

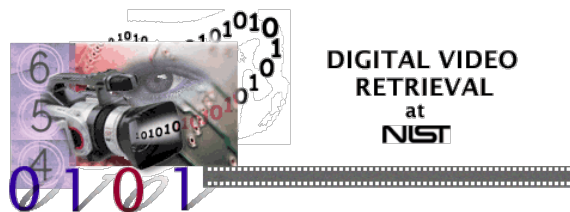
2013 TRECVID Workshop: Surveillance Event Detection (SED) Retrospective + Interactive (rSED+iSED) Task Overview

Jonathan Fiscus (NIST)

Martial Michel (Systems Plus, Inc.)

November 22, 2013

NIST, Gaithersburg, MD, USA

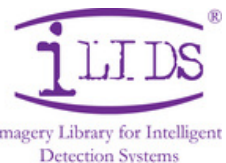


Motivation

- Surveillance Event Detection Motivation
 - SED addresses the need for the advancement of technologies that can perform automatic detection of events in large amounts of surveillance quality video
- rSED Challenges
 - Requires application of several Computer Vision techniques
 - Involves subtleties that are readily understood by humans, difficult to encode for machine learning approaches
 - Can be complicated due to clutter in the environment, lighting, camera placement, traffic, etc.
- iSED Challenges
 - SED remains a difficult task for humans and systems
 - Interactive/relevance feedback have been effectively employed in other related tasks

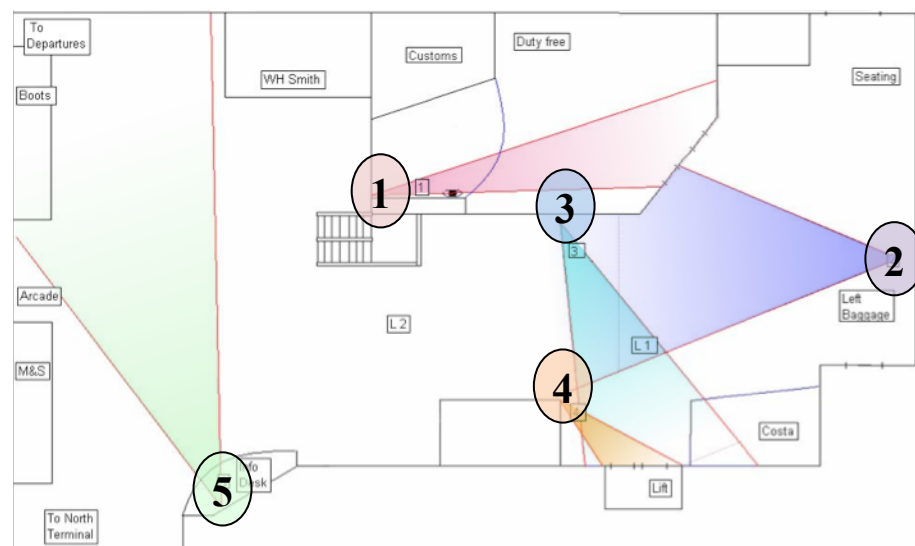
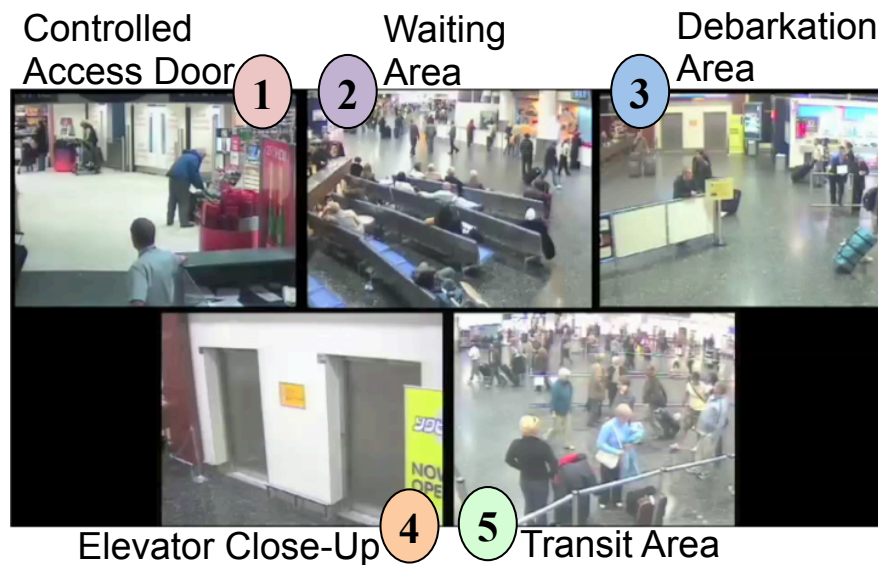
Surveillance Event Detection Tasks

- Retrospective SED (rSED) Task : Given a textual description of an *observable event of interest*, **automatically detect** all occurrences of the event in a non-segmented corpus of video
- Identify each detected event observation by:
 - The **temporal extent** (*beginning and end frames*)
 - A **decision score**: a numeric score indicating how likely the event observation exists with more positive values indicating more likely observations (normalized)
 - An **actual decision**: a boolean value indicating whether or not the event observation should be counted for the primary metric computation
- Interactive (iSED) Task : Given a textual description of an *observable event of interest*, at **test time allow a searcher 25 minutes to filter incorrect event detections** in a non-segmented corpus of video



Evaluation Source Data

- Reused same test data as SED '09, '10, '11 and '12 evaluations
- UK Home Office collected CCTV video from 5 camera views at a busy airport
- Development Set
 - 100 hours of video
 - 10 events annotated on 100% of the data
- Evaluation Set
 - “iLIDS Multiple Camera Tracking Scenario Training set”
 - An identified 15-hours of the 45-hour set evaluated (**same as 2012**)
 - 10 events annotated on 1/3 of the data
 - 7 events evaluated



Events and Instances per Hour (IpH)

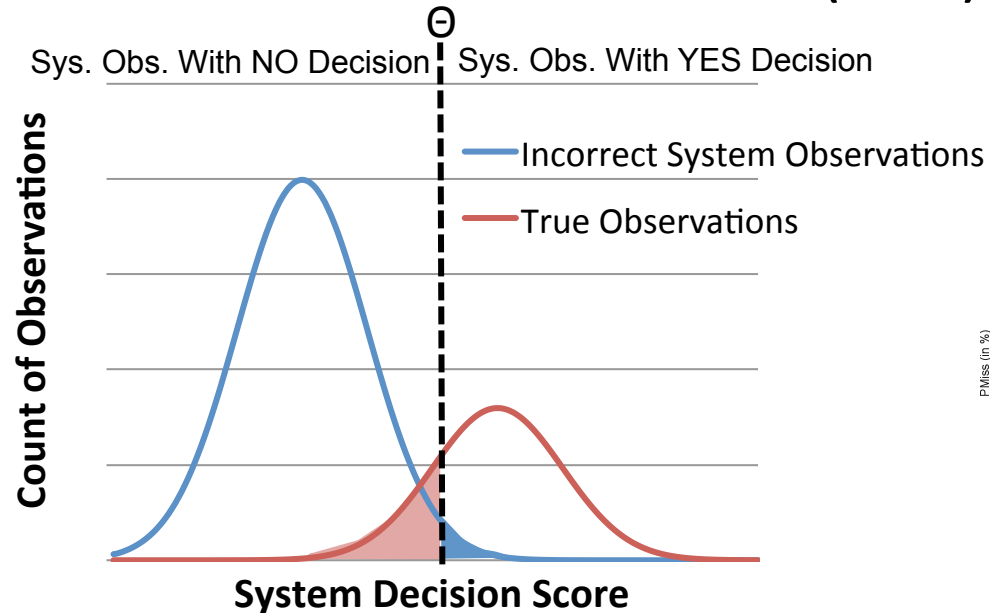
Single Person events		
PersonRuns	7.02 IpH	Someone runs ← <i>Lowest frequency</i>
Pointing	69.74 IpH	Someone points ← <i>Highest frequency</i>
Single Person + Object events		
CellToEar	12.73 IpH	Someone puts a cell phone to his/her head or ear
ObjectPut	40.74 IpH	Someone drops or puts down an object
Multiple People events		
Embrace	11.48 IpH	Someone puts one or both arms at least part way around another person
PeopleMeet	29.46 IpH	One or more people walk up to one or more other people, stop, and some communication occurs
PeopleSplitUp	12.27 IpH	From two or more people, standing, sitting, or moving together, communicating, one or more people separate themselves and leave the frame

Evaluation Protocol & Scoring Process

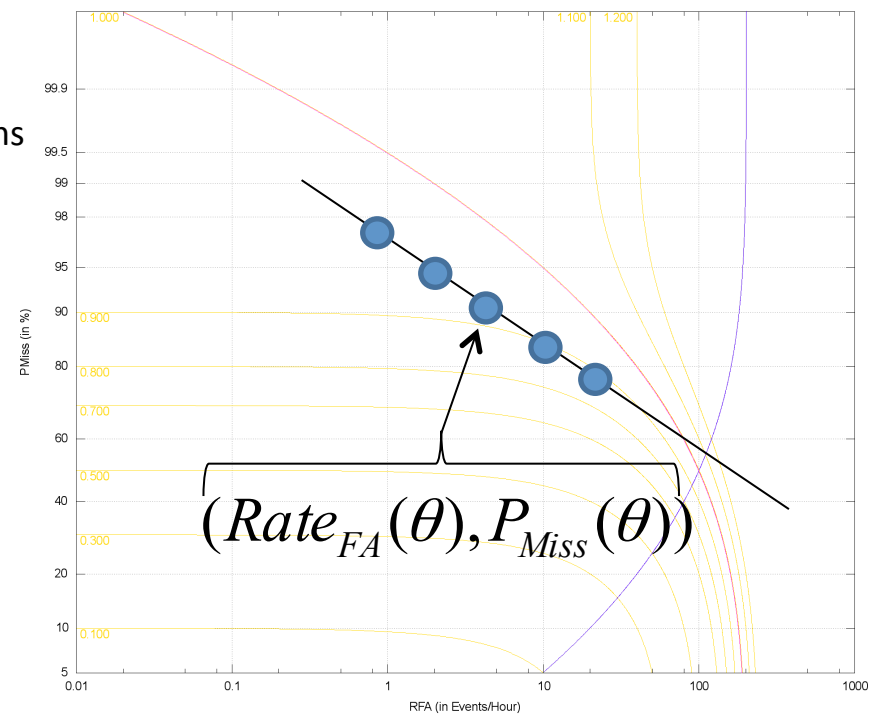
- Evaluation Plan
<http://www.nist.gov/itl/iad/mig/sed.cfm>
- Framework for Detection Evaluation (F4DE) Toolkit
<http://www.nist.gov/itl/iad/mig/tools.cfm>
- Four step evaluation process (for each event)
 1. Segment mapping
 2. Segment scoring
 3. Error metric calculation
 4. Error visualization

Step 4: Error Visualization

Detection Error Tradeoff (DET) Curves ($Prob_{Miss}$ vs. $Rate_{FA}$)



Compute $Rate_{FA}$ and P_{Miss} for all Θ



$$MinNDCR(\theta) = \operatorname{argmin}_{\theta} \left[P_{Miss}(\theta) + \frac{Cost_{FA}}{Cost_{Miss} * R_{TARGET}} * R_{FA}(\theta) \right]$$

$$ActNDCR(Act.Dec.) = P_{Miss}(Act.Dec.) + \frac{Cost_{FA}}{Cost_{Miss} * R_{TARGET}} * R_{FA}(Act.Dec.)$$

For more information about DETCurves: http://www.nist.gov/speech/publications/storage_paper/det.pdf

10 SED 2013 Participants

(with number of systems per event)

		Single Person		Person + object				Multiple People							
		PersonRuns		Pointing		CellToEar		ObjectPut		Embrace		PeopleMeet		PeopleSplit Up	
		iSED	rSED	iSED	rSED	iSED	rSED	iSED	rSED	iSED	rSED	iSED	rSED	iSED	rSED
6 years in a row	Carnegie Mellon University [CMU]	1	5	1	5	1		1	5	1	5	1	5	1	5
5 years in a row	Multimedia Communication and Pattern Recognition Labs, Beijing University of Posts and Telecommunications [BUPT-MCPRL]	1	2	1	2			1	2	1	5	1	5	1	5
2 years in a row	Brno University of Technology [BrnoUT]	1	2	1	2	1	2	1	2	1	2	1	2	1	2
	Dublin City University [dcu-savasa]	3	1	3	7	3	6	3	7						
	IBM Thomas J. Watson Research Center [IBM]	4	3	4	3	4	3	4	3	4	3	4	3	4	3
	The City College of New York Media Lab and SRI [CCNY-SRI]	2		2		2		2		2		2		2	
	Institute of Computer Science and Technology, Peking University [PKU-OS]	1	1	1	1			1	1						
	University of Ottawa(VIVA) and Polytechnique Montréal (LITIV) [VIVAuOttawaLITIVpoly]	1	2	1	2	1	2	1	2	1	2	1	2	1	2
NEW	AT&T Labs Research [ATT]	5		5		5		5		5		5		5	
	Beijing Institute of Technology [BIT]							2							
		19	16	19	22	17	13	21	22	15	17	15	17	15	17

Total Interactive Event Runs

122

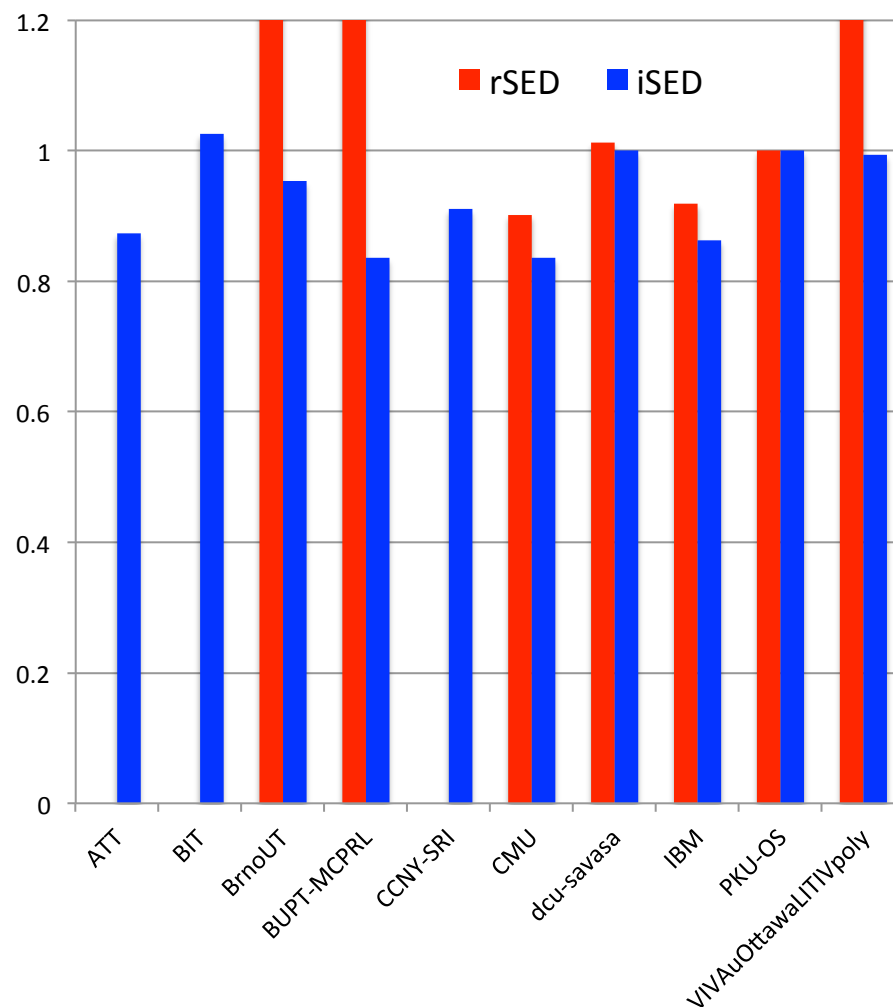
Total Retrospective Event Runs

126

Event-Averaged, Lowest Act NDCR by Site rSED vs. iSED

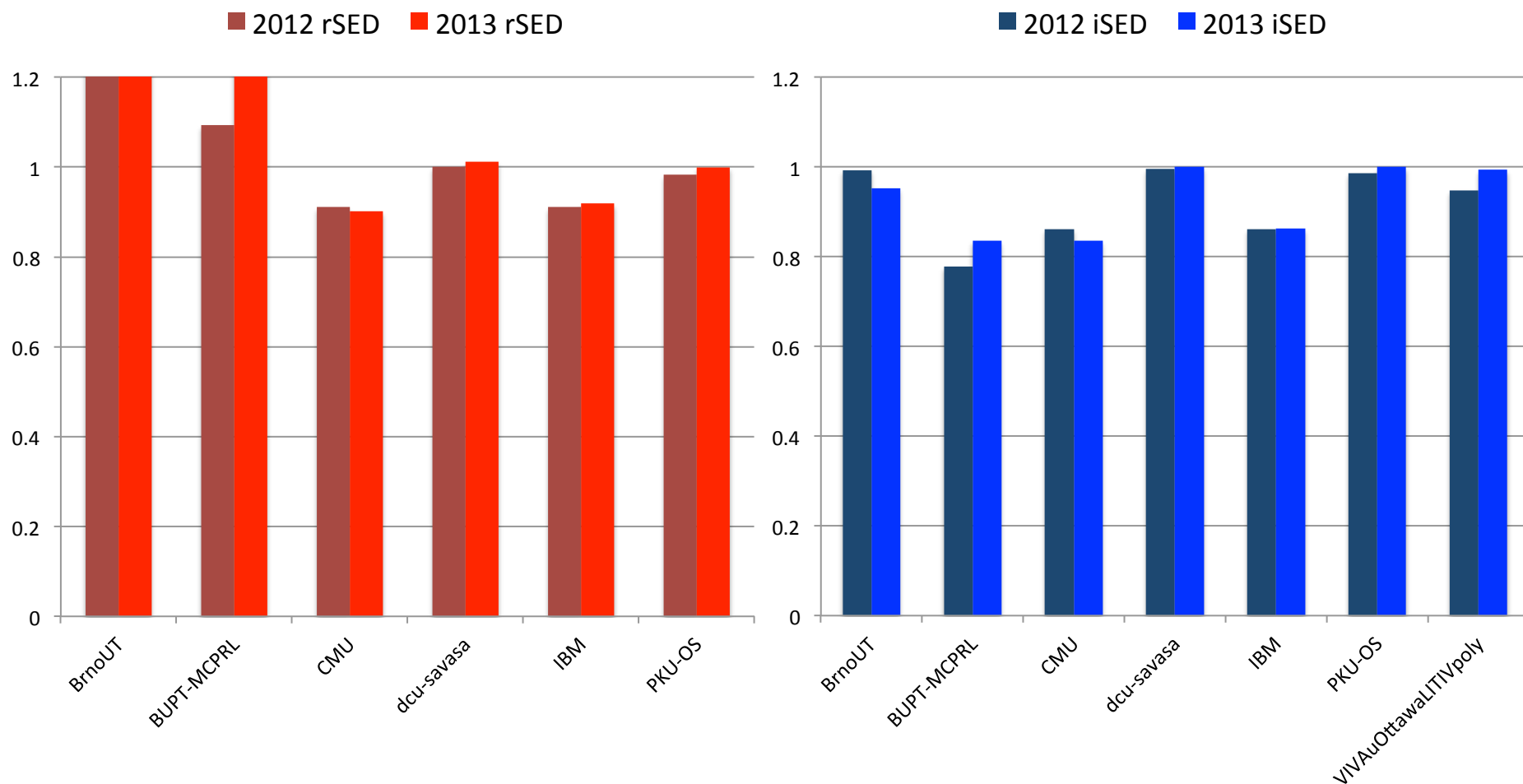
- 7 sites submitted both rSED and iSED runs
- 6 reduced Act.NDCR (from rSED to iSED)

BrnoUT	30%
BUPT-MCPRL	45%
CMU	7%
dcu-savasa	1%
IBM	6%
VIVAuOttawaLITIVpoly	84%



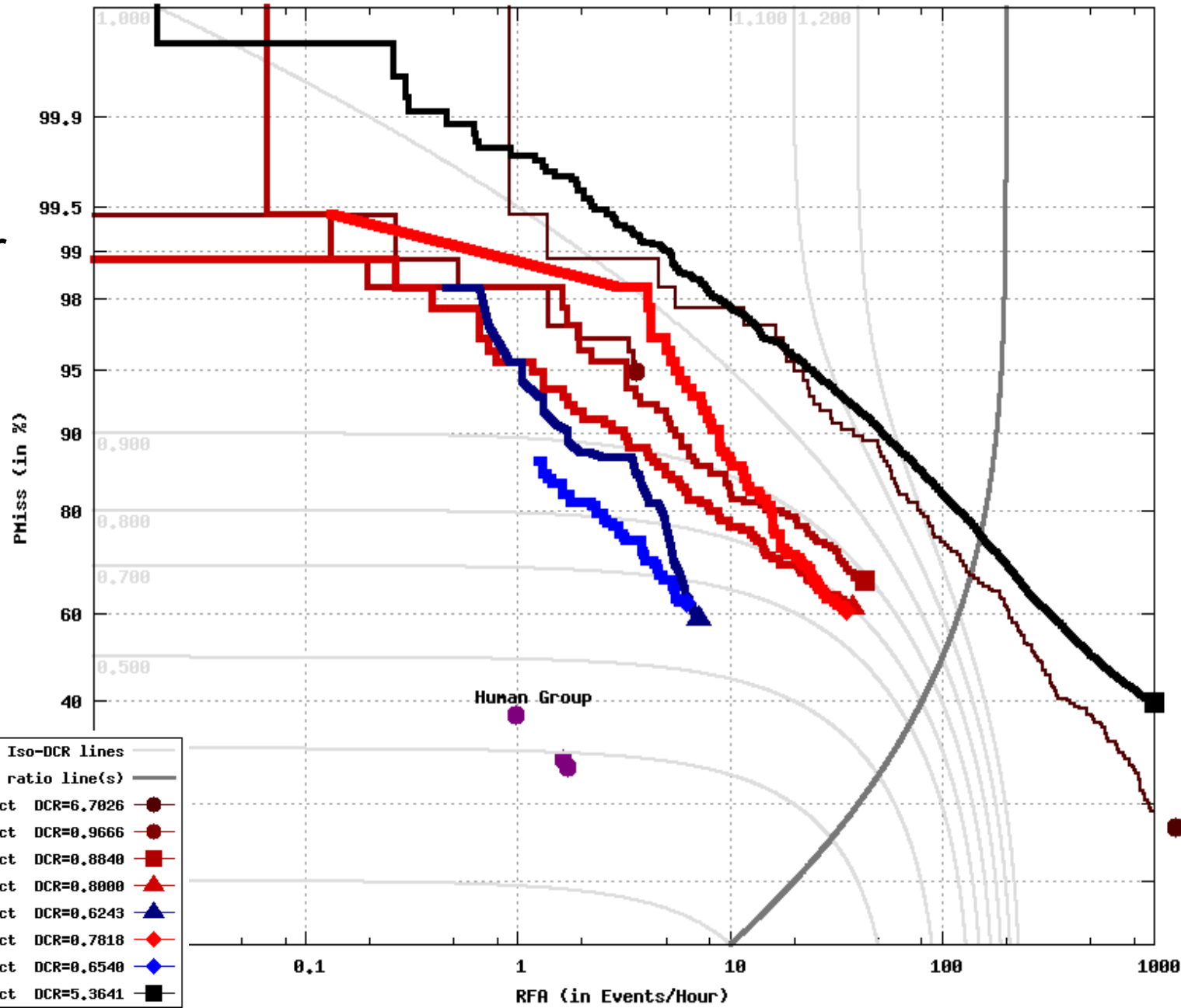
Event-Averaged, Lowest Act NDCR by Site

rSED & iSED: 2012 vs 2013



Embrace Event SED09-13

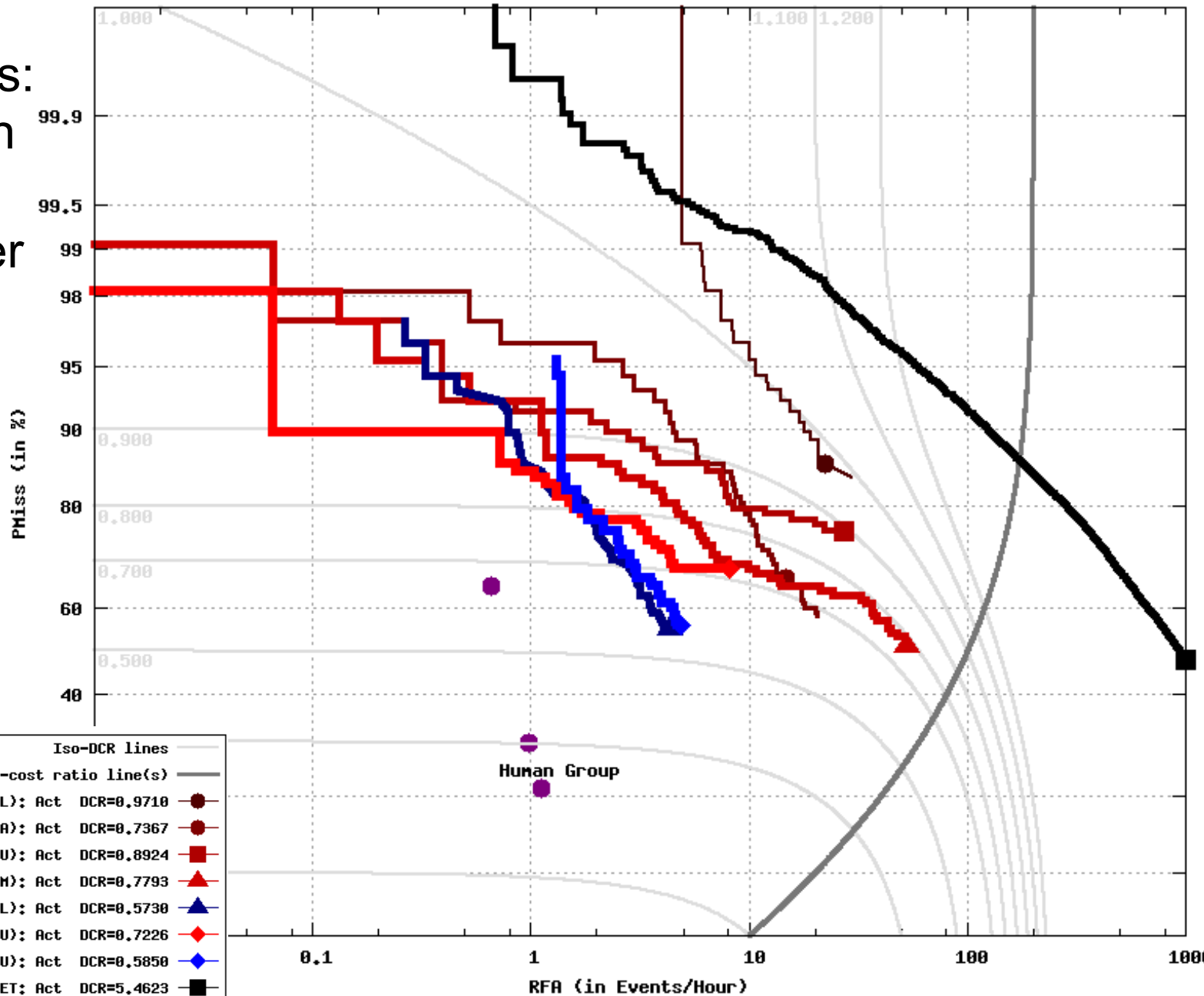
Embrace:
 Comparison
 of Best
 Systems per
 Year using
 Min Act
 NDCR



Iso-DCR lines	
Iso-cost ratio line(s)	
TV09 (CMU): Act	DCR=6.7026
TV10 (IPG-BJTU): Act	DCR=0.9666
TV11 (CMU): Act	DCR=0.8840
TV12-retroED (CHU-IBM): Act	DCR=0.8000
TV12-interactiveED (BUPT-MCPRL): Act	DCR=0.6243
TV13-retroED (IBM): Act	DCR=0.7818
TV13-interactiveED (CMU): Act	DCR=0.6540
RandomDET: Act	DCR=5.3641

PersonRuns Event SED09-13

PersonRuns:
 Comparison
 of Best
 Systems per
 Year using
 Min Act
 NDCR



System	Act	DCR
TV09 (NHKSTRL)	Act	0.9710
TV10 (QMUL-ACTIVA)	Act	0.7367
TV11 (CHU)	Act	0.8924
TV12-retroED (CMU-IBH)	Act	0.7793
TV12-interactiveED (BUPT-MCPRL)	Act	0.5730
TV13-retroED (CHU)	Act	0.7226
TV13-interactiveED (CHU)	Act	0.5850
RandomDET	Act	5.4623

Thoughts for the Future

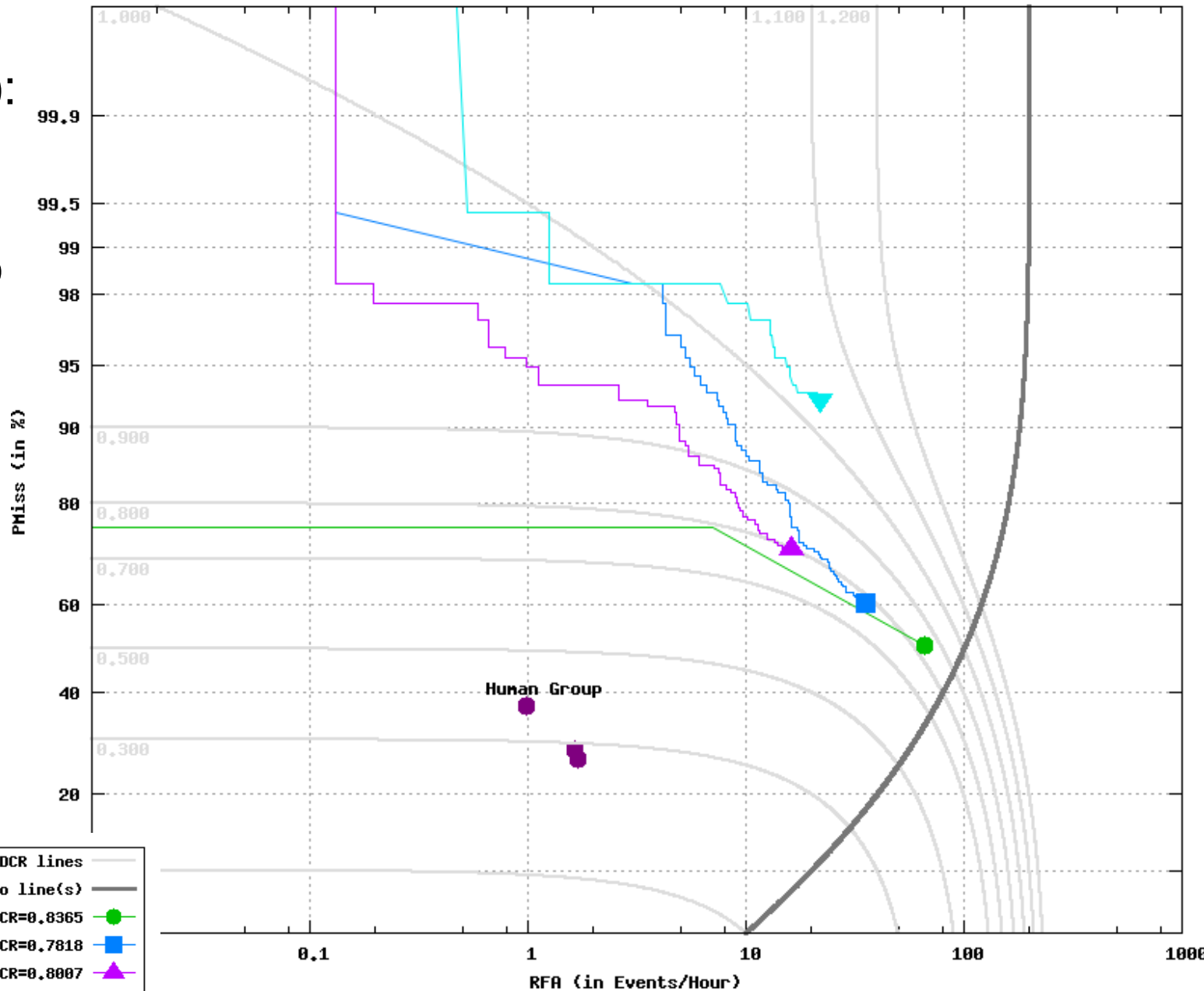
- SED data production is expensive
- Can system combination be used to reduce annotation costs
- Experiment:
 - Use previously developed adjudication infrastructure to combine system outputs
 - If accuracy improves, then it would be a better starting point for annotation

Adjudicating: Top 3

- Combine Top 3 Act NDCR scoring systems submissions per task/event/site (1 max per site)
- Iteratively build a combined system output by performing alignment/merge steps
 - Use the agreement level as the confidence score (3 agree = 1.00, 2 agree = 0.66, 1 agree = 0.33)

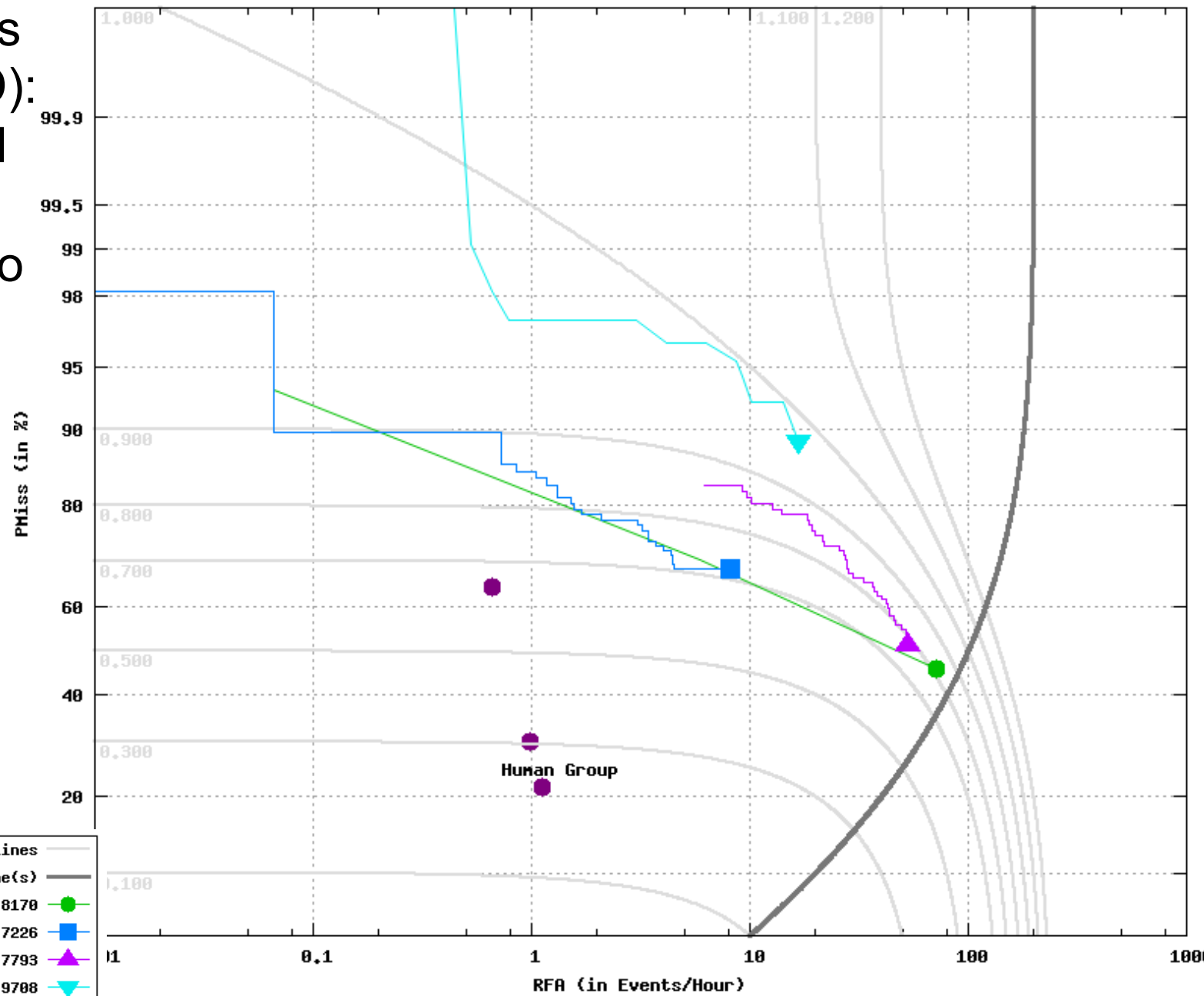
rSED Embrace Adjudication

Embrace
 (2013 rSED):
 Adjudicated
 system
 compared to
 Top 3
 systems



rSED PersonRuns Adjudication

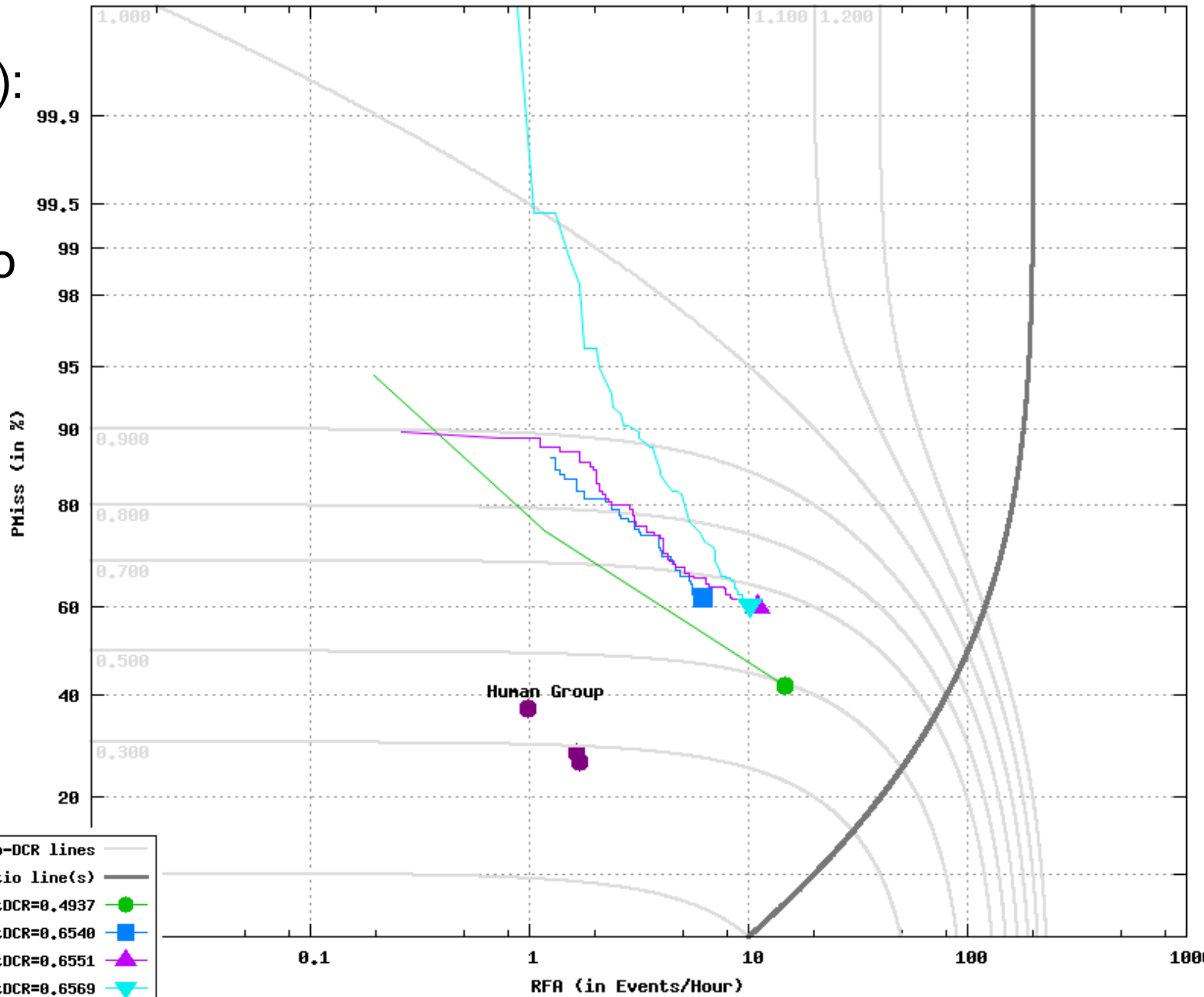
PersonRuns
 (2013 rSED):
 Adjudicated
 system
 compared to
 Top 3
 systems



Iso-DCR lines	—
Iso-cost ratio line(s)	—
Adjudicated : ActDCR=0.8178	●
CMU c-FV_2 : ActDCR=0.7226	■
IBM c-MC_1 : ActDCR=0.7793	▲
PKU-OS p-DTSYS_2 : ActDCR=0.9788	▼

iSED Embrace Adjudication

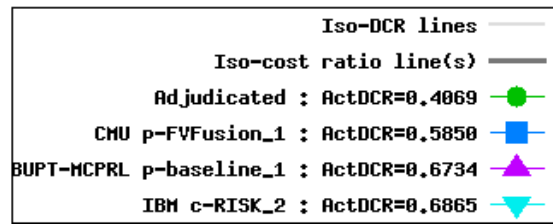
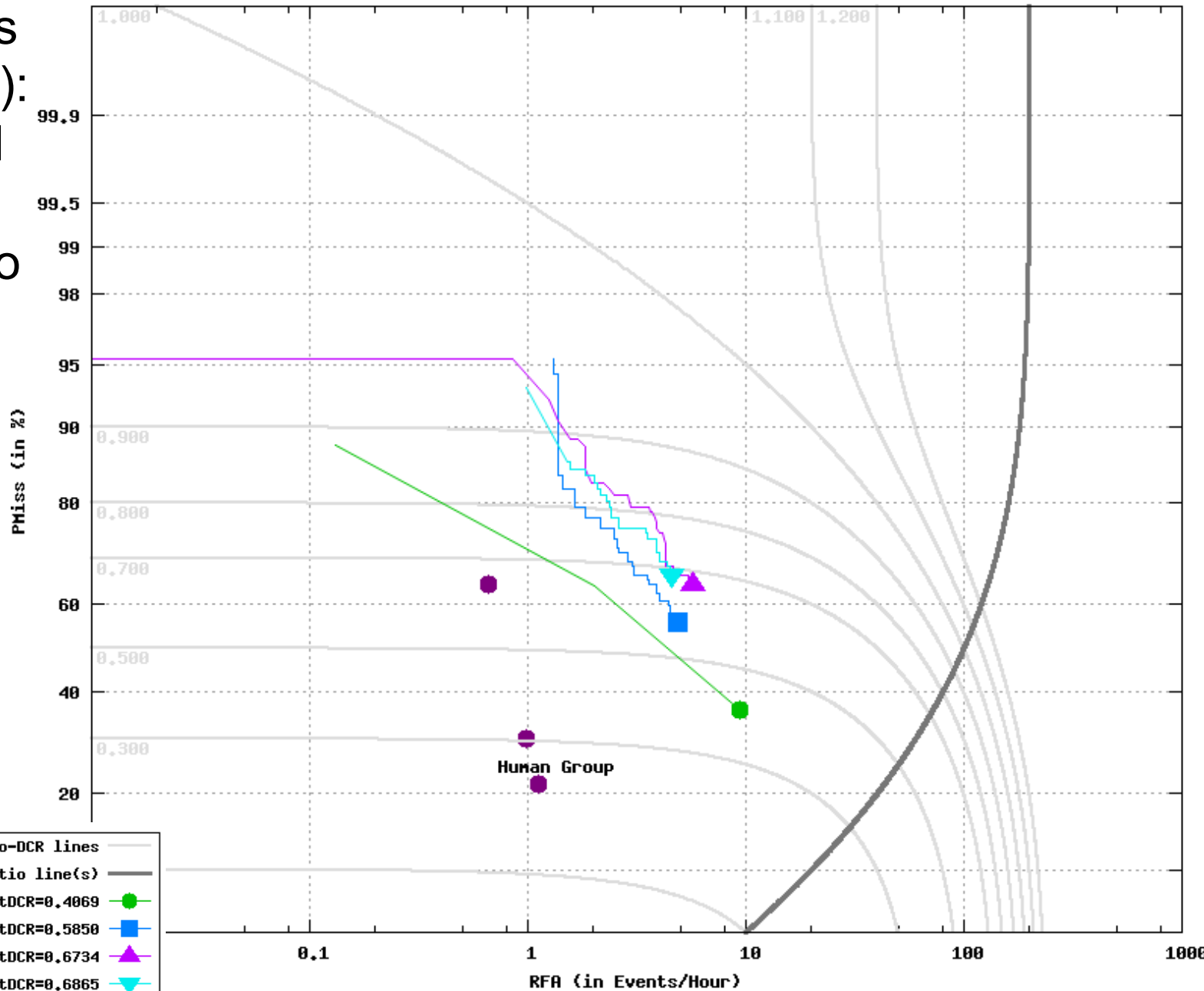
Embrace
 (2013 iSED):
 Adjudicated
 system
 compared to
 Top 3
 systems



- Iso-DCR lines —
- Iso-cost ratio line(s) —
- Adjudicated : ActDCR=0.4937 ●
- CHU p-FVFusion_1 : ActDCR=0.6548 ■
- IBM c-RISK_2 : ActDCR=0.6551 ▲
- BUPT-MCPRL p-baseline_1 : ActDCR=0.6569 ▼

iSED PersonRuns Adjudication

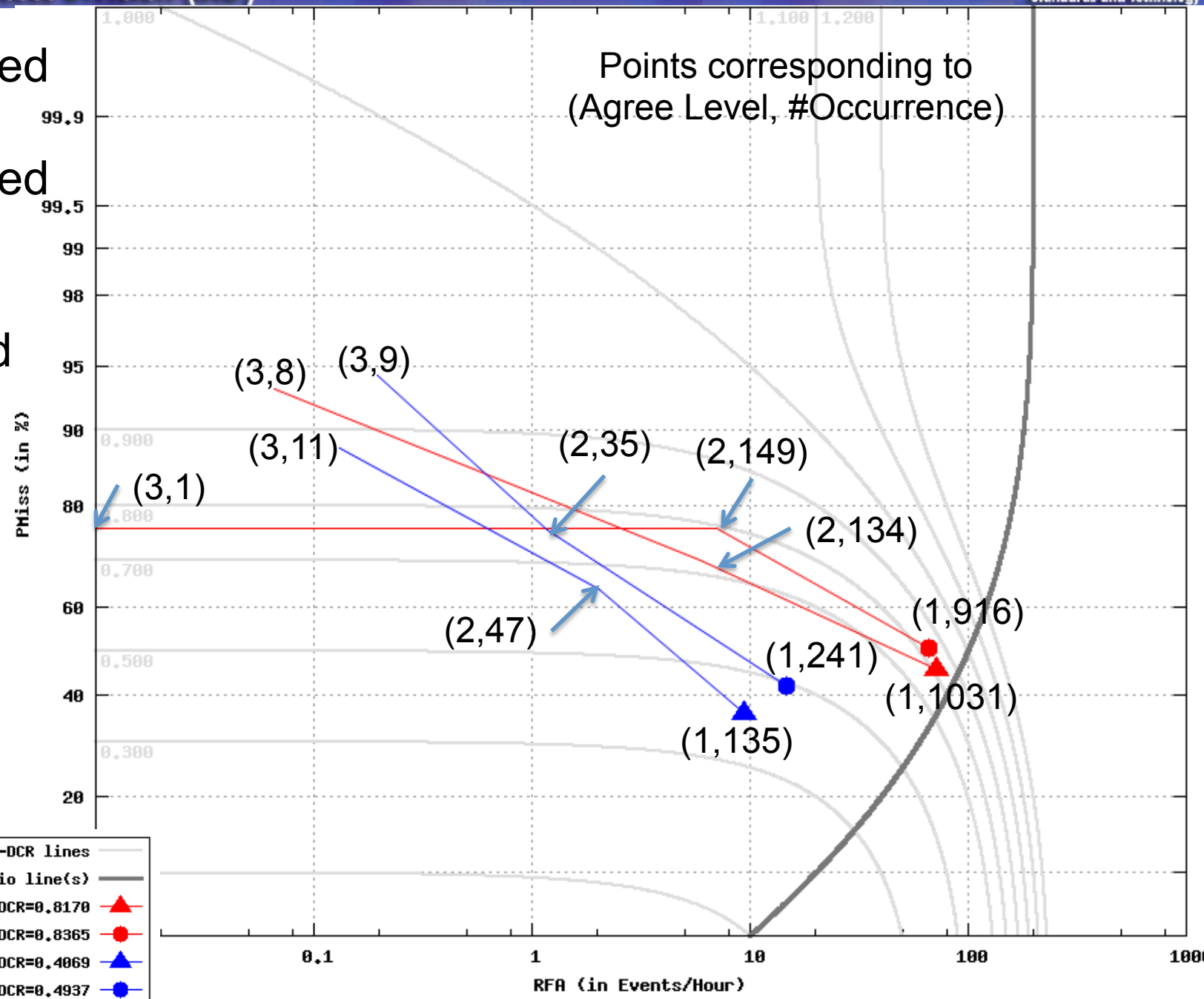
PersonRuns
 (2013 iSED):
 Adjudicated
 system
 compared to
 Top 3



Adjudication “Agree” reviews

Task	Event	Agree Level	#Occurrence	Total Time	Cumulative Time
rSED	Embrace	3	1	13.00s	4431.14s
		2	149	591.12s	
		1	936	3827.92s	
	PersonRuns	3	8	27.12s	2560.56s
		2	114	320.40s	
		1	1031	2213.04s	
iSED	Embrace	3	9	47.76s	1058.44s
		2	35	121.92s	
		1	241	888.76s	
	PersonRuns	3	11	26.98s	695.20s
		2	47	202.00s	
		1	133	466.22s	

Adjudicated
 iSED vs
 Adjudicated
 rSED
 systems
 compared



Iso-DCR lines	
PersonRuns rSED: Act	DCR=0,8170 ▲
Enbrace rSED: Act	DCR=0,8365 ●
PersonRuns iSED: Act	DCR=0,4069 ▲
Enbrace iSED: Act	DCR=0,4937 ●

Future of SED Evaluation

- Community driven REF generation a possibility
 - Participation requiring annotation of a selection of the reference video data on a set of events
 - Number of annotated hours depending on number of participants
 - Reference post extended by iSED post adjudication
- Use of an Automatic Scoring Server
 - requirement to test it during DryRun phase