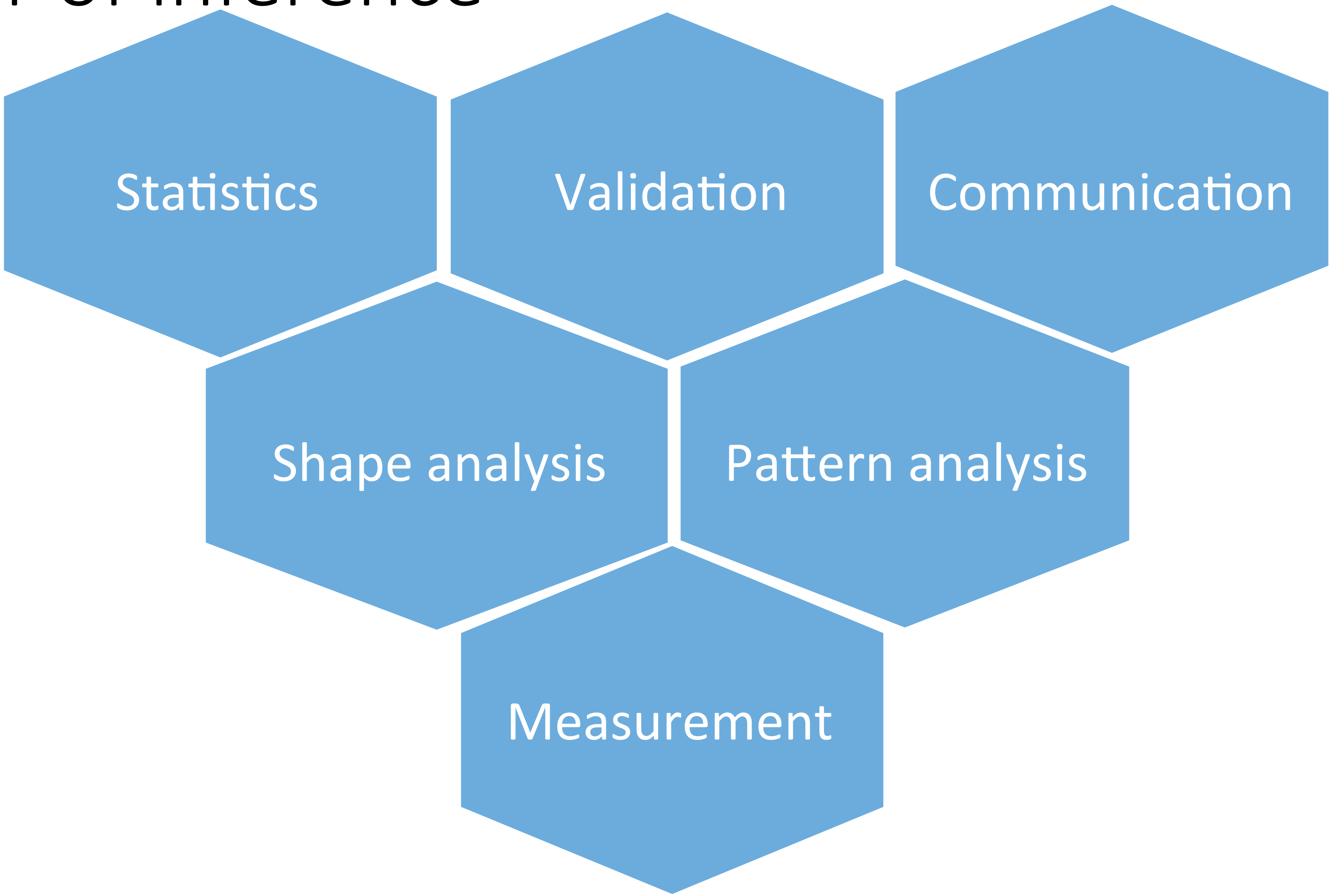
Error propagation in shape analysis

Scott Ferson
Applied Biomathematics

"If a man will begin with certainties, he shall end in doubts; but if he will be content to begin with doubts, he shall end in certainties."

Francis Bacon, *The Advancement Of Learning*

Ladder of inference



Measurement is the base of the ladder

- Extracting quantitative statements about evidence
- Direct or indirect (as of photographs)
  - Image capture
  - Image enhancement ("cleaning")
  - Quantitative image analysis
- Characterizes measurement uncertainty (imprecision)
  - Associated with the protocol (how, who, cleaning, etc.)
  - If absent, must be conservatively characterized post hoc

Without uncertainty statement, not a scientific measurement

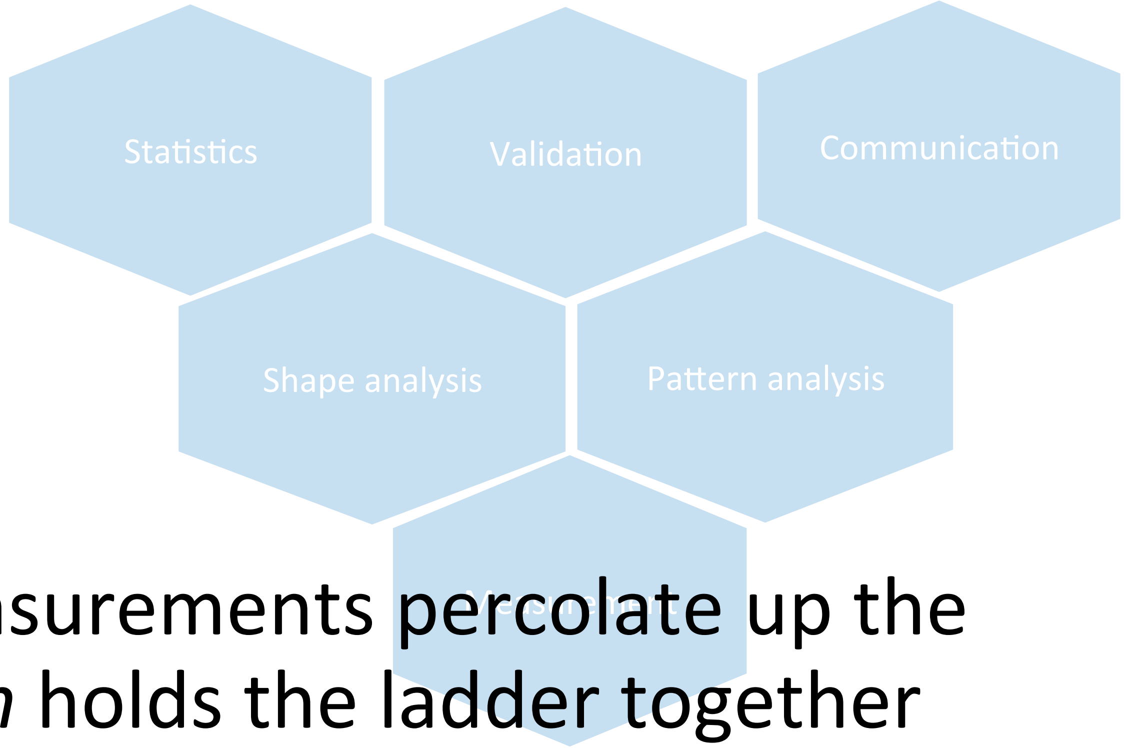
Morphometrics

- Shape analysis
  - Size, shape, and other features of form
  - Landmarks (homologous points)
  - Outlines and contour boundaries
  - 3D trajectories

Hard-object outlines
Signatures
Facial features
Profiles, silhouettes
Shoe or footprints
Tire tread marks

- Pattern analysis
  - Patterns, structure, arrangement, configuration
  - Symmetry, scale, associations, iteration, granularity, detail, texture

Uncertainty percolates up

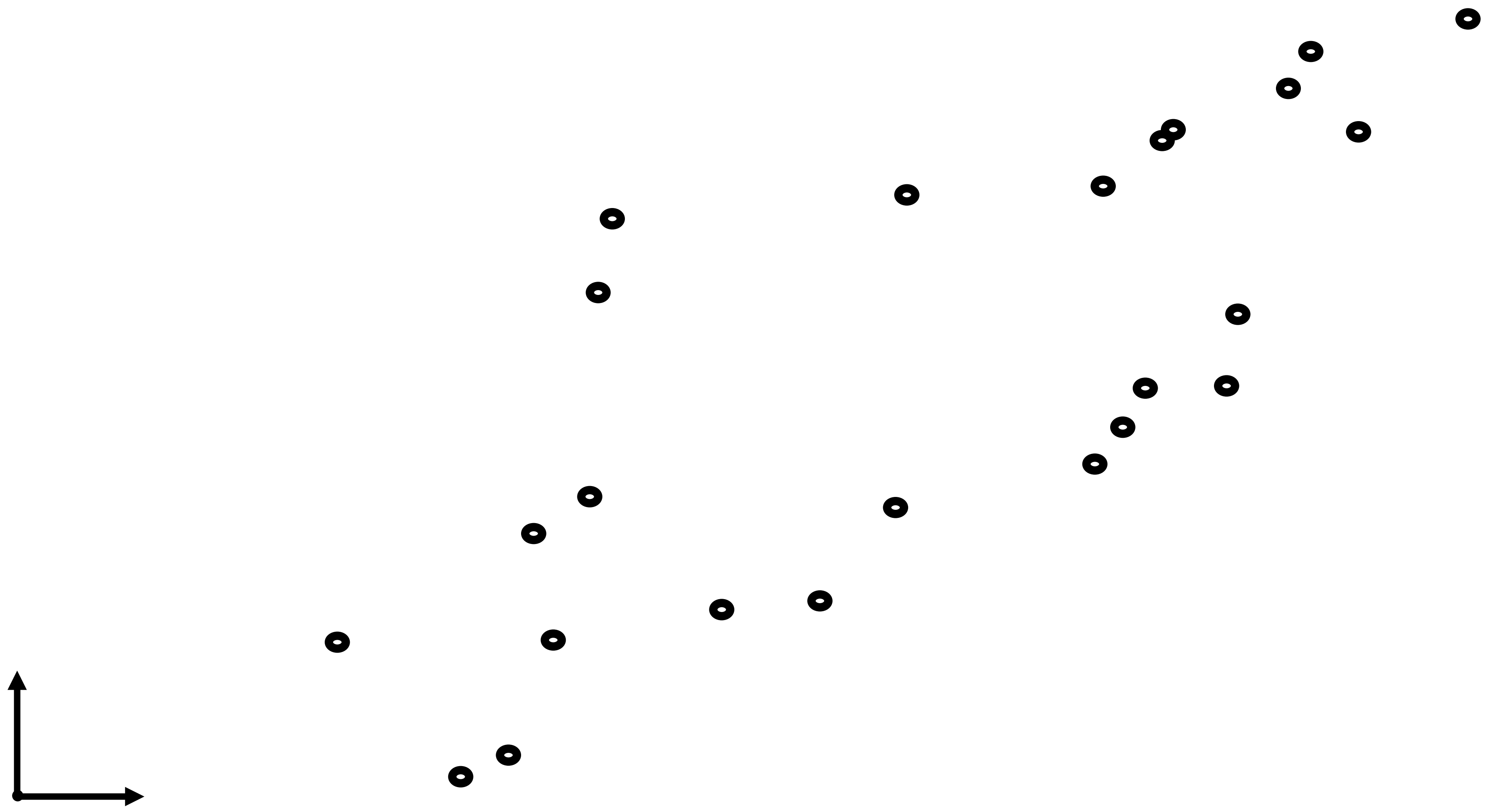


- Uncertainties originating in the measurements percolate up the ladder, and *uncertainty propagation* holds the ladder together
- Statistics: analysis of measurements to get inferences about shapes
- Communication: justifying inferences to judges, juries, public
- Validation: confirming conclusions with real-world observations

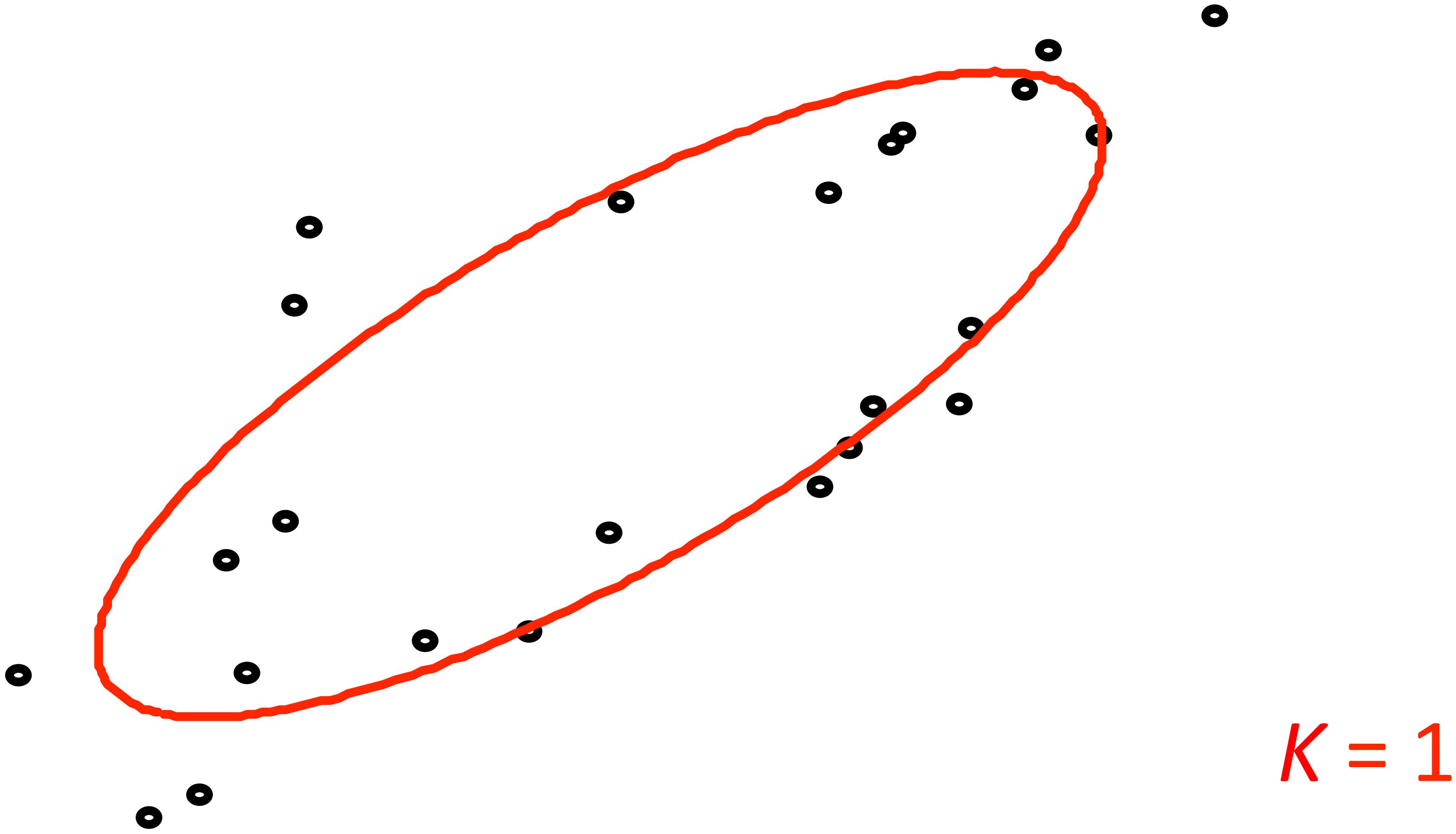
Elliptic Fourier analysis (EFA)

- Fits a closed curve to any ordered set of points in a plane
- Allows any degree of precision desired
- Orthogonal decomposition into a sum of harmonic ellipses
- Very simple computationally (doesn’t even need FFT)
- Used for cases with well-defined landmarks or image contours

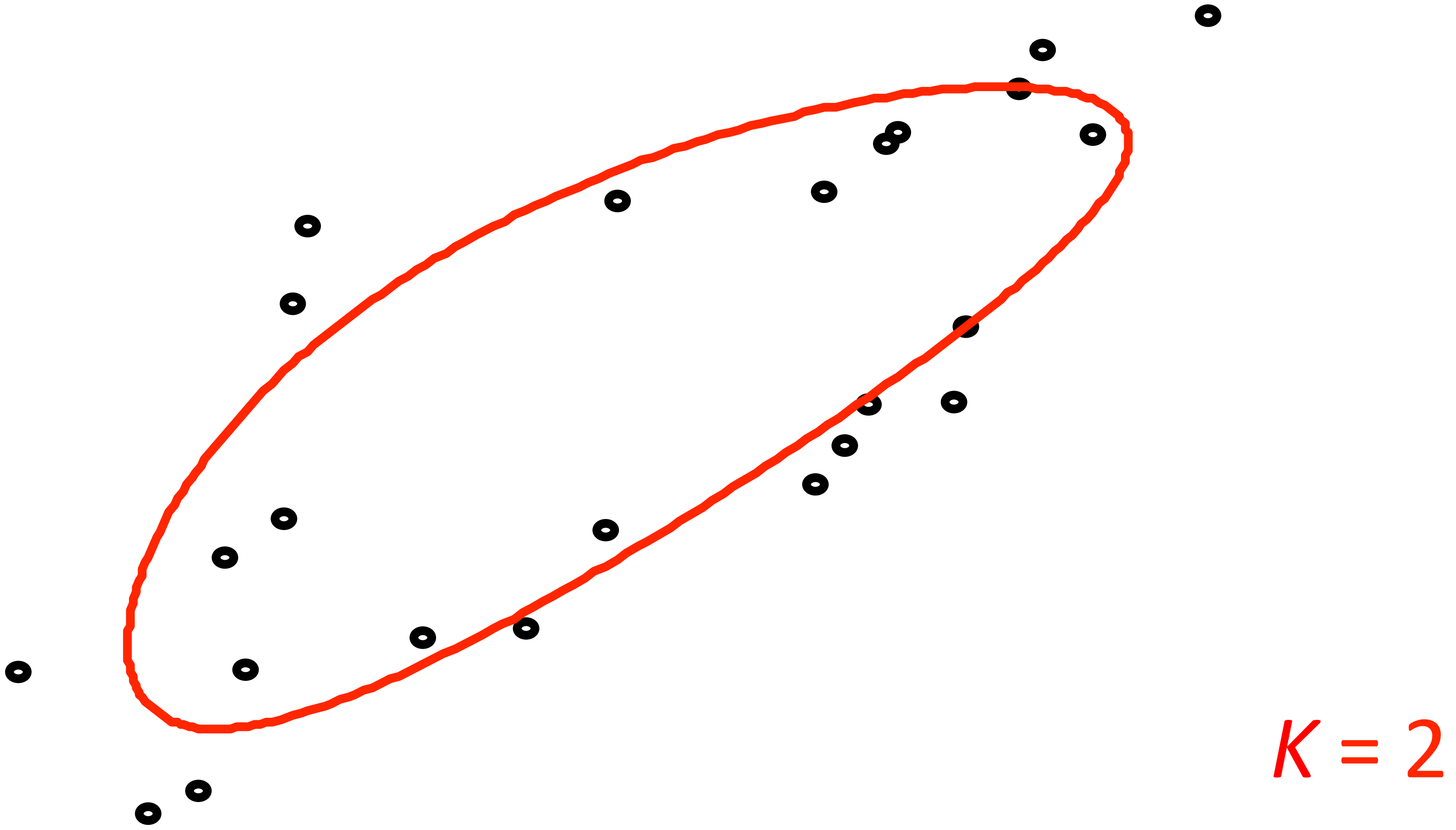
Landmark data e.g., locations of facial features or bone homologies



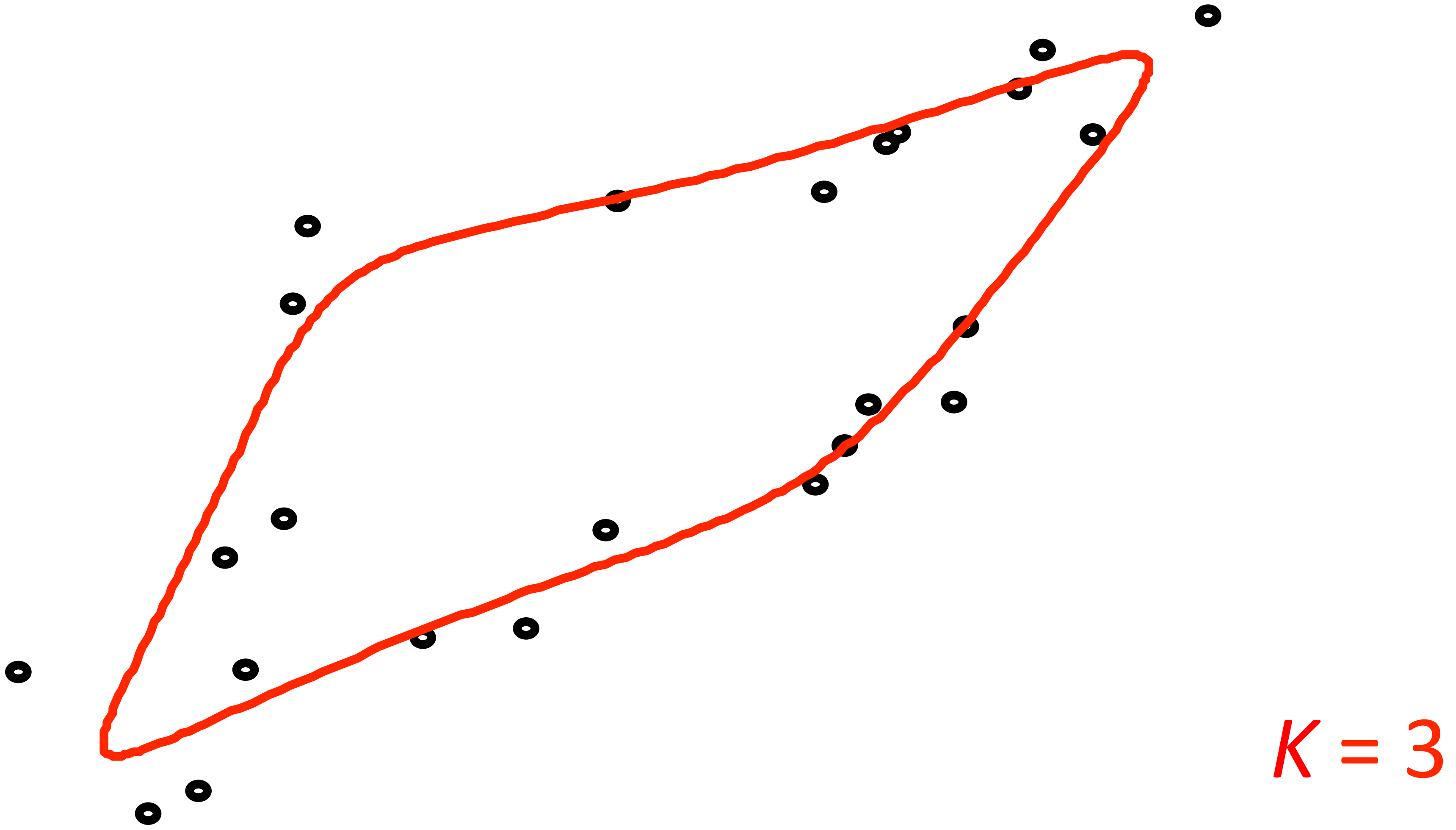
Landmark data



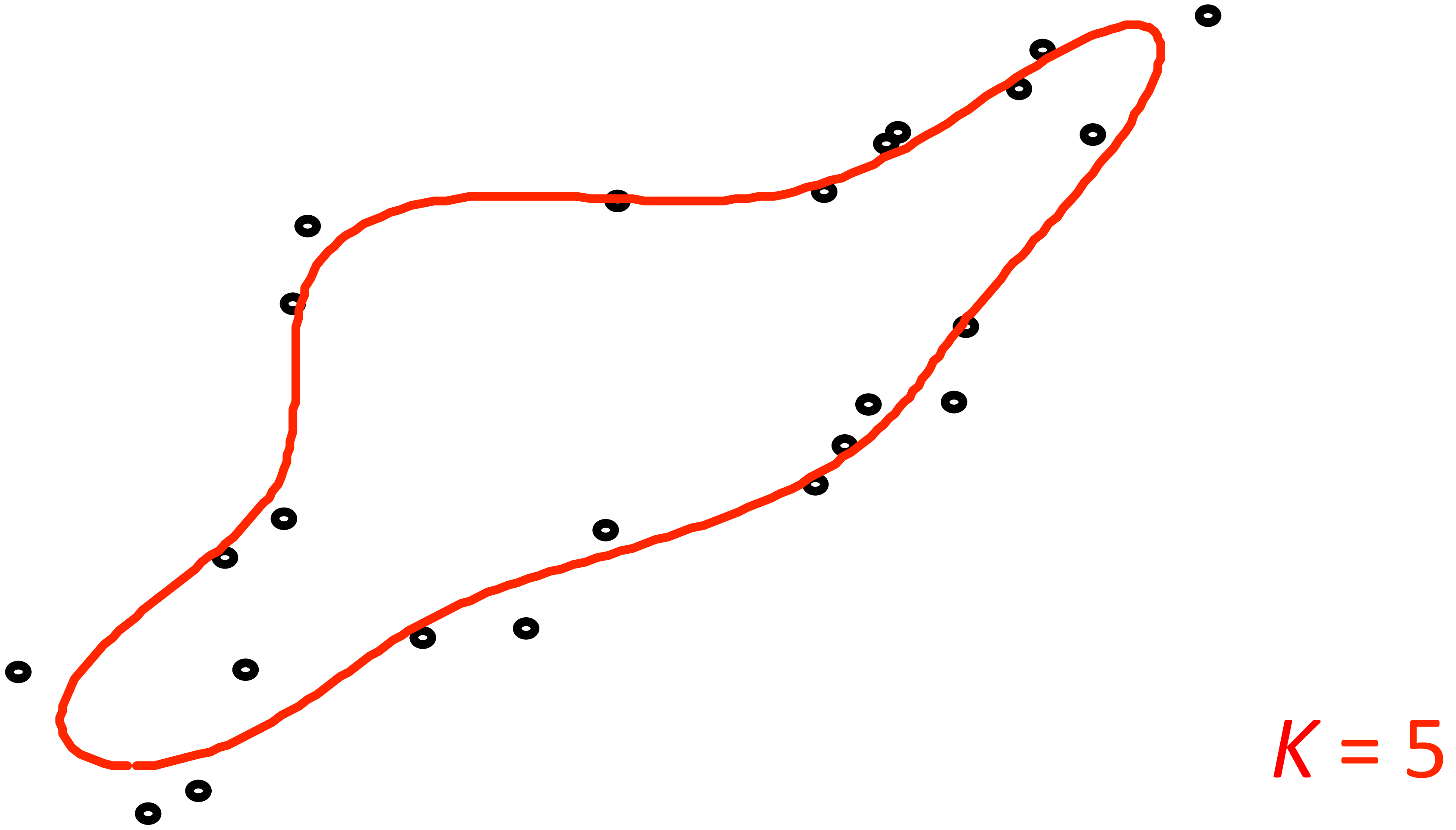
Landmark data



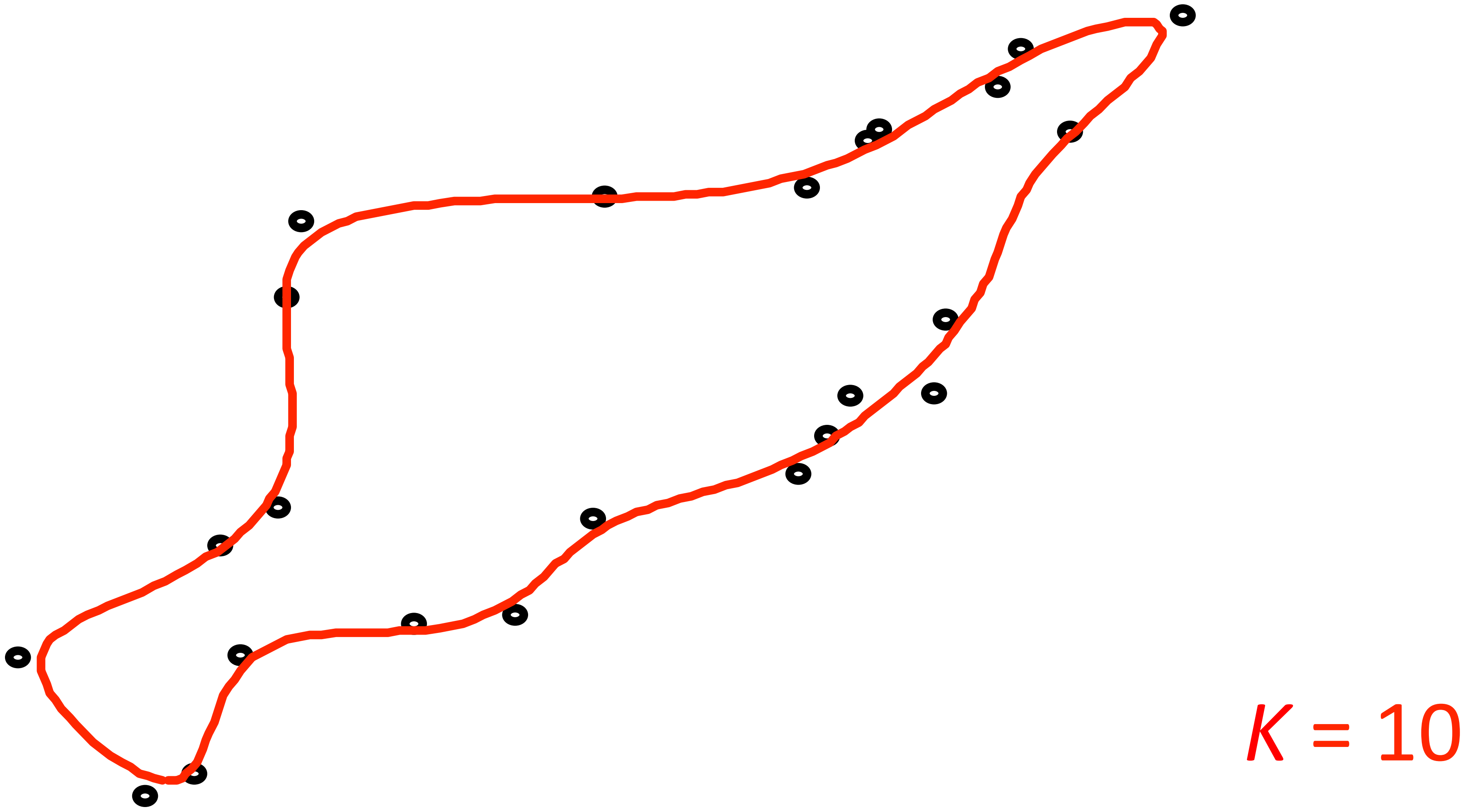
Landmark data



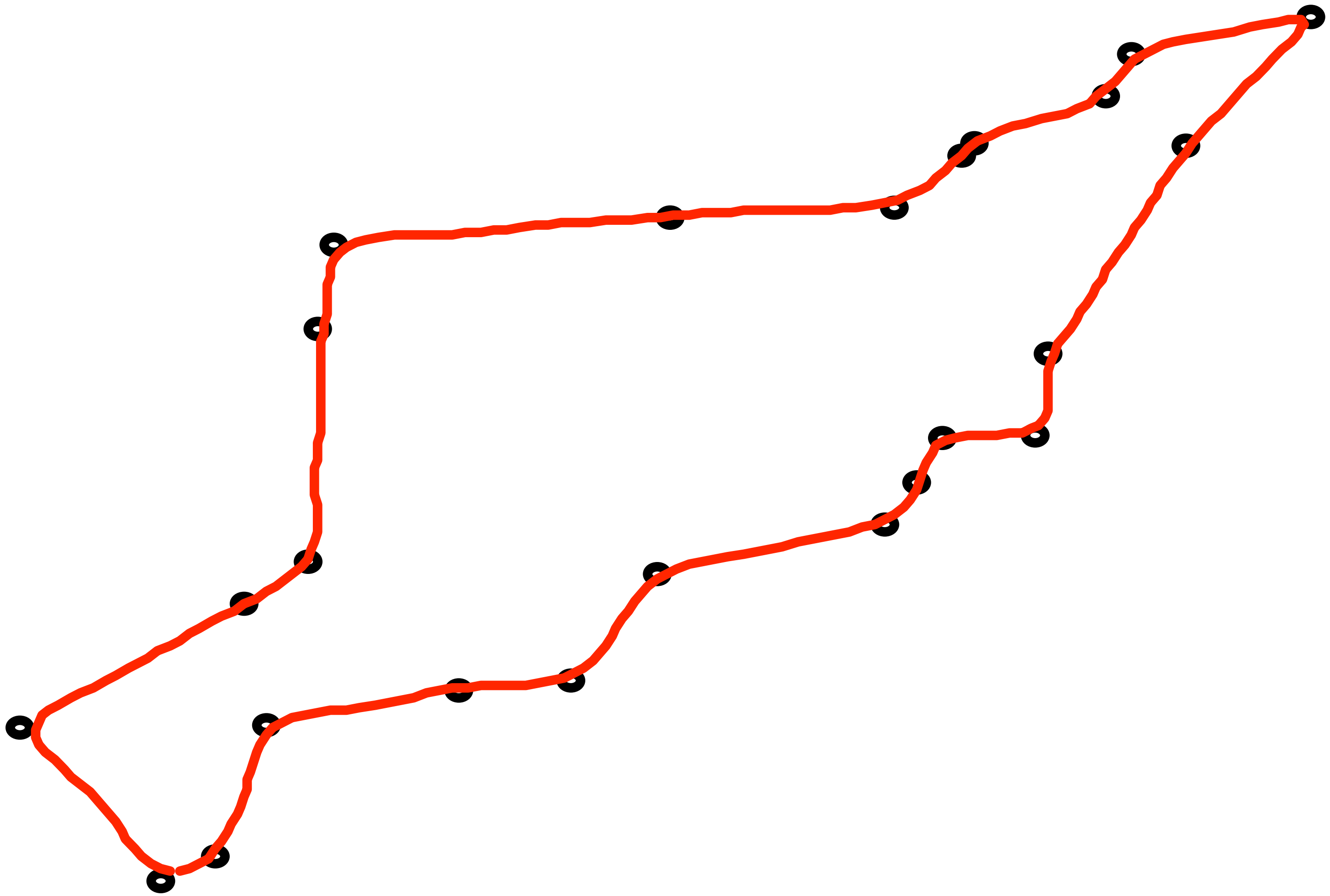
Landmark data



Landmark data

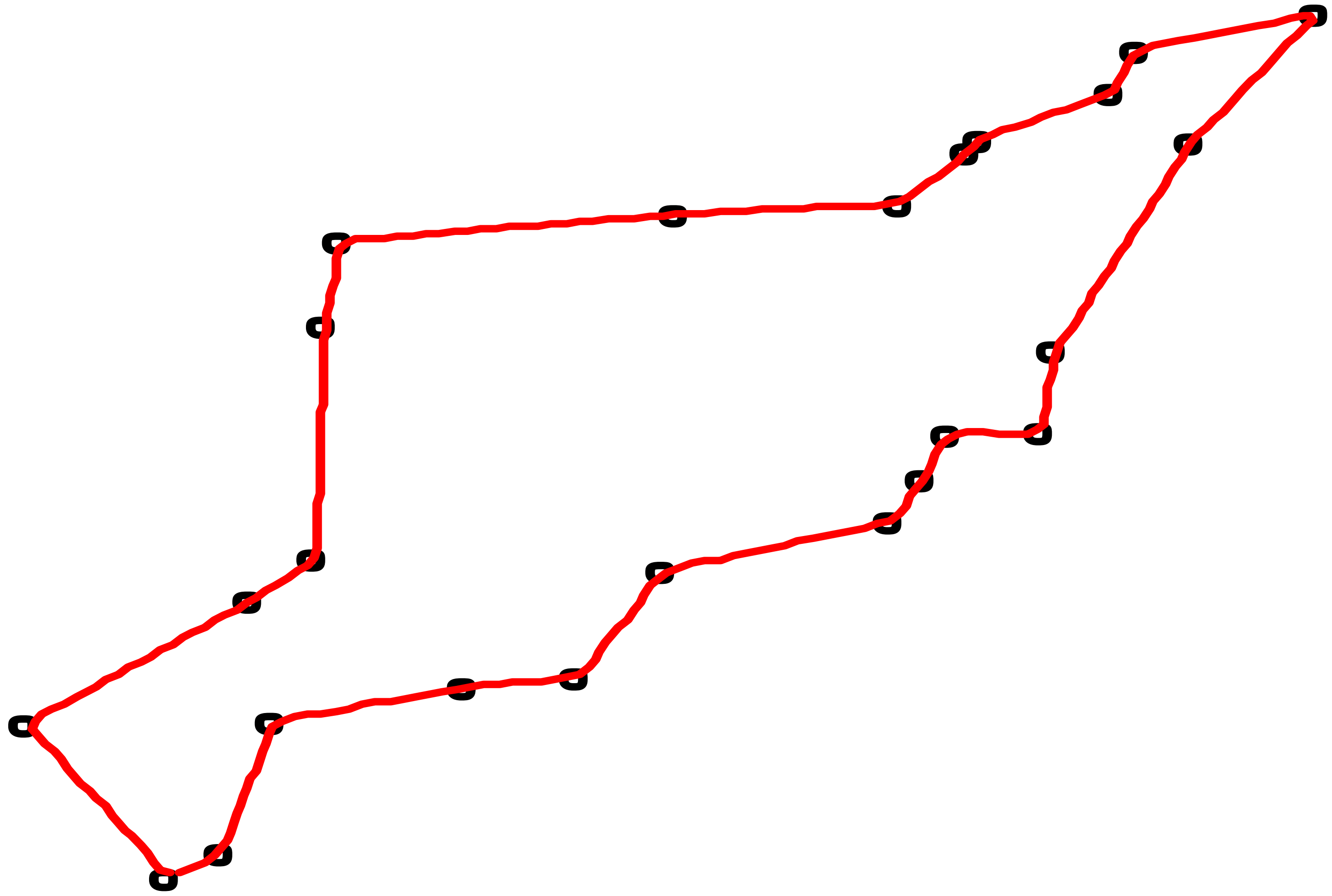


Landmark data

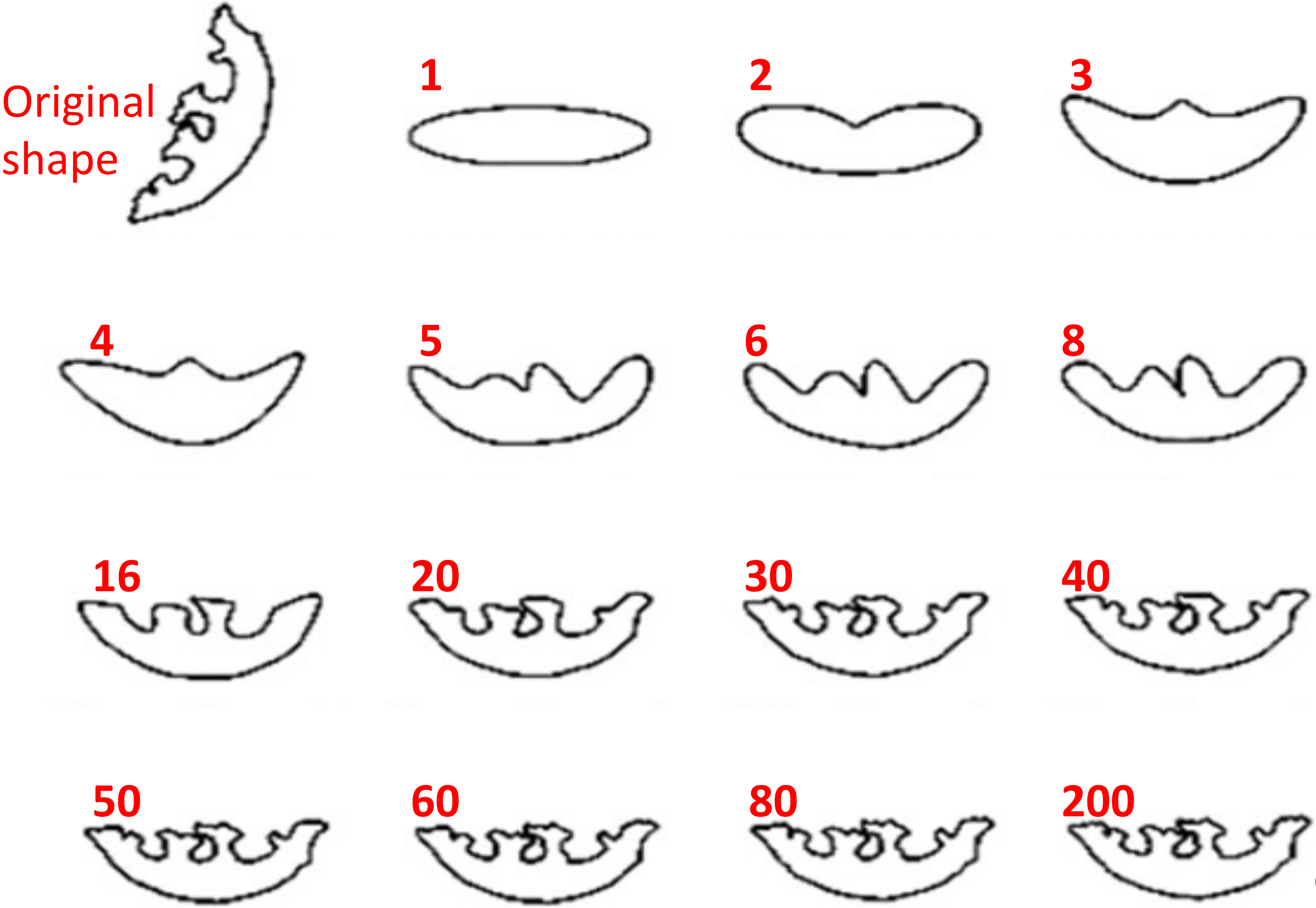


$K = 25$

Landmark data



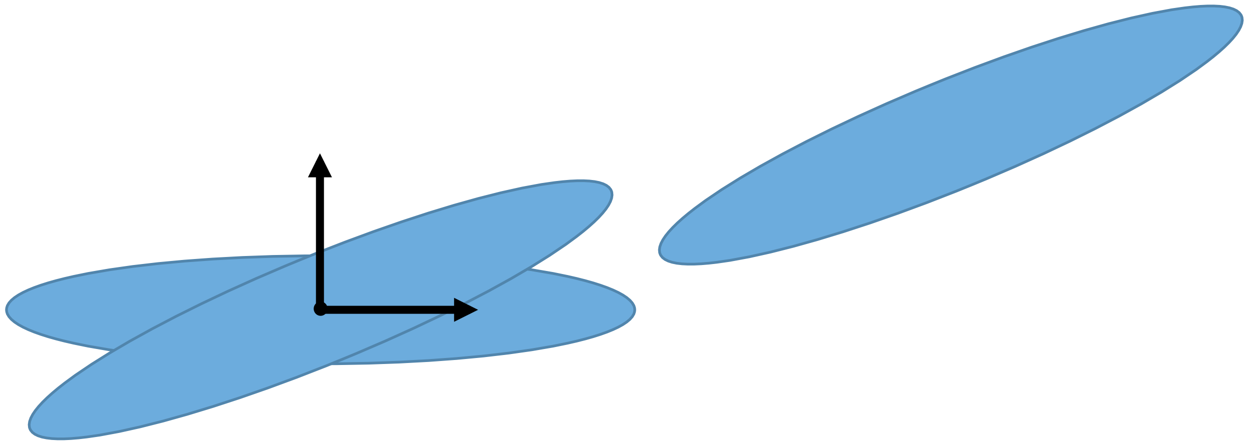
$K = 50$



Carlo et al. (2011)

Normalizations

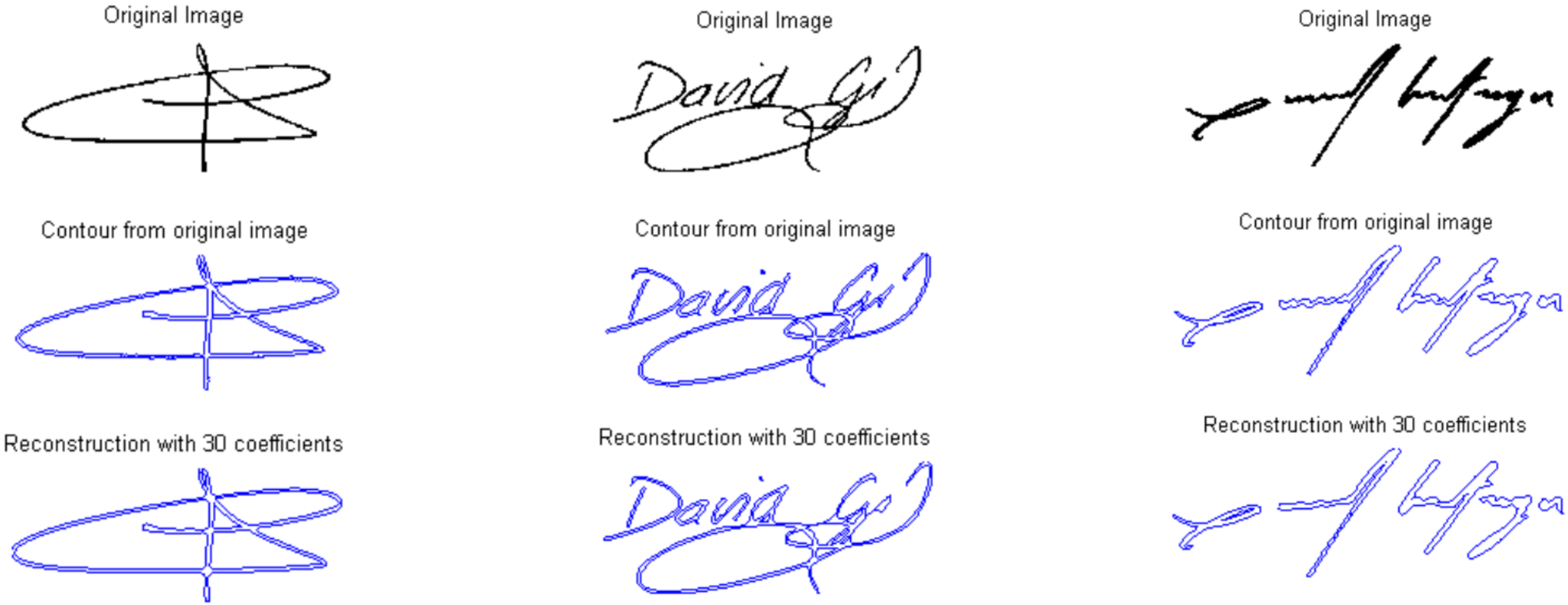
- Several aspects of a contour may be irrelevant to its "shape"
    - Placement
    - Orientation
    - Size or magnification
    - Starting point
    - Direction of the trace



- The EFA coefficients can be normalized to ignore these things
- Analysts can chose which normalizations to use

Signatures

- Marti-Puig et al. (2013) applied EFA to signatures http://repositori.uvic.cat/bitstream/handle/10854/2278/artconlli_a2013_marti_puig_parameterization_written.pdf?sequence=1



What if the landmarks/contour is *poorly* defined?

- Honest characterization of measurement uncertainties may result in contours that are imprecisely or incompletely defined

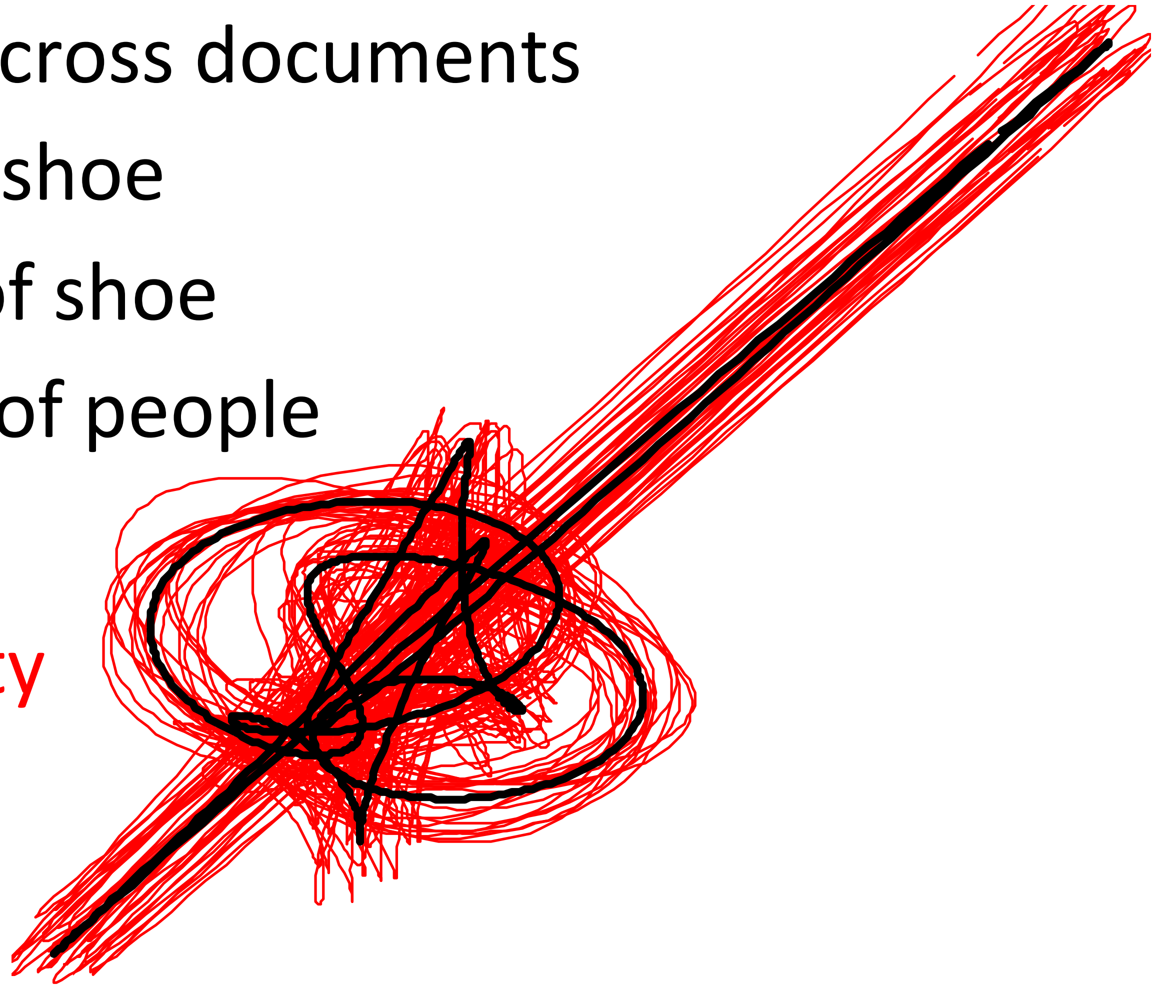
- EFA can be applied in these cases too

- These applications require modern uncertainty projection which distinguishes between
  - Epistemic uncertainty: imprecise or missing measurements
  - Aleatory uncertainty: variation across a population

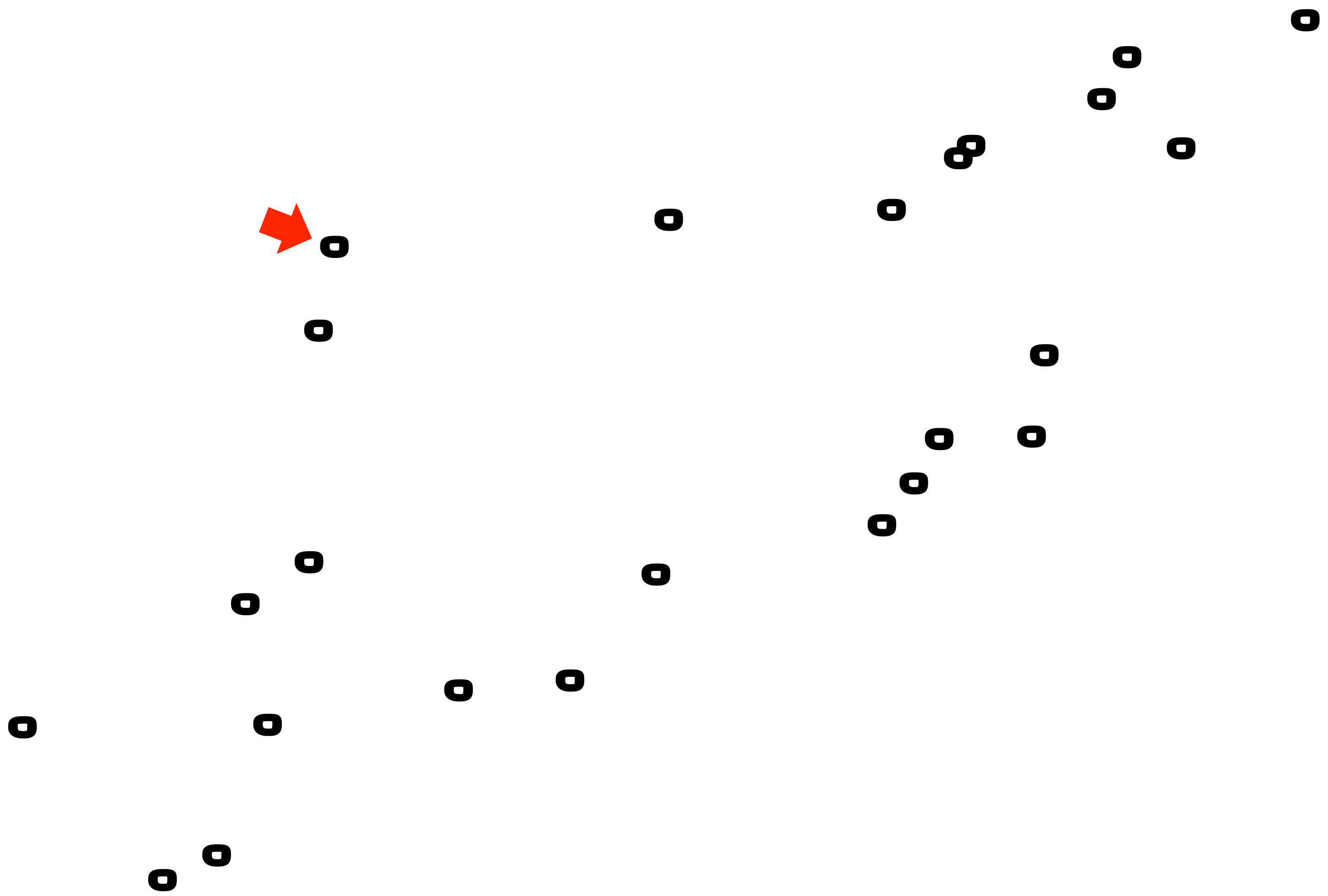
Population of shapes

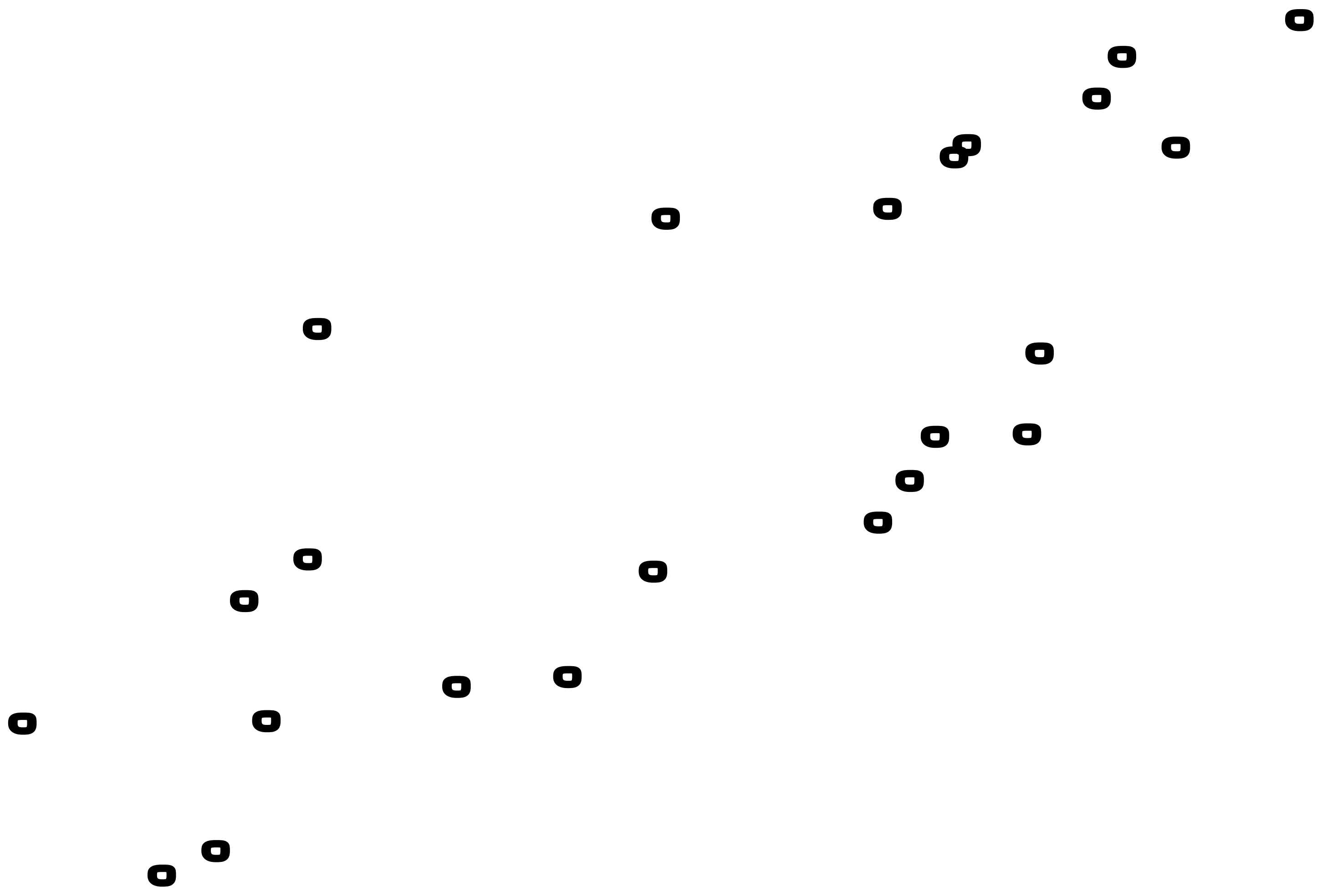
- Different signature samples across documents
- Impressions made by a given shoe
- Impressions made by a kind of shoe
- Shoe impressions by a group of people

Variation is aleatory uncertainty

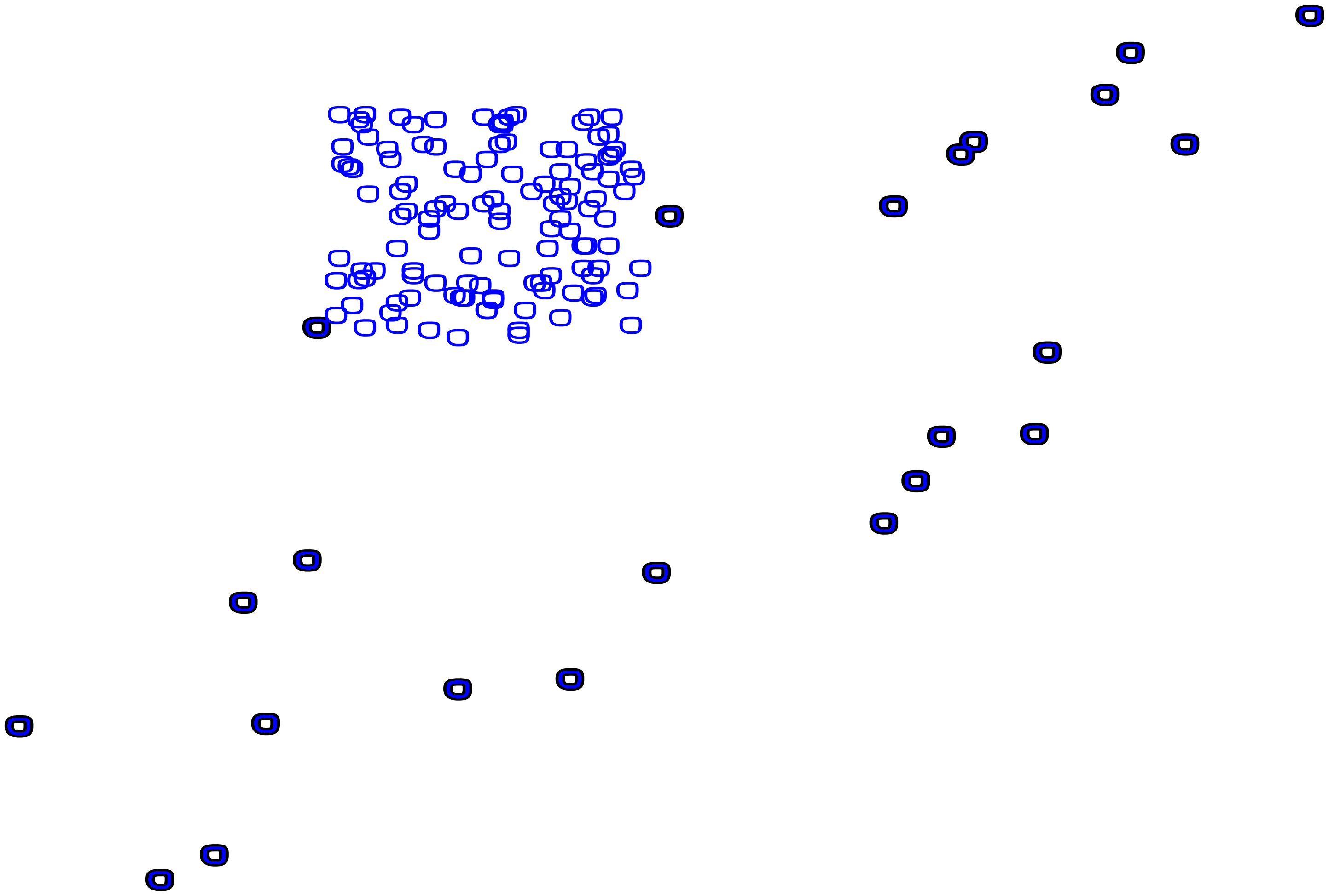


Missing or uncertain landmarks

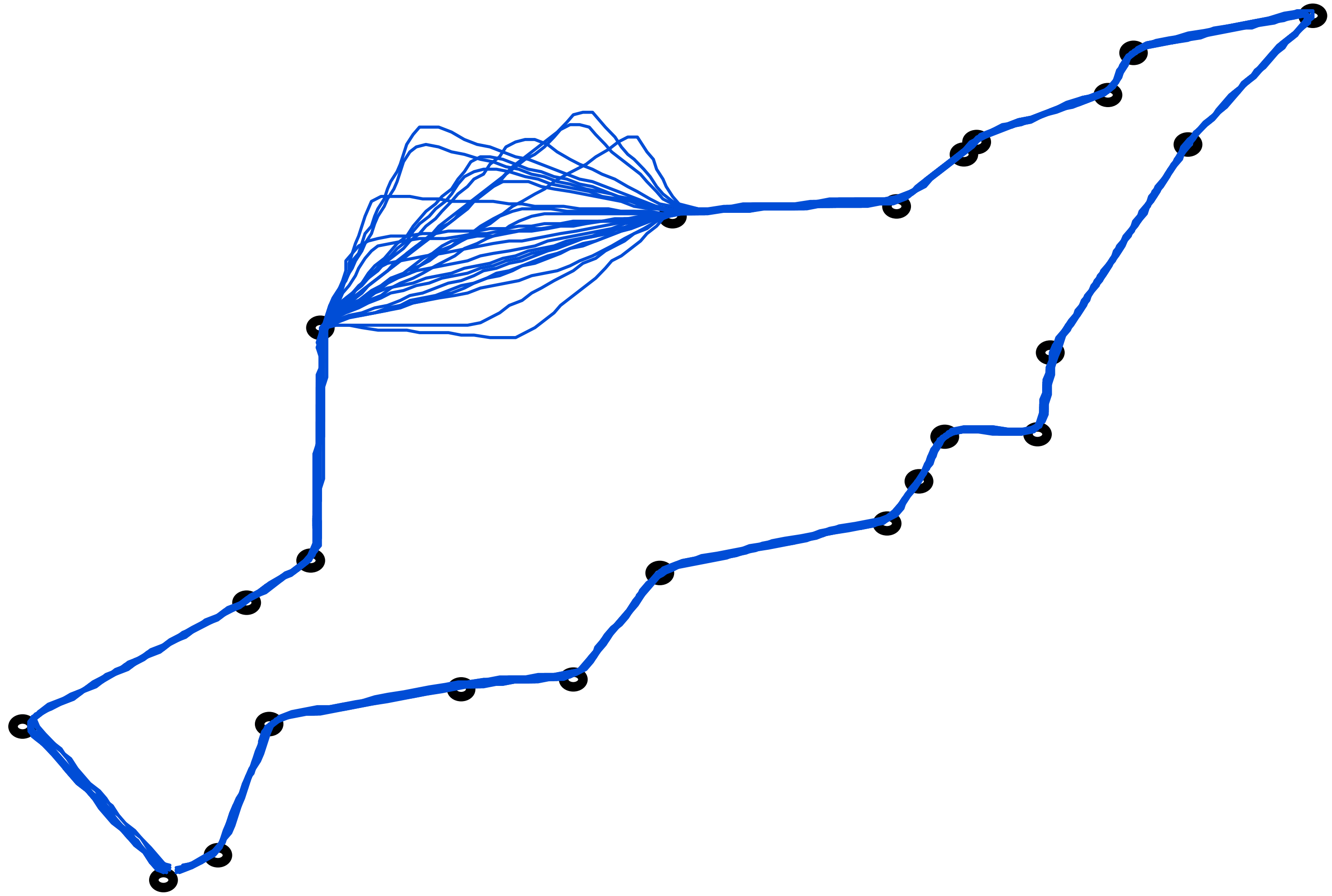




Suppose the location is partially known



Several shapes are epistemically possible



$K = 50$

Digitization uncertainty

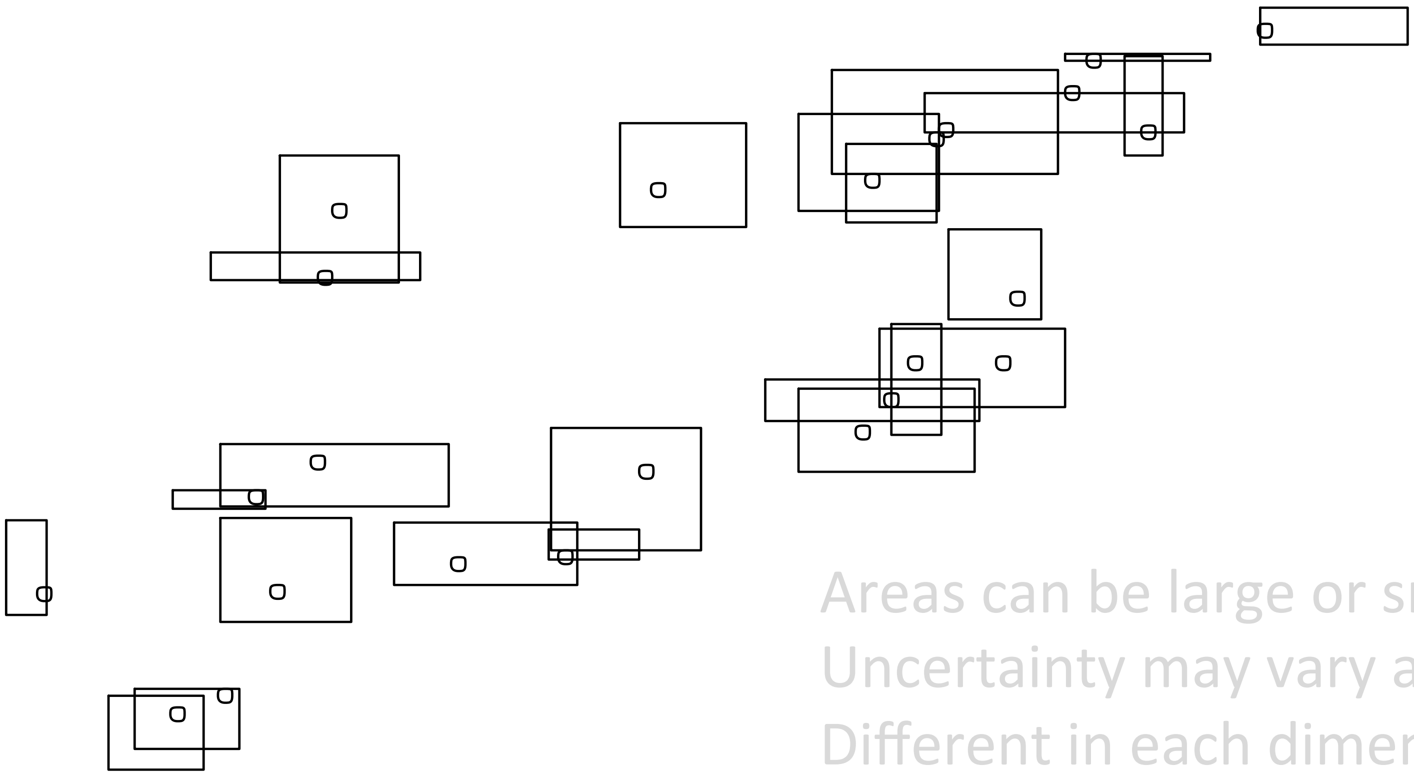
- Imprecision in capturing the points in the plane
  - Different technologies have different very precisions
  - Doubt about the definition of a landmark
- Traditionally ignored
  - Splitting the difference
  - Analyst's judgment

Where's the extremum of a rounded process?
Where's the margin exactly in a bifurcation?
Where's the center of an oculus?
Where is the midline of a ridge?

Automated or human-mediated?
What is the image resolution / pixel size?
Were there photographic distortions?
Did the image cleaning remove some details?

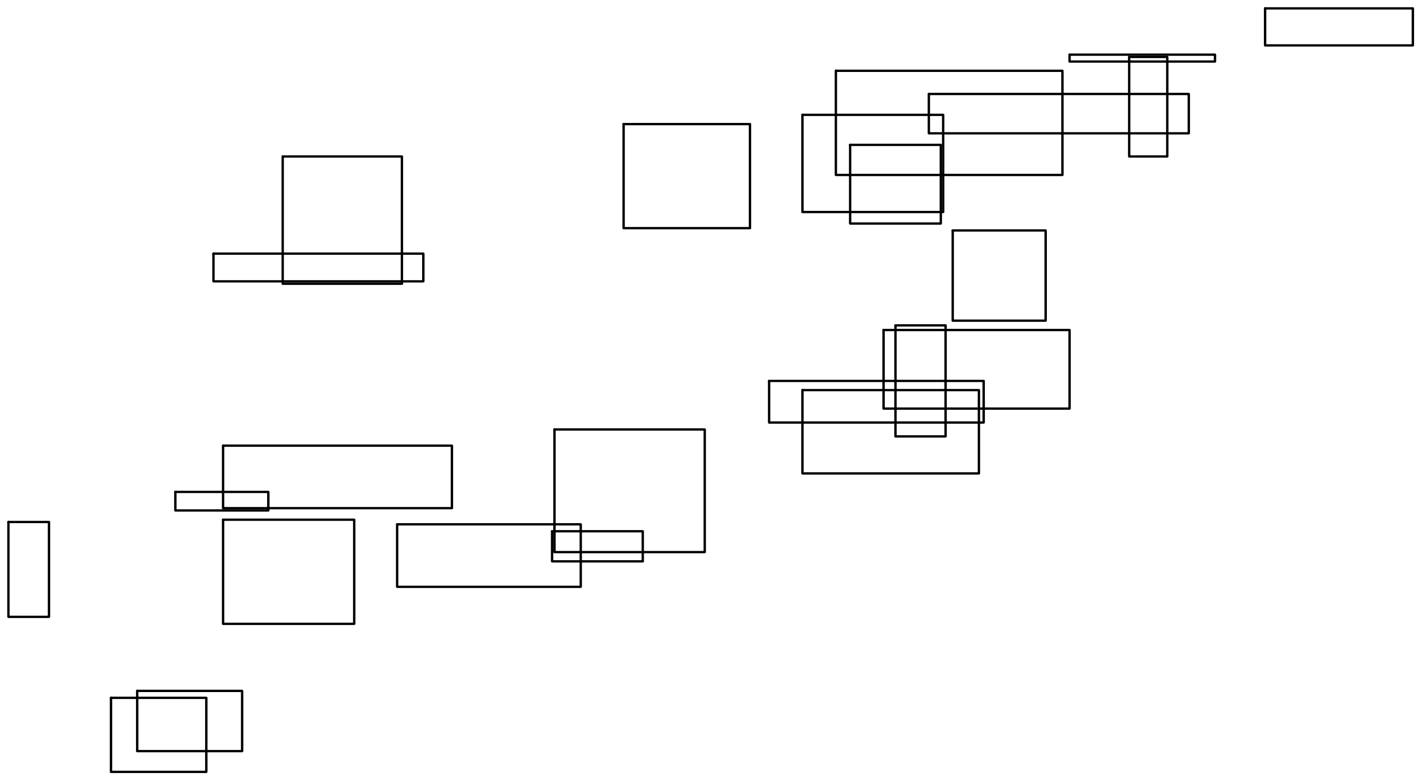
- Can be quite important when propagated to conclusions
- EFA can handle *and account for* this epistemic uncertainty

Analysts often report *points* for uncertain *areas*

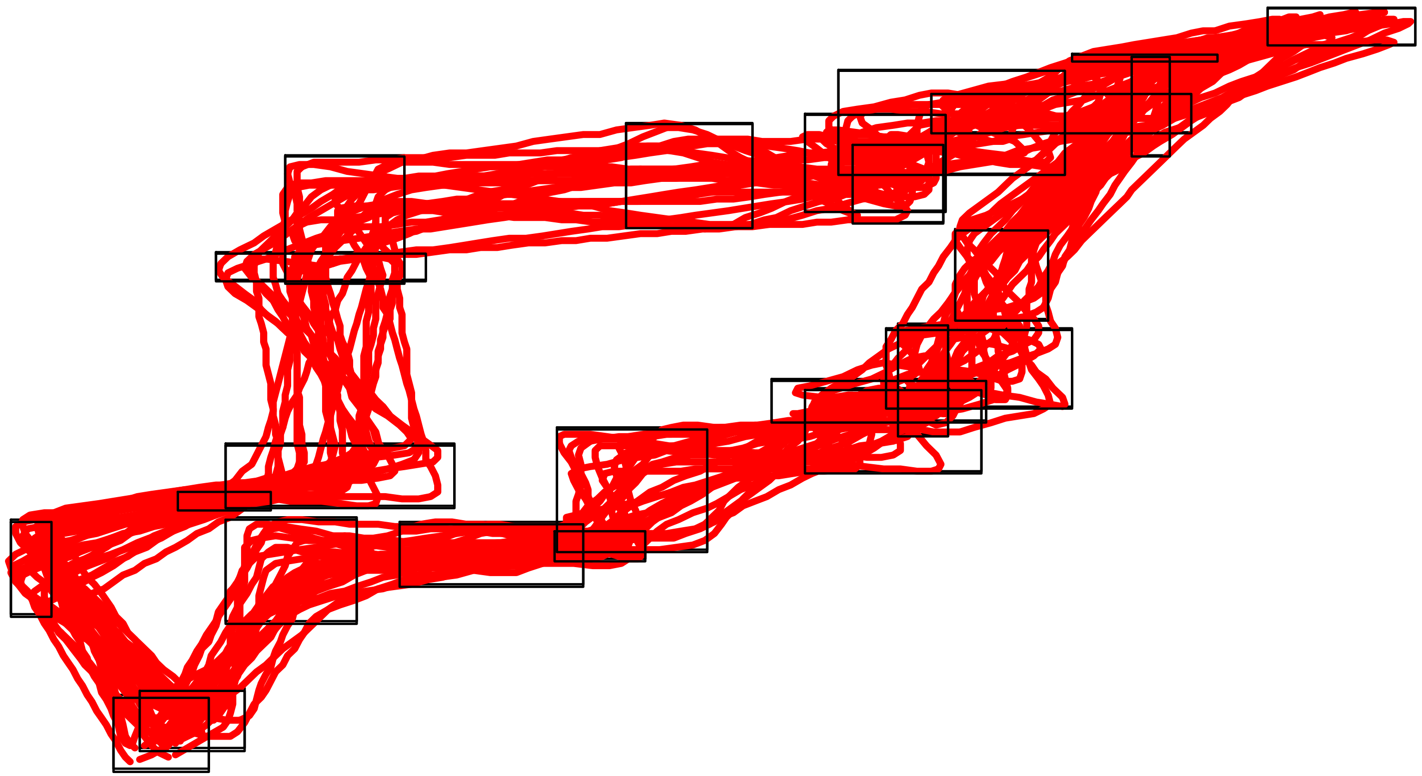


Areas can be large or small (precise)
Uncertainty may vary among landmarks
Different in each dimension
Point values not always in the centers
Can even be overlapping
Much more realistic

Areas can be used to locate landmarks



The result is a *set* of shapes, any of which is possible



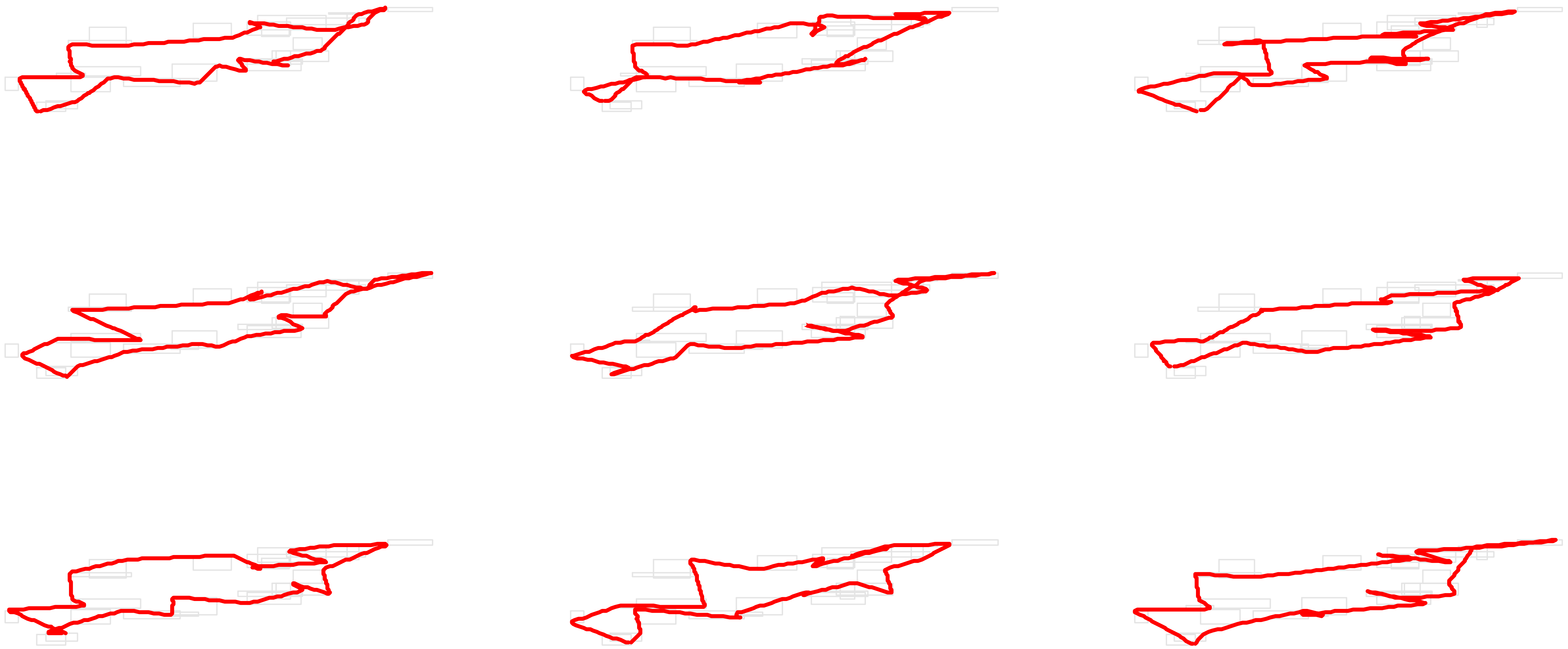
Standard multivariate statistical solutions

- Is a given shape an outlier compared to a population of shapes?
- Which cluster of shapes is a given shape most like?
- Are groups of shapes from the same population of shapes?
- How variable are shapes that are as different as a given shape from a known exemplar shape?

Statistical methods work in the *shape space* of EFA coefficients

Averages, distance, clusters, classifications, regressions on age/gender/race, etc.

Panoply of shapes



Visualizing variation

- Explaining a “match” to juries and judges is hard
- Non-experts lack the necessary experience and context
- Yet humans have extremely sophisticated visual skills
- Reconstituted shapes from EFA can visualize discrepancies as a panoply of shapes, each of which is as different from an exemplar shape as the given shape

Generality

- Can omit some invariance normalizations
  - Sometimes size, orientation, location, sense, start, etc. matter
- Don’t need a continuous trace; works with any point sequence
  - Landmarks that have been assigned an (arbitrary or natural) order
- Doesn’t need to be in a plane
  - 3D trajectories with a third equation for differences in *z*-direction
- Correctly handles aleatory and epistemic uncertainty

Handling error in forensic shape analysis

- Honesty about measurement uncertainties
- Careful propagation of epistemic measurement uncertainties
- Good statistical analyses to account for variation
- Addressing uncertainty means “100% matches” won’t happen

Acknowledgments

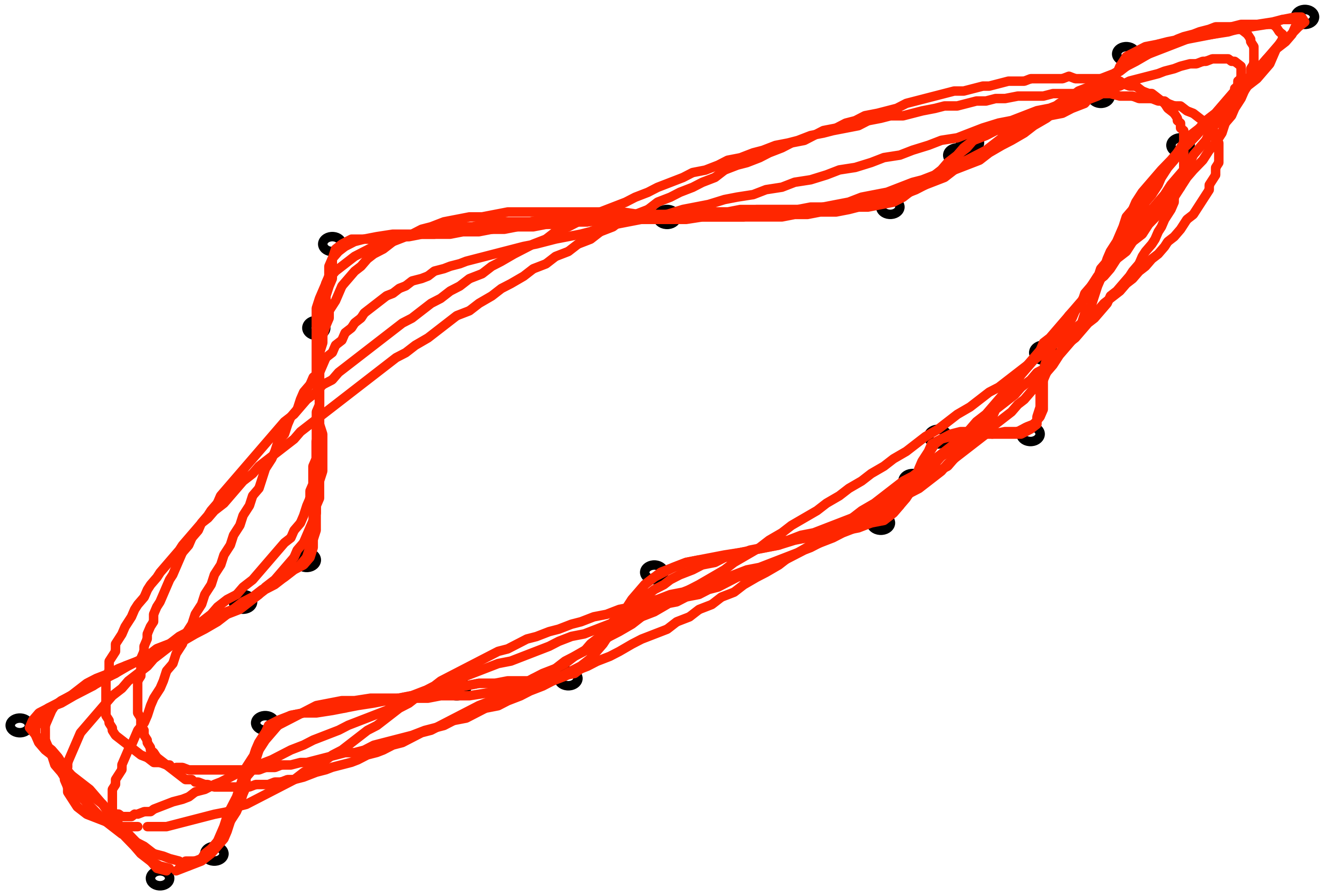
- Andrew Dienstfrey, NIST
- F. James Rohlf, Stony Brook University
- Dennis Slice, University of Florida
- Nick Petraco, John Jay College

"To learn which questions are unanswerable, and *not to answer them*: this skill is most needful in times of stress and darkness."

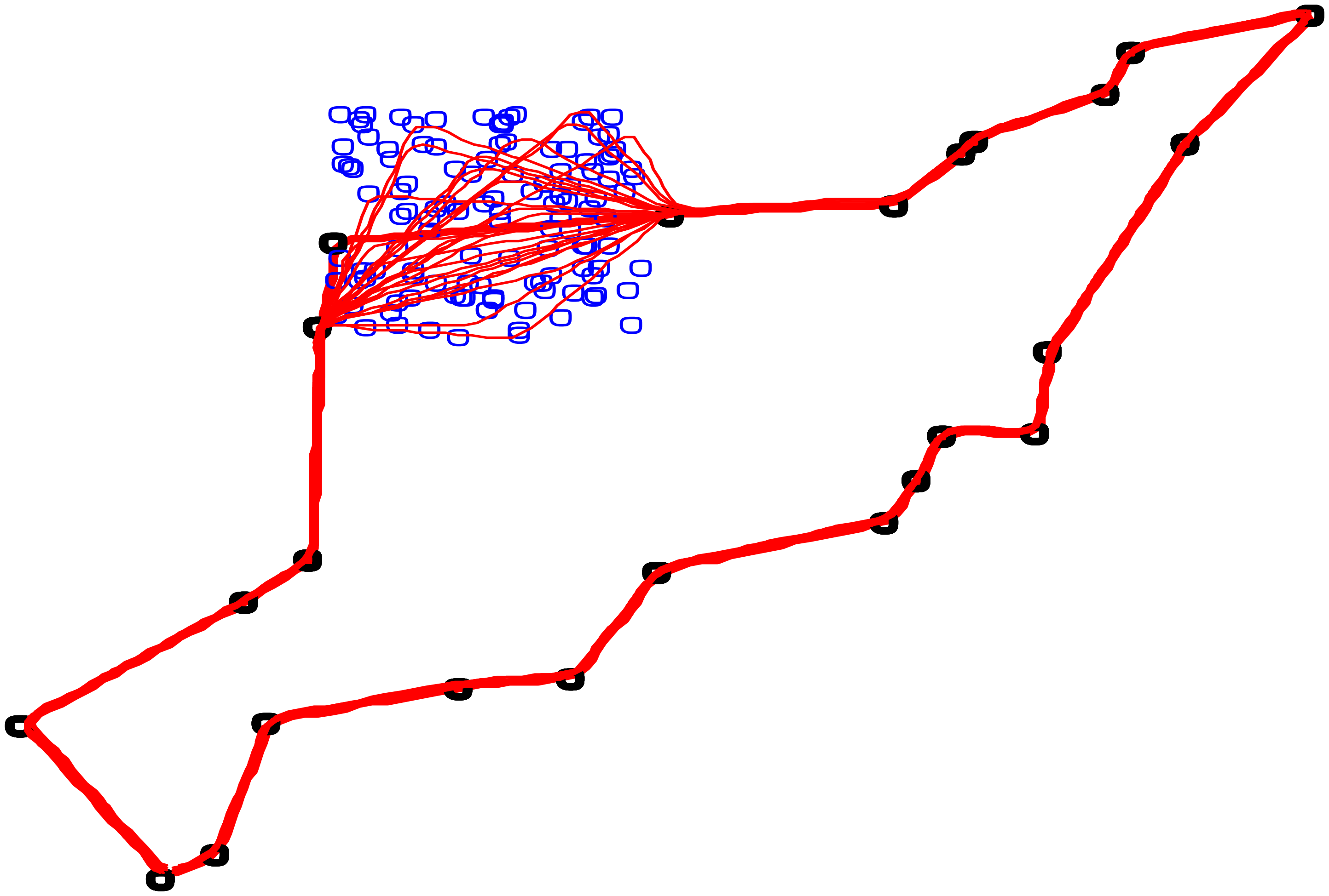
Ursula K. Le Guin, *The Left Hand of Darkness*

End

Landmark data



$K = 1$



Distance between shapes