Feature importance of imager cloud products, microwave radiometry, and atmospheric reanalysis variables in a deep neural network for estimating severe hail likelihood

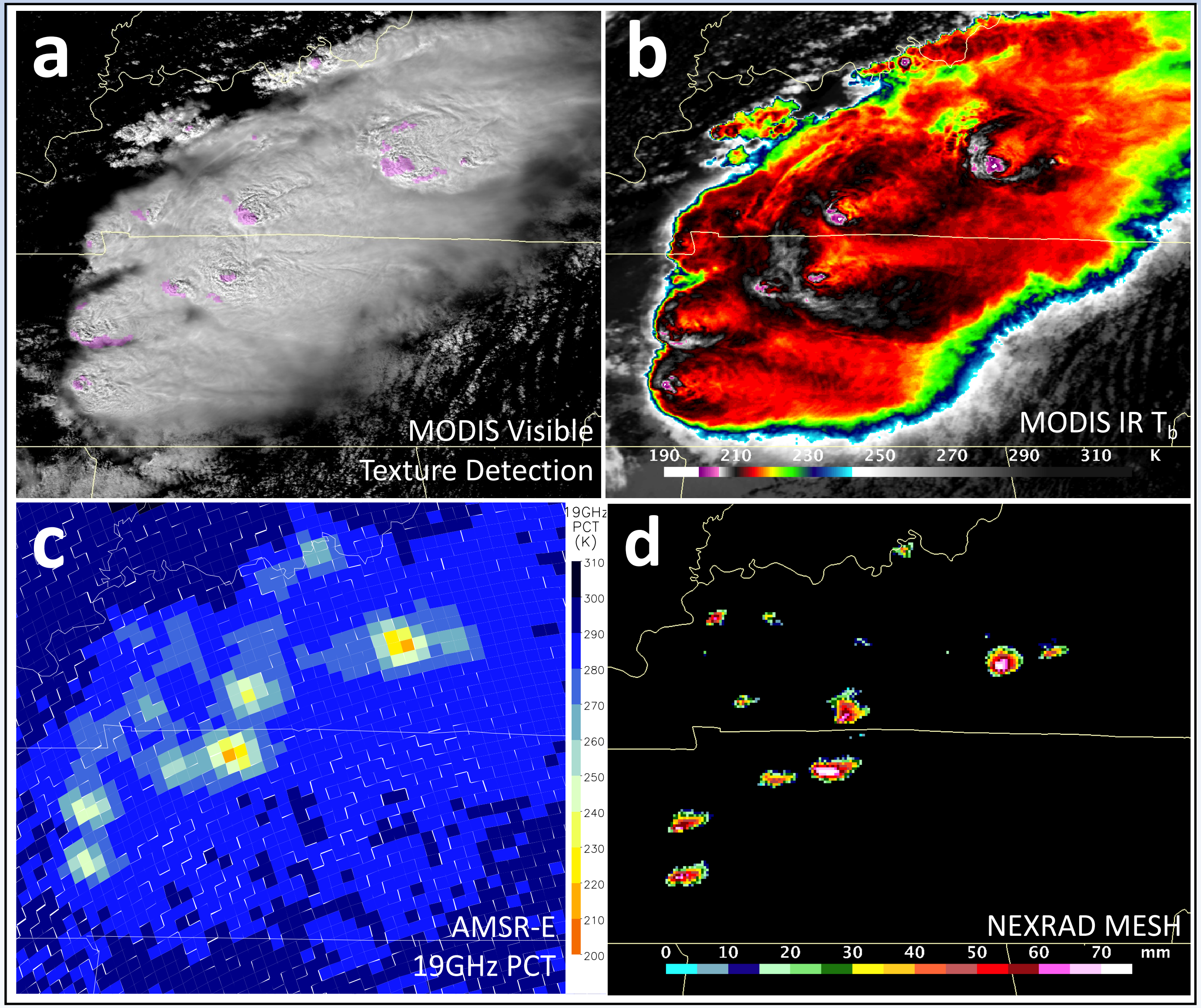
Benjamin Scarino[1], Kristopher Bedka[1], Sarah Bang[2], Kyle Itterly[3], and Daniel Cecil[2]

[1]NASA Langley Research Center, Hampton, VA
[2]NASA Marshall Space Flight Center, Huntsville, AL
[3]Analytical Mechanics Associates, Hampton, VA

Background

- Satellite visible/infrared (IR) imager and microwave radiometer (MWR) signatures of severe weather
    - Highly textured imagery
    - Cold cores < 200 K,
    - Strong 19-GHz PCT depression signifying strong scattering
- NEXRAD measurements of Maximum Expected Size of Hail (MESH) for ground truth
    - More consistent than human spotters and storm reports
    - USA only

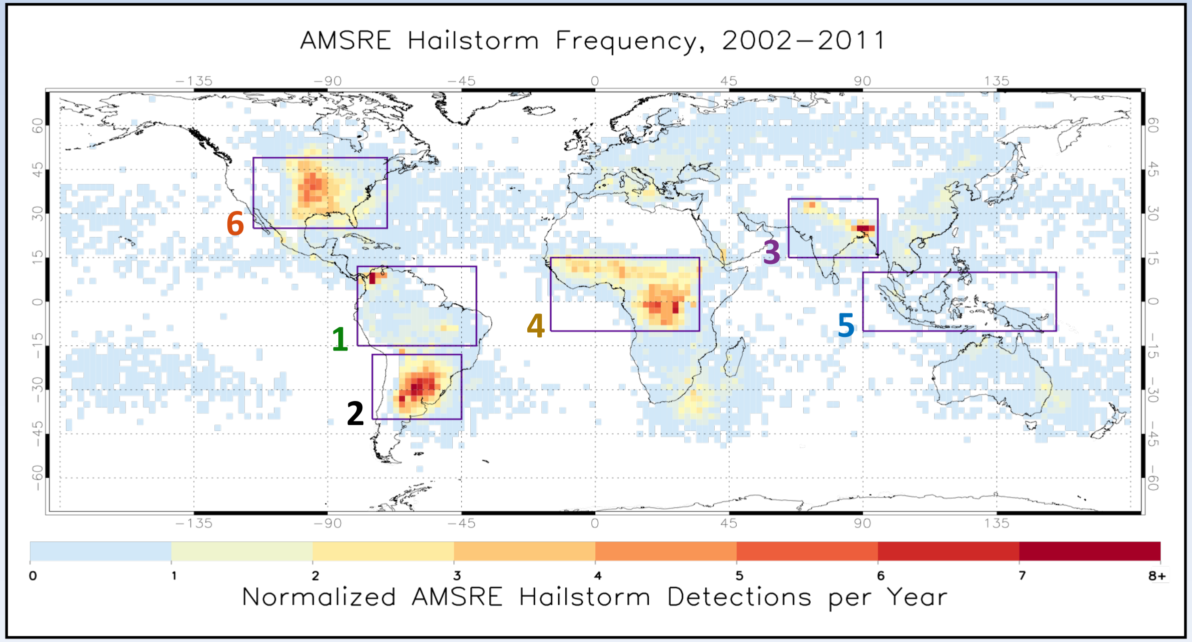
Novel demonstration of how combined satellite IR, MWR, and atmospheric reanalysis data can estimate severe hail likelihood with relatively high accuracy



Severe-weather outbreak over Kentucky and Tennessee, 7 April 2006

Method

- All Aqua orbits coincident with MESH captured using the Langley Automated Sensor Inter-calibration System (LASICS)
  - Satellite *science opportunity* identification tool
  - Local 01:30/13:30 not strongly favorable to hail over USA
- Deep neural network (DNN) trained to map MODIS (IR), AMSR-E (MWR), and/or MERRA-2 reanalysis predictors to MESH over the USA
  - 1,949 samples split 60%, 20%, and 20% across training, validation, and testing sets
  - 6-fold cross-validation
  - Estimates likelihood of 95th percentile MESH > 1.5”
  - True / Null = 48% / 52%
- Evaluate predictor importance using *Recursive Feature Elimination*
- Explore variable space as a function of hail likelihood globally

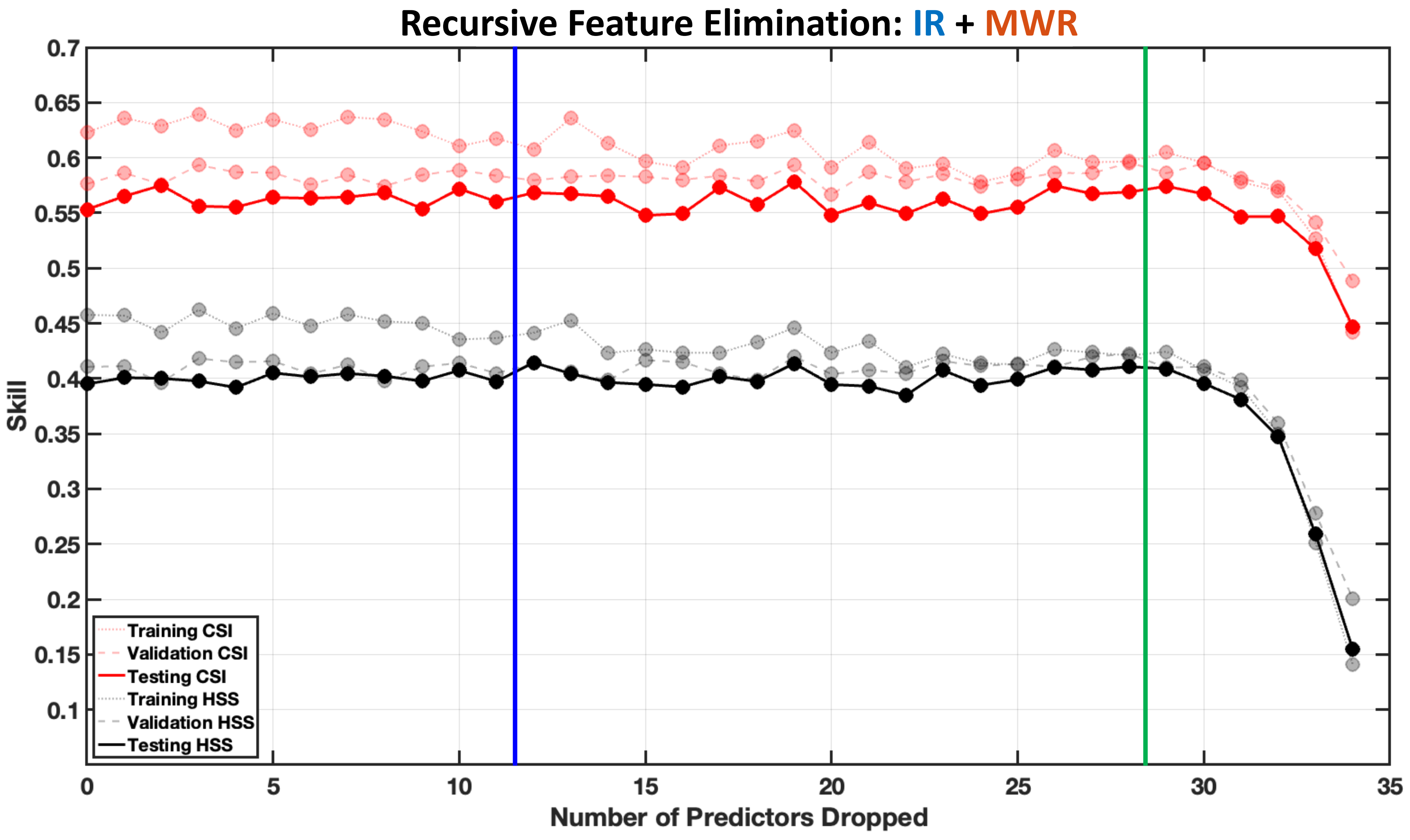


1. Amazon
2. Argentina
3. Bangladesh
4. Tropical Africa
5. Maritime Continent
6. USA

Questions:

1. **How skillfully can a DNN estimate severe hail likelihood?**
2. **How do IR, MWR, and MERRA-2 reanalysis predictors contribute to that skill?**
3. **Is the model globally generalized – how do predictors vary across distinct regions?**

Neural Network Predictor Importance



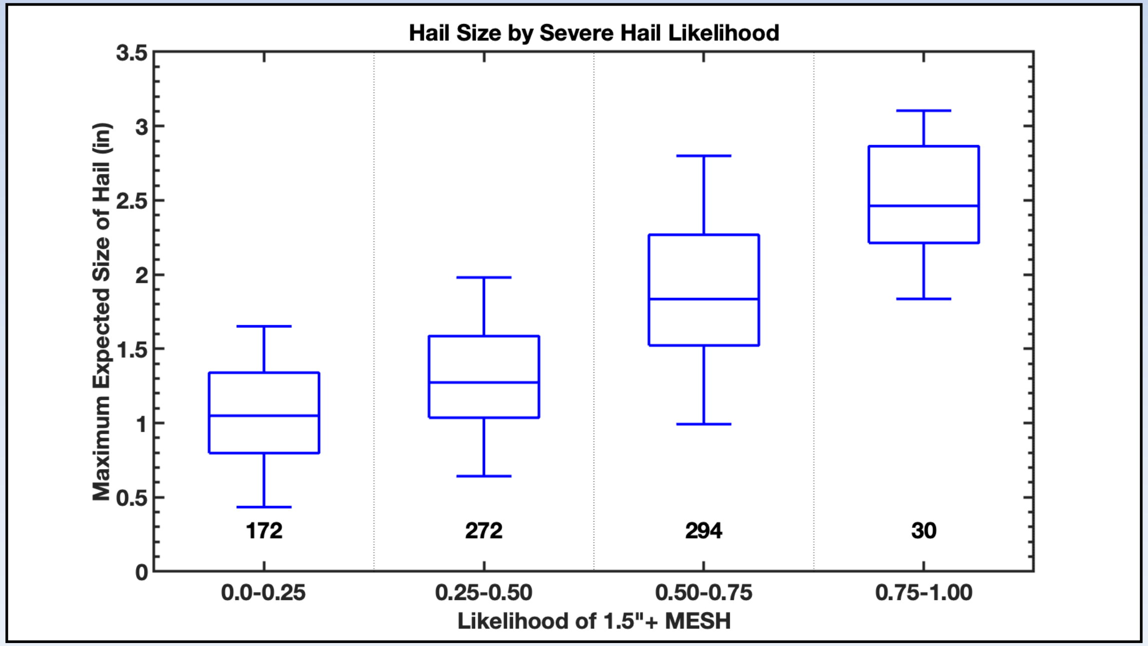
Predictor Importance

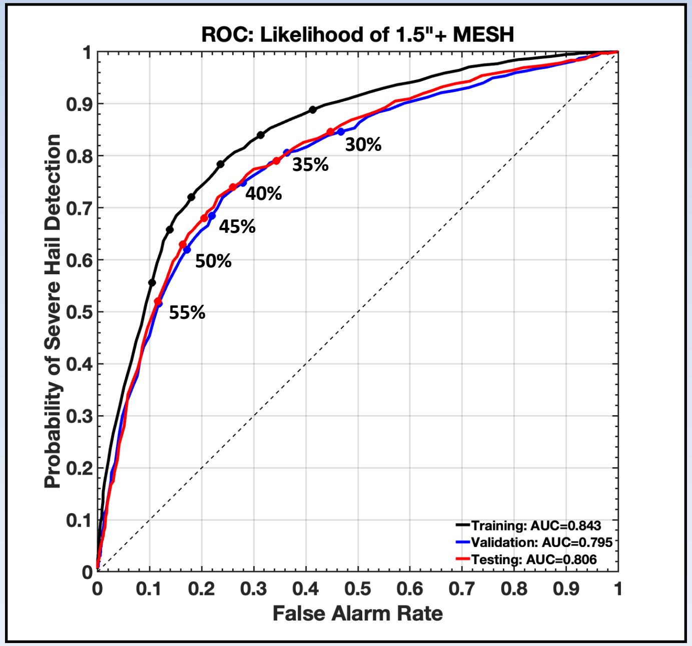
1: Min 19GHz PCT (K)
2: Adj Max 10GHz PCT (K)
3: N 37GHz Pixels < 275 K
4: IR-Tropopause T Diff (K)
5: Cloud-top Height (km)
6: N 85GHz Pixels < 200 K
7: Hail Prob (Min 19GHz PCT)
8: Max 19GHz PCT (K)
9: Max 10GHz PCT (K)
10: Min 10GHz PCT (K)
11: N 85GHz Pixels < 100 K
12: Max 37GHz PCT (K)
13: Anvil Frequency (%)
14: Area of Cold Cloud (km^2)
15: N 85GHz Pixels < 125 K
16: N 37GHz Pixels < 250 K
17: N 85GHz Pixels < 175 K
18: N 85GHz Pixels < 150 K
19: Anvil Height (km)
20: Adj Max 19GHz PCT (K)
21: Min 85GHz PCT (K)
22: OT Probability
23: Deep Convection Score
24: N 37GHz Pixels < 175 K
25: Hail Prob (Min 37GHz PCT)
26: Adj Max 37GHz PCT (K)
27: N 37GHz Pixels < 200 K
28: OT - Anvil BT Diff (K)
29: Min 37GHz PCT (K)
30: Adj Min 19GHz PCT (K)
31: Max 85GHz PCT (K)
32: Adj Min 10GHz PCT (K)
33: OT Area (km^2)
34: N 37GHz Pixels < 225 K
35: Adj Min 37GHz PCT (K)

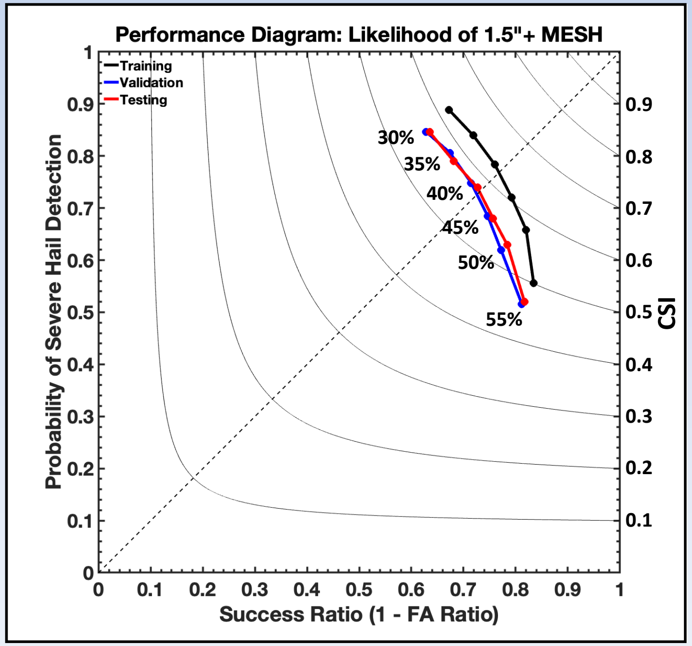
- Train model with all predictors
- Exclude one feature and retrain
- Eliminate least contributing feature and repeat

19 GHz PCT is a strong, unique signal that's very sensitive to ice scattering

Neural Network Skill: IR + MWR Predictors



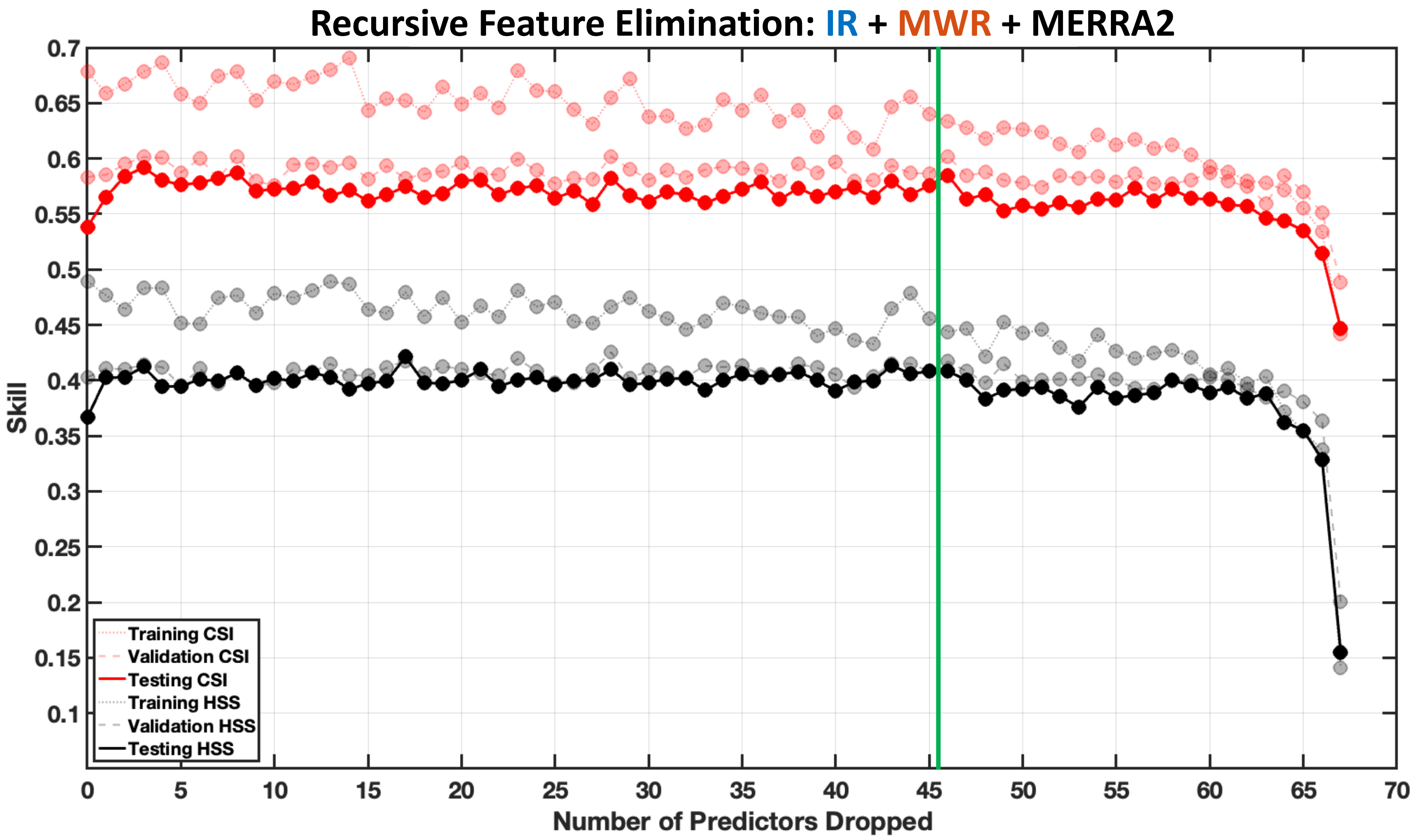




DNN is assigning the highest likelihoods to storms with the greatest MESH

- High fold variance and overfitting perhaps owed to low sample size
- Generalization possible with reasonable skill
- 6-fold Testing set skill exceeds 0.41 (±0.03) HSS and 0.56 (±0.02) CSI at ~50/50 class split
  - 70% (±5%) Recall, 76% (±6%) Precision, 20% (±7%) False Alarm Rate
  - Suggests *balanced* performance

Neural Network Predictor Importance



Predictor Importance

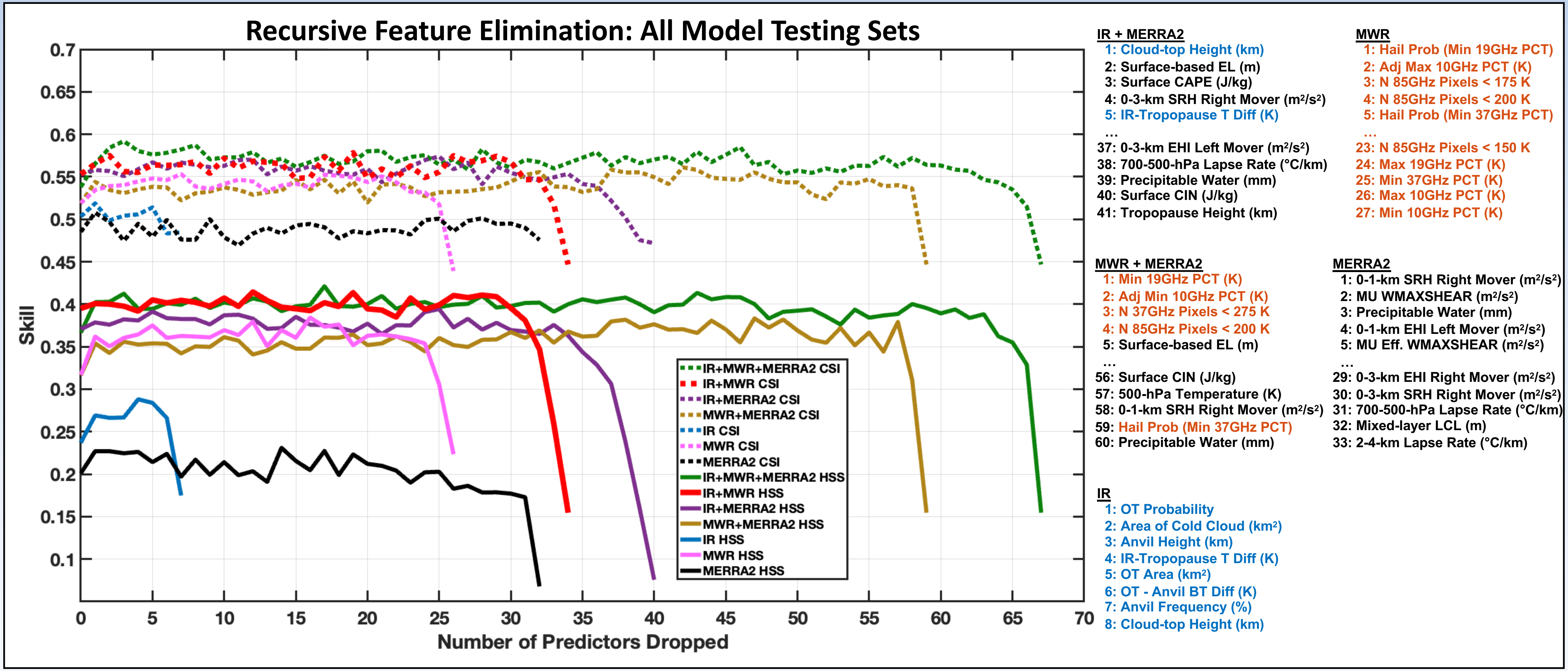
1: Min 19GHz PCT (K)*
2: Min 10GHz PCT (K)
3: N 85GHz Pixels < 175 K
4: Cloud-top Height (km)*
5: Precipitable Water (mm)
6: Adj Min 19GHz PCT (K)
7: Surface CAPE (J/kg)
8: 0-1-km SRH Left Mover (m^2/s^2)
9: 0-3-km EHI Left Mover (m^2/s^2)
10: 0-3-km SRH Left Mover (m^2/s^2)
11: Max 85GHz PCT (K)
12: 0-3-km EHI Right Mover (m^2/s^2)
13: Mixed-layer LCL (m)
14: 0-3-km SRH Right Mover (m^2/s^2)
15: N 85GHz Pixels < 100 K
16: 1000-m Mixed-layer CAPE (J/kg)
17: Min 37GHz PCT (K)
18: Surface-based EL T (K)
19: 0-1-km EHI Right Mover (m^2/s^2)
20: MU Eff. WMAXSHEAR (m^2/s^2)
21: 1000-m Mixed-layer CIN (J/kg)
22: IR-Tropopause T Diff (K)*
23: 2-4-km Lapse Rate (°C/km)
24: OT Probability
25: Hail Prob (Min 37GHz PCT)
26: 0-1-km EHI Left Mover (m^2/s^2)
27: Supercell Composite Parameter
28: Hail Prob (Min 19GHz PCT)
29: Severe Weather Threat Index
30: Significant Tornado Parameter
31: N 37GHz Pixels < 200 K
32: 0-1-km SRH Right Mover (m^2/s^2)
33: Surface CIN (J/kg)
34: N 85GHz Pixels < 125 K
35: 700-500-hPa Lapse Rate (°C/km)
36: Adj Min 10GHz PCT (K)
37: MU WMAXSHEAR (m^2/s^2)
38: Max 10GHz PCT (K)
39: Anvil Frequency (%)
40: Adj Max 37GHz PCT (K)
41: N 37GHz Pixels < 250 K
42: Adj Min 37GHz PCT (K)
43: Deep Convection Score
44: 0-3-km Lapse Rate (°C/km)
45: Max 19GHz PCT (K)
46: Anvil Height (km)
47: OT Area (km^2)
48: 0-6-km Wind Shear (m/s)
49: 0-1-km Wind Shear (m/s)
50: N 37GHz Pixels < 175 K
51: N 85GHz Pixels < 150 K
52: Min 85GHz PCT (K)
53: Tropopause Height (km)
54: Area of Cold Cloud (km^2)
55: Surface-based EL (m)
56: Most Unstable CIN (J/kg)
57: Adj Max 10GHz PCT (K)*
58: 500-hPa Temperature (K)
59: Surface-based LCL (m)
60: Freezing Level (m)
61: Max 37GHz PCT (K)
62: N 37GHz Pixels < 225 K
63: OT - Anvil BT Diff (K)
64: Significant Hail Parameter
65: Most Unstable CAPE (J/kg)
66: N 37GHz Pixels < 275 K*
67: N 85GHz Pixels < 200 K*
68: Adj Max 19GHz PCT (K)

*Top 6 IR + MWR Model Predictors

- A feature may be deemed “unimportant” because of irrelevancy or redundancy
- That is, information may be important, but is already contained in remaining features

Significantly correlated features may “demote” one another

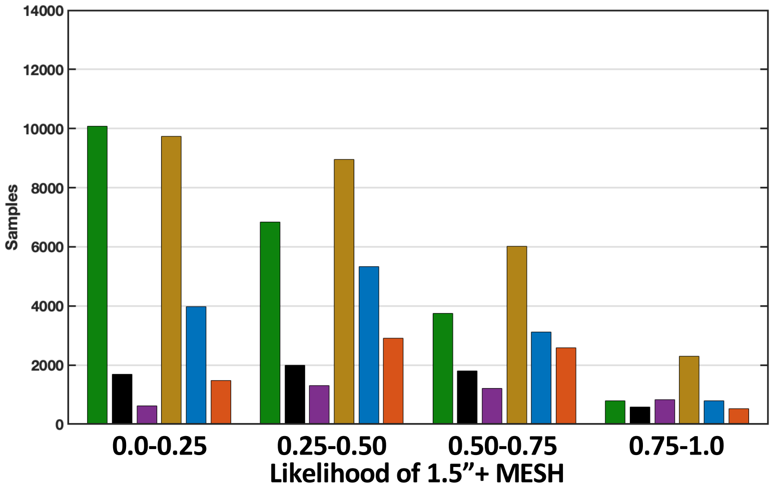
Neural Network Predictor Importance



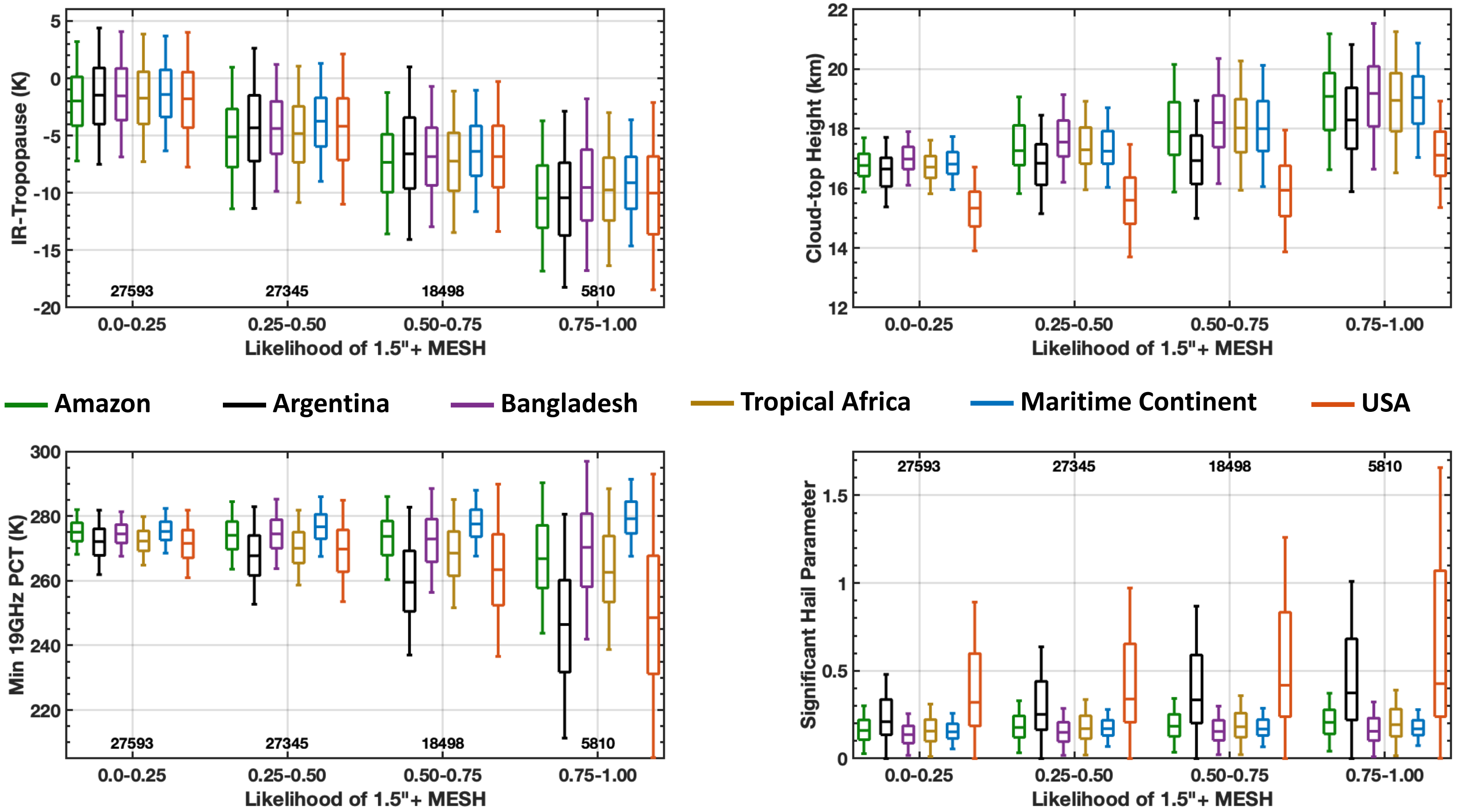
- IR+MWR model shows similar features sensitivity and skill to that of the IR+MWR+MERRA-2 model
- IR- or MERRA-2-only models have poorest skill
- Min 19-GHz PCT remains consistently important

A DNN with MWR predictors alone is the most capable single-source model

Variable Space by Region



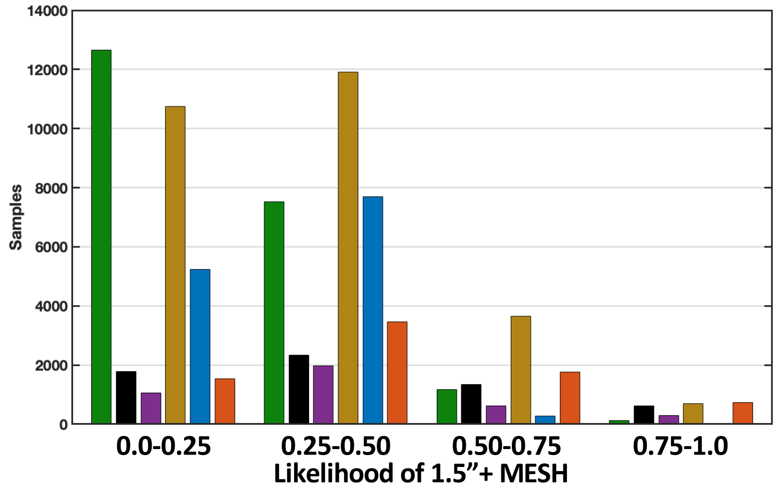
By IR + MWR Model Likelihoods



Training a model over USA with MERRA-2 is going to reinforce MERRA-2 – also cloud heights and cold area

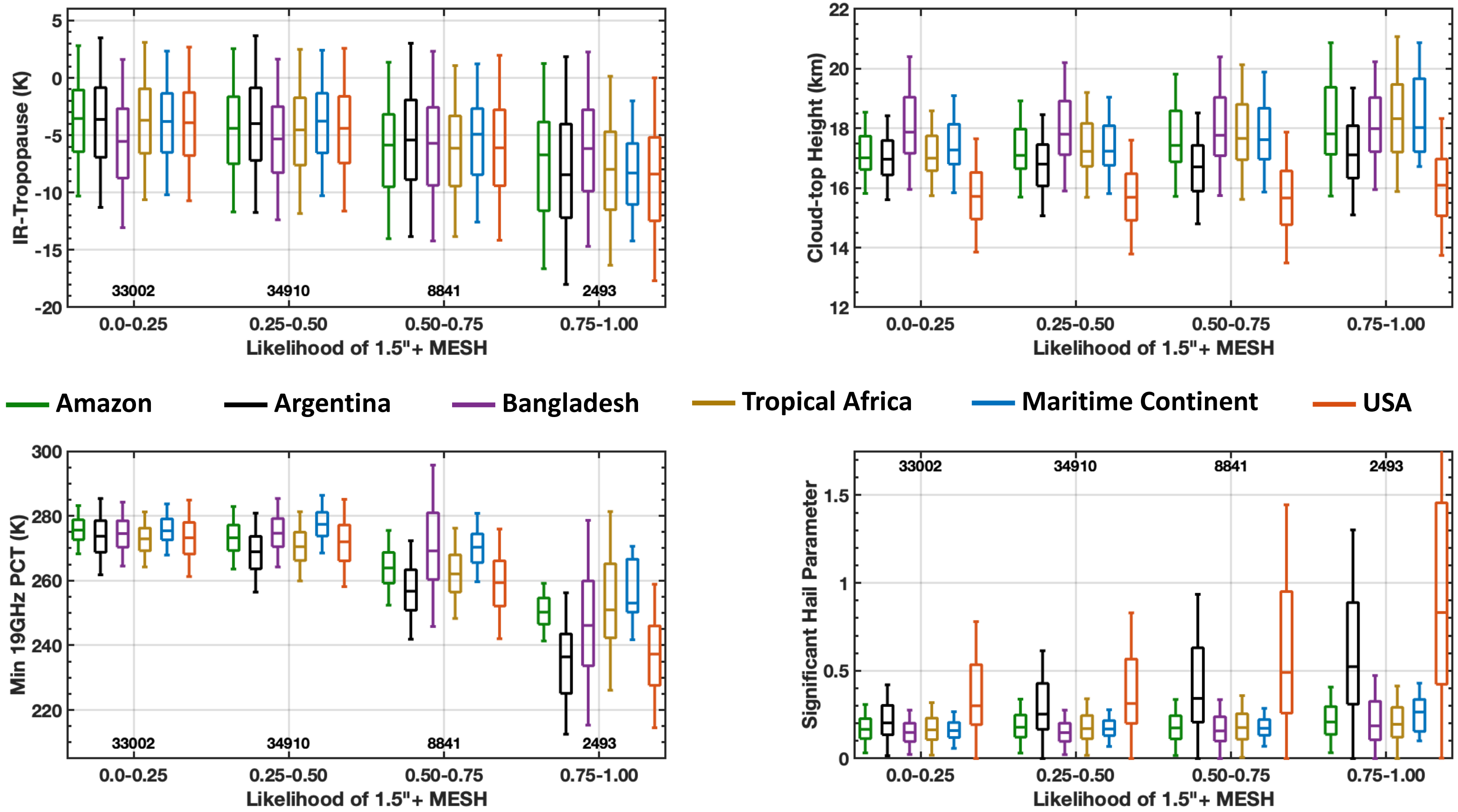
- Are predictor variables well grouped?
- Height and environment not generalized predictors
- Maritime Continent has peculiarly high likelihoods

Variable Space by Region



Larger features are preferred when model is trained on MWR – dominates otherwise strong IR patterns

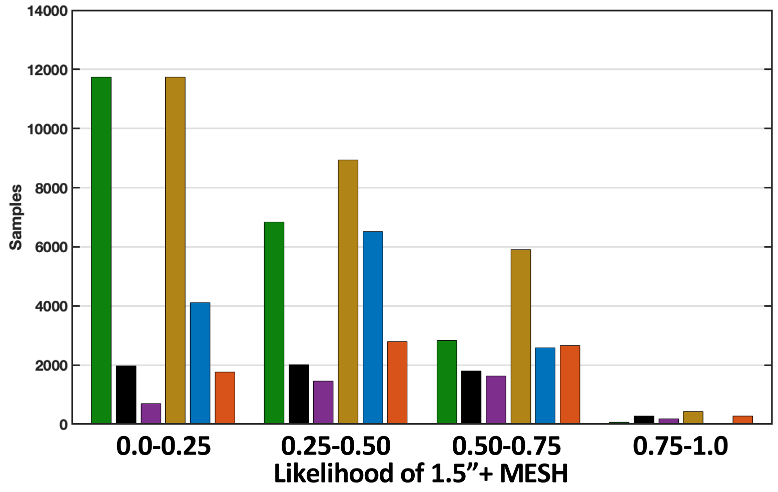
By MWR Model Likelihoods

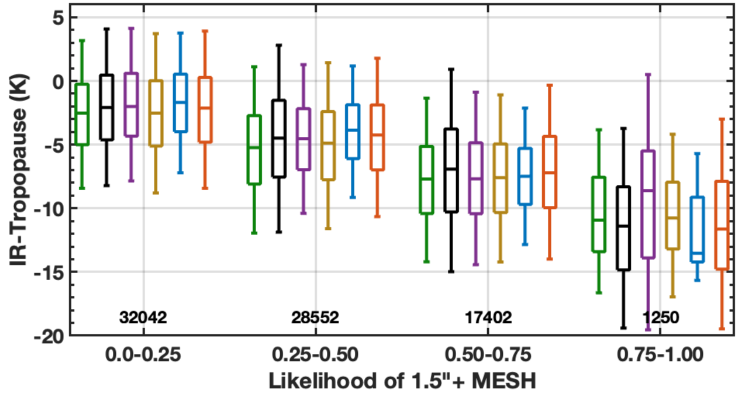


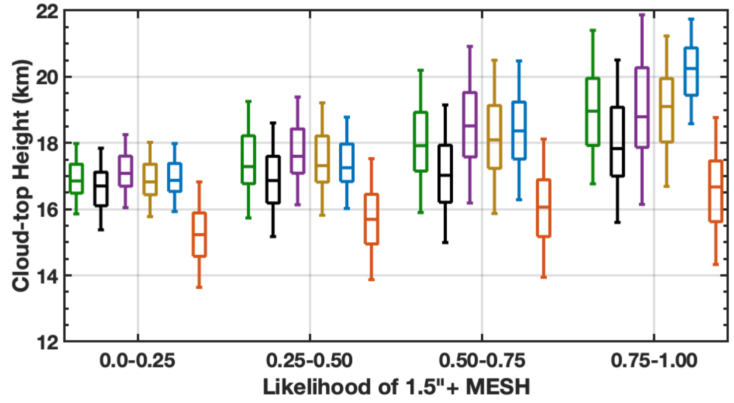
- Better grouping of MWR
- IR-Trop signal is downplayed

Variable Space by Region

By Alternative IR + MWR Model Likelihoods (no heights, cold cloud area, or 37-GHz PCT)

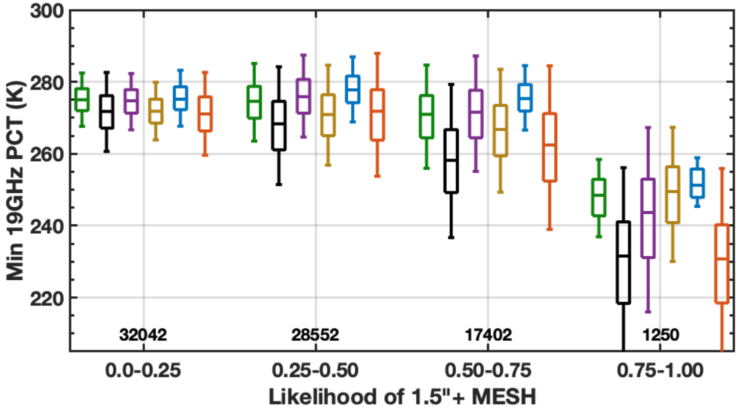


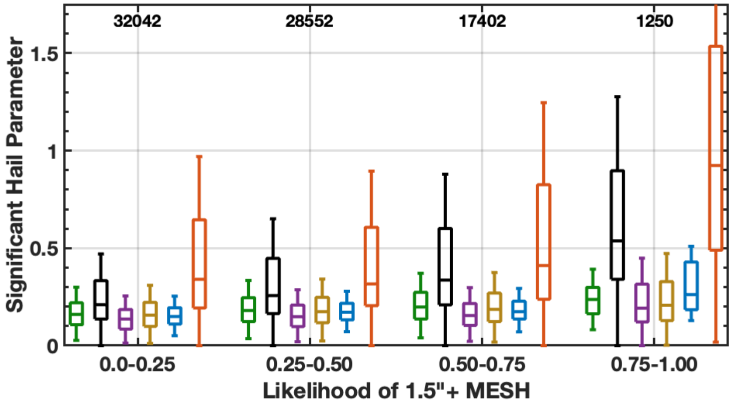




Amazon Argentina Bangladesh Tropical Africa Maritime Continent USA

Removing latitude-dependent variables helps improves generalization – but constrains predictive dynamics range





- Sample size limitations hurting skill of more generalize model
- Poorly generalized predictors can be specifically important – not a lot of sample size to work with if losing key predictors

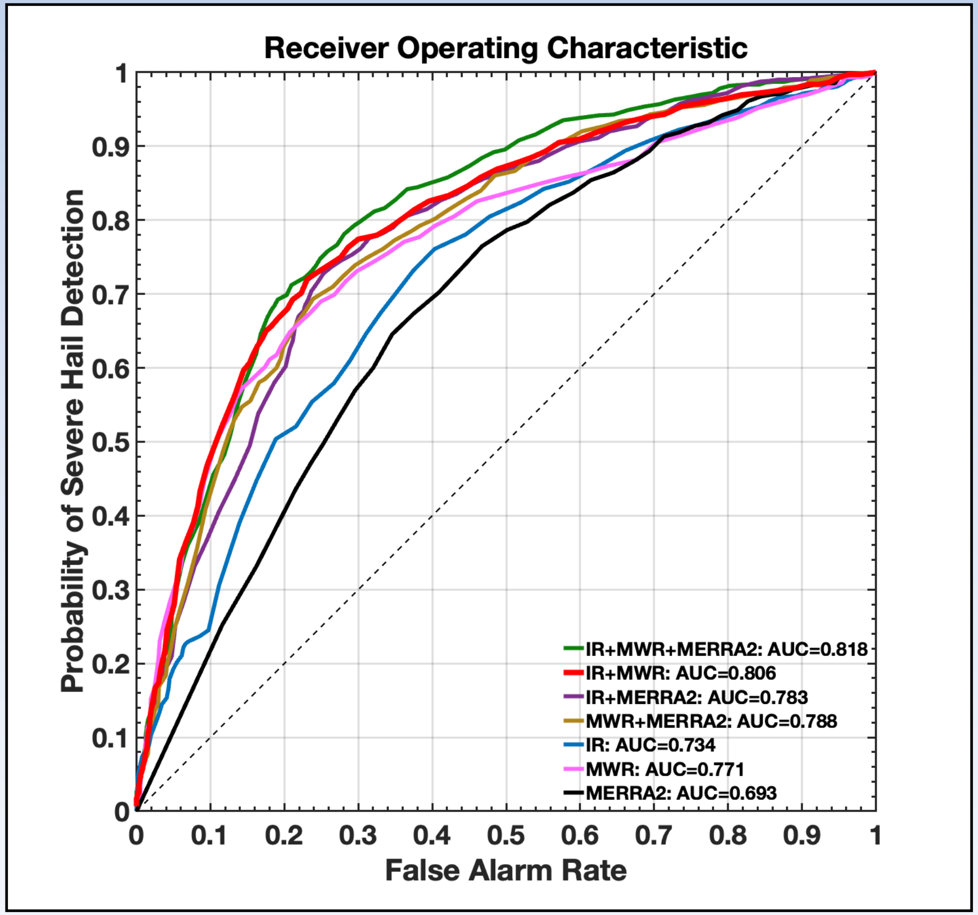
Summary

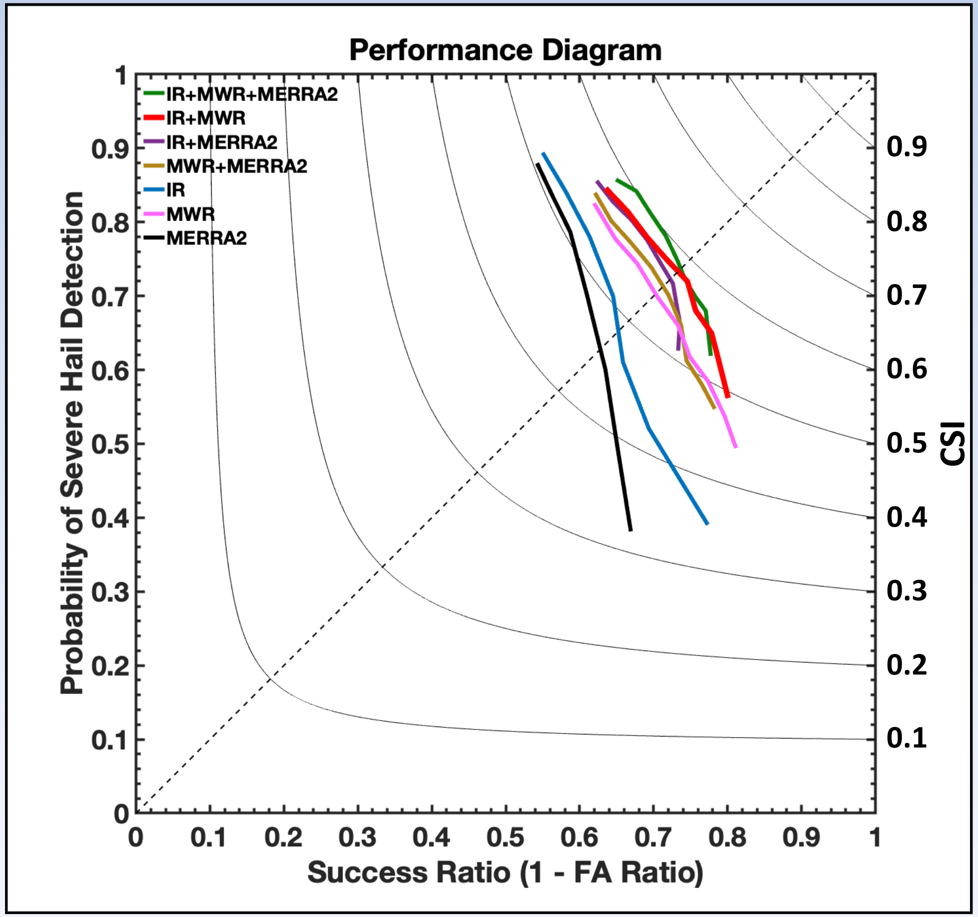
Key Predictors		
Thermodynamic	**Kinematic**	**Scale**
IR-Tropopause T Diff (K)	Storm Relative Helicity (m^2/s^2)	19GHz PCT (K)
Surface-based EL (m)	Energy Helicity Index (m^2/s^2)	10GHz PCT (K)
Surface-based CAPE (J/kg)	WMAXSHEAR (m^2/s^2)	N 85GHz Pixels
Precipitable Water (mm)		Area of Cold Cloud (km^2)

- IR+MWR DNN skillfully estimates severe hail likelihood
- MERRA-2 reanalysis predictors offer slightly better skill, but are less globally generalized
- Larger features are preferred when model is trained on MWR – which is also more latitudinally-independent than models that include IR in training

Additional Slides

Neural Network Skill: All Model Testing Sets



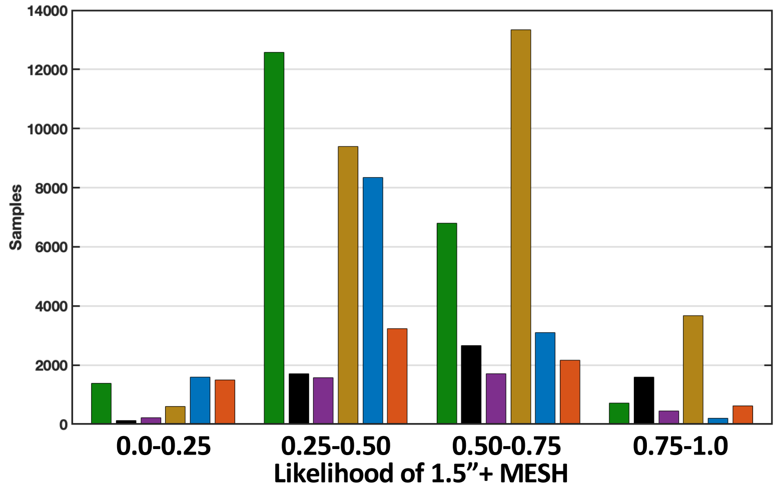


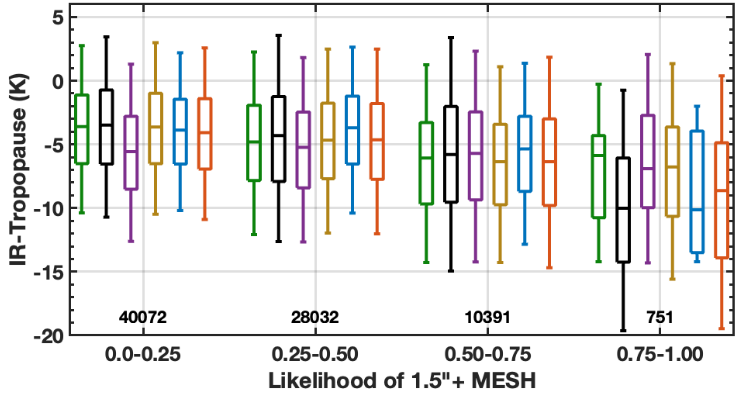
- Model most skillful when combining all IR + MWR satellite observations and MERRA-2
- IR + MWR combination model remains reasonably skillful

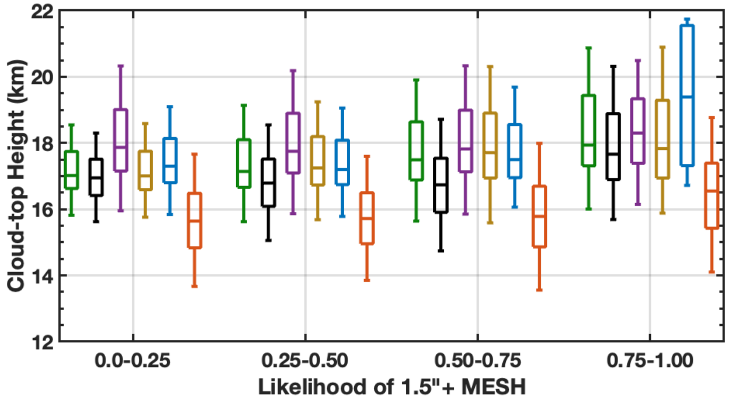
Important to consider regionally dependency or atmospheric environment

Variable Space by Region

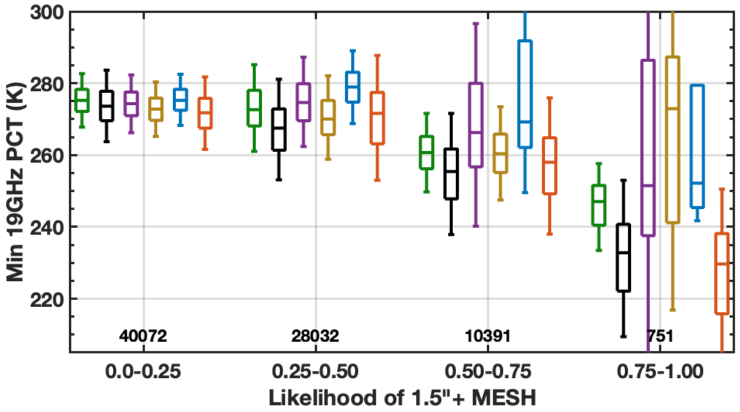
By IR + MWR + MERRA2 Model Likelihoods

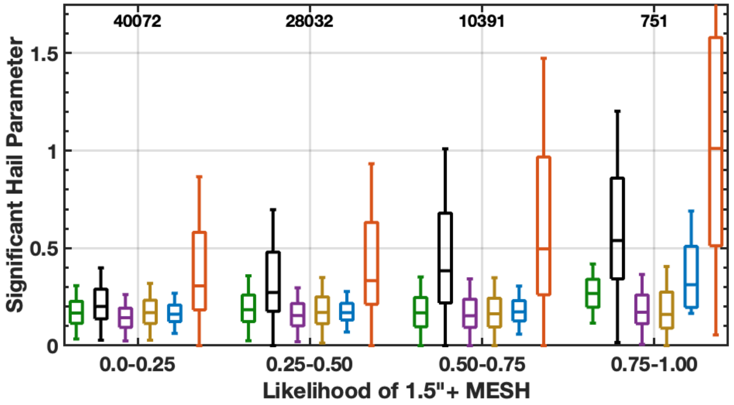




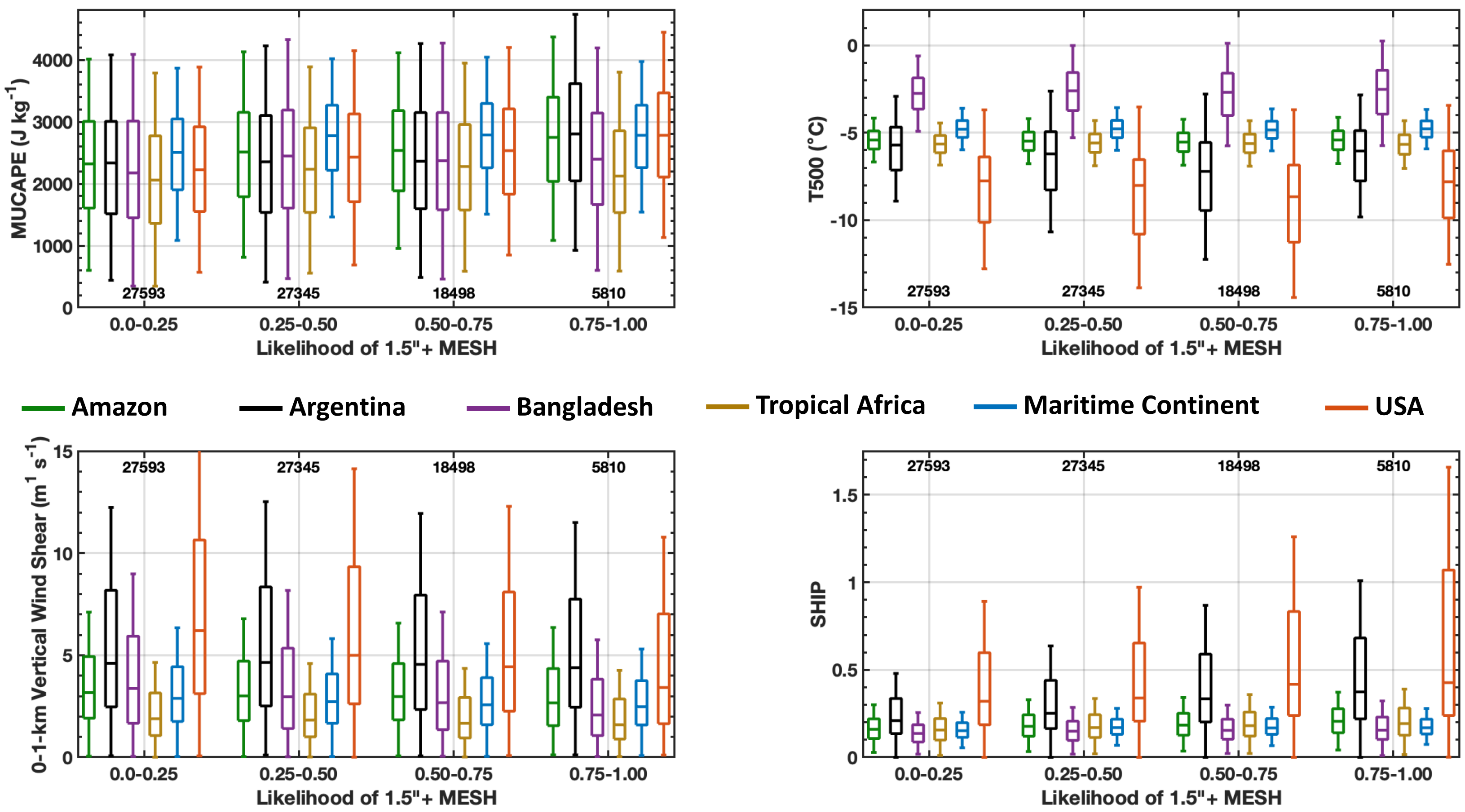


Amazon Argentina Bangladesh Tropical Africa Maritime Continent USA

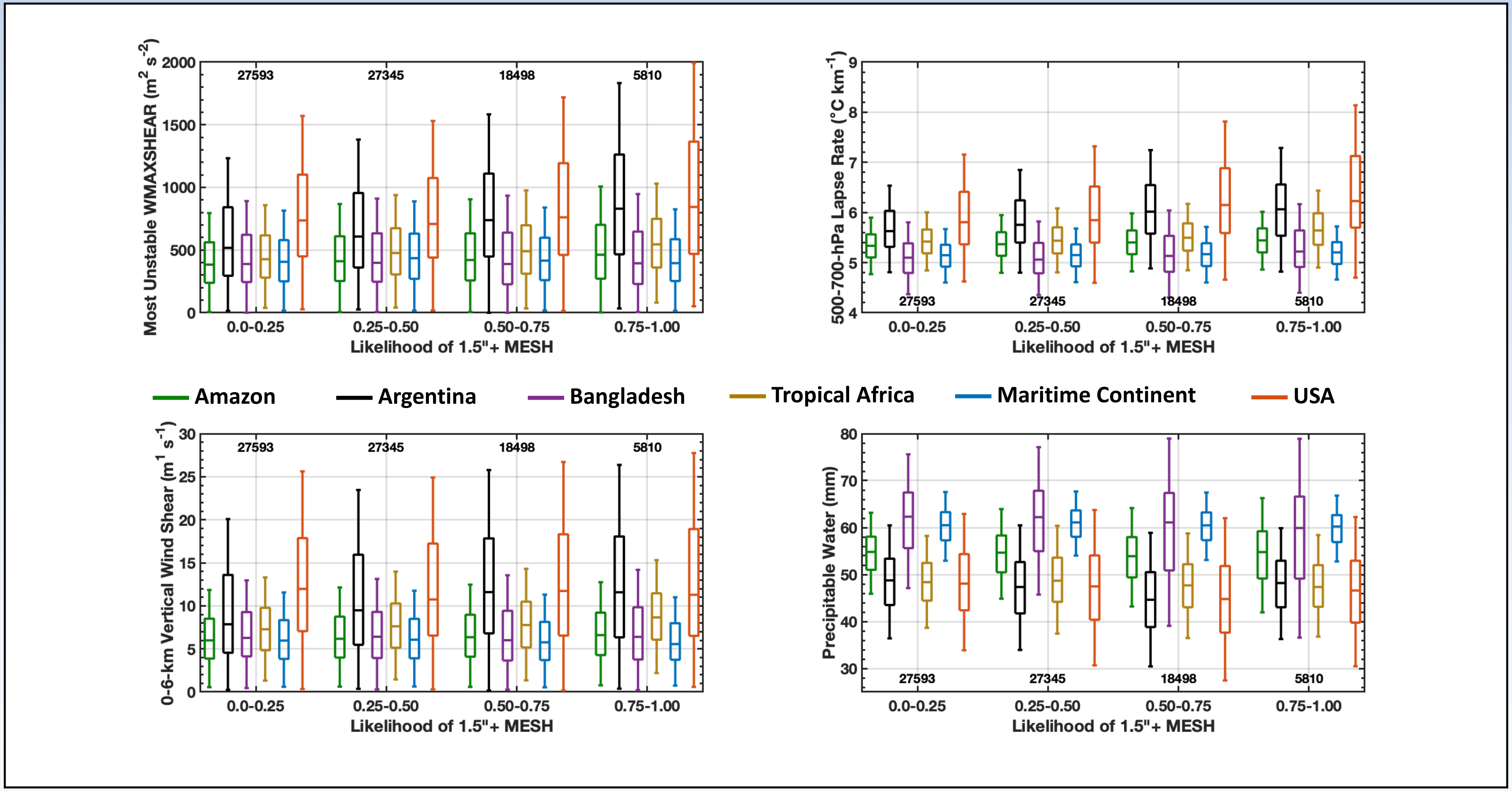




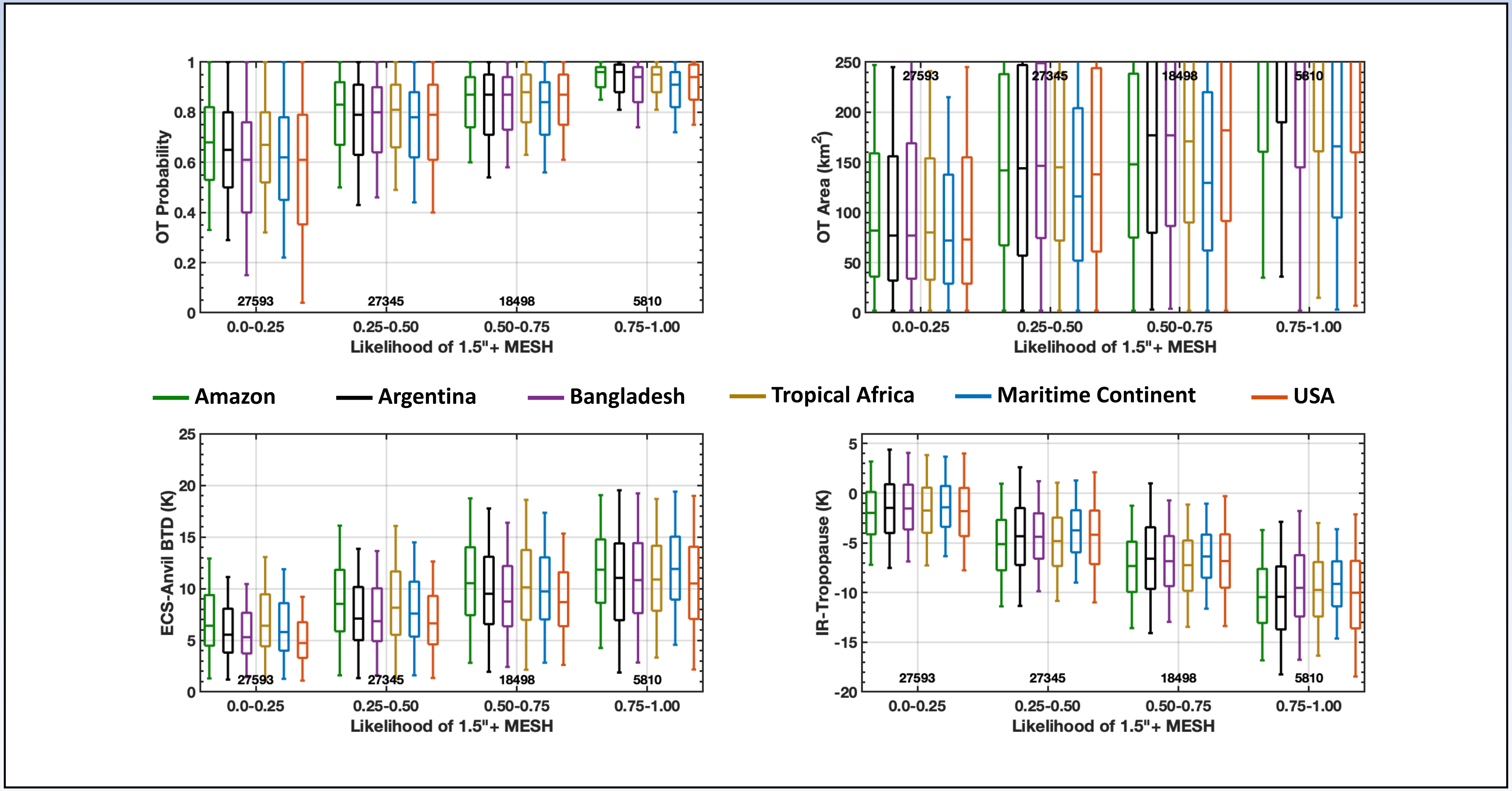
Additional Satellite MERRA2 by Region



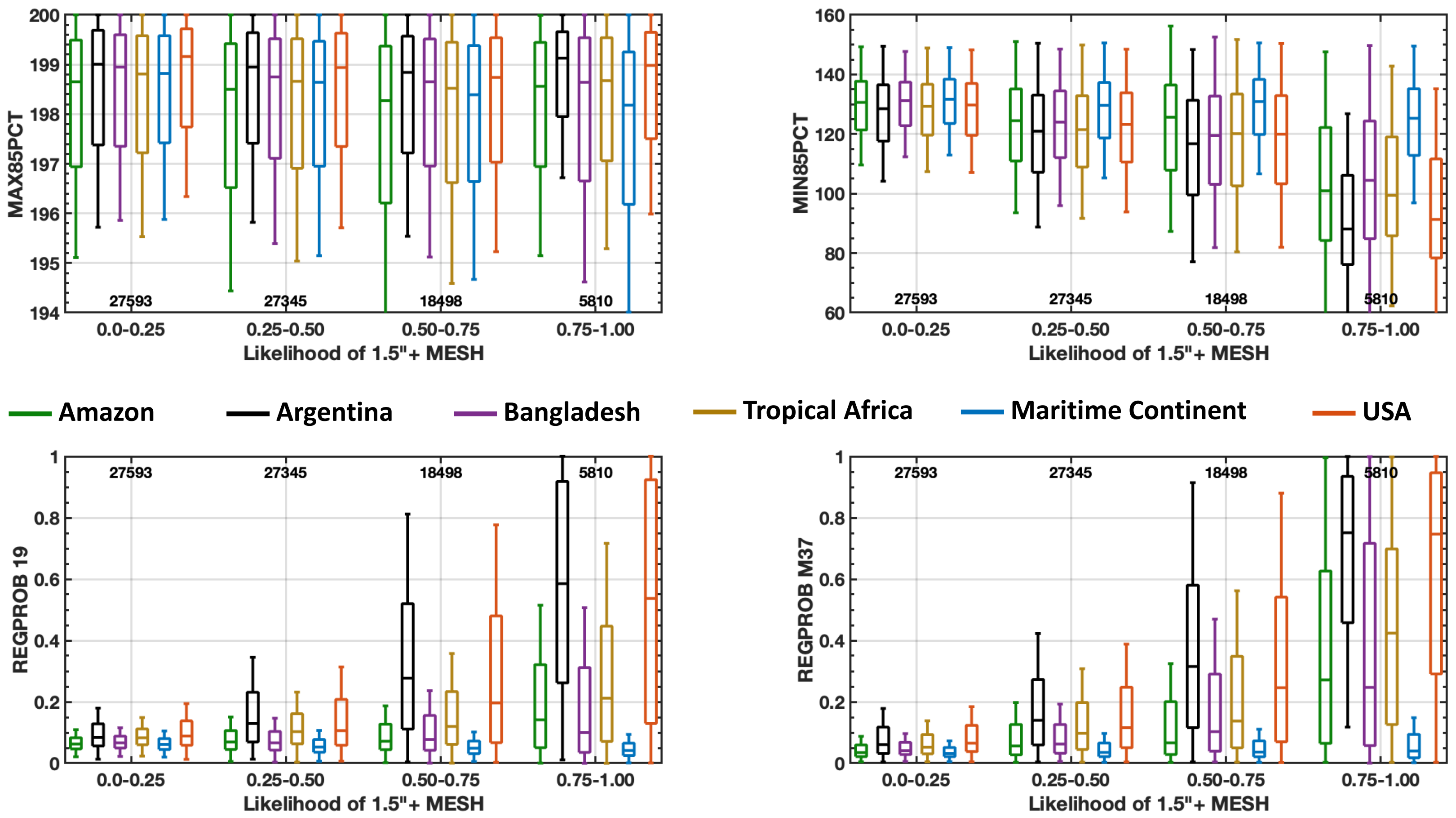
Additional Satellite MERRA2 by Region



Additional Satellite IR by Region



Additional Satellite MWR by Region



Additional Satellite MWR by Region