STABLE COMPARISON OF TIME SERIES USING TOPOLOGY

Bala Krishnamoorthy Elizabeth Thompson
Washington State University

arXiv: 2501.02817

WADEPS

COMPARING TIME SERIES

→ stock price v/s volume of trading

(image: www)

COMPARING TIME SERIES

→ stock price v/s volume of trading

(image: www)

→ rainfall amount v/s river levels

(image: www)

COMPARING TIME SERIES

→ stock price v/s volume of trading

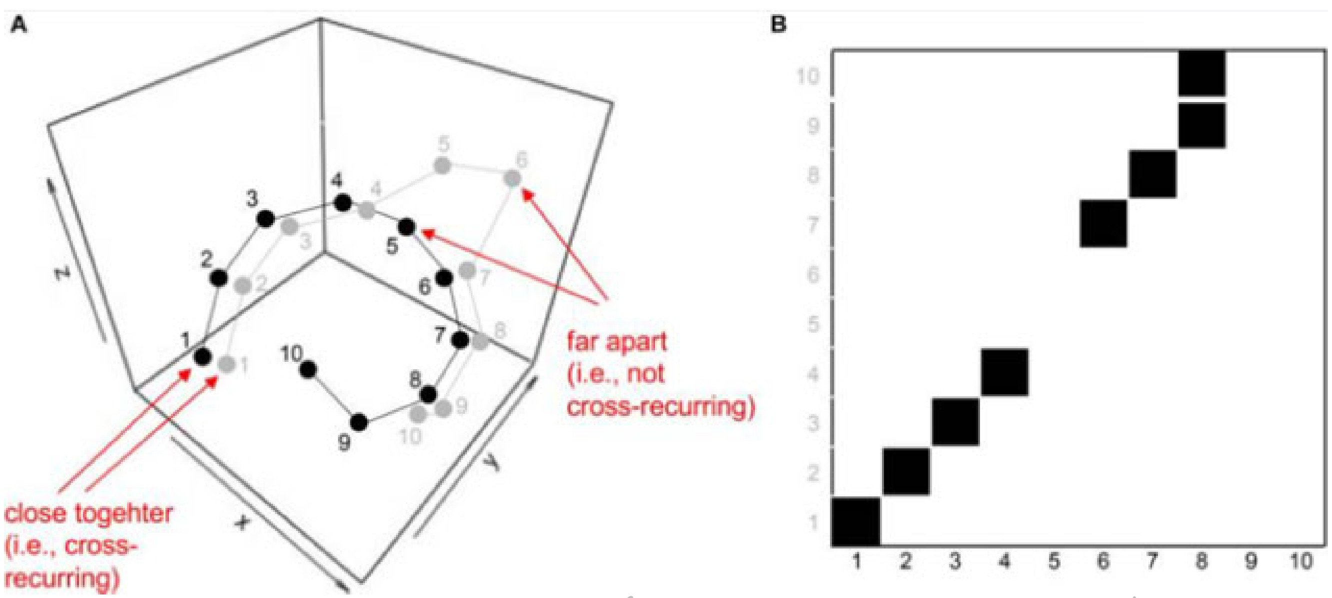
(image: www)

→ rainfall amount v/s river levels

(image: www)

? How to compare?

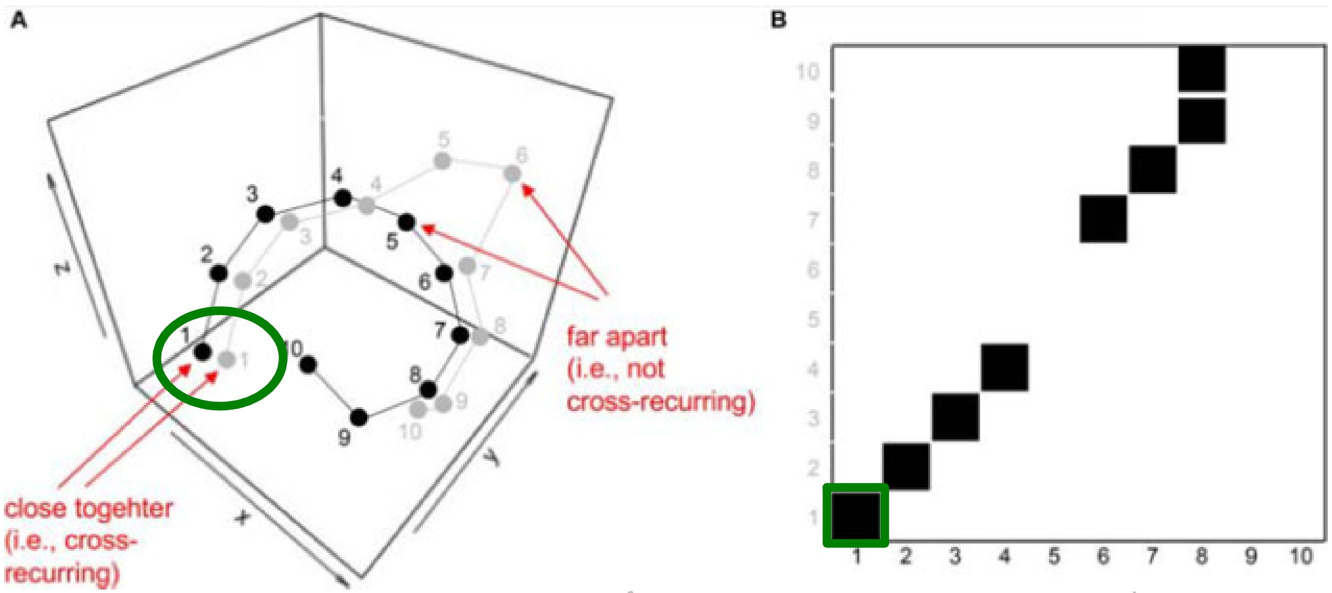
CROSS-RECURRENCE (%DET)



Cross-recurrence matrix $\mathcal{C}$

(Wallot & Leonardi, 2018)

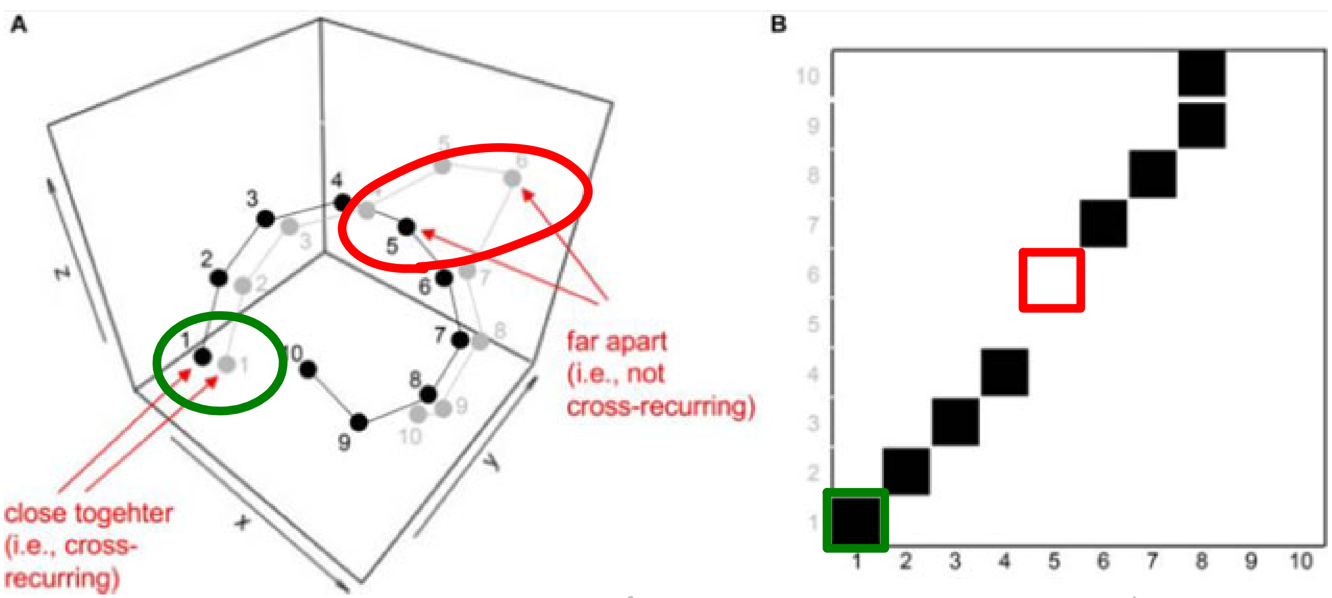
Cross-Recurrence (%DET)



Cross-recurrence matrix $\mathcal{C}$

(Wallot & Leonardi, 2018)

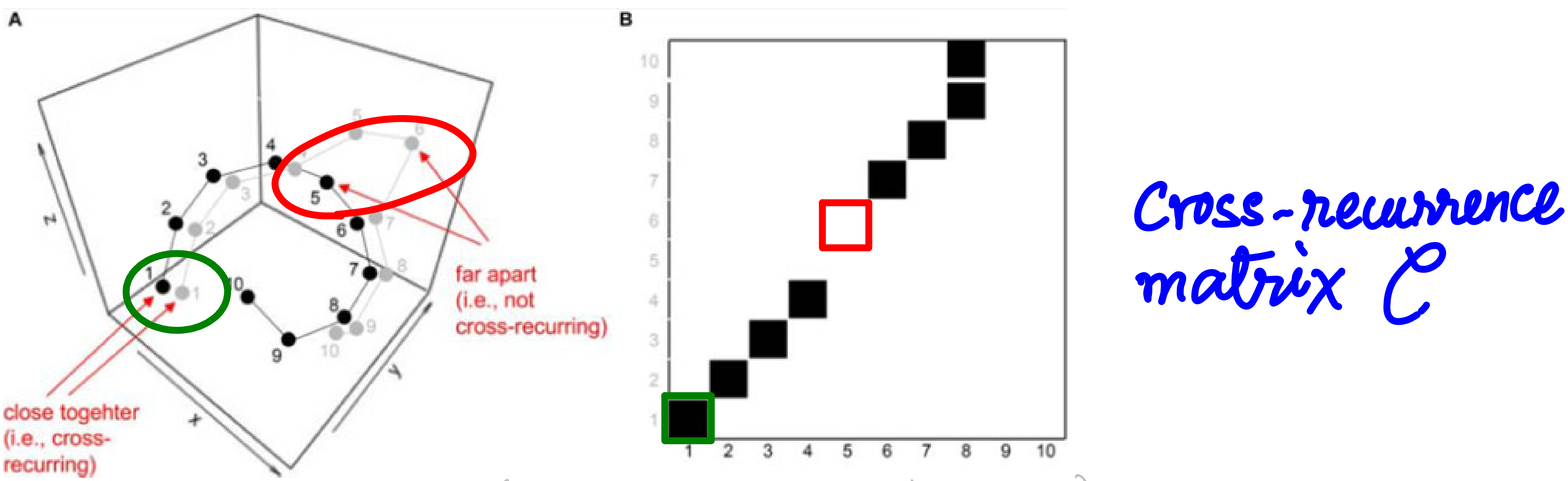
CROSS-RECURRENCE (%DET)



(Wallot & Leonardi, 2018)

Cross-recurrence matrix $\mathcal{C}$

CROSS-RECURRENCE (%DET)

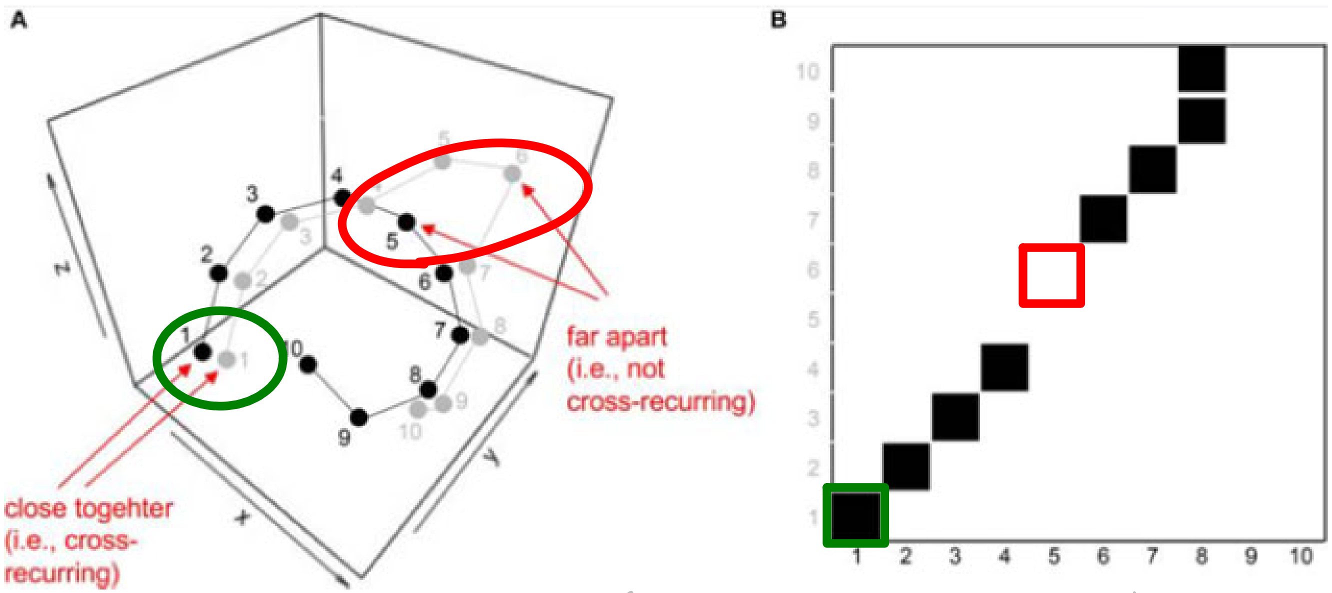


(Wallot & Leonardi, 2018)

→ % Determinism (%DET)

– % of 1's in diagonal strips of $\mathcal{C}$

CROSS-RECURRENCE (%DET)



(Wallot & Leonardi, 2018)

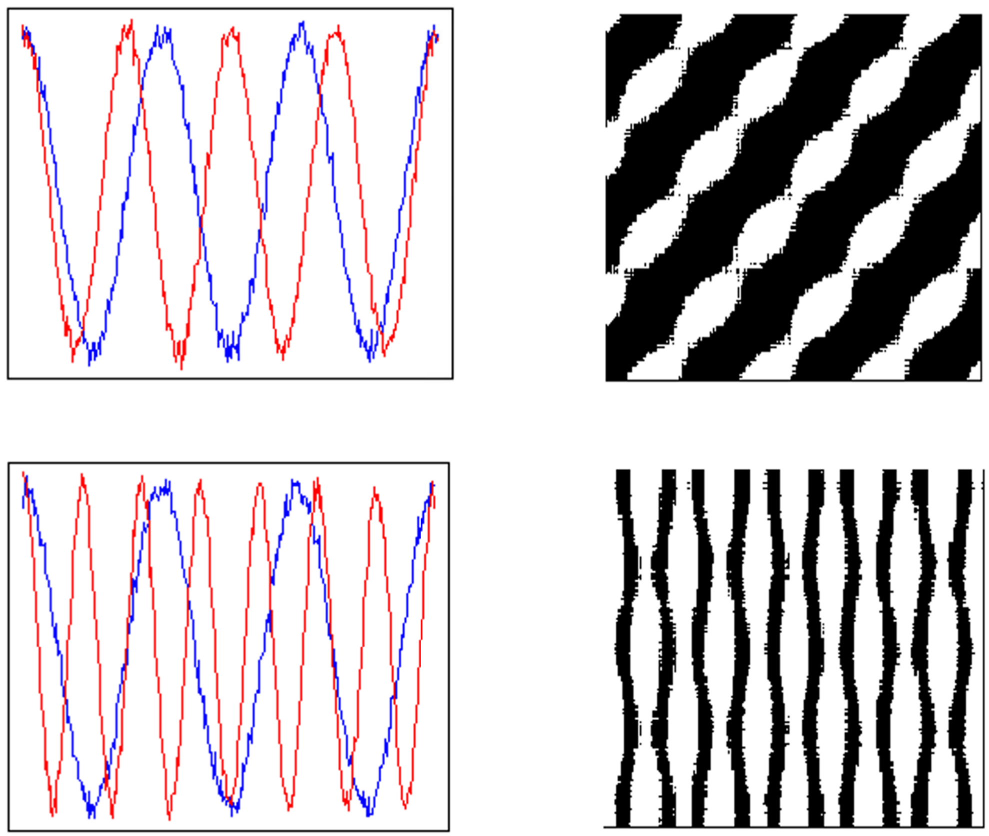
Cross-recurrence matrix $\mathcal{C}$

→ % Determinism (%DET)

– % of 1's in diagonal strips of $\mathcal{C}$

→ four parameters: time lag τ, embed. dimension, dist. threshold (tol), # diagonal strips (min DL)

%DET: INSTABILITY

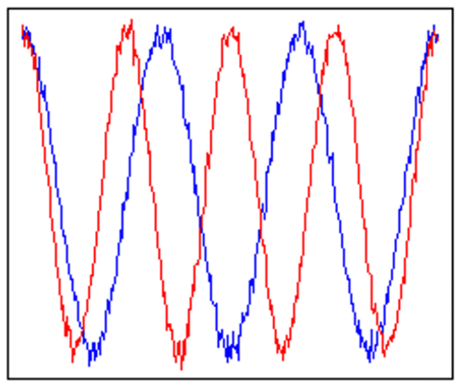


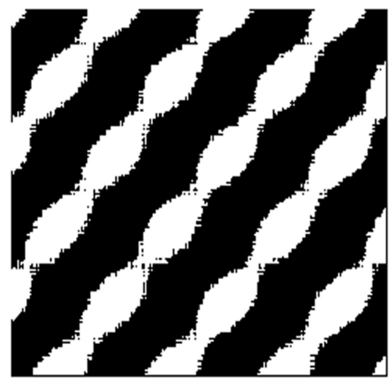
%DET=95

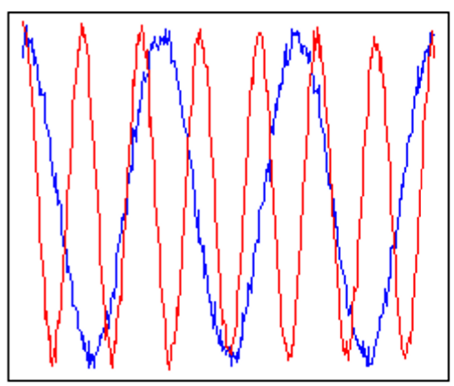
minDL=7

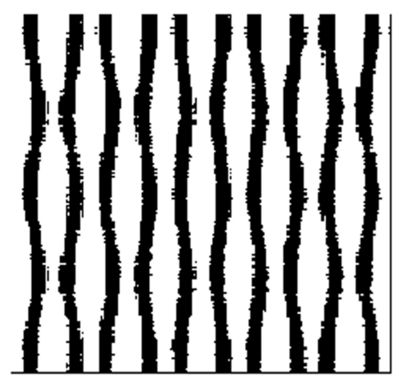
%DET=83

%DET: INSTABILITY









%DET = 95

minDL = ~~7~~ 5

%DET = 8~~3~~ 97!

STABLE MEASURE?

? Define a measure of similarity that is provably stable?

- small change in input $\Rightarrow$ small change in measure

STABLE MEASURE?

? Define a measure of similarity that is provably stable?

- small change in input $\Rightarrow$ small change in measure

? Minimize # parameters?

STABLE MEASURE?

? Define a measure of similarity that is provably stable?

- small change in input $\Longrightarrow$ small change in measure

? Minimize # parameters?

? Effects of large dimensions?

STABLE MEASURE?

? Define a measure of similarity that is provably stable?
- small change in input $\Rightarrow$ small change in measure

? Minimize # parameters?

? Effects of large dimensions?

? Does it work well in practice?

RESULTS

✓ score$(f_1|f_2)$: conditional periodicity of f_1 given f_2

– persistent homology of a single point cloud

RESULTS

✓ score$(f_1|f_2)$: conditional periodicity of f_1 given f_2

- persistent homology of a single point cloud

✓ theoretical stability under

- small changes to periodicity
- Gaussian noise

RESULTS

- ✓ $score(f_1|f_2)$: conditional periodicity of f_1 given f_2
  - persistent homology of a single point cloud
- ✓ theoretical stability under
  - small changes to periodicity
  - Gaussian noise
- ✓ single parameter : embedding dimension
  - min embedding dimension to control precision of $score(f_1|f_2)$

RESULTS

- ✓ $score(f_1|f_2)$: conditional periodicity of f_1 given f_2
  - persistent homology of a single point cloud
- ✓ theoretical stability under
  - small changes to periodicity
  - Gaussian noise
- ✓ single parameter : embedding dimension
  - min embedding dimension to control precision of $score(f_1|f_2)$
- ✓ stable under PCA truncation

RESULTS

✓ $\mathrm{score}(f_1|f_2)$: conditional periodicity of f_1 given f_2
- persistent homology of a single point cloud

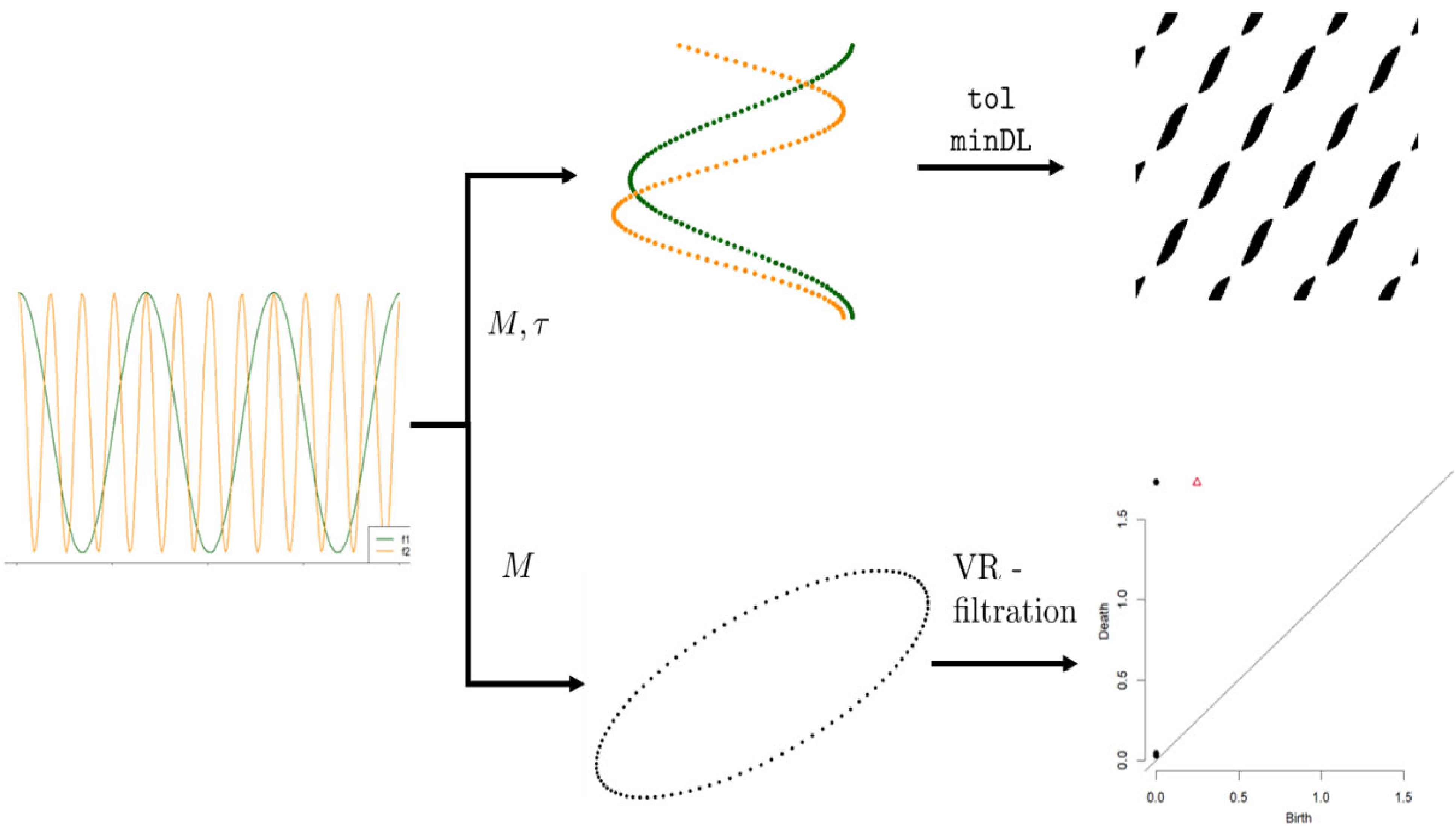
✓ theoretical stability under
- small changes to periodicity
- Gaussian noise

✓ single parameter: embedding dimension
- min embedding dimension to control precision of $\mathrm{score}(f_1|f_2)$

✓ stable under PCA truncation

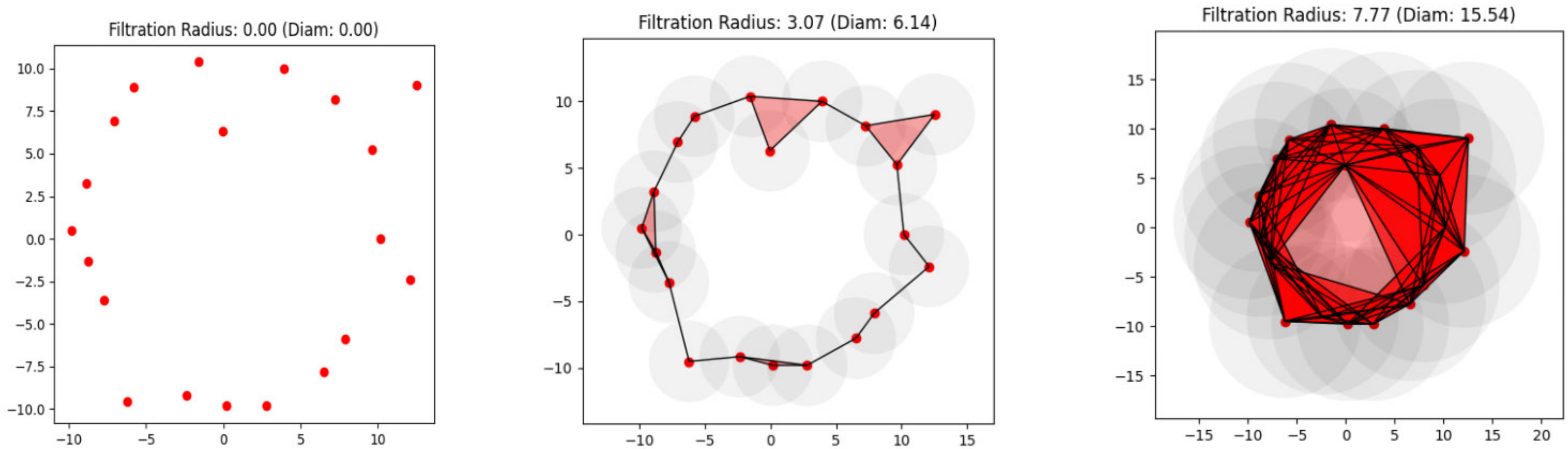
✓ computational evidence

SCORE($f_1|f_2$) VS %DET



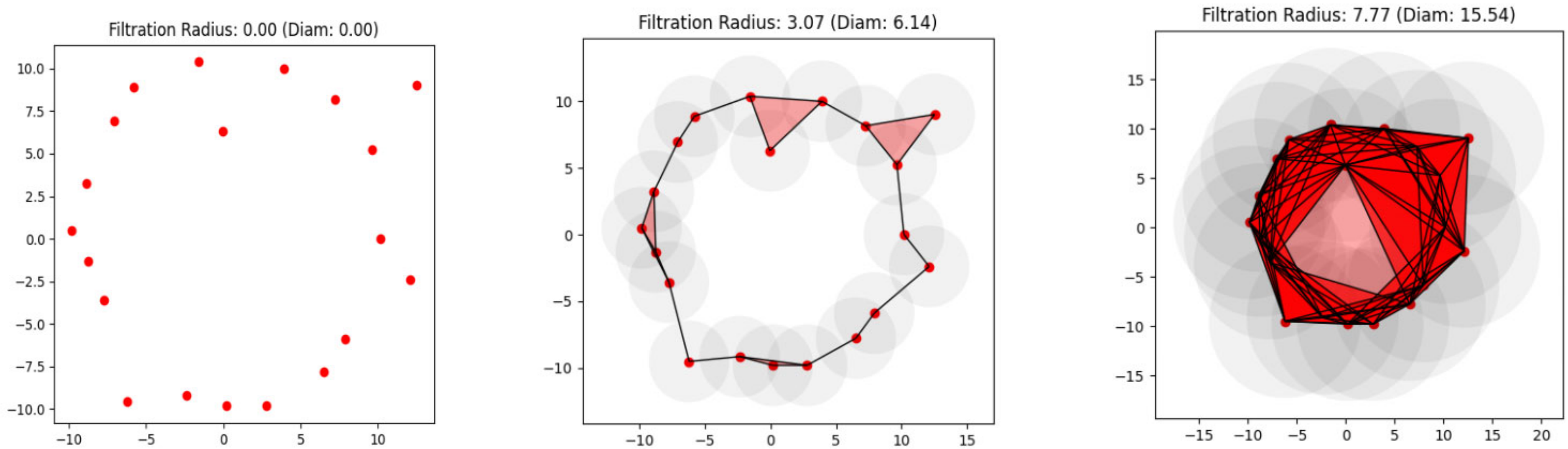
Vietoris-Rips (VR) Persistence

Edelsbrunner et al., 2002

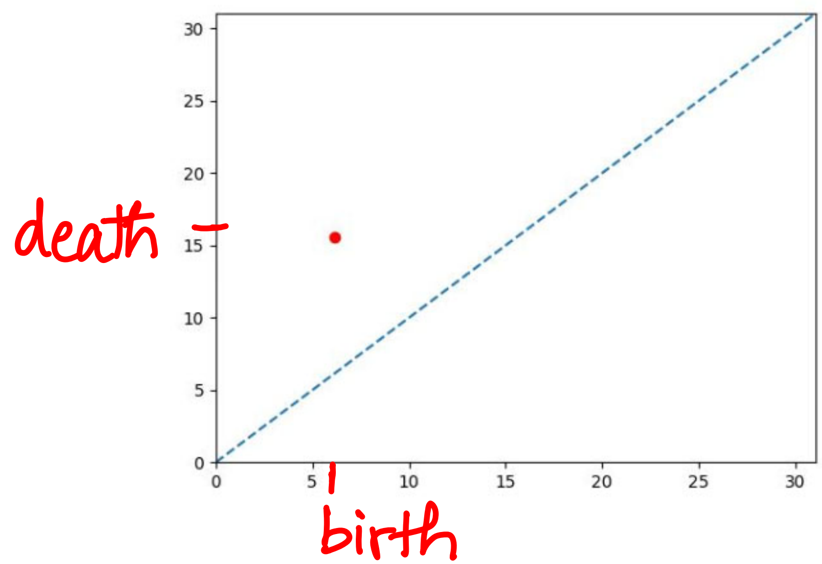


VIETORIS-RIPS (VR) PERSISTENCE

Edelsbrunner et al., 2002



persistence diagram (PD):

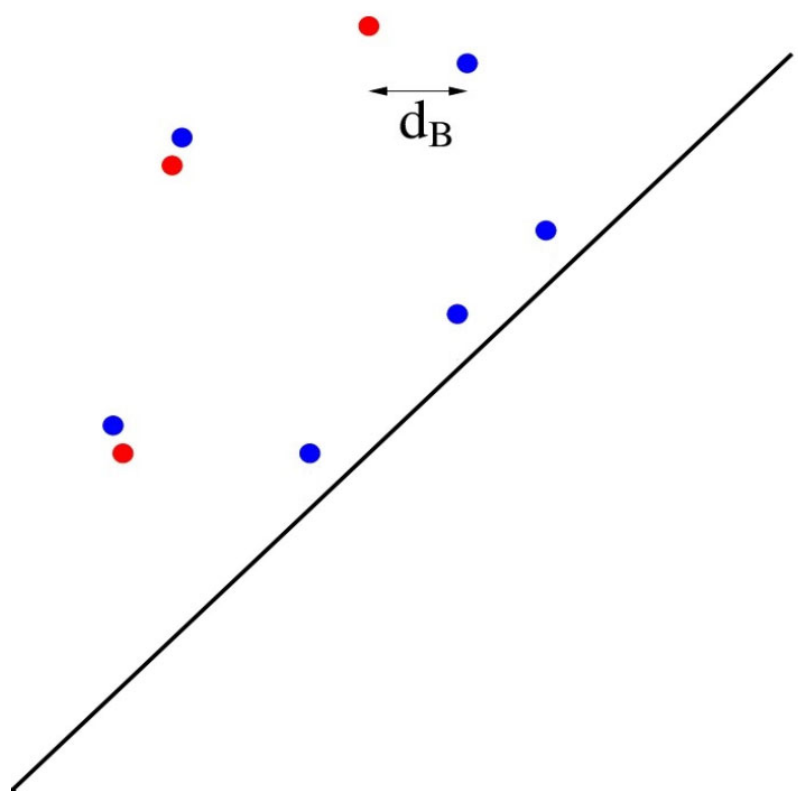


Bottleneck Distance

for two PDs X and Y

$$d_B(X,Y) = \min_{\gamma : X \to Y} \max_{x \in X} \|x - \gamma(x)\|_\infty$$

bijection



PERSISTENCE STABILITY

Chazal, de Silva, Oudot (2014)

Chazal, de Silva, Glisse, Oudot (2016)

$$d_B\big(\mathrm{dgm}(VR(X)), \mathrm{dgm}(VR(Y))\big) \leq 2\, d_{GH}(X,Y) \leq 2\, d_H(X,Y)$$

Gromov-Hausdorff

Hausdorff

Persistence Stability

Chazal, de Silva, Oudot (2014)

Chazal, de Silva, Glisse, Oudot (2016)

$$d_B\big(\mathrm{dgm}(VR(X)), \mathrm{dgm}(VR(Y))\big) \leq 2\, d_{GH}(X,Y) \leq 2\, d_H(X,Y)$$

Gromov-Hausdorff

under all isometric embeddings of X, Y into common metric space

Hausdorff

X, Y in common metric space

PERSISTENCE STABILITY

Chazal, de Silva, Oudot (2014)
Chazal, de Silva, Glisse, Oudot (2016)

$$d_B\big(dgm(VR(X)), dgm(VR(Y))\big) \leq 2\,d_{GH}(X,Y) \leq 2\,d_H(X,Y)$$

Gromov-Hausdorff

under all isometric embeddings of X, Y into common metric space

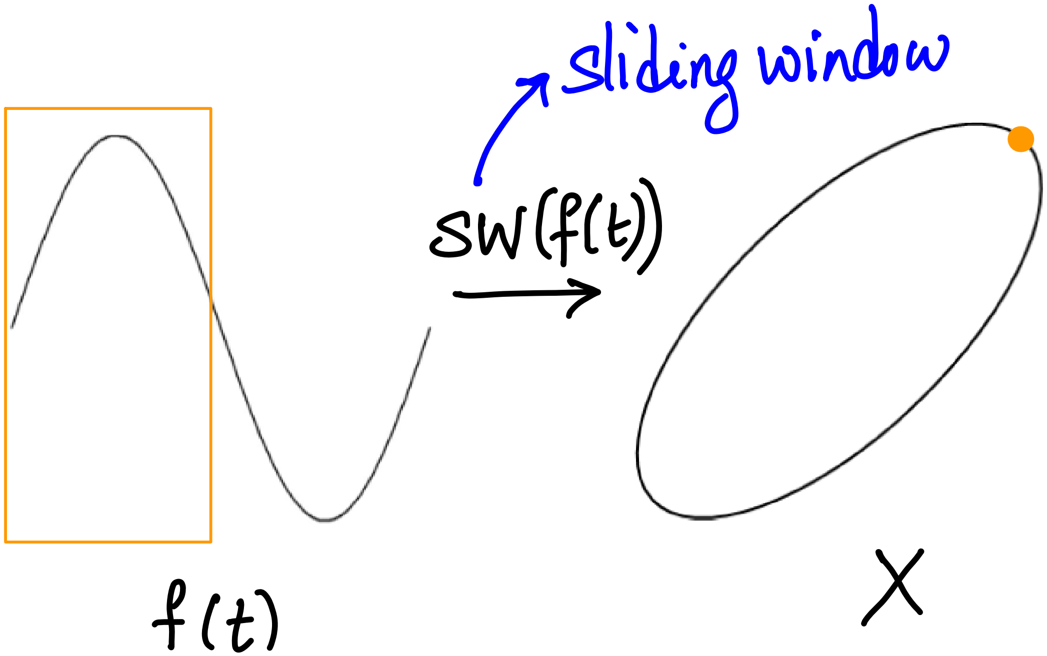
Hausdorff

X, Y in common metric space

PH: VR Persistent Homology

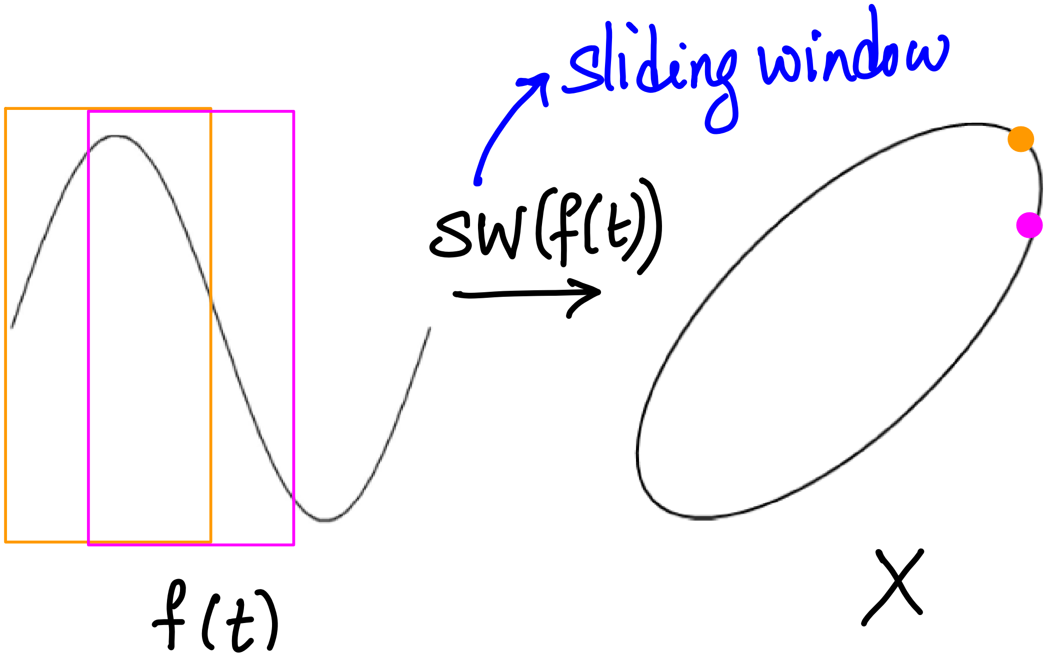
PH on Time Series

Perea & Harer (2015)
Perea et al. (2015)



PH ON TIME SERIES

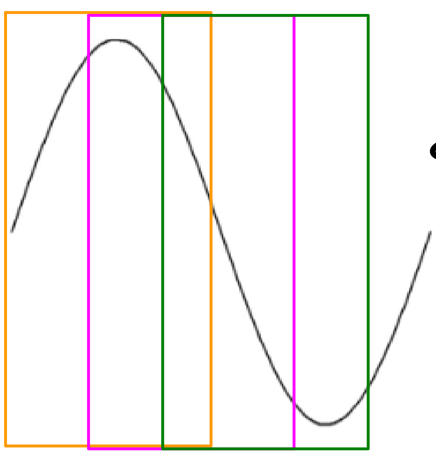
Perea & Harer (2015)
Perea et al. (2015)



PH ON TIME SERIES

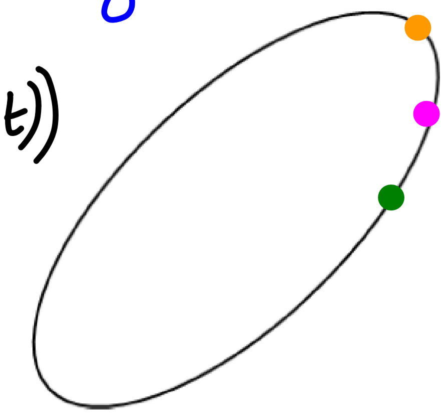
Perea & Harer (2015)
Perea et al. (2015)

sliding window



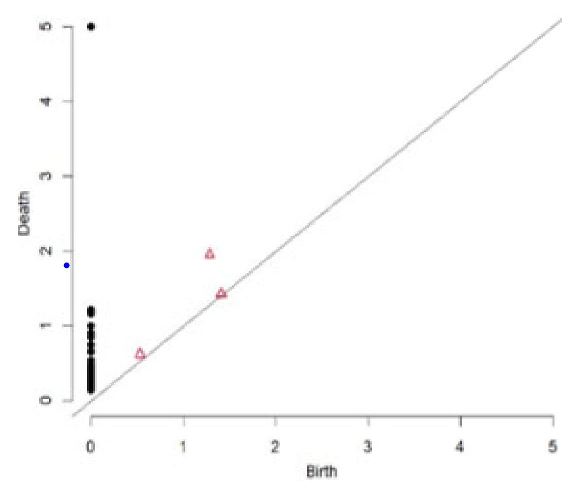
f(t)

$\xrightarrow{SW(f(t))}$



X

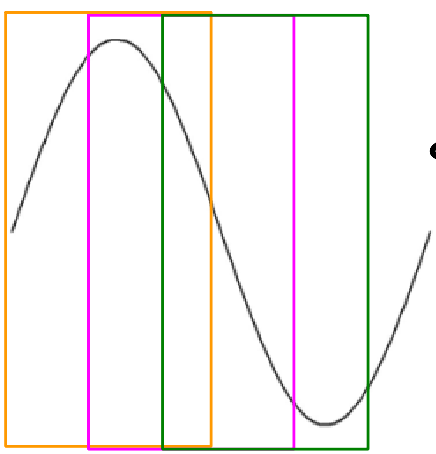
$\xrightarrow{VR}$



dgm(X)

PH on Time Series

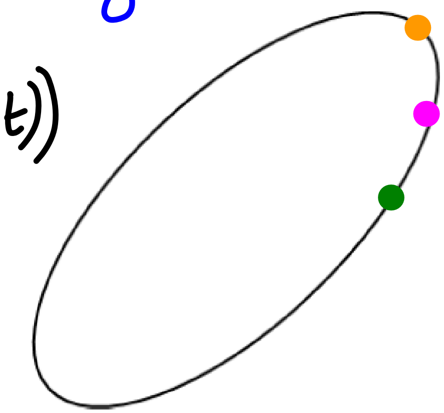
Perea & Harer (2015)
Perea et al. (2015)



$f(t)$

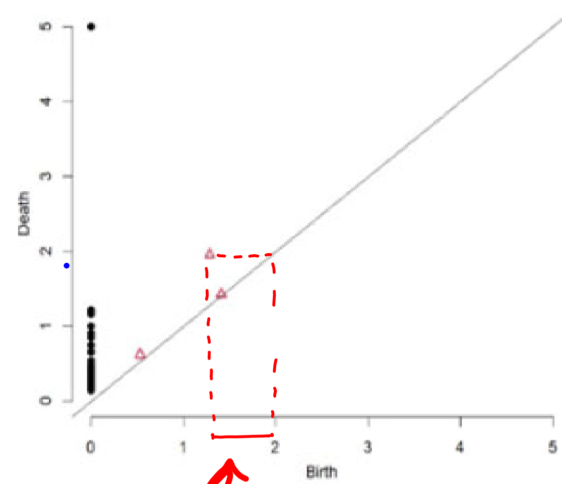
sliding window

$\xrightarrow{SW(f(t))}$



X

$\xrightarrow{VR}$



$dgm(X)$

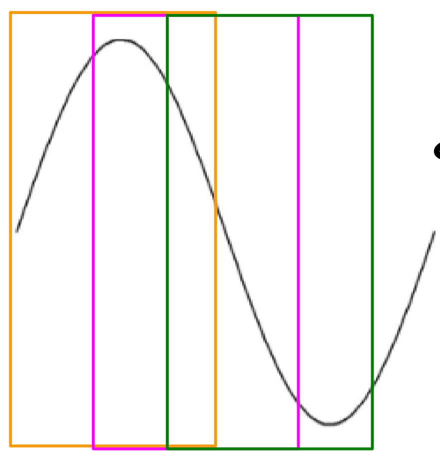
$mp(dgm(X))$

max 1-persistence

PH on Time Series

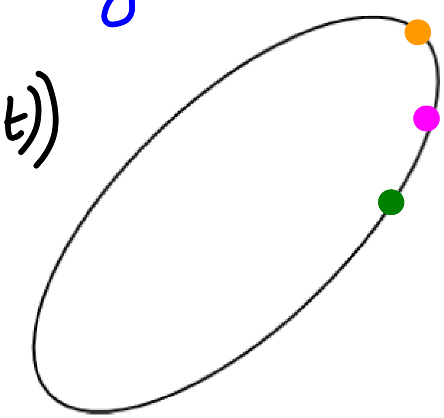
Perea & Harer (2015)
Perea et al. (2015)

sliding window



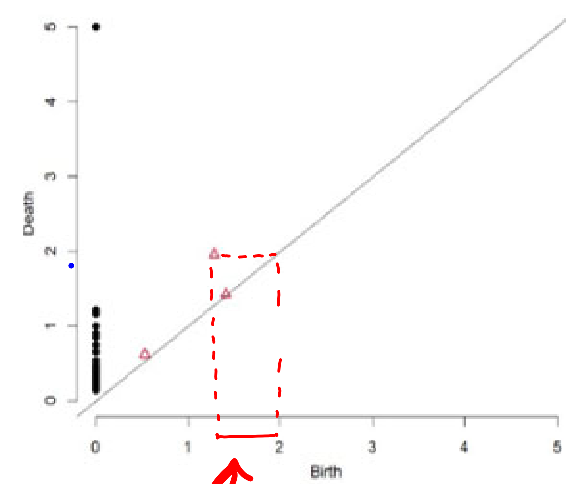
$f(t)$

$\xrightarrow{SW(f(t))}$



X

$\xrightarrow{VR}$



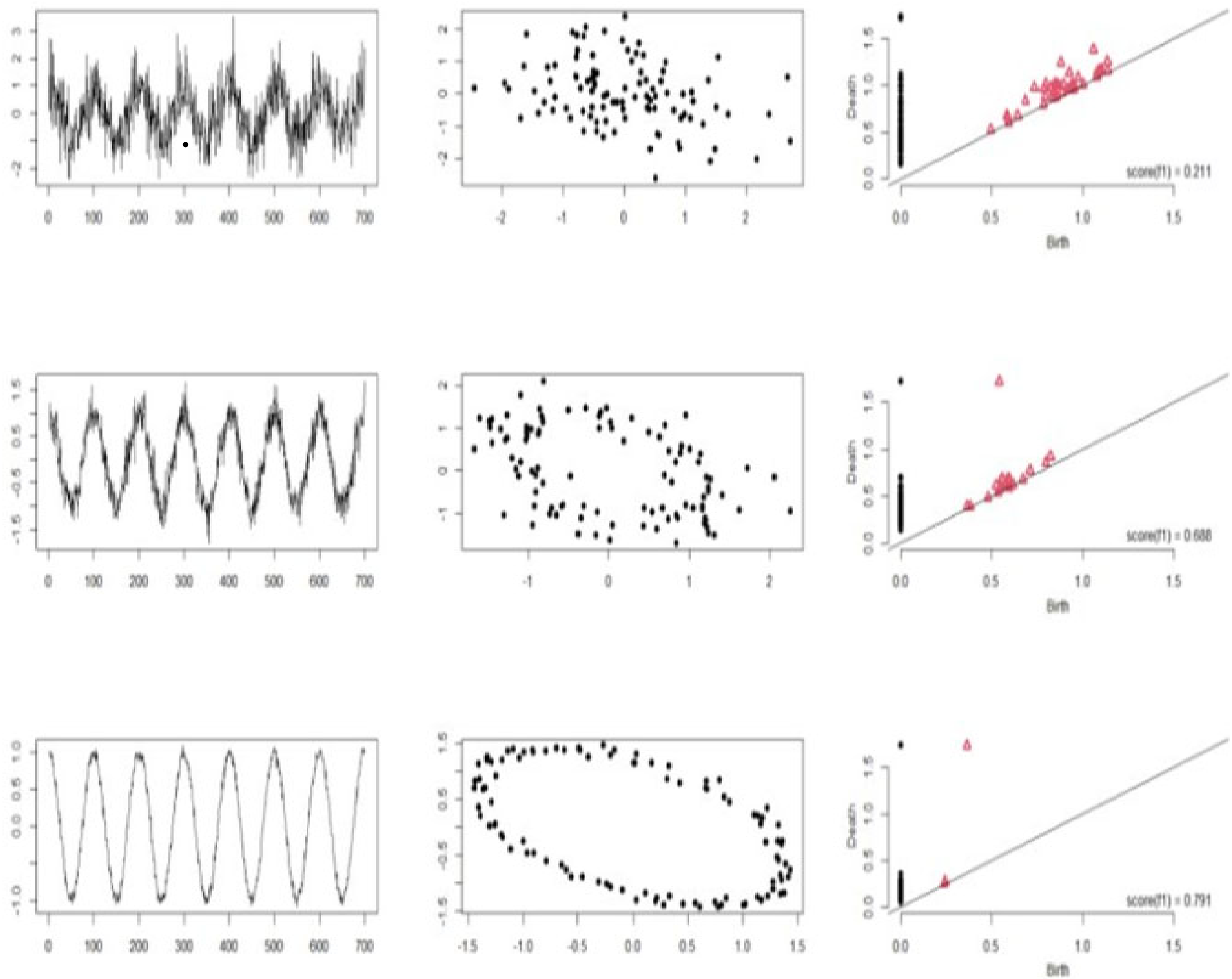
$dgm(X)$

$mp(dgm(X))$

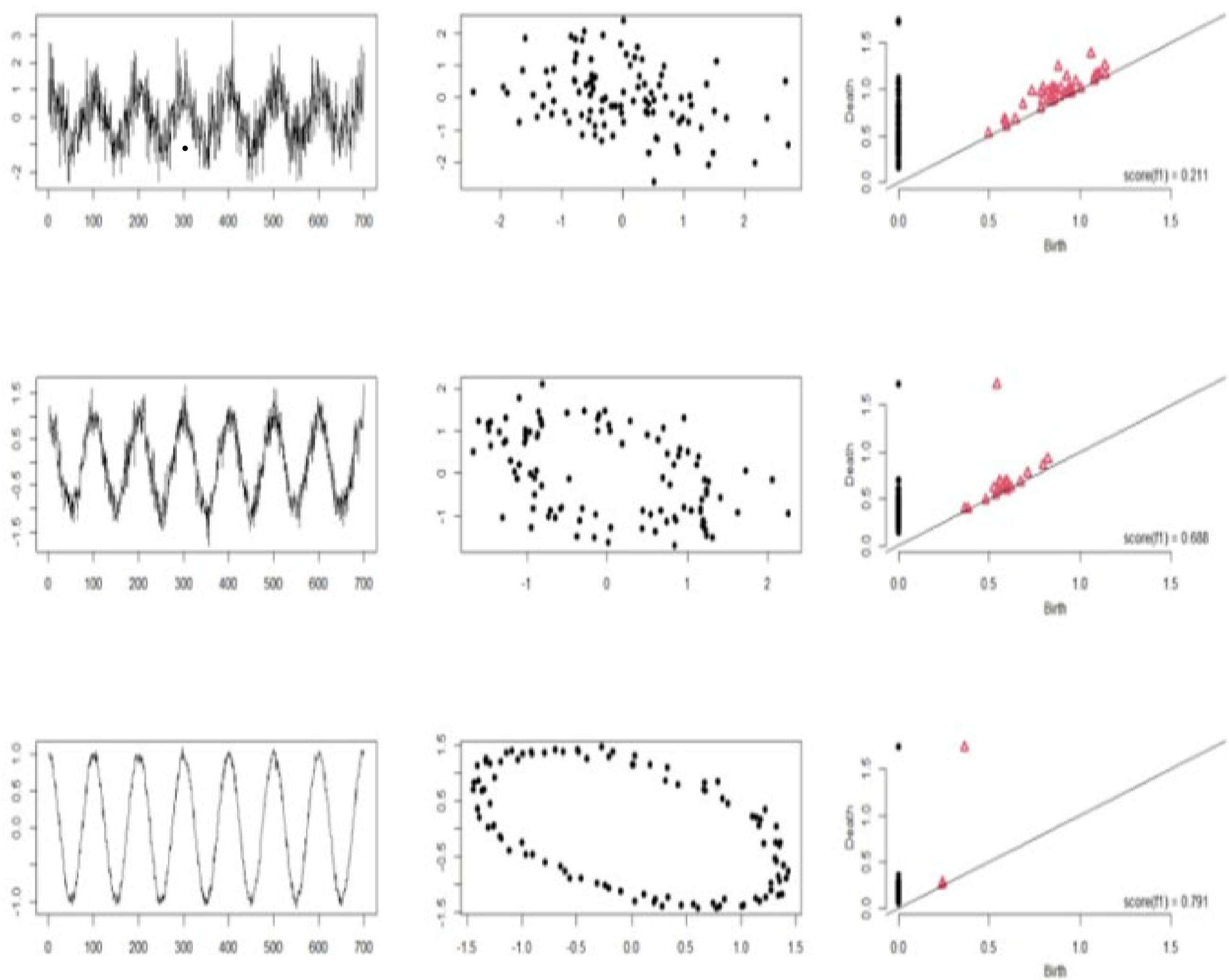
max 1-persistence

$$\rightarrow \quad score(f) = \frac{mp(dgm(X))}{\sqrt{3}}$$

PH on Time Series



PH ON TIME SERIES



→ stability : closeness of $SW(f)$ and $SW(S_N f)$

N-truncated Fourier series of f

CONDITIONAL PERIODICTY SCORE

<u>Def</u> $f_1, f_2 : [0, 2\pi] \to \mathbb{R}$ continuous, periodic time series

f_2 more-periodic than f_1 $\left(\frac{2\pi}{\omega_2} \leq \frac{2\pi}{\omega_1} \right)$

CONDITIONAL PERIODICTY SCORE

<u>Def</u> $f_1, f_2 : [0, 2\pi] \to \mathbb{R}$ continuous, periodic time series

f_2 more-periodic than f_1 $\left(\frac{2\pi}{\omega_2} \leq \frac{2\pi}{\omega_1} \right)$

Conditional SW Embedding of f_1 given f_2:

$$SW_{M,\tau} f_{1|2}(t) = \left(f_1(t), \ldots, f_1(t + M\tau)\right)^T \text{ for } \tau = \frac{2\pi}{\omega_2(M+1)}.$$

CONDITIONAL PERIODICTY SCORE

Def $f_1, f_2 : [0, 2\pi] \to \mathbb{R}$ continuous, periodic time series

f_2 more-periodic than f_1 $\left(\frac{2\pi}{\omega_2} \leq \frac{2\pi}{\omega_1} \right)$

Conditional SW Embedding of f_1 given f_2:

$$SW_{M,\tau} f_{1|2}(t) = \left(f_1(t), \ldots, f_1(t + M\tau) \right)^T \text{ for } \tau = \frac{2\pi}{\omega_2(M+1)}.$$

$$\text{score}(f_1 | f_2) = \frac{mp\left(dgm\left(SW f_{1|2}(T)\right)\right)}{\sqrt{3}}, \quad T \in \left[0, \frac{2\pi}{\omega_1}\right].$$

CONDITIONAL PERIODICTY SCORE

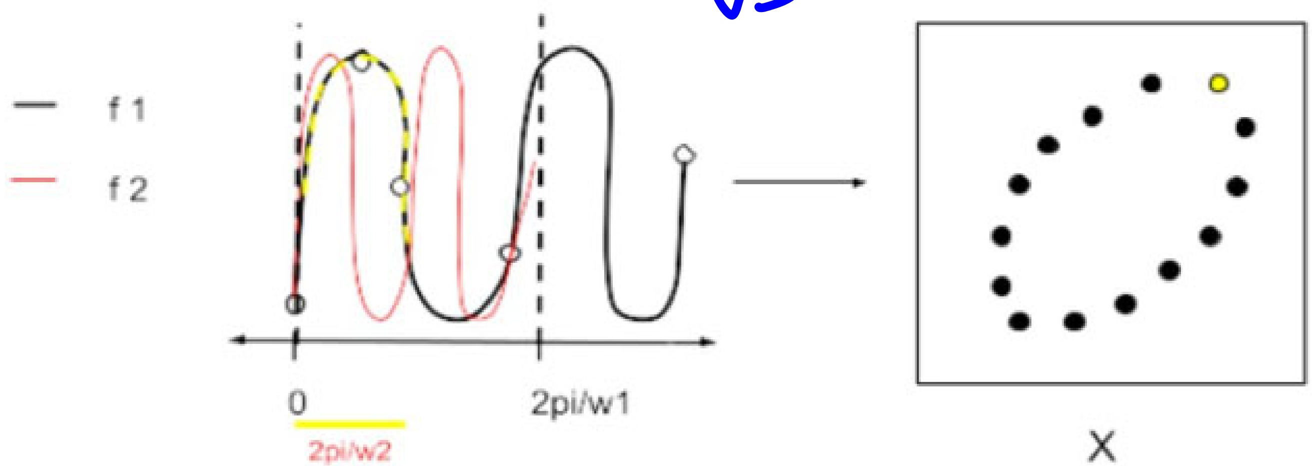
Def $f_1, f_2 : [0,2\pi] \to \mathbb{R}$ continuous, periodic time series

f_2 more-periodic than f_1 $\left(\frac{2\pi}{\omega_2} \leq \frac{2\pi}{\omega_1}\right)$

Conditional SW Embedding of f_1 given f_2:

$$SW_{M,\tau} f_{1|2}(t) = \left(f_1(t), \dots, f_1(t+M\tau)\right)^T \text{ for } \tau = \frac{2\pi}{\omega_2(M+1)}.$$

$$\text{score}(f_1|f_2) = \frac{mp\left(dgm\left(SW f_{1|2}(T)\right)\right)}{\sqrt{3}}, \quad T \in \left[0, \frac{2\pi}{\omega_1}\right].$$



REDUCTION TO PERIODICTY SCORE

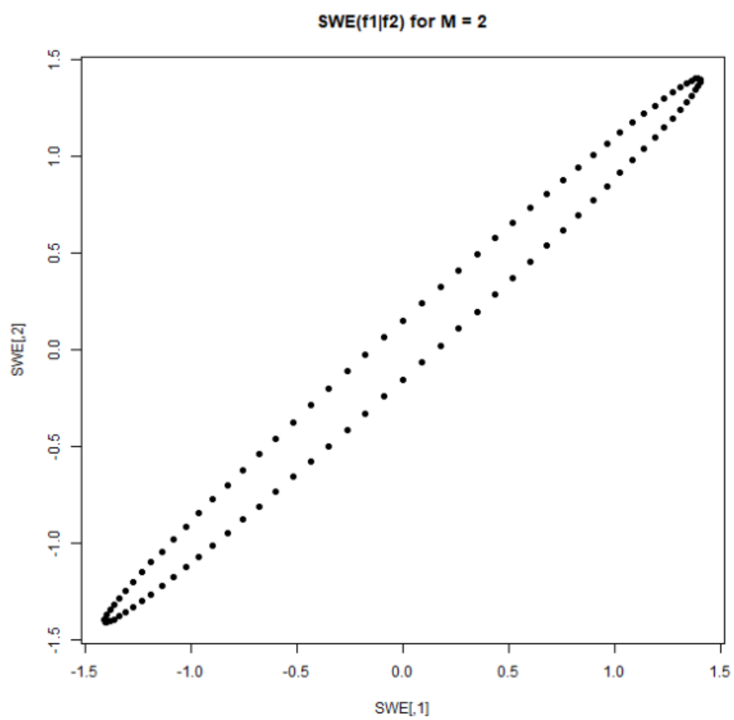
Proposition 1

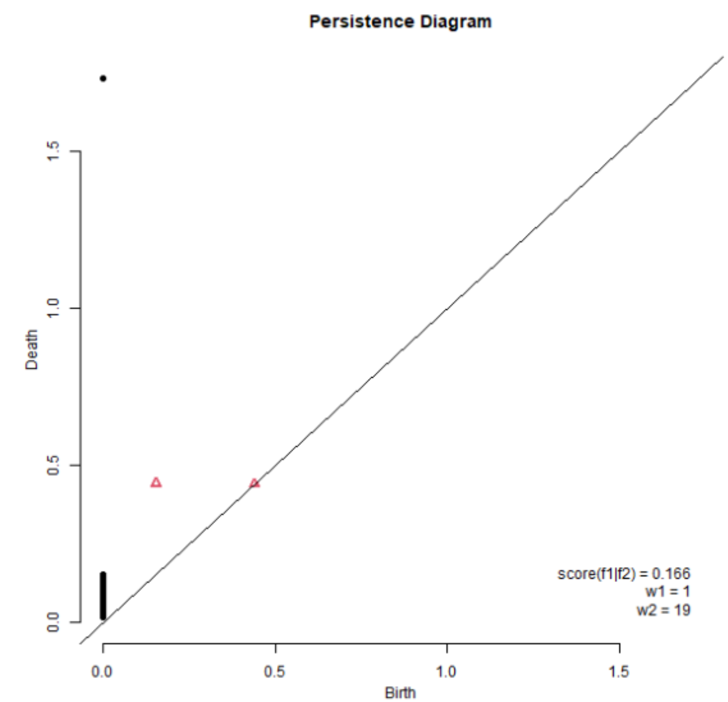
$$\lim_{\frac{2\pi}{\omega_2} \to \frac{2\pi}{\omega_1}^-} \mathrm{score}(f_1 | f_2) = \mathrm{score}(f_1)$$

Reduction to Periodicty Score

Proposition 1

$$\lim_{\frac{2\pi}{\omega_2} \to \frac{2\pi}{\omega_1}^-} \mathrm{score}(f_1|f_2) = \mathrm{score}(f_1)$$

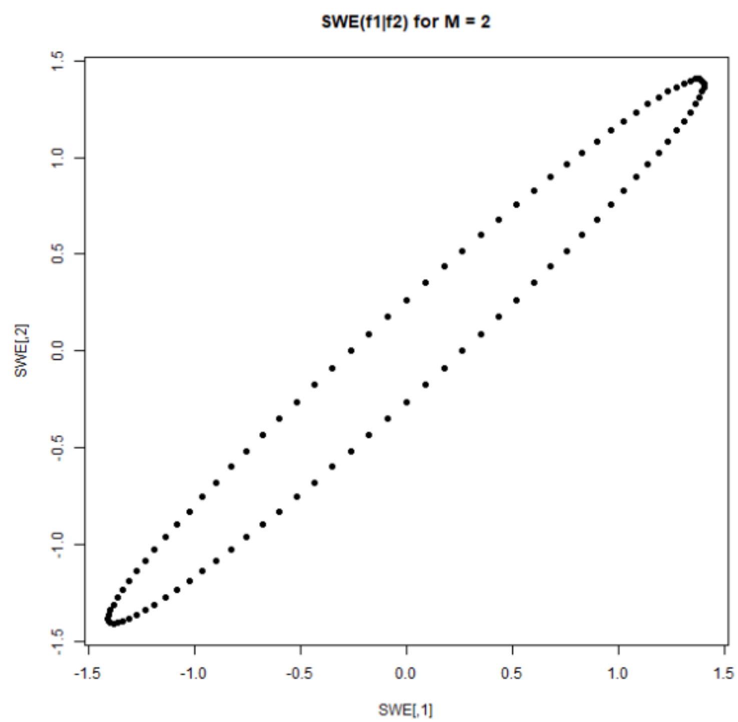


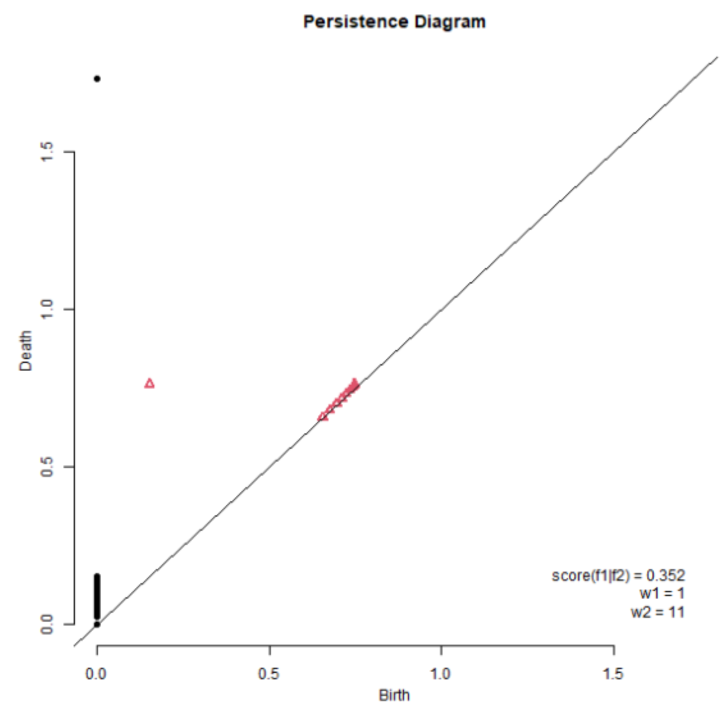


REDUCTION TO PERIODICTY SCORE

Proposition 1

$$\lim_{\frac{2\pi}{\omega_2} \to \frac{2\pi}{\omega_1}^-} \text{score}(f_1|f_2) = \text{score}(f_1)$$

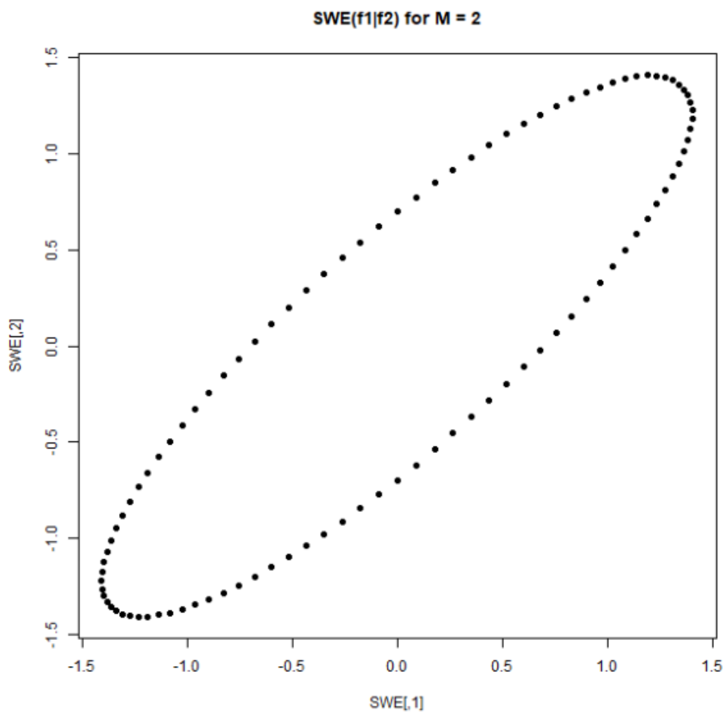


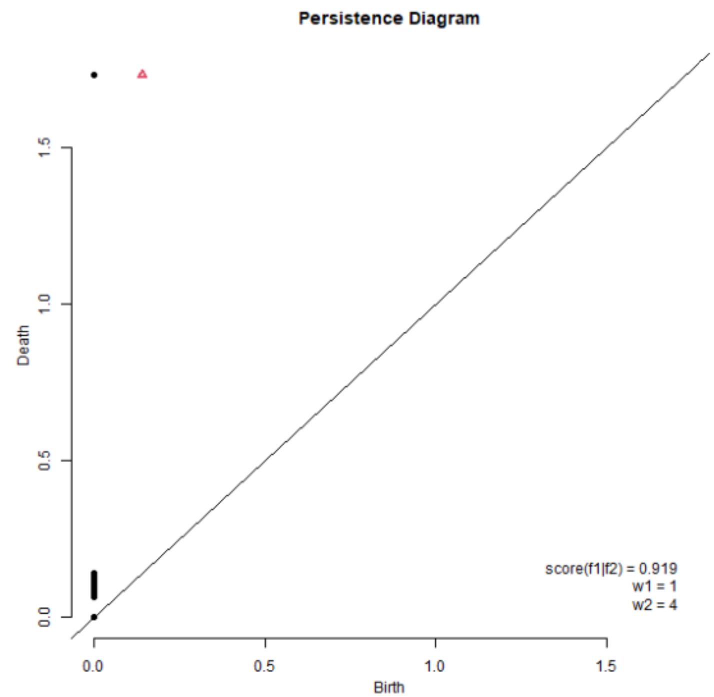


REDUCTION TO PERIODICTY SCORE

Proposition 1

$$\lim_{\frac{2\pi}{\omega_2} \to \frac{2\pi}{\omega_1}^-} \text{score}(f_1|f_2) = \text{score}(f_1)$$





STABILITY RESULTS

✓ small change in periodicity of $f_2 \Rightarrow$ score$(f_1|f_2)$ changes only a little

STABILITY RESULTS

✓ small change in periodicity of $f_2 \Rightarrow$ $\text{score}(f_1|f_2)$ changes only a little

Theorem 2 $f_1, f_{21}, f_{22} : [0, 2\pi] \to \mathbb{R} : \frac{2\pi}{\omega_1} \geq \frac{2\pi}{\omega_{21}} \geq \frac{2\pi}{\omega_{22}}$

$X_1 = SW_{M,\tau_1} f_{1|21}(T)$, $X_2 = SW_{M,\tau_2} f_{1|22}(T)$

$$d_H(X_1, X_2) \leq \sqrt{M+1} \left| \frac{2\pi}{\omega_{21}} - \frac{2\pi}{\omega_{22}} \right| \sqrt{\sum_i |f_1'(c_i)|^2}$$

STABILITY RESULTS

✓ small change in periodicity of $f_2 \Rightarrow$ score$(f_1|f_2)$ changes only a little

Theorem 2 $f_1, f_{21}, f_{22} : [0, 2\pi] \to \mathbb{R} : \frac{2\pi}{\omega_1} \geq \frac{2\pi}{\omega_{21}} \geq \frac{2\pi}{\omega_{22}}$

$X_1 = SW_{M,\tau_1} f_{1|21}(T), \quad X_2 = SW_{M,\tau_2} f_{1|22}(T)$

$$d_H(X_1, X_2) \leq \sqrt{M+1} \left| \frac{2\pi}{\omega_{21}} - \frac{2\pi}{\omega_{22}} \right| \sqrt{\sum_i |f_1'(c_i)|^2}$$

$$\left| \text{score}(f_1|f_{21}) - \text{score}(f_1|f_{22}) \right| \leq 4\sqrt{\frac{M+1}{3}} \left| \frac{2\pi}{\omega_{21}} - \frac{2\pi}{\omega_{22}} \right| \sqrt{\sum_i |f_1'(c_i)|^2}$$

STABILITY RESULTS

✓ Small Gaussian noise added to $f_1 \Rightarrow$ $\text{score}(f_1|f_2)$ changes only a little w.h.p

STABILITY RESULTS

✓ Small Gaussian noise added to $f_1 \Rightarrow$ score$(f_1|f_2)$ changes only a little w.h.p

Lemma 3 $f_1^{\sigma}(t) = f_1(t) + \epsilon_t$ for $\epsilon_t \sim N(0, \sigma^2)$

For $\delta \in (0,1)$, it is at least $(1-\delta)\cdot 100\%$ likely that

$$\left|\text{score}(f_1|f_2) - \text{Score}_{\sigma}(f_1|f_2)\right| \le 4\sigma\sqrt{\frac{M+1}{3\delta}}$$

STABILITY RESULTS

✓ Using top K principal components of PCA projection of SW $f_{1|2}(t)$ preserves score$(f_1|f_2)$

STABILITY RESULTS

✓ Using top K principal components of PCA projection of SW $f_{1|2}(t)$ preserves score$(f_1|f_2)$

<u>Theorem 4</u> $\phi : \mathbb{R}^{M+1} \to \mathbb{R}^K$: PCA projection with eigen(vectors/values) $\{c_k, \lambda_k\}_{k=1}^{N}$.

$$\left|\text{score}(f_1|f_2) - \text{Score}_\phi(f_1|f_2)\right| \le \sqrt{\frac{8}{3}}\ \sqrt[4]{\sum_{i=K+1}^{N} \lambda_i^2}$$

"unused" (N−K) eigenvalues

MIN. EMBEDDING DIMENSION

✓ Embedding dim. $\mathcal{M}$ above which score $(f_1|f_2)$ does not change much with dimension

MIN. EMBEDDING DIMENSION

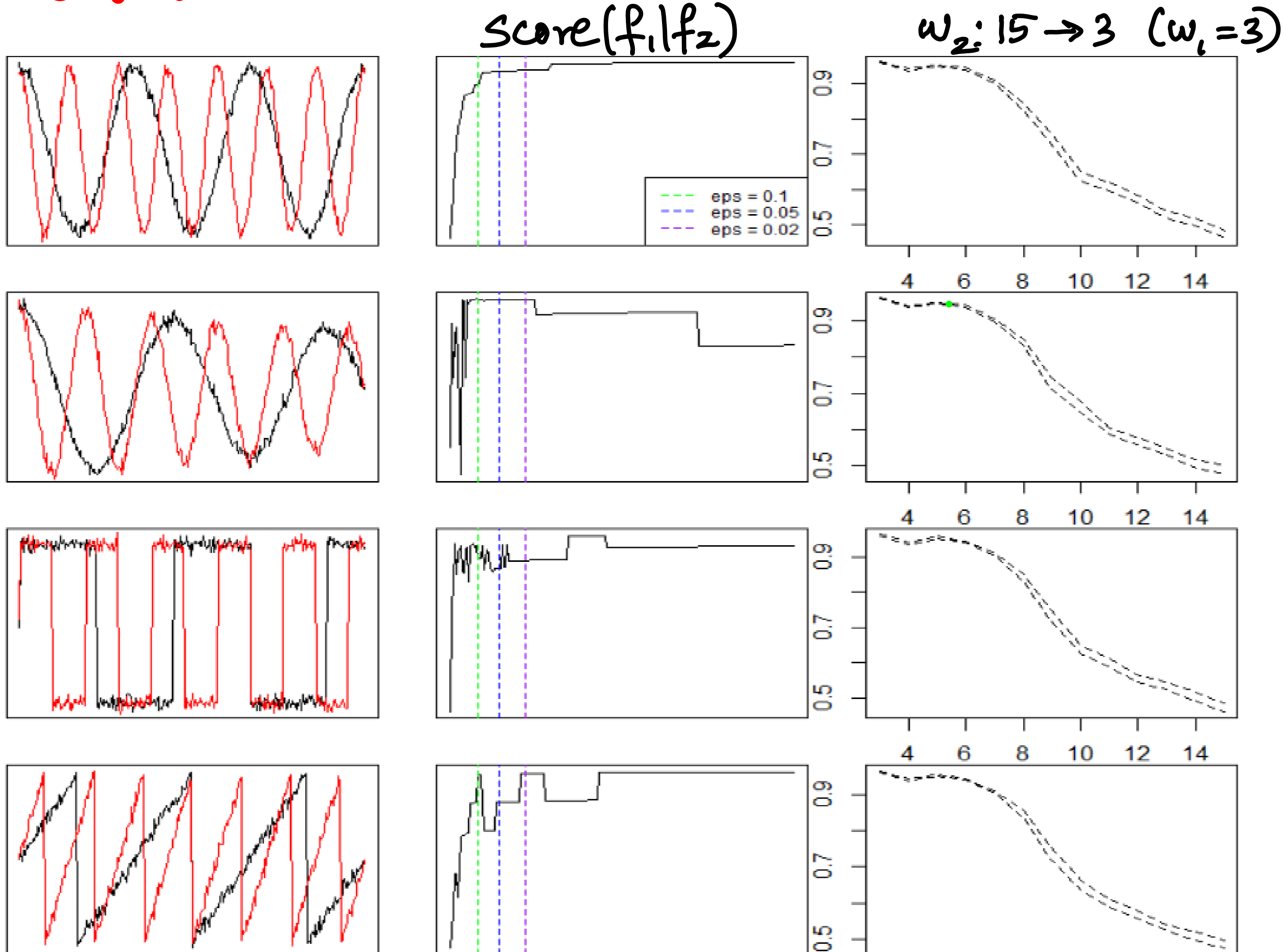
✓ Embedding dim. $\mathcal{M}$ above which score$(f_1|f_2)$ does not change much with dimension

<u>Theorem 5</u> For $\epsilon > 0$, with $\mathcal{M} = \left\lceil \frac{2\pi}{\omega_2 \epsilon} \right\rceil$, for any $M_2 > M_1 \geq \mathcal{M}$

$$\left| \text{score}_{M_1}(f_1|f_2) - \text{Score}_{M_2}(f_1|f_2) \right| \leq \epsilon \cdot g(M_1, f_1) + h(M_1, M_2, f_1)$$

constants w.r.t. ϵ

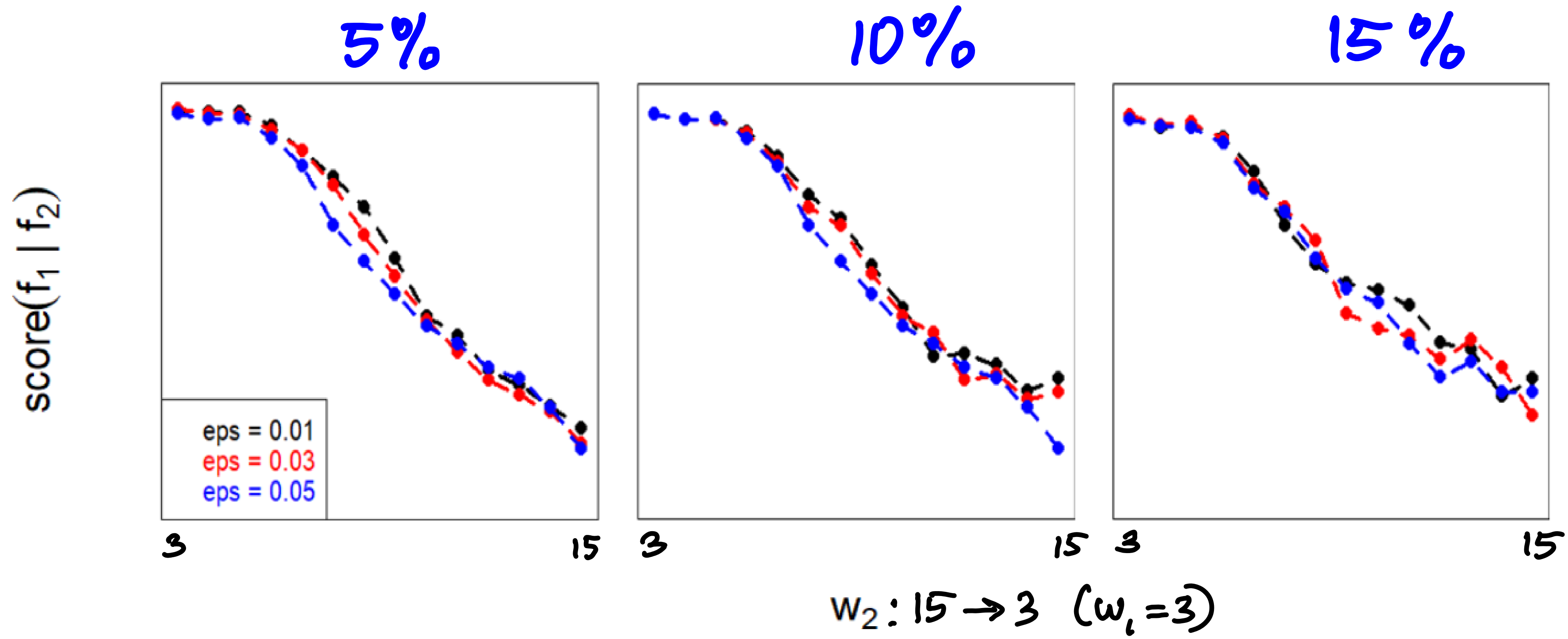
COMPUTATIONAL RESULTS



$\mathcal{M}$ for ϵ = 0.1, 0.05, 0.03

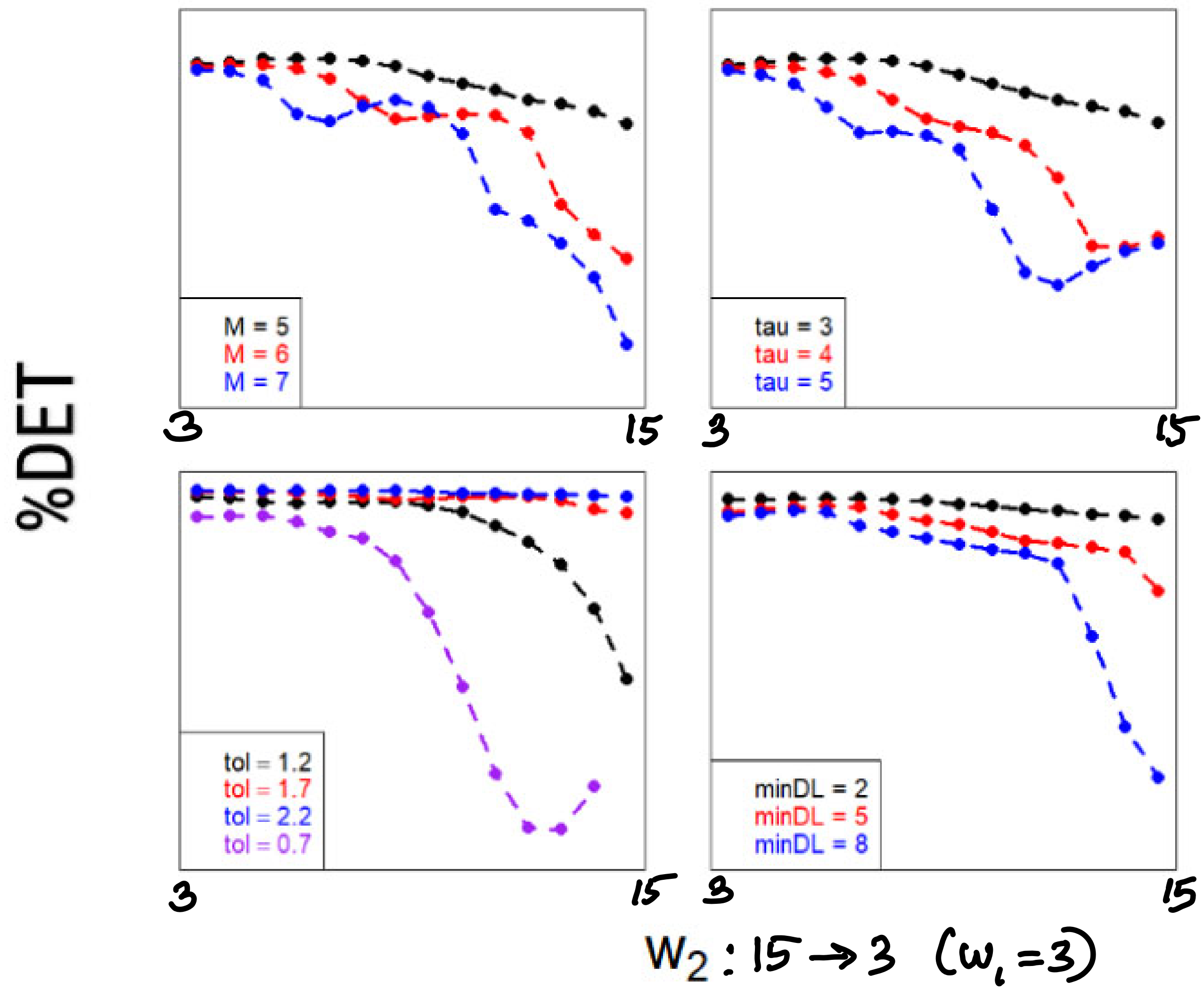
COMPUTATIONAL RESULTS

$score(f_1|f_2)$ under Gaussian noise



COMPUTATIONAL RESULTS

%DET under 10% Gaussian noise



OPEN QUESTIONS

? Testing on real data?

OPEN QUESTIONS

? Testing on real data?

? (Bound on) min # PC's needed to "preserve" $\text{score}_\phi(f_1 \| f_2)$?

OPEN QUESTIONS

? Testing on real data?

? (Bound on) min # PC's needed to "preserve" $\text{score}_\phi(f_1 | f_2)$?

Thank You!