Shape Distribution Modelling and Wasserstein Variance Decompositions with Linear Optimal Transport

Michael Wilson

Florida State University

mwilson5@.fsu.edu

May 23, 2024

Approach

Data:

- Y: Images from manufacturing process
- x: scalar input to manufacturing process

Approach:

1. Estimate Shape Mixture Distribution Components: Size-and-shape Mode Based Clustering (using data pooled across frames)
2. Represent each frame as probability vector (based on cluster proportions)
3. Regress probability vectors on scalar inputs

Note: in the next slide, we present the unclustered frames, along with the covariate values in the title. Frames are sorted by covariate value, in ascending order. (One frame is missing from the data)

Unclustered Frames

Size-and-shape Mode-based Clustering

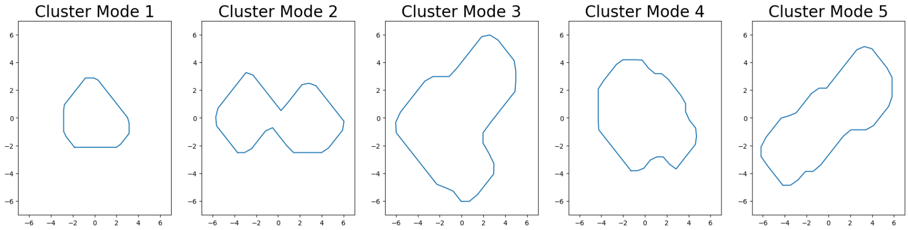


Figure: Size-and-shape modes selected by K-modes Kernel Mixture Clustering

- In the next slide, we plot the frames, with the color of each curve corresponding to cluster (Black curves are considered 'unclustered')

Clustered Frames

Plots of individual Clusters

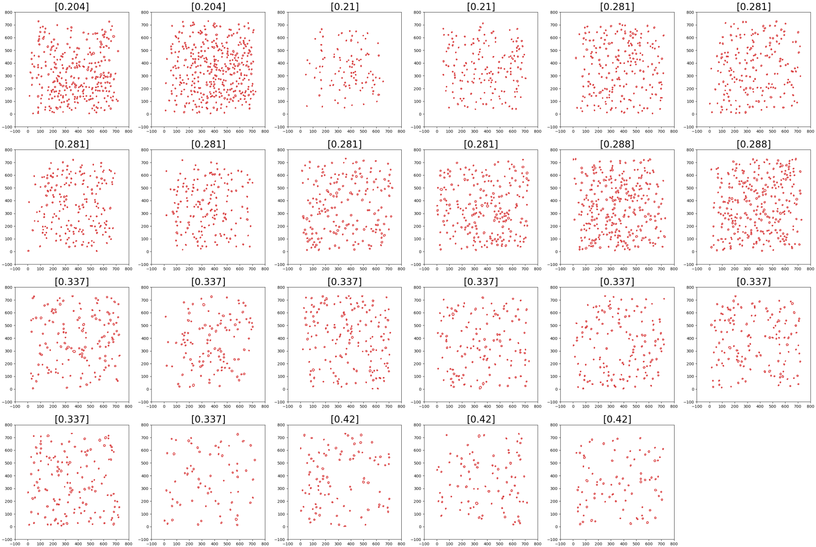
- In the next 5 slides, we plot only one cluster across all frames - a simple way to visualize the relationship between covariate and cluster counts.

Only Cluster 1

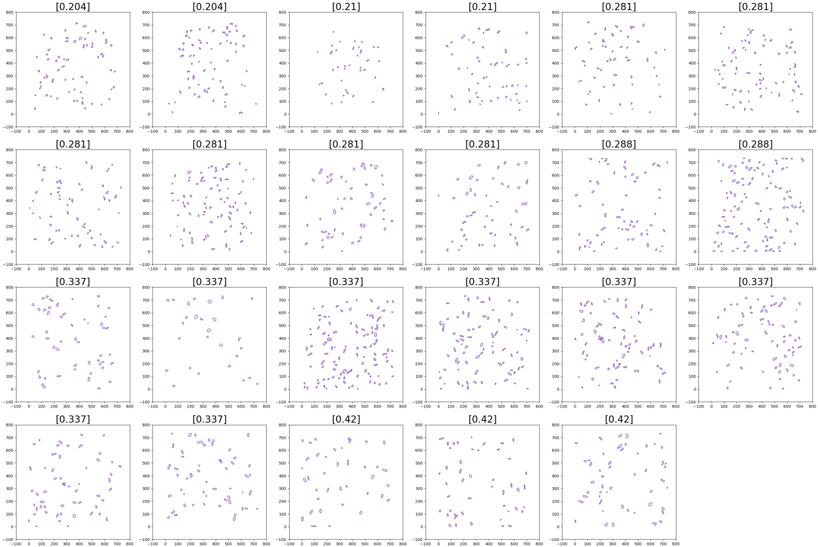
Only Cluster 2

Only Cluster 3

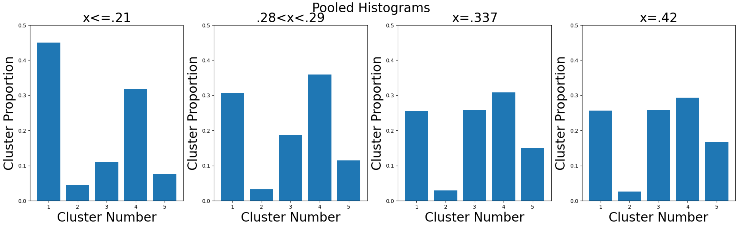
Only Cluster 4



Only Cluster 5

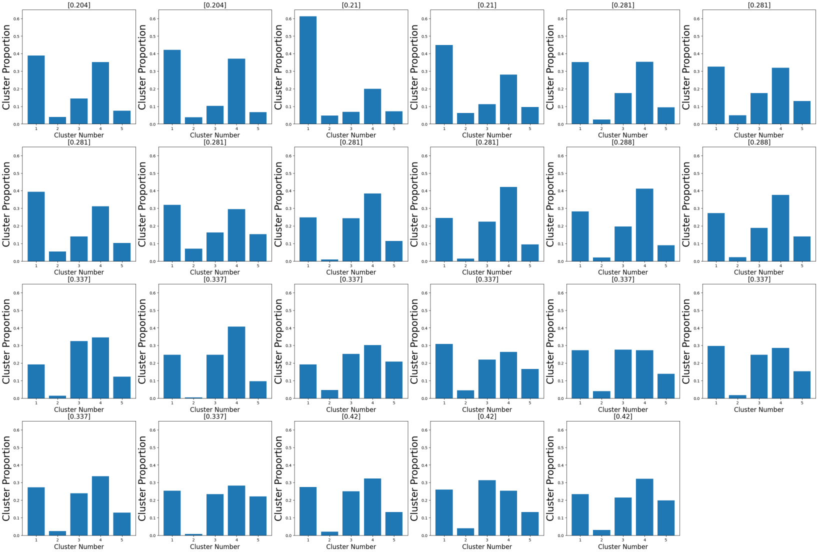


Pooled Histograms

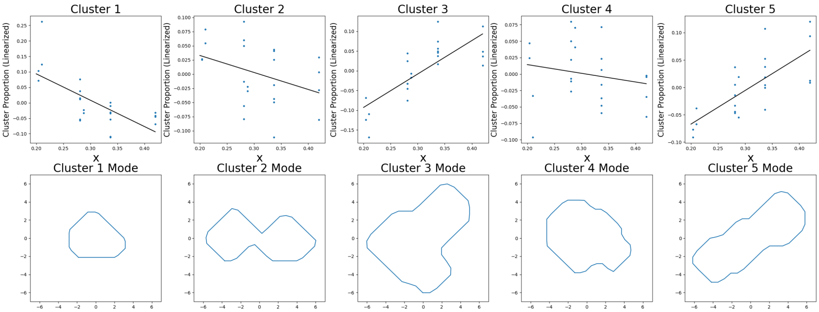


Here we pool frames by covariate, and plot probability vectors for each group of covariates. In the next slide, we plot the probability vectors for each frame individually, with covariate in title.

Histograms for Individual Frames



Scatter Plots



- Here we plot the scatter plots for each cluster proportion, against the covariate values, along with the OLS line of best fit. Corresponding cluster modes are included in the bottom row. (Note, y-values in scatter plots correspond to components of a linearized version of the square-root probability vectors.)
- In the next 5 slides, we present the model summaries for these 5 models.

Model Summary for Cluster 1

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.480
Model:	OLS	**Adj. R-squared:**	0.455
Method:	Least Squares	**F-statistic:**	19.38
Date:	Thu, 23 May 2024	**Prob (F-statistic):**	0.000248
Time:	14:08:29	**Log-Likelihood:**	32.445
No. Observations:	23	**AIC:**	-60.89
Df Residuals:	21	**BIC:**	-58.62
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	0.2639	0.061	4.302	0.000	0.136	0.391
x1	-0.8517	0.193	-4.403	0.000	-1.254	-0.449

Omnibus:	6.280	**Durbin-Watson:**	1.106
Prob(Omnibus):	0.043	**Jarque-Bera (JB):**	4.140
Skew:	0.772	**Prob(JB):**	0.126
Kurtosis:	4.391	**Cond. No.**	16.5

Model Summary for Cluster 2

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.140
Model:	OLS	**Adj. R-squared:**	0.099
Method:	Least Squares	**F-statistic:**	3.405
Date:	Thu, 23 May 2024	**Prob (F-statistic):**	0.0791
Time:	14:15:42	**Log-Likelihood:**	36.450
No. Observations:	23	**AIC:**	-68.90
Df Residuals:	21	**BIC:**	-66.63
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0929	0.052	1.803	0.086	-0.014	0.200
x1	-0.3000	0.163	-1.845	0.079	-0.638	0.038

Omnibus:	1.207	**Durbin-Watson:**	1.491
Prob(Omnibus):	0.547	**Jarque-Bera (JB):**	1.012
Skew:	-0.298	**Prob(JB):**	0.603
Kurtosis:	2.163	**Cond. No.**	16.5

Model Summary for Cluster 3

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.607
Model:	OLS	**Adj. R-squared:**	0.589
Method:	Least Squares	**F-statistic:**	32.47
Date:	Thu, 23 May 2024	**Prob (F-statistic):**	1.18e-05
Time:	14:11:34	**Log-Likelihood:**	38.496
No. Observations:	23	**AIC:**	-72.99
Df Residuals:	21	**BIC:**	-70.72
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	-0.2625	0.047	-5.568	0.000	-0.361	-0.164
x1	0.8474	0.149	5.698	0.000	0.538	1.157

Omnibus:	0.027	**Durbin-Watson:**	1.322
Prob(Omnibus):	0.987	**Jarque-Bera (JB):**	0.144
Skew:	0.065	**Prob(JB):**	0.930
Kurtosis:	2.635	**Cond. No.**	16.5

Model Summary for Cluster 4

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.037
Model:	OLS	**Adj. R-squared:**	-0.009
Method:	Least Squares	**F-statistic:**	0.7993
Date:	Thu, 23 May 2024	**Prob (F-statistic):**	0.381
Time:	14:11:51	**Log-Likelihood:**	38.617
No. Observations:	23	**AIC:**	-73.23
Df Residuals:	21	**BIC:**	-70.96
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	0.0408	0.047	0.870	0.394	-0.057	0.138
x1	-0.1323	0.148	-0.894	0.381	-0.440	0.175

Omnibus:	0.317	**Durbin-Watson:**	1.305
Prob(Omnibus):	0.853	**Jarque-Bera (JB):**	0.228
Skew:	-0.213	**Prob(JB):**	0.892
Kurtosis:	2.764	**Cond. No.**	16.5

Model Summary for Cluster 5

OLS Regression Results

Dep. Variable:	y	**R-squared:**	0.537
Model:	OLS	**Adj. R-squared:**	0.515
Method:	Least Squares	**F-statistic:**	24.36
Date:	Thu, 23 May 2024	**Prob (F-statistic):**	6.98e-05
Time:	14:12:06	**Log-Likelihood:**	42.612
No. Observations:	23	**AIC:**	-81.22
Df Residuals:	21	**BIC:**	-78.95
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	-0.1899	0.039	-4.817	0.000	-0.272	-0.108
x1	0.6137	0.124	4.935	0.000	0.355	0.872

Omnibus:	0.688	**Durbin-Watson:**	2.330
Prob(Omnibus):	0.709	**Jarque-Bera (JB):**	0.633
Skew:	0.355	**Prob(JB):**	0.729
Kurtosis:	2.606	**Cond. No.**	16.5