UNIVERSITY OF
EASTERN FINLAND

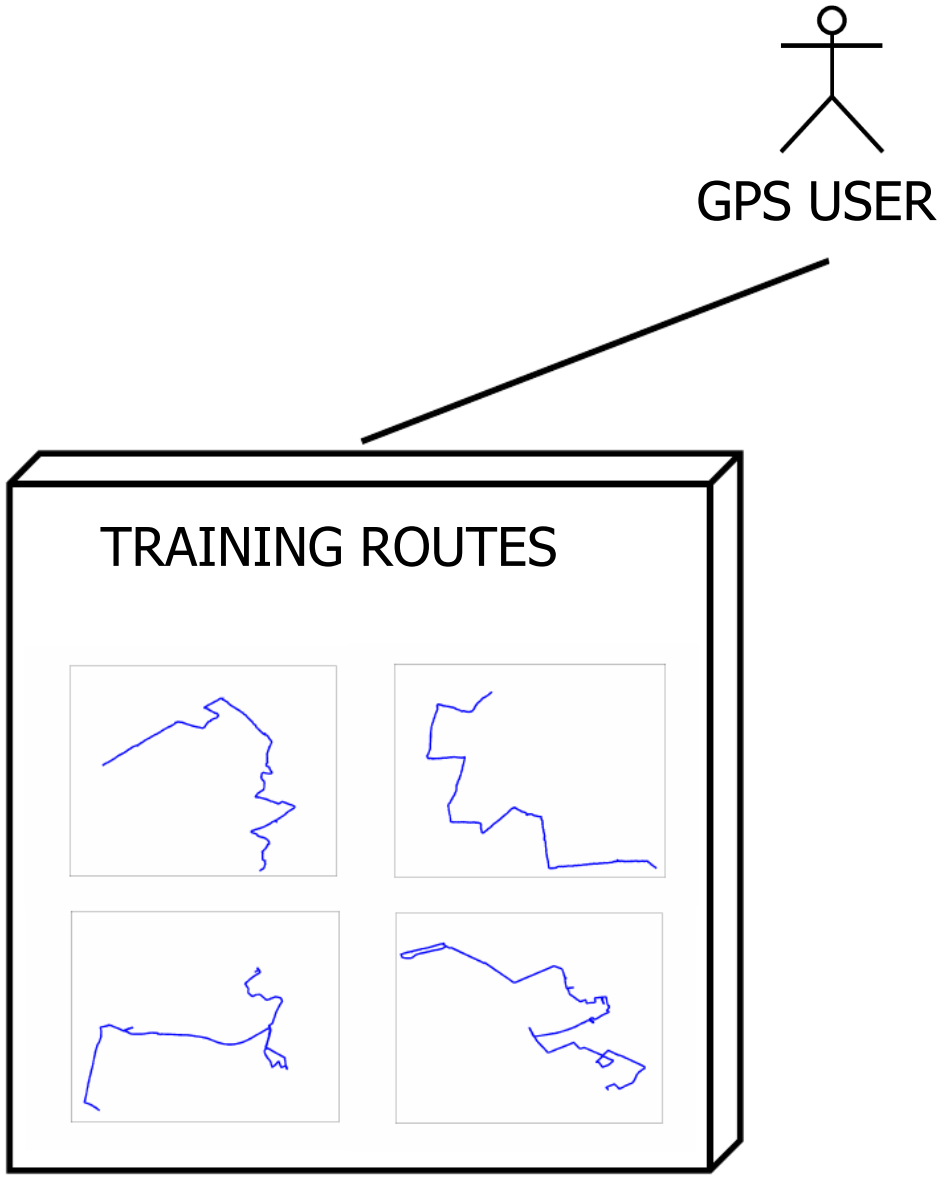
December 1, 2016

GPS Trajectory Biometrics: From Where You Were to How You Move

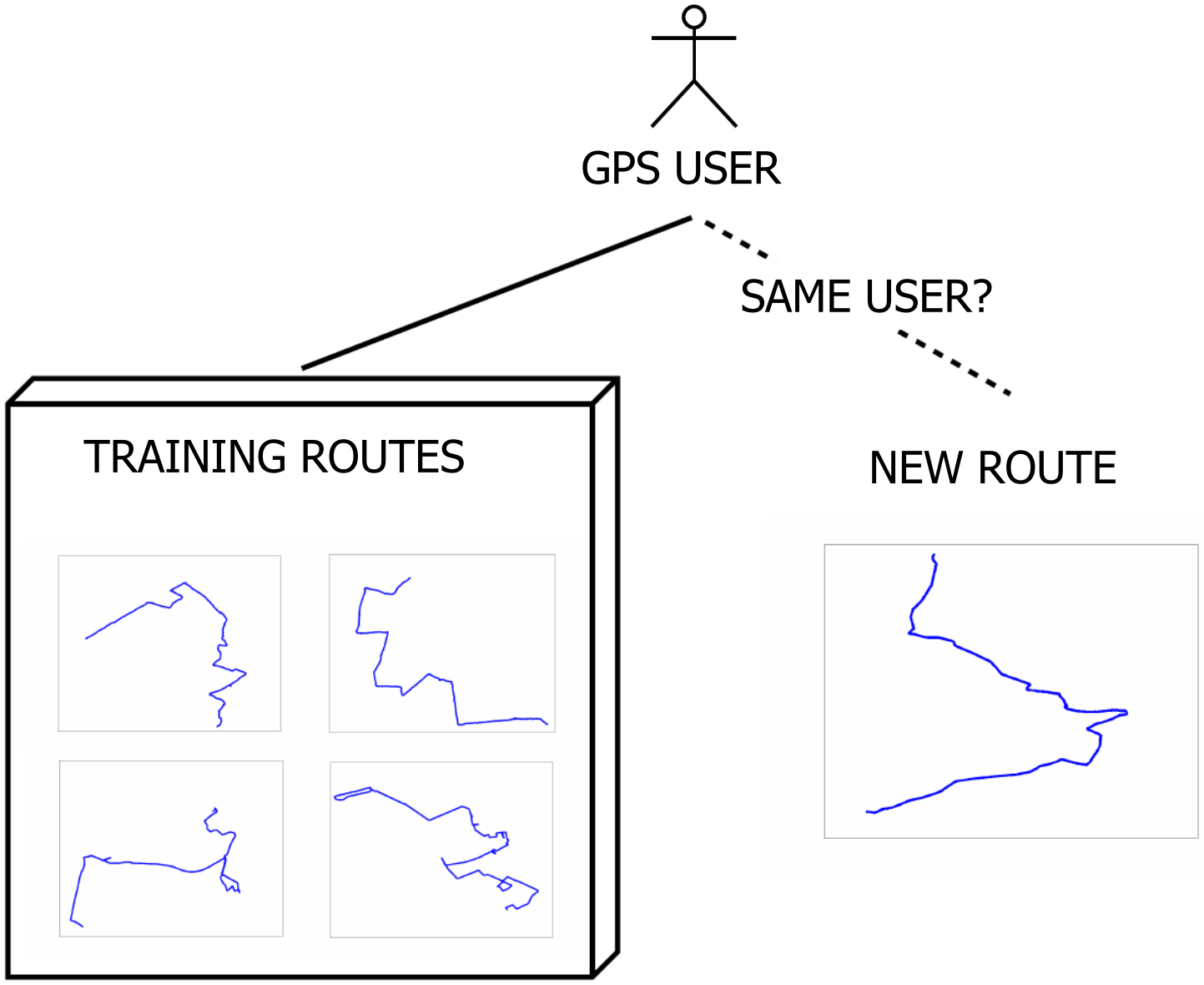
Sami Sieranoja, Tomi Kinnunen, Pasi Fränti

School of Computing
University of Eastern Finland

Problem: Recognize user of a route



Problem: Recognize user of a route



Previous work

Depends on the location: *where* people move

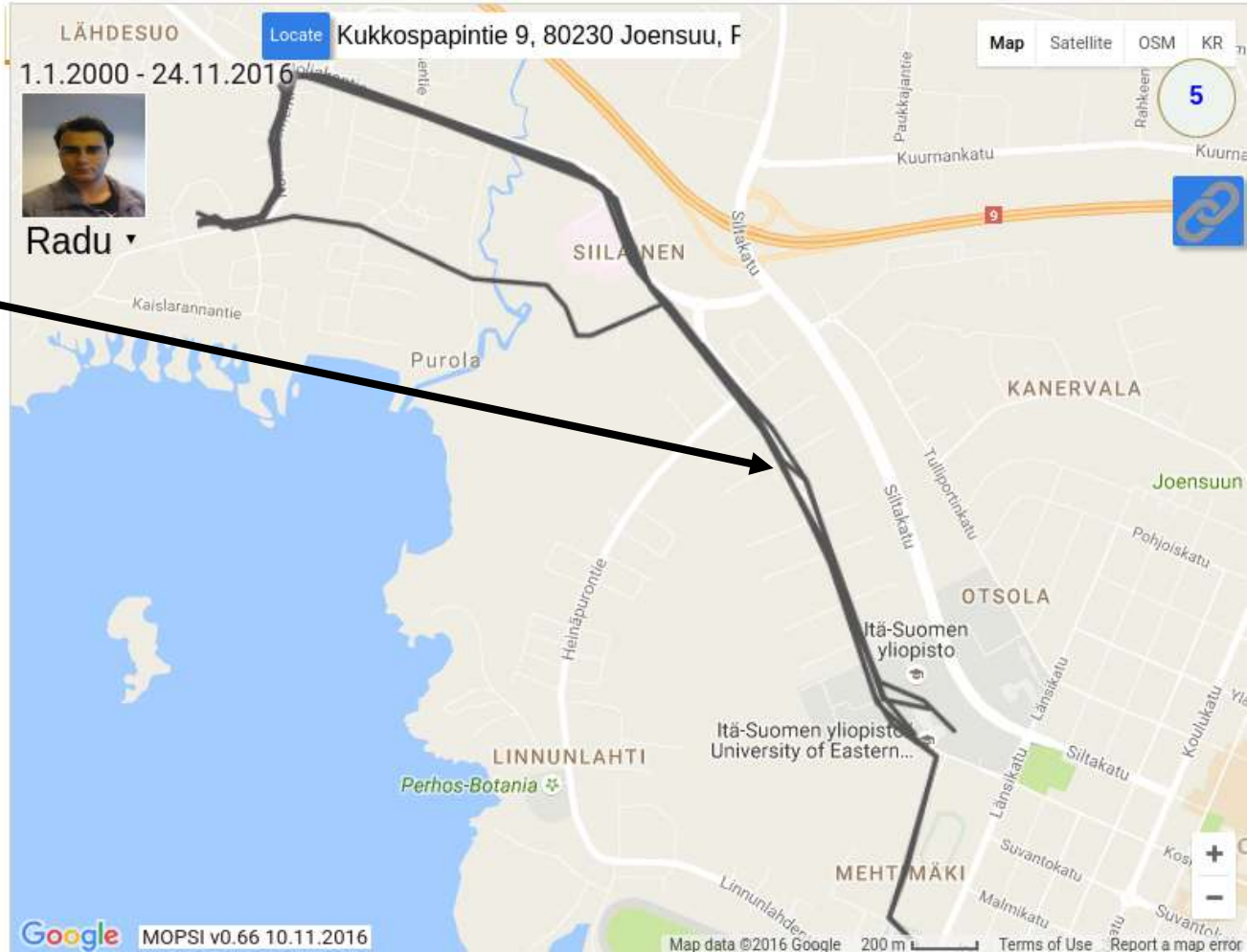
[1] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the Crowd: The privacy bounds of human mobility," Scientific reports, vol. 3, 2013.

[2] L. Rossi, J. Walker, and M. Musolesi, "Spatio-Temporal Techniques for User Identification by means of GPS Mobility Data," CoRR, vol. abs/1501.06814, 2015.

Recognition where people move

Easy task

Work to home routes of a user



24.1.2014
Route 1: 16:24 - 17:08 3 km 554 m

23.1.2014
Route 2: 09:49 - 10:17 2 km 997 m

22.11.2013
Route 3: 10:07 - 10:27 4 km 024 m

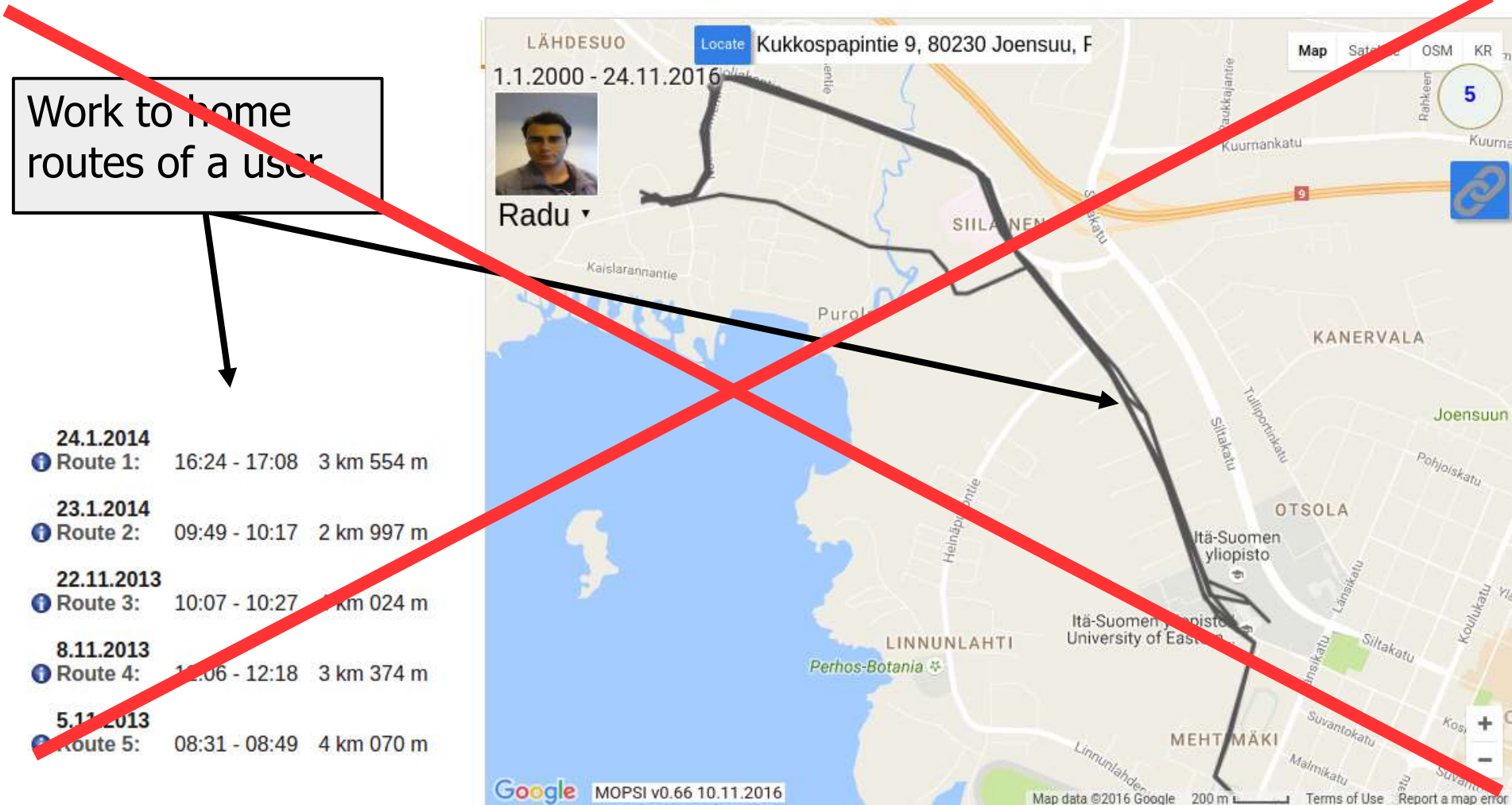
8.11.2013
Route 4: 12:06 - 12:18 3 km 374 m

5.11.2013
Route 5: 08:31 - 08:49 4 km 070 m

When moving in a new area

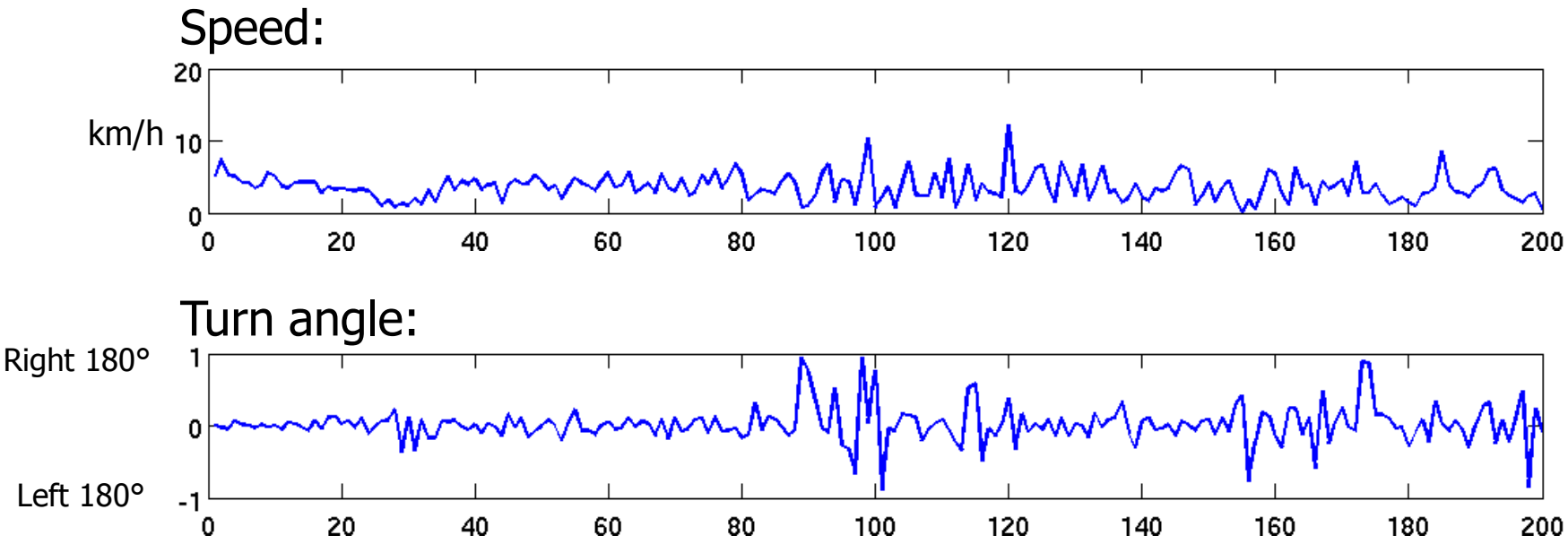
Does not work anymore

Work to home routes of a user



Proposed: *how* people move

Use only local variations in speed and turn angles of a route:



Movement types

Consider only routes with speeds in the walking, running, cycling range.

Implications

Forensics:

- Track movement of “burned phone”. Unknown user.
- Subject gives the phone to another person.
- Detection of the change of tracked subject.

Privacy:

- One might want to reconsider sharing one's routes in online services like sports tracker.
- Handle GPS routes with similar case as other private information like voice or fingerprints.

Methods: Feature extraction

Two basic features:

- Speed (km/h)
- Turn angle.

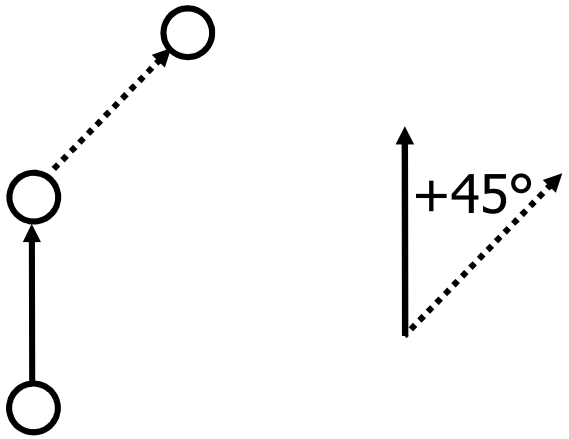
Model how frequently and by how much they change within a route segment.

Process using a sliding window. For window contents: *Discrete fourier transform* (DFT).

Feature extraction: Turn angles

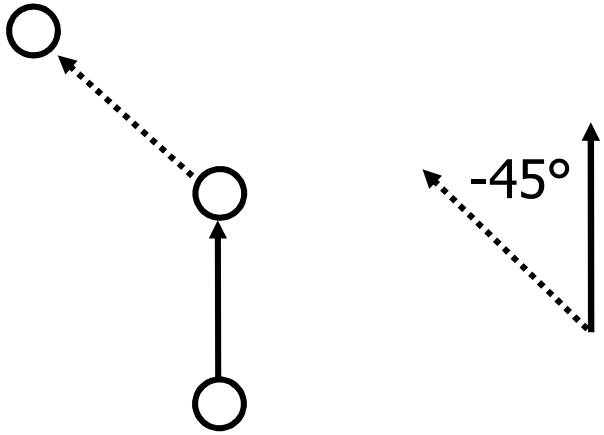
Normalize turns to [-1,+1] range.

Right Turn:



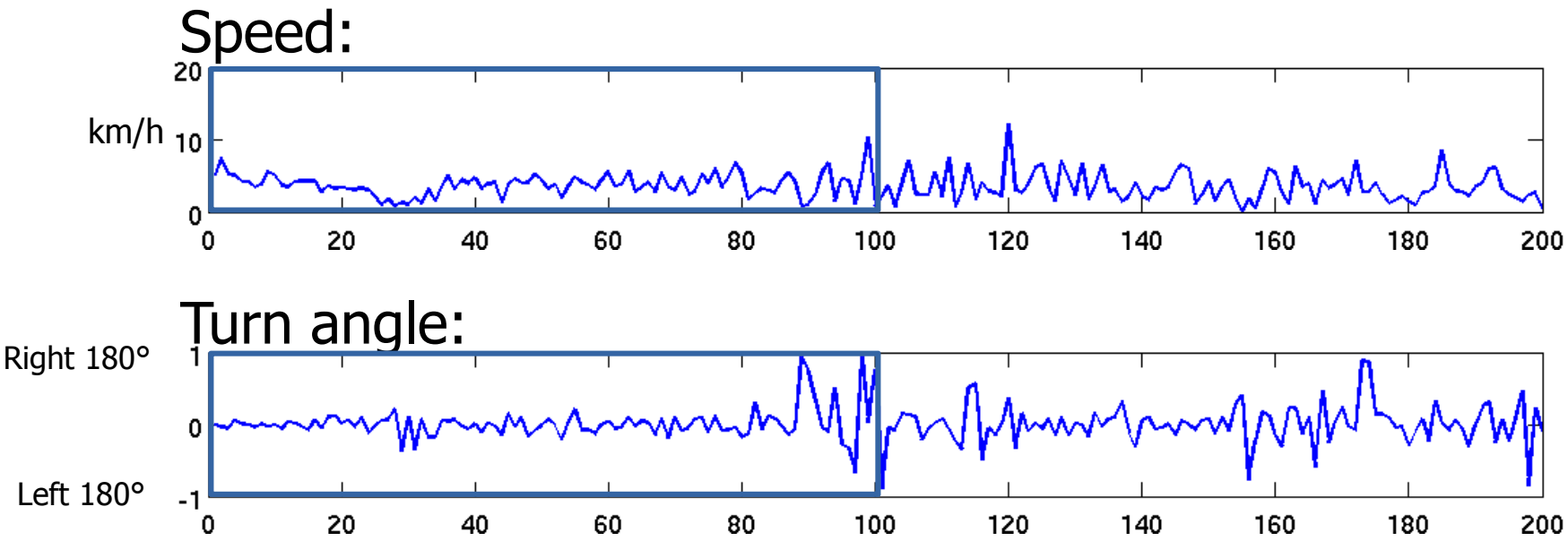
+45/180=+0.25

Left Turn:



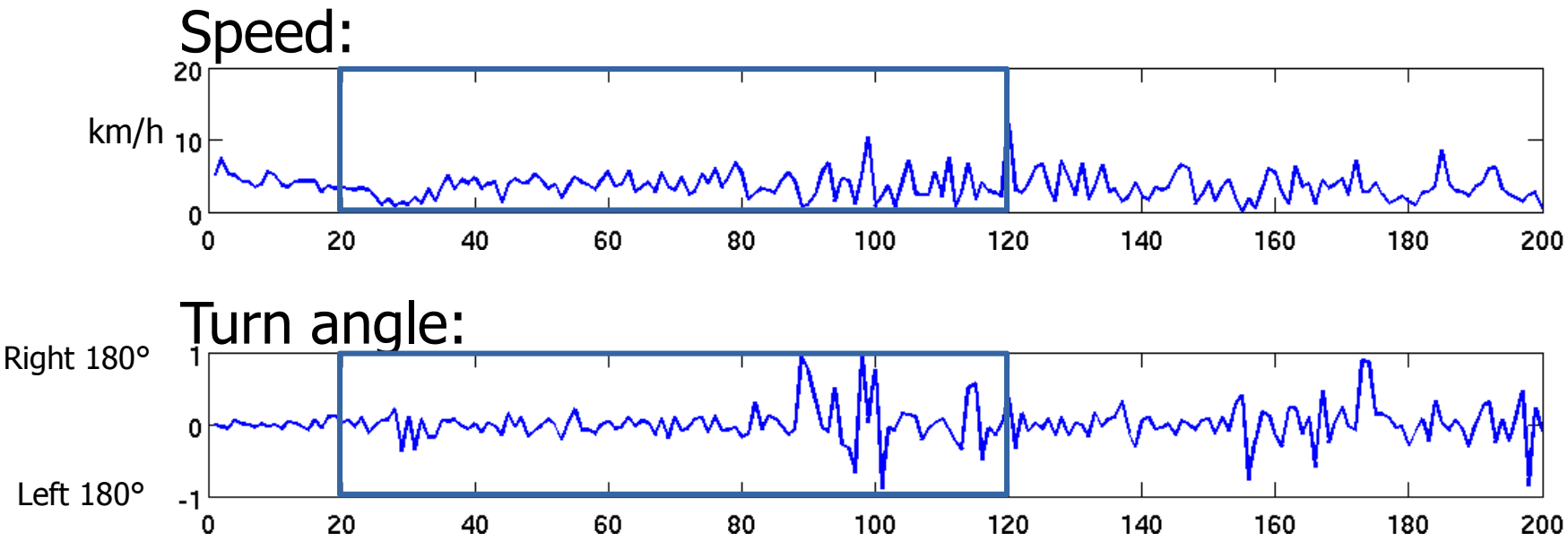
-45/180=-0.25

Feature extraction: Sliding window



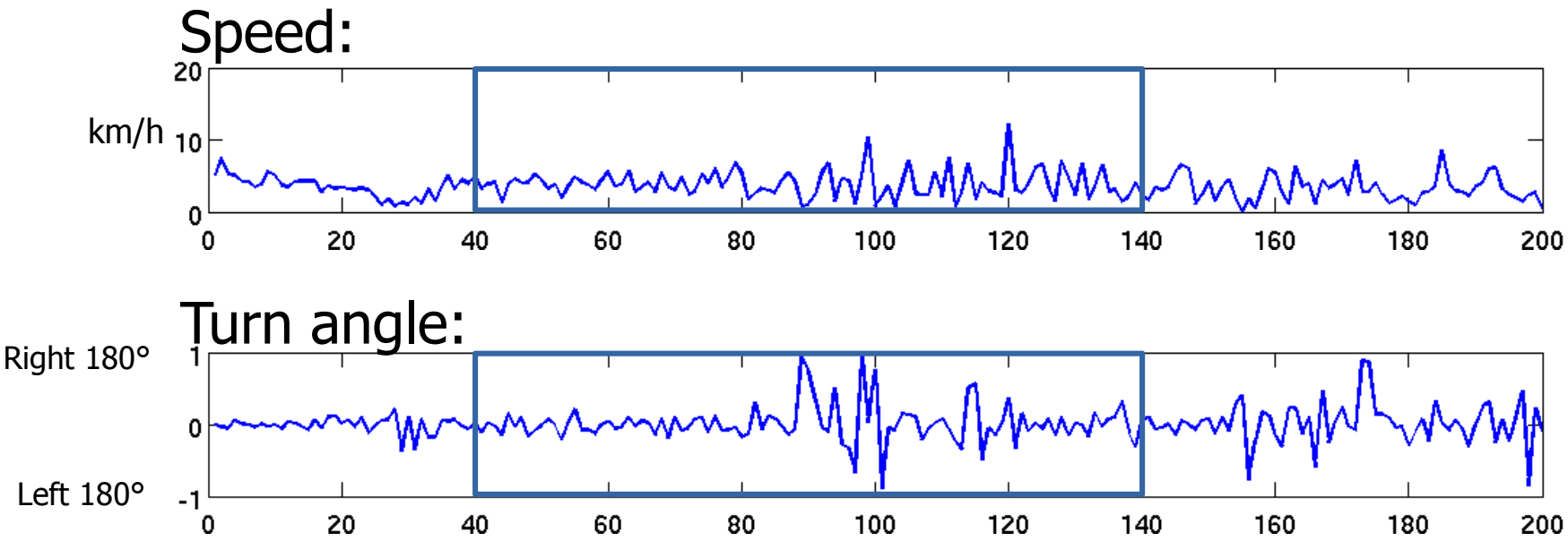
Step size =1 (99% overlap)
Window size = 100 seconds

Feature extraction: Sliding window



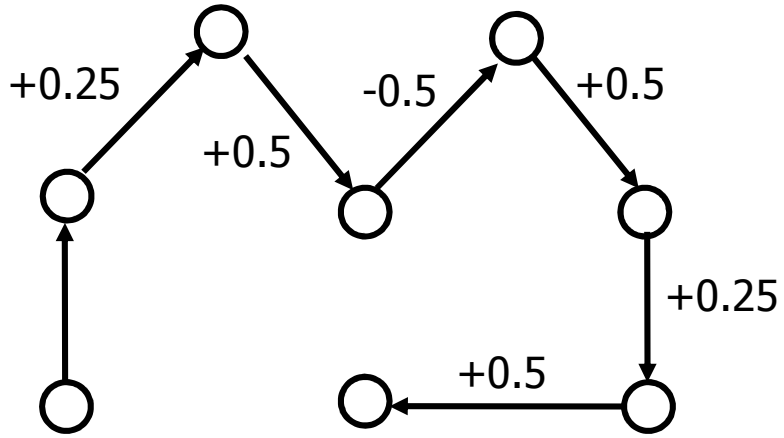
Step size =1 (99% overlap)
Window size = 100 seconds

Feature extraction: Sliding window



Step size =1 (99% overlap)
Window size = 100 seconds

Turn angle window processing



Process sequence with
Integration (cumulative sum):

[+0.25 +0.5 -0.5 +0.5 +0.25 +0.5]

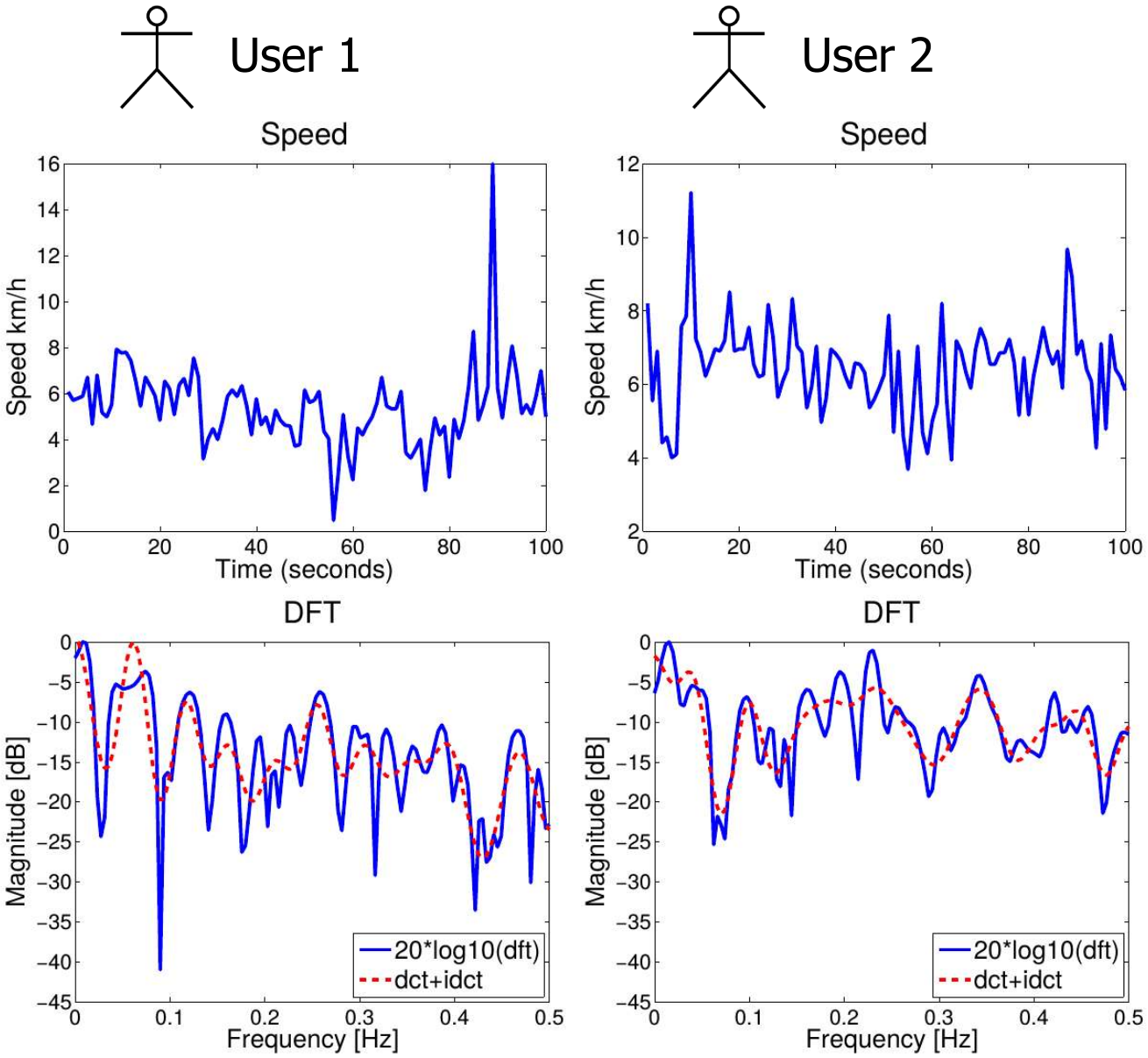
[0.25 0.75 0.25 0.75 1.0 1.5]

Similar to using direction (bearing), except:

- Not restricted to certain range
- Small +2° turn can cause large change: 359° => 1°

Feature extraction: Fourier transform

Parametrize log magnitude spectrum using *Discrete cosine transform* (DCT).



Feature extraction: Concatenation

- DCT parametrization produces 24 dimensional vector for each feature type (speed, turn angle).
- Features are concatenated to 2*24=48 dimensional vector for each window.
- Processing with sliding window produces a sequence of vectors for each route.

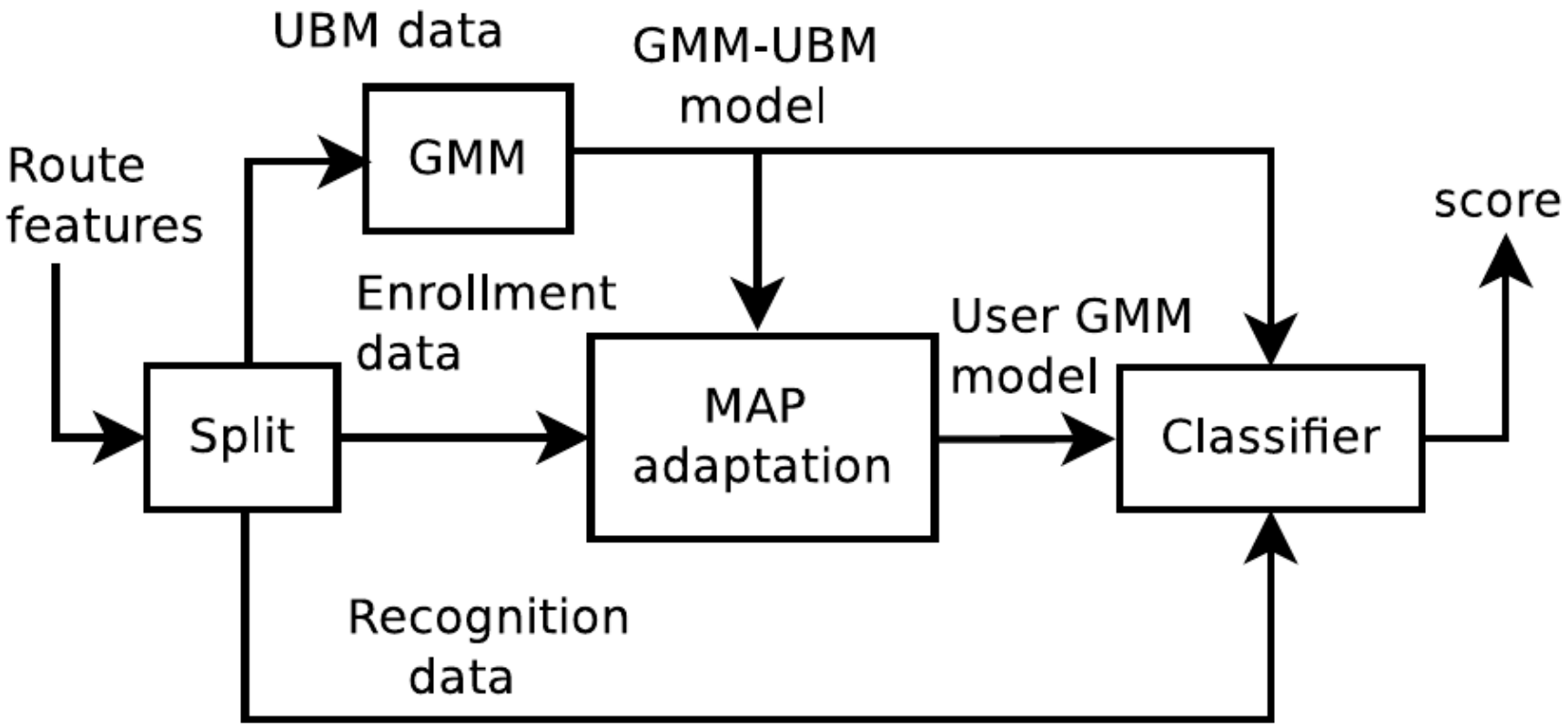
Methods: Modeling

- *Gaussian mixture model -universal background model classifier* (GMM-UBM) [2,3]. Widely used in speech processing for somewhat similar DFT based features.

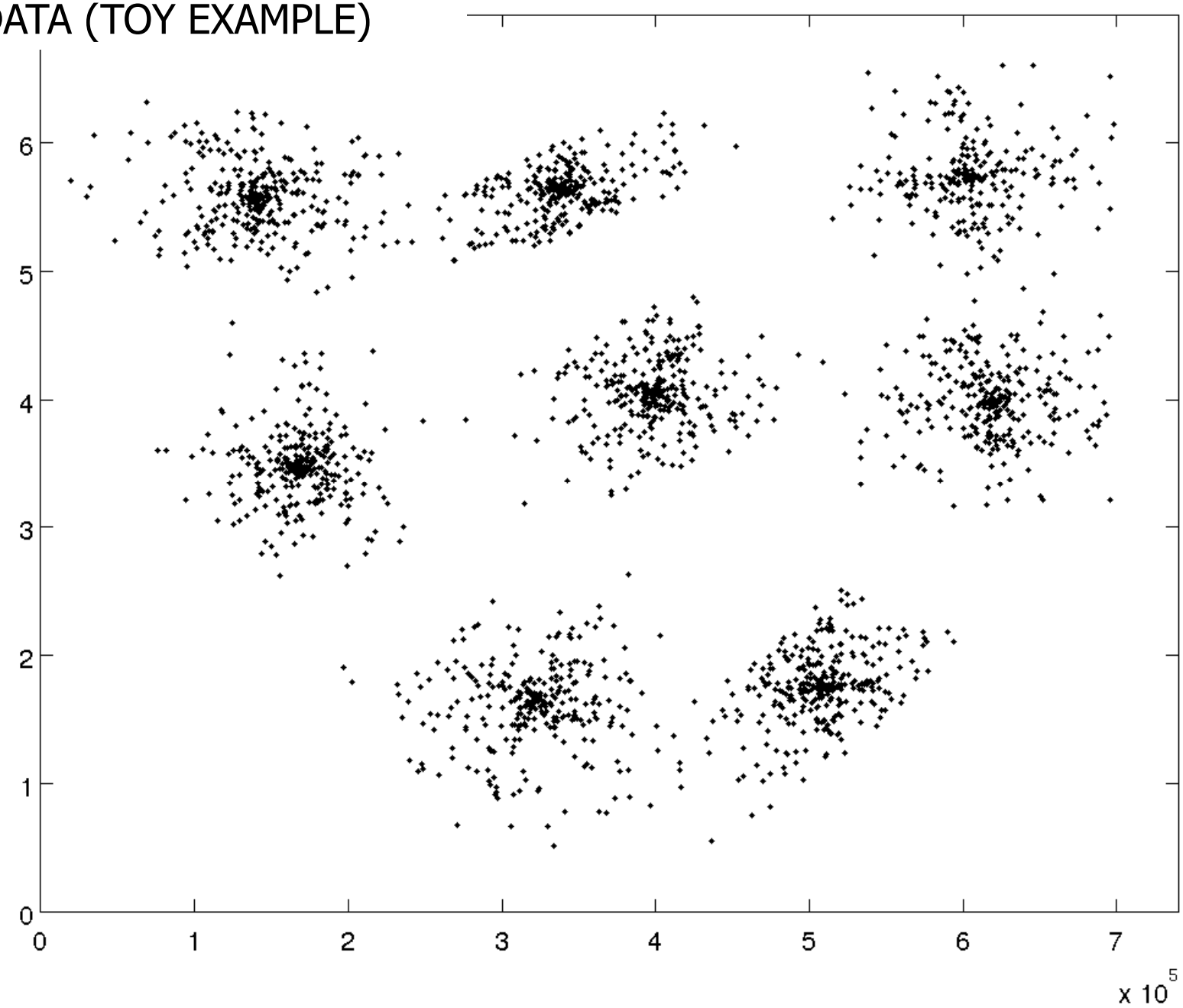
[2] Reynolds, D., Quatieri, T., Dunn, R.: Speaker verification using adapted gaussian mixture models. Digital Signal Processing 10(1), 19–41 (January 2000)

[3] Gauvain, J.L., Lee, C.H.: Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains. IEEE Transactions on Speech and audio processing 2(2), 291–298 (1994)

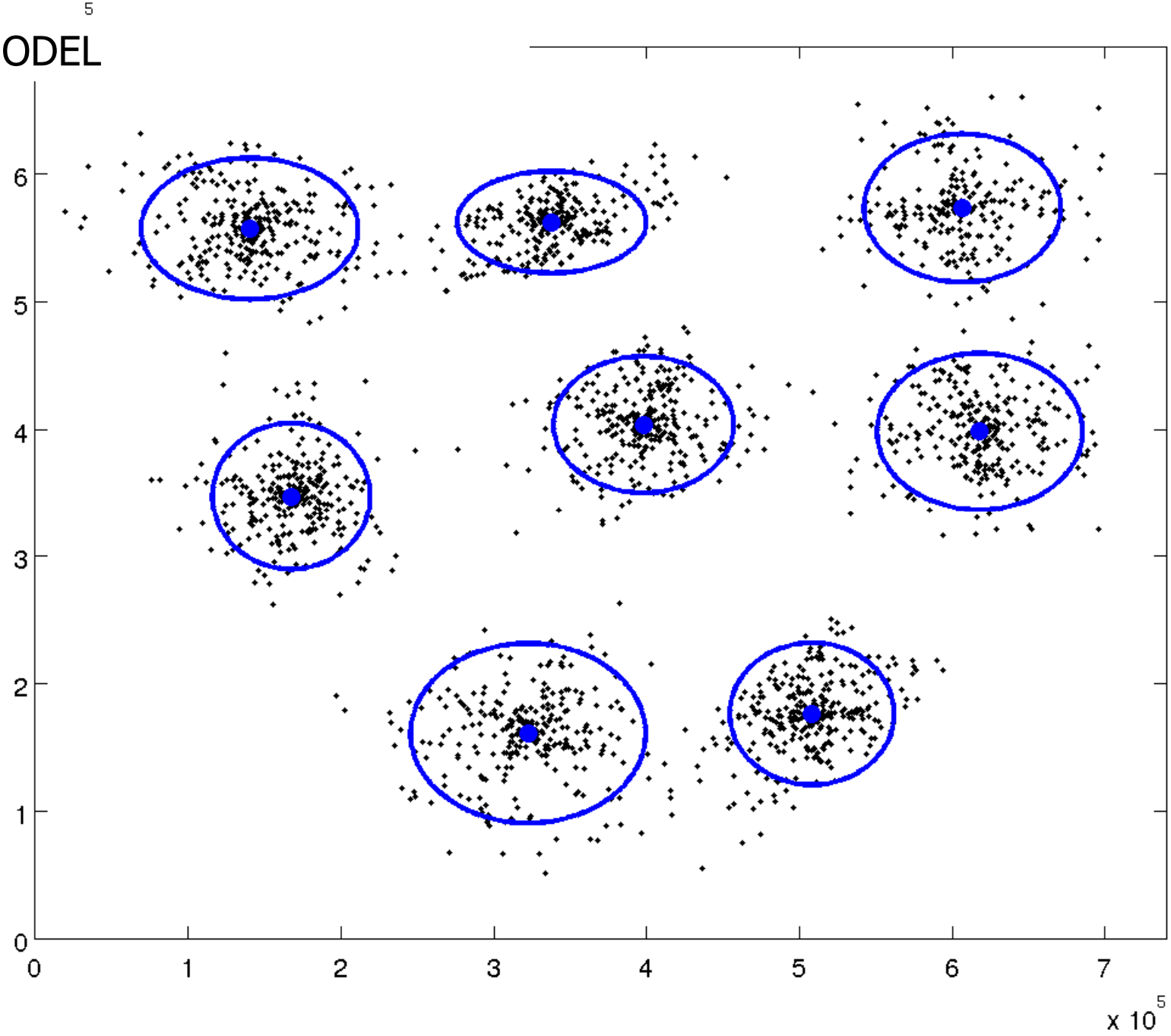
Methods: Modeling



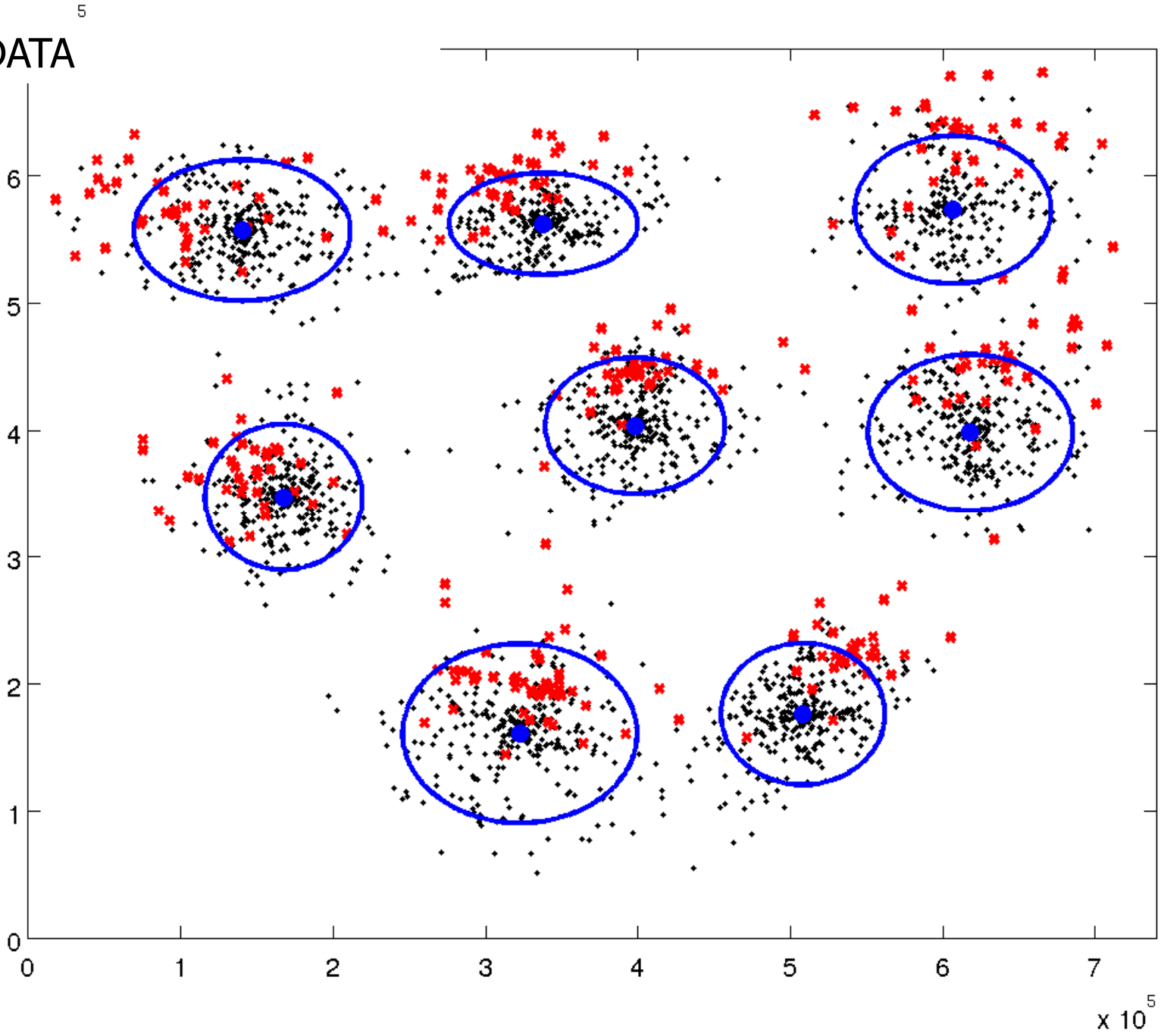
UBM DATA (TOY EXAMPLE)



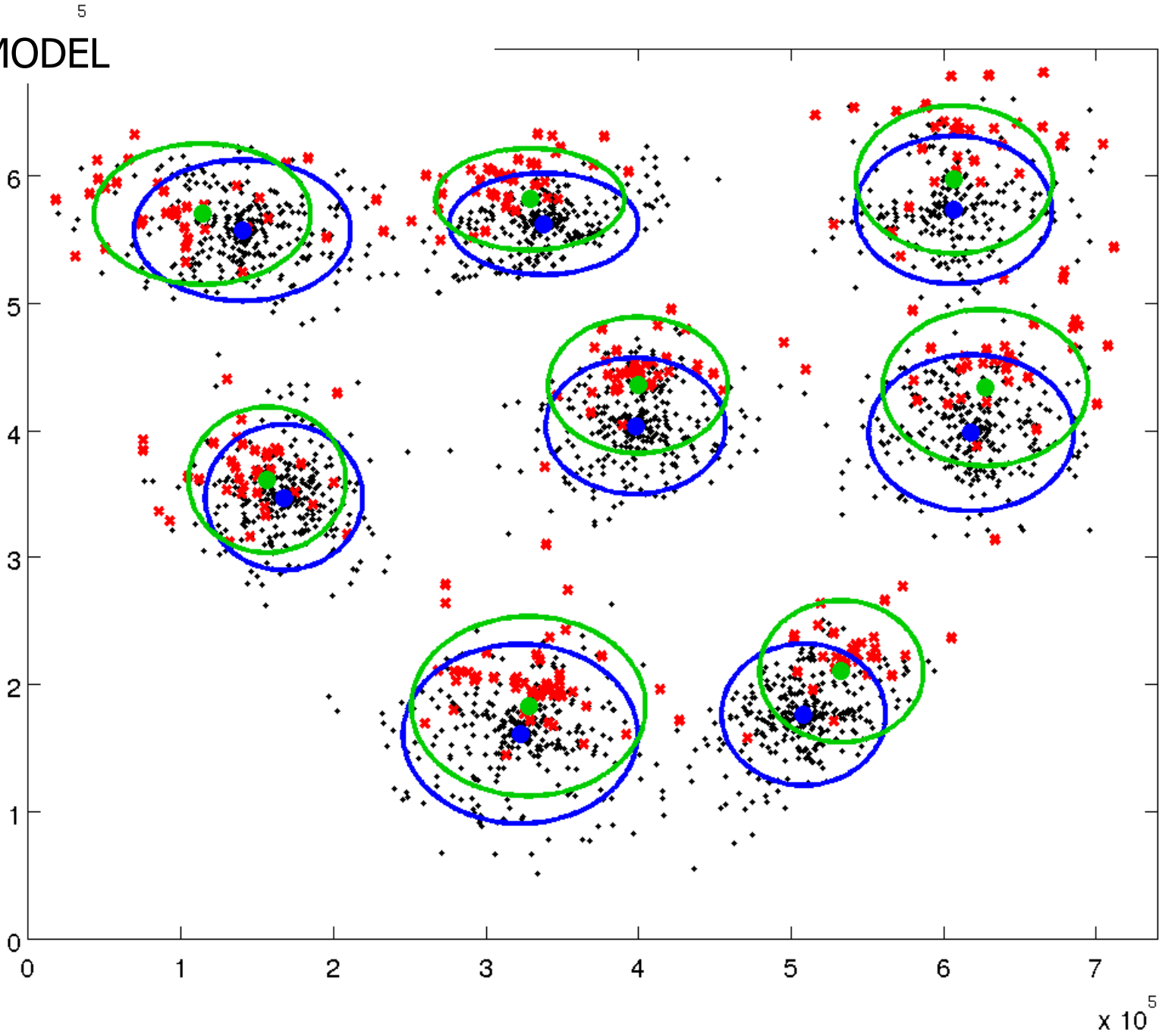
UBM MODEL



USER DATA



USER MODEL



Datasets

Two source:

- Openstreetmap [1]: Collection of GPS traces
- Geolife [2]: GPS dataset collected by Microsoft Research Asia (2007-2012).

[1] http://wiki.openstreetmap.org/wiki/Planet.gpx
[2] https://www.microsoft.com/en-us/download/details.aspx?id=52367

Datasets: filtering

- Two variants: 30 and 60 min trial route length.
- Uniform sampling interval.
- Remove overlapping routes per user.

	OSM30m	OSM60m	Geolife30m	Geolife60m
Target users	156	51	34	20
Sampling interval	1s	1s	2s	2s
Route length	30min	60min	30min	60min
Training data / user	2h	4h	2h	4h
Test data / user	2h	4h	2h	4h

Experimental procedure

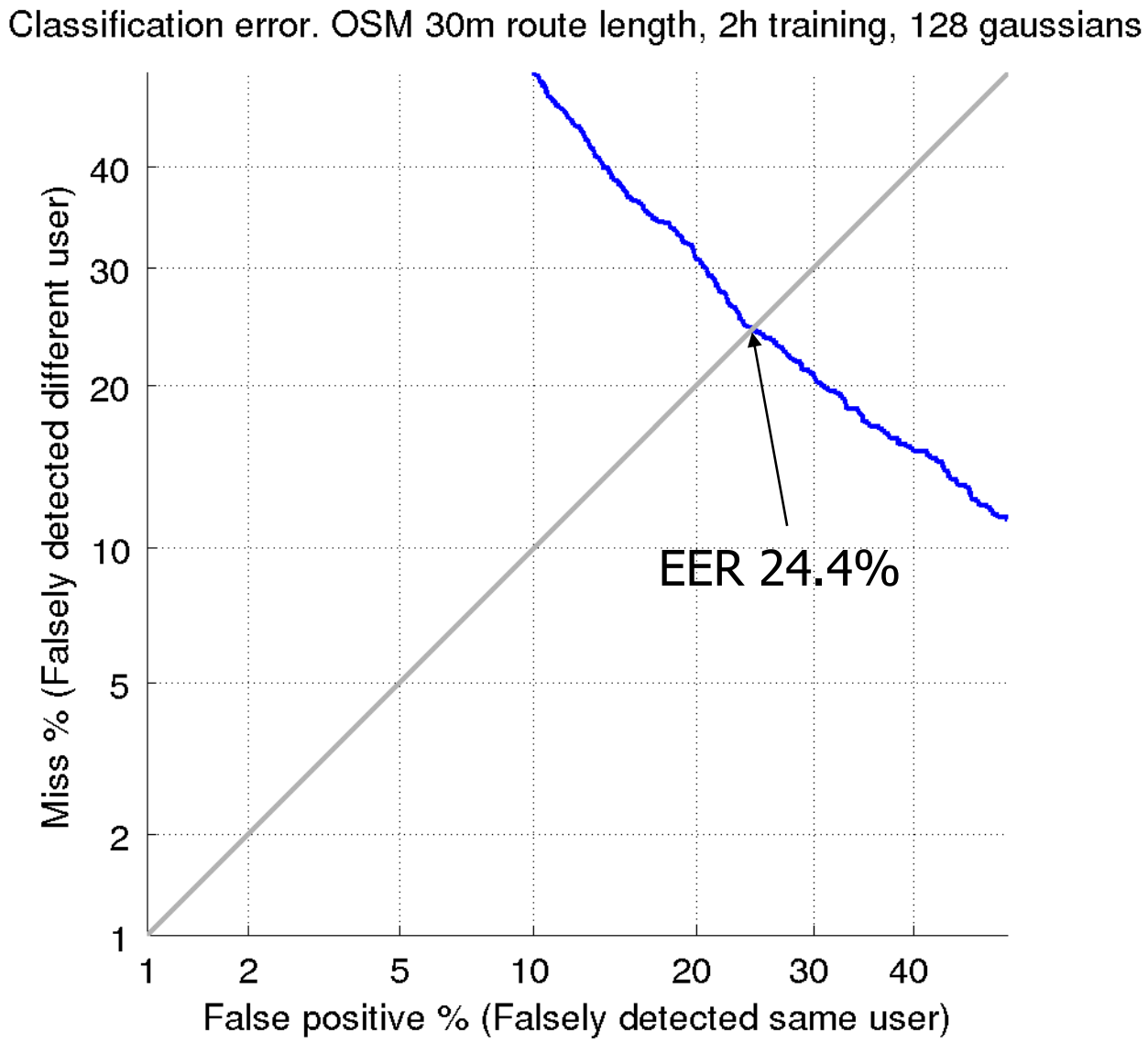
Calculate classifier scores for pairs of route and user model.

Two cases:

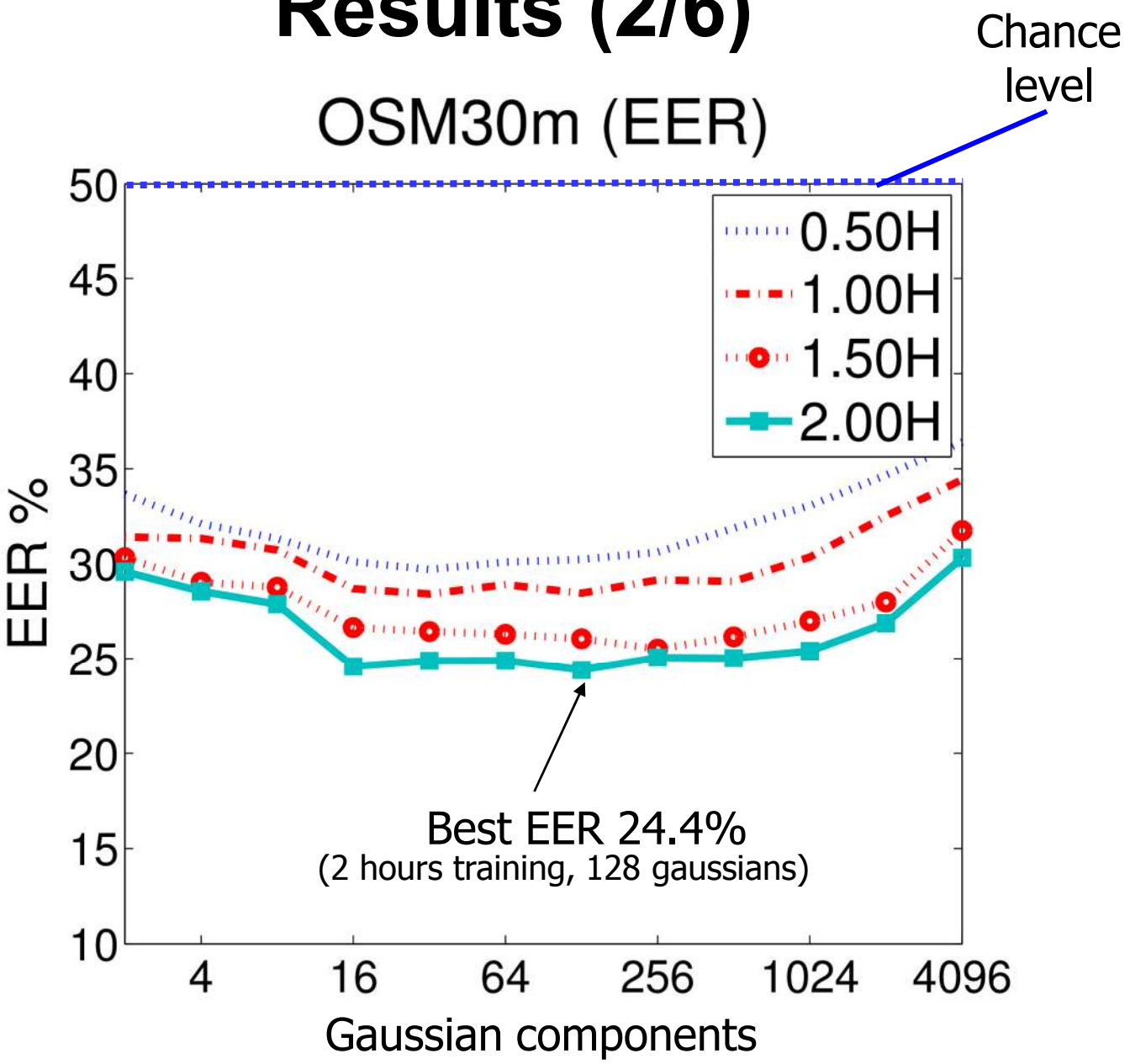
- Model and route belong to the same user
- Model and route belong to different user

Performance measured as *equal error rate* (EER): Rate at which both acceptance and rejection errors are equal.

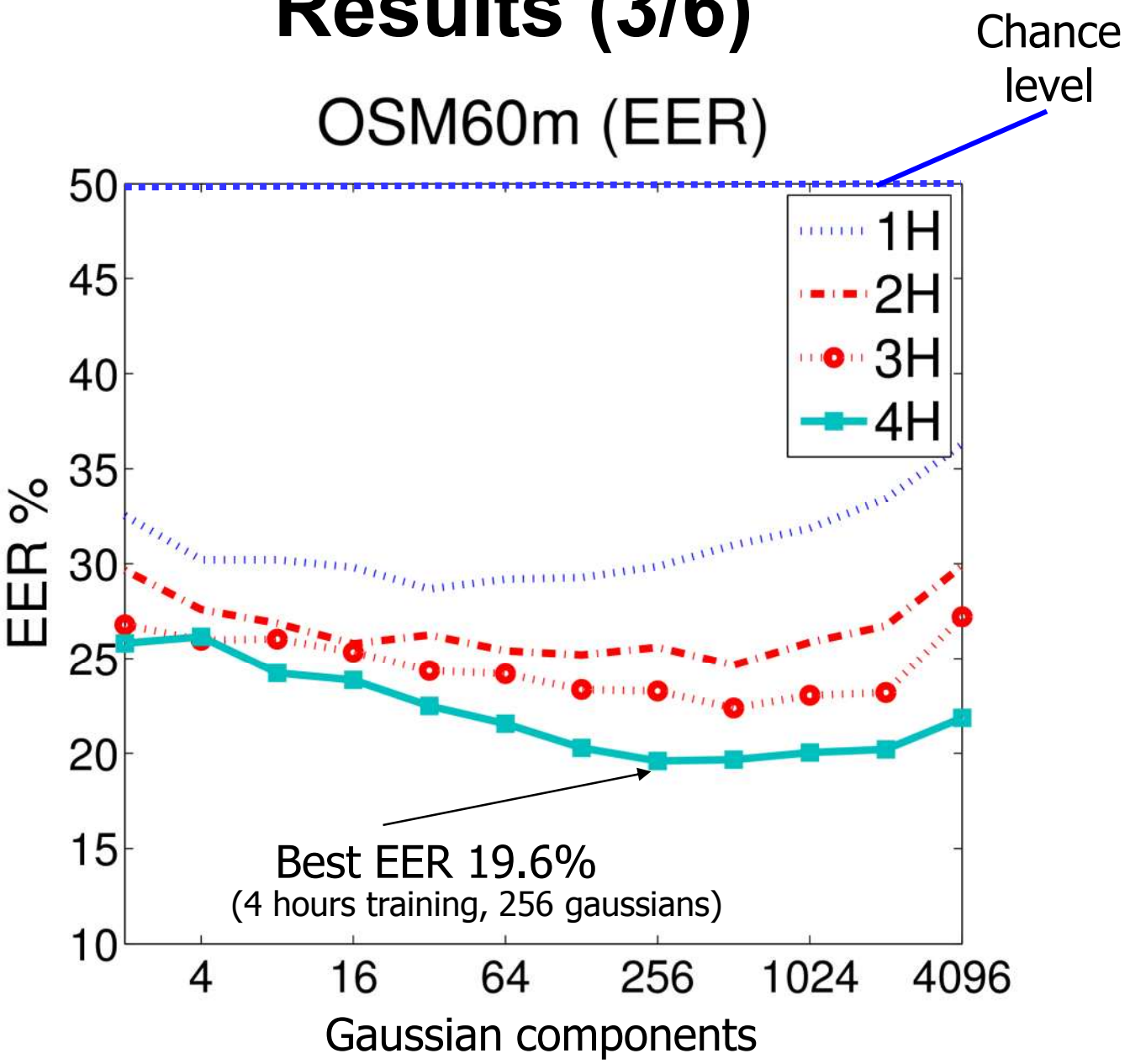
Results (1/6)


Classification error. OSM 30m route length, 2h training, 128 gaussians
Miss % (Falsely detected different user)
False positive % (Falsely detected same user)
EER 24.4%
1
2
5
10
20
30
40

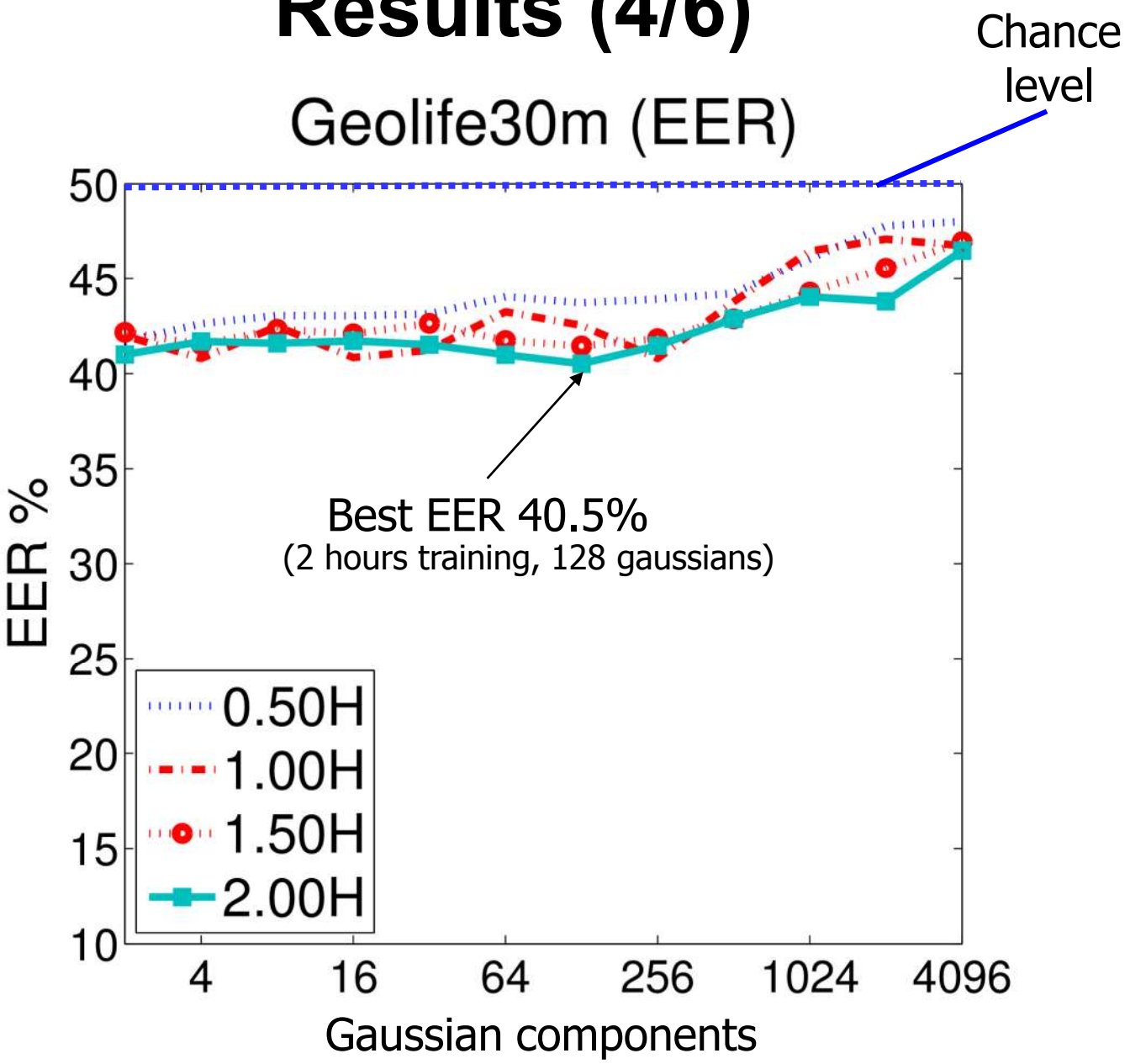
Results (2/6)


Chance level
OSM30m (EER)
0.50H
1.00H
1.50H
2.00H
EER %
50
45
40
35
30
25
20
15
10
4
16
64
256
1024
4096
Gaussian components
Best EER 24.4%
(2 hours training, 128 gaussians)

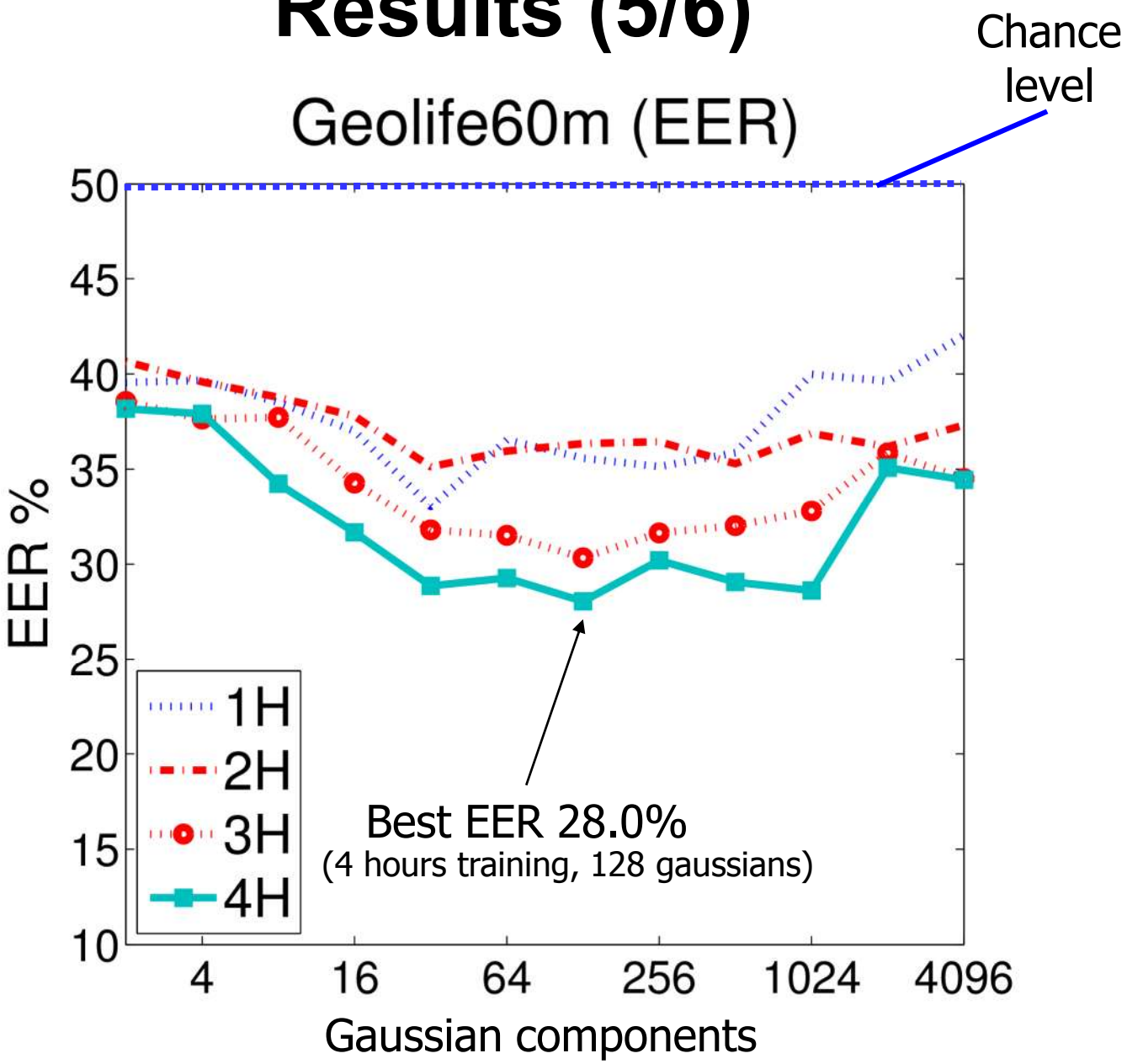
Results (3/6)


Chance level
OSM60m (EER)
EER %
50
45
40
35
30
25
20
15
10
1H
2H
3H
4H
Best EER 19.6%
(4 hours training, 256 gaussians)
4
16
64
256
1024
4096
Gaussian components

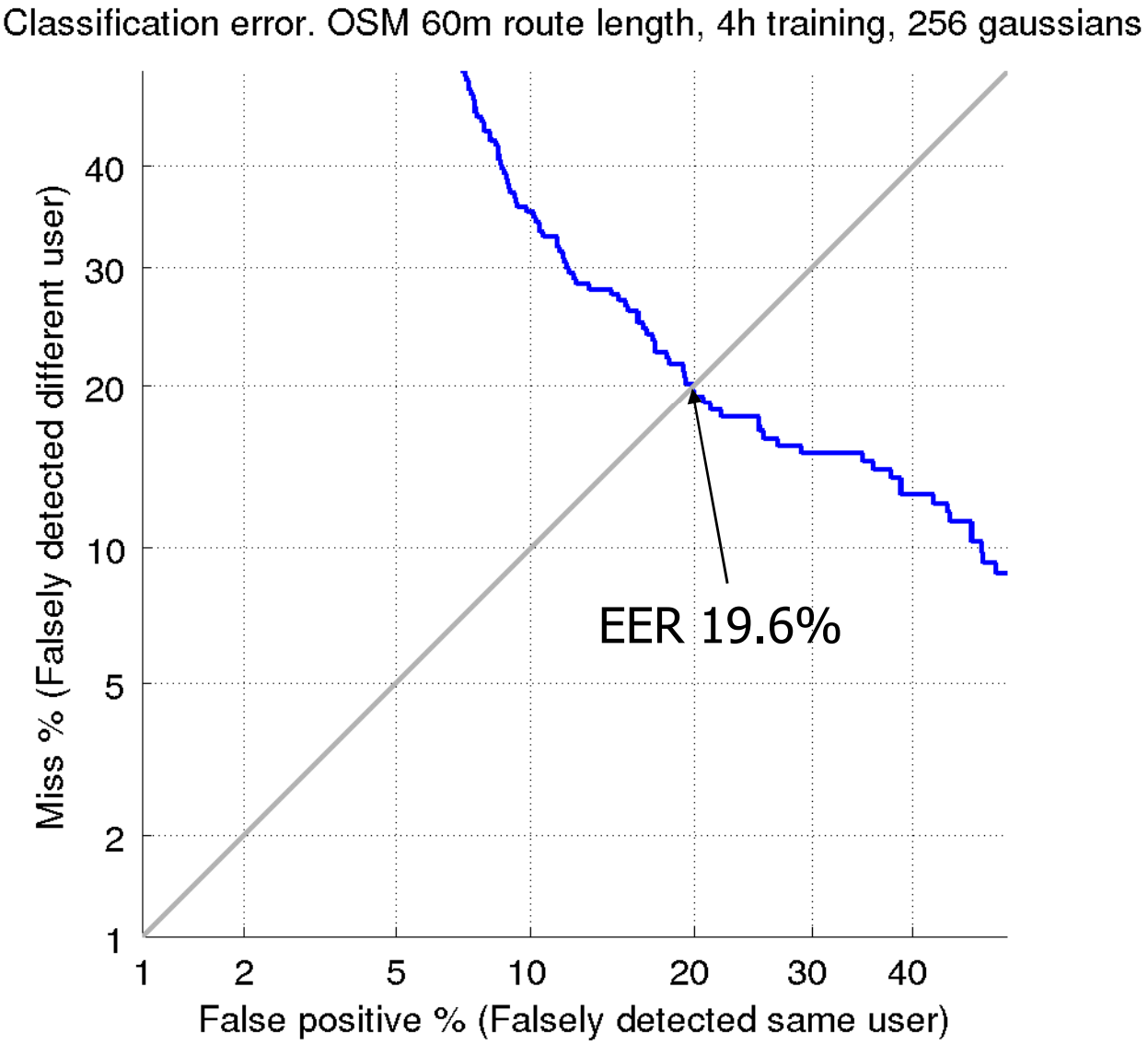
Results (4/6)


Chance level
Geolife30m (EER)
Best EER 40.5%
(2 hours training, 128 gaussians)
EER %
50
45
40
35
30
25
20
15
10
0.50H
1.00H
1.50H
2.00H
4
16
64
256
1024
4096
Gaussian components

Results (5/6)


Chance level
Geolife60m (EER)
EER %
50
45
40
35
30
25
20
15
10
1H
2H
3H
4H
Best EER 28.0%
(4 hours training, 128 gaussians)
4
16
64
256
1024
4096
Gaussian components

Results (6/6): Best overall EER



Conclusions

- Achieves accuracy of 19.6% *equal error rate* (EER) in the best case.
- Local variations in GPS data do possess some user specific characteristics.
- Method might be useful in some applications such as screening subjects in forensics.

Datasets will be available in: http://cs.uef.fi/~samisi/gps/