Ein Innovationszentrum der Daimler AG
und der Universität Ulm

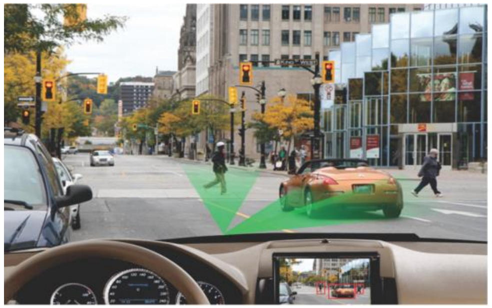
Driver Head Pose Estimation by Regression

Y. Tessema, M. Hoeffken and Dr. U. Kressel

Motivation – Head Pose Estimation

Driver

Environment



Driver Observation

Advanced Driver Assistance Systems

Driver Action

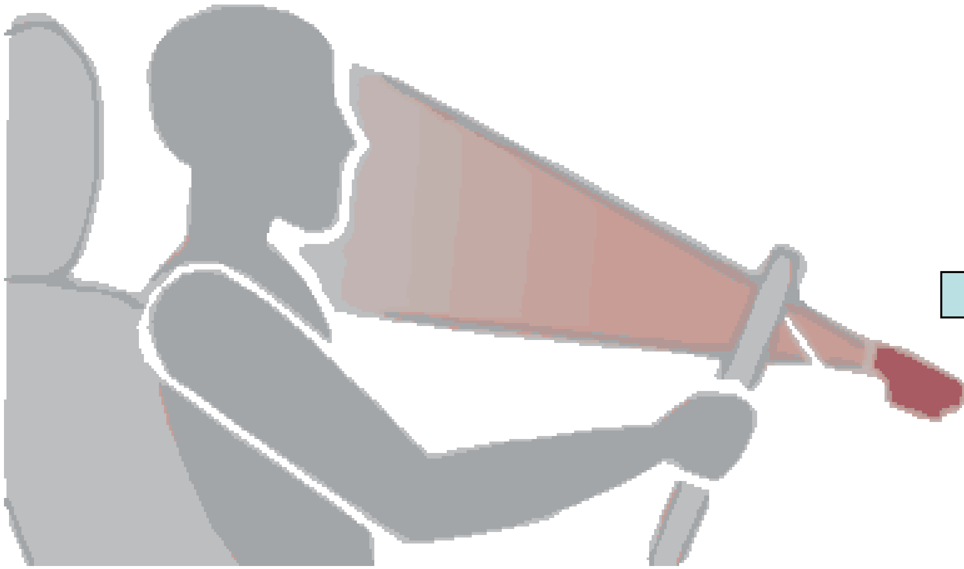
Vehicle Dynamics

Vehicle

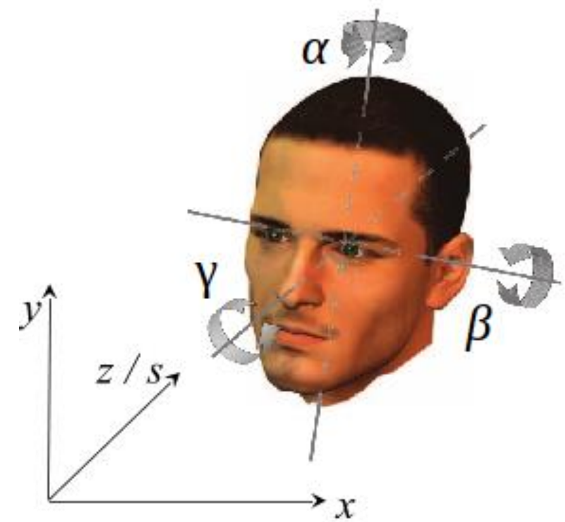
Estimating the driver‘s head pose

- To analyze the driver‘s attention
- Field of view of the driver
- Autonomous driving

Motivation – Head Pose Estimation



Stereo Camera

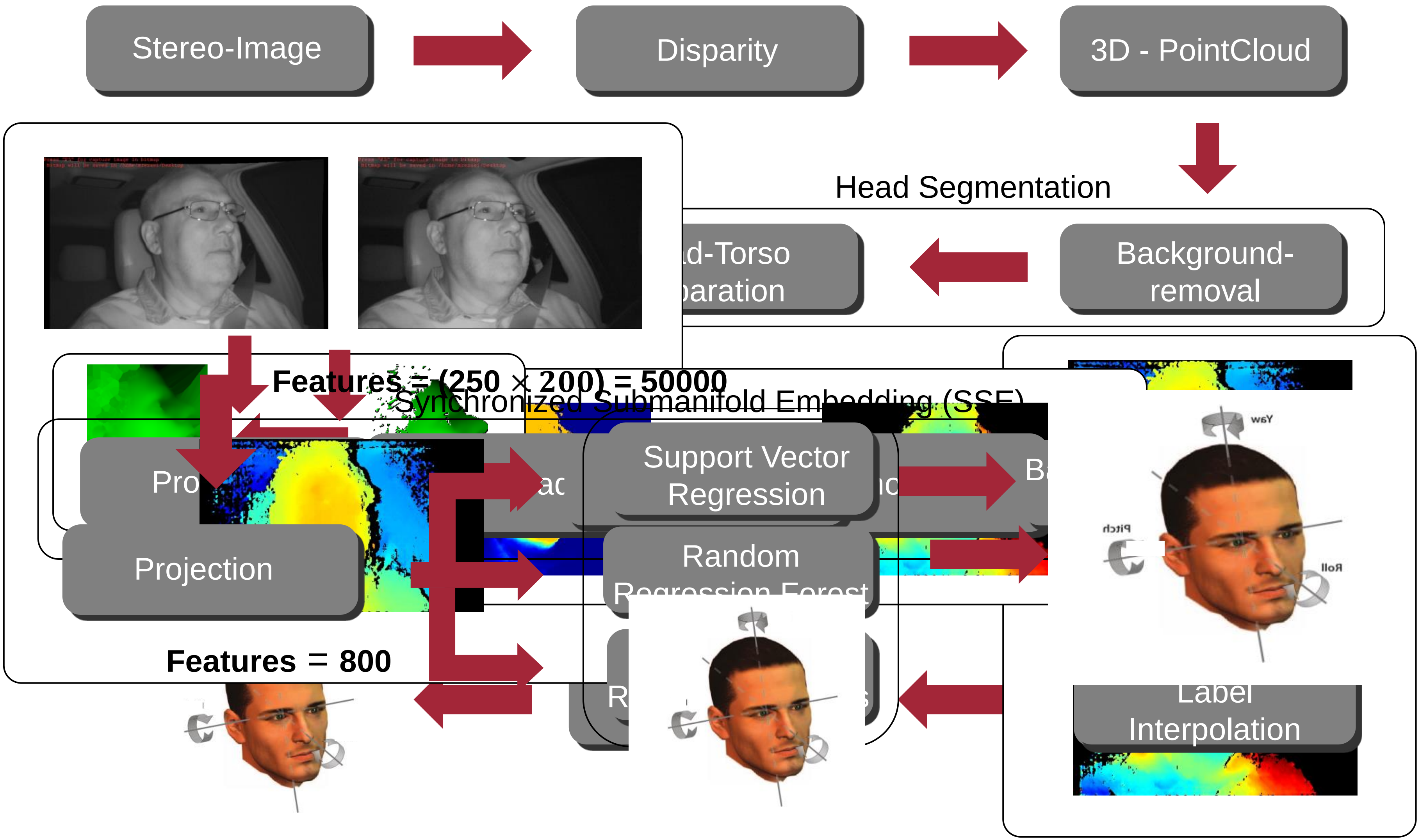


x,y,z / s - Position
α - Yaw
β - Pitch
γ - Roll

Outline

1. Motivation
2. Processing Chain Overview
3. Training and Testing
4. Results and Evaluations
5. Further work

Processing Chain - Overview



Data Gathering – Drive Simulator

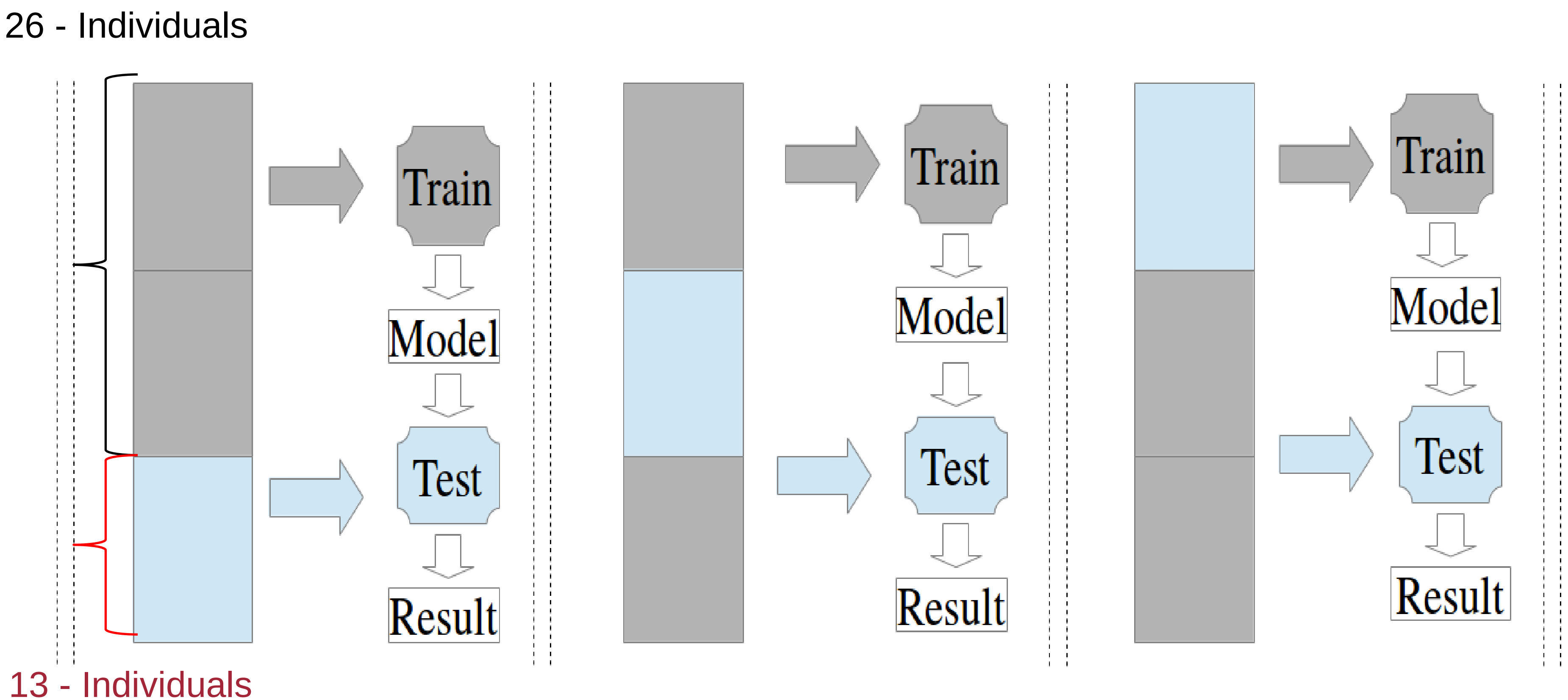
Laser bird sensor

- 39 Individuals
- Ca. 1hr drive time
- Includes women/men of different size, race and age groups.
- Number of samples ≈ 70000

- Head orientation : Quaternions
- 4 – dimensional predictor variable

Training and Testing

- 3 - fold cross validation is used for parameter study of the algorithms.



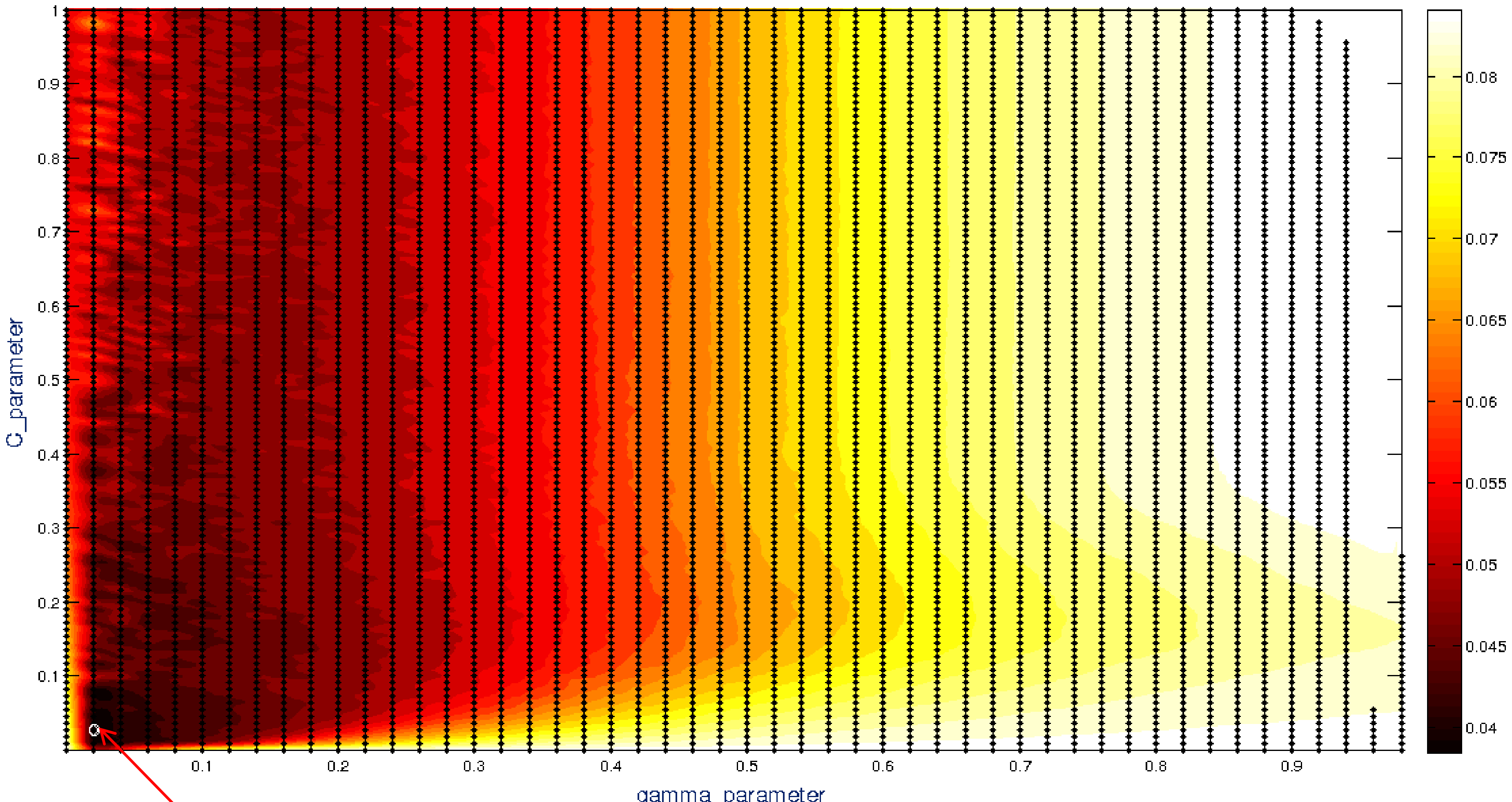
Parameter Study

Algorithms	Feature dimension	Fixed Parameters	Optimized Parameters	Studied Ranges *(grid_step, grid_min, grid_max)*
Support Vector Regression	[69410 × **10**]	Kernel type: - Linear - RBF	Tolerable deviation – Є Cost Parameter – C RBF- gamma (Υ)	Є – (9e-4, 1e-4, 1e-2) C – (9e-4, 1e-4, 1) Υ – (5e-4, 1e-4,1)
Random Regression Forests	[69410 × **10**]	Number of trees (100) Maximum Depth (20)	Minimum sample count (s) Predictor Variable (m)	s – (1, 2, 500) m – (1,1,10) m – (1,50,800)
Extremely Randomized Trees	[69410 × **10**]	Number of trees (100) Maximum Depth (20)	Minimum sample count (s) Predictor Variable (m)	s – (1,2,500) m – (1,1,10)

Accuracy Measurement: $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$

Massive Parallel computation

Optimal Parameters - Support Vector Regression

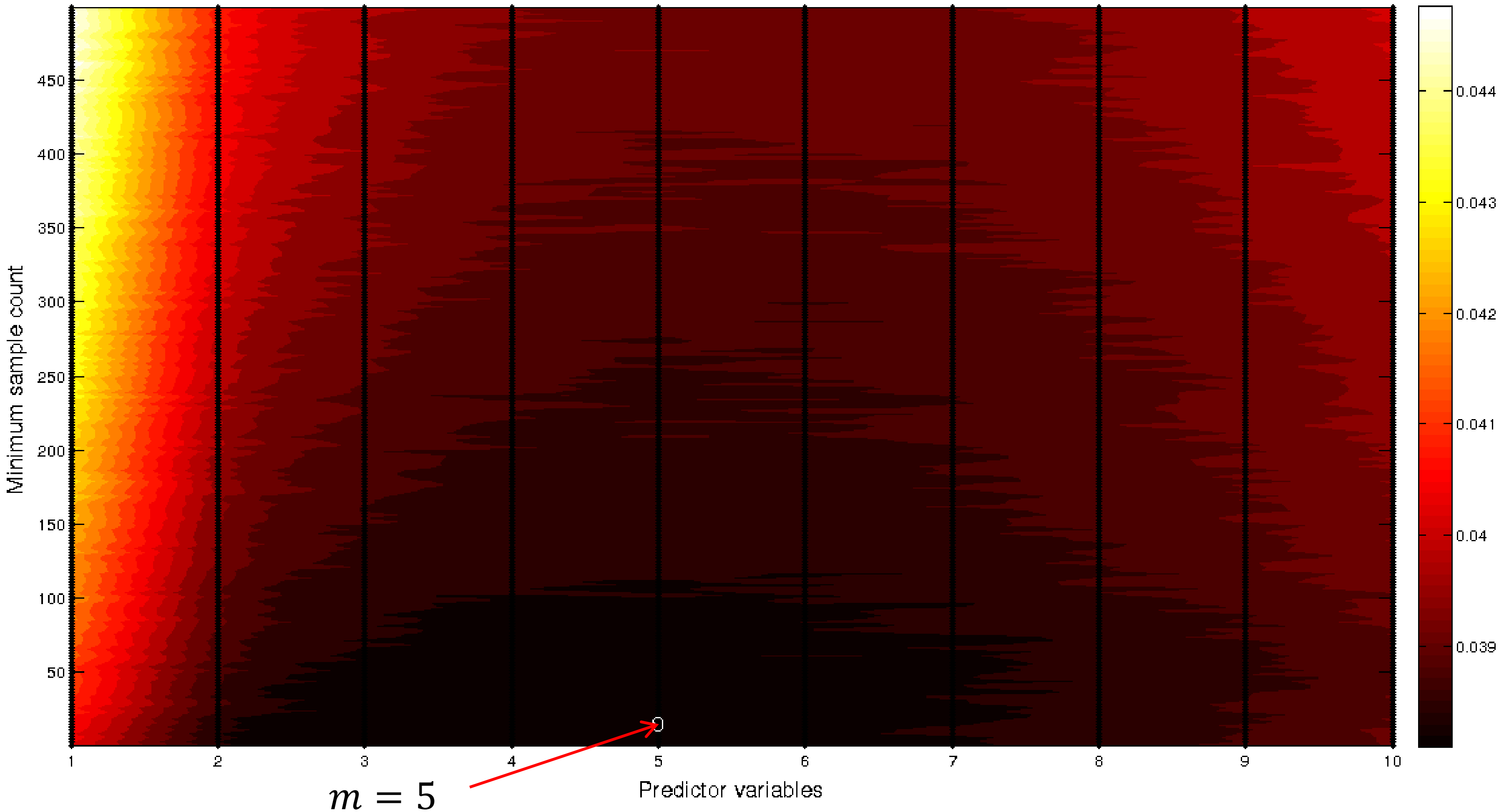


$\epsilon = 0.0082$

$C = 0.0151$

$\gamma = 0.011$

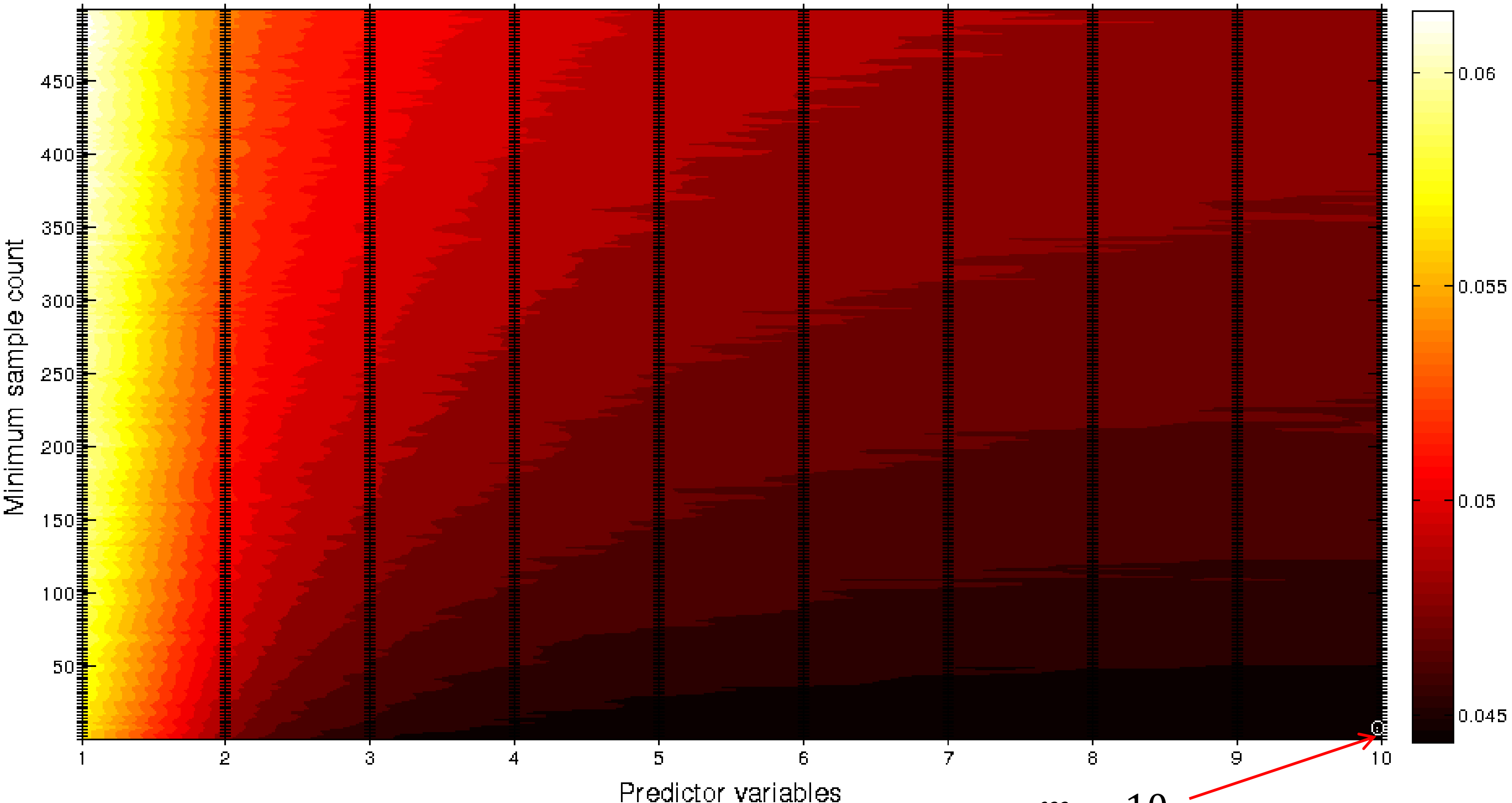
Optimal Parameters – Random Regression Forests



$m = 5$

$s = 17$

Optimal Parameters - Extremely Randomomized Trees

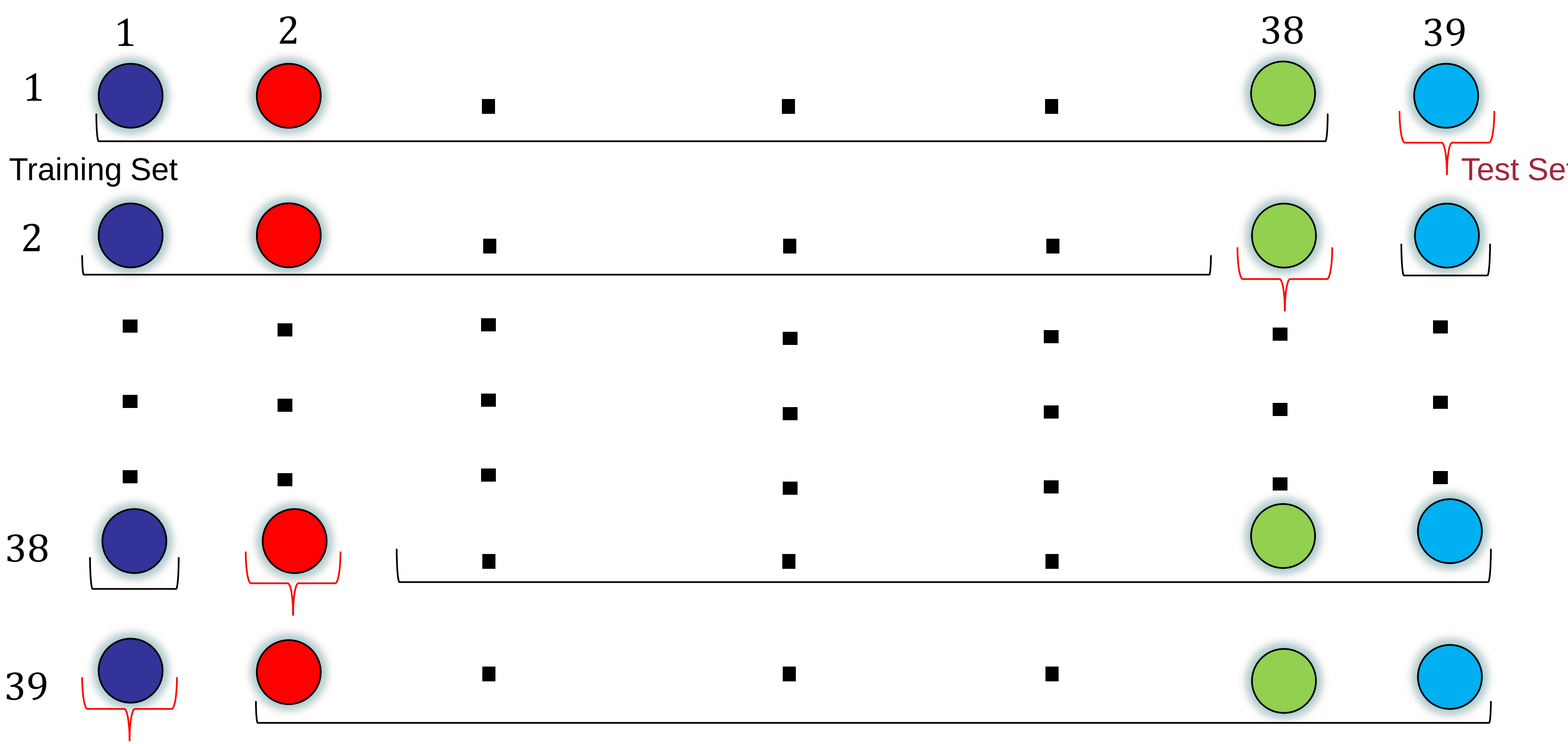


$m = 10$

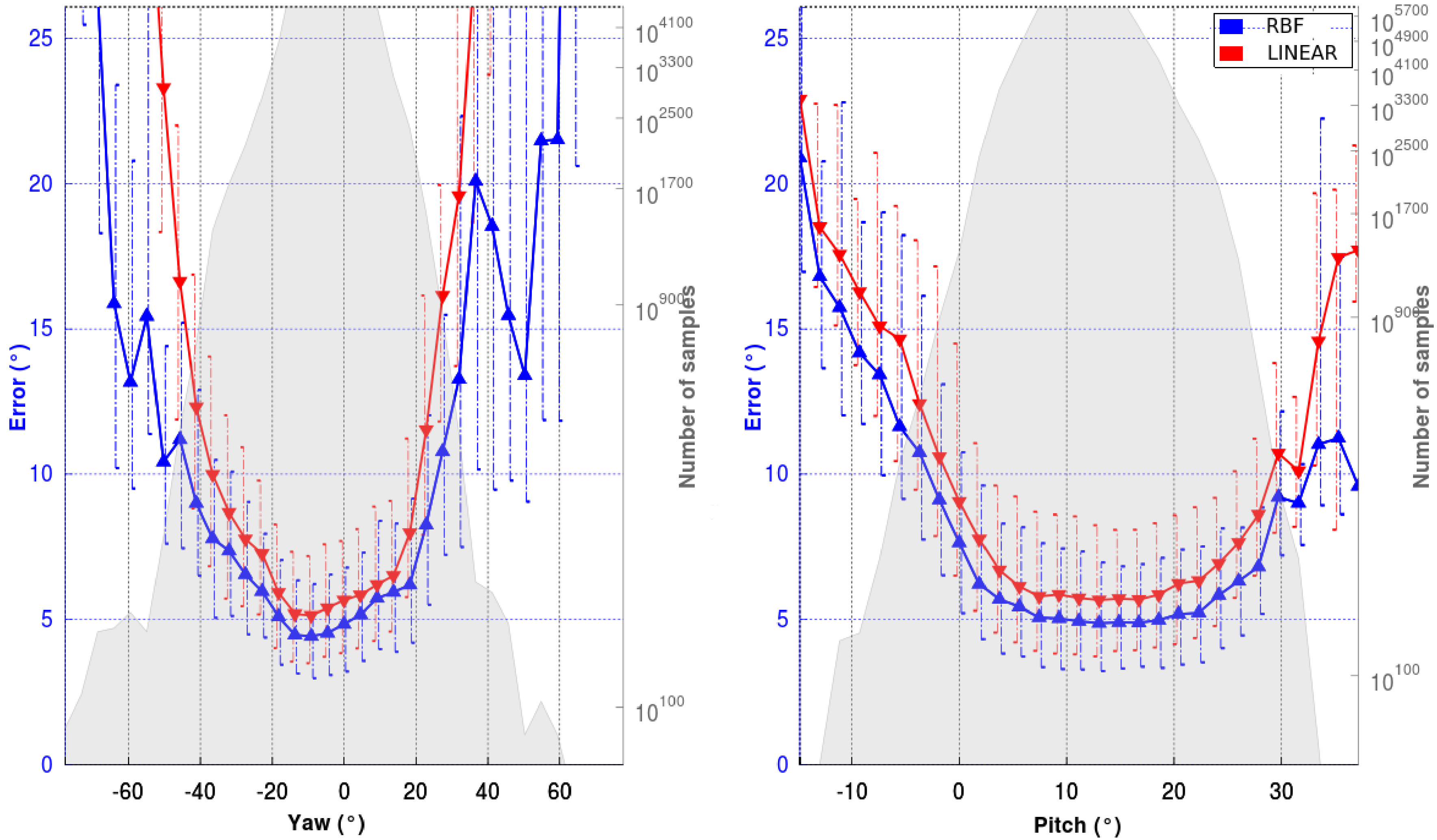
$s = 8$

Leave - One - Individual - Out

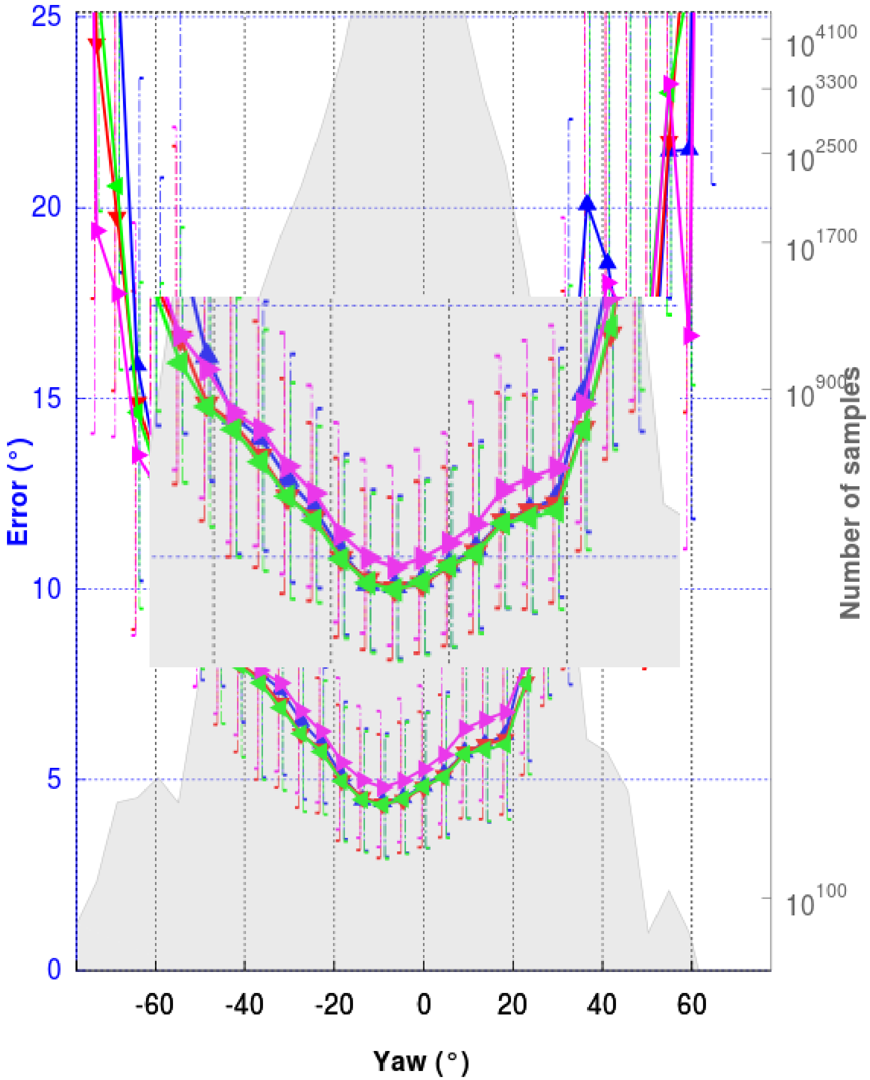
- 39 - fold cross validation
- Used for training the algorithms with the optimal parameters and evaluation is done on the resulting models.

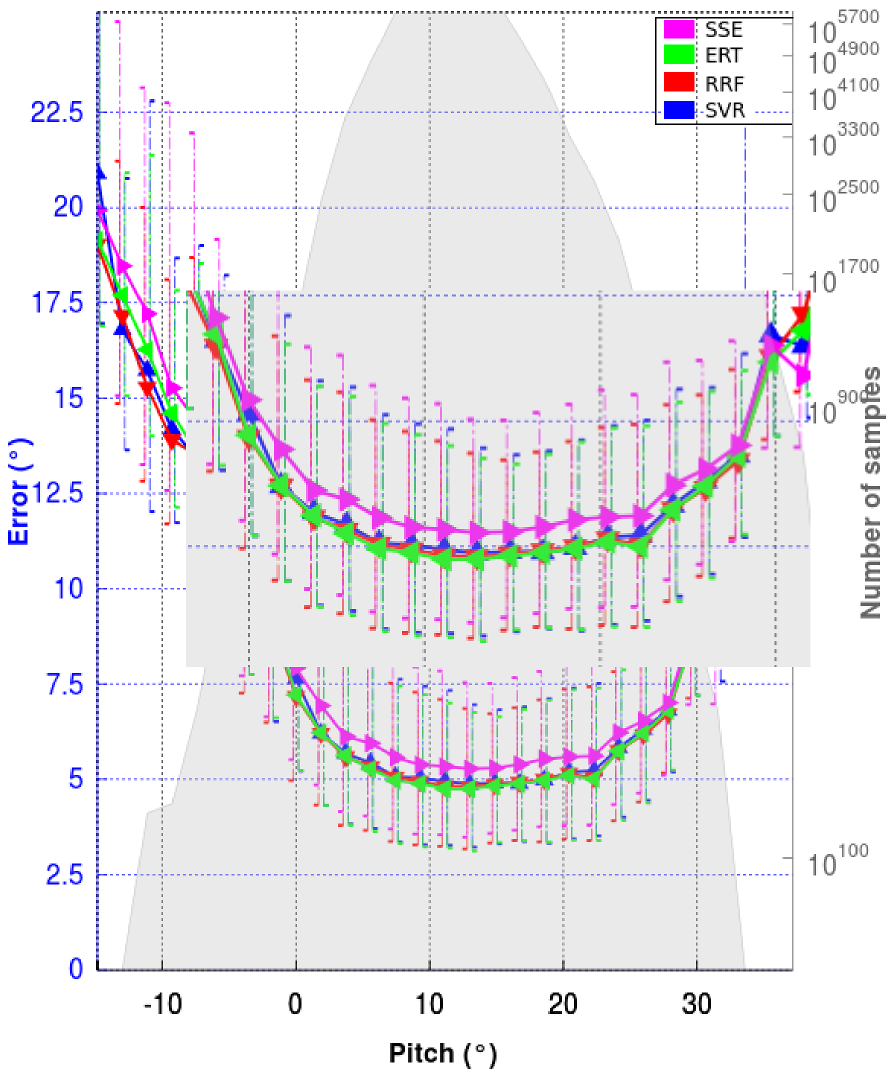


Support Vector Regression – Linear Vs RBF kernel

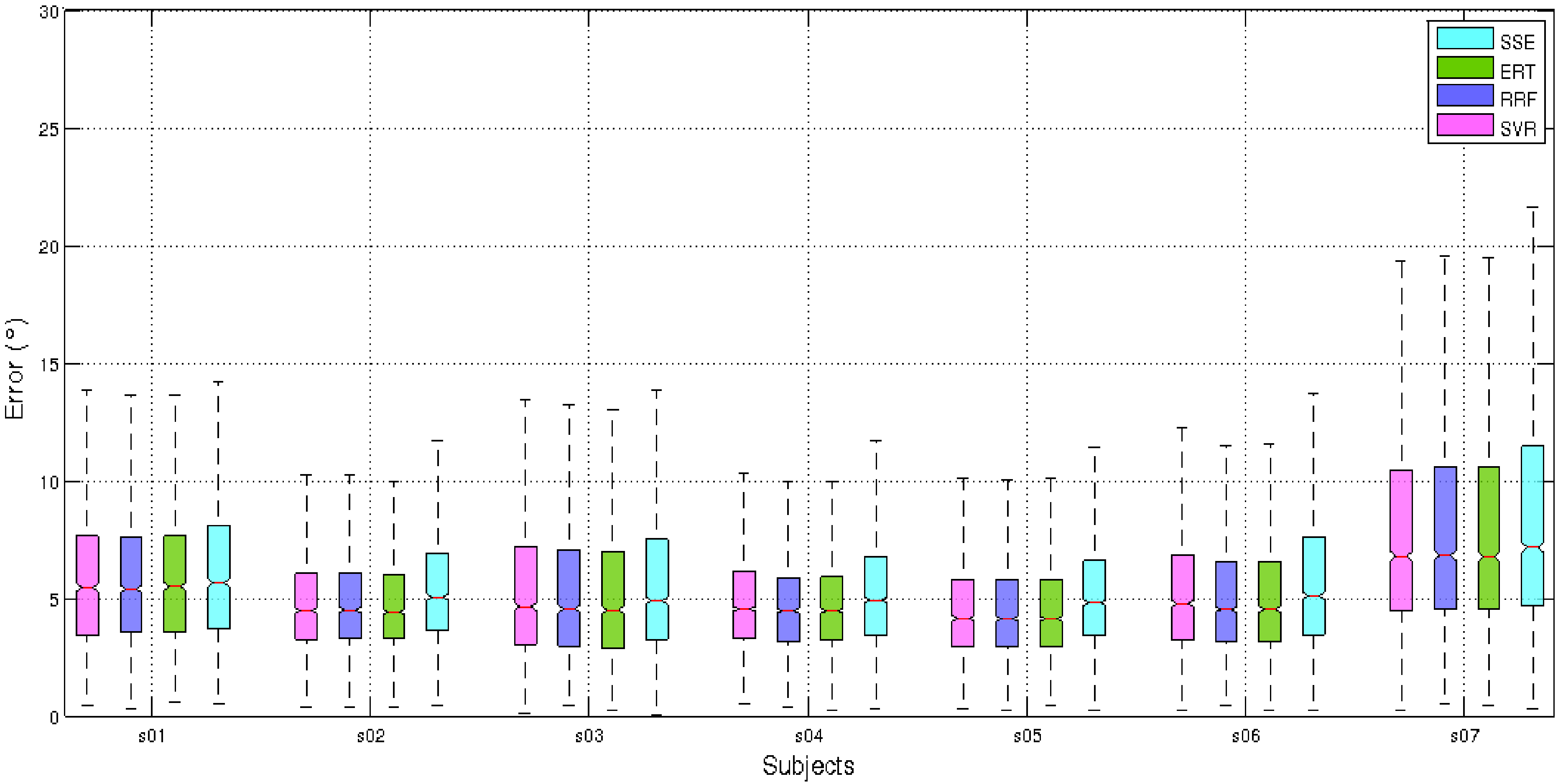


Comparison - SVR Vs RRF Vs ERT Vs SSE





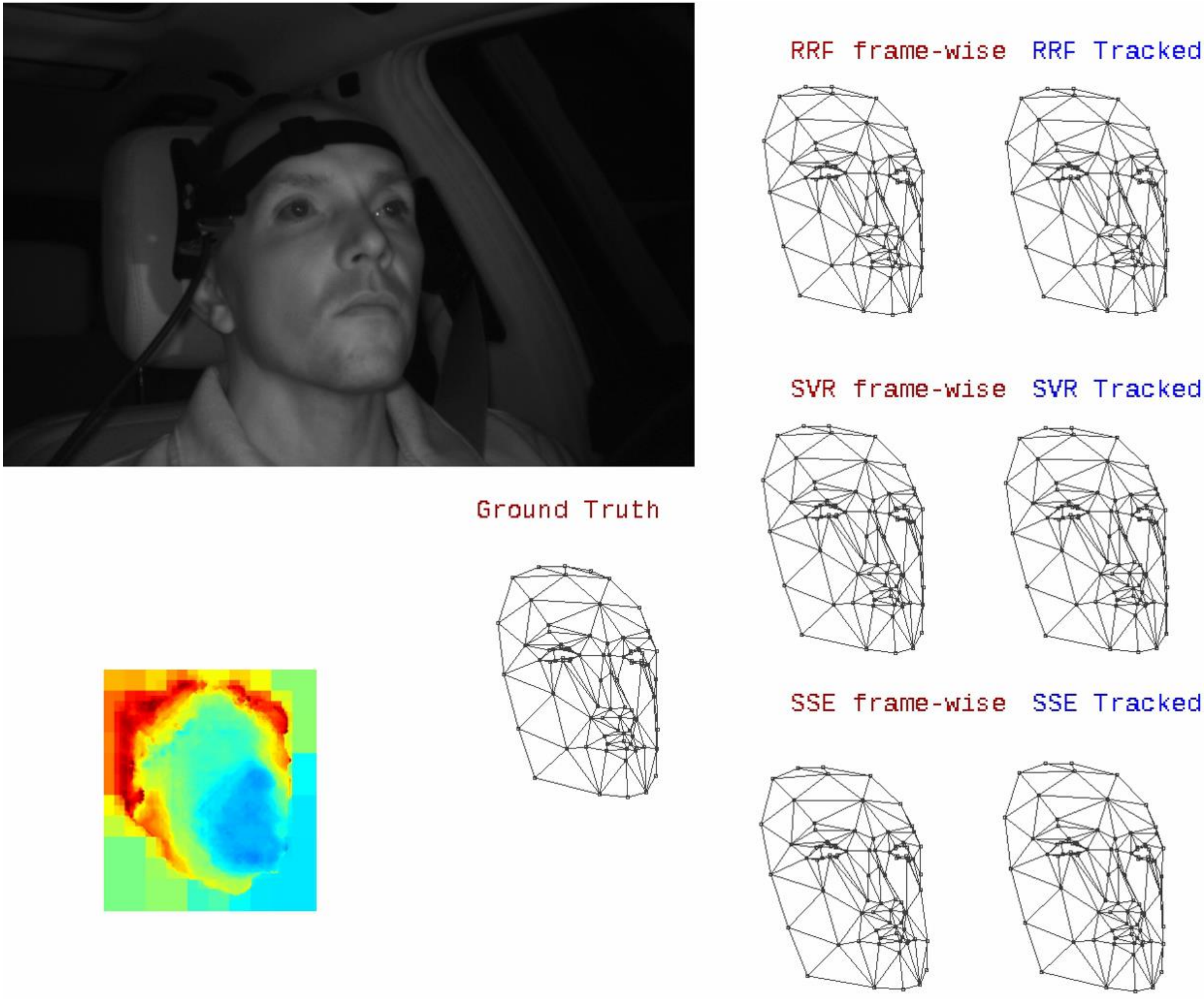
Comparison – Individual Errors – SVR Vs RRF Vs ERT Vs SSE



driveu

Time Cost

Algorithms	Time - Cost	
	Observed Prediction Time per frame in ms	Average Time – Complexity (O-notation)
SVR – Linear	1.126	$O(p)$
SVR – RBF	´4.658	$O(vp)$
RRF	1.194	$O(Bd)$
ERT	1.348	$O(Bd)$
SSE	1.52	$O(\log n)$


RRF frame-wise
RRF Tracked
SVR frame-wise
SVR Tracked
Ground Truth
SSE frame-wise
SSE Tracked

Further Work

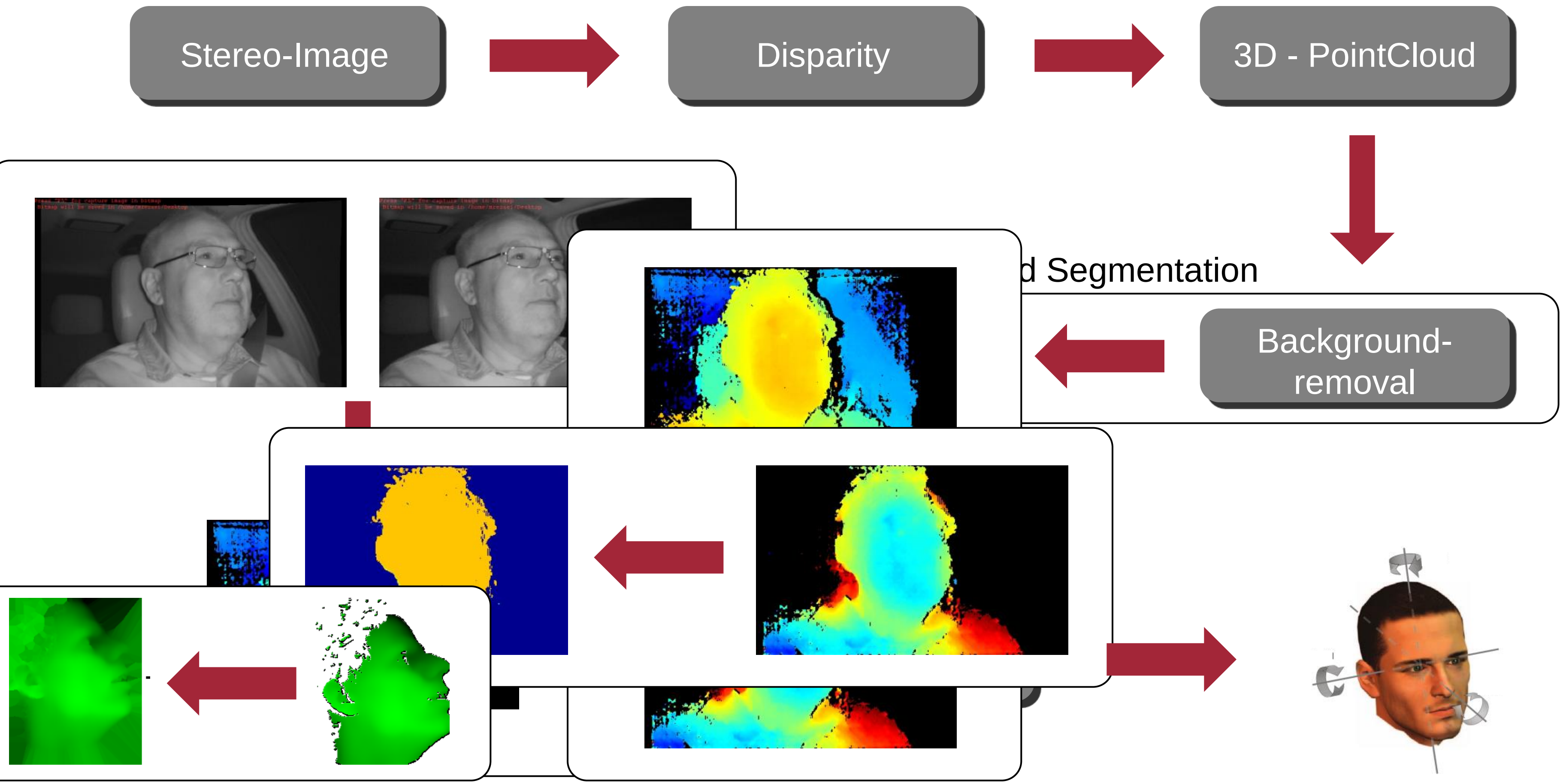
- Better gound truth information extraction method – Iterative Closest Point (ICP)

Thank You

Outline

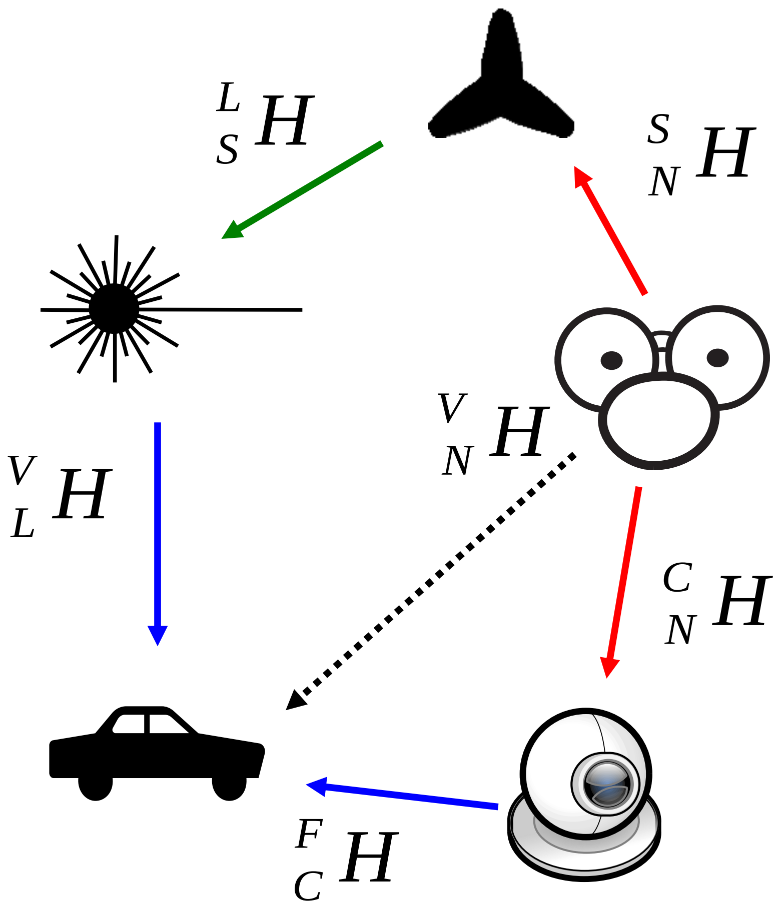
1. Motivation
2. Processing Chain Overview
3. Proposed Regression Algorithms
    - Support Vector Regression
    - Random Regression Forest
    - Extremely Randomized Trees
4. Training and Testing
5. Results and Evaluations
6. Further Work

Processing Chain - Overview



driveu

LaserBIRD → Ground Truth



Ground Truth

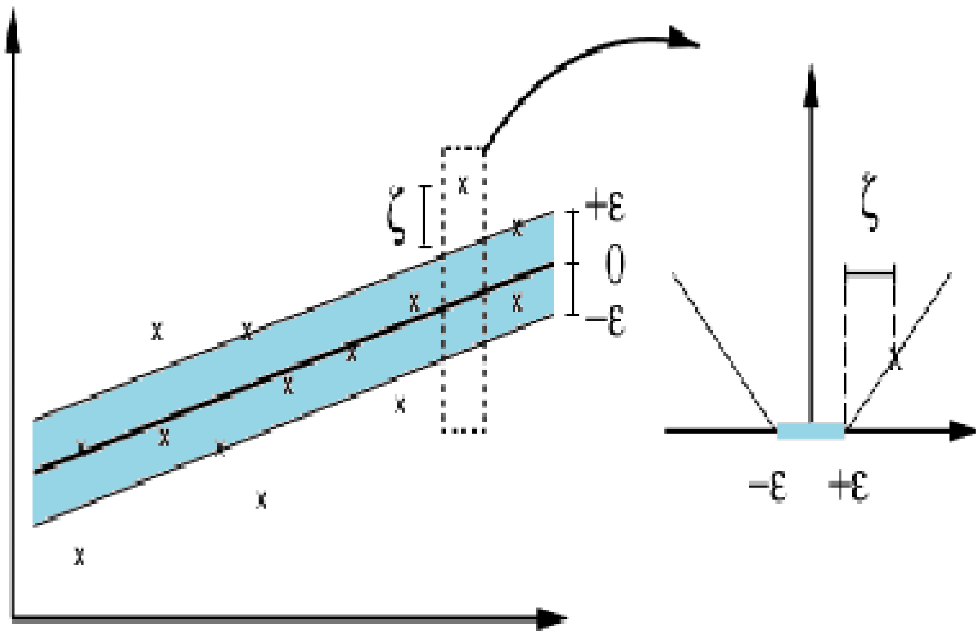
$$ {}^{C}_{N}H = {}^{V}_{C}H^{-1}\,{}^{V}_{L}H\,{}^{L}_{S}H\,{}^{S}_{N}H $$

Legend:

- Laserscanner
- LaserSensor
- Vehicle
- Camera
- Face (Nose)

Support Vector Regression

$$f(x) = \langle w, x \rangle + b$$



minimize $\frac{1}{2}||w||^2$

subjected to $\begin{cases} y_i - \langle w, x \rangle - b \leq \epsilon \\ \langle w, x \rangle + b - y_i \leq \epsilon \end{cases}$

minimize $\frac{1}{2}||w||^2 + C \sum_{i=1}^{l} (\xi_i + \xi_i^*)$

subjected to $\begin{cases} y_i - \langle w, x \rangle - b \leq \epsilon + \xi_i \\ \langle w, x \rangle + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i + \xi_i^* \geq 0 \end{cases}$

The Prediction Function:

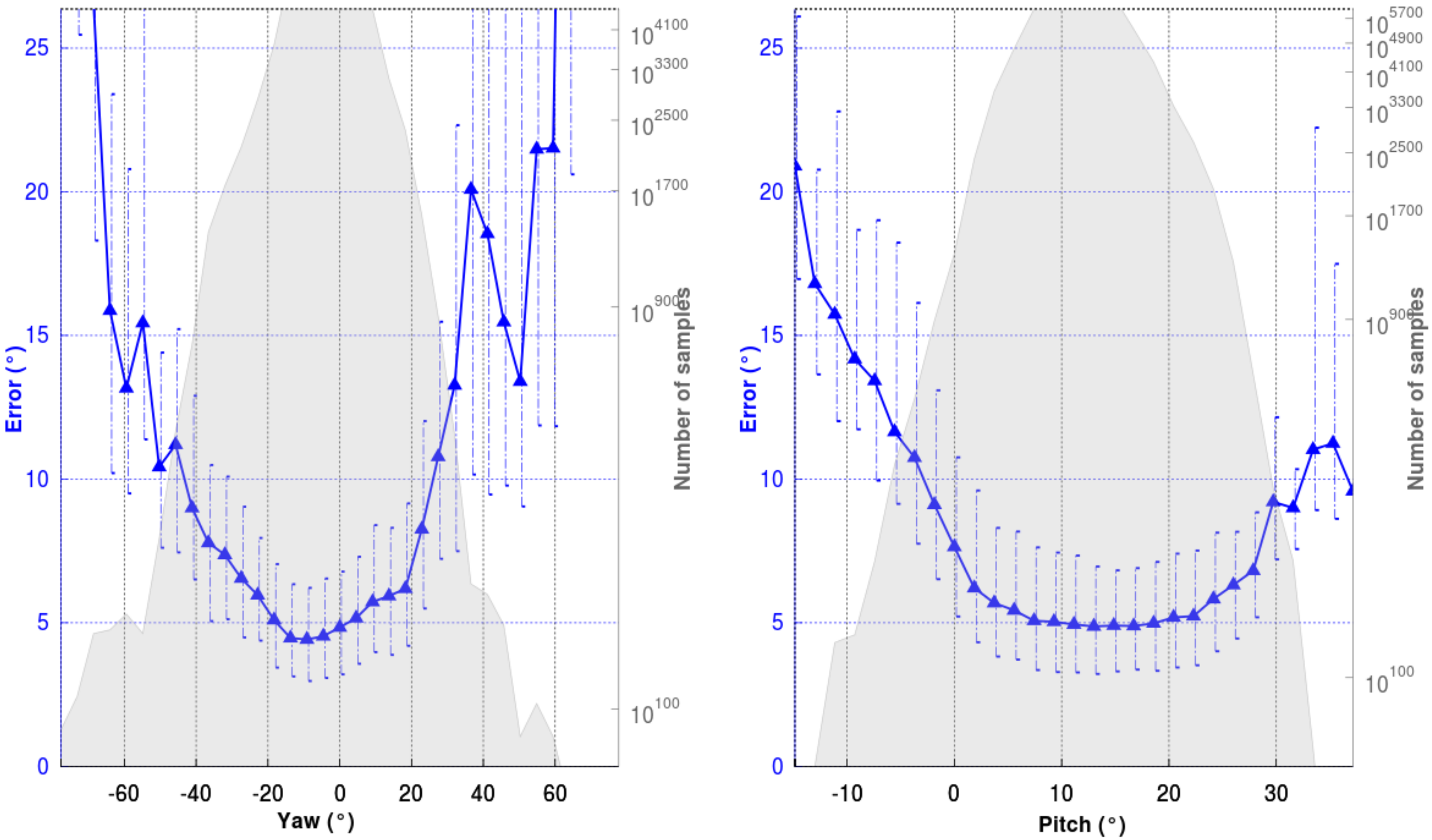
$$\mathrm{w} = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) \Phi(x_i) \quad \text{and} \quad f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) k(x_i, x_j) + b$$

$$\alpha, \alpha^* \in [0, C]$$

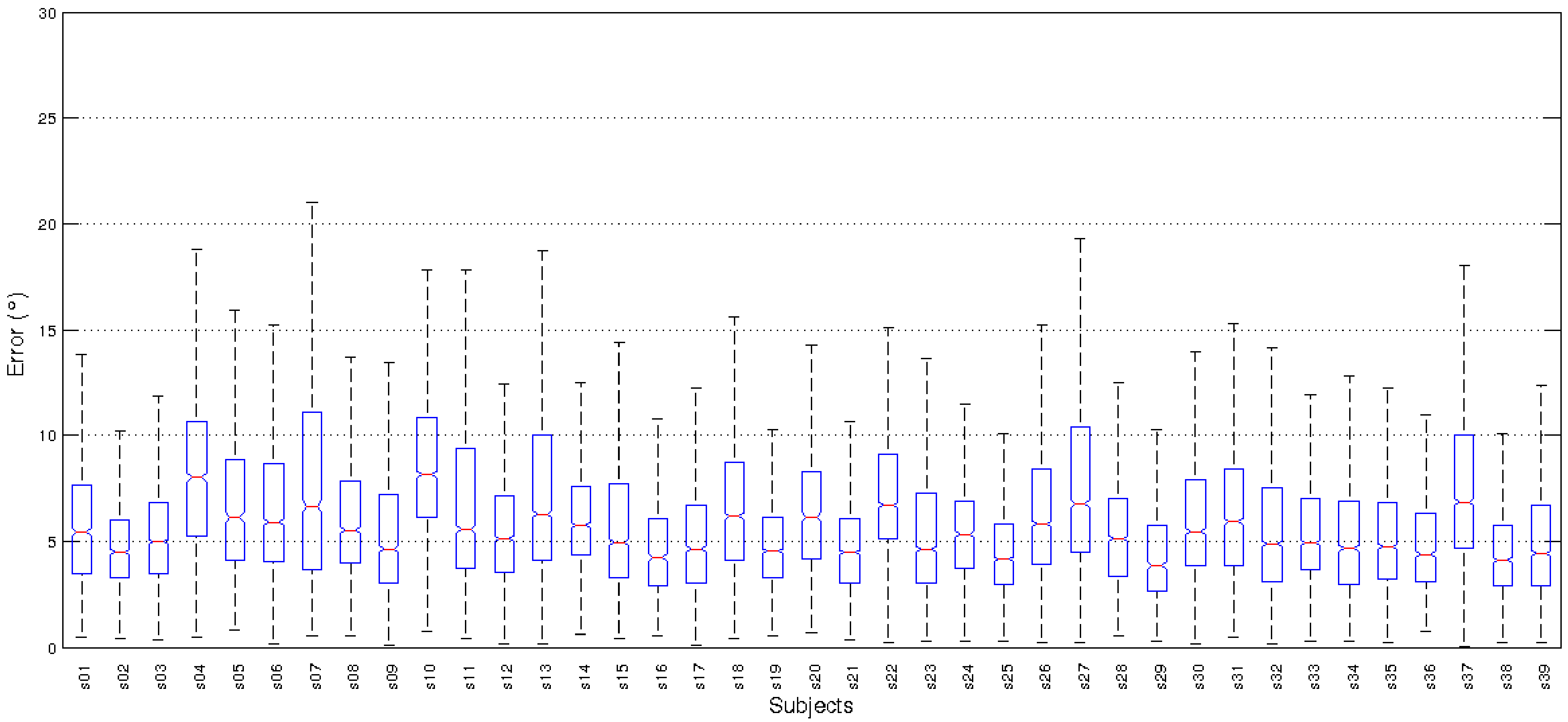
Comparison - SVR Vs RRF Vs ERT Vs SSE

Algorithms	Mean error in °	Accuracy in %		
		$\leq$ 15°	$\leq$ 10°	$\leq$ 5°
SVR – Linear	8.42	90.64	78.60	34.94
SVR – RBF	6.96	94.38	85.34	45.42
RRF	6.82	94.93	86.17	46.23
ERT	6.79	95.03	86.33	46.67
SSE	7.41	93.91	83.12	40.02

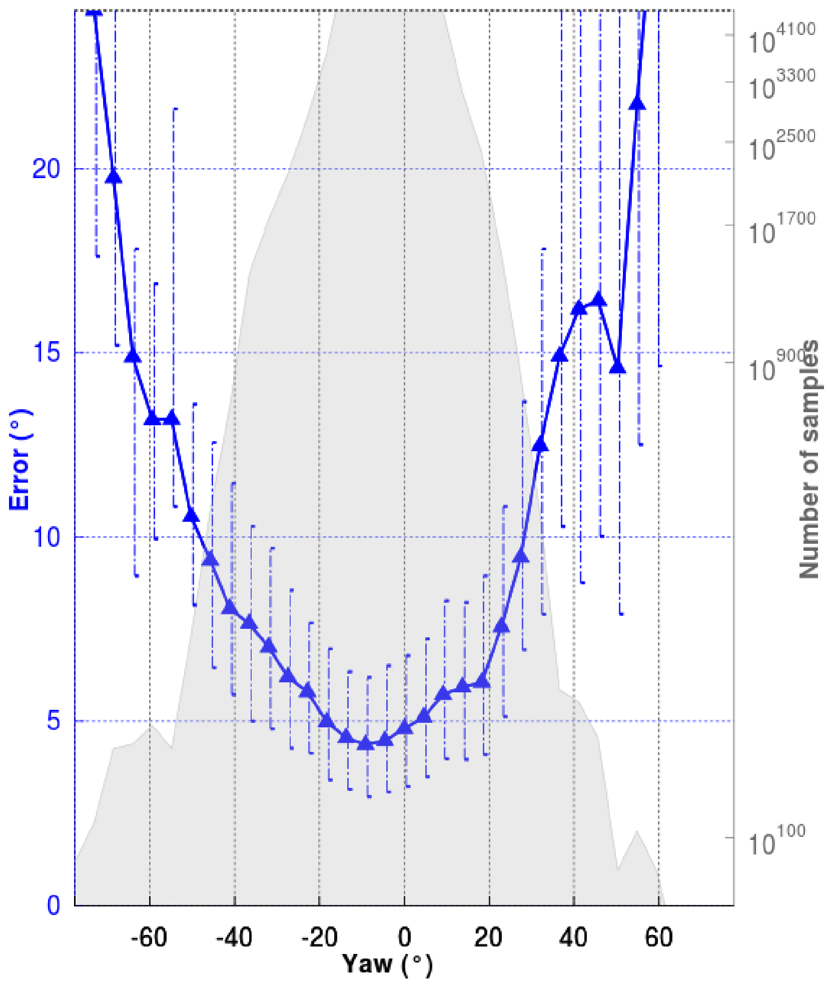
Evaluation - Support Vector Regression – Projected Error

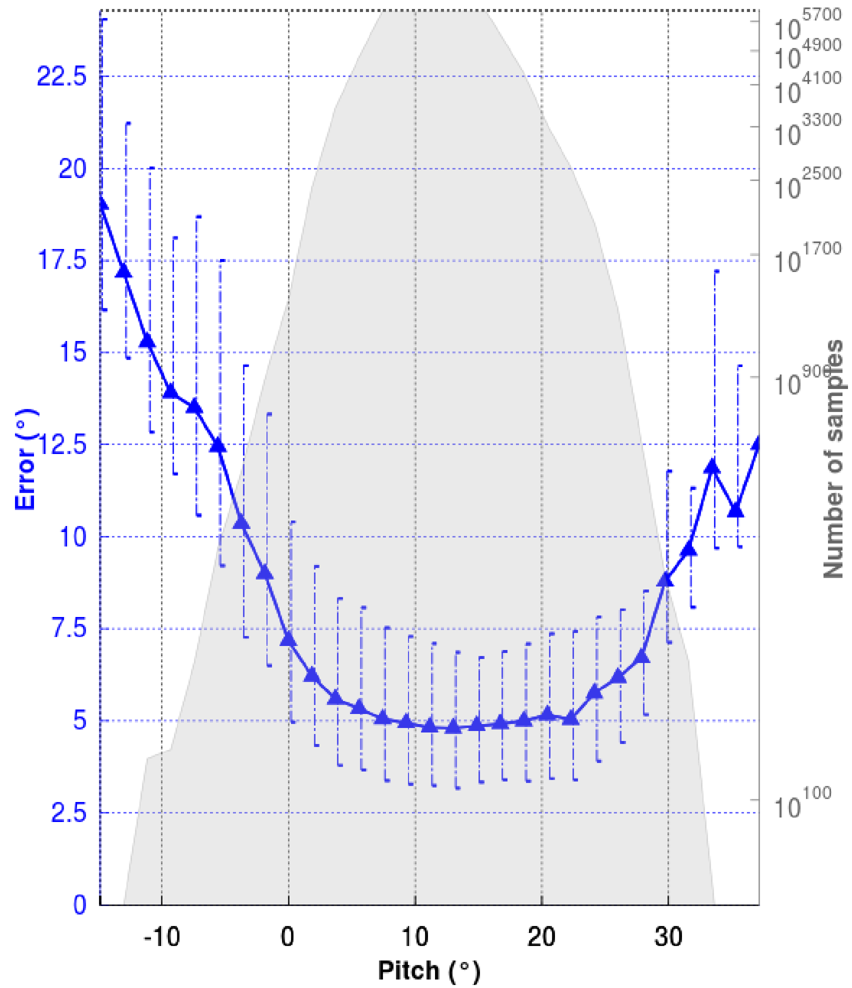


Evaluation – Support Vector Regression – Individual Errors

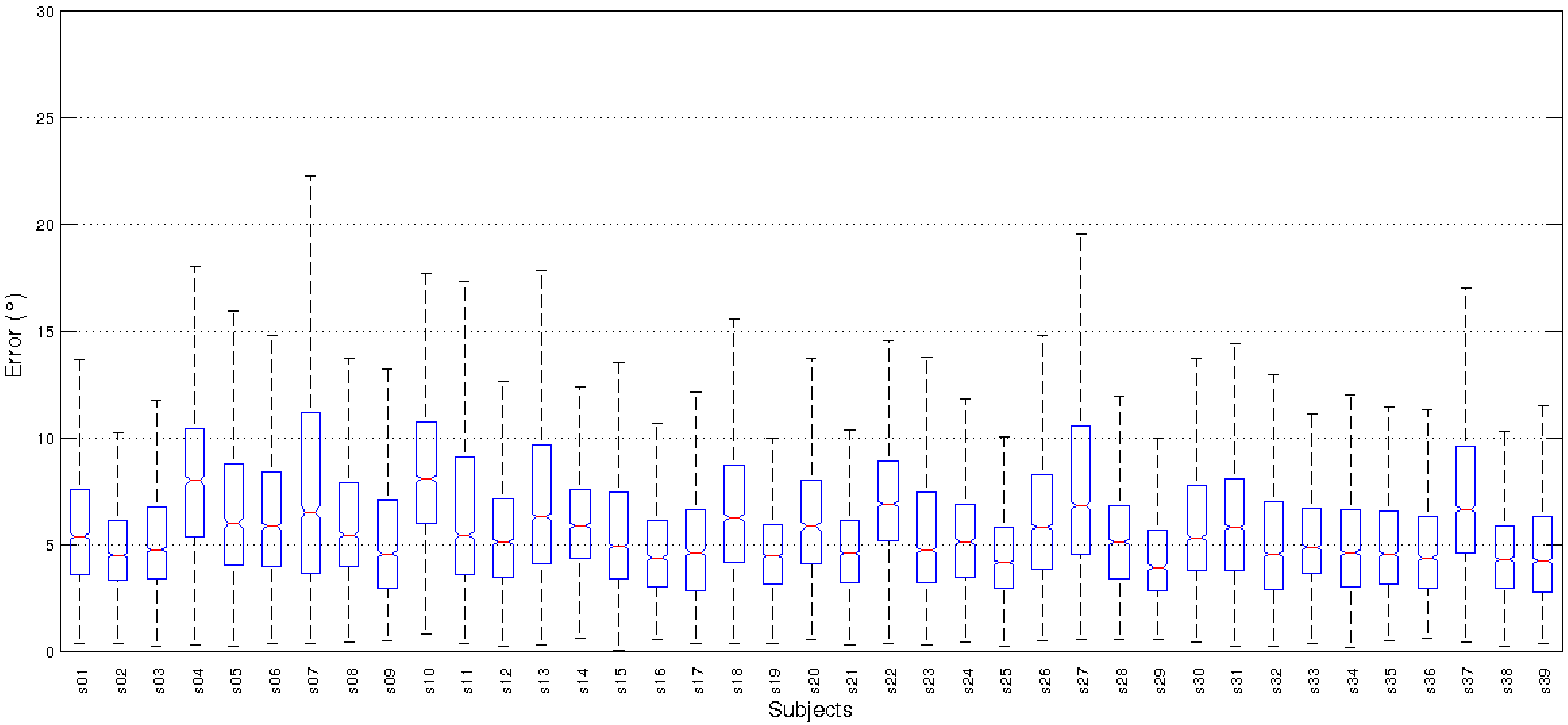


Evaluation – Random Regression Forests – Projected Error

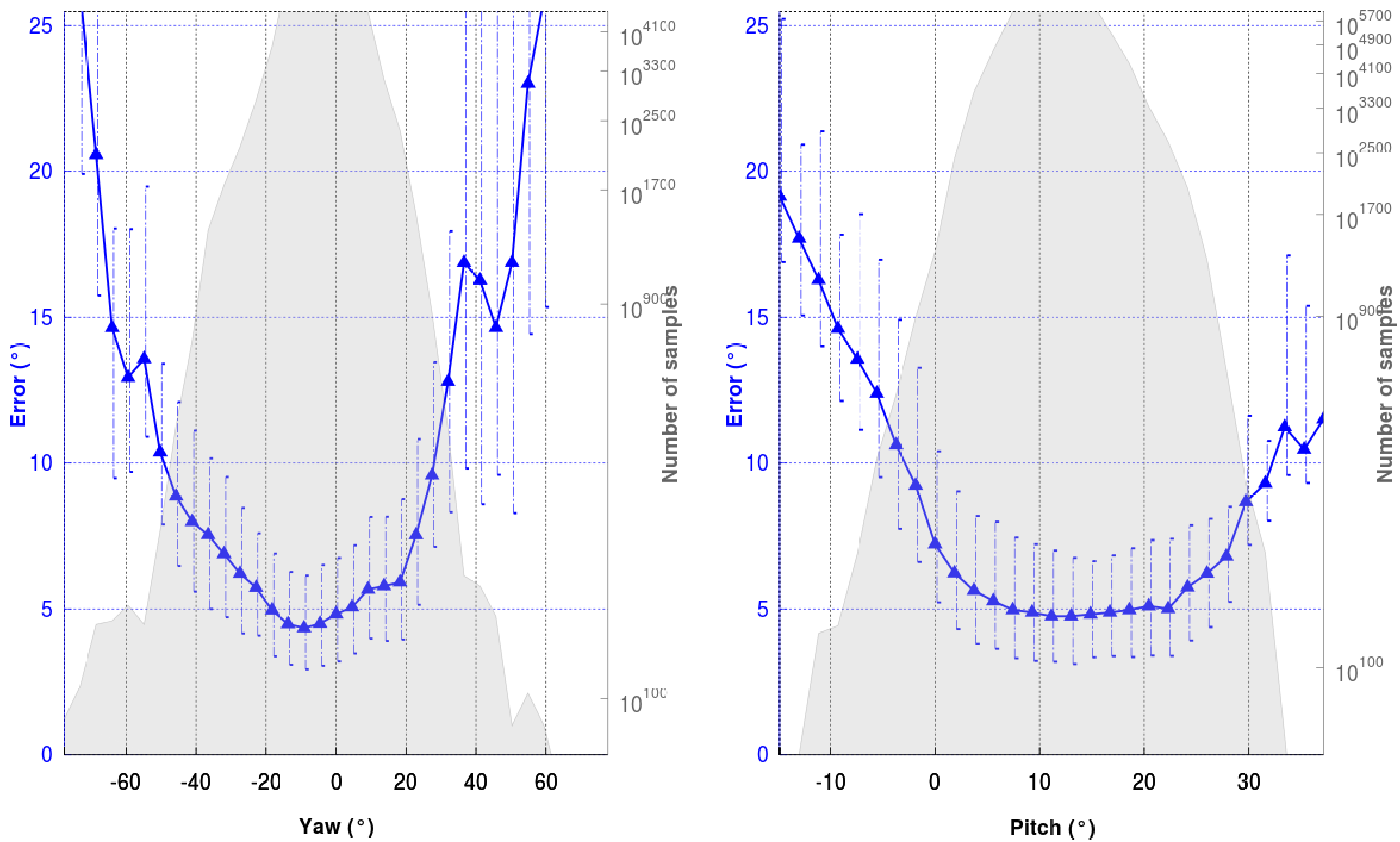




Evaluation – Random Regression Forests – Individual Error

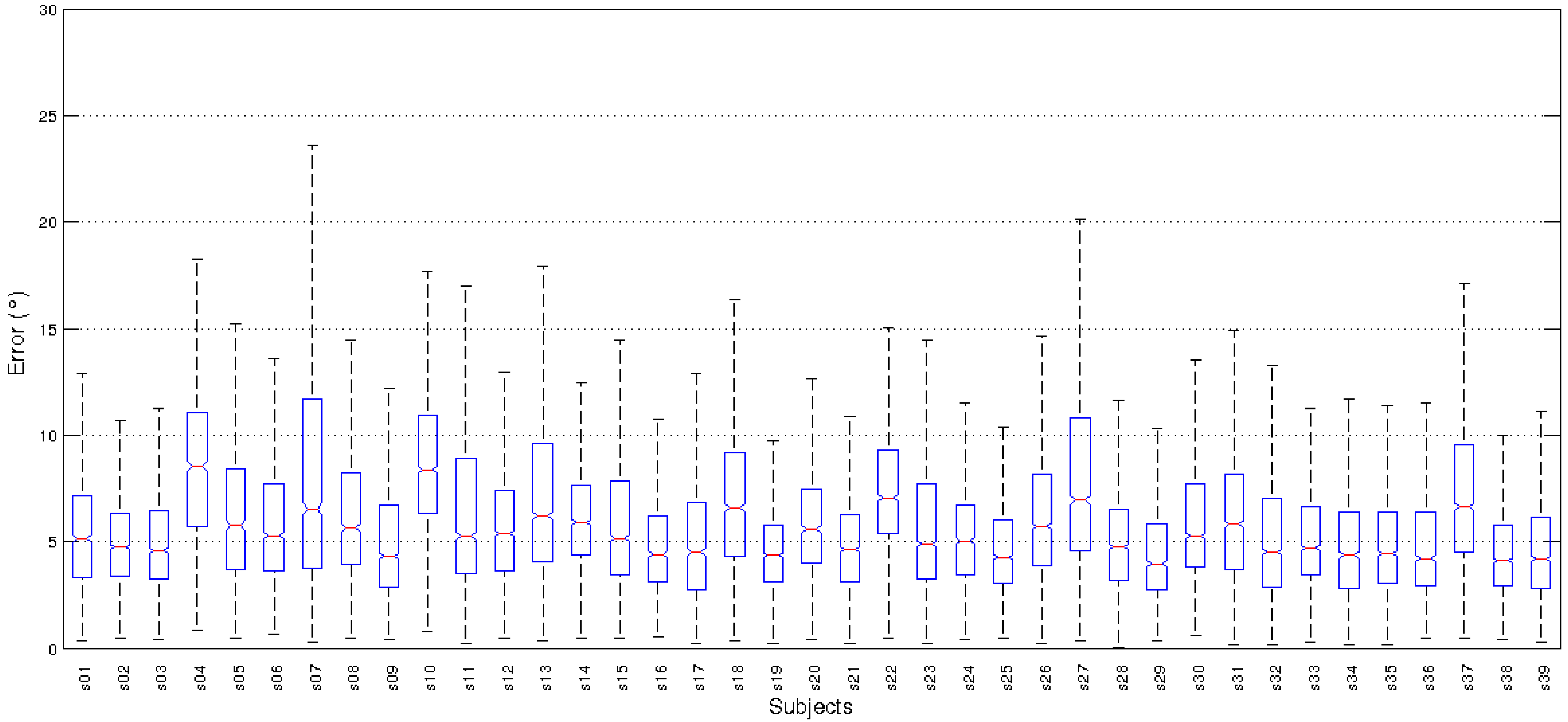


Evaluation – Extremely Randomized Trees – Projected Error



-

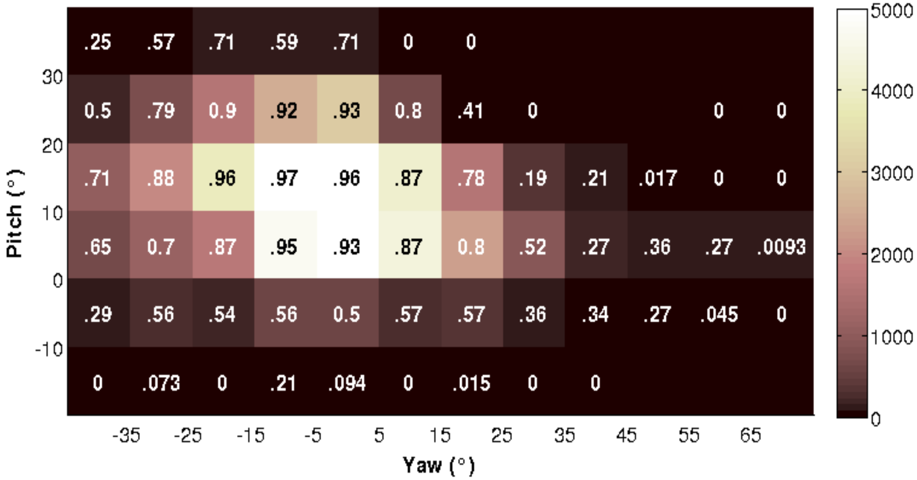
Evaluation – Extremely Randomized Trees – Individual Error



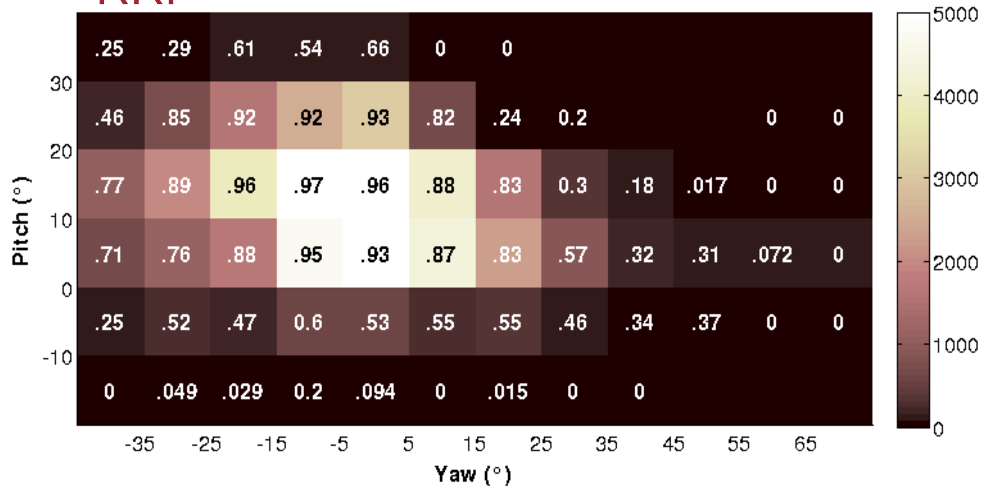
driveu

Comparison - SVR Vs RRF Vs ERT Vs SSE

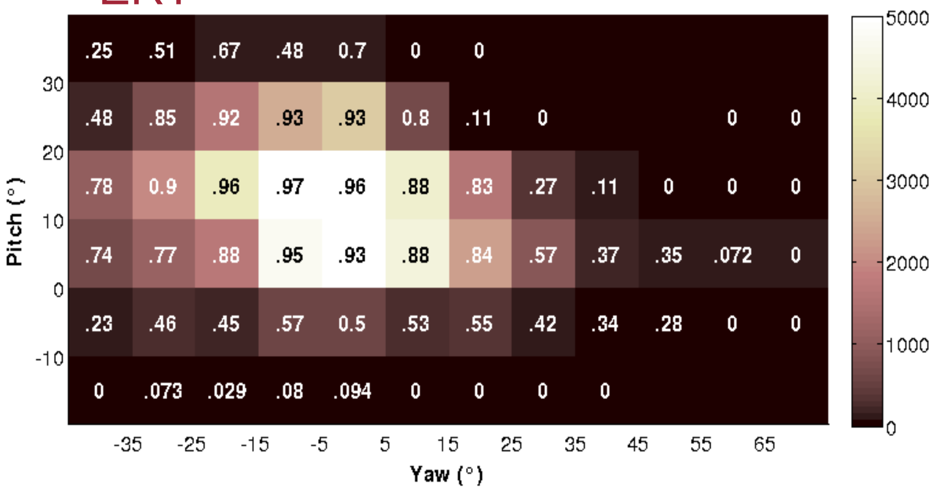
SVR



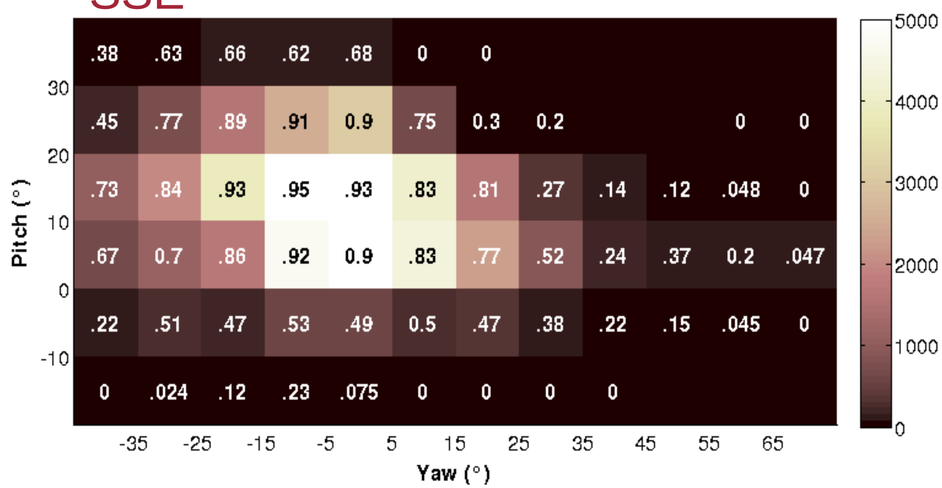
RRF



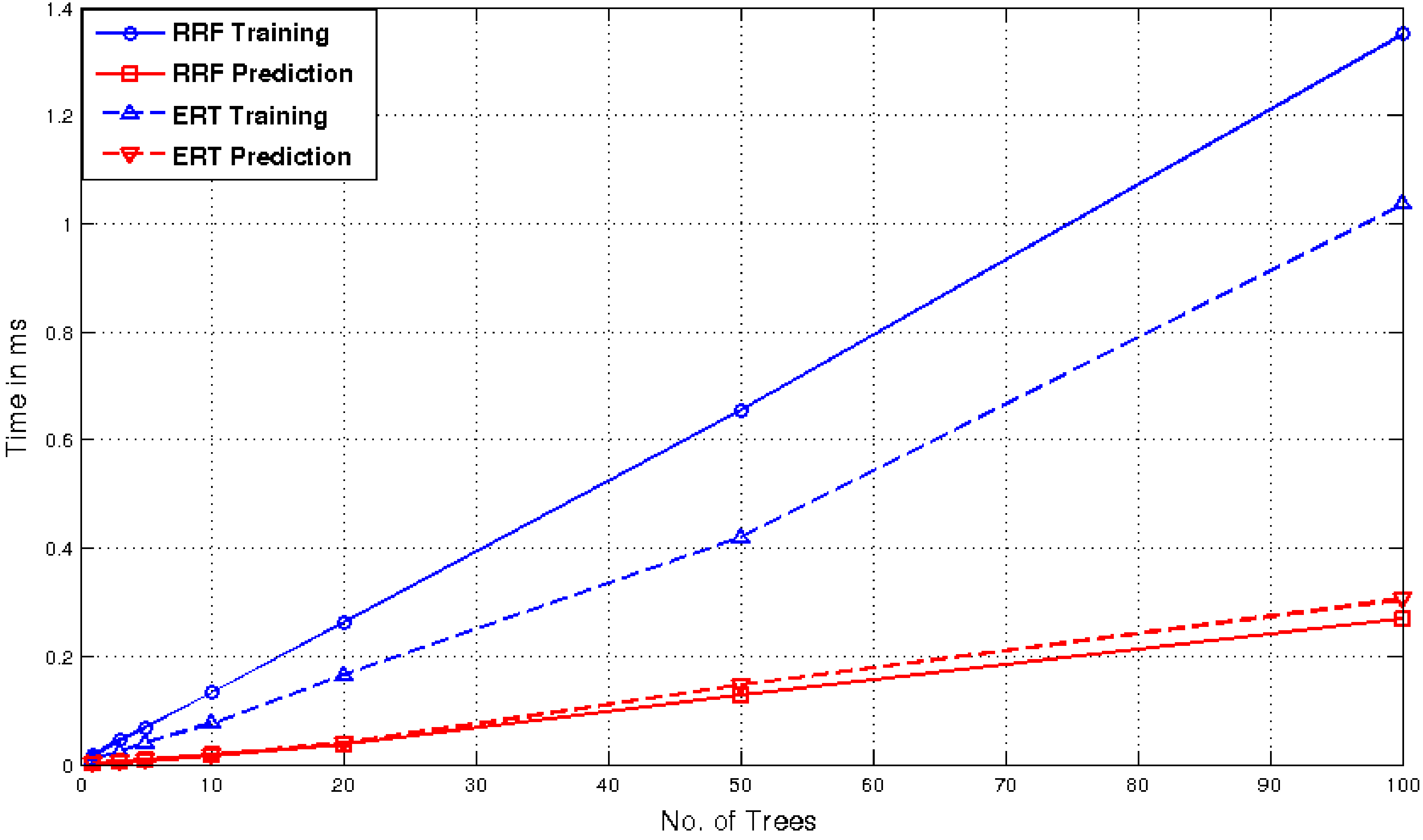
ERT



SSE



Time Cost


RRF Training
RRF Prediction
ERT Training
ERT Prediction
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Time in ms
0
10
20
30
40
50
60
70
80
90
100
No. of Trees