

---

# Contents

---

<b>1</b>	<b>What is OODA?</b>	<b>1</b>
1.1	Case Study: Curves as Data Objects	3
1.2	Case Study: Shapes as Data Objects	10
1.2.1	The Segmentation Challenge	10
1.2.2	General Shape Representations	12
1.2.3	Skeletal Shape Representations	14
1.2.4	Bayes Segmentation via Principal Geodesic Analysis	15
<b>2</b>	<b>Breadth of OODA</b>	<b>19</b>
2.1	Amplitude and Phase Data Objects	19
2.2	Tree Structured Data Objects	24
2.3	Sounds as Data Objects	26
2.4	Images as Data Objects	29
<b>3</b>	<b>Data Object Definition</b>	<b>31</b>
3.1	OODA Foundations	31
3.1.1	OODA terminology	31
3.1.2	Object and Feature Space Example	32
3.1.3	Scree Plots	36
3.1.4	Formalization of Modes of Variation	39
3.2	Mathematical Notation	39
3.3	Overview of Object and Feature Spaces	40
3.3.1	Example: Probability Distributions as Data Objects	43
<b>4</b>	<b>Exploratory and Confirmatory Analyses</b>	<b>49</b>
4.1	Exploratory Analysis - Discover Structure in Data	49
4.1.1	Example: Tilted Parabolas FDA	50
4.1.2	Example: Twin Arches FDA	55
4.1.3	Case Study: Lung Cancer Data	58
4.1.4	Case Study: Pan Cancer Data	63
4.2	Confirmatory Analysis - Is It Really There?	67
4.3	Further Major Statistical Tasks	72
<b>5</b>	<b>OODA Preprocessing</b>	<b>75</b>
5.1	Visualization of Marginal Distributions	75
5.1.1	Case Study: Spanish Mortality Data	76

5.1.2	Case Study: Drug Discovery Data	79
5.2	Standardization - Appropriate Linear Scaling	91
5.2.1	Example: Two Scale Curve Data	91
5.2.2	Overview of Standardization	95
5.3	Transformation - Appropriate Nonlinear Scaling	97
5.4	Registration - Appropriate Alignment	100
<b>6</b>	<b>Data Visualization</b>	<b>103</b>
6.1	Heat-Map Views of Data Matrices	103
6.2	Curve Views of Matrices and Modes of Variation	111
6.3	Data Centering and Combined Views	114
6.4	Scatterplot Matrix Views of Scores	125
6.5	Alternatives to PCA Directions	129
<b>7</b>	<b>Distance Based Methods</b>	<b>133</b>
7.1	Fréchet Centers In Metric Spaces	135
7.2	Multi-Dimensional Scaling For Object Representation	140
7.3	Important Distance Examples	144
7.3.1	Conventional Norms	144
7.3.2	Wasserstein Distances	145
7.3.3	Procrustes Distances	147
7.3.4	Generalized Procrustes Analysis	150
7.3.5	Covariance Matrix Distances	152
<b>8</b>	<b>Manifold Data Analysis</b>	<b>157</b>
8.1	Directional Data	157
8.2	Introduction to Shape Manifolds	159
8.3	Statistical Analysis of Shapes	161
8.4	Landmark Shapes	167
8.4.1	Shape Tangent Space	170
8.4.2	Case Study: Digit 3 Data	171
8.4.3	Case Study: DNA Molecule Data	173
8.4.4	Principal Nested Shape Spaces	175
8.4.5	Size-and-shape space	177
8.4.6	Further methodology	178
8.5	Central Limit Theory on Manifolds	178
8.6	Backwards PCA	180
8.7	Covariance Matrices as Data Objects	183
<b>9</b>	<b>FDA Curve Registration</b>	<b>187</b>
9.1	Fisher-Rao Curve Registration	188
9.1.1	Example: Shifted Betas Data	188
9.1.2	Introduction to Warping Functions	194
9.1.3	Fisher-Rao Mathematics	195
9.2	Principal Nested Spheres Decomposition	207

<b>10 Graph Structured Data Objects</b>	<b>211</b>
10.1 Arterial Trees as Data Objects	212
10.1.1 Combinatoric Approaches	212
10.1.2 Phylogenetics	213
10.1.3 Dyck Path	216
10.1.4 Persistent Homology	217
10.1.5 Comparison of Tree Analysis Methods	221
10.2 Networks as Data Objects	222
10.2.1 Graph Laplacians	222
10.2.2 Example: A Tale of Two Cities	224
10.2.3 Extrinsic and Intrinsic Analysis	226
10.2.4 Case Study: Corpus Linguistics	226
10.2.5 Labeled versus unlabeled nodes	228
 <b>11 Classification - Supervised Learning</b>	 <b>231</b>
11.1 Classical Methods	233
11.2 Kernel Methods	242
11.3 Support Vector Machines	249
11.4 Distance Weighted Discrimination	253
11.5 Other Classification Approaches	258
 <b>12 Clustering - Unsupervised Learning</b>	 <b>261</b>
12.1 K-Means Clustering	262
12.2 Hierarchical Clustering	265
12.3 Visualization Based Methods	274
12.3.1 Hybrid Clustering Methods	276
 <b>13 High Dimensional Inference</b>	 <b>277</b>
13.1 DiProPerm - Two Sample Testing	277
13.2 Statistical Significance in Clustering	283
13.2.1 High Dimensional SigClust	288
 <b>14 High Dimensional Asymptotics</b>	 <b>297</b>
14.1 Random Matrix Theory	298
14.2 High Dimension Low Sample Size	303
14.3 High Dimension Medium Sample Size	312
 <b>15 Smoothing and SiZer</b>	 <b>315</b>
15.1 Why not Histograms? - Hidalgo Stamps Data	316
15.2 Smoothing Basics - Bralower Fossils Data	322
15.3 Smoothing Parameter Selection	324
15.4 Statistical Inference and SiZer	326
15.4.1 Case Study: British Family Incomes Data	327
15.4.2 Case Study: Bralower Fossils Data	330
15.4.3 Case Study: Mass Flux Data	331

15.4.4 Case Study: Kidney Cancer Data	332
15.4.5 Additional SiZer Applications and Variants	333
<b>16 Robust Methods</b>	<b>335</b>
16.1 Robustness Controversies	336
16.2 Robust Methods for OODA	337
16.2.1 Case Study: Cornea Curvature Data	344
16.2.2 Case Study: Genome Wide Association Data	349
16.3 Other Robustness Areas	353
<b>17 PCA Details and Variants</b>	<b>355</b>
17.1 Viewpoints of PCA	356
17.1.1 Data Centering	358
17.1.2 Singular Value Decomposition	366
17.1.3 Gaussian Likelihood View	373
17.1.4 PCA Computational Issues	373
17.2 Two Block Decompositions	374
17.2.1 Partial Least Squares	375
17.2.2 Canonical Correlations	378
17.2.3 Joint and Individual Variation Explained	383
<b>18 OODA Context and Related Areas</b>	<b>385</b>
18.1 History and Terminology	385
18.2 OODA Analogy with Object-Oriented Programming	386
18.3 Compositional Data Analysis	388
18.4 Symbolic Data Analysis	389
18.5 Other Research Areas	391
<b>Bibliography</b>	<b>395</b>
<b>Index</b>	<b>441</b>