results from edge detection on images.

What's the Issue?

- How *straight* is this conduit?
- Can we use data from this?
- Will our results be skewed?
- How do we quantify "straightness"?

Again, this is a **Data Quality Problem**.



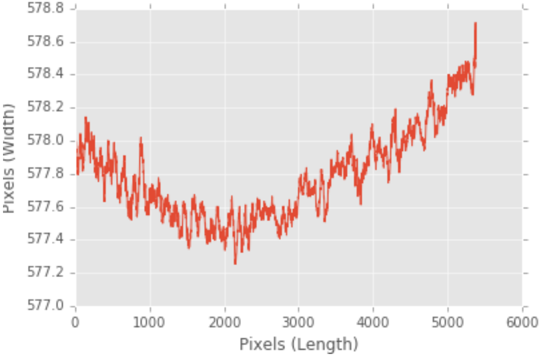


Algorithm

1. Given some single-trial from edge detection on conduit image, find a smoothing fit.
2. Repeat for all trials.
3. Plot the *Curvature* and *Average Residual* of each smoothed trial fit on the plane.
4. Use Classification Algorithms to determine the space of non-curvy lines, i.e. the space of good quality data.

General Fit

By averaging the left and right edges of our axisymmetric conduit we can find the centerline.

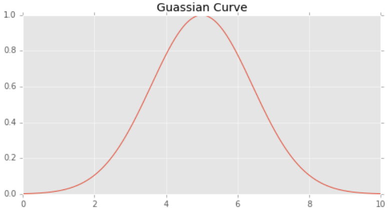


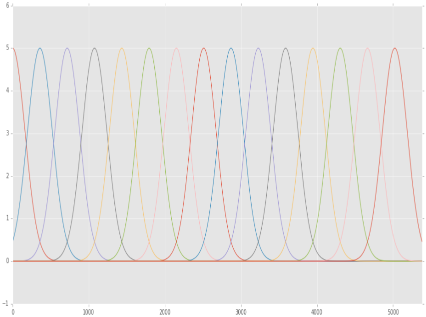
Regression Model

We use the regression model

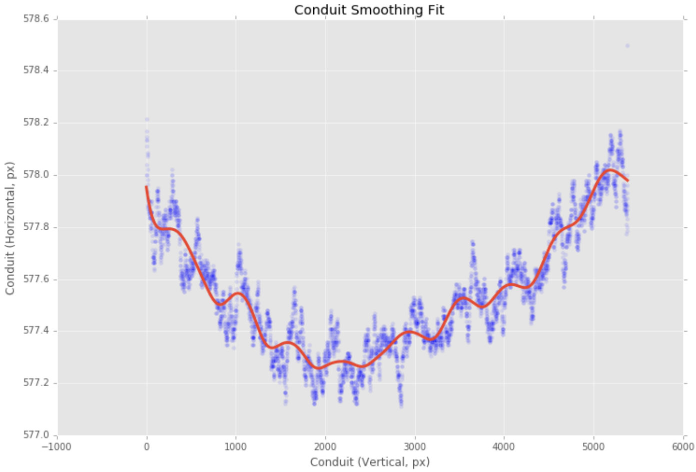
$$h(x) = \beta_1 + \beta_2 x + \sum_{i=3}^{n} \beta_i f(x - x_i)$$

where $f(y) = \beta e^{-(sy)^2}$.





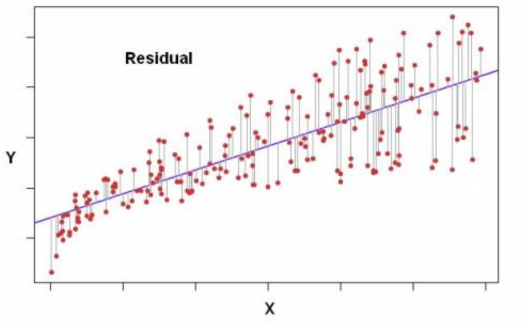
Example Fit



Conduit Smoothing Fit
578.6
578.4
578.2
578.0
577.8
577.6
577.4
577.2
577.0
Conduit (Horizontal, px)
−1000
0
1000
2000
3000
4000
5000
6000
Conduit (Vertical, px)

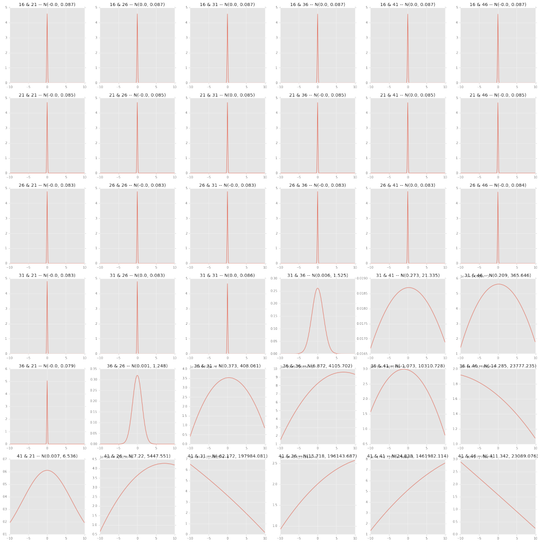
Fit Parameter Determination

- The shape parameter is how wide each gaussian is.
- The count parameter is how many gaussians we use in the regression.
- To find our parameters we examine every fit's residuals, take their mean, and the find the standard deviation of that set. Our ideal parameters have low standard deviation.

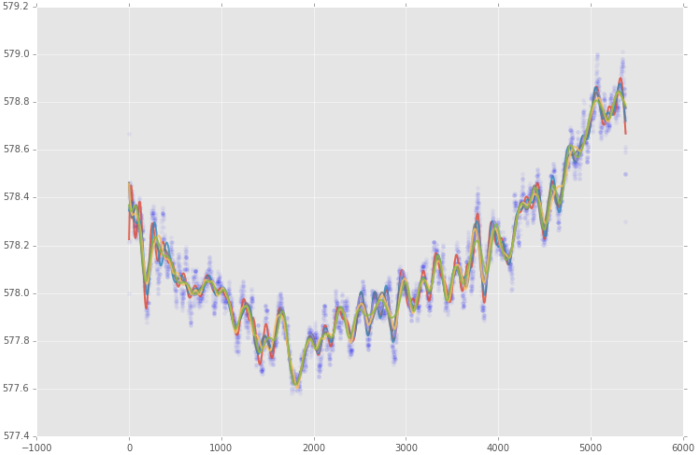


- Examined using available data from the lab.

Sample Parameter Grid



Fit Comparison



Conduit Metrics

Now that we have a way to map our scattered points to a smooth line, we measure the quality of our collected data.

- Curvature - Based on second derivative
- Noise - Based on residuals. Included as if this is too high, it means the fit was poor, which can either indicate a blurry photo, or extreme curves.

Curvature

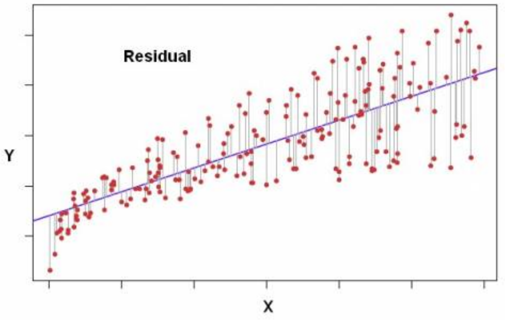
For this application, defined as

$$\frac{1}{n \cdot \text{scale}} \cdot \sqrt{\int_a^b \left(s''(x)\right)^2 dx}$$

which is a non-dimensionalized number based on pixels and scaling factors.

Noise

Residuals are defined as $y - \hat{y}$.



For this application, we define the noise as

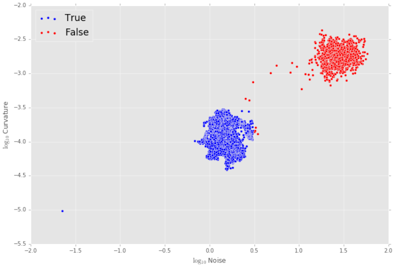
$$\overline{|y - \hat{y}|}.$$

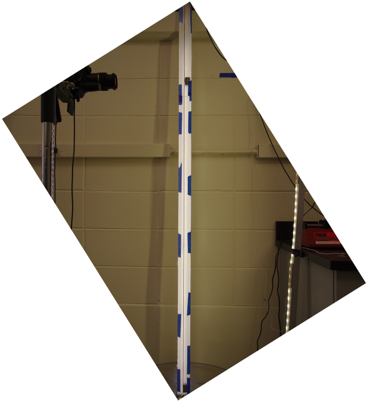
Plot

True
False
$\log_{10}$ Curvature
$\log_{10}$ Noise
−2.0
−2.5
−3.0
−3.5
−4.0
−4.5
−5.0
−5.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0

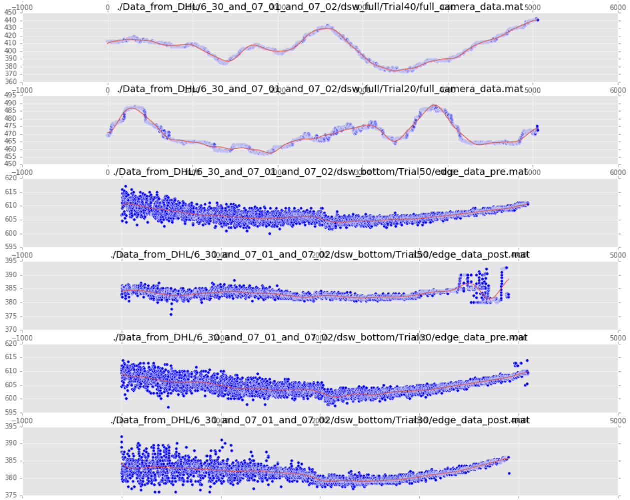
Interpreting our Plot

- How do we map this to our conduit?
- What do the clusters tell us?
- Where do we go from here?





Example Fits

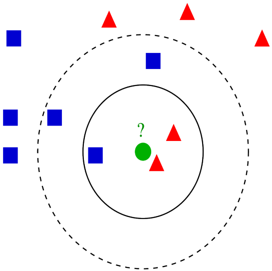


A (Brief) Introduction to Machine Learning

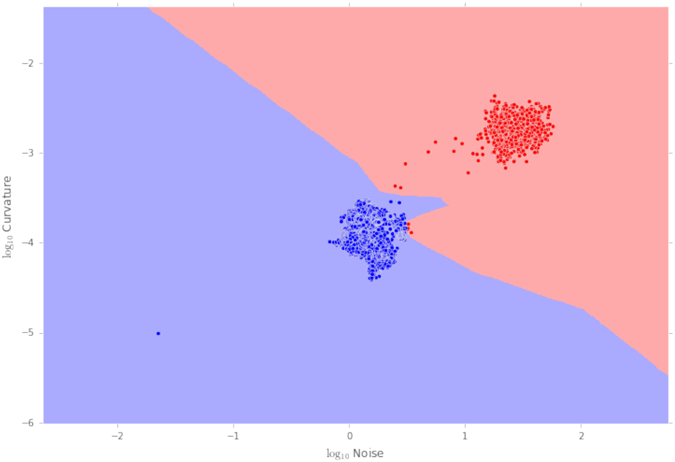
- For this problem, we need to use a subset of Machine Learning referred to as **classification**.
- Classification problems are those in which we need to sort a dataset into different categories.
- This perfectly describes our situation.
- These algorithms require a training and testing dataset.
- Our results will improve over time.

k-Nearest Neighbor Algorithm

k nearby points "vote" on the class of the point in question.

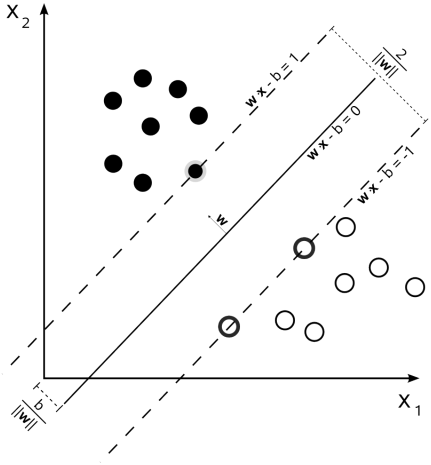


Nearest Neighbor Application

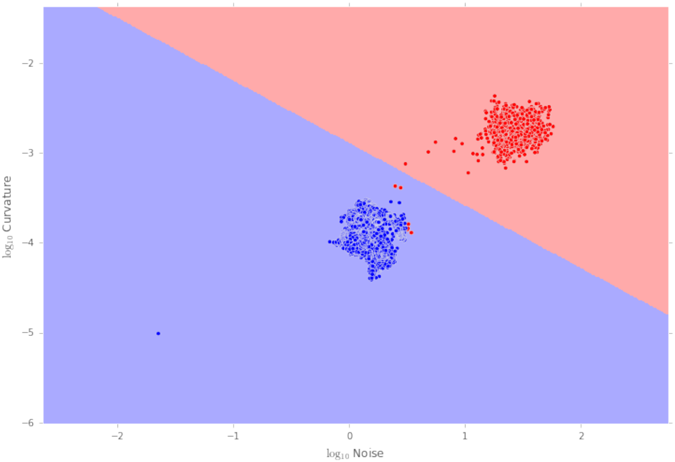


Support Vector Machine Algorithm

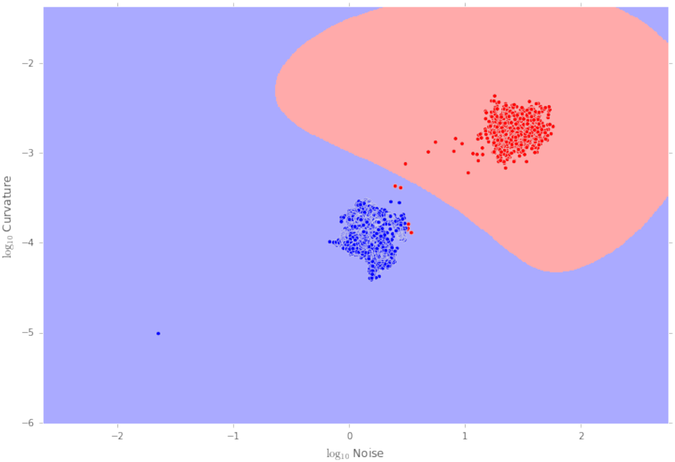
Tries to find the optimal dividing line between classes.



Support Vector Machine with Linear Kernel

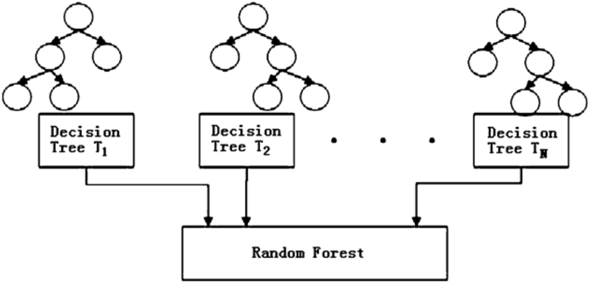


Support Vector Machine with RBF Kernel

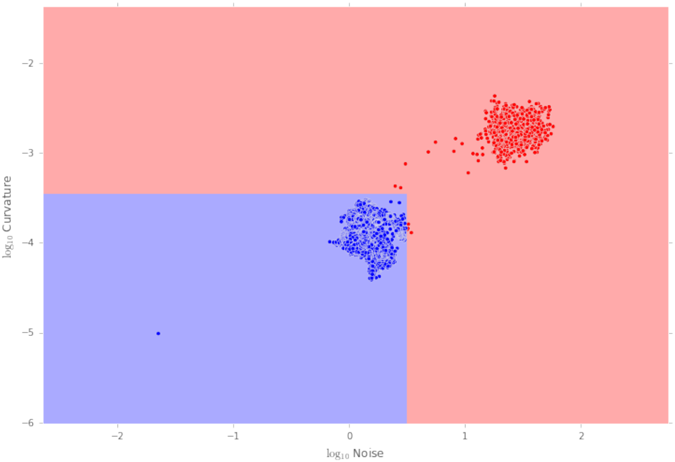


Randomized Forest Algorithm

Creates a "forest" of decision trees and classifies based on mean result.



Randomized Forest Application

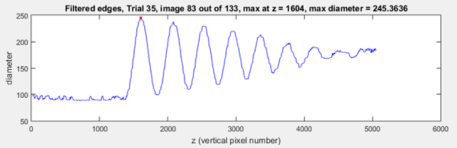


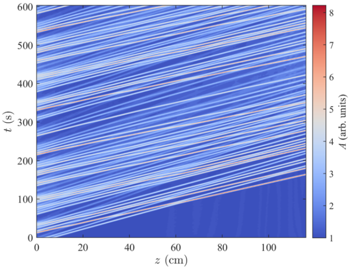
So What?

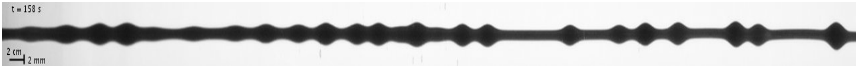
- **Poor quality data yields poor quality results.**
- We now can ensure good quality data.

Moving Forward

- Now that we can ensure good quality data in our analysis we can move forward with some neat projects, like the Soliton Gas.
- A width problem instead of centerline.
- Where exactly is a specific soliton in a gas?
- Poisson?







Acknowledgements

- Mark Hoefer
- Michelle Maiden
- Peter Wills
- Funded by NSF EXTREEMS-QED