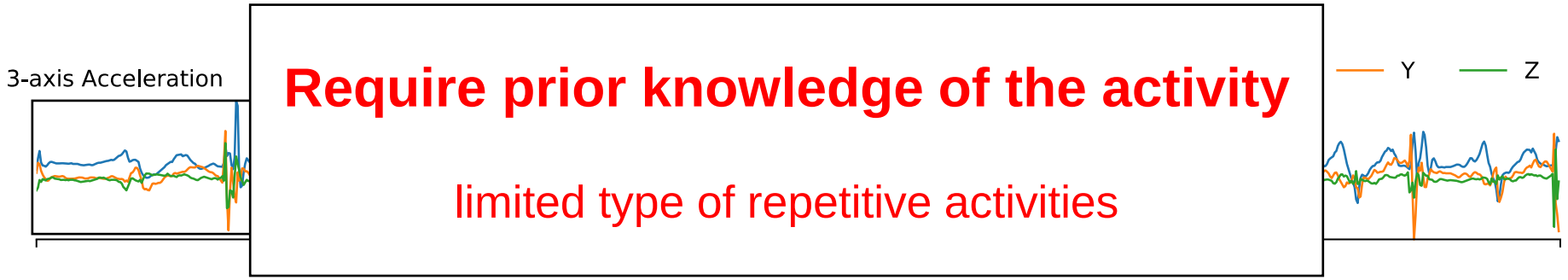
mSIMPAD: Repetitive Activity Detection

What?

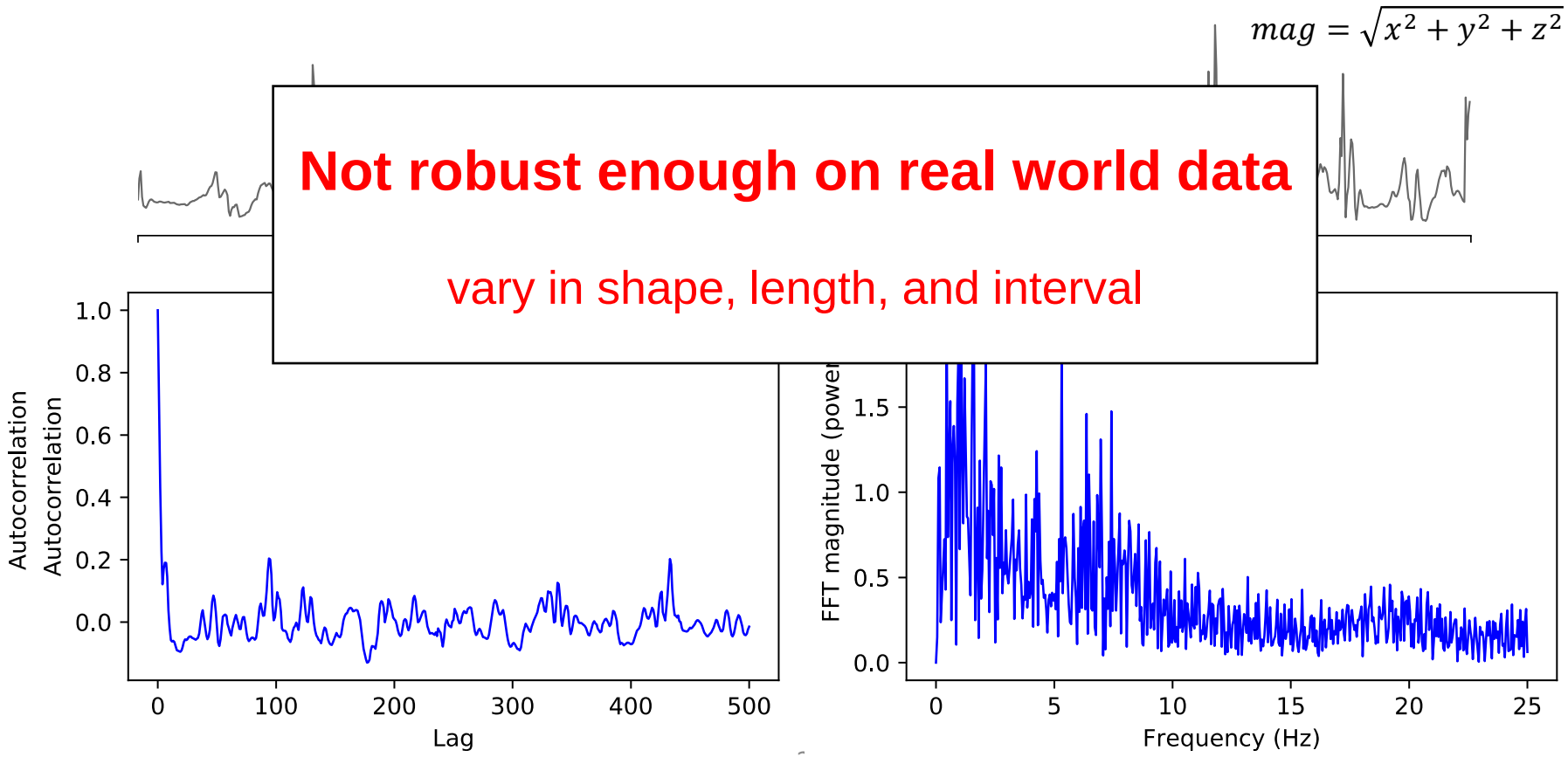
- Identify if a person performs repetitive activity using data collected from mobile devices.



Typical human activity recognition (HAR) methods:

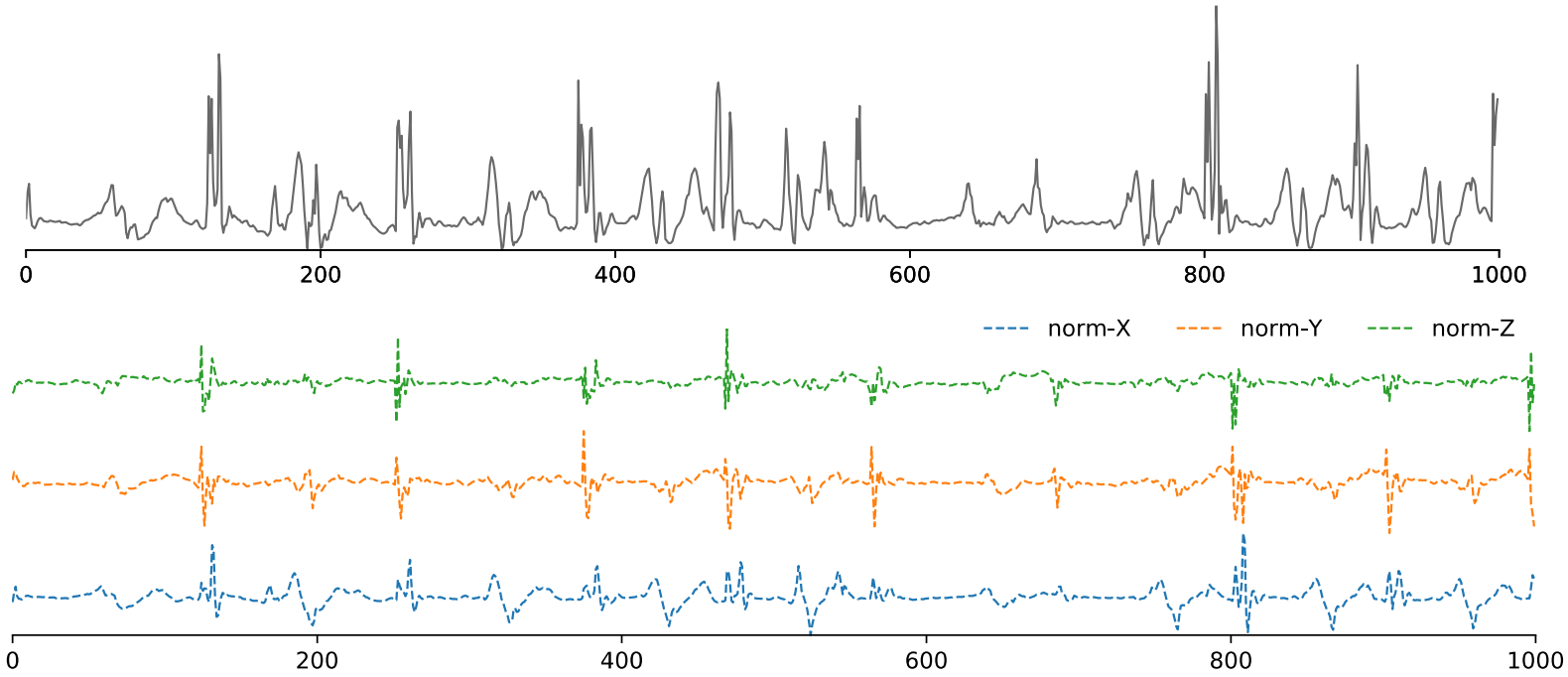
Possible Solution II: Repetitive Activity Detection

- Determine if the input data contains repetitive patterns with Autocorrelation / Frequency Domain Analysis



Proposed Solution

- Find if there contains any repetitive patterns in the time series
  - **Generality** - unlimited type of repetitive activities
  - **Robustness** - handle variable shape, length, and interval



Problem Formulation

- **Definitions**
  - A time series $T = [t_1, t_2, \dots, t_n]$
  - A su
  - $T_{i,l}$ i

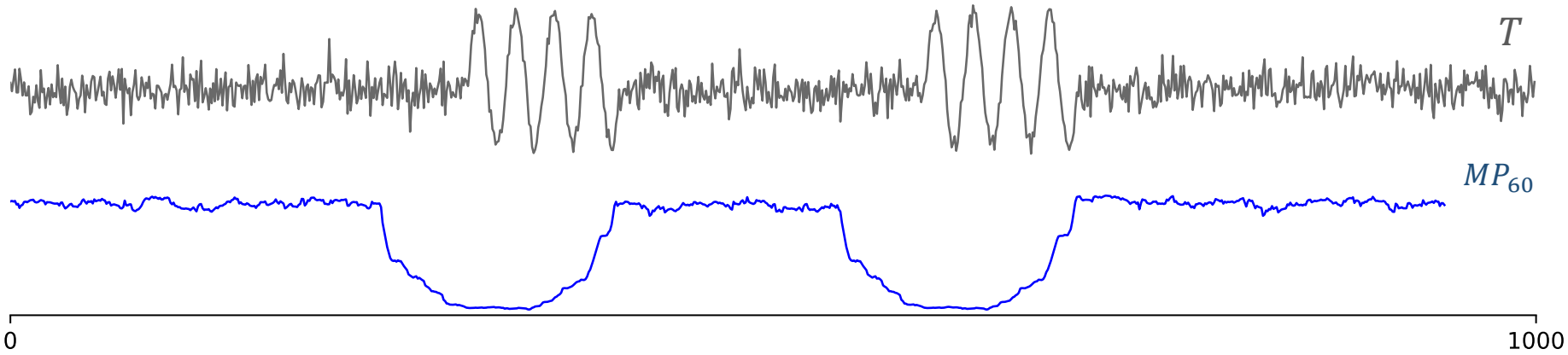
mSIMPAD: ***M****ultiple-length* ***S****uccessive s****IM****ilar* ***PA****tterns* ***D****etector*

$$dist(T_{i,l}, T_{j,l}) = \sqrt{\sum_{p=1}^{} \left(\frac{}{\sigma_{i,l}} \right.} \qquad \left. \frac{}{\sigma_{j,l}} \right)$$

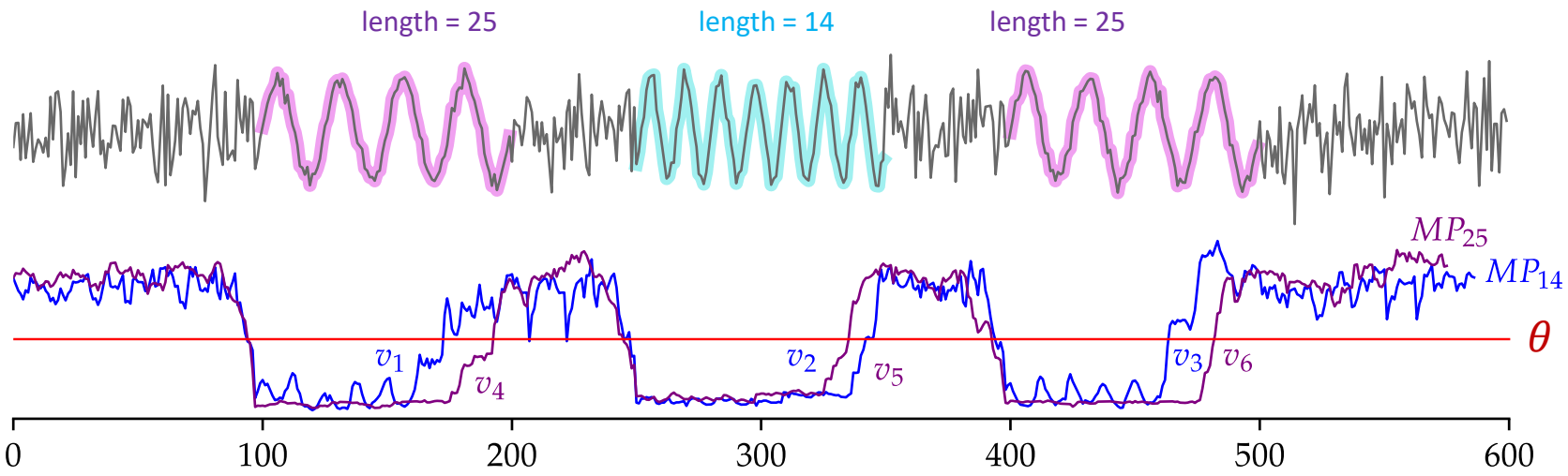
- Given T, $L = \left(l_1, \dots, l_{|L|}\right)$, and m, identify all subsequences in T that contains successive similar patterns (**SSP**)

Range-Constrained Matrix Profile (RCMP)

- **RCMP** is a variation of **Matrix Profile** [Yeh'16], a data structure that annotate a time series.
  - A value of RCMP MP_l at index i is the **distance** between $T_{i,l}$ to its **nearest neighbor** $T_{j,l}$ within the **user-defined range** m.



SSP with Different Lengths



What can we learn from this?

- MP containing SSP has relatively lower value.
- Regions that lower than $\boldsymbol{\theta}$ are called **valleys**.
- Valley with **larger area depth** is a better fit to the pattern.

Choosing The Best Set of Valleys

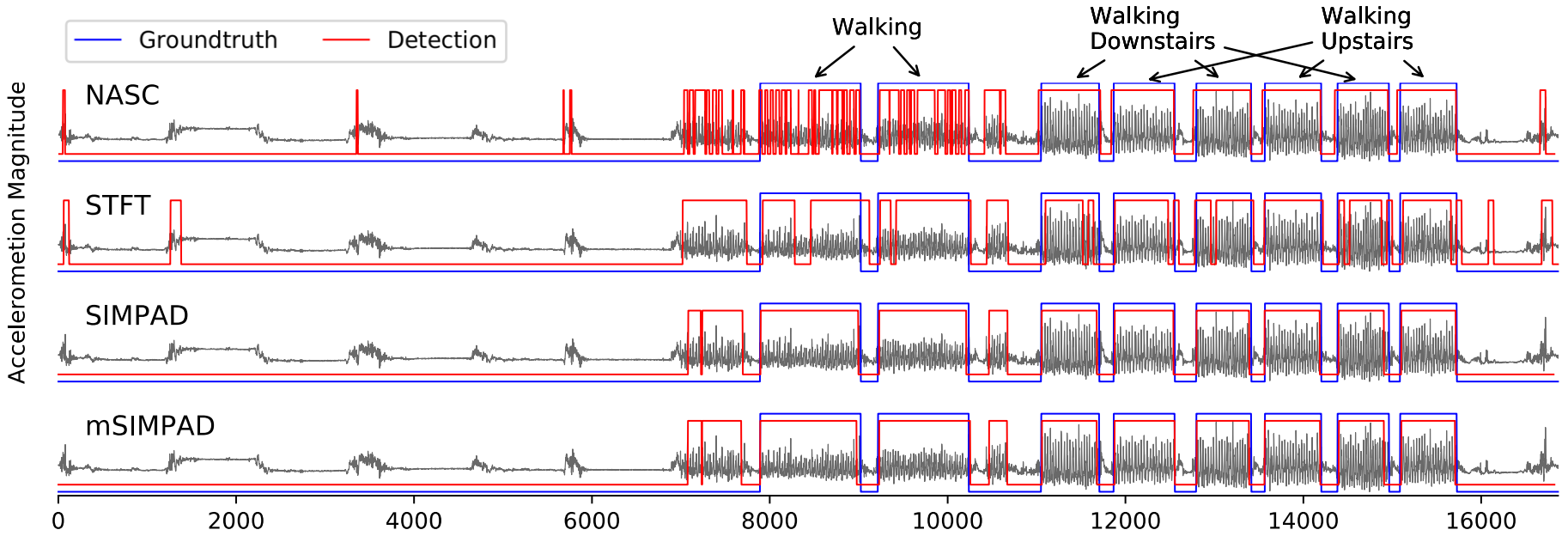
- Finding the best match of valleys = selecting valleys that maximize the total sum
- Let $V = \{v_1, v_2, \dots, v_{|V|}\}$ be the set of valleys found in MP at all l.
- $idx(v_i)$ denotes the corresponding indices of v_i.

$$\max_{V_{opt} \subseteq V} \sum_{v_i \in V_{opt}} v_i$$

$$\text{s.t. } \forall (v_i, v_j) \in V_{opt} : idx(v_i) \cap idx(v_j) = \emptyset$$

- Reduce to a maximum-weight independent set problem, and solved with a branch-and-bound approach

Example of Repetitive Activity Detection



Example of the detection result on a walking dataset [Brajdic'13].

- **Fewer** false positives
- More **accurate** and **coherent** detected regions

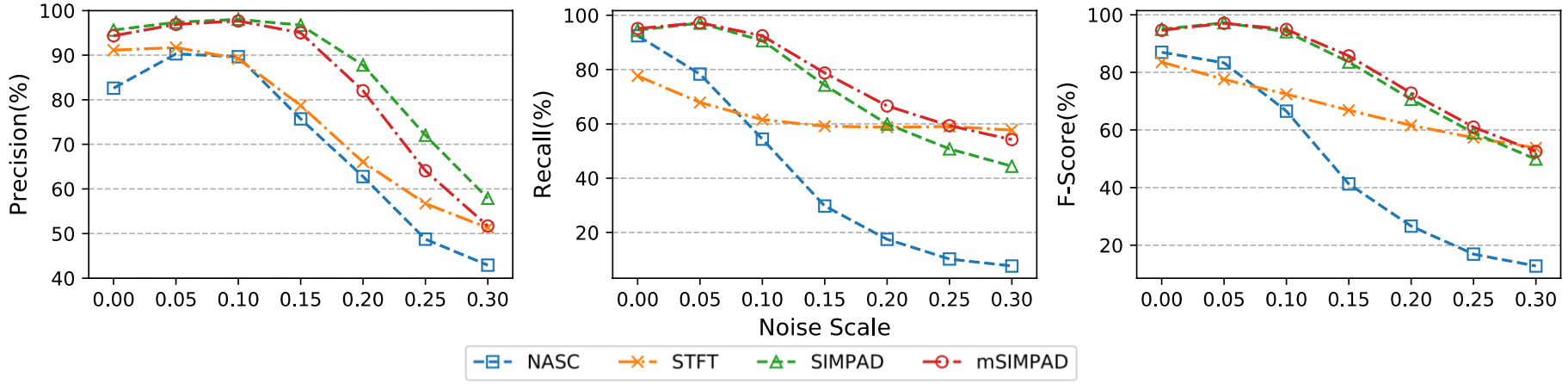
Evaluation on Detection Performance

Dataset	Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
HAPT	*NASC*	94.10 ± 3.49	91.26 ± 3.05	91.56 ± 10.88	91.07 ± 6.57
	STFT	94.03 ± 2.37	91.30 ± 3.35	91.73 ± 5.44	91.41 ± 3.34
	SIMPAD	**96.44 ± 2.26**	**95.63 ± 4.33**	94.50 ± 3.89	**94.96 ± 2.98**
	mSIMPAD	96.16 ± 2.60	94.35 ± 5.36	**95.10 ± 3.45**	94.62 ± 3.44
PAMAP2	*NASC*	81.70 ± 12.32	**99.39 ± 0.85**	66.47 ± 21.95	77.12 ± 22.26
	STFT	78.79 ± 9.12	99.31 ± 0.90	62.45 ± 6.75	76.50 ± 4.76
	SIMPAD	84.11 ± 5.59	99.12 ± 1.14	71.24 ± 5.66	82.78 ± 3.74
	mSIMPAD	**84.62 ± 5.65**	98.28 ± 1.78	**72.90 ± 5.58**	**83.59 ± 3.62**

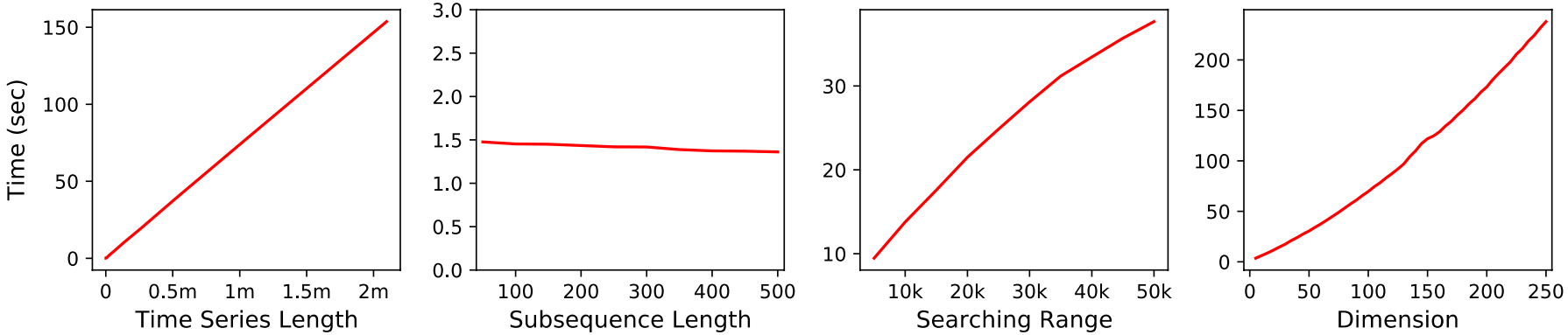
Performance on HAPT [Anguita'13] and PAMAP2 [Reiss'12] where the values given as mean ± SD.

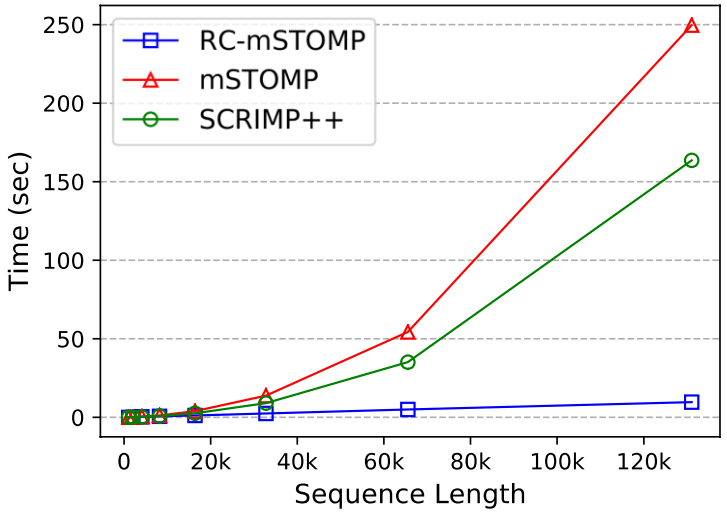
- **Outperforms** baseline methods
- Robust to **noisy** and **poor-quality** data

Downsampling input series to 20Hz



Evaluation on Running Time





- Compute large amount of data in minutes
- Scales linearly to the input size
  - support real-time applications