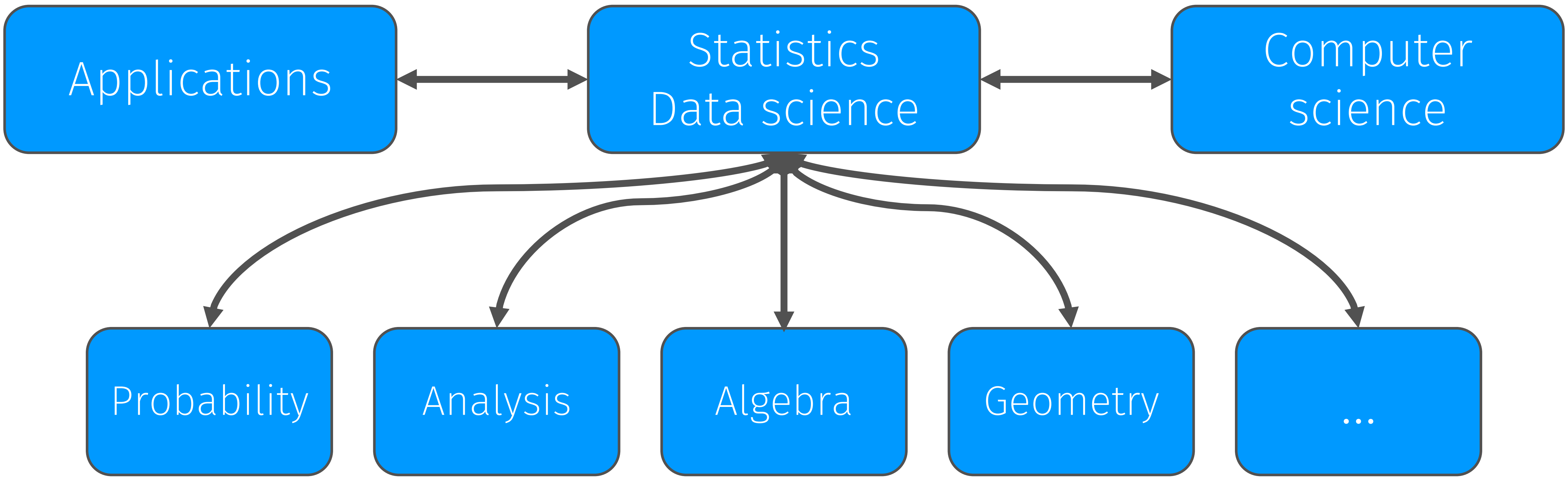
Why does a statistician need Analysis?

Some insights on the covariance operator

Steven Golovkine

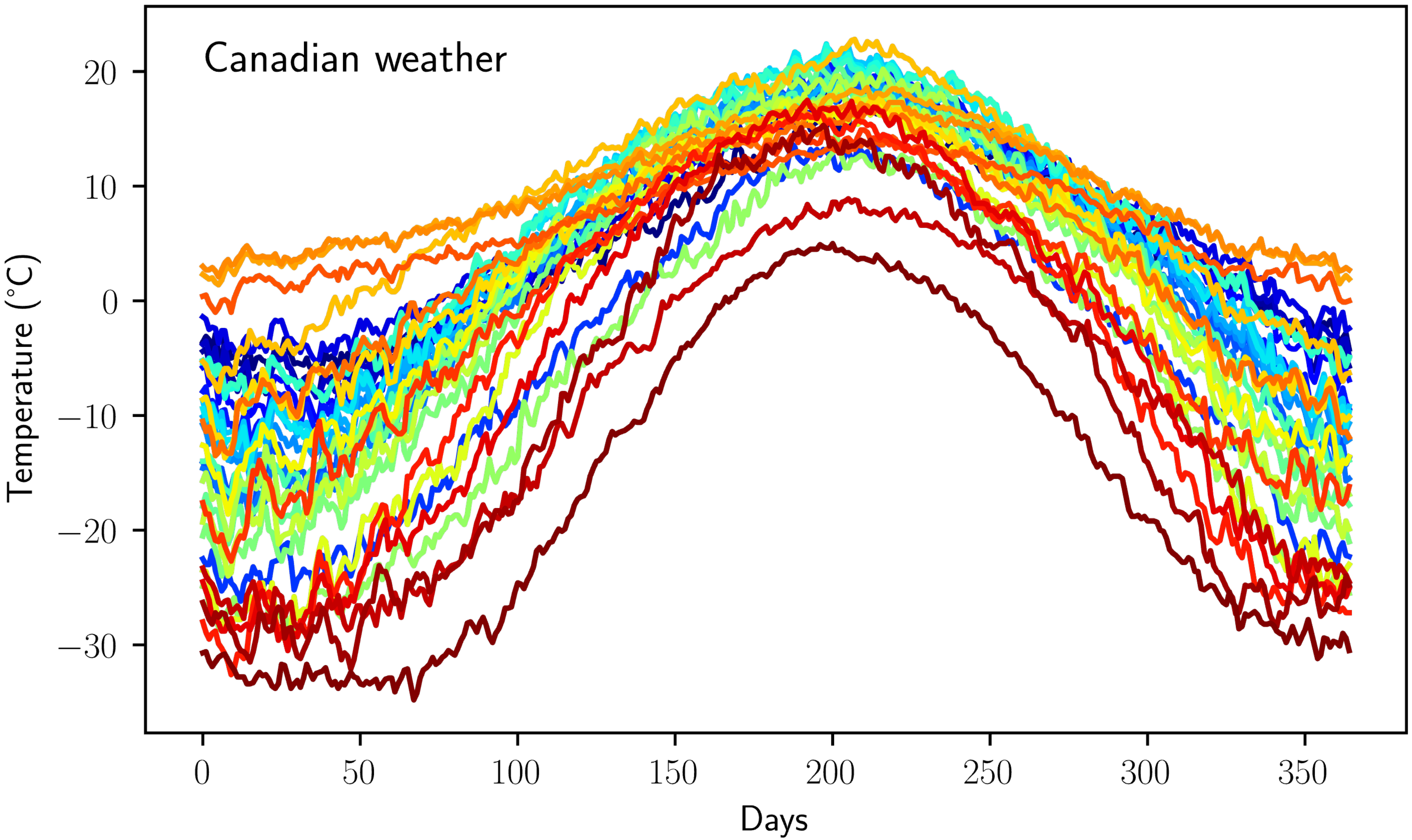
May 13th, 2025

UNIVERSITÉ LAVAL
Faculté des sciences et de génie
Département de mathématiques et de statistique

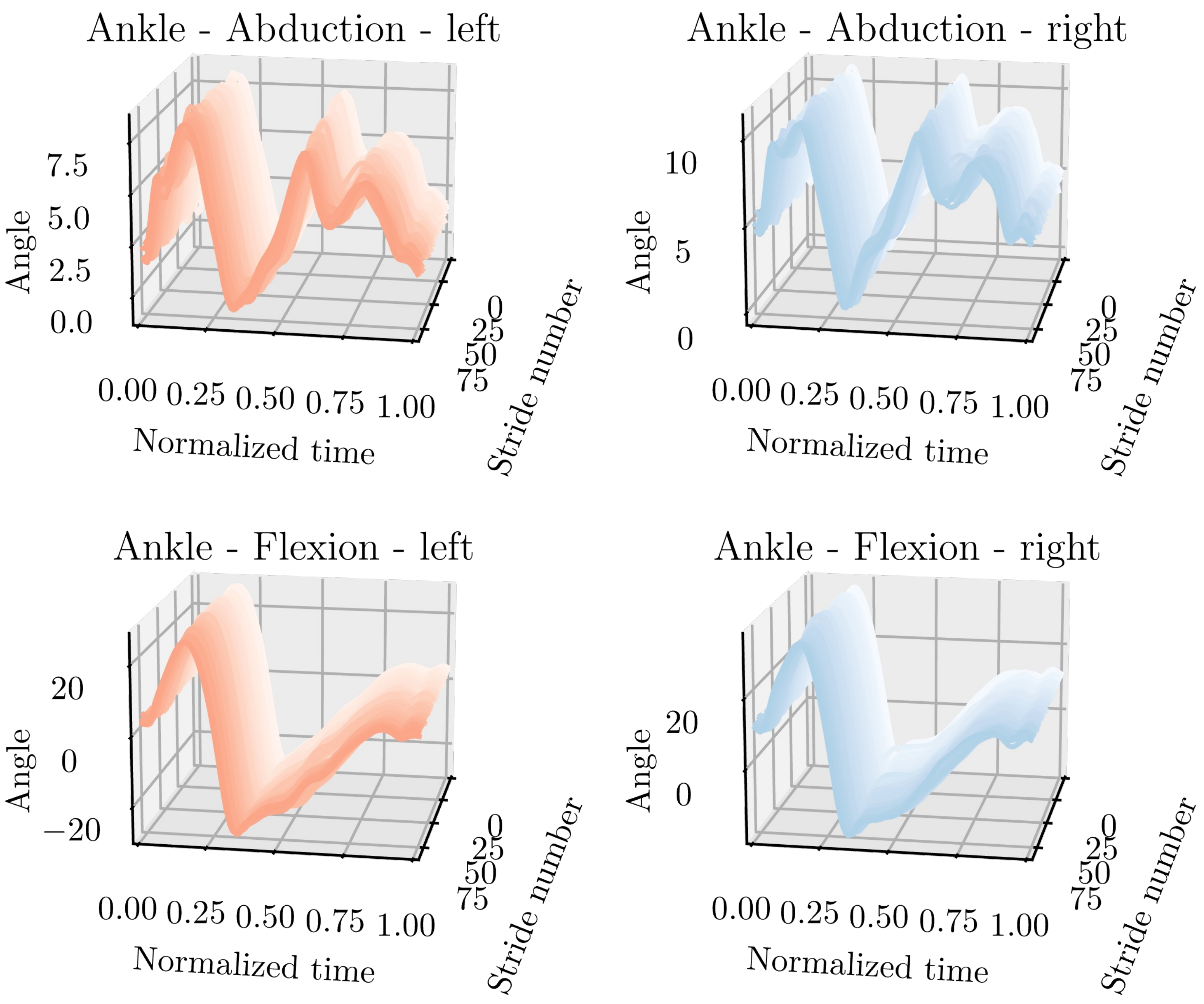


The best thing about being a statistician is that you get to play in everyone's backyard. — John Tukey[1]

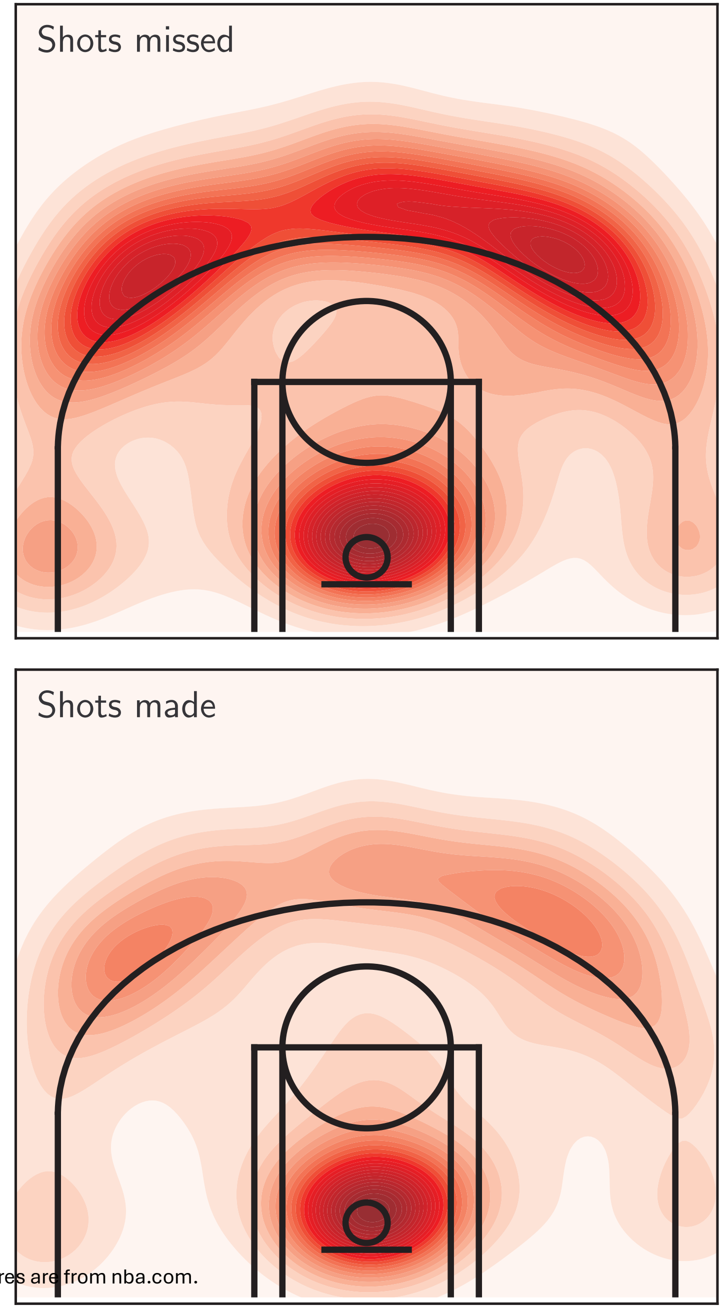
[1]see www.princeton.edu/pr/news/00/q3/0727-tukey.htm or The New York Times, July 28, 2000.



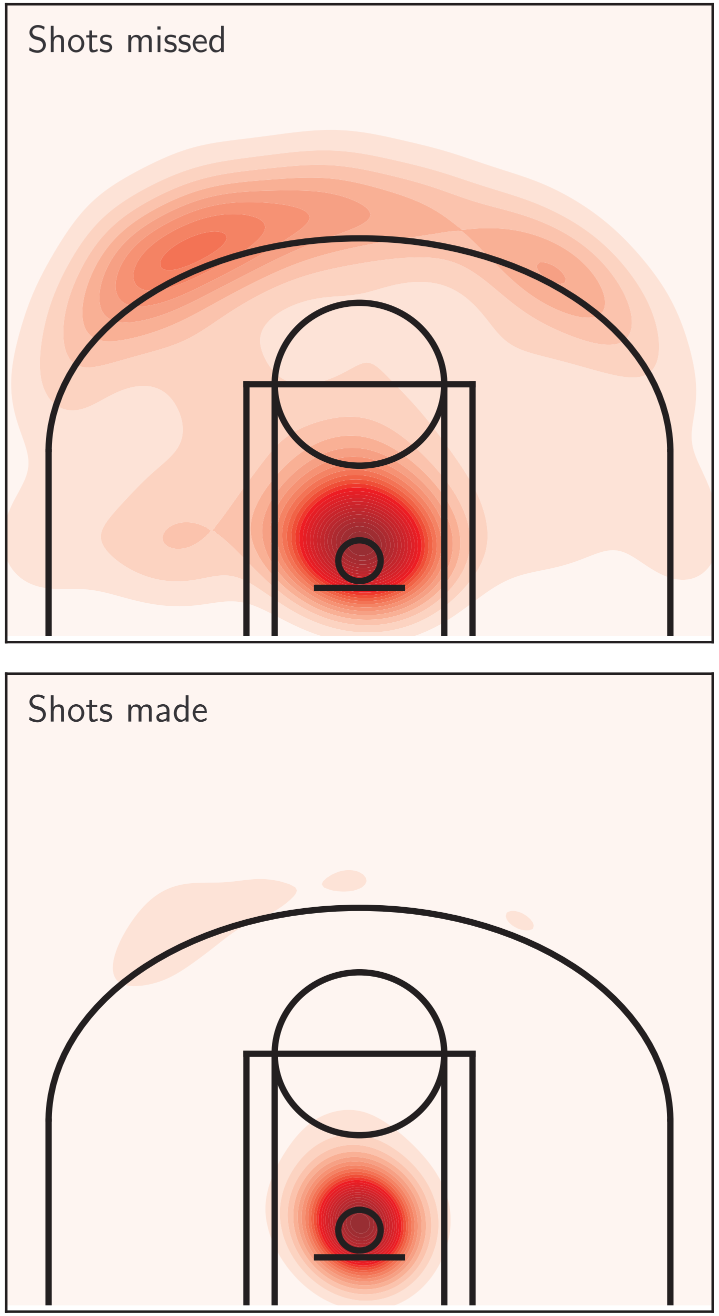
Canadian weather
Temperature (°C)
Days
20
10
0
−10
−20
−30
0
50
100
150
200
250
300
350


Ankle - Abduction - left
Angle
7.5
5.0
2.5
0.0
0.00 0.25 0.50 0.75 1.00
Normalized time
0
25
50
75
Stride number
Ankle - Abduction - right
Angle
10
5
0
0.00 0.25 0.50 0.75 1.00
Normalized time
0
25
50
75
Stride number
Ankle - Flexion - left
Angle
20
0
−20
0.00 0.25 0.50 0.75 1.00
Normalized time
0
25
50
75
Stride number
Ankle - Flexion - right
Angle
20
0
0.00 0.25 0.50 0.75 1.00
Normalized time
0
25
50
75
Stride number

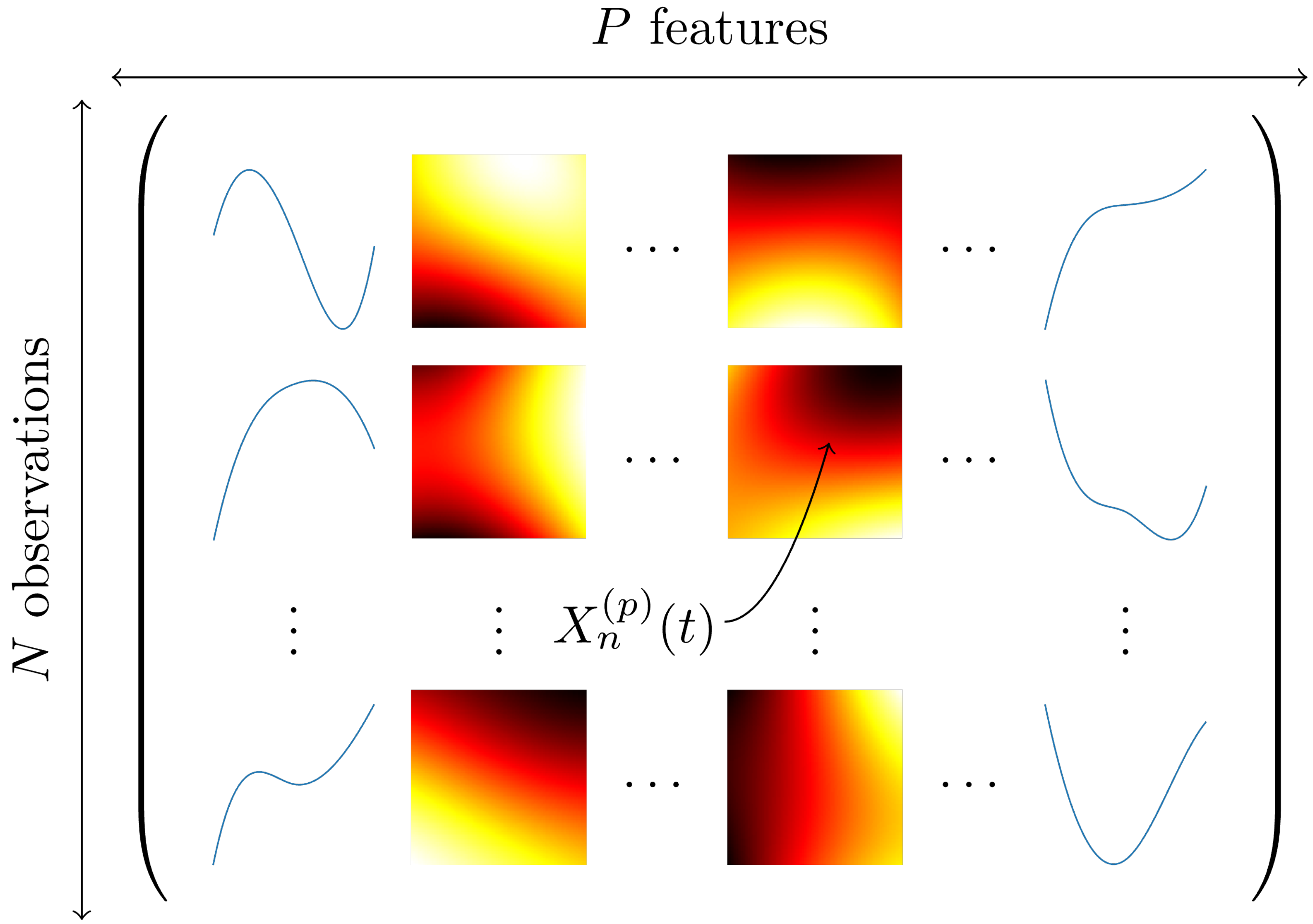
Stephen Curry - Guard



LeBron James - Forward



Shots density are from the authors. Player's pictures are from nba.com.



P features
N observations
$X_n^{(p)}(t)$

$$\mathcal{H} = \underbrace{\mathcal{L}^2(\mathcal{T}_1) \times \cdots \times \mathcal{L}^2(\mathcal{T}_P)}_{P \text{ terms}}, \quad \mathcal{T}_p \subset \mathbb{R}^d, \quad p = 1, \ldots, P.$$

$$\langle\langle f, g \rangle\rangle = \sum_{p=1}^{P} \int_{\mathcal{T}_p} f^{(p)}(t_p) g^{(p)}(t_p) \mathrm{d}t_p, \quad f, g \in \mathcal{H}.$$

Happ, C. and Greven, S., Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains (2018).

For a stochastic process X

$$\mu(t) = \mathbb{E}(X(t)), \quad t \in \mathcal{T}$$

$$C(s,t) = \mathbb{E}\left(\{X(s) - \mu(s)\}\{X(t) - \mu(t)\}\right), \quad s,t \in \mathcal{T}$$

The covariance operator Γ is an integral operator with kernel C

$$(\Gamma f)^{(p)}(t_p) = \langle\langle C_{p,\cdot}(t_p, \cdot), f(\cdot) \rangle\rangle, \quad t_p \in \mathcal{T}_p$$

Happ, C. and Greven, S., Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains (2018).

Hilbert-Schmidt theorem

$$\Gamma\phi_k = \lambda_k\phi_k \quad \text{and} \quad \lambda_k \to 0 \quad \text{for} \quad k \to \infty$$

Happ, C. and Greven, S., Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains (2018).

Mercer's theorem

$$C_{p,p}(s,t) = \sum_{k=1}^{\infty} \lambda_k \phi_k^{(p)}(s) \phi_k^{(p)}(t)$$

Happ, C. and Greven, S., Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains (2018).

Karhunen-Loève decomposition

$$X(t) = \sum_{k=1}^{\infty} c_k \phi_k(t) \quad \text{with} \quad c_k = \langle\!\langle X, \phi_k \rangle\!\rangle$$

Happ, C. and Greven, S., Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains (2018).

Estimation

1. For each components :
   - Estimate the univariate components and scores.
2. Concatenate the scores.
3. Perform an eigenanalysis of the covariance of the scores.
4. Estimate the multivariate components.

Happ, C. and Greven, S., Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains (2018).

Gram matrix

$$\mathbf{M} = \Big[\langle\!\langle X_i, X_j \rangle\!\rangle \Big]_{i,j=1,\ldots,N}$$

Duality diagram

$$L_X : f \mapsto \begin{pmatrix} \langle\!\langle X_1, f \rangle\!\rangle \\ \vdots \\ \langle\!\langle X_N, f \rangle\!\rangle \end{pmatrix} \qquad L_X^\star : u \mapsto \begin{pmatrix} \sum_{n=1}^{N} u_n X_n^{(1)}(t_1) \\ \vdots \\ \sum_{n=1}^{N} u_n X_n^{(P)}(t_P) \end{pmatrix}$$

Golovkine, S. et al., On the use of the Gram matrix for multivariate functional principal components analysis (2025).

$$
\begin{tikzcd}
\mathcal{H}^\star \arrow[r, "L_X"] & \mathbb{R}^N \arrow[d, "\mathbf{M}"] \arrow[dl, bend left, "L_X^\star(\mathbf{M}u)"] \\
\mathcal{H} \arrow[u, "\Gamma"] \arrow[ur, bend left, "L_X(\Gamma f)"] & \mathbb{R}^{N\star} \arrow[l, "L_X^\star"]
\end{tikzcd}
$$

Golovkine, S. et al., On the use of the Gram matrix for multivariate functional principal components analysis (2025).

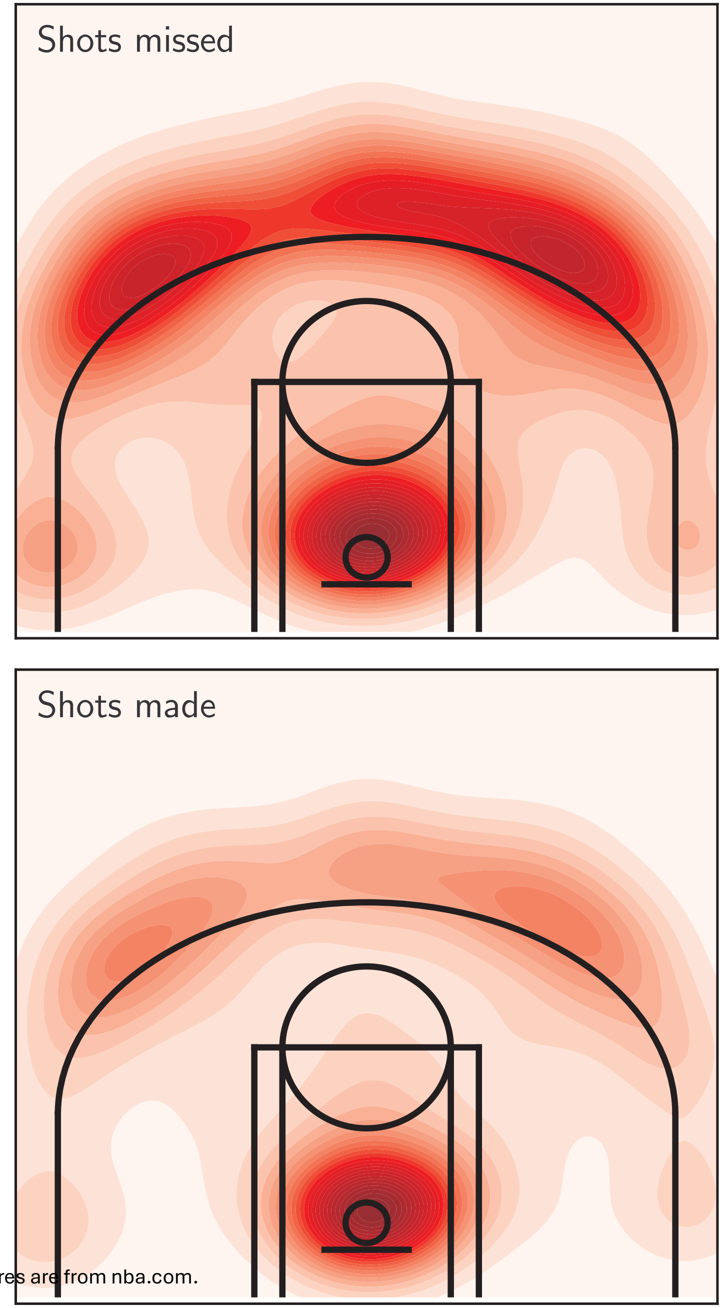
Relationships between $\mathbf{\Gamma}$ and $\mathbf{M}$

$$\lambda_k = l_k, \quad k = 1, 2, \ldots, N$$

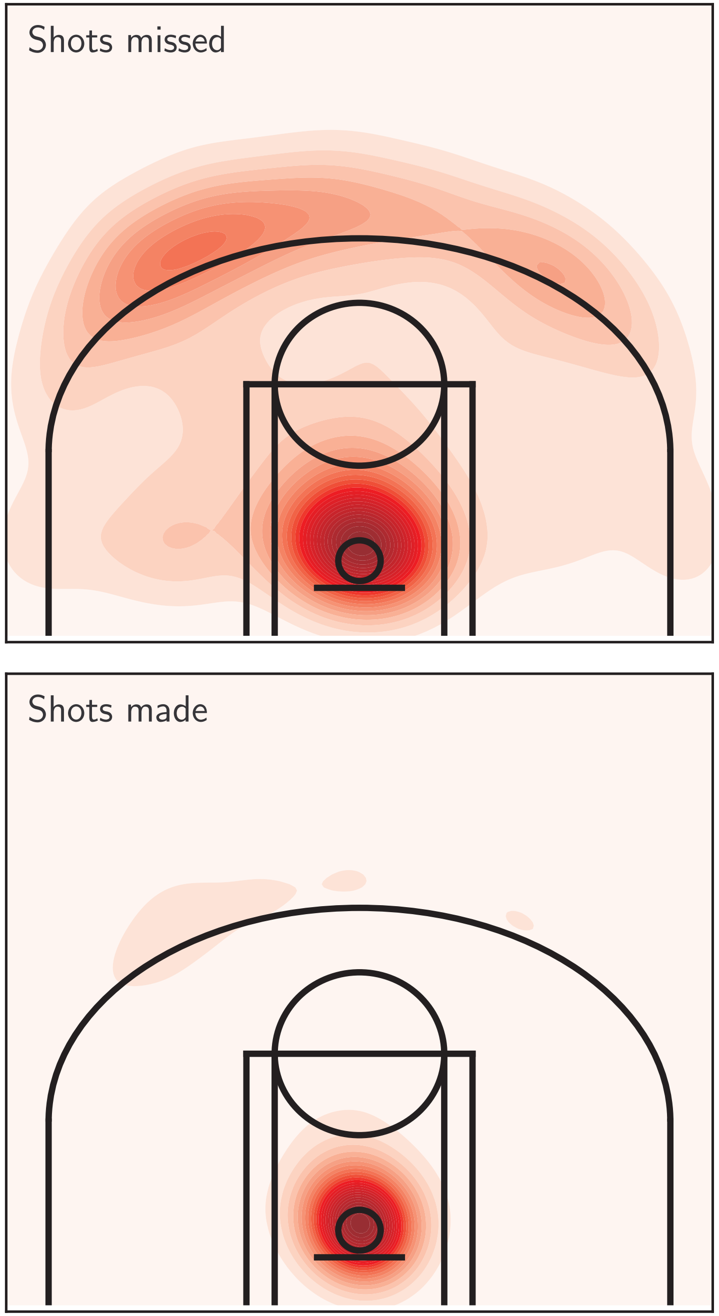
$$\phi_k(\cdot) = \frac{1}{\sqrt{N l_k}} \sum_{n=1}^{N} [\boldsymbol{u}_k]_n X_n(\cdot), \quad k = 1, 2, \ldots, N$$

$$c_{nk} = \sqrt{N l_k} [\boldsymbol{u}_k]_n, \quad n = 1, 2, \ldots, N, \quad k = 1, 2, \ldots, N$$

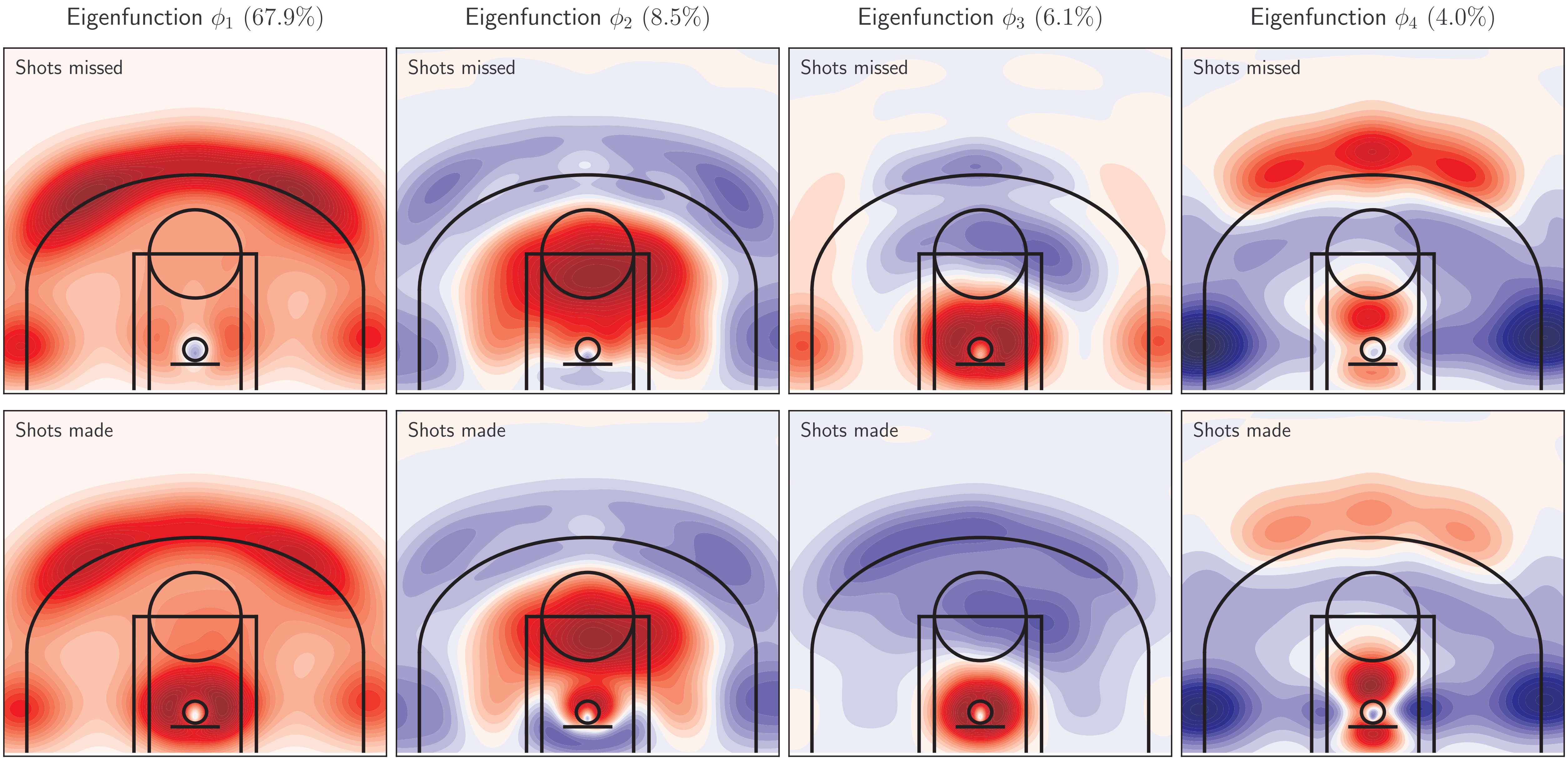
Stephen Curry - Guard



LeBron James - Forward



Shots density are from the authors. Player's pictures are from nba.com.



Pictures are from the authors.

RKHS

$$\mathcal{H}(C) = \{\Gamma^{1/2}(f), f \in \mathcal{H}\}$$

$$\mathcal{H}(C) = \left\{ f(\cdot) = \sum_{k=1}^{\infty} \alpha_k \sqrt{\lambda_k} \phi_k(\cdot), \quad \{\alpha_k\}_{k=1}^{\infty} \in \ell_2(\mathbb{N}) \right\}$$

Larraz, B., On reproducing kernel methods in functional statistics (2018).

The spectral density operator $\mathcal{F}$

with kernel

$$f_\theta(s,t) = \frac{1}{2}\sum_{h\in\mathbb{Z}} C_h(s,t)\exp(-ih\theta)$$

Differential operators

$$L = w_0(t)I + w_1(t)D + \cdots + w_{q-1}(t)D^{q-1} + D^q(t)$$

$$LX(t) \approx 0$$

Lots of stuff to do in FDA

and we need analysis for that…