Analyzing Nonlinear Variations in Common Shape Curves

Rima Izem

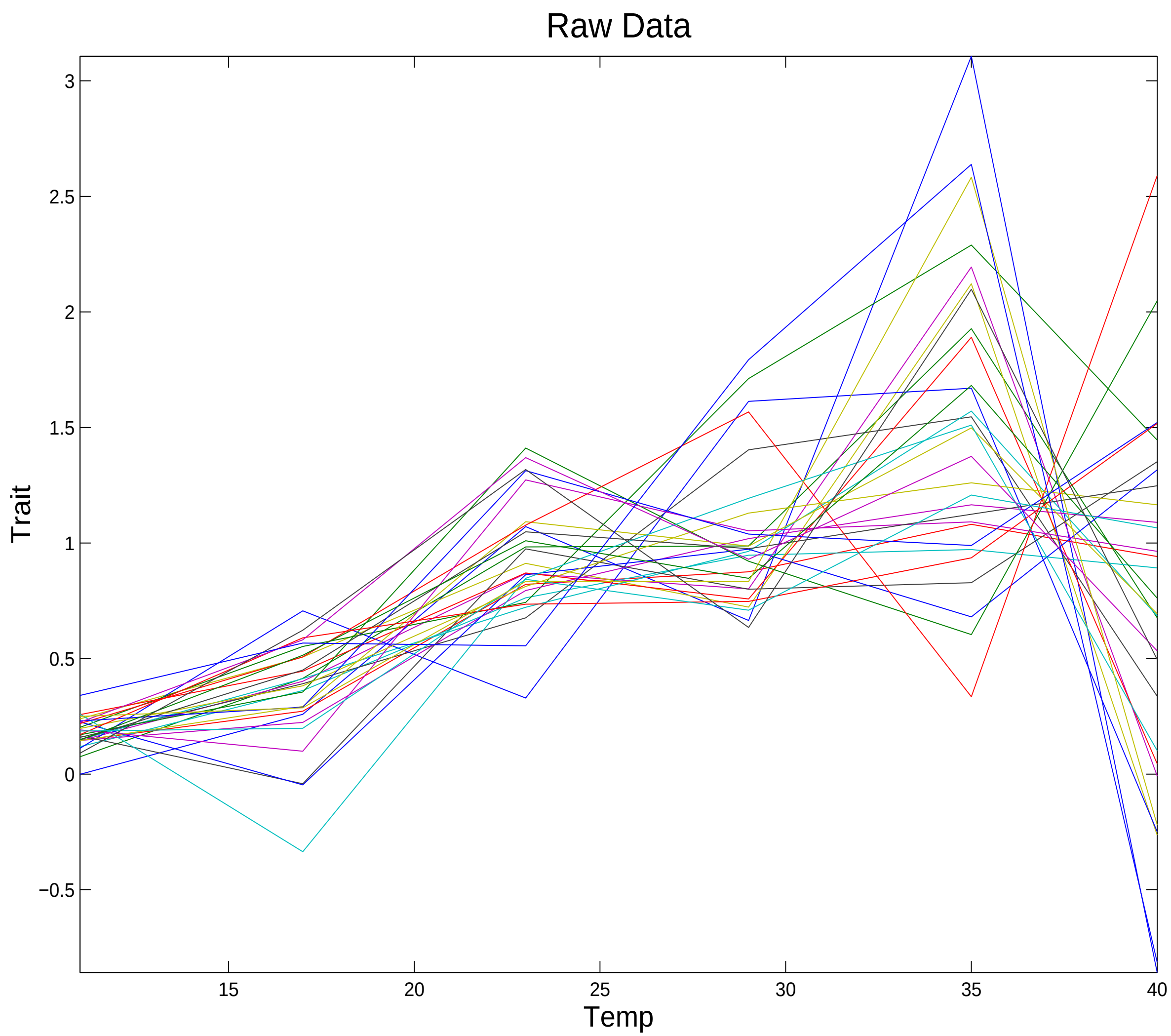
Joint work with

J.S Marron, Department of Statistics

and

J.G Kingsolver, Department of Biology

August 20, 2003

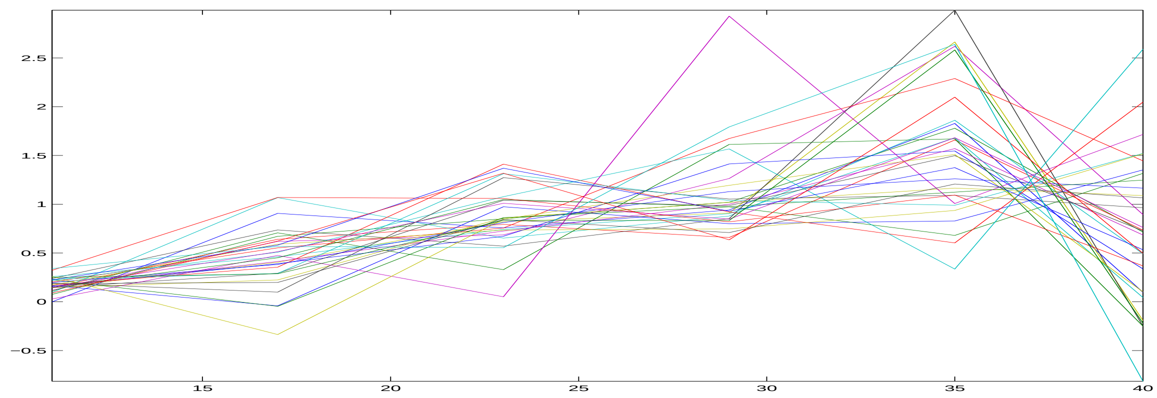


Raw Data
Trait
Temp
3
2.5
2
1.5
1
0.5
0
−0.5
15
20
25
30
35
40

Outline

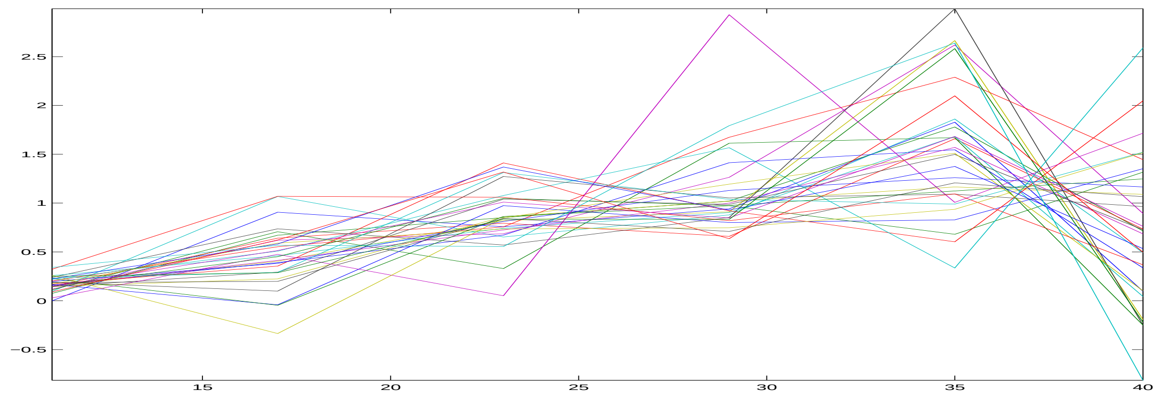
- Data
- Motivation and model
- Toy example
- Results

- Performance z. Ex: $z =$ growth rate.
- Temperature t. Ex: $d = 6$, and $t_1, \ldots, t_6 \in [11, 41]$ Celsius.
- Thermal performance curve $(\mathbf{TPC})$: $z = f(t)$.
- Reaction Norm Curve: $z = f(e)$.



Feature(1): Curves' variation $\equiv$ Genetic variation

- Family = individuals of similar genotype.
  Ex: offsprings of same parents, clones,...
- Population = n families, Ex: $n = 32$.



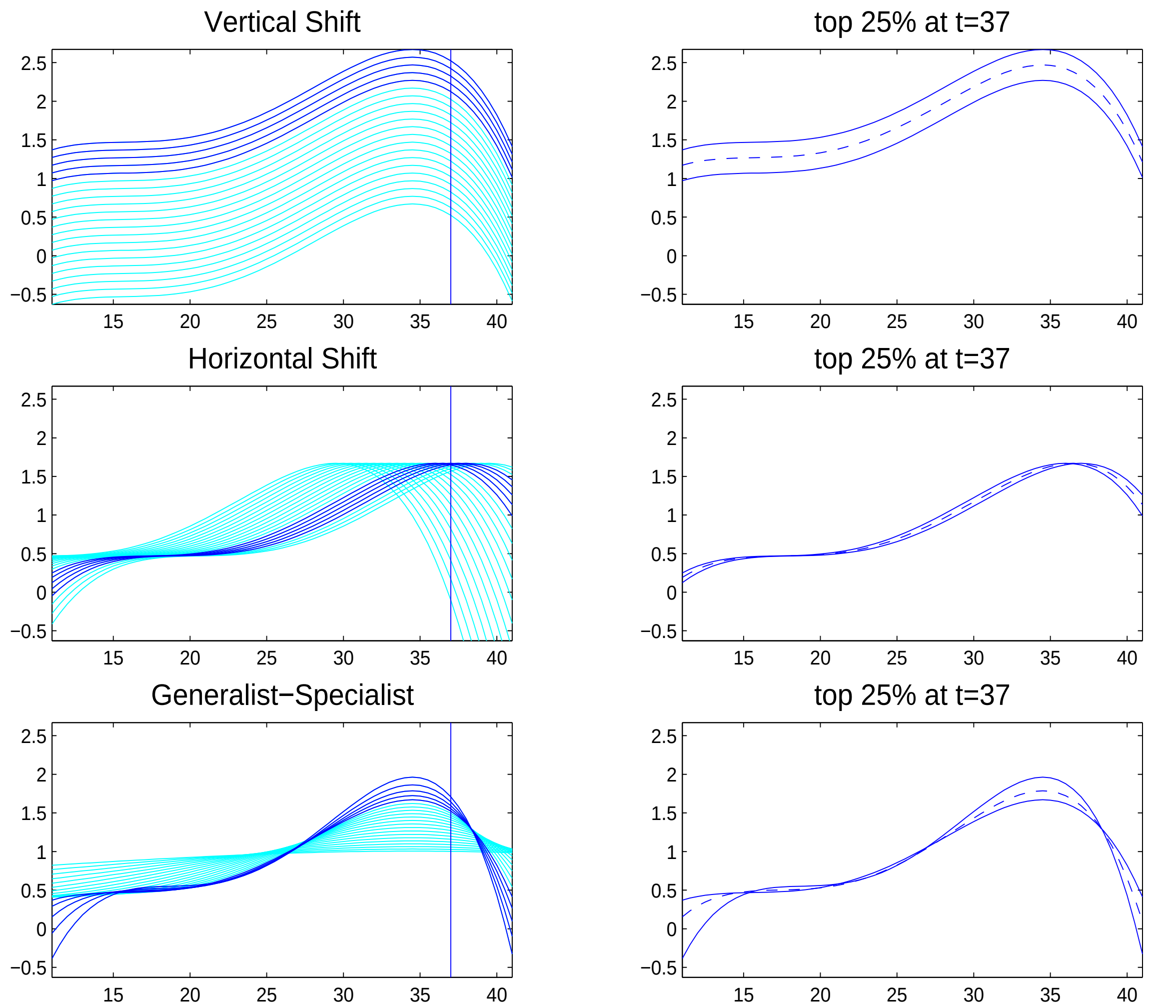
Feature (2) : Template shape $f(t)$ of TPC:

- $\uparrow$, reaches a max and $\downarrow$.
- Rate of $\uparrow$ slower than rate of $\downarrow$.
- Tendency toward a unique max.

Outline

- Data
- Motivation and model
- Toy example
- Results

- Goal: Analyze three TPC modes of variation.
- Motivation: **evolution** of gen. variation under **selection**.
- Mode of variation $\equiv$ Constraint to evolution



Mode of variation **before** selection – Genetic variation**after** selection

PCA	Template Modes of Variations
No assumption on shape	Template shape of TPCs.
Linear PC of G	**Nonlinear** modes of variation of G.
Meaning of each PC?	Biological modes.
RSS quantify a PC	$\widetilde{RSS}$ quantify a mode.

Model

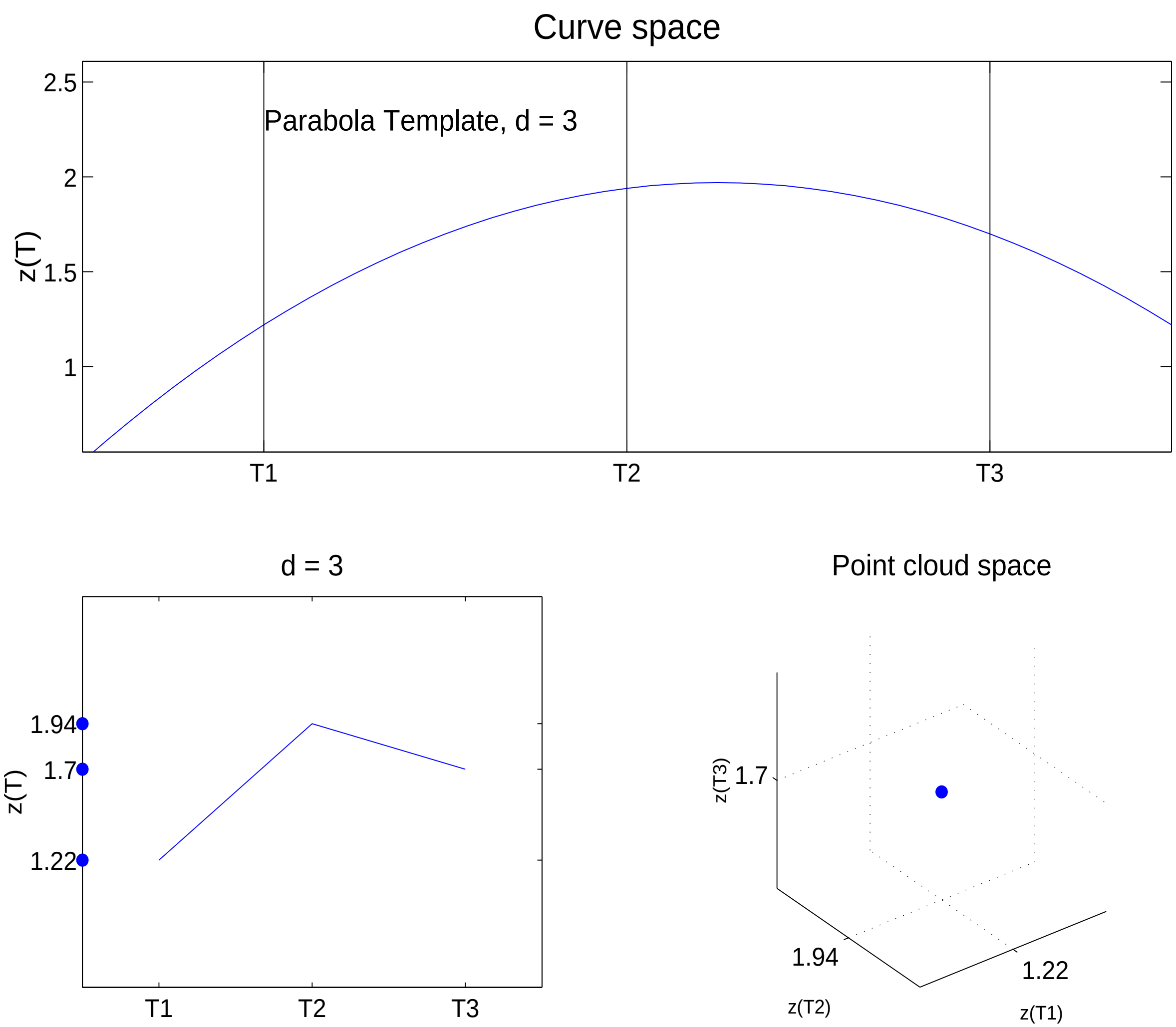
$$Z_i(t_j) = R_i(t_j) + \epsilon_{i,j}, 1 \leq i \leq n \text{ and } 1 \leq j \leq d.$$

$$R_i(t) = w_i f\left(w_i(t - m_i)\right) + h_i$$

	Parameter	Variation
Vertical shift	h_i	linear
Horizontal shift	m_i	nonlinear
Gen-Spec	w_i	nonlinear

Outline

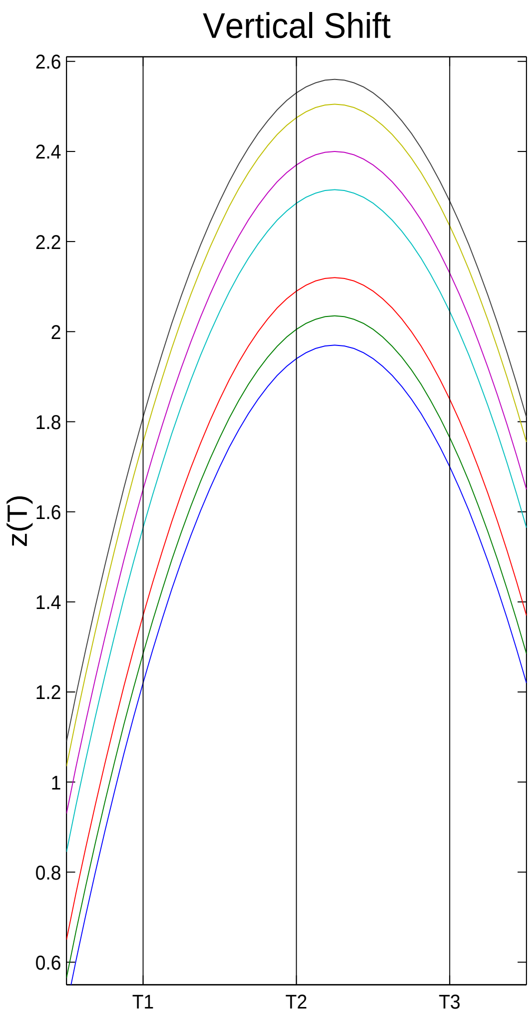
- Data
- Motivation and model
- Toy example
  - One variation (Vertical Shift, Horizontal Shift).
  - More than one variation (two, three, etc).
- Results

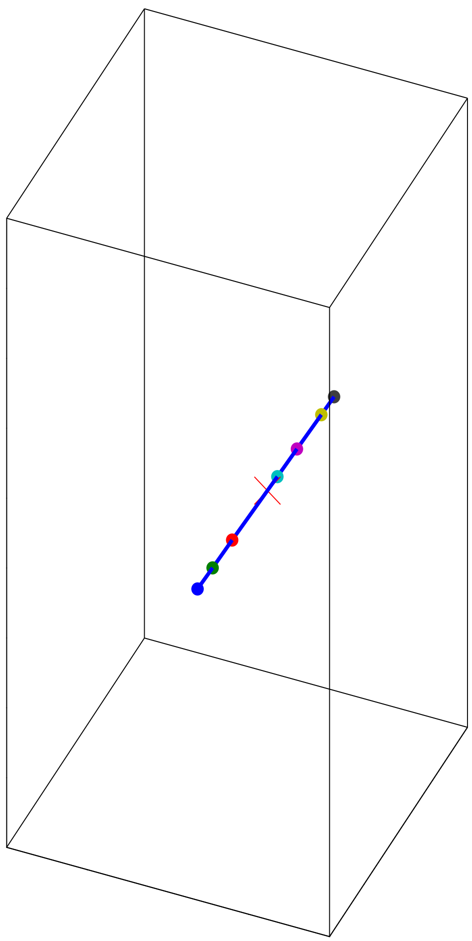

Curve space
2.5
2
1.5
1
z(T)
Parabola Template, d = 3
T1
T2
T3
d = 3
1.94
1.7
1.22
z(T)
T1
T2
T3
Point cloud space
1.7
z(T3)
1.94
z(T2)
1.22
z(T1)

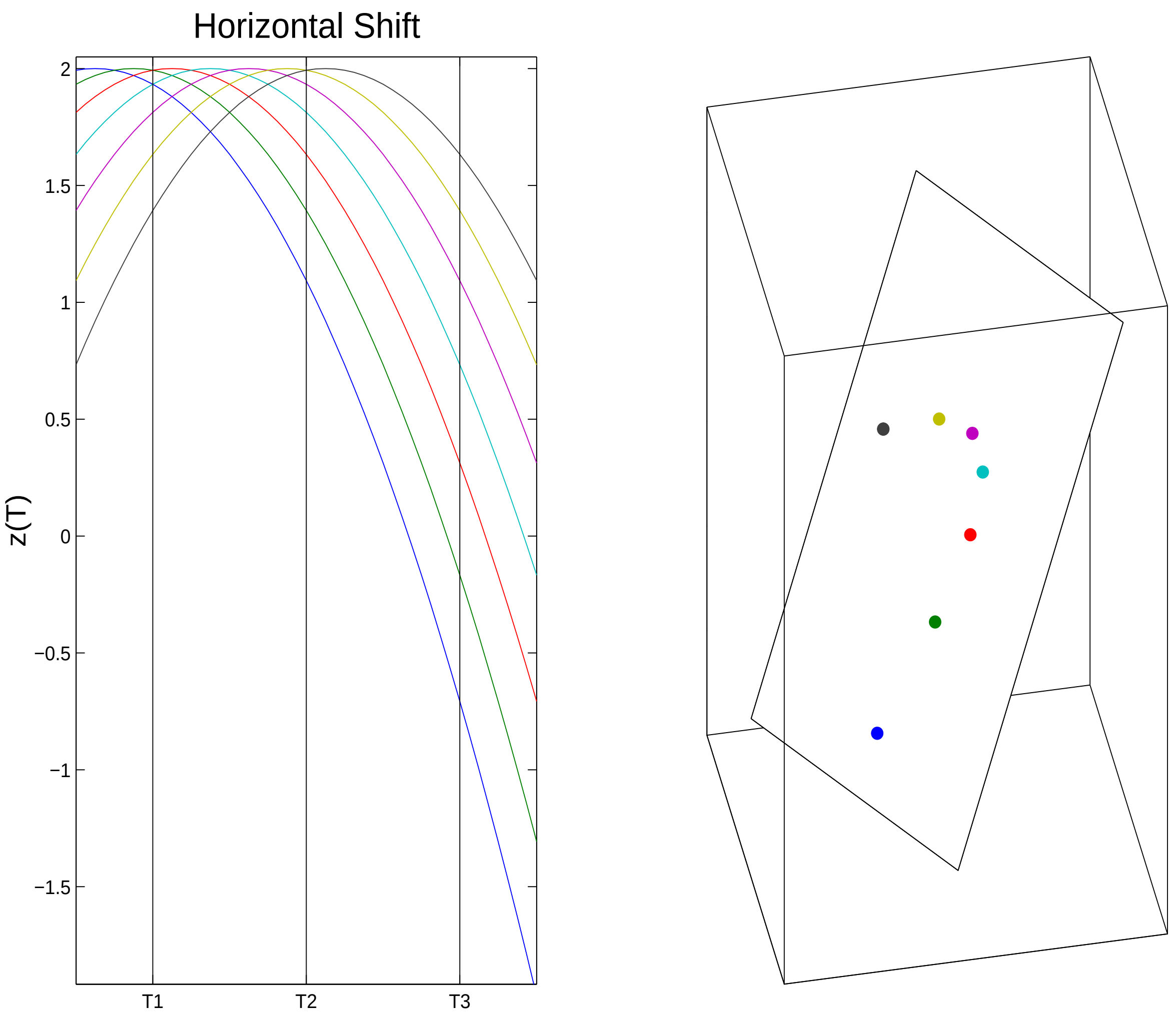
Toy example

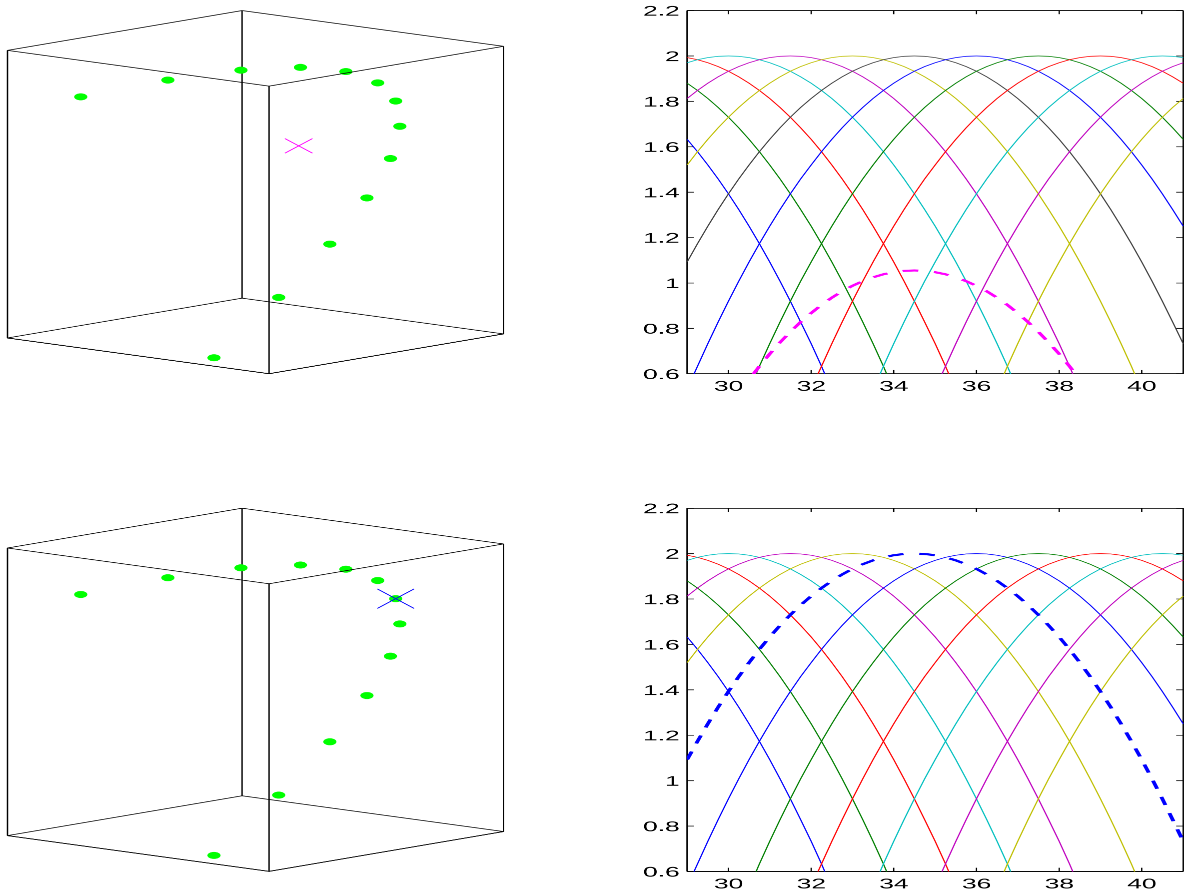
- $f(t) = -t^2 + c$.
- Model,

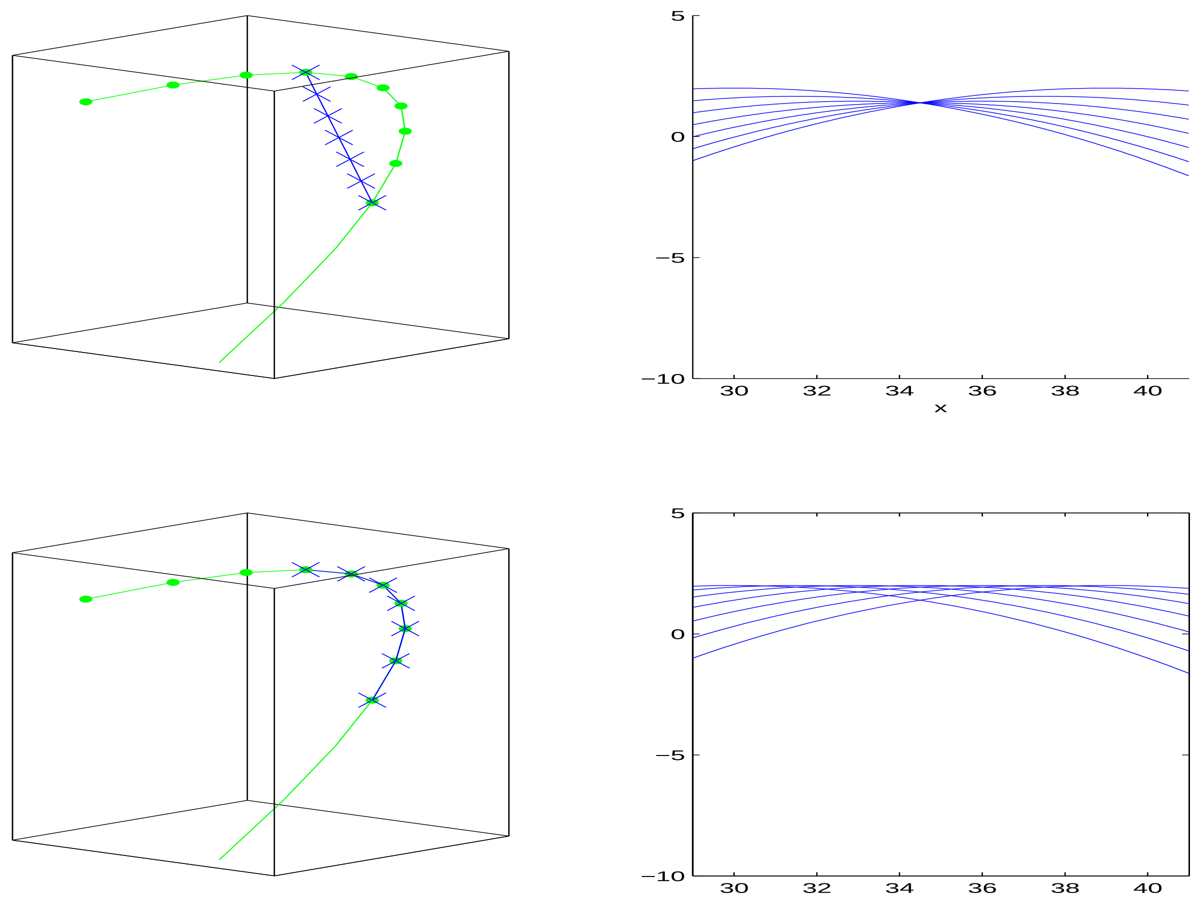
$$\begin{aligned} z_{i,j} &= w_i f(w_i(t_j - m_i)) + h_i + \epsilon_{i,j} \\ &= -w_i^3(t_j - m_i)^2 + w_i c + h_i + \epsilon_{i,j}. \end{aligned}$$


Vertical Shift
z(T)
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
T1
T2
T3




Horizontal Shift
z(T)
2
1.5
1
0.5
0
−0.5
−1
−1.5
T1
T2
T3


2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
30
32
34
36
38
40


5
0
−5
−10
30
32
34
36
38
40
x
5
0
−5
−10
30
32
34
36
38
40

Linear case (conventional PCA), one mode

- Center of the projections: $\bar{R}$, and $\bar{R} = \bar{Z}$.
- Spread: $SSM = \sum_{i=1}^{n} ||R_i - \bar{Z}||^2$.
- $RSS = \frac{SSM}{SSM+SSE}$ quantifies the variation.

Nonlinear case (Template Mode of Variation), one mode

- $\widetilde{R}$ is the sample geodesic mean.
- d_g is the geometric distance.
- $\widetilde{RSS}$ quantifies nonlinear variations.

One mode of variation

Linear (PCA)	Nonlinear (TMV)
$RSS = \frac{SSM}{SSM+SSE}$	$\widetilde{RSS} = \frac{\widetilde{SSM}}{\widetilde{SSM}+SSE}$
$SSM = \sum_{i=1}^{n} \|\|R_i - \bar{Z}\|\|^2$	$\widetilde{SSM} = \sum_{i=1}^{n} \left(d_g(R_i, \widetilde{R})\right)^2$
$\bar{Z} = \bar{R}$	$\bar{Z} \neq \widetilde{R}$
$SSM \perp SSE$	$\widetilde{SSM}$(not $\perp$)SSE

More than one mode of variation

- Data fall in a manifold $d' \geq 2$: define new d_g
- Decompose model variation.
- Quantify each variation.

Two modes of variation: linear + nonlinear

- (h, m) movie. Define

$$d_g((h_1, m_1), (h_2, m_2)) = \sqrt{d_m(m_1, m_2)^2 + (h_1 - h_2)^2}.$$

- (h, w) movie. Define

$$d_g((h_1, w_1), (h_2, w_2)) = \sqrt{d_w(w_1, w_2)^2 + (h_1 - h_2)^2}.$$

Two modes of variations: nonlinear + nonlinear

(w, m) movie, define

$$d_g((w_1, m_1), (w_2, m_2)) = \sqrt{\frac{1}{2}\left(M^2 + W^2\right)}, \text{ where}$$

$$\begin{aligned} M^2 &= d_{m,(w=w_1)}(m_1, m_2)^2 + d_{m,(w=w_2)}(m_1, m_2)^2 \\ W^2 &= d_{w,(m=m_1)}(w_1, w_2)^2 + d_{w,(m=m_2)}(w_1, w_2)^2 \end{aligned}$$

Two modes of variations (w, m) (contd)

Other decomposition,

$$d_g((w_1, m_1), (w_2, m_2)) = \sqrt{\frac{1}{4}\left(D_m + D_w + Diff_m + Diff_w\right)},$$ where

$$\begin{aligned}
D_m &= \left(d_{m,(w=w_1)}(m_1, m_2) + d_{m,(w=w_2)}(m_1, m_2)\right)^2 \\
D_w &= \left(d_{w,(m=m_1)}(w_1, w_2) + d_{w,(m=m_2)}(w_1, w_2)\right)^2 \\
Diff_m &= \left(d_{m,(w=w_1)}(m_1, m_2) - d_{m,(w=w_2)}(m_1, m_2)\right)^2 \\
Diff_w &= \left(d_{w,(m=m_1)}(w_1, w_2) - d_{w,(m=m_2)}(w_1, w_2)\right)^2
\end{aligned}$$

Three modes of variation (w,m,h)

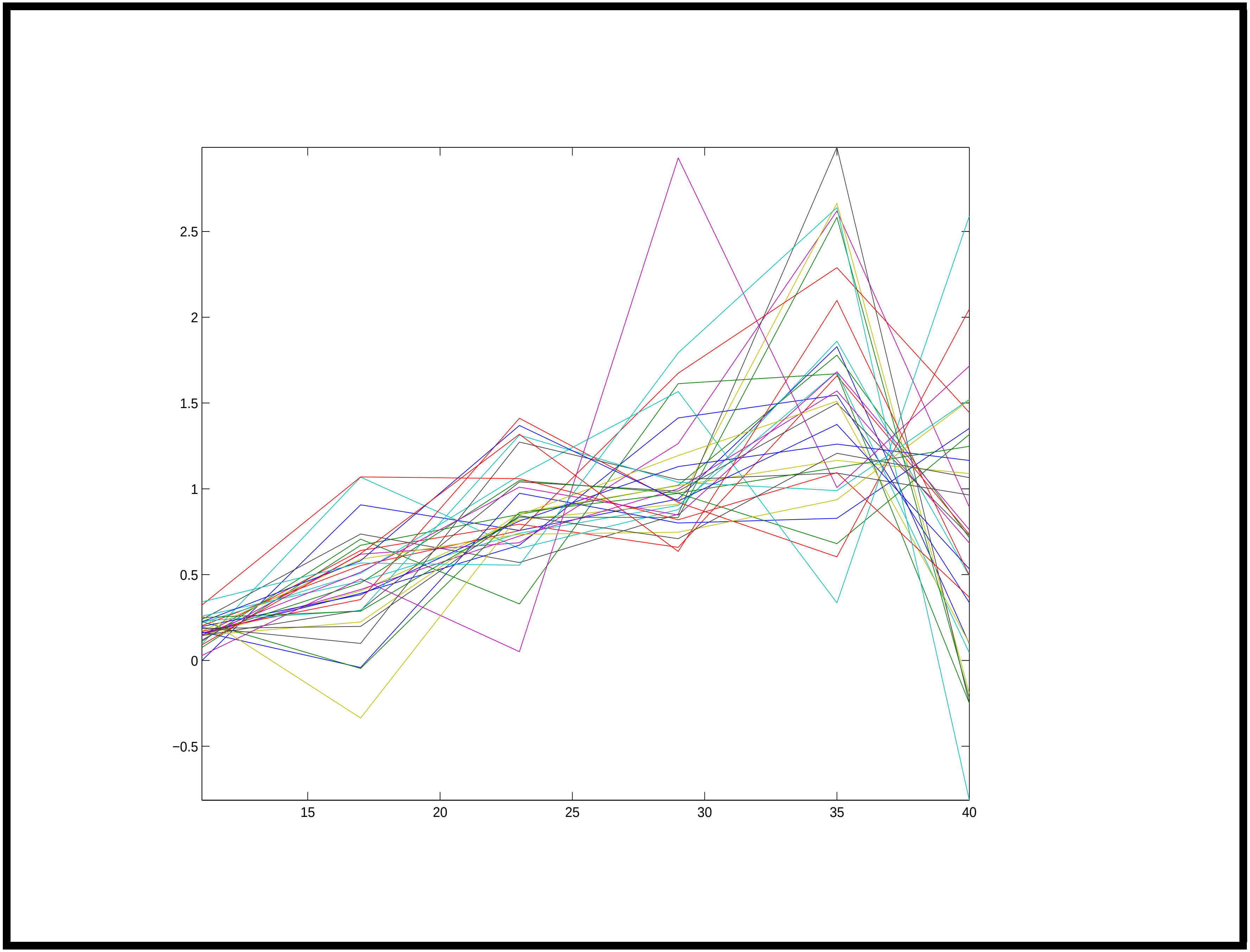
$$d_g((w_1, m_1, h_1), (w_2, m_2, h_2)) = \sqrt{d_g((w_1, m_1), (w_2, m_2))^2 + (h_1 - h_2)^2}$$

Geodesic mean, SS, and RSS

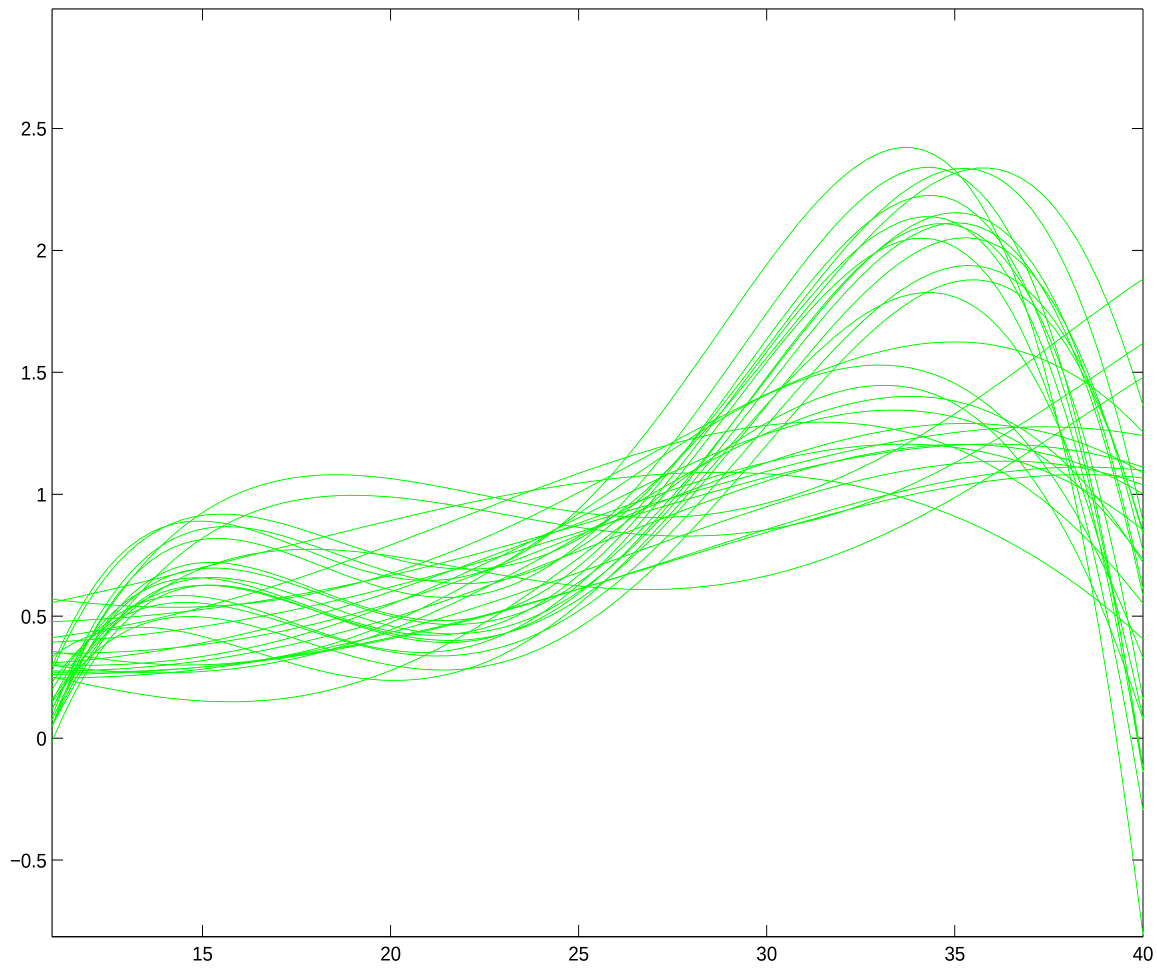
- Geodesic sample mean $\tilde{R} \doteq \text{Argmin}_a \sum_i d_g(a, R_i)^2$.
- Spread: $\widetilde{SSM} = \sum_i d_g(\tilde{R}, R_i)^2$.
- $\widetilde{RSS} = \frac{\widetilde{SSM}}{\widetilde{SSM}+SSE}$.

Outline

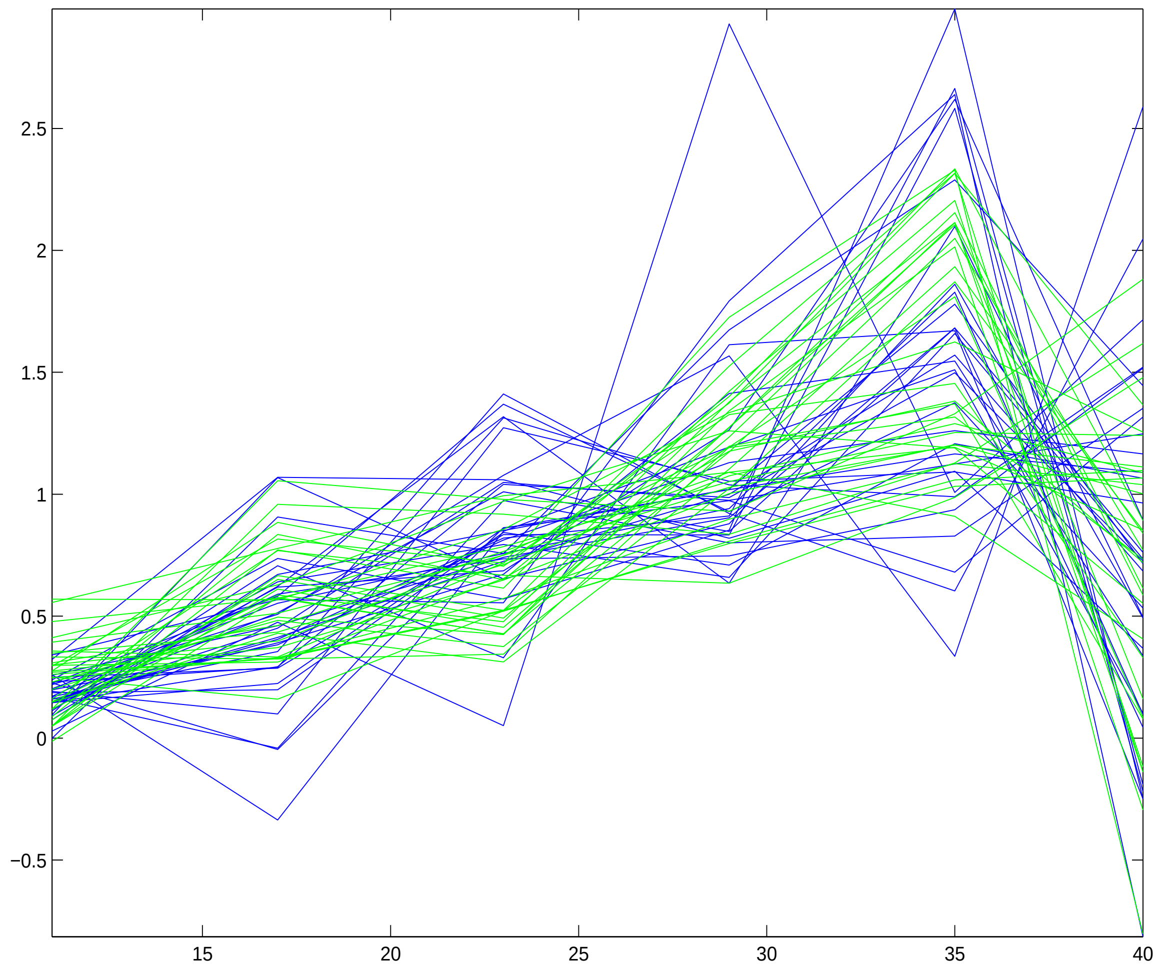
- Data
- Motivation and model
- Toy example
- Results

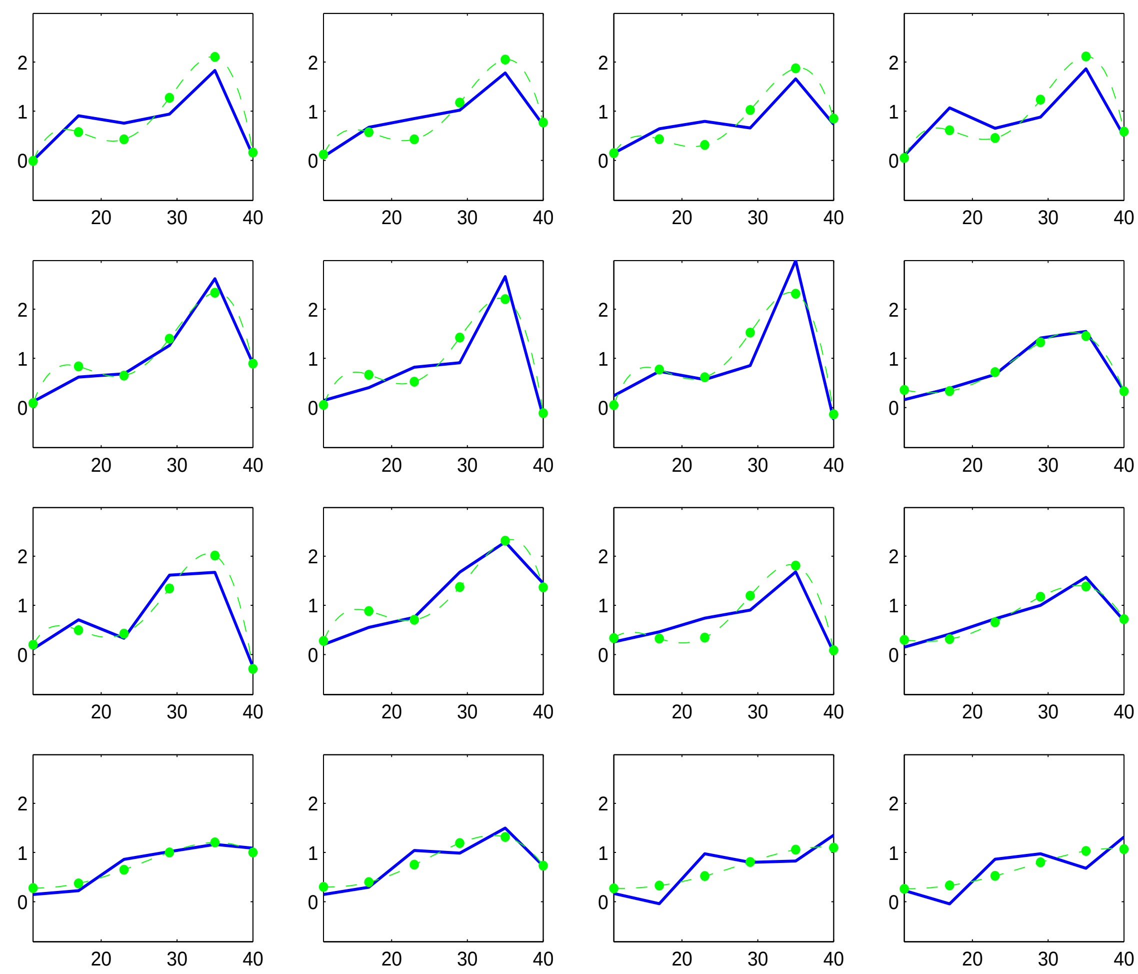


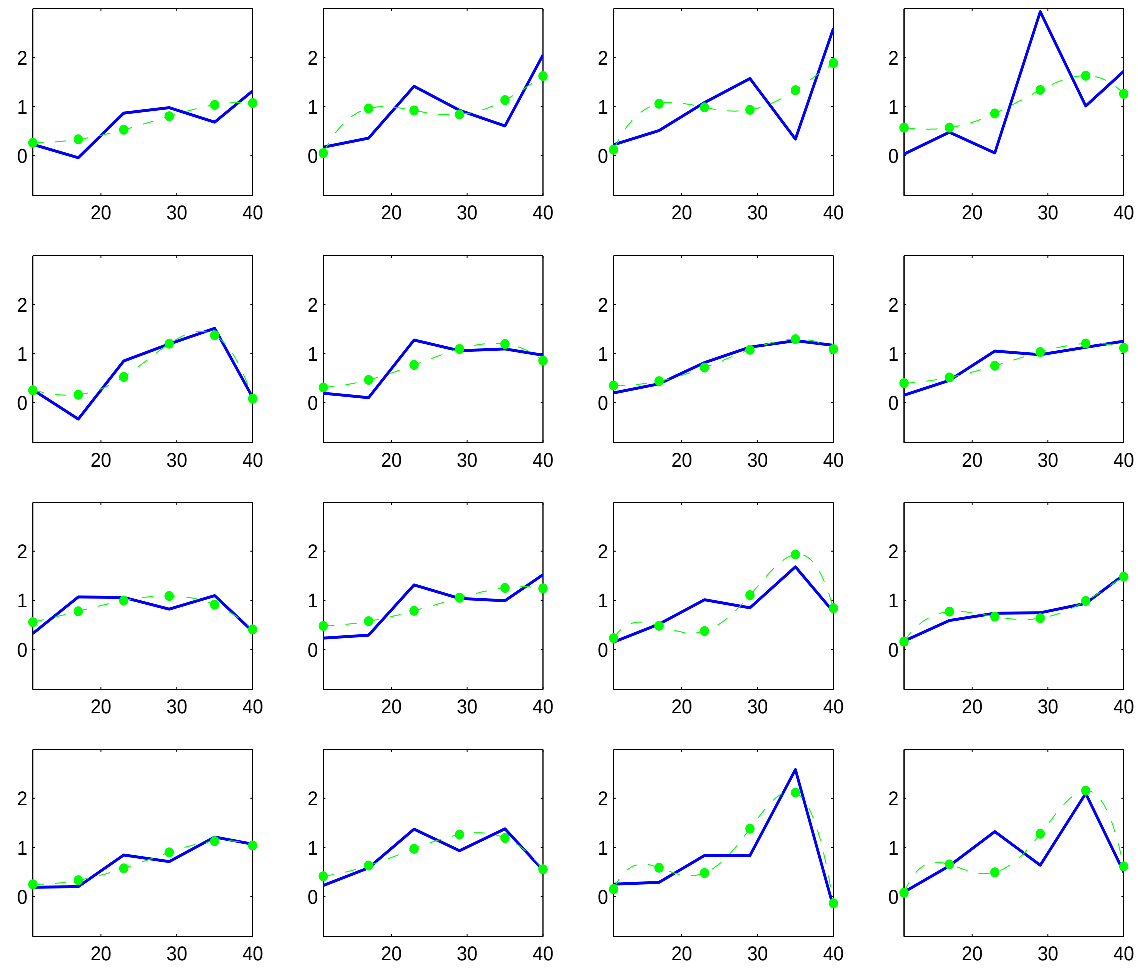
2.5
2
1.5
1
0.5
0
−0.5
15
20
25
30
35
40

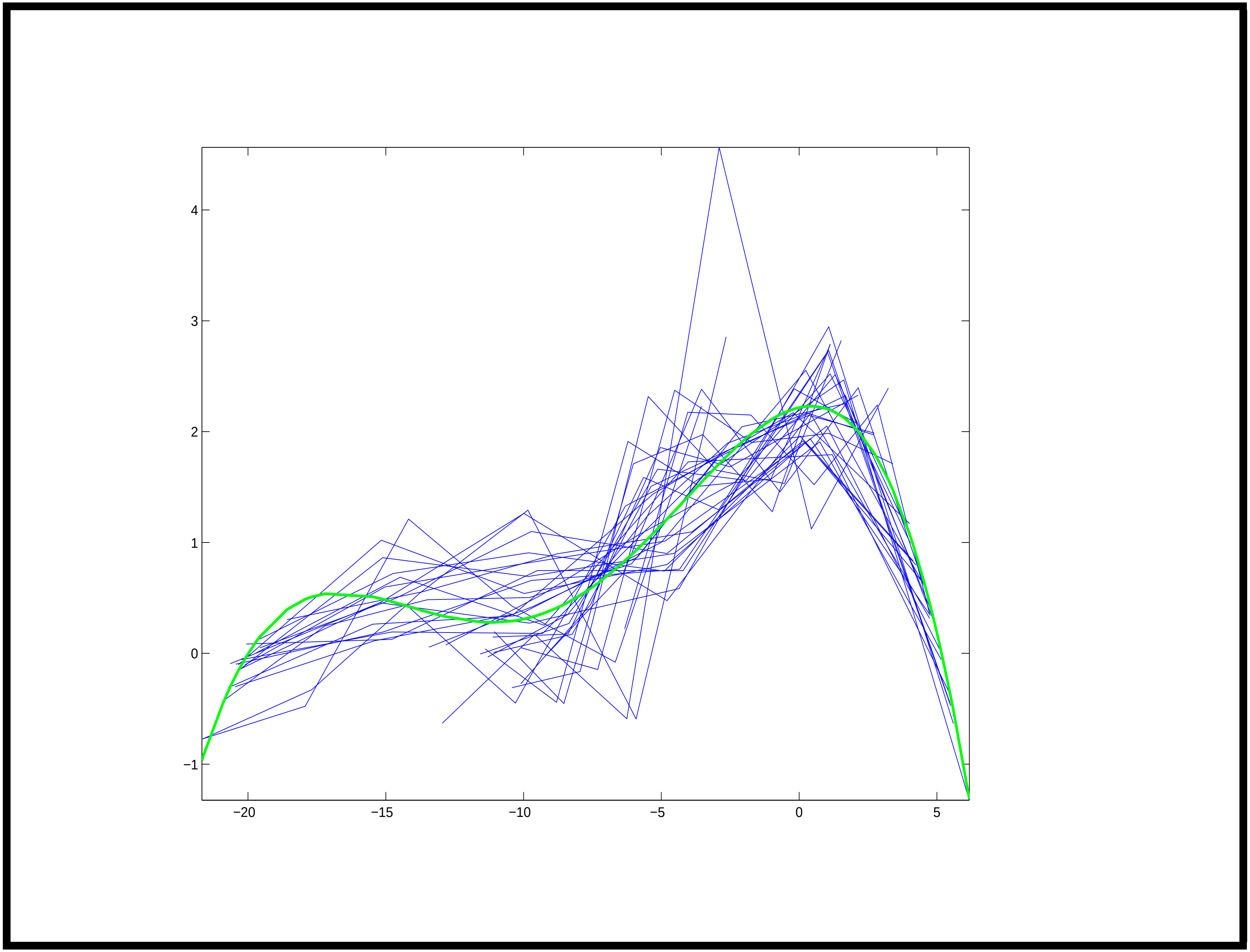


2.5
2
1.5
1
0.5
0
−0.5
15
20
25
30
35
40


2.5
2
1.5
1
0.5
0
−0.5
15
20
25
30
35
40


0 1 2
20 30 40


2
1
0
20
30
40



4
3
2
1
0
−1
−20
−15
−10
−5
0
5

Mode of variation	$\widetilde{RSS}$
Generalist Specialist	43.40%
Horizontal Shift	20.85%
Vertical Shift	7.36%
Horiz-Gen.Spec	3.78%
Model total	75.28%
Error	24.72%

Conclusion for Evolutionary Biology: Selecting for higher growth rate in a range of temperature will not result in higher growth rate at all temperatures.

In Functional Data Analysis: Assuming a common shape of curves, the Template Modes of Variation allows to decompose curves' variance into linear and nonlinear modes and to quantify each mode. Linear modes are a particular case of this method of decomposition.

Future work: for $d' \geq 2$, use Differential Geometry theory

- Convergence of sample geodesic mean to the geodesic mean.
- Conditions of existence and uniqueness of $\widetilde{R}$.
- Sensitivity analysis of the decomposition to common shape.

THANK YOU