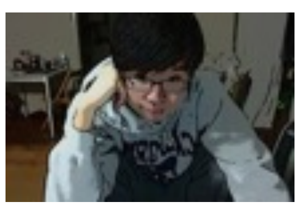
Investigating Sleep-Wake Signals: A Persistent Homology Approach



Hau-Tieng Wu
Dept. of Math.
Duke Univ.

Yu-Min Chung 鍾佑民
Department of Mathematics and Statistics
University of North Carolina at Greensboro

羅友倫醫師
林口長庚醫院

Peter Hu
CS, NTNU, Taiwan

Topological Data Analysis and Applications in Dynamical Systems
05/18-23 2019 SIAMDS

Myself

Computation of Inertial Manifolds

2005

- Indiana University, 2013 Ph.D

Mike Jolly

Ricardo Rosa
UFRJ, Brazil

- University of Kansas, 2014 Postdoc

Erik van Vleck

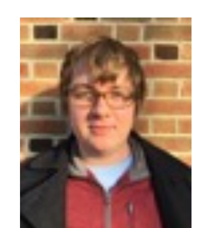
Andrew Steyer

- College of William and Mary Postdoc

Chi-Kwong Li

Liu Yue

Honor Thesis Student
Emily Schaal



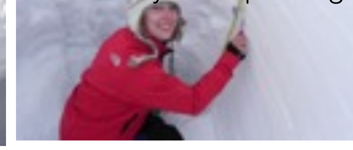
Honor Thesis Student
Dmitriy Zhigunov

- Firn Analysis

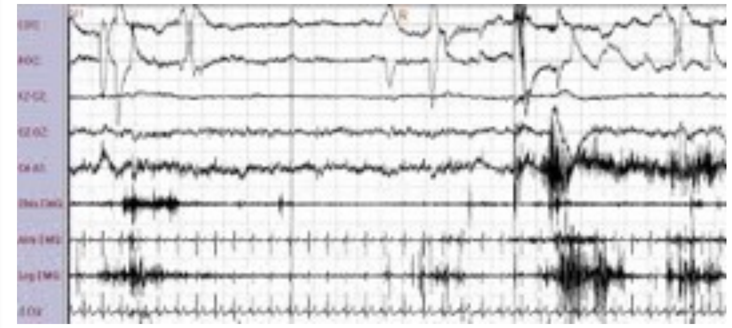
Sarah Day
College of William & Mary

Zoe Courville
Dartmouth College

Kaitlin Keegan
Centre for Ice and Climate
University of Copenhagen



- Sleep-Wake classification



Hau-Tieng Wu
Dept. of Math.
Duke Univ.

羅友倫醫師
林口長庚醫院

- Texture classification

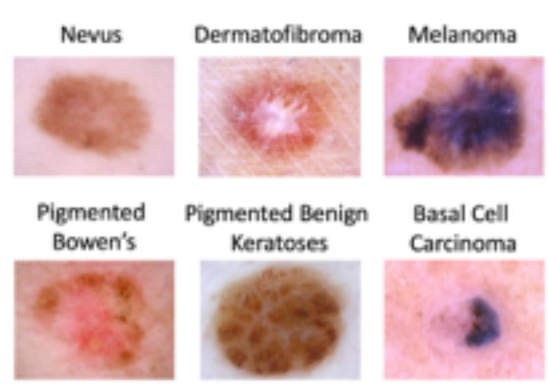
Austin Lawson
UNCG

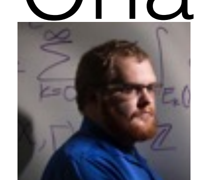
- Human Red Blood Cell

Madalena Costa
Harvard Medical School

Lawrence Leemis
Dept. Math, WM

- Skin Lesion Challenge



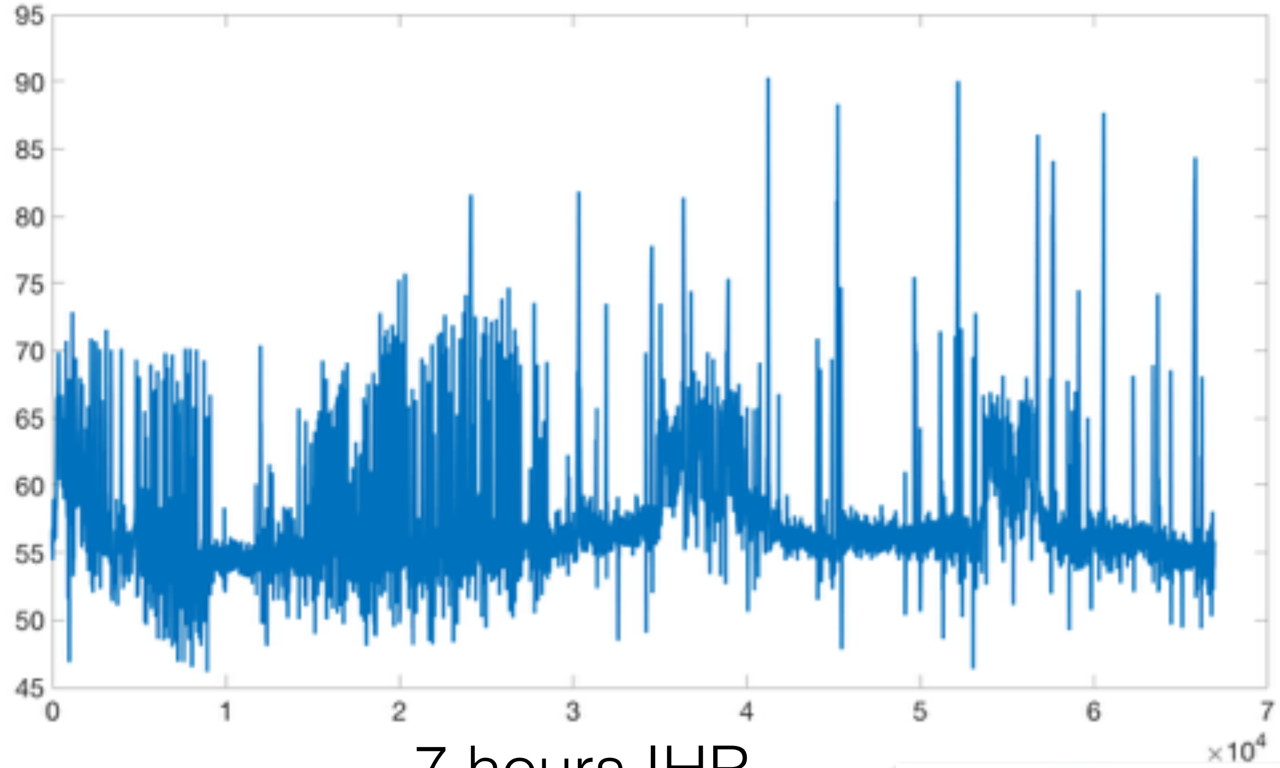


Peter Hu
CS, NTNU, Taiwan

Clifford Smyth
UNCG

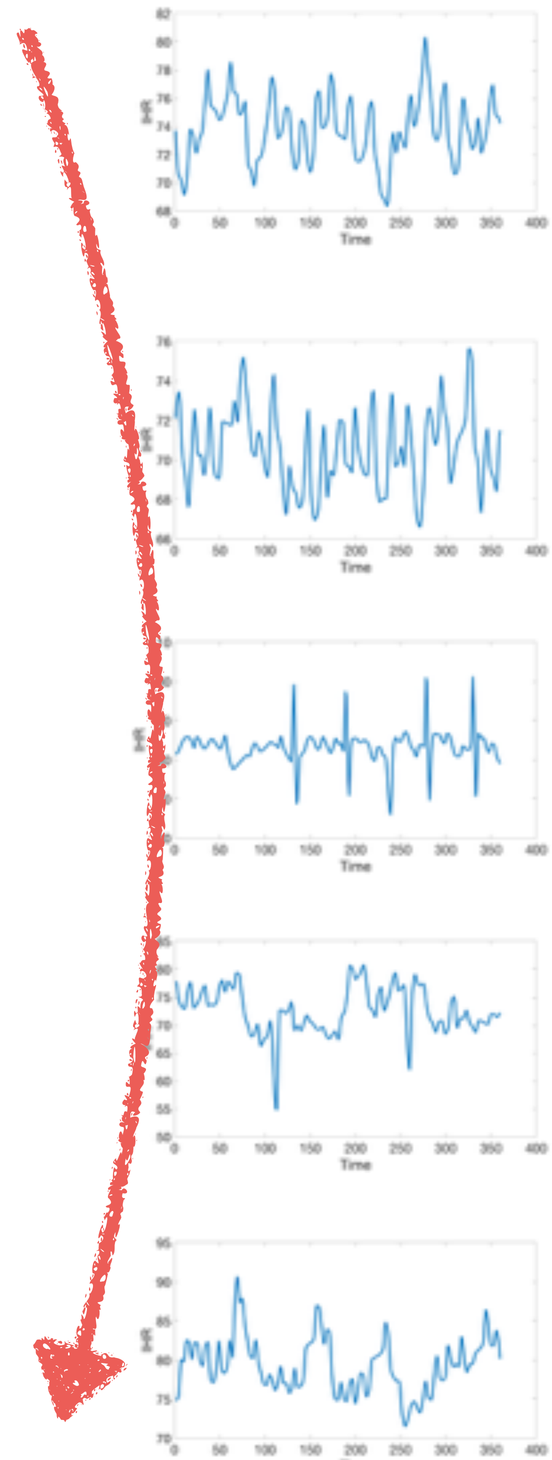
THE UNIVERSITY of NORTH CAROLINA
GREENSBORO

125
CELEBRATING 125 YEARS OF OPPORTUNITY & EXCELLENCE



7 hours IHR

Instantaneous Heart Rate (IHR)



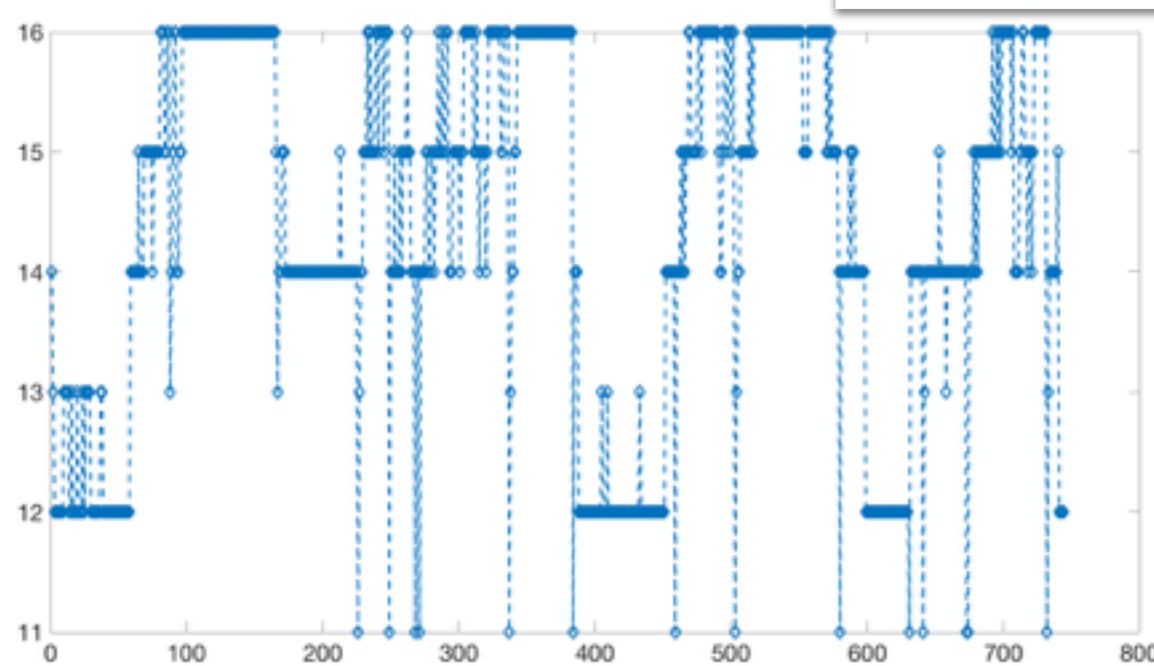
NREM1

NREM2

NREM3

REM

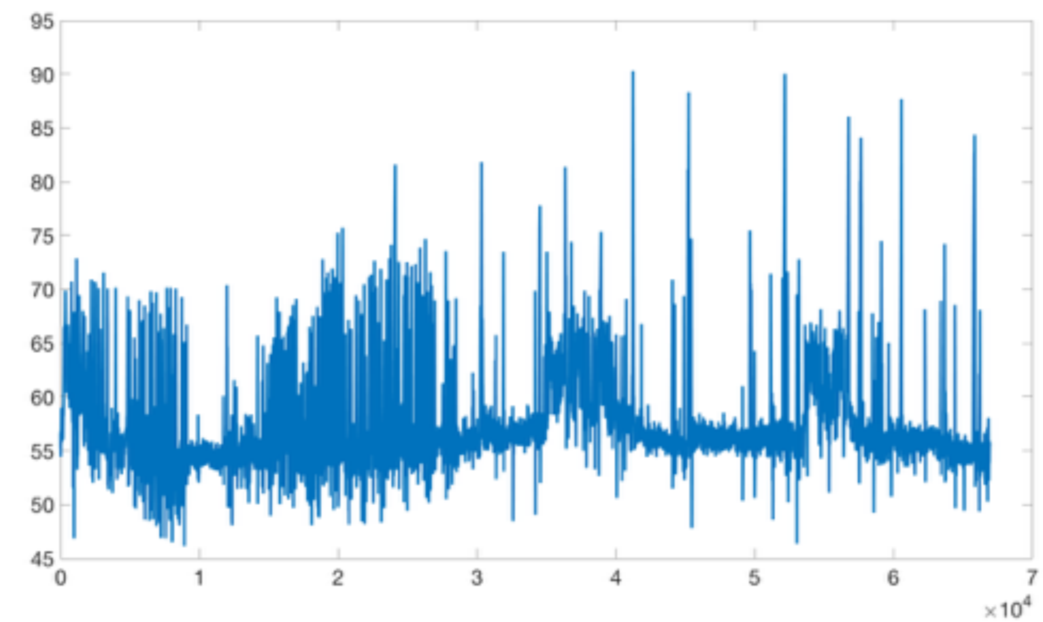
Wake

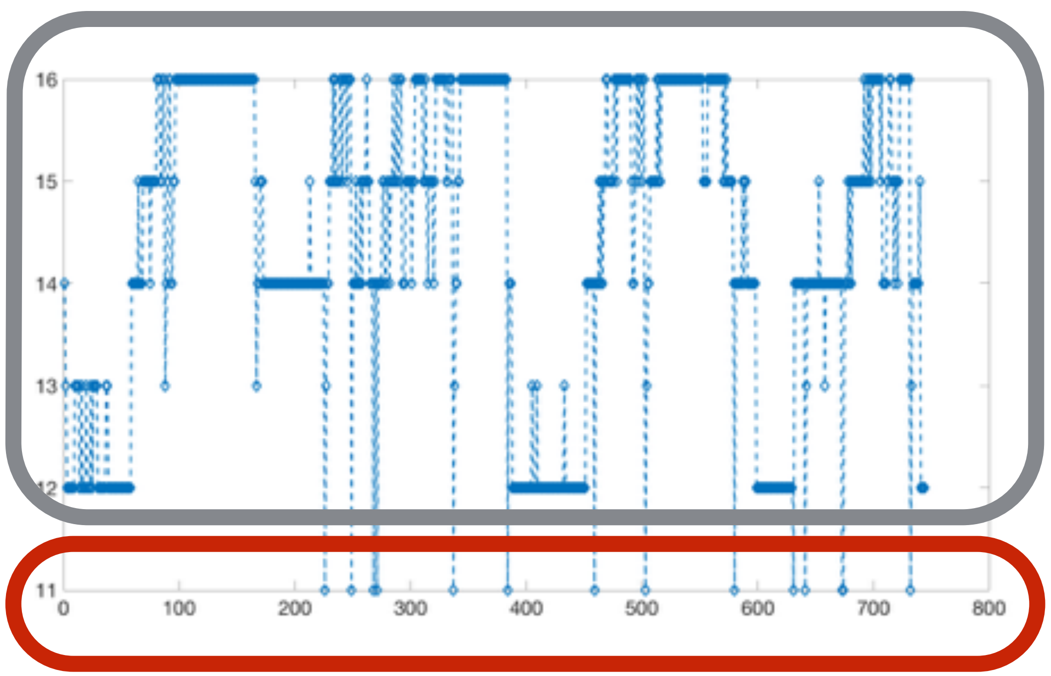


Annotations based on 30-second IHR

THE UNIVERSITY of NORTH CAROLINA
GREENSBORO

Main Task





Sleep

NREM1
NREM2
NREM3
NREM4
REM

Wake

Sleep-Wake Classification

[Xiao-Yan-Song-Yang-Yang '13] wake, REM, NonREM

[Mendez-Matteucci '10] REM, NonREM

[Lewicke-Sazonov-Corwin-Neuman-Schuckers '08]

[Long-Fonseca-Haakma-Aarts-Foussier '12]

[Aktaruzzaman-Migliorini-Tenhunen-Himanen-Bianchi-Sassi '15]

[Malik-Lo-Wu '18] wake, sleep

…

THE UNIVERSITY of NORTH CAROLINA GREENSBORO

CELEBRATING 125 YEARS OF OPPORTUNITY & EXCELLENCE

Challenges

- Imbalanced datasets, e.g. #waking=14 #sleeping= 760

Identify the awake signals

- Heterogeneity of datasets

Build a unified model across different datasets

Data

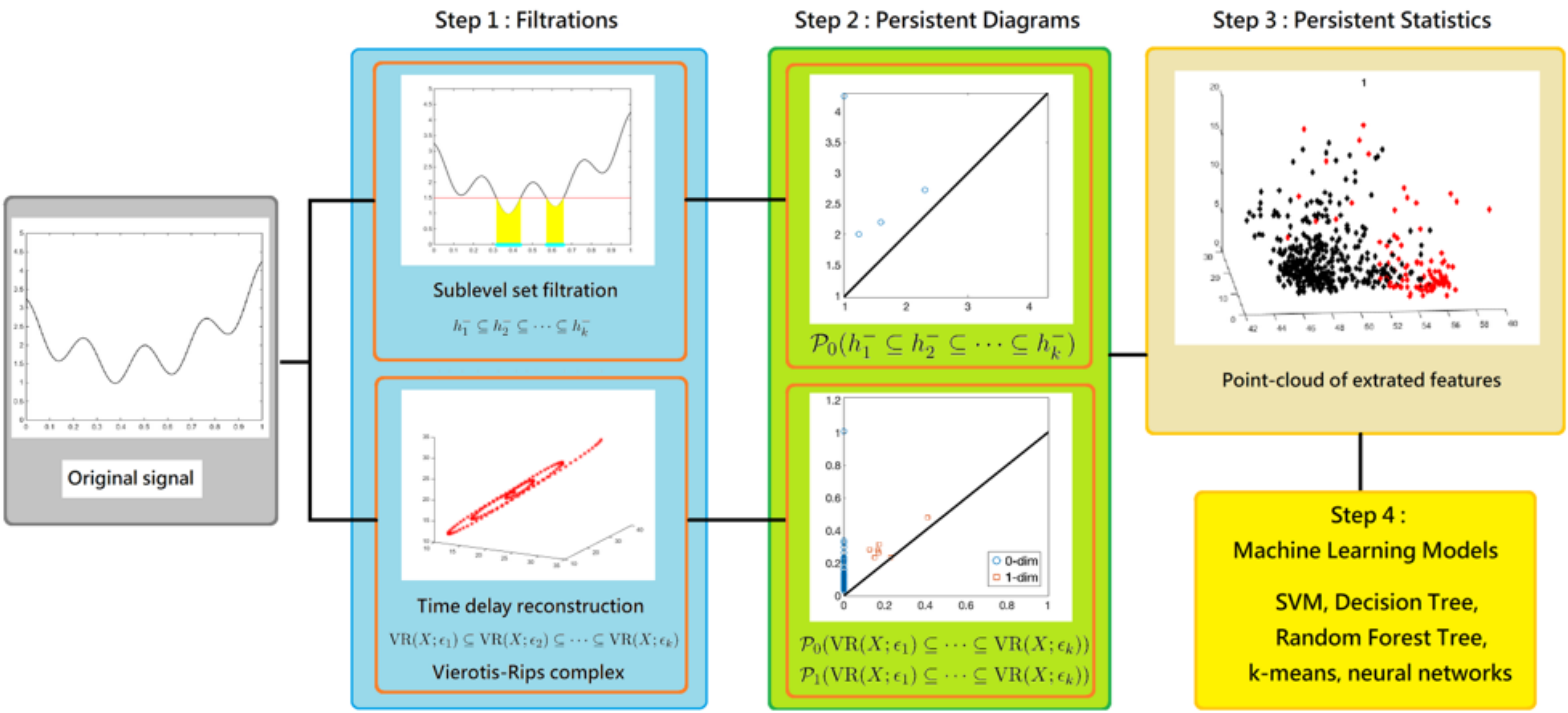
- **CGMH** (長庚紀念醫院)
  - 80 subjects as training
  - 27 subjects as testing

- **DREAMS(**http://www.tcts.fpms.ac.be/~devuyst/Databases/DatabaseSubjects)
  - 25 subjects

- St. Vincent's University Hospital / **U**niversity **C**ollege **D**ublin **S**leep **A**pnea **D**ata**b**ase(https://physionet.org/pn3/ucddb/)
  - 25 subjects

Each subject has about 7 hour IHR.

Method



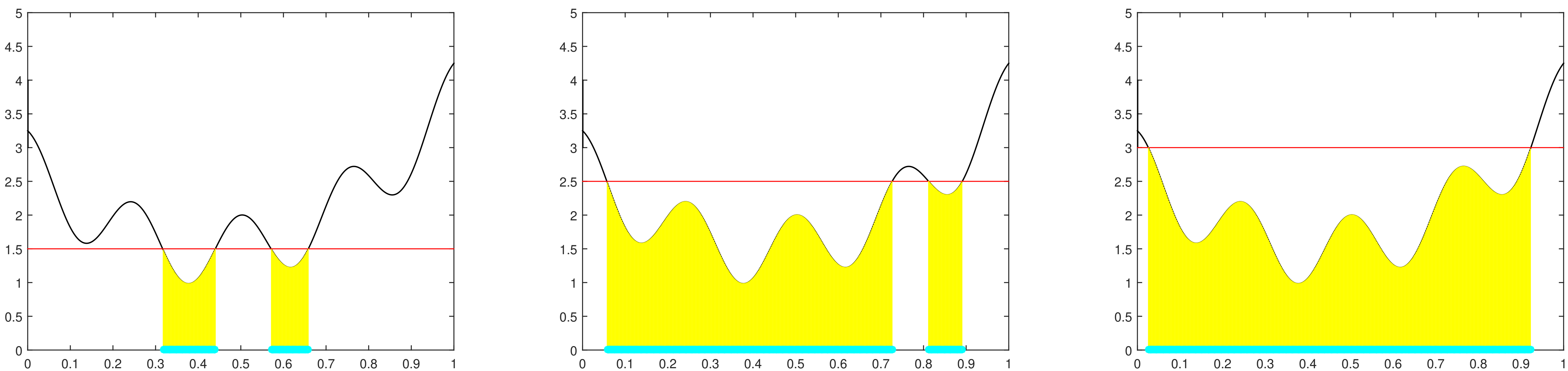
YMC-Hu-Wu-Lo, preprint

Step 1: Filtrations

Sublevel Set Filtration:

$$f_t^- := \{x \in P |\ f(x) \leq t\}$$

$$f_{t_0}^- \subset f_{t_1}^- \subset \cdots \subset f_{t_n}^- \qquad t_0 < t_1 < \cdots < t_n$$



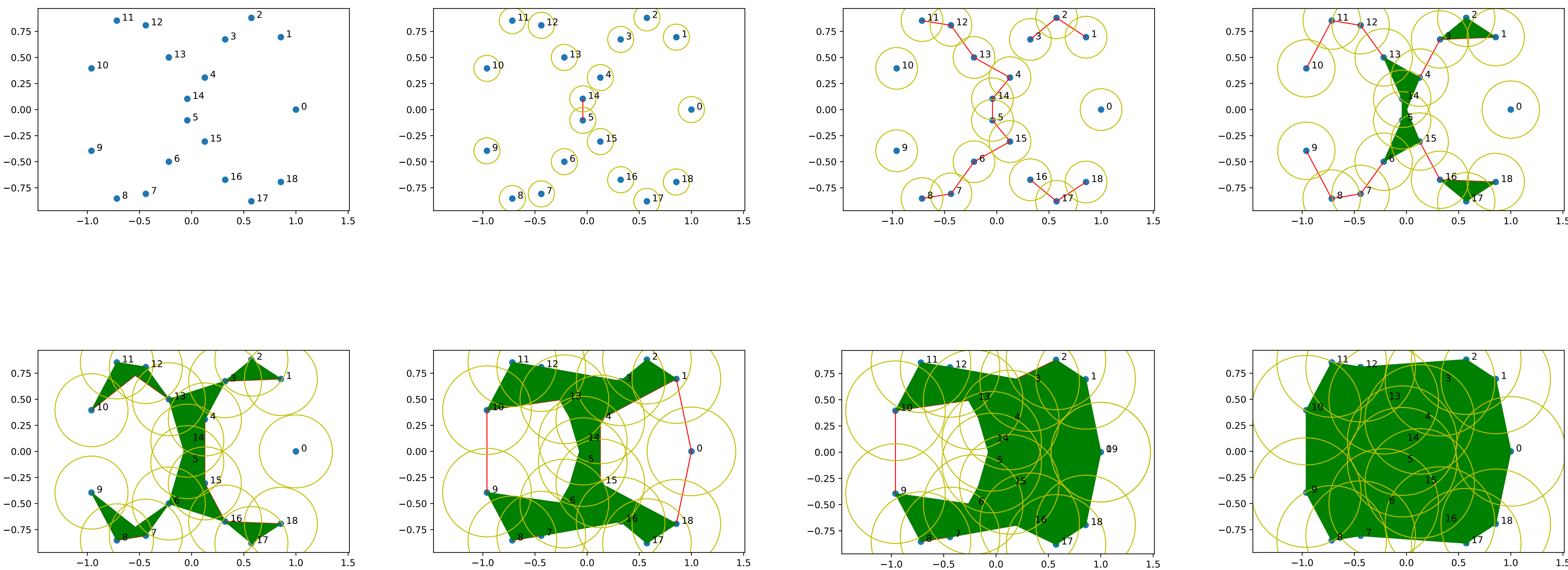
Shape of the function

THE UNIVERSITY *of* NORTH CAROLINA
GREENSBORO

CELEBRATING 125 YEARS OF OPPORTUNITY & EXCELLENCE

Step 1: Filtrations

VR Complex Filtration:



Shape of the point cloud

THE UNIVERSITY *of* NORTH CAROLINA
GREENSBORO

Celebrating 125 Years of Opportunity & Excellence

Step 2: Persistence Diagrams

DIPHA (A Distributed Persistent Homology Algorithm)

Project Founder:

Jan Reininghaus (Email: jan.reininghaus@gmail.com)

Contributors:

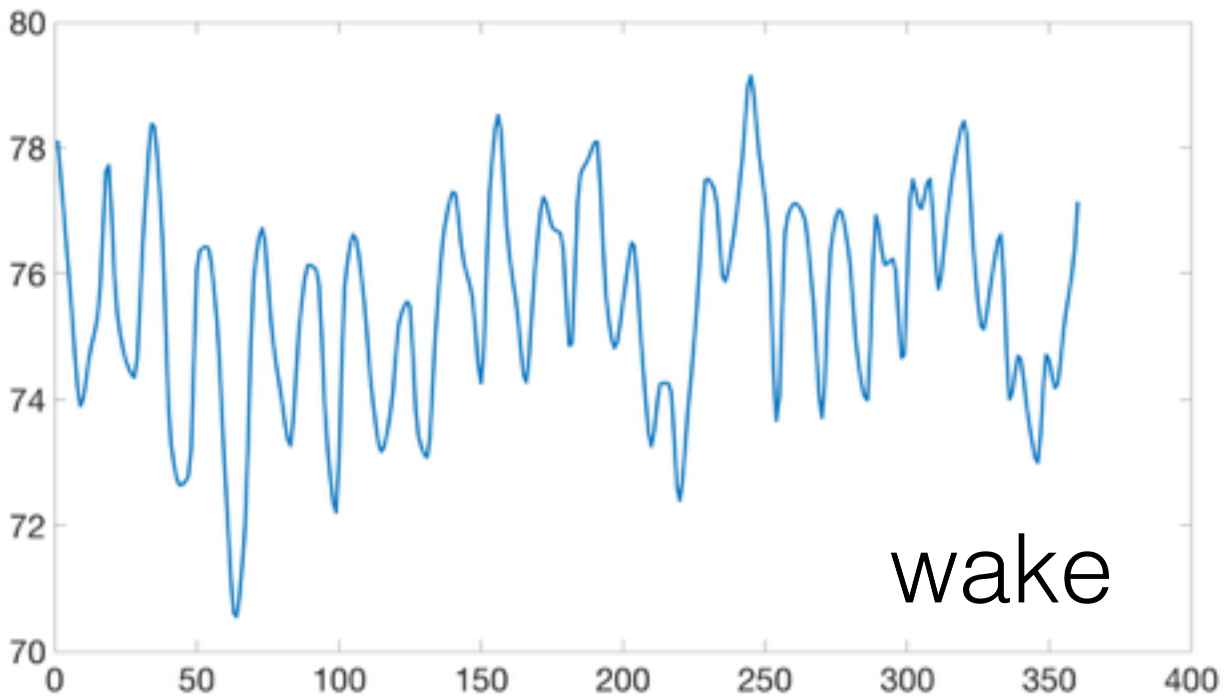
Ulrich Bauer, Michael Kerber

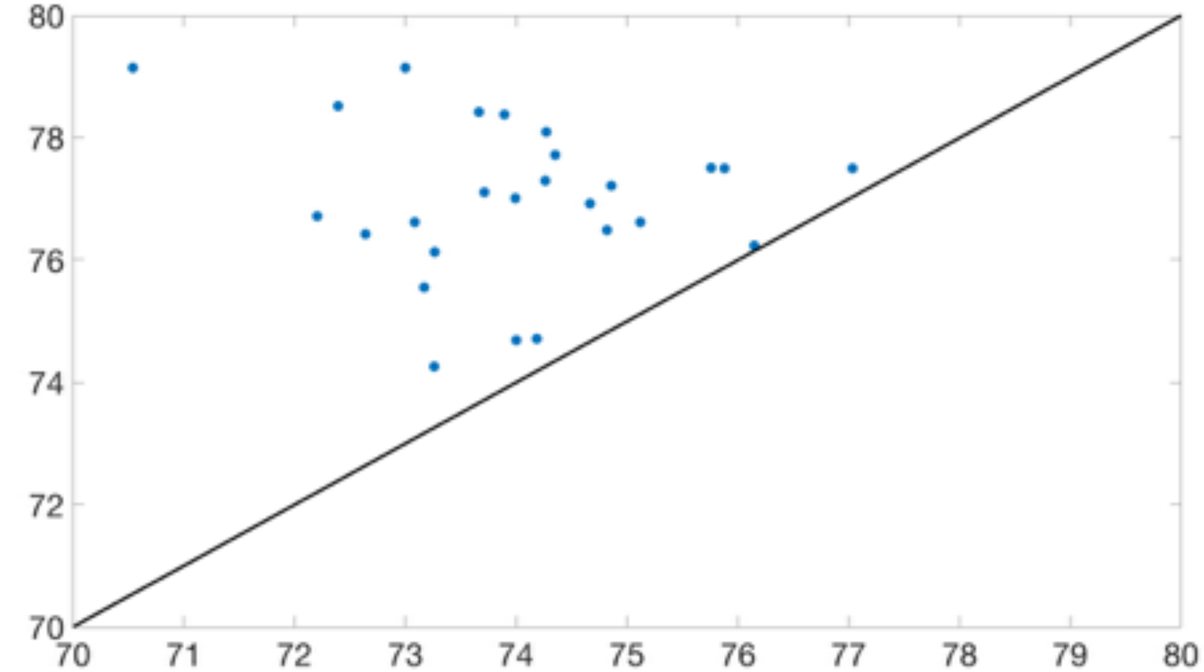
Ripser

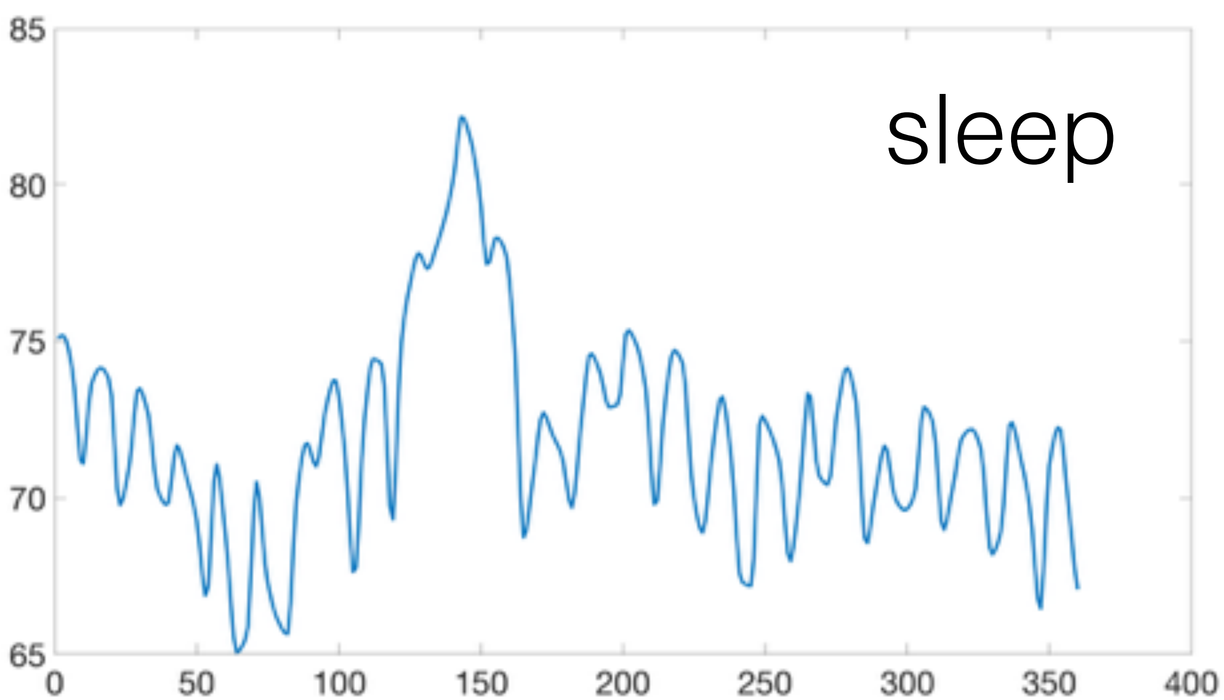
Description

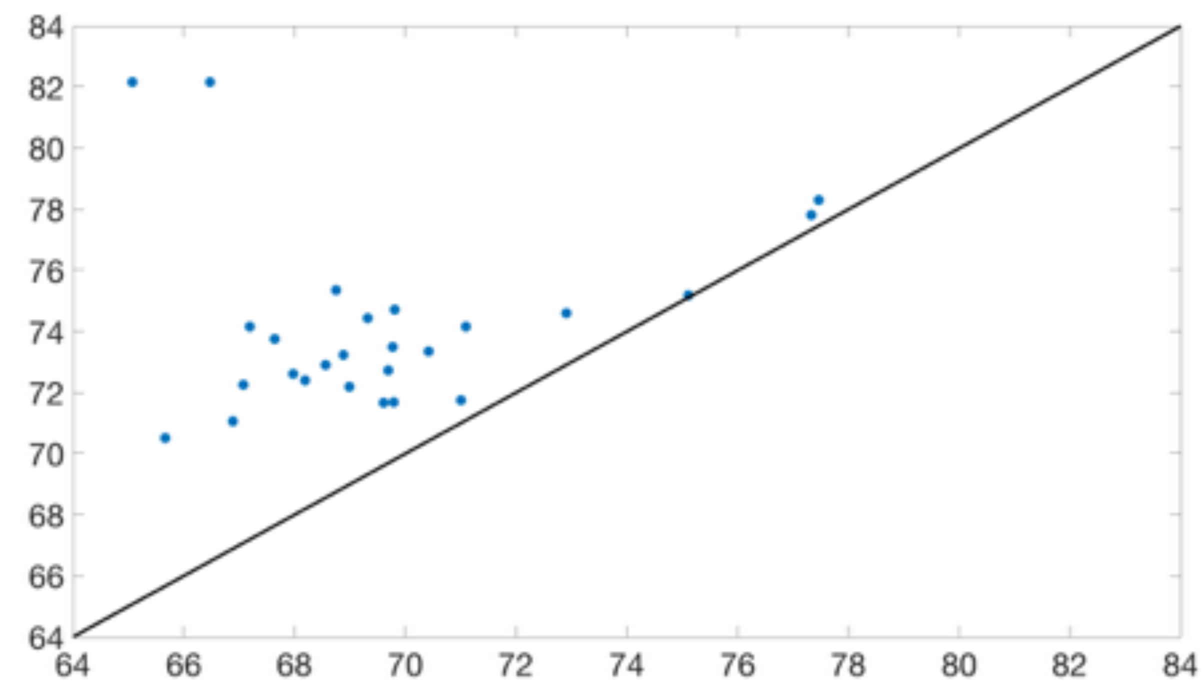
Ripser is a lean C++ code for the computation of Vietoris–Rips persistence barcodes. It can do just this one thing, but does it extremely well.

Step 2: Persistence Diagrams









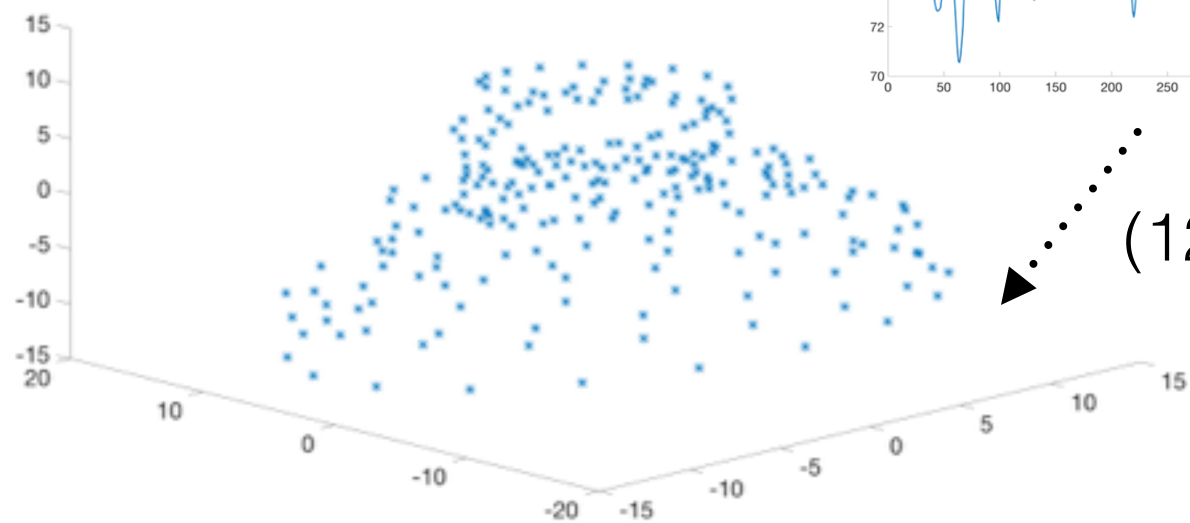
THE UNIVERSITY of NORTH CAROLINA
GREENSBORO

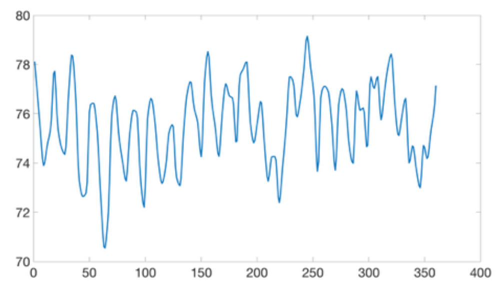
Sublevel Set Filtration

$\mathcal{P}(H_{90})$

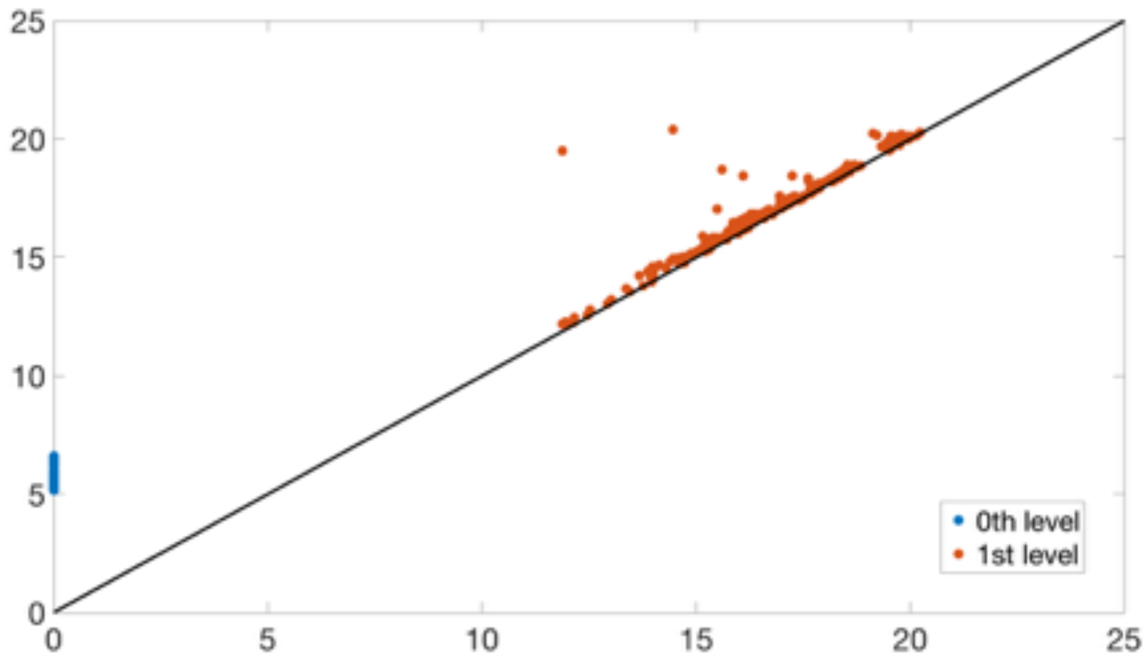
CELEBRATING 125 YEARS OF OPPORTUNITY & EXCELLENCE

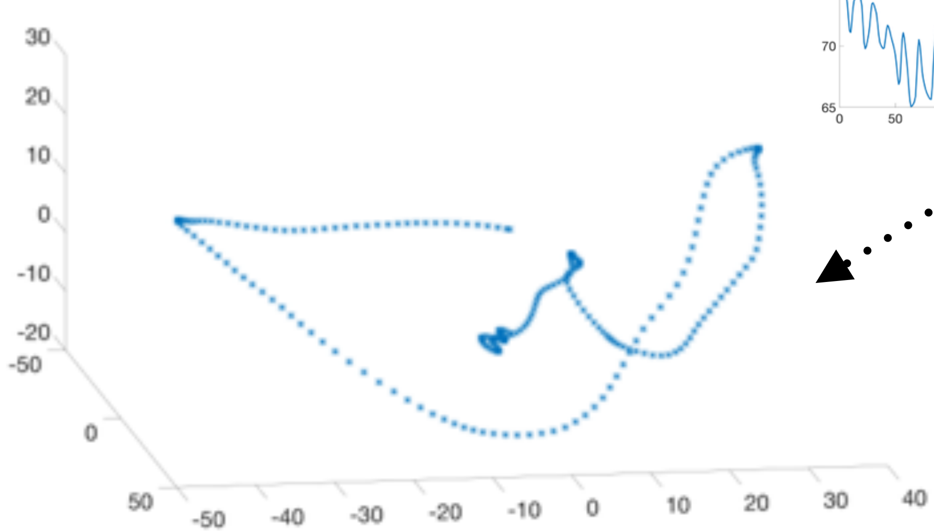
Step 2: Persistence Diagrams

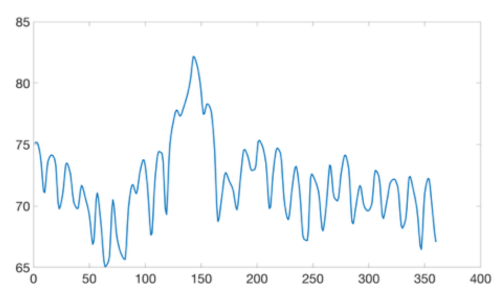




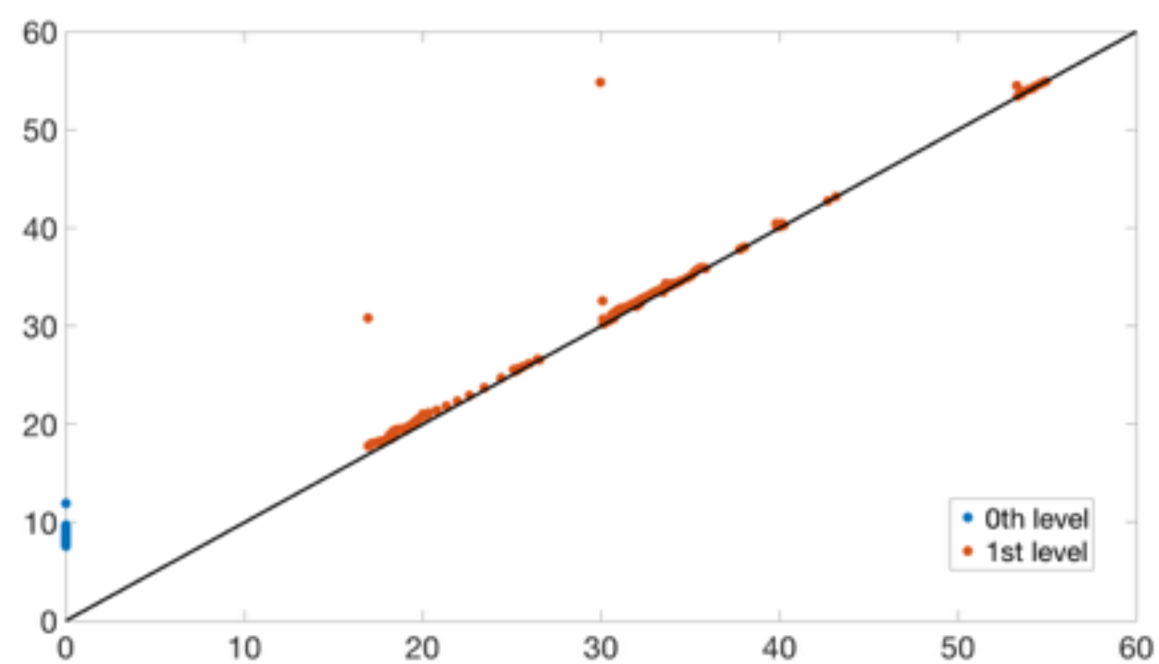
(120,1)







(120,1)



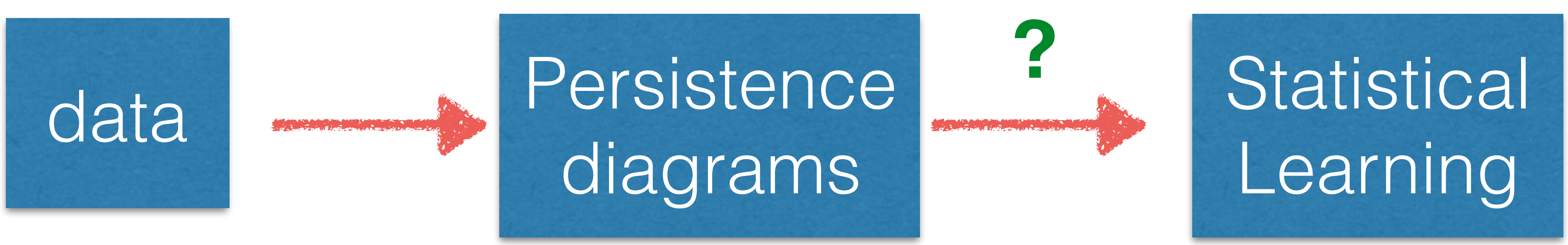
VR Complex Filtration

$\mathcal{P}(R_{120,1})$

Persistent Diagrams

data → Persistence diagrams

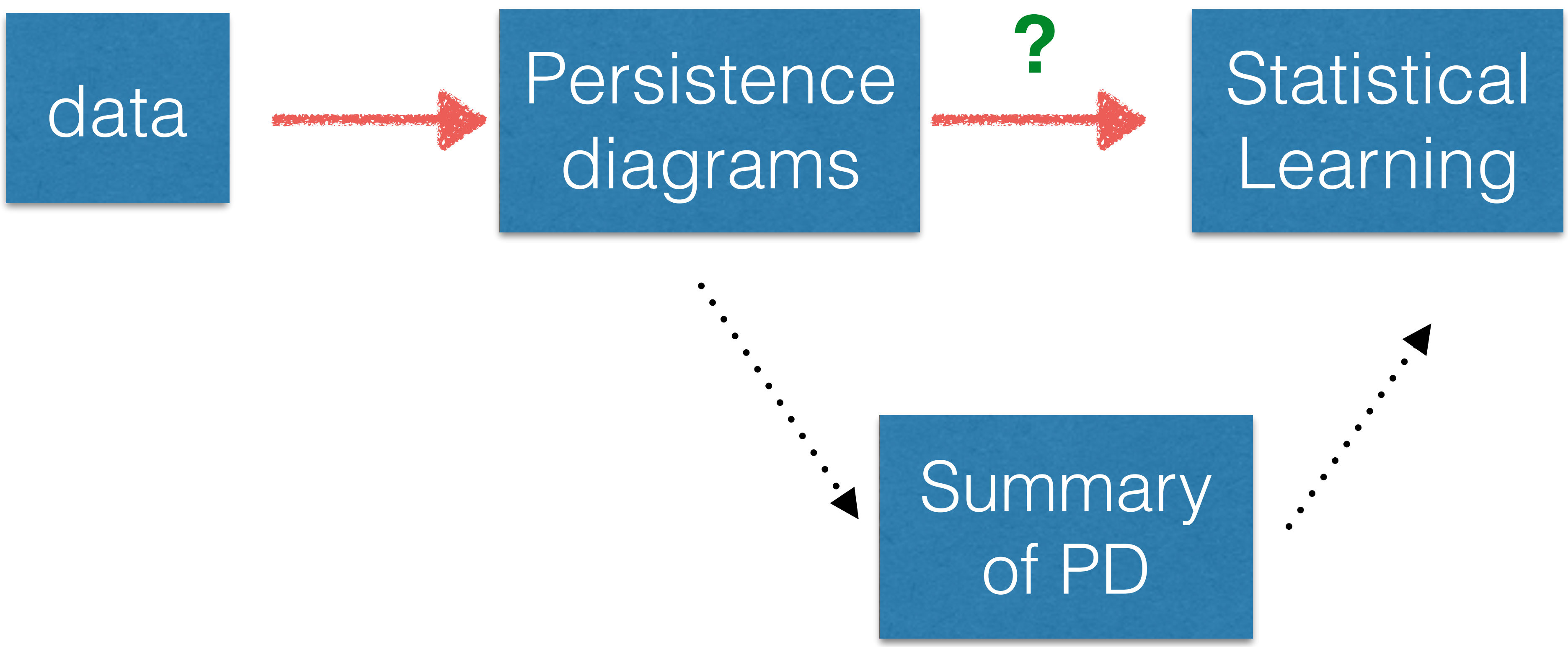
Persistent Diagrams



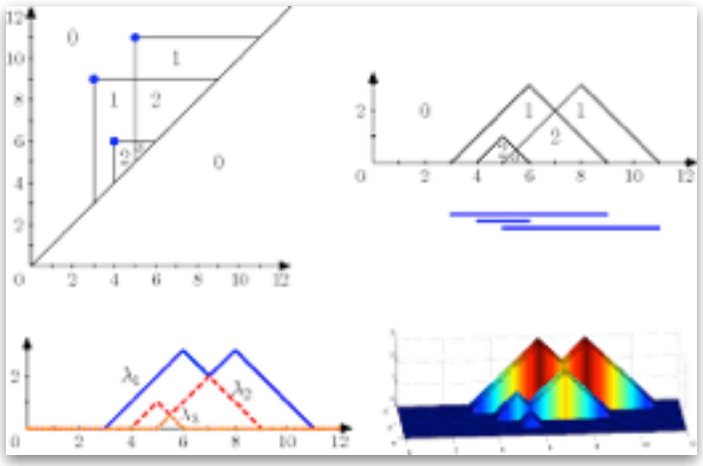
THE UNIVERSITY *of* NORTH CAROLINA
GREENSBORO

CELEBRATING
125 YEARS OF
OPPORTUNITY
& EXCELLENCE

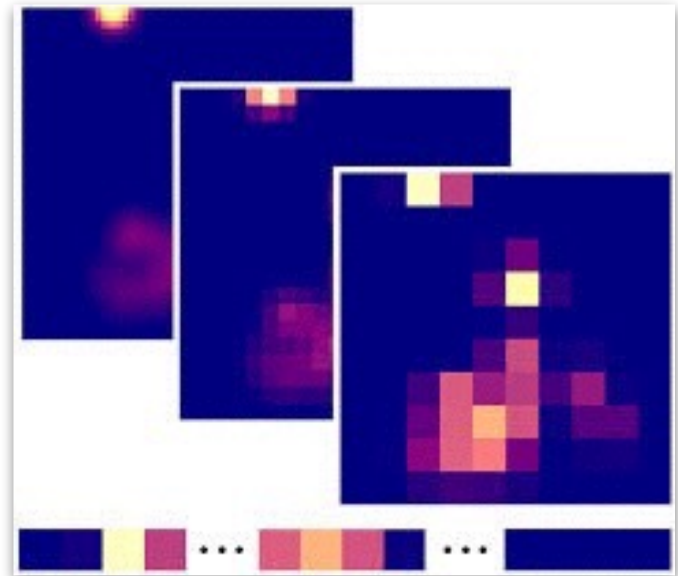
Persistent Diagrams



Summary of PD



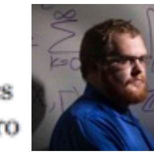
Bubenik '15 J. Mach. Learn. Res.



Adam, et. al. '17 J. Mach. Learn. Res.

Persistence Curves: A canonical framework for summarizing persistence diagrams

Yu-Min Chung and Austin Lawson
Department of Mathematics and Statistics
University of North Carolina at Greensboro



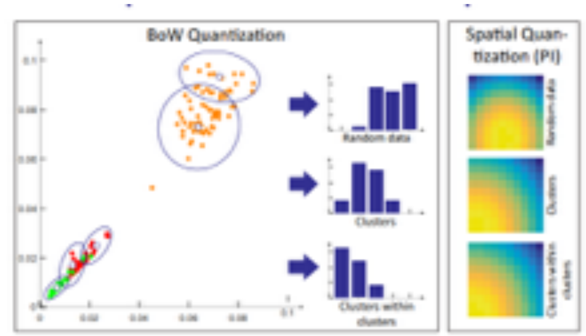
This CVPR2015 paper is the Open Access version, provided by the Computer Vision Foundation.
The authoritative version of this paper is available in IEEE Xplore.

A Stable Multi-Scale Kernel for Topological Machine Learning

Jan Reininghaus, Stefan Huber, IST Austria
Ulrich Bauer, IST Austria, TU München
Roland Kwitt, University of Salzburg, Austria

Persistence Codebooks for Topological Data Analysis

Bartosz Zieliński[1] Mateusz Juda[1] Matthias Zeppelzauer[2]



Persistence paths and signature features in topological data analysis

Ilya Chevyrev, Vidit Nanda, and Harald Oberhauser

IOP Publishing — Inverse Problems
Inverse Problems 27 (2011) 124007 (22pp) doi:10.1088/0266-5611/27/12/124007

Probability measures on the space of persistence diagrams

Yuriy Mileyko[1], Sayan Mukherjee[2] and John Harer[3]

Approximating Continuous Functions on Persistence Diagrams Using Template Functions

Jose A. Perea
Elizabeth Munch
Department of Computational Mathematics, Science, and Engineering; and Department of Mathematics
Firas A. Khasawneh
Department of Mechanical Engineering
Michigan State University
East Lansing, MI 48824, USA

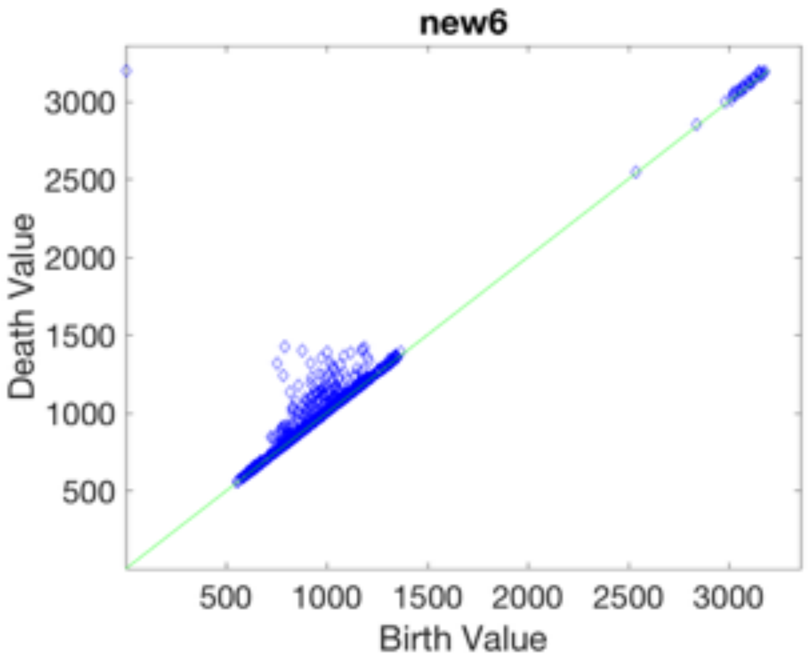
Persistence weighted Gaussian kernel for topological data analysis

Genki Kusano[1] GENKSN@GMAIL.COM
Kenji Fukumizu[2] FUKUMIZU@ISM.AC.JP
Yasuaki Hiraoka[1] HIRAOKA@WPI-AIMR.TOHOKU.AC.JP
[1]Tohoku University, [2]The Institute of Statistical Mathematics

THE UNIVERSITY of NORTH CAROLINA GREENSBORO

CELEBRATING 125 YEARS OF OPPORTUNITY & EXCELLENCE

Persistence Statistics



Single Number

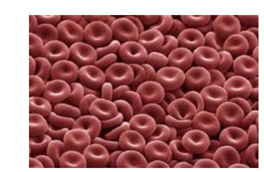
Step 3: Persistence Statistics

Topological Roughness

[YMC-Day-Costa, preprint]

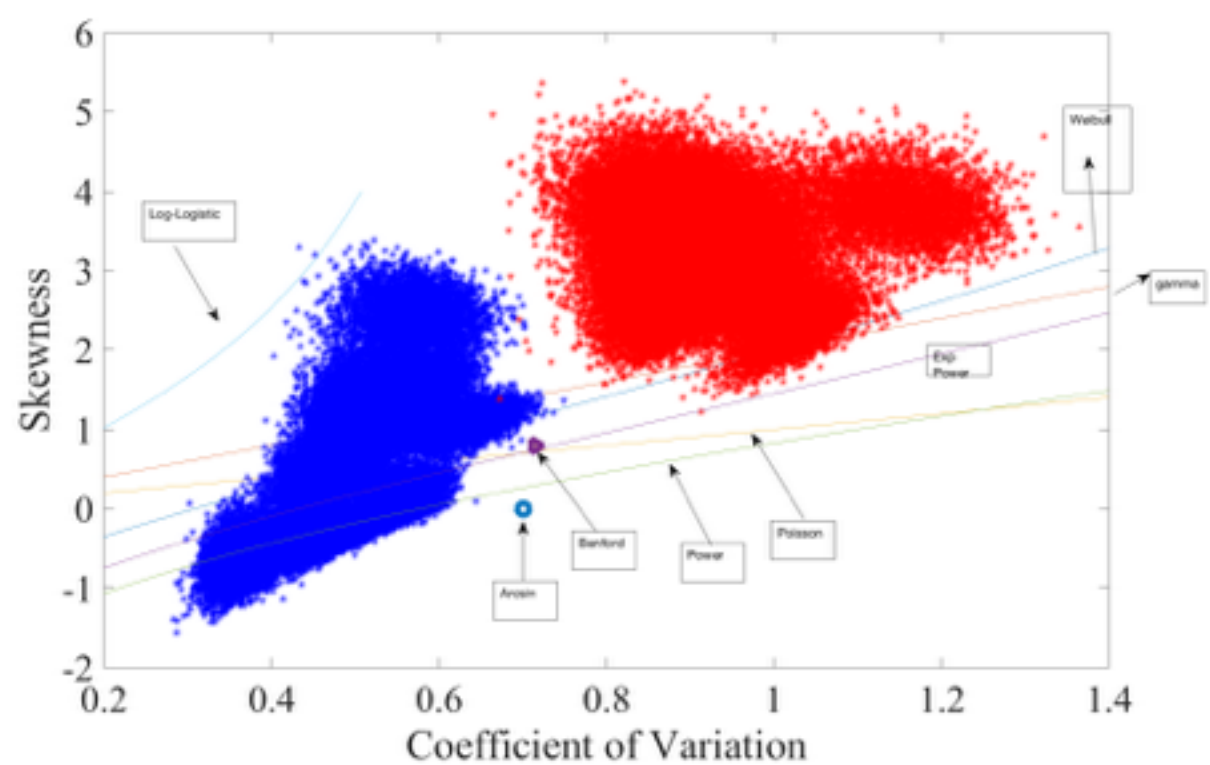
Label	Symbol	Name	Spatial	Dimensionless	Scale	Translation	Size	Stable
PS1	L	Total Lifespan	✓	✗	✗	✓	✗	✓
PS2	E	Lifespan Entropy	✓	✓	✓	✓	✗	✓
PS3	Λ	Total Multiplicative Lifespan	✓	✓	✓	✗	✗	*
PS4	E_Λ	Multiplicative Lifespan Entropy	✓	✓	✓	✗	✗	*
PS5	CV	Midlife Coefficient of Variation	✓	✓	✓	✗	✓	✗
PS6	SK	Midlife Skewness	✓	✓	✓	✓	✓	✗
PS7	KU	Midlife Kurtosis	✓	✓	✓	✓	✓	✗
-	R_{rms}	Intensity root mean squared[17]	✗	✗	✗	✓	✓	✓
-	R_{sk}	Intensity skewness [17]	✗	✗	✗	✓	✓	✓
-	R_{ku}	Intensity kurtosis [17]	✗	✗	✗	✓	✓	✓

TABLE 1. Properties of Persistence Statistics (PS). *Up to authors' knowledge, it is still an open question. Scale, Translation, and Size refer to Scale Invariant, Translation Invariant, and Size Invariant, respectively.



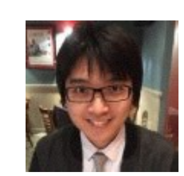
Blue: Old Cell

Red: New Cell



Madalena Costa
Harvard Medical School

Sarah Day
College of William & Mary

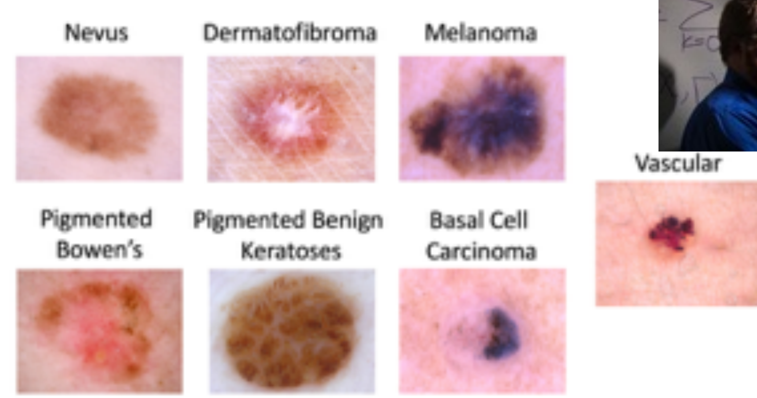


THE UNIVERSITY of NORTH CAROLINA GREENSBORO

CELEBRATING 125 YEARS OF OPPORTUNITY & EXCELLENCE

Step 3: Persistence Statistics

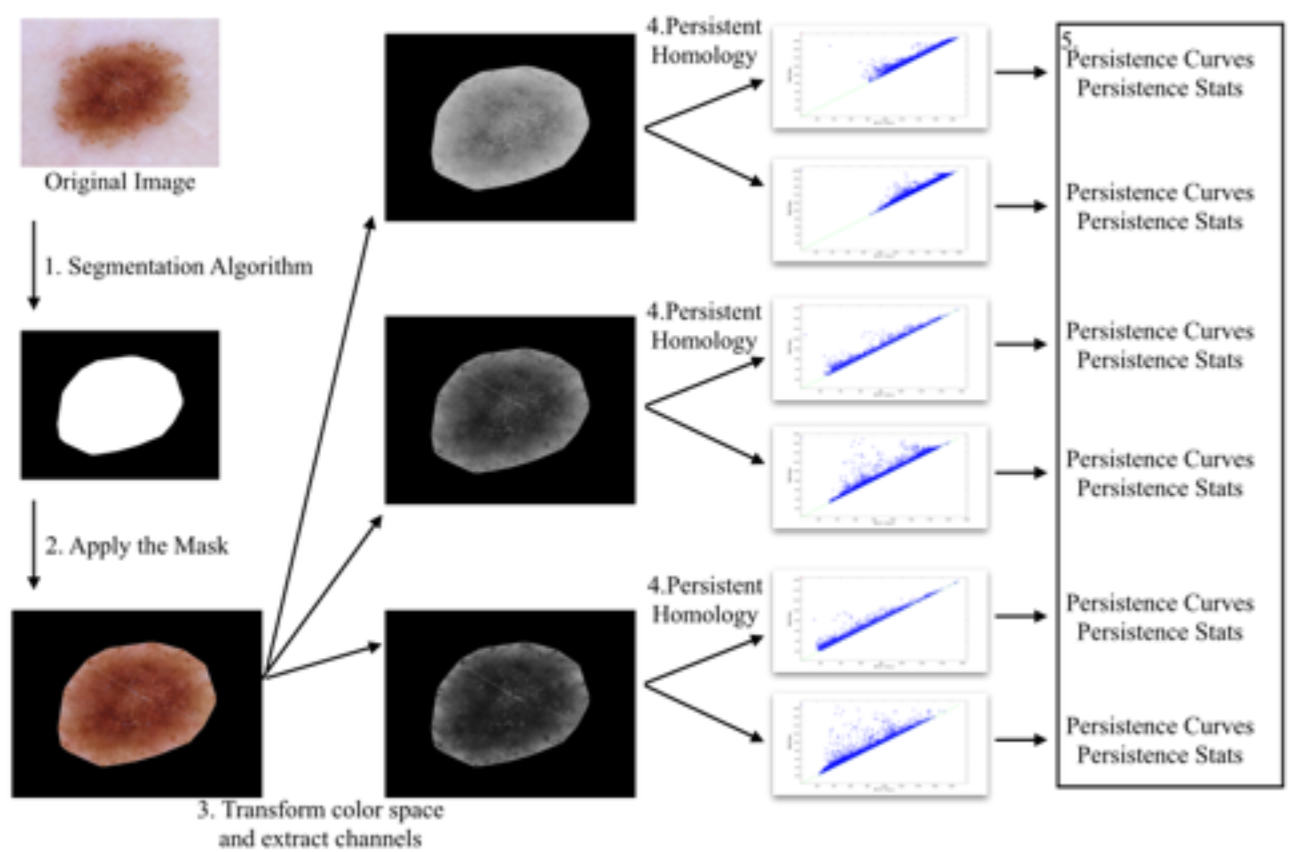
- Skin Lesion



Peter Hu
CS, NTNU, Taiwan

Clifford Smyth
UNCG

IEEE Big Data 2018



THE UNIVERSITY of NORTH CAROLINA
GREENSBORO

Celebrating 125 Years of Opportunity & Excellence

Step 3: Persistence Statistics

$$M = \{d + b \mid (b, d) \in \mathcal{P}\}, \text{ and } L = \left\{ \frac{d-b}{\sum d - b} \mid (b, d) \in \mathcal{P} \right\}.$$

$$\boxed{\Phi(\mathcal{P}) := \begin{bmatrix} \mathcal{S}(M) \\ \mathcal{S}(L) \\ E(L) \end{bmatrix} \in \mathbb{R}^{19}}$$

$$E(L) = -\sum_{l \in L} l \log(l).$$

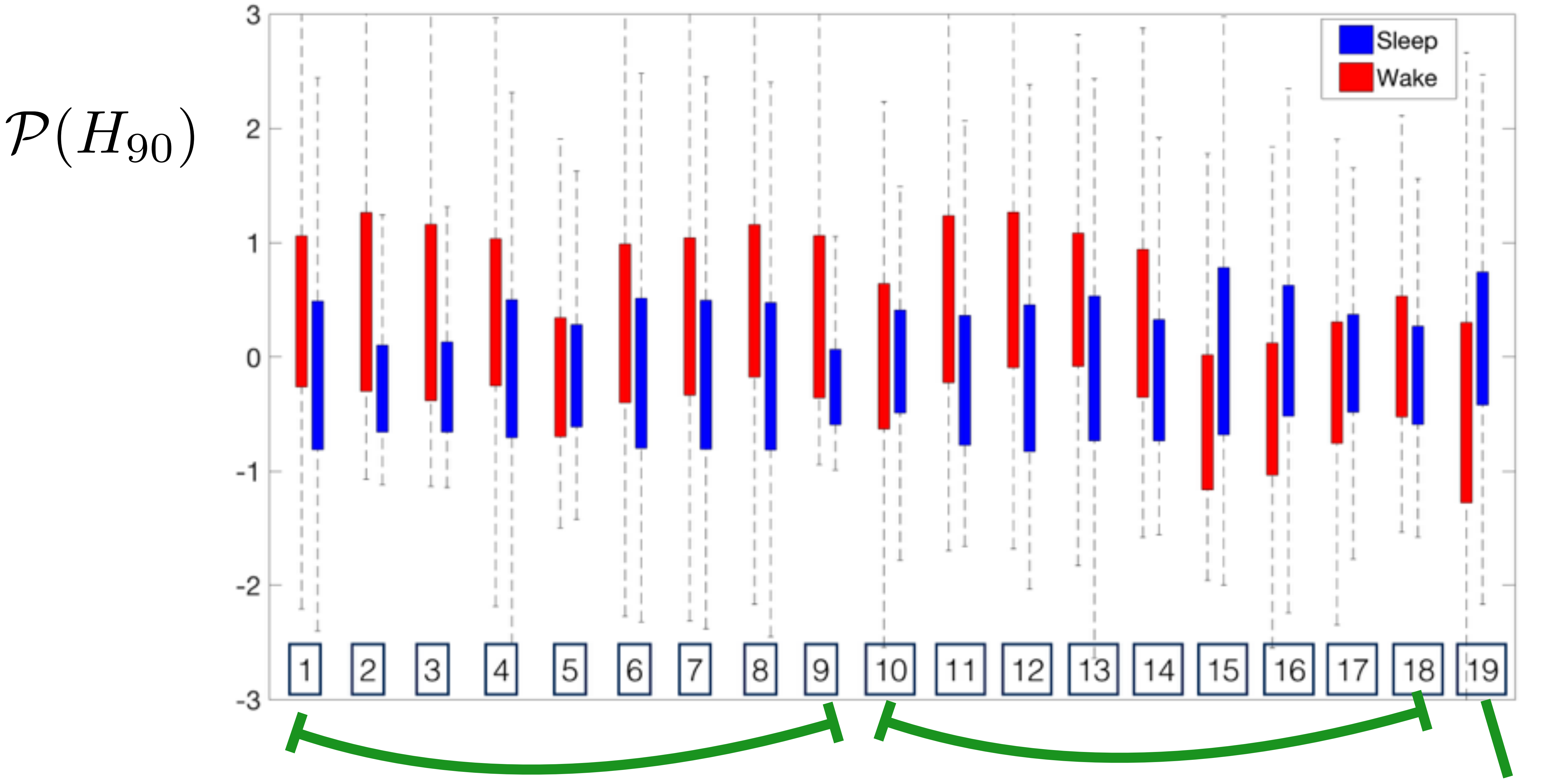
Label	Measurement
1	mean
2	standard deviation
3	coefficient of variation
4	skewness
5	kurtosis
6-8	25, 50, 75 percentile
9	interquartile range

TABLE 1. Statistical measurements considered in $\mathcal{S}$.

Pattern Recognition

An entropy-based persistence barcode

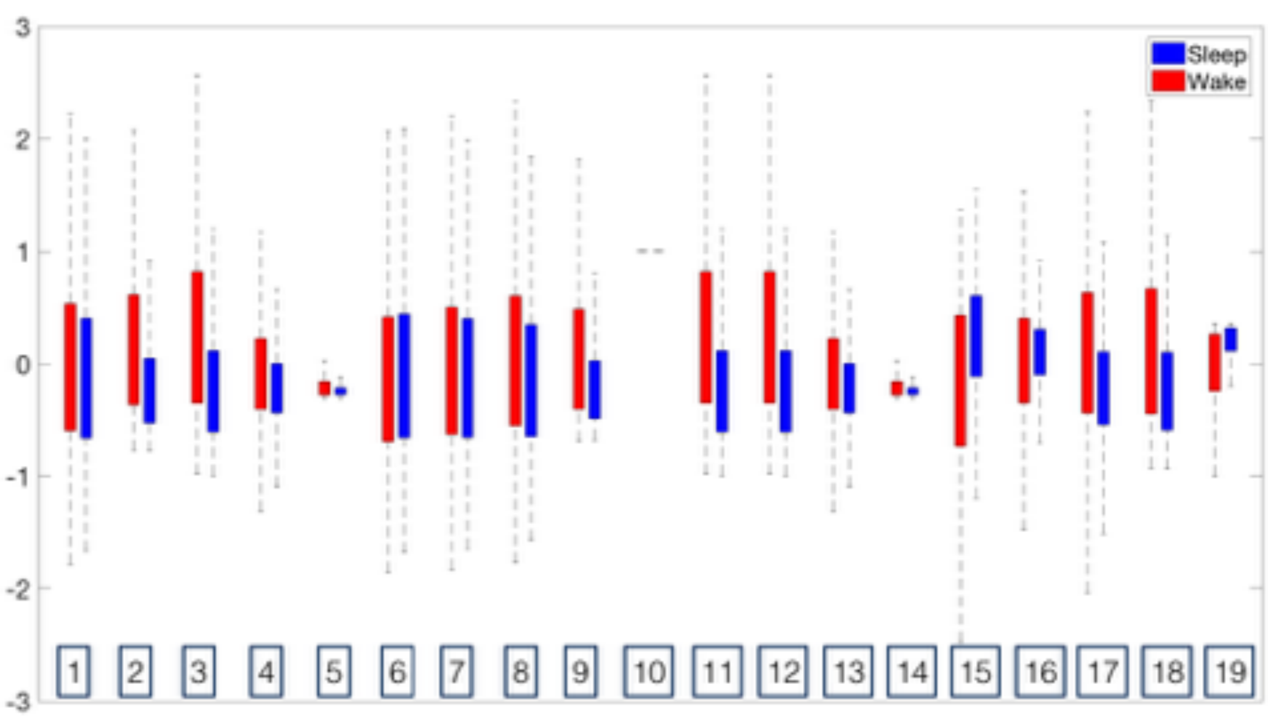
Step 3: Persistence Statistics



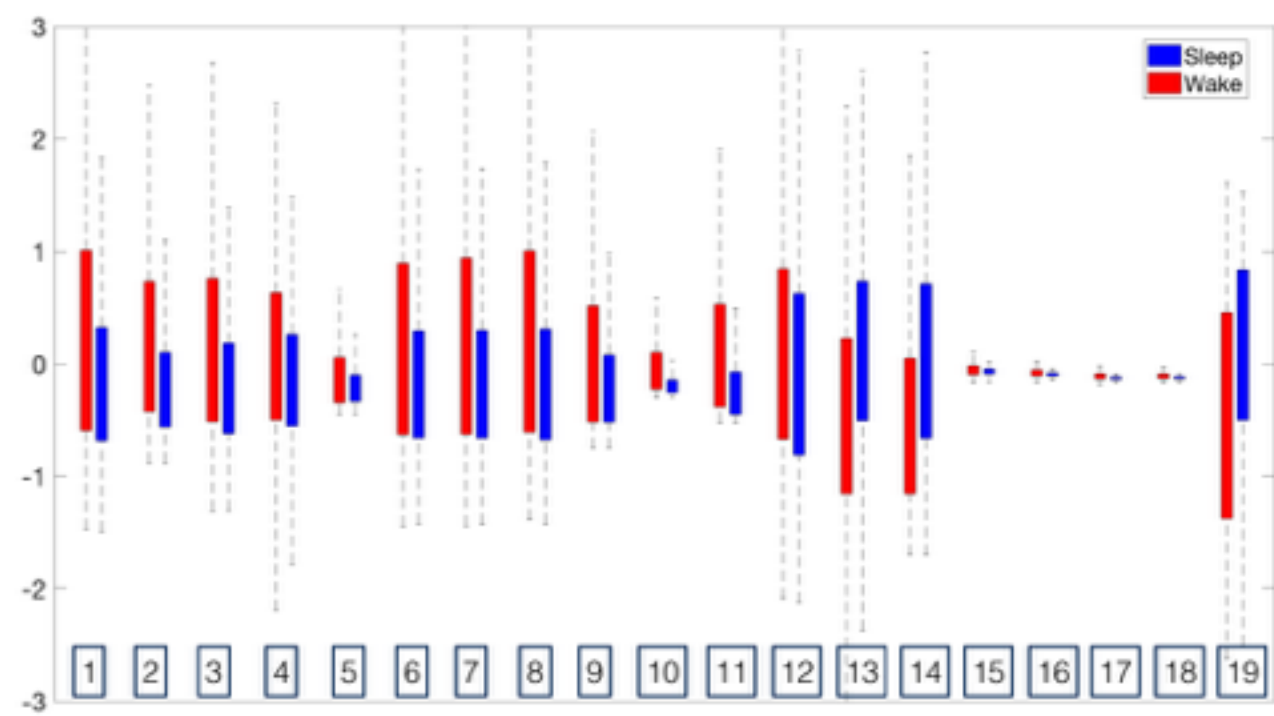
Label	Measurement
1	mean
2	standard deviation
3	coefficient of variation
4	skewness
5	kurtosis
6-8	25, 50, 75 percentile
9	interquartile range

TABLE 1. Statistical measurements considered in $\mathcal{S}$.

Step 3: Persistence Statistics



$$\Phi(\mathcal{P}_0(R_{120,1}))$$

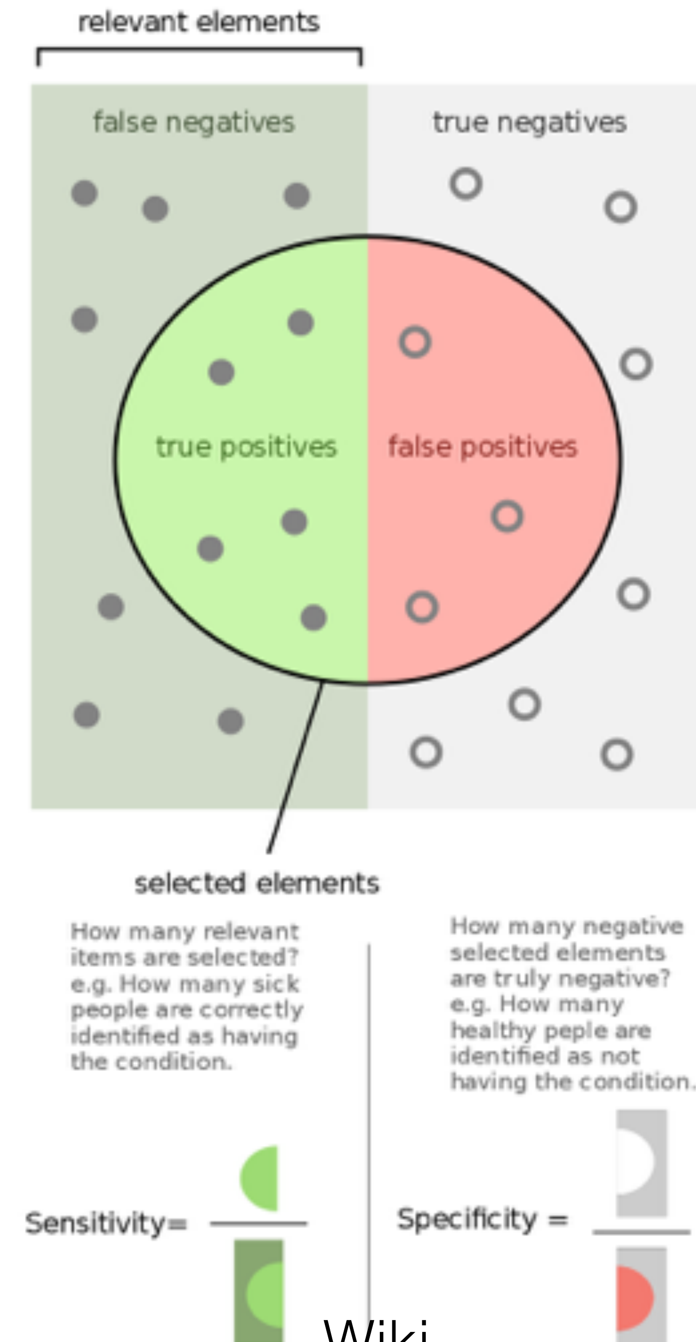


$$\Phi(\mathcal{P}_1(R_{120,1}))$$

Step 4: Machine Learning

Evaluation Metrics

Actual \ Predict	Wake	Sleep
Wake	TP	FN
Sleep	FP	TN



Wiki

- $SE = TP/(TP + FN)$
- $SP = TN/(TN + FP)$

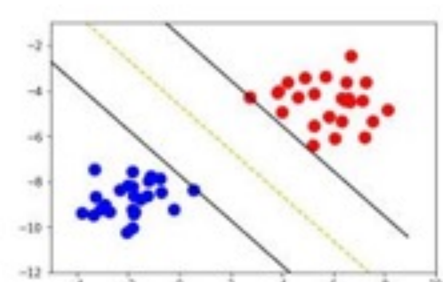
Step 4: Machine Learning

Training process

|W| << |S|

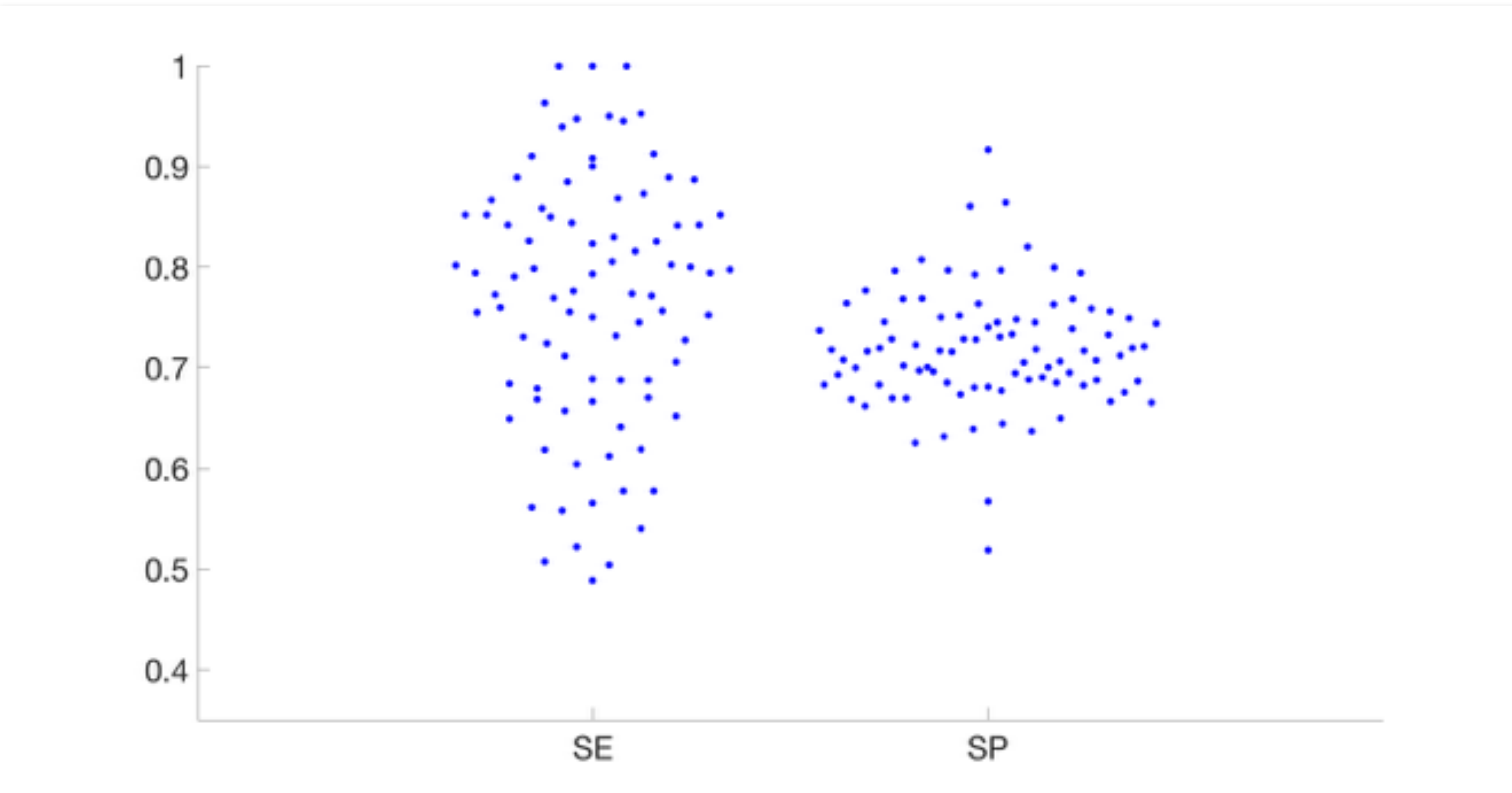
1. Use all W
2. Adopt a subsampling process on S
3. Select |W| epochs from S

Support Vector Machine (SVM)



Step 4: Machine Learning

Leave-One-Out-Cross-Validation



$76\% \pm 13.94\%$ $71.89\% \pm 5.8\%$

5 fold CV

Trial	SE	SP
1	72.5129	69.0295
2	74.6411	71.0923
3	70.4866	73.3053
4	71.0153	71.2564
5	66.8810	72.7040

Result

YMC-Hu-Wu-Lo, preprint

(B) $(\Phi(\mathcal{P}_0(R_{120,1})),\ \Phi(\mathcal{P}_1(R_{120,1})),\ \Phi(\mathcal{P}(H_{90})))$

	CGMH-training	CGMH-validation	DREAMS Subjects	UCDSADB
TP	6824 ± 32	2301 ± 57	2059 ± 132	2111 ± 200
FP	15587 ± 177	2036 ± 289	4846 ± 647	5125 ± 755
TN	38960 ± 177	11490 ± 289	10501 ± 647	9787 ± 755
FN	2327 ± 32	875 ± 57	1079 ± 132	1821 ± 200
SE (%)	74.6 ± 0.4	72.5 ± 1.8	65.6 ± 4.2	53.7 ± 5.1
SP (%)	71.4 ± 0.3	74 ± 1.9	68.4 ± 4.2	65.6 ± 5.1
ACC (%)	71.9 ± 0.2	73.4 ± 1.2	68 ± 2.8	63.1 ± 3
PR (%)	20.5 ± 0.2	36.4 ± 1.1	30 ± 1.5	29.3 ± 1.1
F1	0.432 ± 0.001	0.484 ± 0.006	0.411 ± 0.006	0.379 ± 0.004
AUC				
Kappa	0.287 ± 0.002	0.333 ± 0.011	0.232 ± 0.016	0.148 ± 0.01

TABLE 6. SVM performance.

TABLE 1. Performance statistics for the CNN model trained on CGMH-training

	CGMH -training	CGMH -validation	DREAMS Subjects	UCDSADB
TP	4,464	1,800	1,777	1,838
FP	2,143	1,763	2,151	2,853
TN	31,550	14,906	14,532	12,883
FN	3,315	1,633	1,572	2,400
SE (%)	57.4	52.4	53.1	43.4
SP (%)	93.6	89.4	87.1	81.9
ACC (%)	86.8	83.1	81.4	73.7
PR (%)	67.6	50.5	45.2	39.2
F1	0.62	0.51	0.49	0.41
AUC	0.90	0.83	0.81	0.72
Kappa	0.54	0.41	0.38	0.24

J. Malik, Y.-L. Lo, H.-T. Wu '18
Physiological Measurement

Deep learning

THE UNIVERSITY of NORTH CAROLINA
GREENSBORO

CELEBRATING 125 YEARS OF OPPORTUNITY & EXCELLENCE

Result

YMC-Hu-Wu-Lo, preprint

	CGMH-training	CGMH-validation	DREAMS	UCDSADB
TP	6316 ± 728	1864 ± 132	1994 ± 19	1880 ± 23
FP	18131 ± 5166	4095 ± 562	3453 ± 83	4335 ± 78
TN	36416 ± 5166	11431 ± 562	11894 ± 83	10577 ± 78
FN	2834 ± 728	1312 ± 132	1145 ± 19	2052 ± 23
SE (%)	69.0 ± 8.0	58.7 ± 4.1	63.5 ± 0.6	47.8 ± 0.6
SP (%)	66.8 ± 9.5	73.6 ± 3.6	77.5 ± 0.5	70.9 ± 0.5
ACC (%)	67.1 ± 7.0	71.1 ± 2.3	75.1 ± 0.4	66.1 ± 0.3
PR (%)	26.6 ± 3.1	31.4 ± 1.5	36.6 ± 0.4	30.3 ± 0.2
F1	0.380 ± 0.021	0.408 ± 0.006	0.465 ± 0.002	0.371 ± 0.002
AUC				
Kappa	0.217 ± 0.039	0.240 ± 0.014	0.318 ± 0.004	0.154 ± 0.003

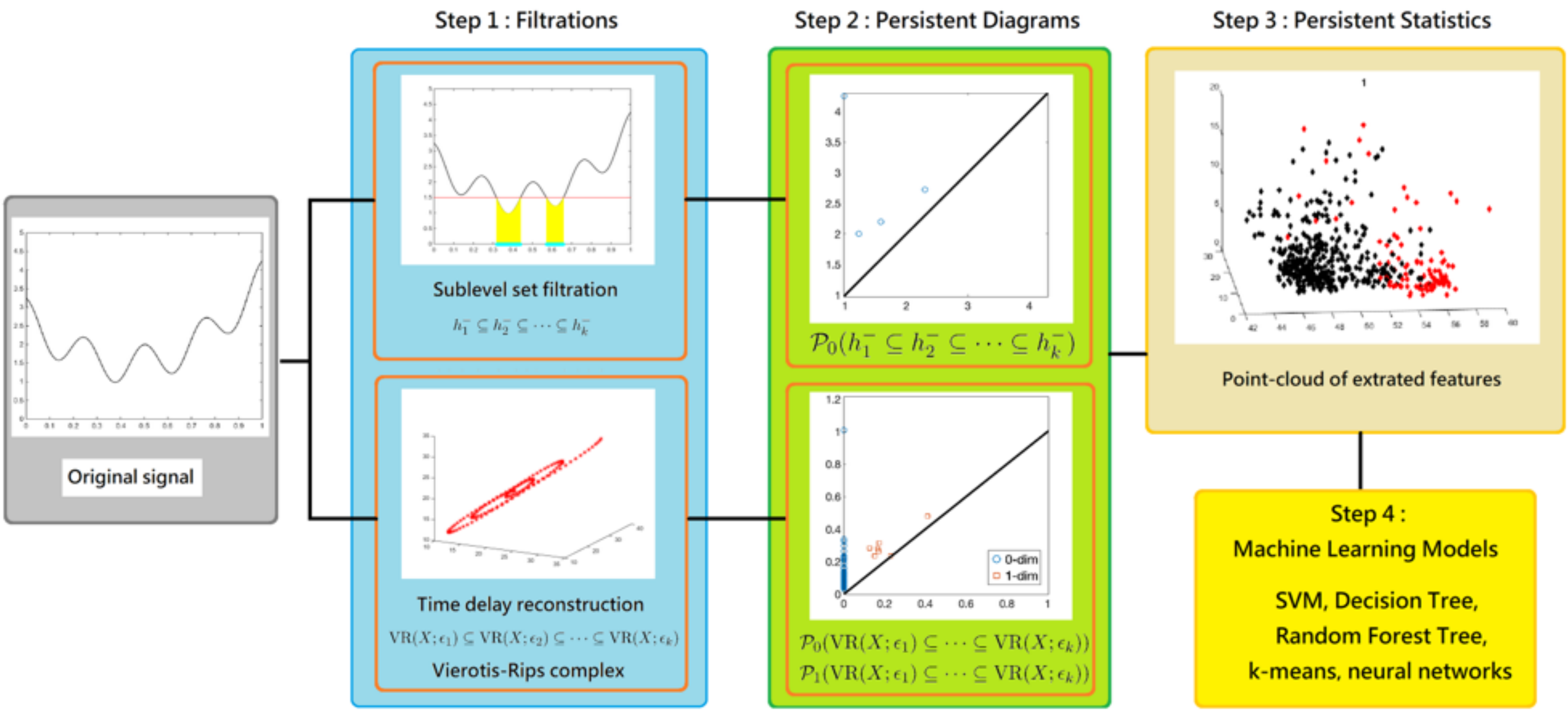
(B) $(\Phi(\mathcal{P}_0(R_{120,1})),\ \Phi(\mathcal{P}_1(R_{120,1})),\ \Phi(\mathcal{P}(H_{90})))$

TABLE 3. SVM performance (trained on DREAMS dataset).

J. Malik, Y.-L. Lo, H.-T. Wu ’18
Physiological Measurement

Training \ Validation		CGMH -validation	DREAMS Subjects	UCDSADB
DREAMS Subjects	TP	1,517	2,095	1,166
	FP	1,928	1,396	1,713
	TN	14,741	15,287	14,023
	FN	1,916	1,254	3,072
	SE	44.2	62.6	27.5
	SP	88.4	91.6	89.1
	ACC	80.9	86.8	76.0
	PR	44.0	60.0	40.5
	F1	0.44	0.61	0.33
	AUC	0.77	0.89	0.67
	Kappa	0.33	0.53	0.19

Summary


Step 1 : Filtrations
Step 2 : Persistent Diagrams
Step 3 : Persistent Statistics
Original signal
Sublevel set filtration
$h_1^- \subseteq h_2^- \subseteq \cdots \subseteq h_k^-$
Time delay reconstruction
$\mathrm{VR}(X;\epsilon_1) \subseteq \mathrm{VR}(X;\epsilon_2) \subseteq \cdots \subseteq \mathrm{VR}(X;\epsilon_k)$
Vierotis-Rips complex
$\mathcal{P}_0(h_1^- \subseteq h_2^- \subseteq \cdots \subseteq h_k^-)$
0-dim
1-dim
$\mathcal{P}_0(\mathrm{VR}(X;\epsilon_1) \subseteq \cdots \subseteq \mathrm{VR}(X;\epsilon_k))$
$\mathcal{P}_1(\mathrm{VR}(X;\epsilon_1) \subseteq \cdots \subseteq \mathrm{VR}(X;\epsilon_k))$
Point-cloud of extrated features
Step 4 :
Machine Learning Models
SVM, Decision Tree,
Random Forest Tree,
k-means, neural networks

THE UNIVERSITY of NORTH CAROLINA
GREENSBORO
Department of Mathematics & Statistics

Department of Mathematics and Statistics

Statistical Consulting Center

AWM

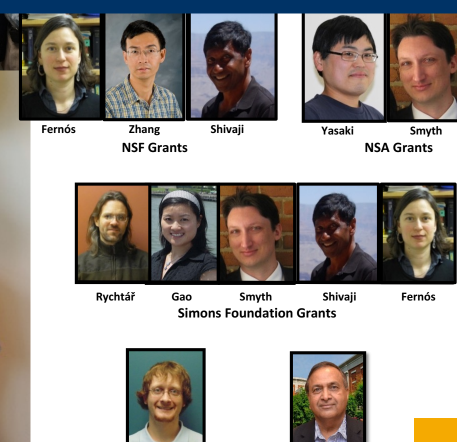
UNCG MATH SCHOLARSHIPS
Department of Mathematics & Statistics

NSF-REU

Math Help Center

State Math Contest

Recent Major External Award Recipients



Fernós Zhang Shivaji
NSF Grants

Yasaki Smyth
NSA Grants

Rychtář Gao Smyth Shivaji Fernós
Simons Foundation Grants

Lewis
MAA
NExT Fellow

Gupta
Fellow of ASA,
Sankhyiki Bhushan Award

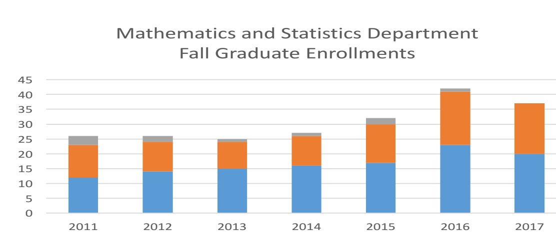
Graduate Tea

uncg.edu/mat

Math Club

Graduate Programs

Thank you!



PhD in Computational Mathematics

- Unique in North Carolina
- Began in 2008
- Integrates a computational component in each dissertation
- Prepares graduates for careers in industry or academia
- 12 Ph.D. alumni

MA in Mathematics

Mathematics	Applied Statistics	Data Analytics	Actuarial Mathematics	Teaching College Mathematics
Prepares students for PhD in Mathematics	Prepares students to work in industry or pursue a PhD in Statistics	Prepares students to work with "Big Data"	Prepares students for career as an actuary.	Prepares students for College Teaching

THE UNIVERSITY of NORTH CAROLINA
GREENSBORO

CELEBRATING 125 YEARS OF OPPORTUNITY & EXCELLENCE