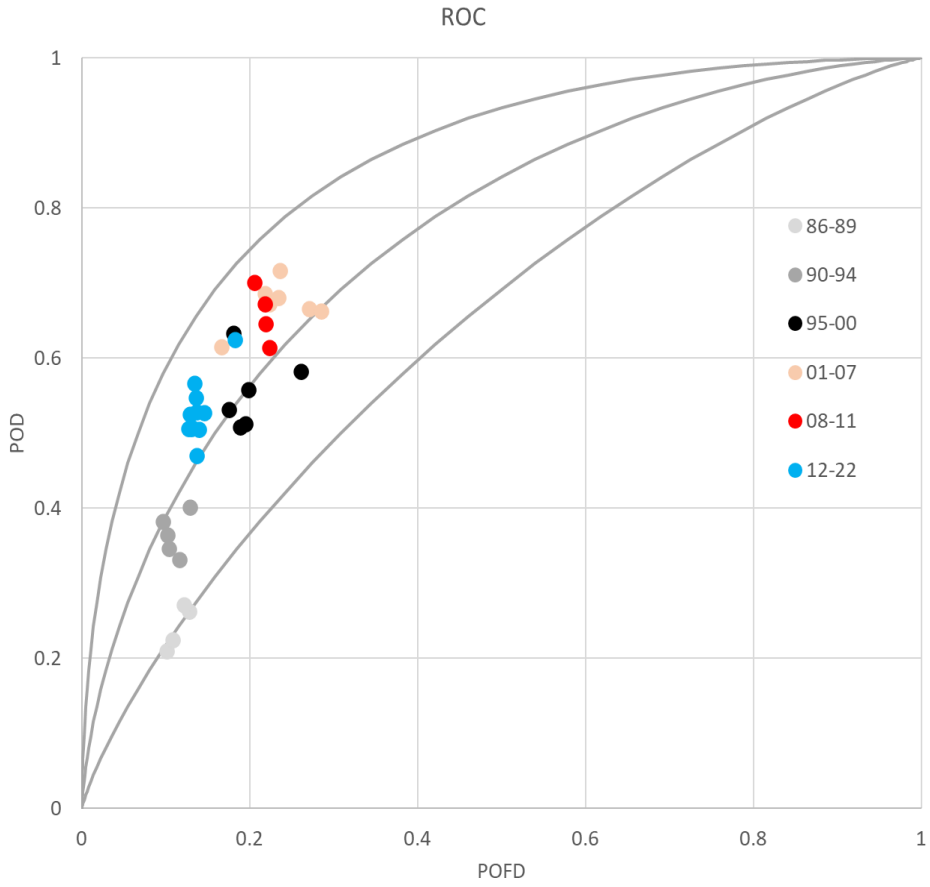
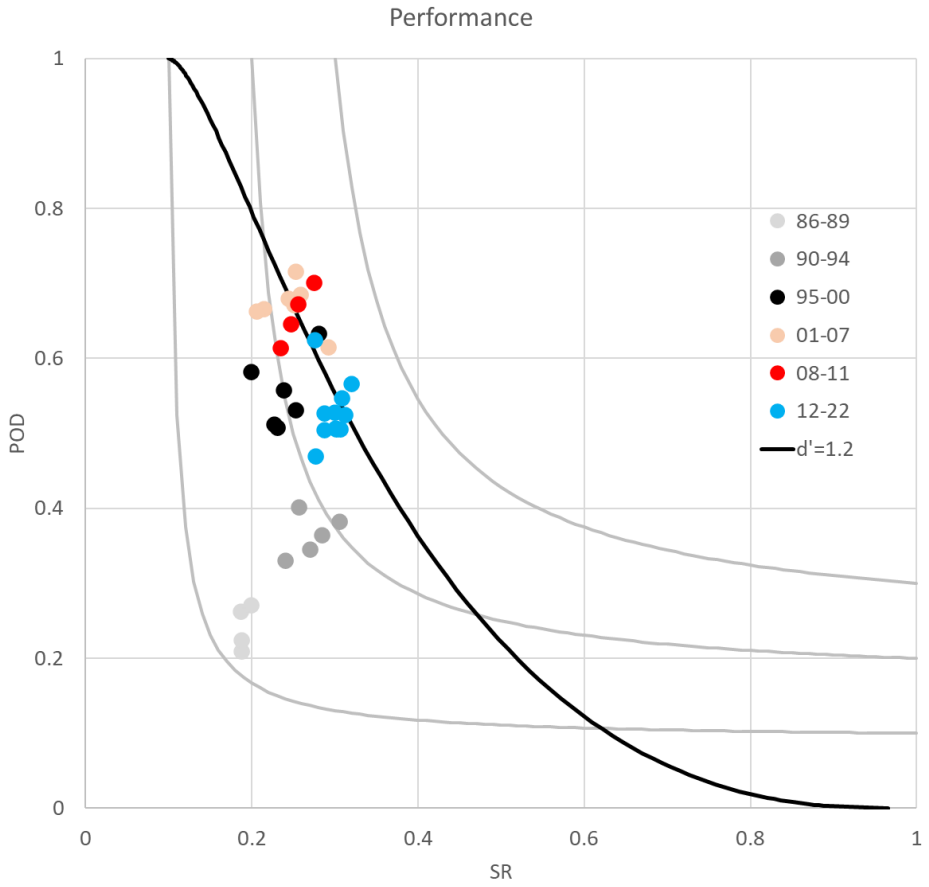
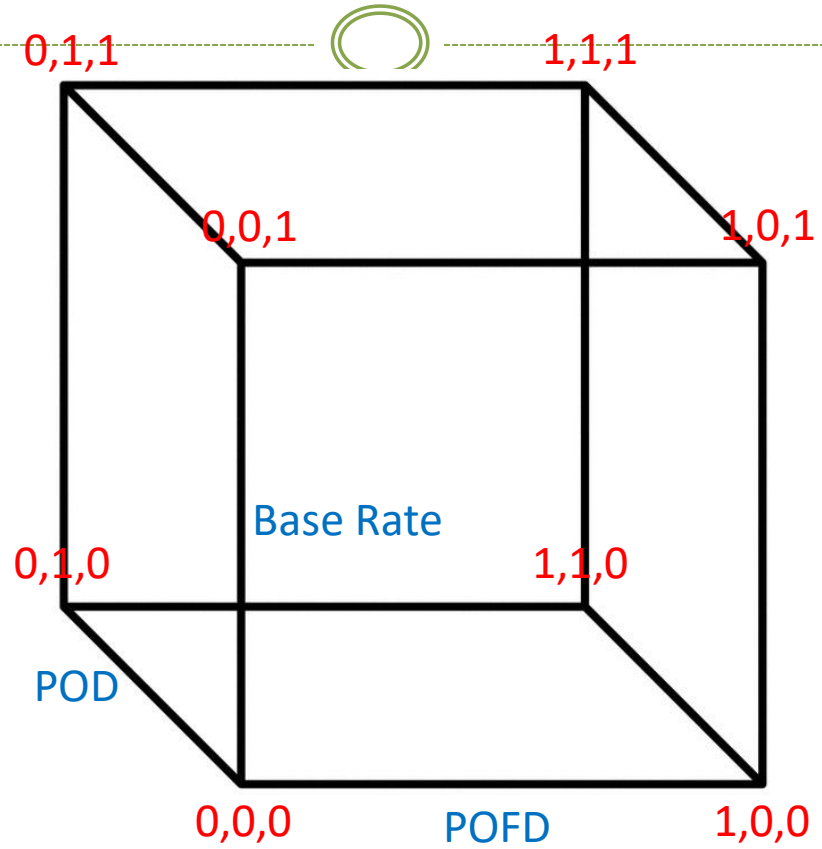
If a Picture's Worth a Thousand Words, Visualizing Aspects of Performance Might Be Worth a Chapter

HAROLD BROOKS
NOAA/NSSL
HAROLD.BROOKS@NOAA.GOV
@HEBROOKS87
WITH HELP FROM
MICHAEL BALDWIN, MONTE FLORA, PAT SKINNER, CHRIS KERR, JEFF MILNE


ROC
POD
POFD
0
0.2
0.4
0.6
0.8
1
86-89
90-94
95-00
01-07
08-11
12-22


Performance
POD
SR
0
0.2
0.4
0.6
0.8
1
86-89
90-94
95-00
01-07
08-11
12-22
d'=1.2

Unit cube of evaluation



The 2x2 table

Forecast	Event: Y	Event: N
Y	a	b
N	c	d

- Base Rate = a+c
- Prob of detection (POD)=a/(a+c)
- Prob of false detection (POFD)=b/(b+d)
- False alarm ratio (FAR)=b/(a+b)
- Success ratio (SR)=a/(a+b)
- Bias (B) = (a+b)/(a+c)=SR/POD

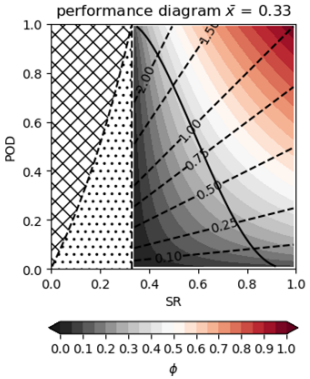
- Critical success index (CSI)=a/(a+b+c)
  - CSI=HM[POD,SR]/(2-HM[POD,SR])
- $\phi = \frac{ad-bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$
- $F_\beta = (1+\beta^2)\frac{(SR)(POD)}{(\beta^2 SR+POD)}$
  - $\beta$is weight of importance of POD to SR
  - Equivalent toTversky index, related to GM

Simple Decision Rule

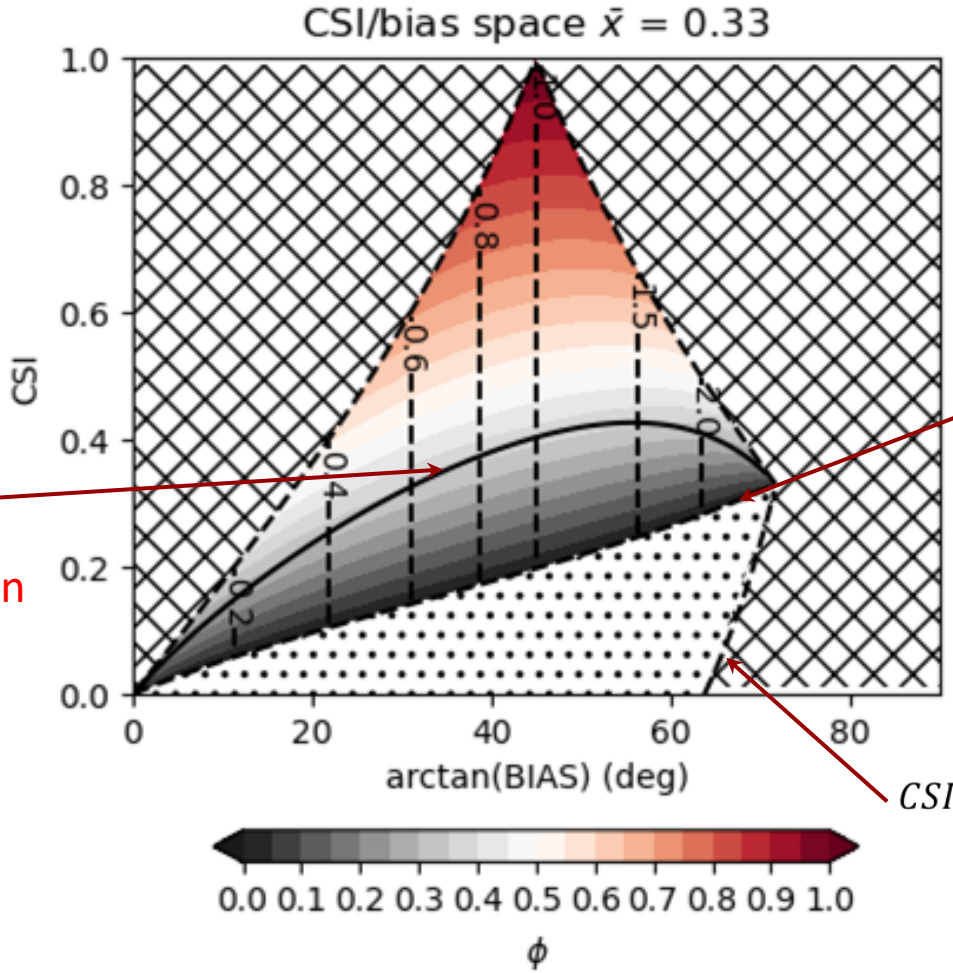
		Event	
		Y	N
Forecast	Y	A	B
	N	C	D

Misclassification Cost Ratio (α). Act if $p > \frac{(B-D)}{(B-D)+(C-A)} = \frac{Cost(FA)}{Cost(FA)+Cost(ME)}$

$$\frac{Cost(FA)}{Cost(ME)} = \frac{\Delta POD}{\Delta POFD}\frac{\bar{x}}{1-\bar{x}}$$



$\bar{x} = 0.33$



$d' = 1$

$CSI_{no\ skill} = \frac{B\bar{x}}{1+B-B\bar{x}}$

$CSI_{min} = \bar{x}(B + 1) - 1$

Forecast Less Often

Forecast More Often

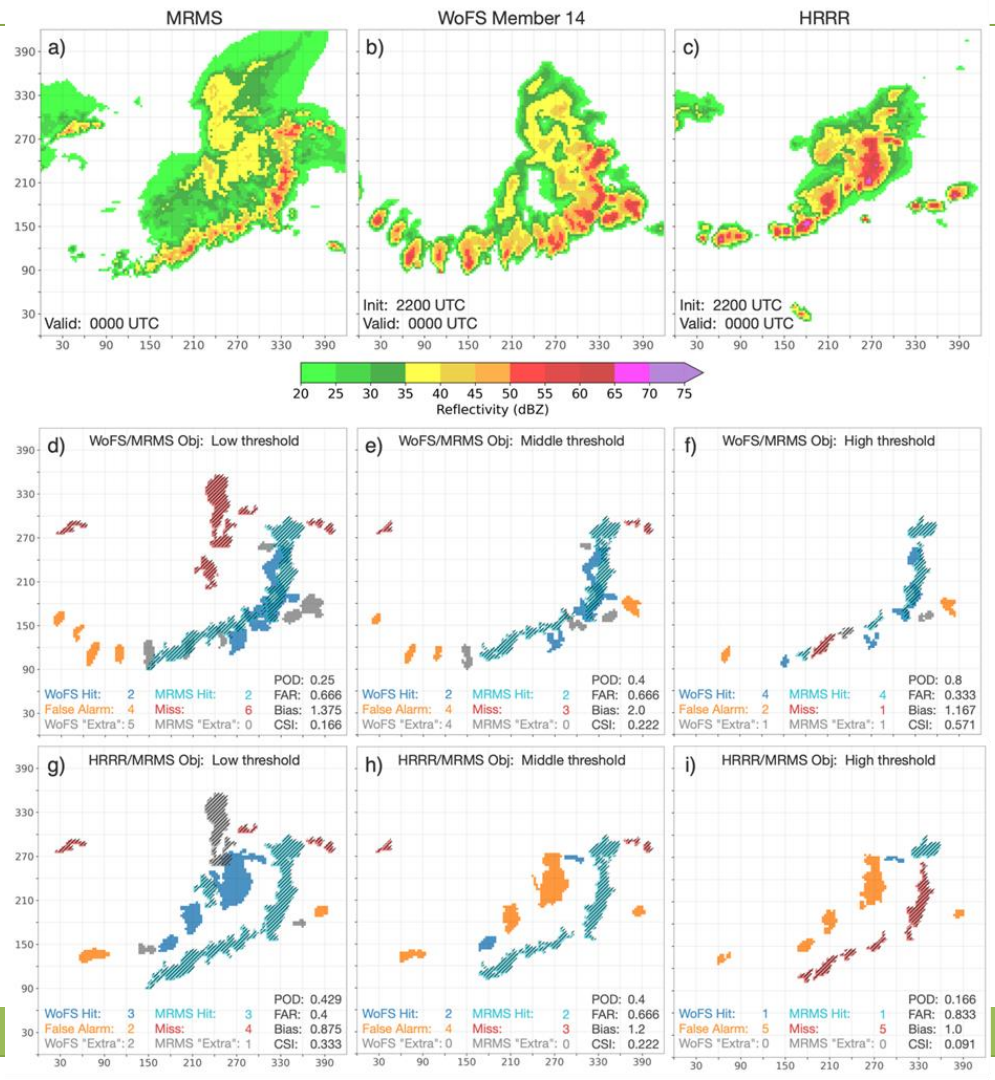
Our problem of interest

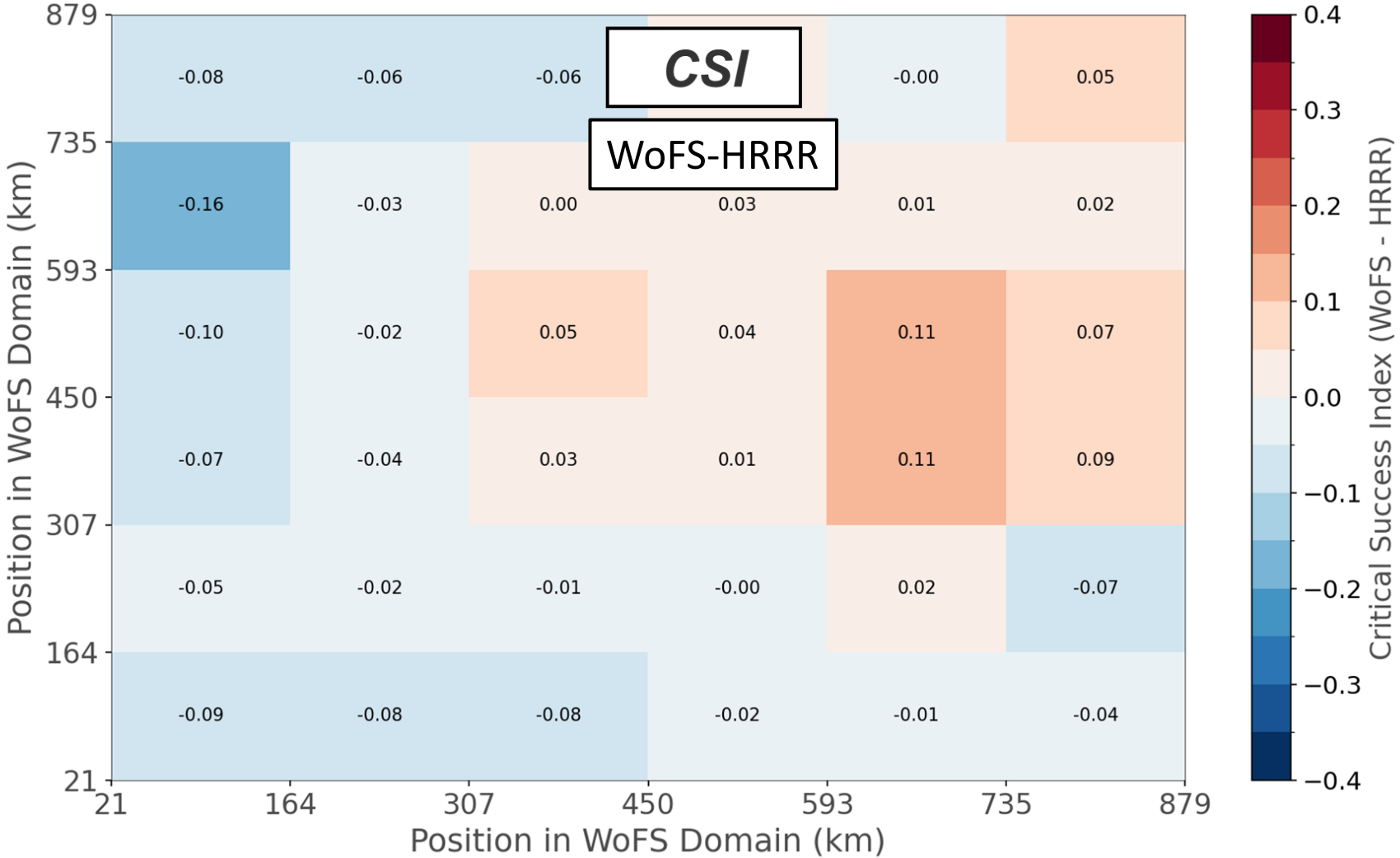
- Forecasting for severe thunderstorms, tornadoes ~0.5-6 h
- Proxies for truth
- Can have deterministic, probabilistic systems

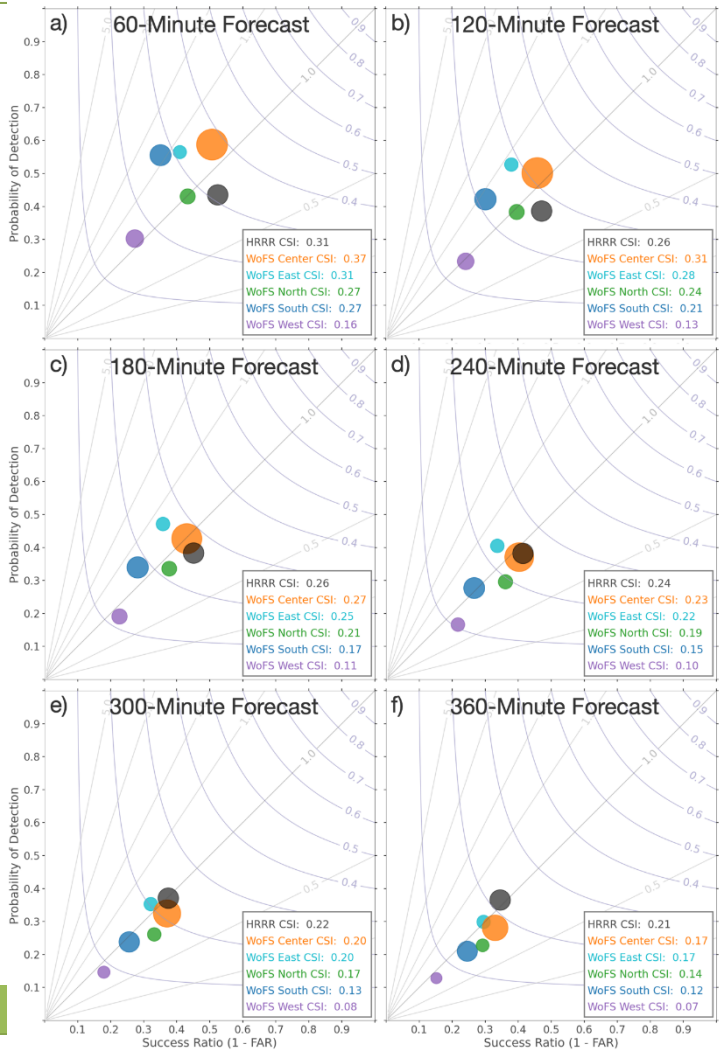
- Objects from Multi-Radar Multi-Sensor (MRMS)

Model description

- HRRR (3 km)
  - Covers contiguous US (1799x1059 grid points)
  - Runs out to 48 hours at 0000 and 1200 UTC
  - Hourly data assimilation
- WOFS (3 km)
  - Regional, moves around (300x300 grid points)
  - Uses HRRR 1200 UTC forecasts as initialization
  - Assimilates radar and satellite data not in HRRR every 15 minutes
  - 0-6 hr, available every 5 minutes, starting at 1700 UTC


MRMS
WoFS Member 14
HRRR
a)
b)
c)
Valid: 0000 UTC
Init: 2200 UTC
Valid: 0000 UTC
Init: 2200 UTC
Valid: 0000 UTC
20 25 30 35 40 45 50 55 60 65 70 75
Reflectivity (dBZ)
d) WoFS/MRMS Obj: Low threshold
WoFS Hit: 2 MRMS Hit: 2 POD: 0.25
False Alarm: 4 Miss: 6 FAR: 0.666
WoFS "Extra": 5 MRMS "Extra": 0 Bias: 1.375
CSI: 0.166
e) WoFS/MRMS Obj: Middle threshold
WoFS Hit: 2 MRMS Hit: 2 POD: 0.4
False Alarm: 4 Miss: 3 FAR: 0.666
WoFS "Extra": 4 MRMS "Extra": 0 Bias: 2.0
CSI: 0.222
f) WoFS/MRMS Obj: High threshold
WoFS Hit: 4 MRMS Hit: 4 POD: 0.8
False Alarm: 2 Miss: 1 FAR: 0.333
WoFS "Extra": 1 MRMS "Extra": 1 Bias: 1.167
CSI: 0.571
g) HRRR/MRMS Obj: Low threshold
WoFS Hit: 3 MRMS Hit: 3 POD: 0.429
False Alarm: 2 Miss: 4 FAR: 0.4
WoFS "Extra": 2 MRMS "Extra": 1 Bias: 0.875
CSI: 0.333
h) HRRR/MRMS Obj: Middle threshold
WoFS Hit: 2 MRMS Hit: 2 POD: 0.4
False Alarm: 4 Miss: 3 FAR: 0.666
WoFS "Extra": 0 MRMS "Extra": 0 Bias: 1.2
CSI: 0.222
i) HRRR/MRMS Obj: High threshold
WoFS Hit: 1 MRMS Hit: 1 POD: 0.166
False Alarm: 5 Miss: 5 FAR: 0.833
WoFS "Extra": 0 MRMS "Extra": 0 Bias: 1.0
CSI: 0.091


CSI
WoFS-HRRR
-0.08 -0.06 -0.06 -0.00 0.05
-0.16 -0.03 0.00 0.03 0.01 0.02
-0.10 -0.02 0.05 0.04 0.11 0.07
-0.07 -0.04 0.03 0.01 0.11 0.09
-0.05 -0.02 -0.01 -0.00 0.02 -0.07
-0.09 -0.08 -0.08 -0.02 -0.01 -0.04
879
735
593
450
307
164
21
Position in WoFS Domain (km)
21 164 307 450 593 735 879
Position in WoFS Domain (km)
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
Critical Success Index (WoFS - HRRR)


a) 60-Minute Forecast
HRRR CSI: 0.31
WoFS Center CSI: 0.37
WoFS East CSI: 0.31
WoFS North CSI: 0.27
WoFS South CSI: 0.27
WoFS West CSI: 0.16
b) 120-Minute Forecast
HRRR CSI: 0.26
WoFS Center CSI: 0.31
WoFS East CSI: 0.28
WoFS North CSI: 0.24
WoFS South CSI: 0.21
WoFS West CSI: 0.13
c) 180-Minute Forecast
HRRR CSI: 0.26
WoFS Center CSI: 0.27
WoFS East CSI: 0.25
WoFS North CSI: 0.21
WoFS South CSI: 0.17
WoFS West CSI: 0.11
d) 240-Minute Forecast
HRRR CSI: 0.24
WoFS Center CSI: 0.23
WoFS East CSI: 0.22
WoFS North CSI: 0.19
WoFS South CSI: 0.15
WoFS West CSI: 0.10
e) 300-Minute Forecast
HRRR CSI: 0.22
WoFS Center CSI: 0.20
WoFS East CSI: 0.20
WoFS North CSI: 0.17
WoFS South CSI: 0.13
WoFS West CSI: 0.08
f) 360-Minute Forecast
HRRR CSI: 0.21
WoFS Center CSI: 0.17
WoFS East CSI: 0.17
WoFS North CSI: 0.14
WoFS South CSI: 0.12
WoFS West CSI: 0.07
Probability of Detection
Success Ratio (1 - FAR)

Comparing 1 and 3 km WOFS

- 1 km has smaller domain
- Compare only if both 1 and 3 have objects
- 30 minutes-3 hours, every 30 minutes

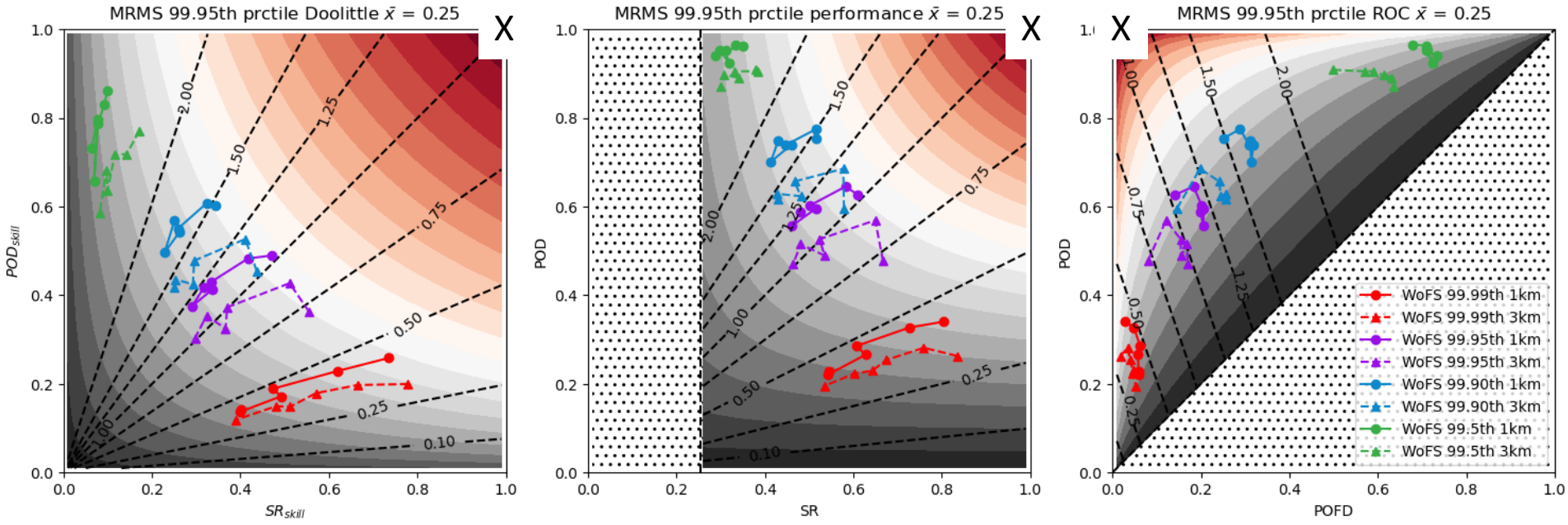
Skill scores based on ROC, performance diagram elements

$$POD_{skill} = \frac{POD - POD_{chance}}{POD_{perfect} - POD_{chance}} = \frac{ad - bc}{(a + c)(c + d)}$$

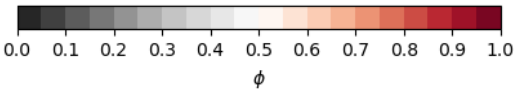
$$SR_{skill} = \frac{SR - SR_{chance}}{SR_{perfect} - SR_{chance}} = \frac{ad - bc}{(a + b)(b + d)}$$

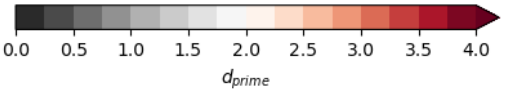
$$POFD_{skill} = \frac{POFD - POFD_{chance}}{POFD_{perfect} - POFD_{chance}} = \frac{ad - bc}{(a + b)(b + d)} = SR_{skill}$$

- Doolittle (1885), Kraemer (1988)

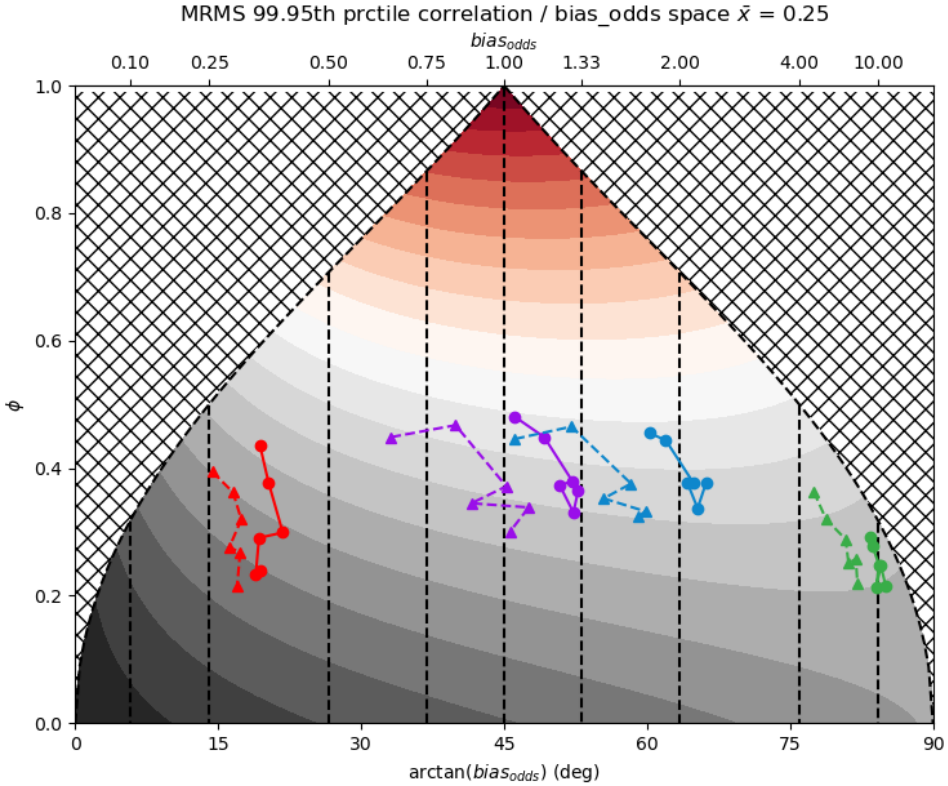


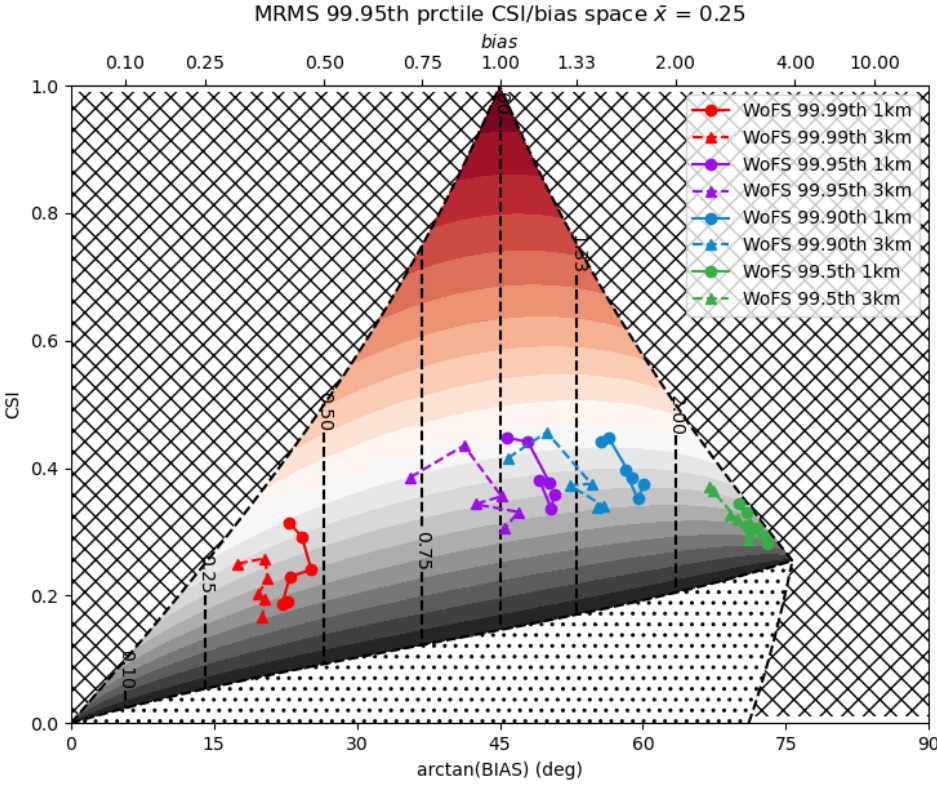
Solid-1 km, Dashed-3 km, X is perfect

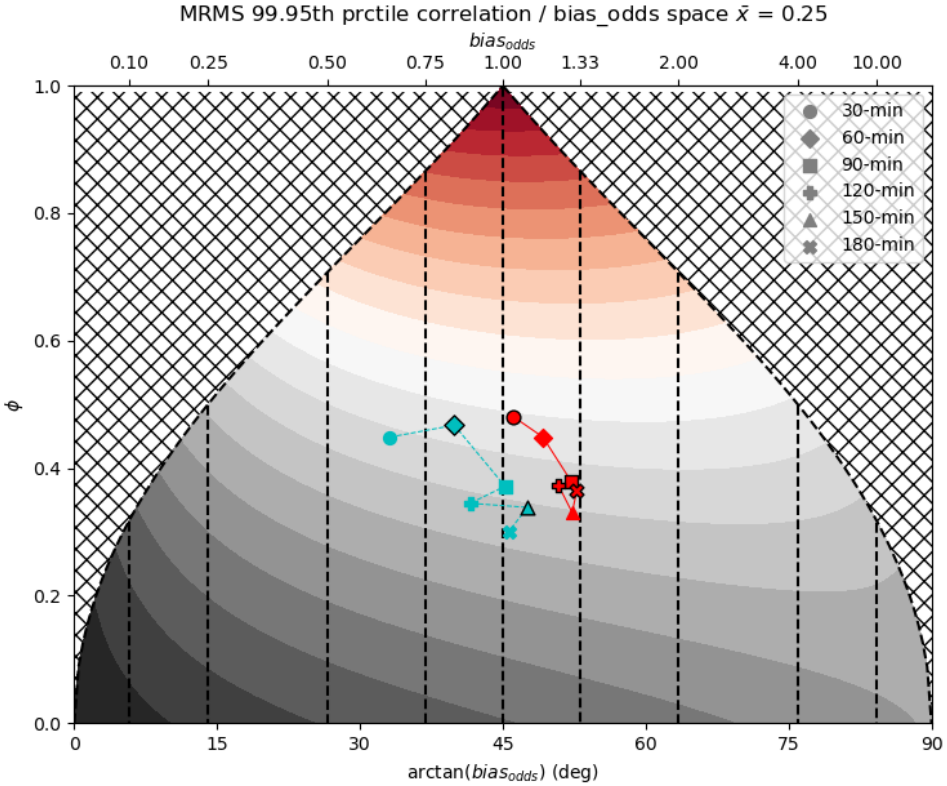


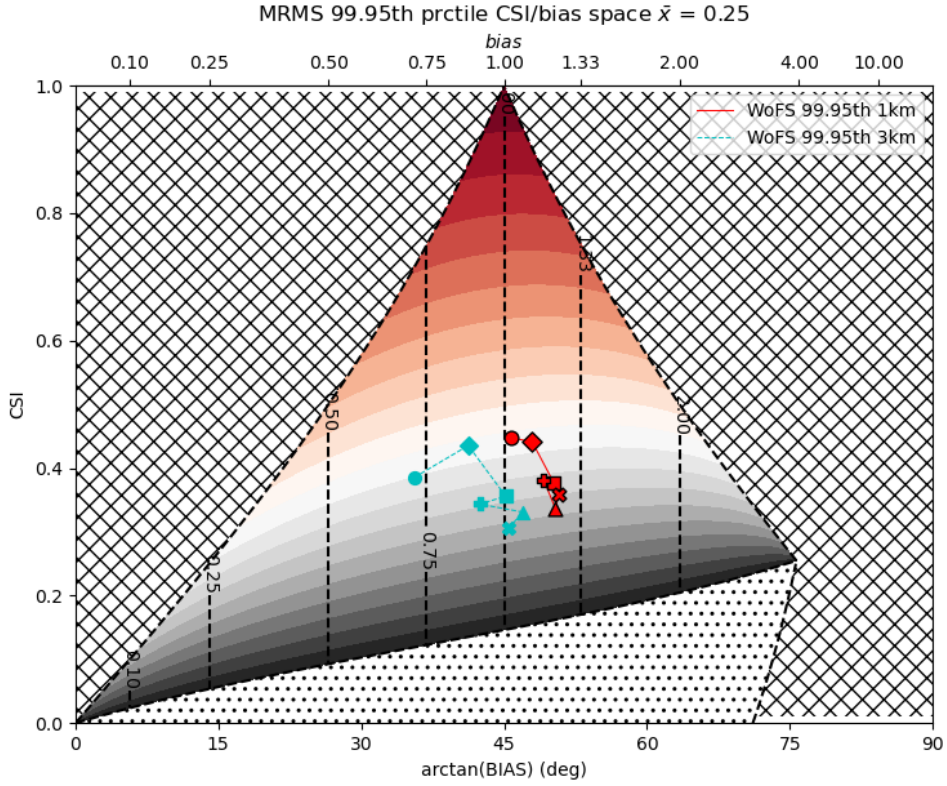


Lines are same forecast threshold at different lead times (old-away from perfect)

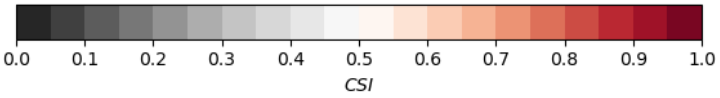

MRMS 99.95th prctile correlation / bias_odds space $\bar{x}$ = 0.25
$bias_{odds}$
0.10 0.25 0.50 0.75 1.00 1.33 2.00 4.00 10.00
ϕ
arctan($bias_{odds}$) (deg)
0.0 0.2 0.4 0.6 0.8 1.0
0 15 30 45 60 75 90
CSI
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

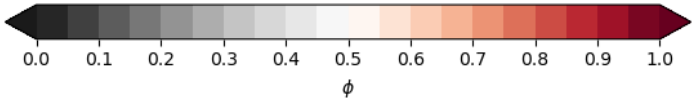

MRMS 99.95th prctile CSI/bias space $\bar{x}$ = 0.25
bias
0.10 0.25 0.50 0.75 1.00 1.33 2.00 4.00 10.00
CSI
arctan(BIAS) (deg)
0.0 0.2 0.4 0.6 0.8 1.0
0 15 30 45 60 75 90
WoFS 99.99th 1km
WoFS 99.99th 3km
WoFS 99.95th 1km
WoFS 99.95th 3km
WoFS 99.90th 1km
WoFS 99.90th 3km
WoFS 99.5th 1km
WoFS 99.5th 3km
ϕ
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

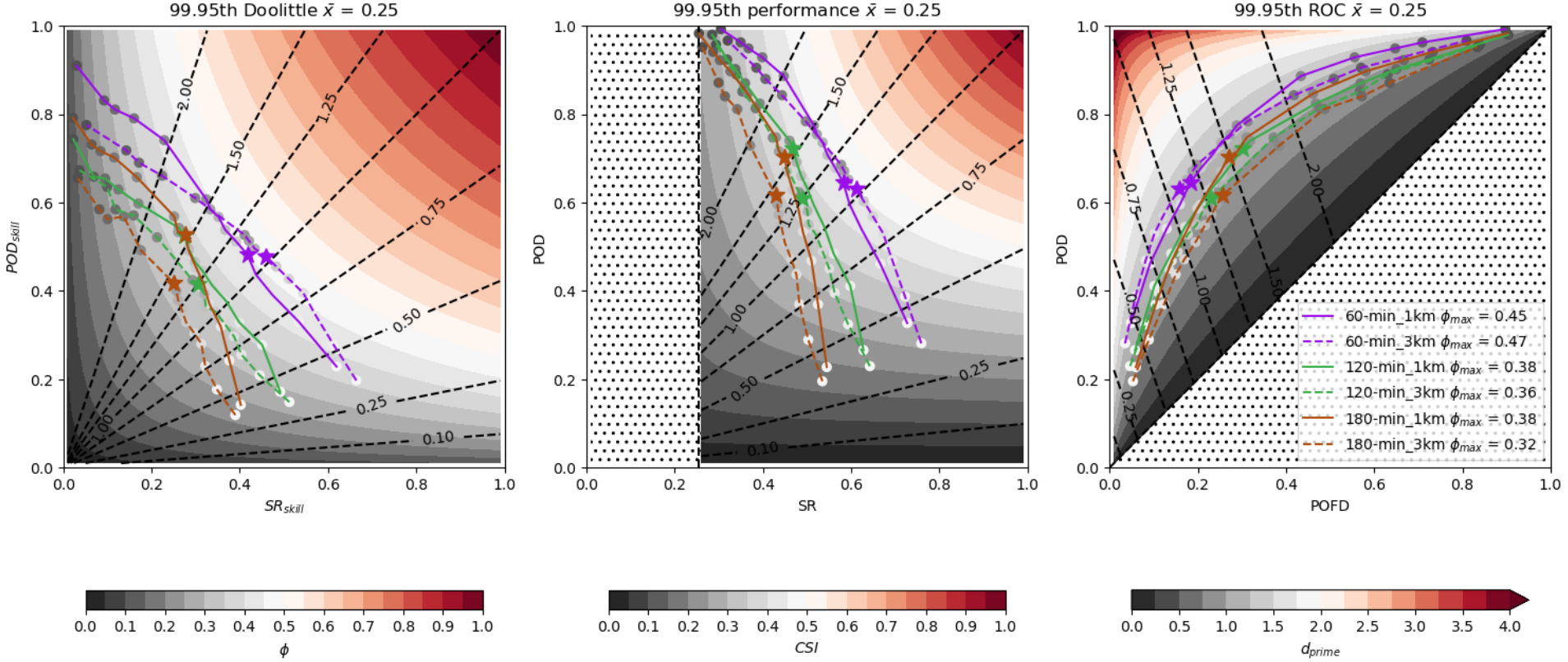


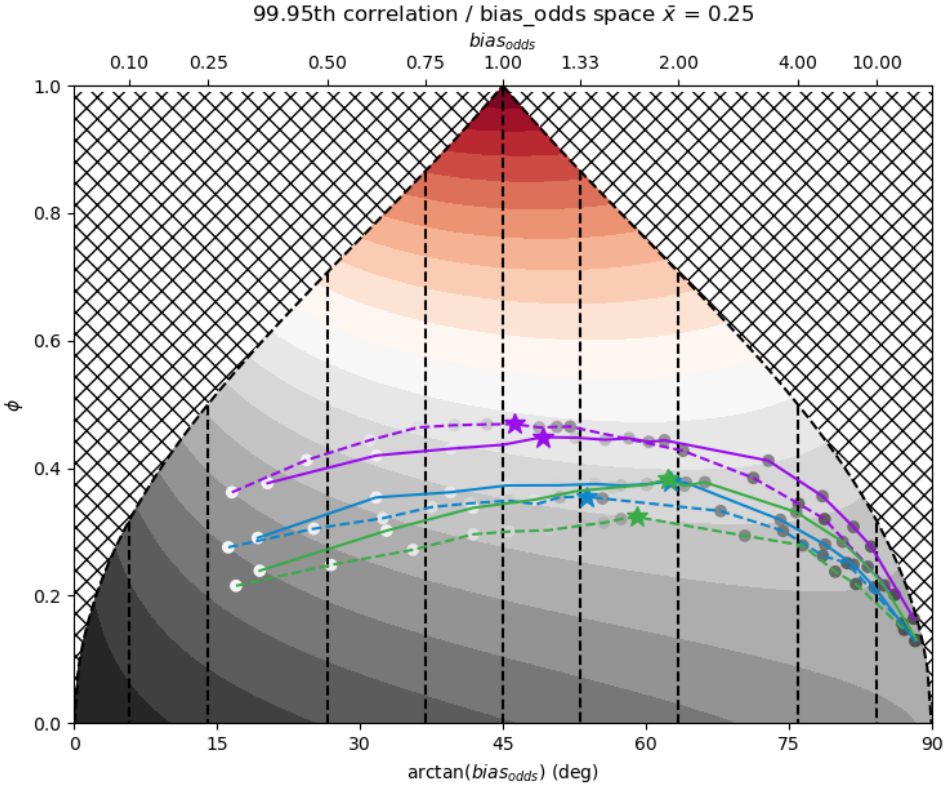


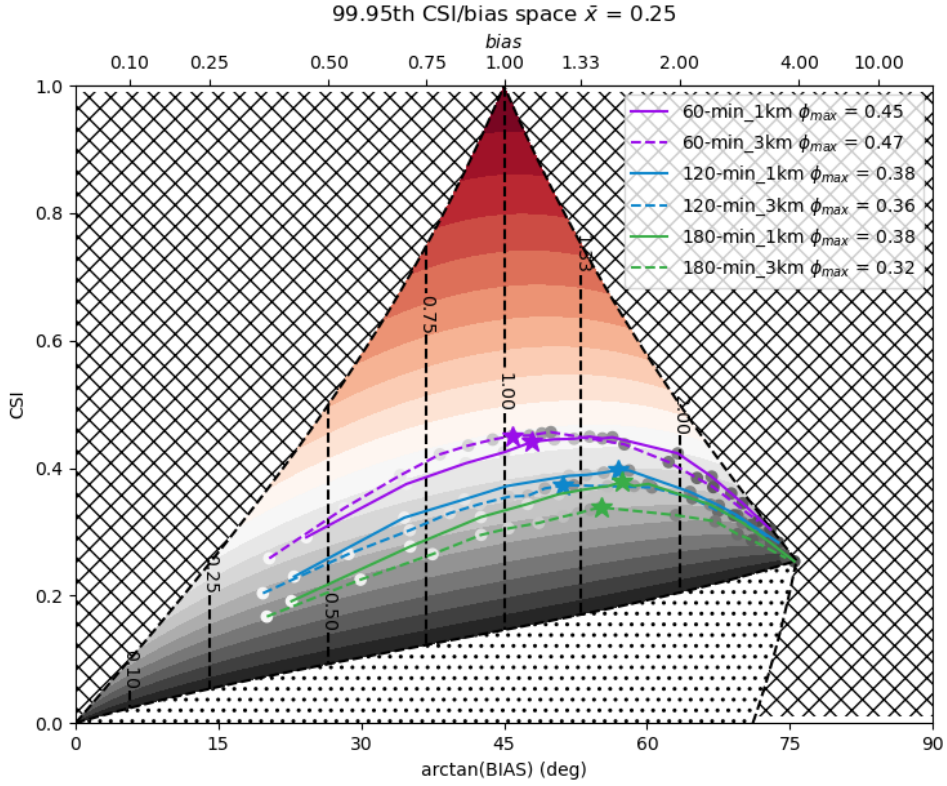
Using same forecast threshold as observed



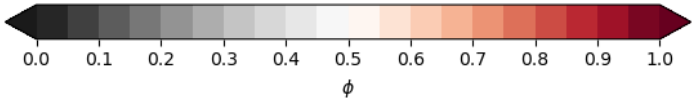


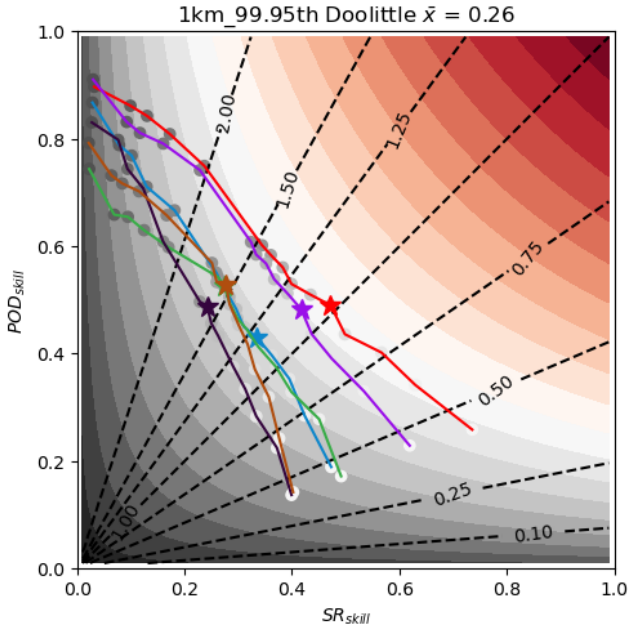

99.95th Doolittle $\bar{x}$ = 0.25
99.95th performance $\bar{x}$ = 0.25
99.95th ROC $\bar{x}$ = 0.25
POD_{skill}
SR_{skill}
POD
SR
POFD
2.00
1.50
1.25
0.75
0.50
0.25
0.10
1.00
60-min_1km ϕ_{max} = 0.45
60-min_3km ϕ_{max} = 0.47
120-min_1km ϕ_{max} = 0.38
120-min_3km ϕ_{max} = 0.36
180-min_1km ϕ_{max} = 0.38
180-min_3km ϕ_{max} = 0.32
ϕ
CSI
d_{prime}

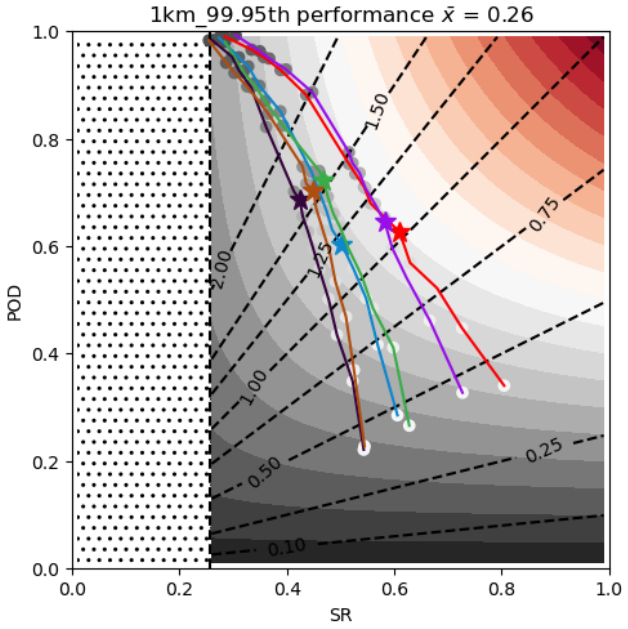


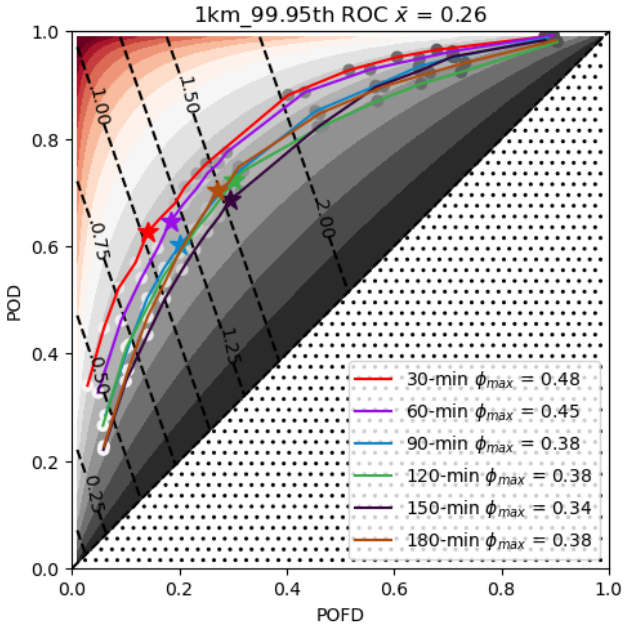


All forecast thresholds at one time

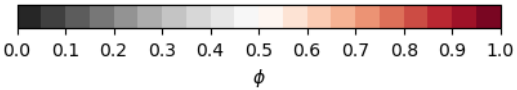


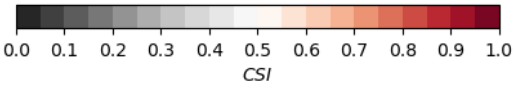


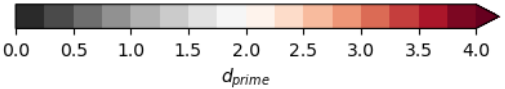




Only 1 km forecasts, every 30 minutes



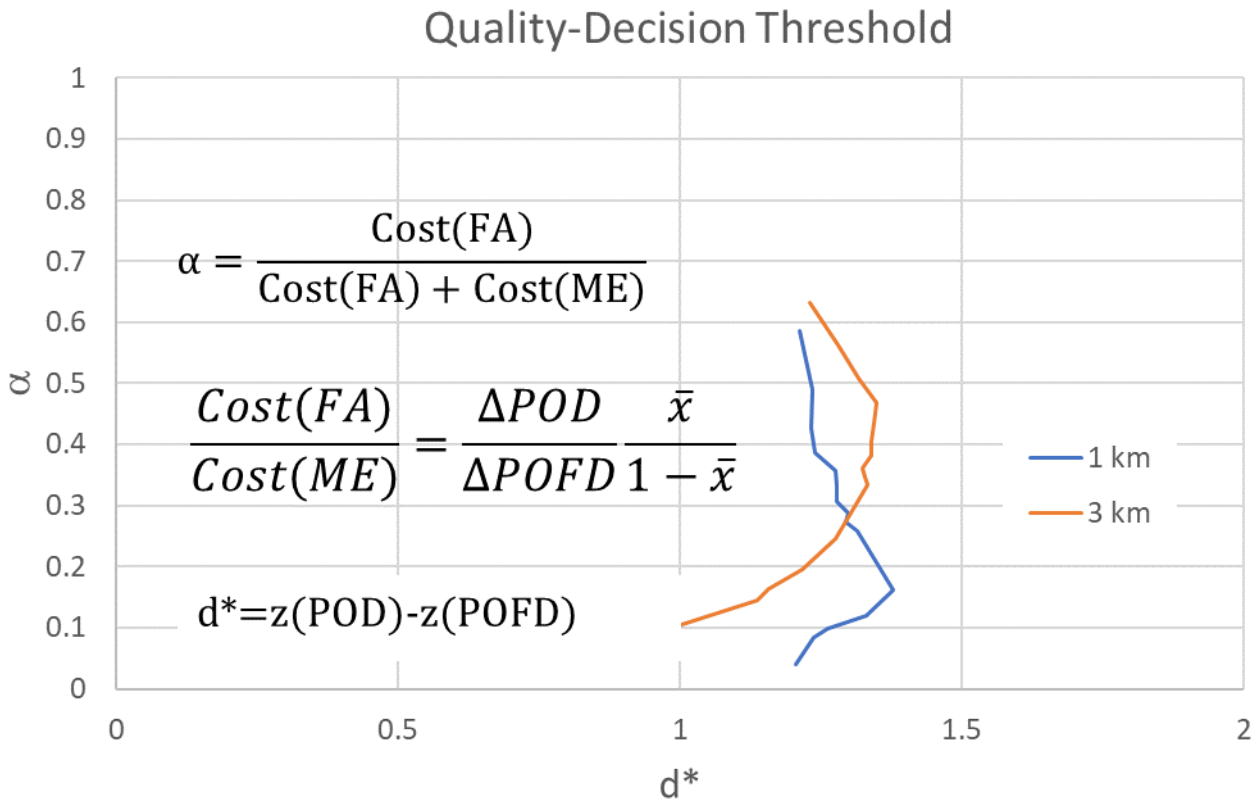


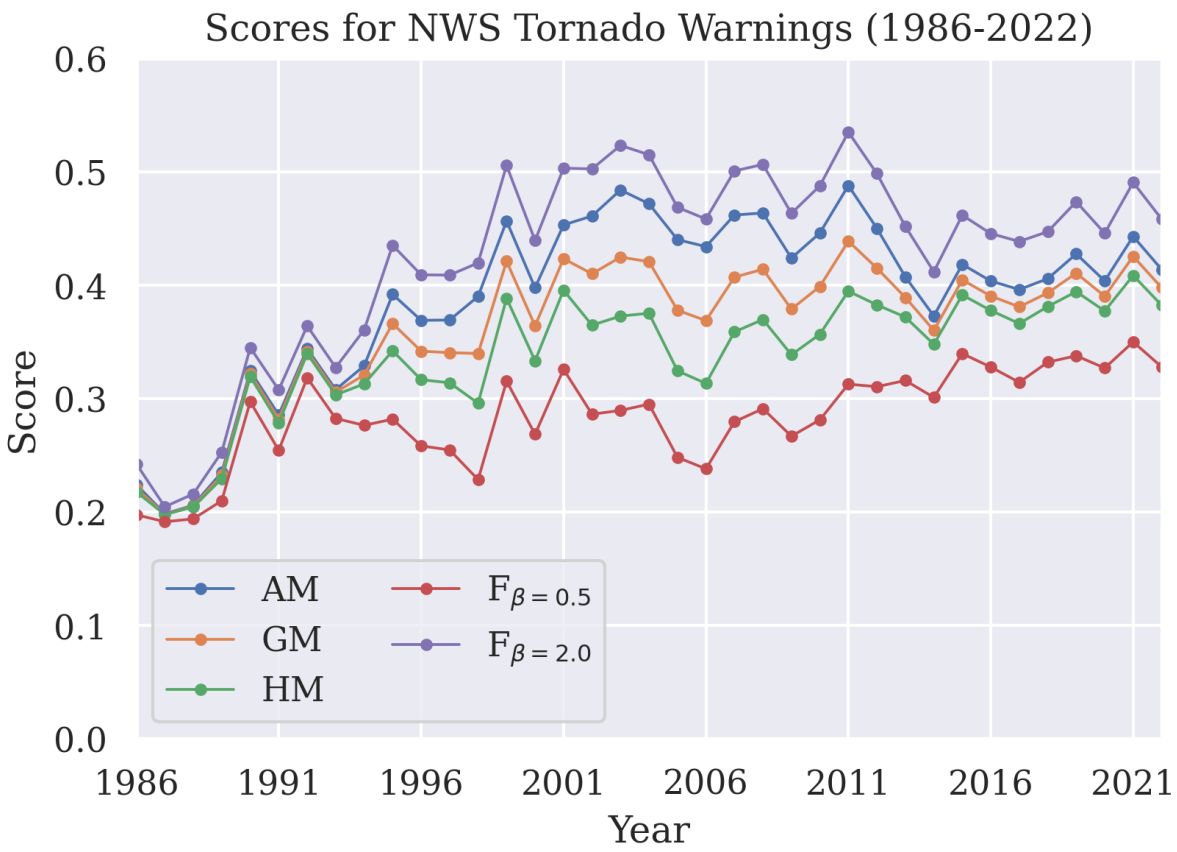


Value propositions

- CSI assumes errors are of same importance
  - Note shape of HM and GM has scores far from B=1 look different
- F and Tversky explicitly weight errors

- Implicit vs explicit value of errors

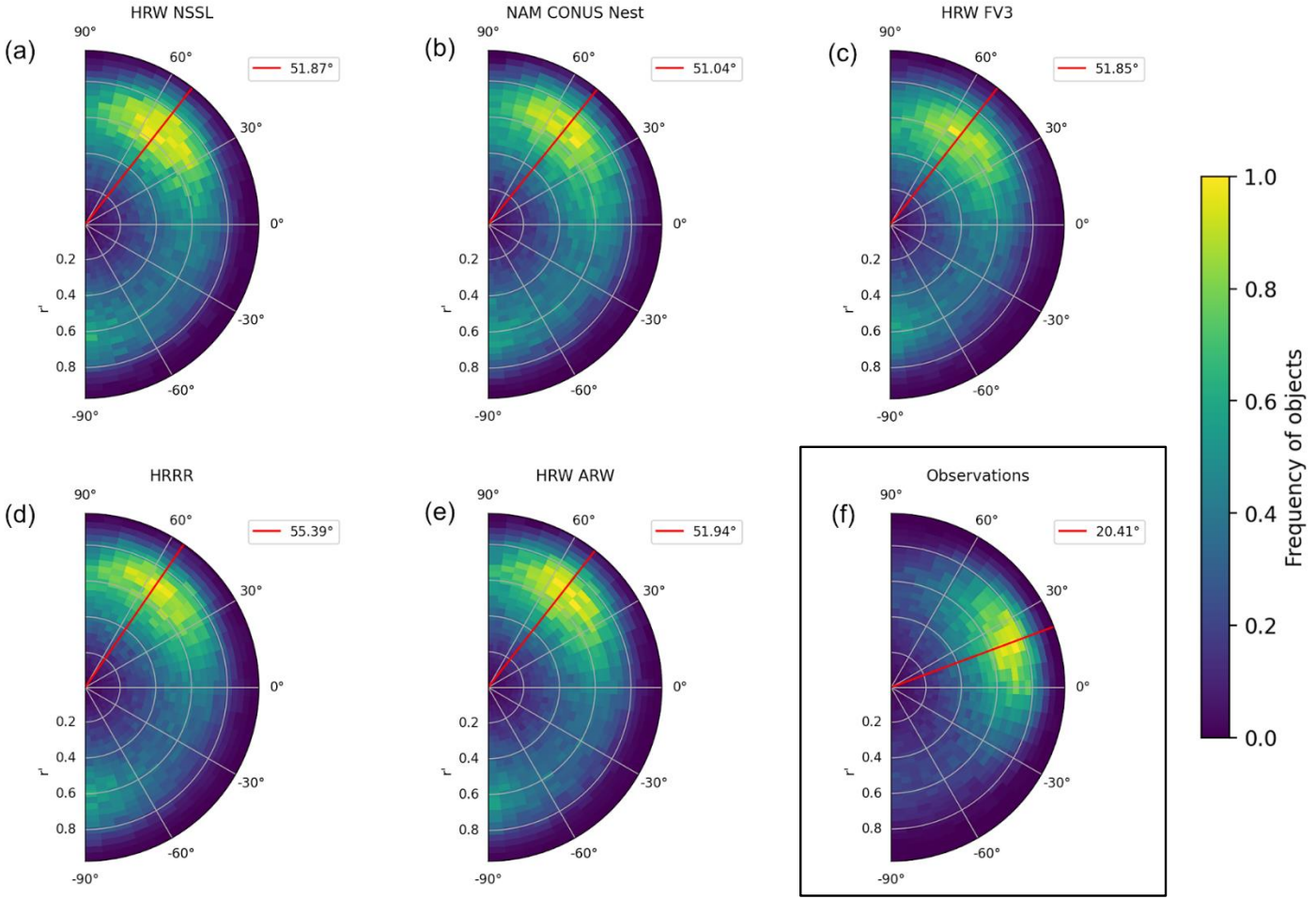

Quality-Decision Threshold
$\alpha = \frac{\text{Cost(FA)}}{\text{Cost(FA)} + \text{Cost(ME)}}$
$\frac{Cost(FA)}{Cost(ME)} = \frac{\Delta POD}{\Delta POFD} \frac{\bar{x}}{1-\bar{x}}$
d*=z(POD)-z(POFD)
1 km
3 km
α
d*
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.5
1
1.5
2


Scores for NWS Tornado Warnings (1986-2022)
Score
Year
0.0
0.1
0.2
0.3
0.4
0.5
0.6
1986
1991
1996
2001
2006
2011
2016
2021
AM
GM
HM
$F_{\beta=0.5}$
$F_{\beta=2.0}$

Climatology of models

- Do models look like observations?
- Radar reflectivity from operational models

Aspect ratio v orientation



Wrapping Up

- Variety of graphics show variety of quality/value aspects
- Acknowledge choices
- Climatology of models