Performance of the Traffic Flow Management Convective Forecast (TCF)

Jason Baker, Kevin Stone, and Matthew Wandishin

FPAW Fall 2023

14 November 2023

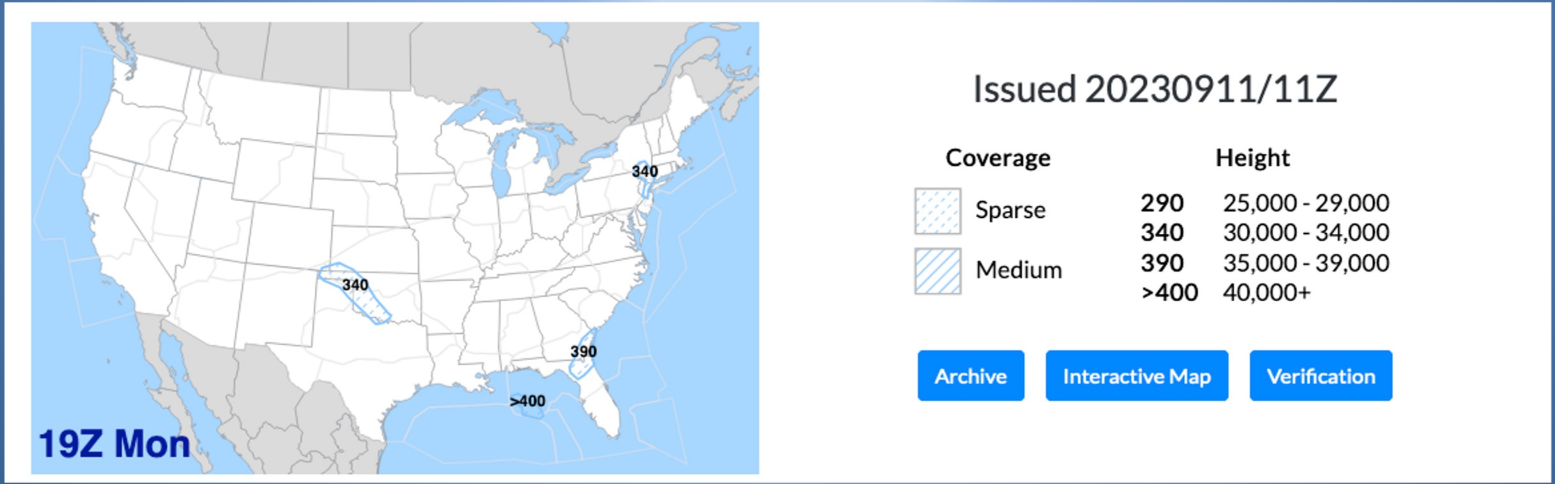
Outline

- Motivation
- Background on TCF
- Background on verification techniques
- TCF performance results and case studies
- Summary

Motivation

- TCF is one of the primary tools used in the NAS for convective weather forecasting
- TCF has evolved from Collaborative Convective Forecast Product (CCFP) and Collaborative Aviation Weather Statement (CAWS)
  - Forecast polygons are now smaller; do metrics align with feedback?
- Metrics that support operations
  - Easy to access with a consistent approach
  - Criteria rarely verifies; does criteria need to change?
- How collaboration impacts the forecast
- TCF's role within an Infocentric NAS
  - New Product on Aviation Weather Display arriving in 2024
  - Convective Weather Avoidance Polygons and TCF overlap/conflict?

TCF Background



TCF Definition

- TCF products are polygons denoting areas of convection meeting certain criteria

1. Composite radar reflectivity of at least 40 dBZ;
2. Echo tops at or above FL250;
3. Coverage (1 & 2) of at least 25% of the polygon area;
4. Forecaster confidence of at least 50% (High) that criteria (1, 2, & 3) will be met.

TCF Purpose

Purpose of TCF

1. Intended to provide an accurate representation of the **convection of most significance for strategic decisions of air traffic flow management**;

TCF Purpose

Purpose of TCF

1. Intended to provide an accurate representation of the convection of most significance for strategic decisions of air traffic flow management;

2. Intended to provide a **common forecast baseline**, as consistent as possible, shared and **collaborated among all meteorological organizations** responsible for providing forecasts of convection to Air Traffic Management (ATM) within the FAA/Industry Collaborative Decision Making (CDM) processes and/or within commercial aviation organizations; and

TCF Purpose

Purpose of TCF

1. Intended to provide an accurate representation of the convection of most significance for strategic decisions of air traffic flow management;

2. Intended to provide a common forecast baseline, as consistent as possible, shared and collaborated among all meteorological organizations responsible for providing forecasts of convection to Air Traffic Management (ATM) within the FAA/Industry Collaborative Decision Making (CDM) processes and/or within commercial aviation organizations; and

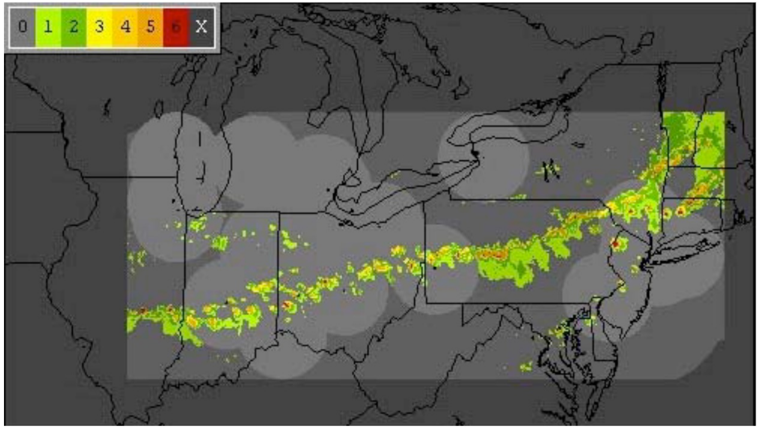
3. Intended for use as the **authoritative source** of convective weather forecast information for Traffic Flow Management **strategic planning and decisions** which are **collaborated** between the government and industry.

TCF Creation and Observations

TCF Auto: polygons produced by algorithm based on input from the HRRR and HIRES WRF-ARW models

TCF Prelim/Final: initially drawn by Aviation Weather Center forecaster (Prelim), then adjusted through collaborative process with Center Weather Service Units and Industry

Observations: Corridor Integrated Weather System (CIWS) radar mosaic (VIL and Echo Top)



Verification Background

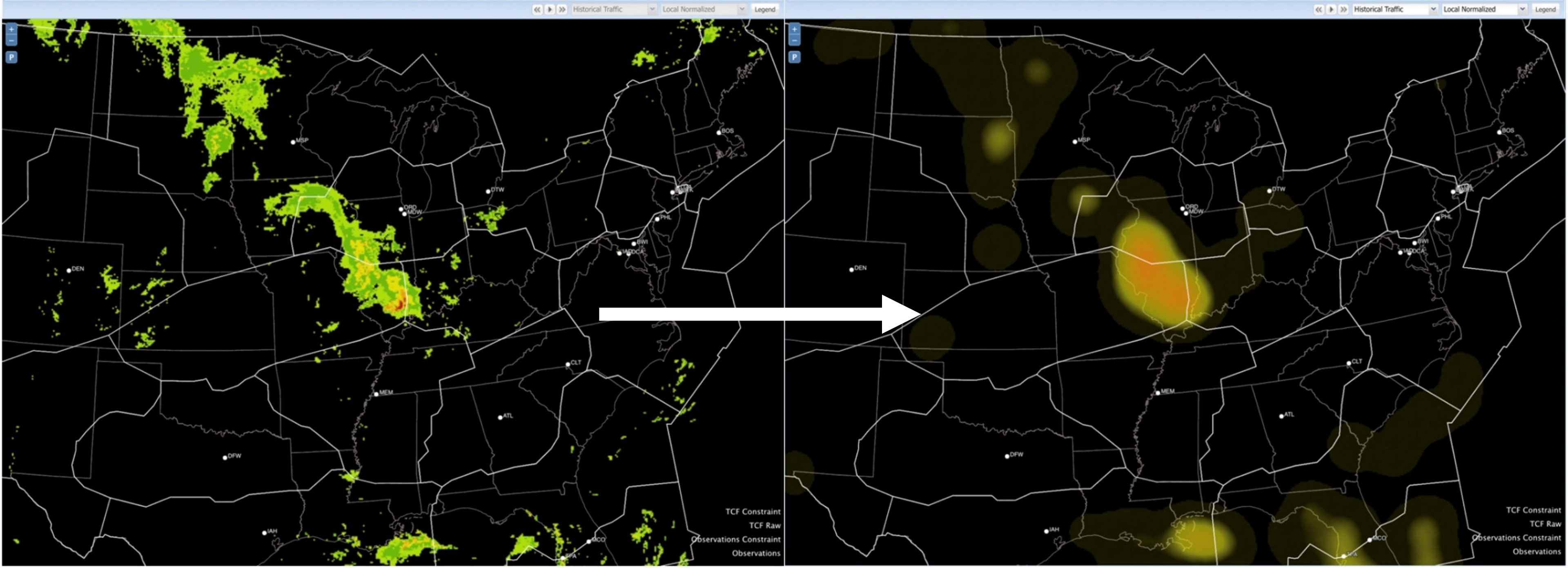
Verification Approaches

Verification includes two perspectives:

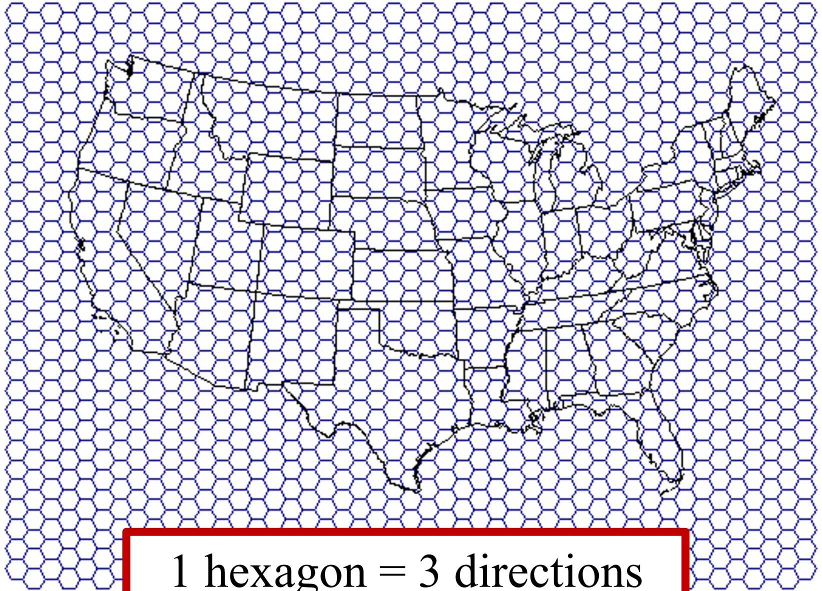
1. Accuracy in constraint space
2. Conditions relative to polygon definitions

Verification Approaches -- Constraint

From weather to constraint

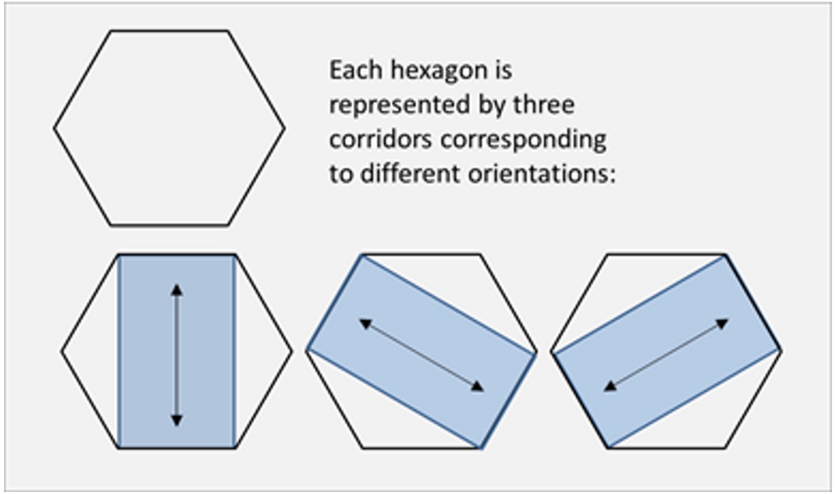


Verification Approaches -- Constraint



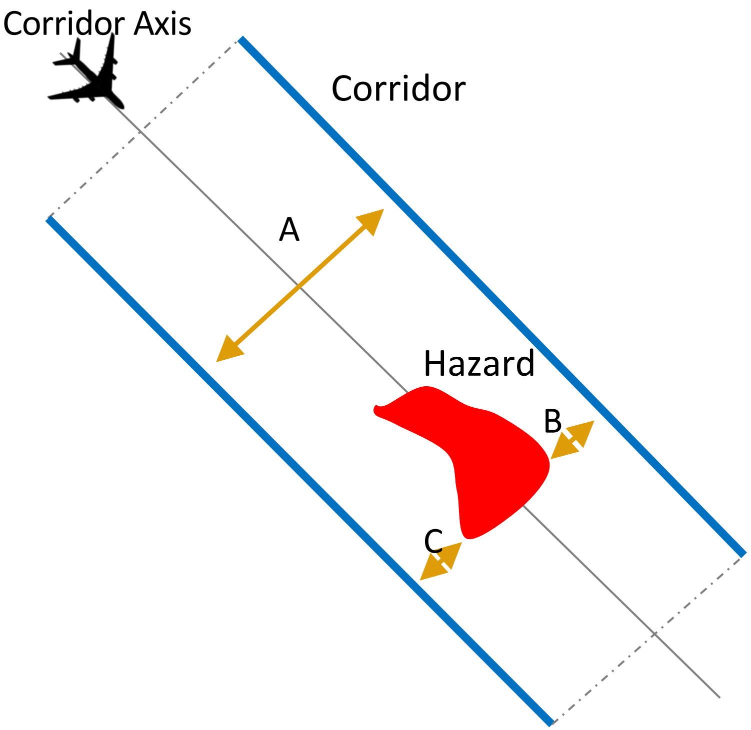
1 hexagon = 3 directions of corridors
Supports computation of traffic in any direction

- Hexagonal Grid with 80 x 40 nmi corridors
  - Approximates width of jet routes
  - Captures sensitivity to orientation of hazard



Verification Approaches -- Constraint

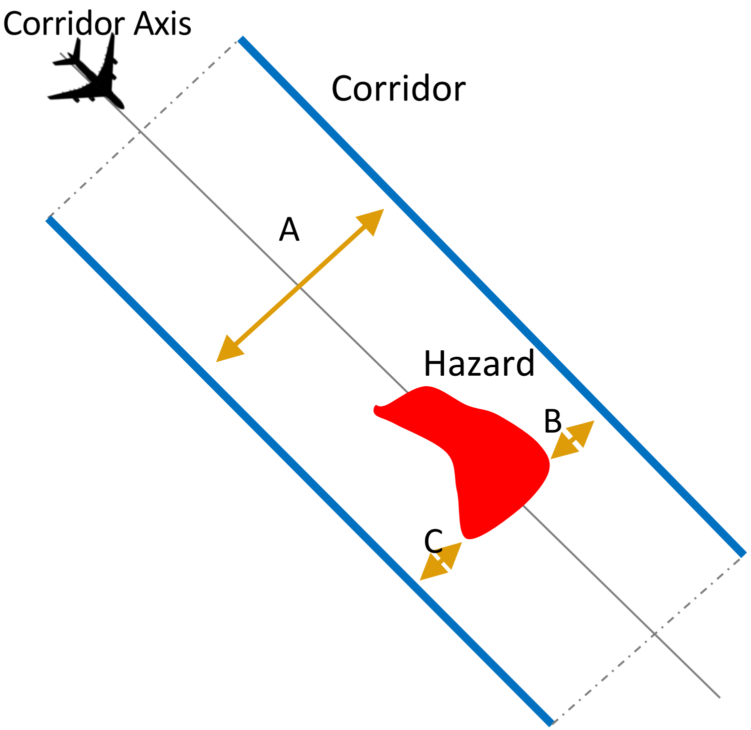
Within each corridor, compare the available space without weather (A) to the space available in the presence of weather (B+C).



Verification Approaches -- Constraint

It is often remarked that polygon placement need not be perfect, just close enough.

Note that moving the hazard in the along-corridor direction will have no effect on the constraint.

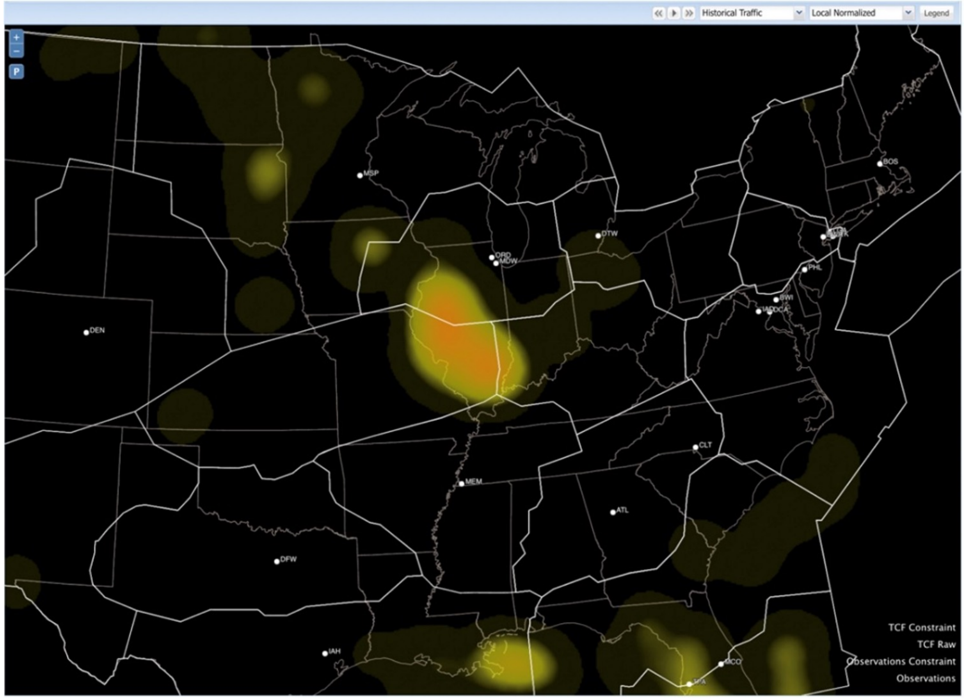


Verification Approaches -- Constraint

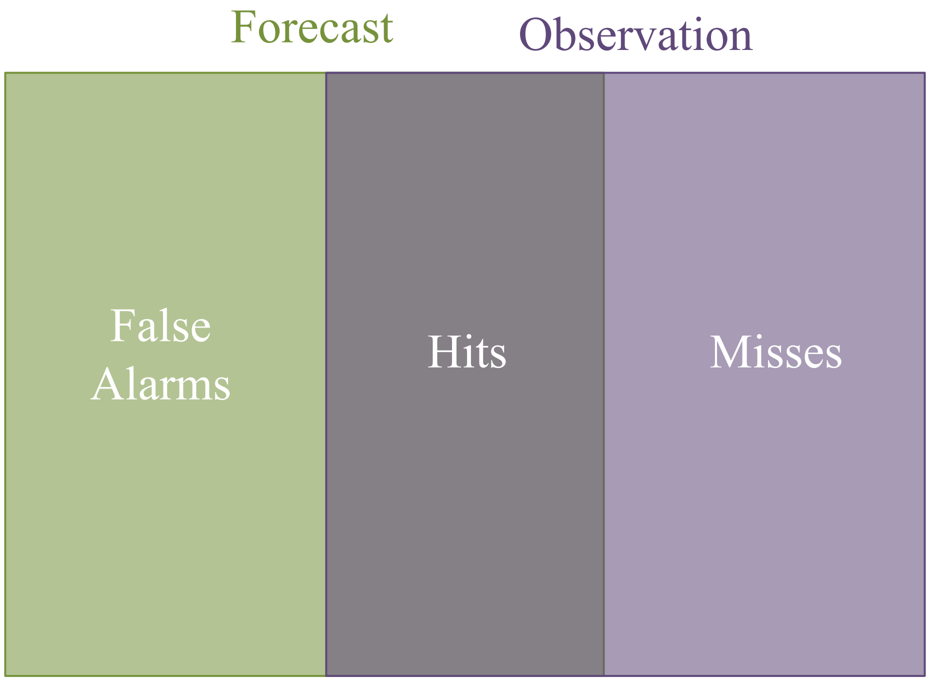
Both TCF polygons and CIWS obs are translated into constraint.

A threshold is then applied resulting in yes/no areas of forecast and observed constraint.

These binary fields are then brought together for verification.

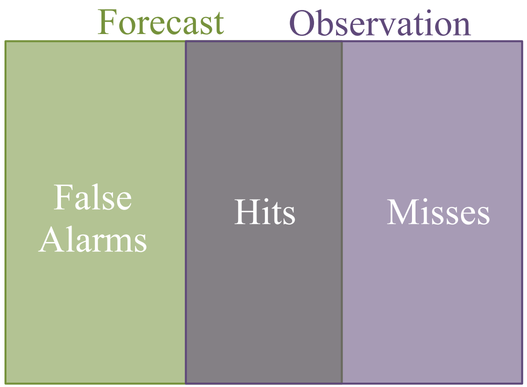


Verification Scores



Verification Scores

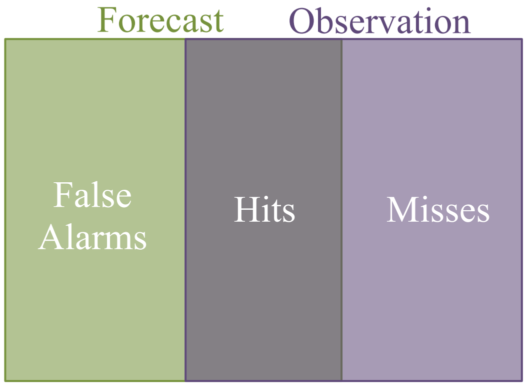
$$\mathrm{POD} = \frac{\text{Hits}}{\text{Observation}}$$



Verification Scores

$$\mathrm{POD} = \frac{\text{Hits}}{\text{Observation}}$$

$$\mathrm{SR} = \frac{\text{Hits}}{\text{Forecast}} = 1 - \mathrm{FAR}$$

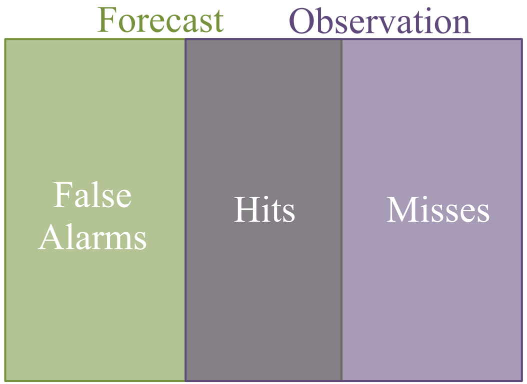


Verification Scores

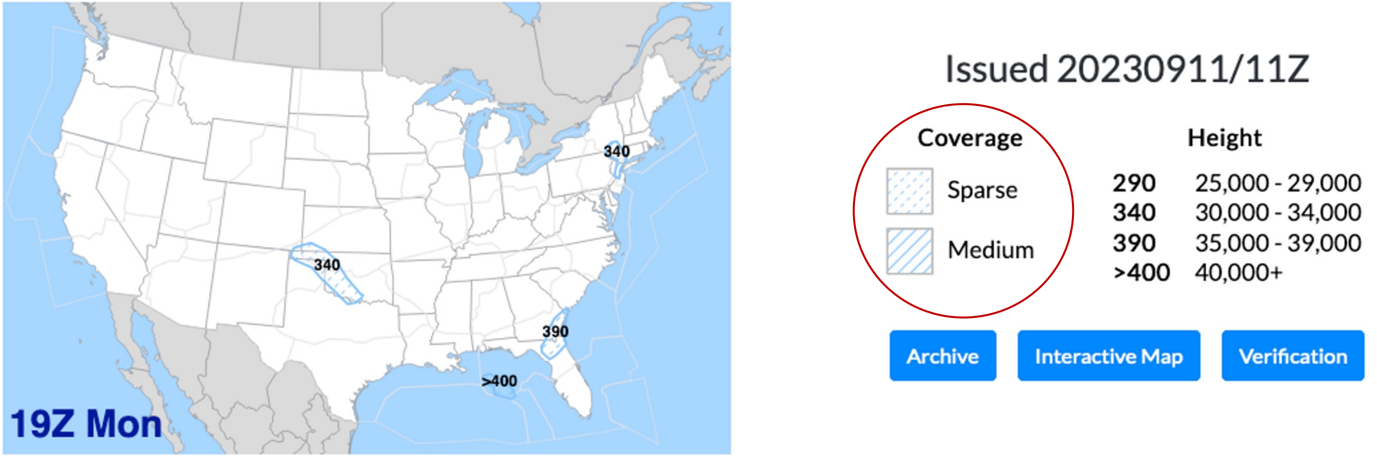
$$\mathrm{POD} = \frac{\text{Hits}}{\text{Observation}}$$

$$\mathrm{SR} = \frac{\text{Hits}}{\text{Forecast}} = 1 - \mathrm{FAR}$$

$$\mathrm{CSI} = \frac{\text{Intersection}}{\text{Union}}$$



Verification by polygon definition

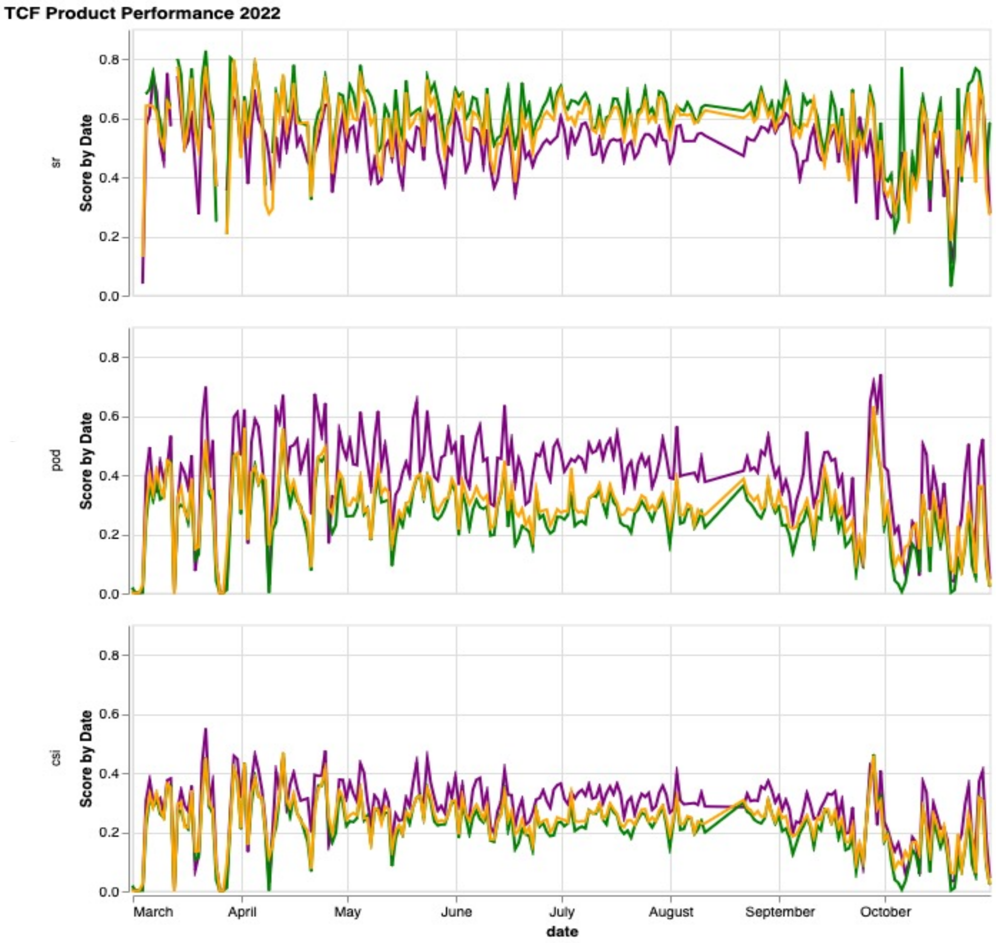


Sparse: 25 – 39% coverage

Medium: ≥ 40%

TCF Performance Results

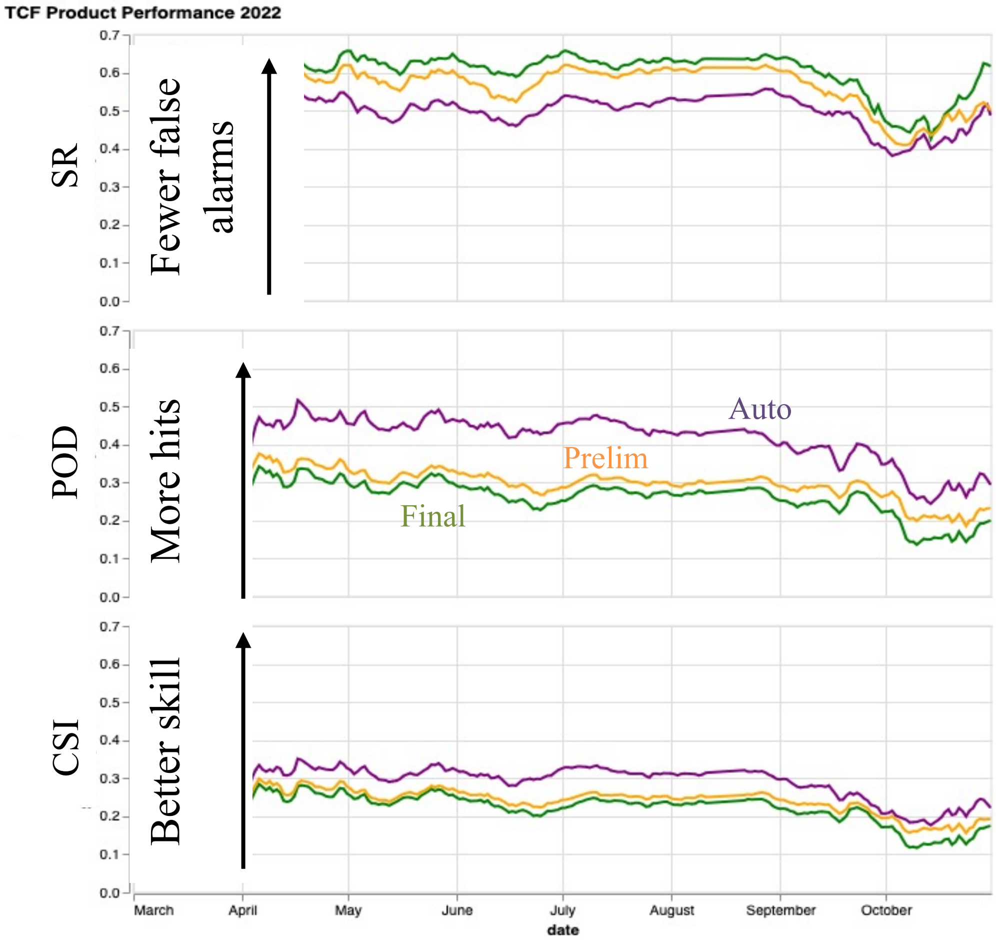
TCF Performance



Too noisy to see trends in the daily scores.

TCF Performance

Plotted scores are averaged over 2 weeks to reduce noise.

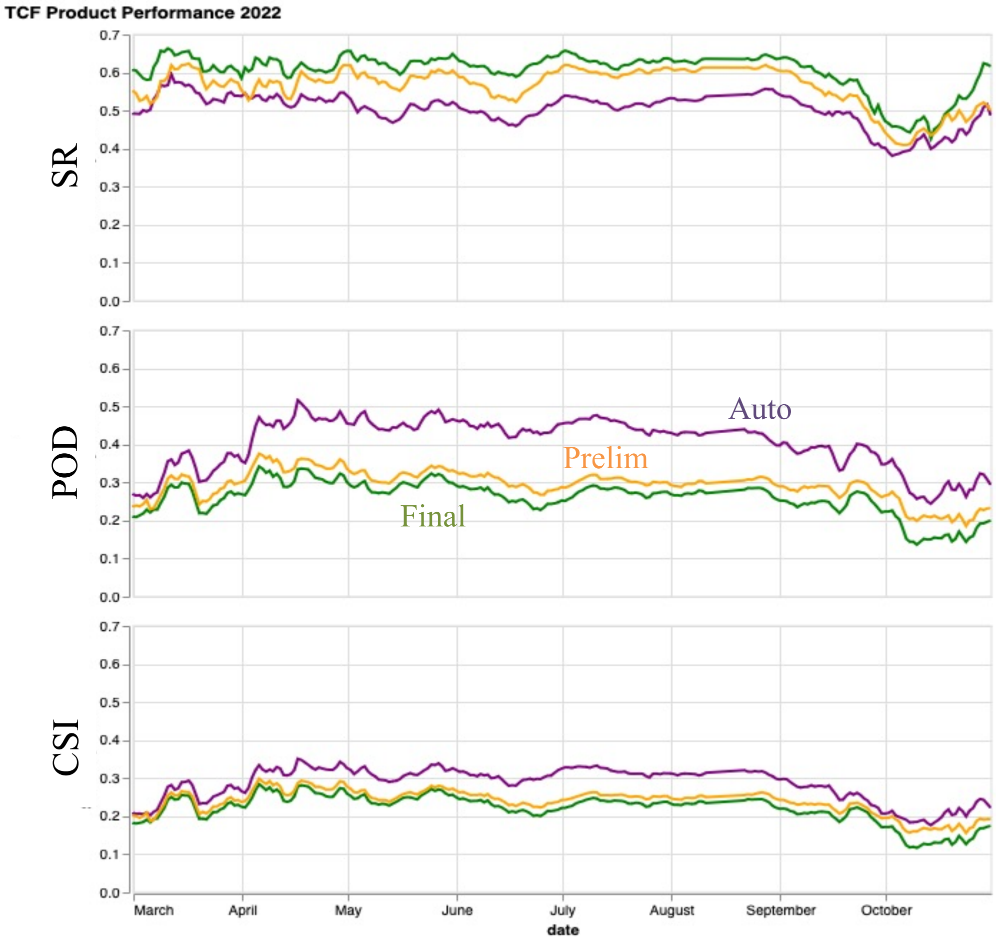


TCF Performance

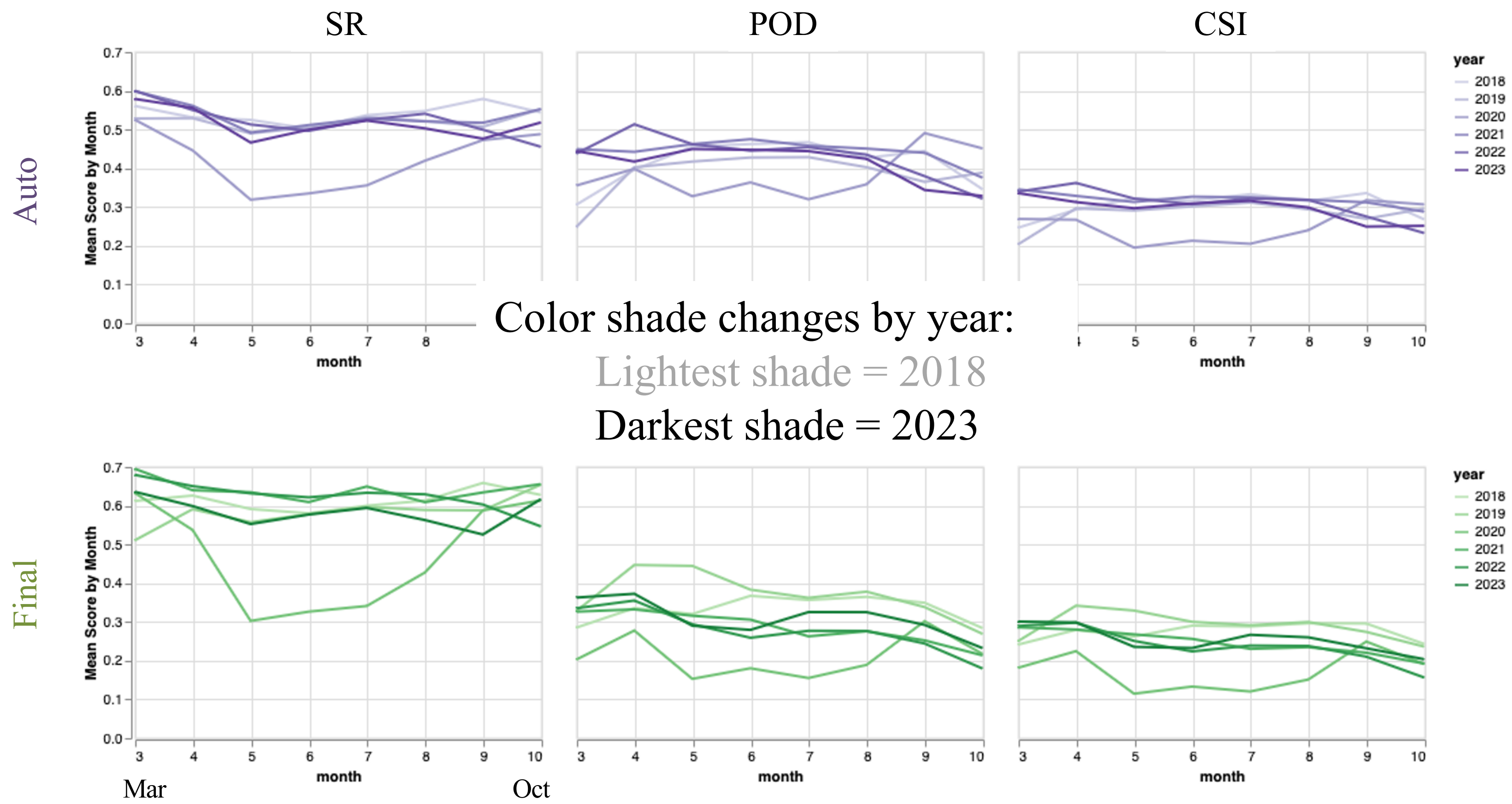
Compared to Auto, Final has
- fewer false alarms (SR is higher)
- fewer hits
- less skill

Prelim lies in between but is (usually) more like the Final than the Auto.

Performance is noisier at the beginning and end of the convective season, fairly constant through the year.

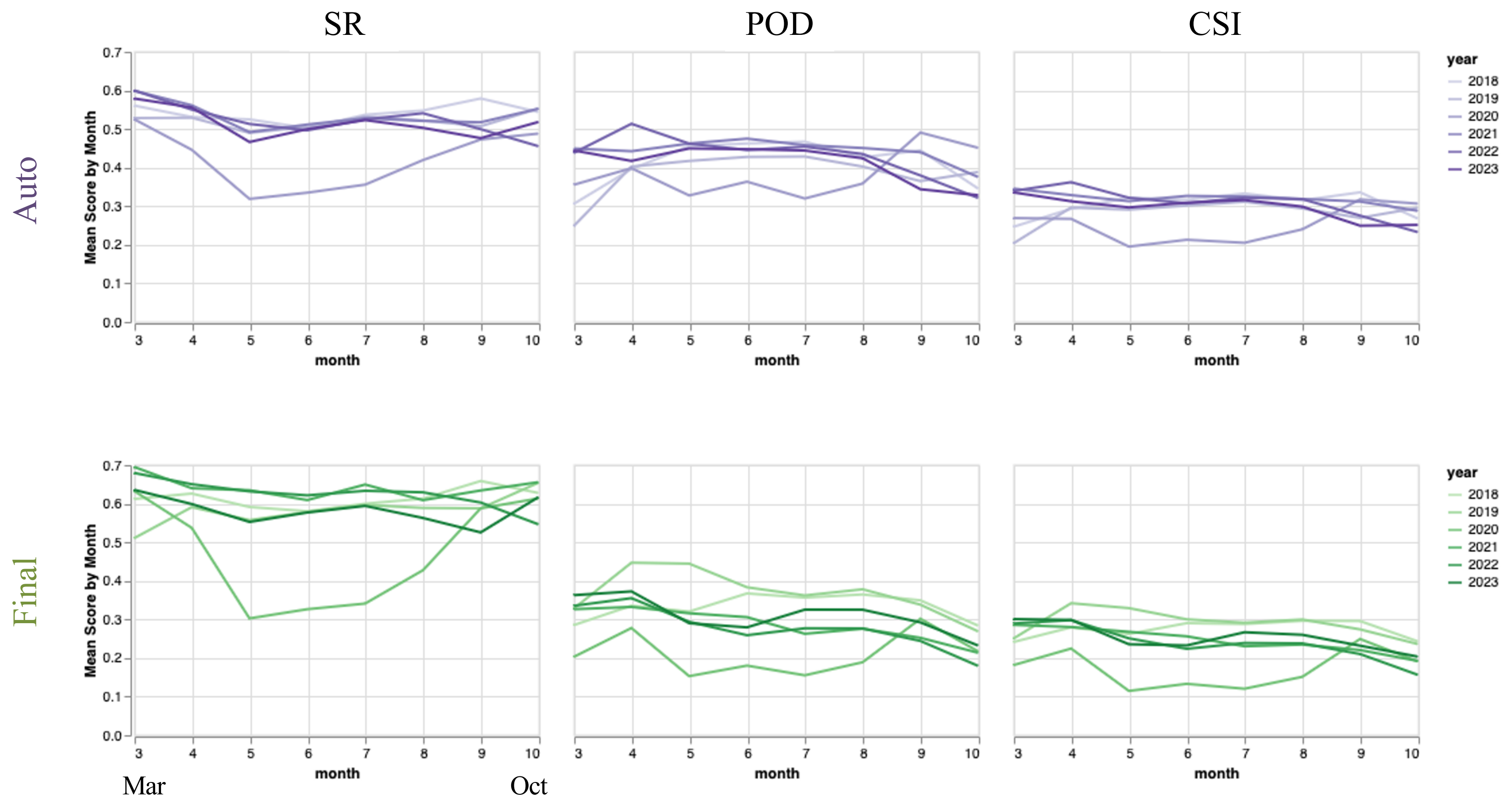


TCF Performance by Month and Year

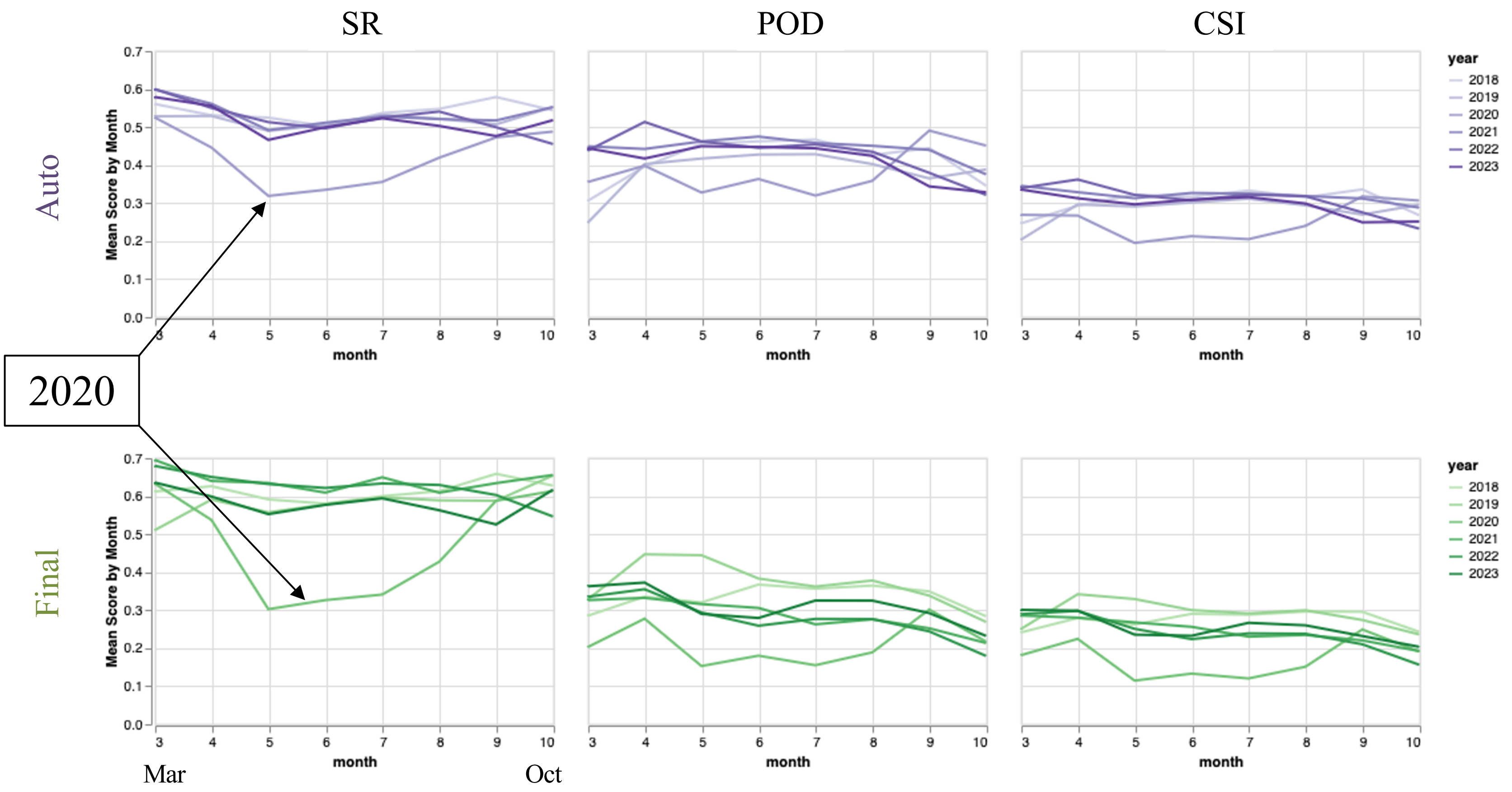


Color shade changes by year:
Lightest shade = 2018
Darkest shade = 2023

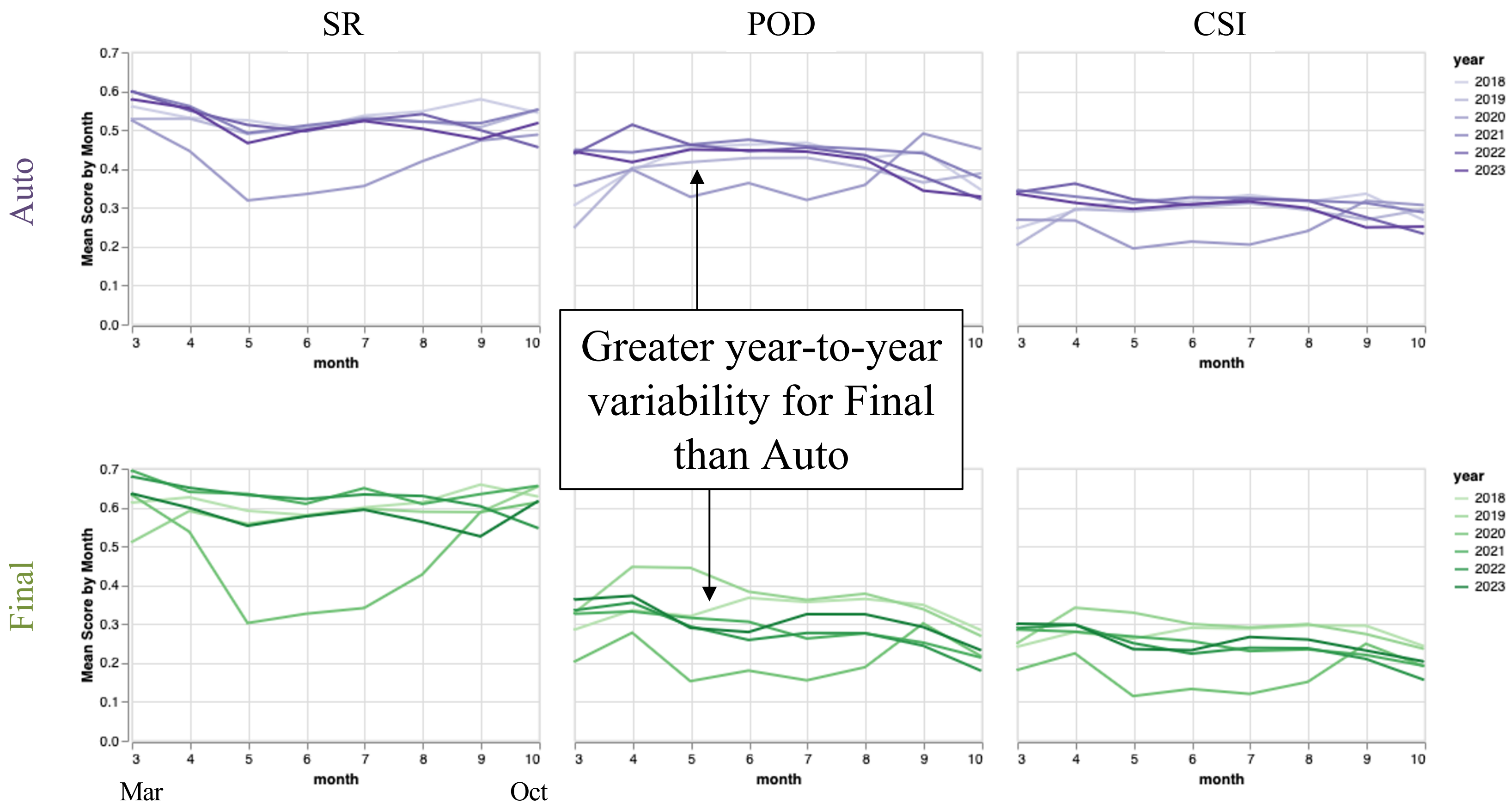
TCF Performance by Month and Year



TCF Performance by Month and Year

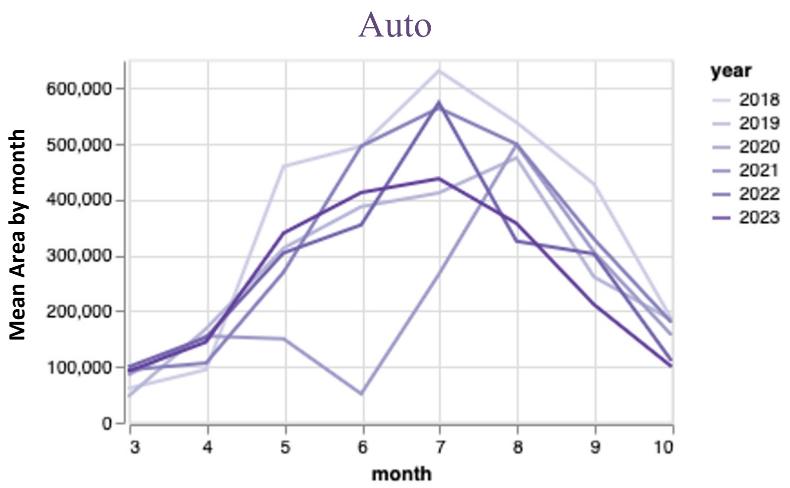


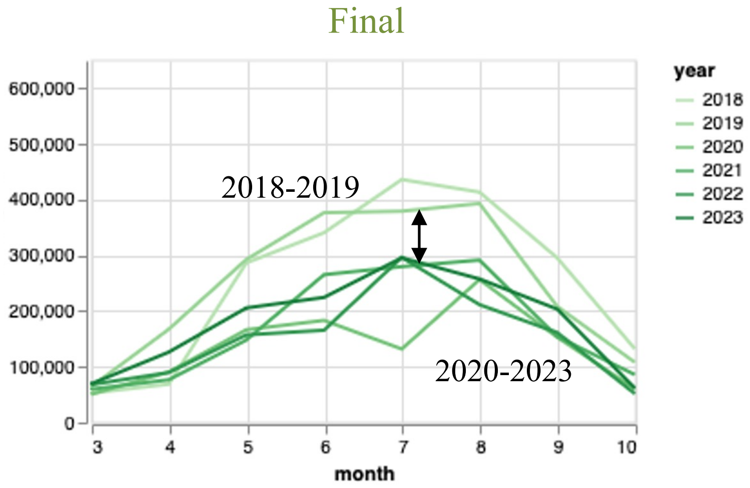
TCF Performance by Month and Year



TCF Area by Month and Year

- Both products have a substantial peak in total area in mid-Summer
- Both products have considerable spread across years
- Spread in Auto appears random; Final has a distinct change after 2019.

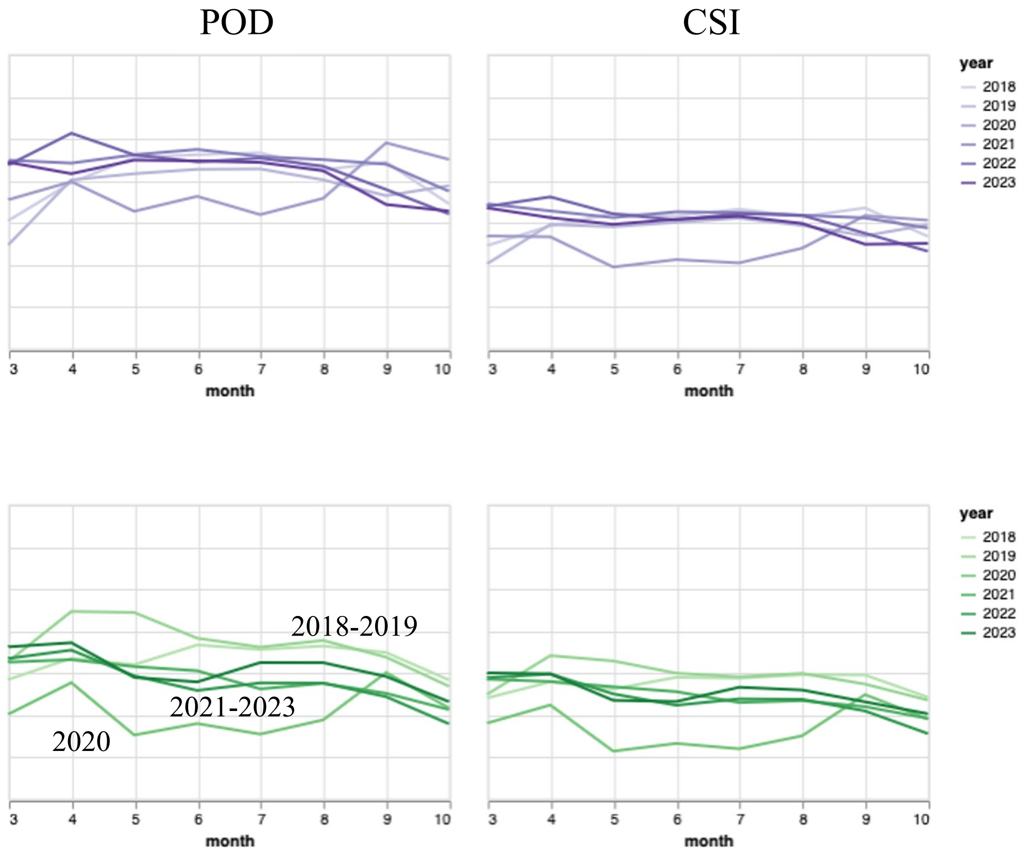




TCF Performance by Month and Year

There is no temporal trend in the Auto performance.

The decrease in average total polygon area after 2019 in the Final led to a drop in amount of convection captured and a smaller drop in skill.



TCF Performance

So the Auto consistently outperforms Final because the larger polygon areas capture more storms.

But is this the whole story?

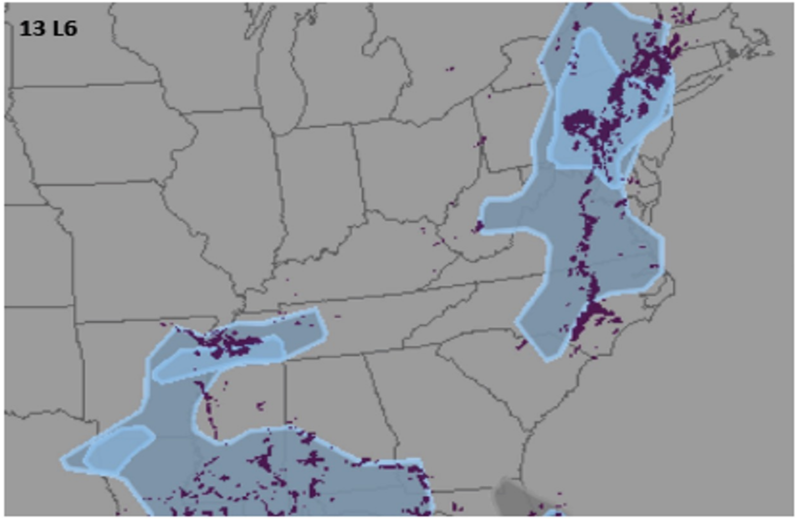
1. CSI assumes that missed events (important convection outside of polygons) and false alarms (empty space inside polygons) are equally undesirable. Is that accurate?
2. Let’s look more at the false alarm space, i.e., polygon coverage.

Case study highlighting Auto empty polygon

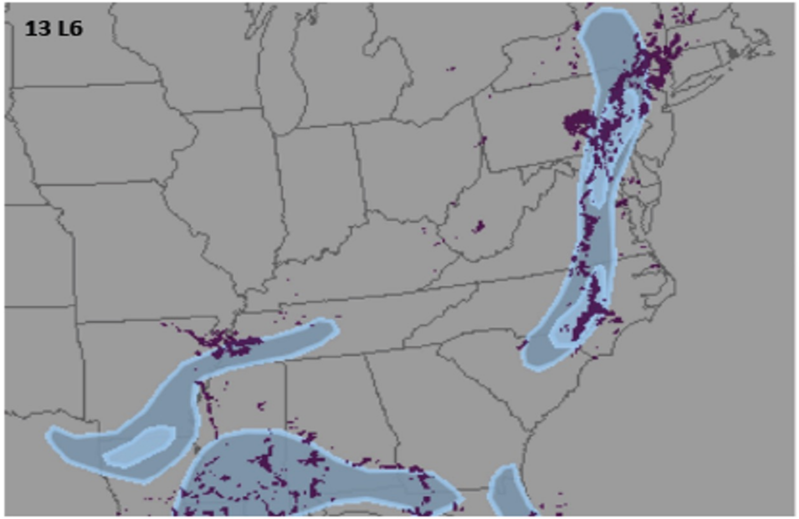
9 July

Convection widespread from New England down to NC. The narrower Final polygons better communicate the weather scenario. Both products capture the scattered storms along the Gulf Coast. And the atmosphere curiously failed to produce the convection over northwest LA expected by both Auto and Final.

Auto



Final

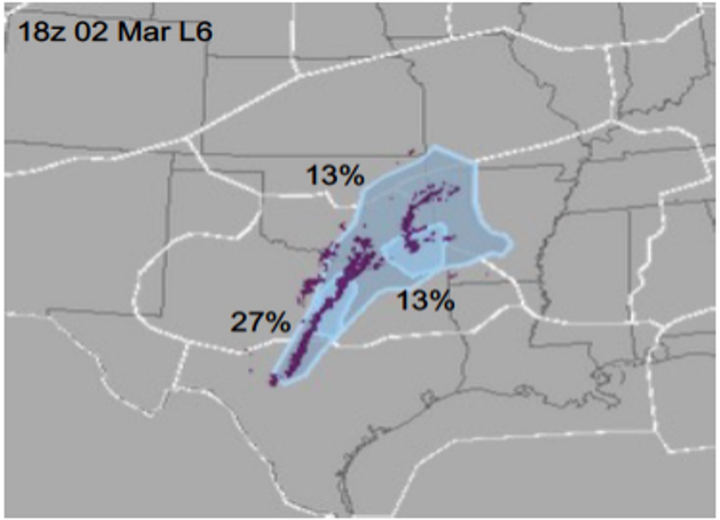


Another case study highlighting false alarm space

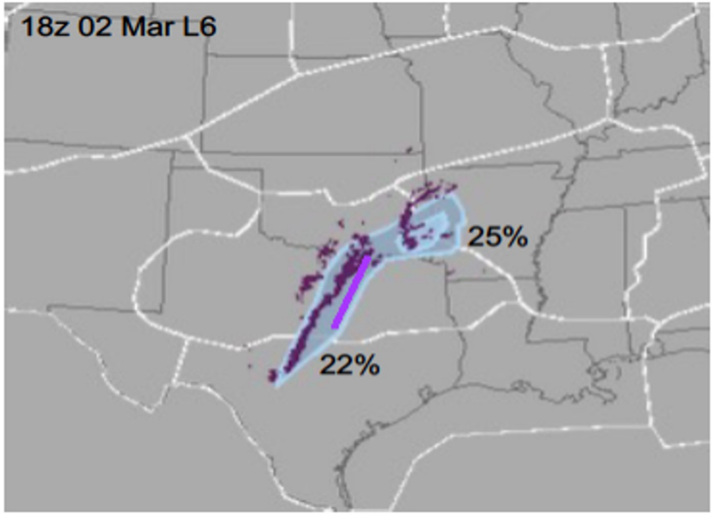
2 March

Auto split the Medium coverage area (from previous forecast) to poor effect. Final decreased the Medium coverage area but missed orientation of the storms. Both capture the main line but missed the trailing convection. The line forecast verified, but what is the proper interpretation?

Auto



Final

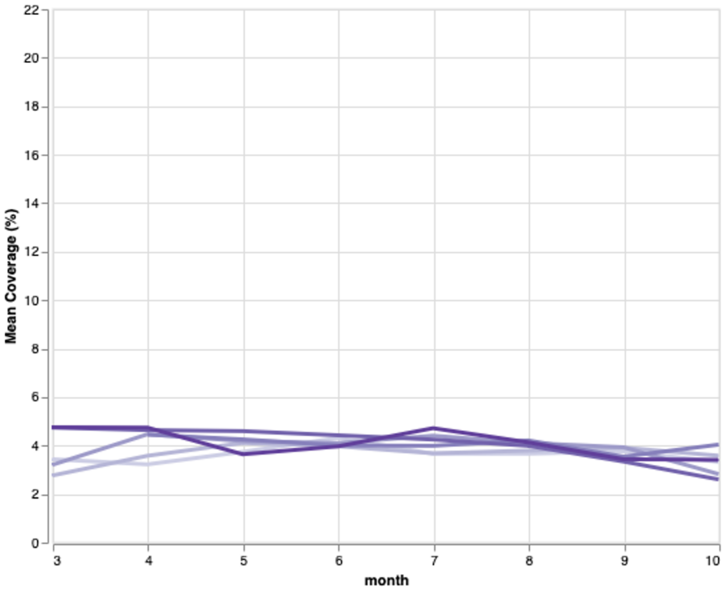


TCF coverage: Sparse

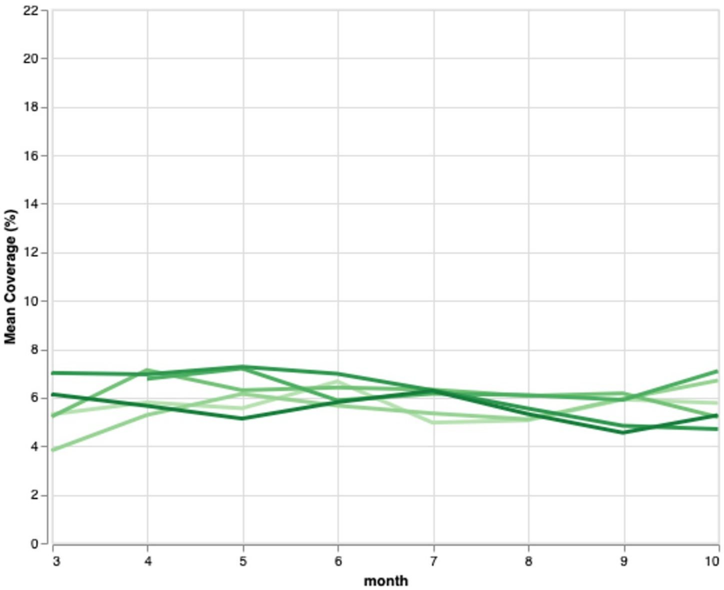
Coverage in Final polygons is about 50% greater than in Auto polygons.

Both are well below the official threshold (25%).

Auto



Final

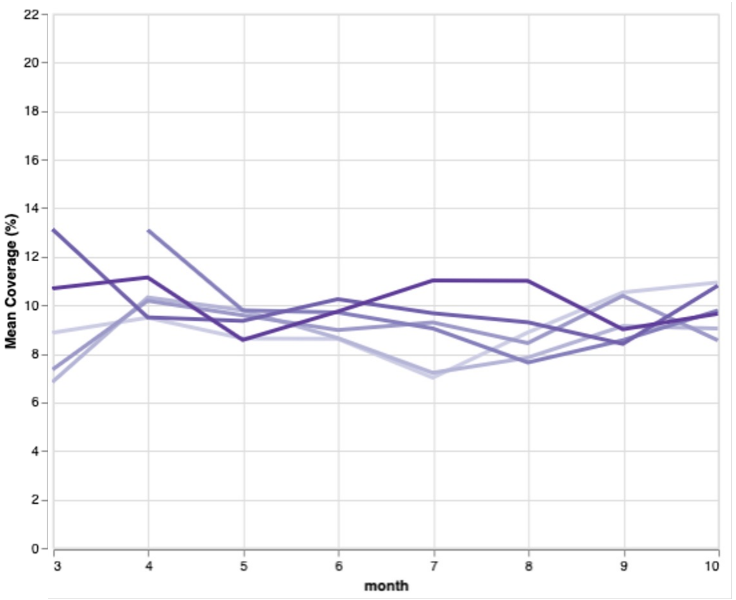


TCF coverage: Medium

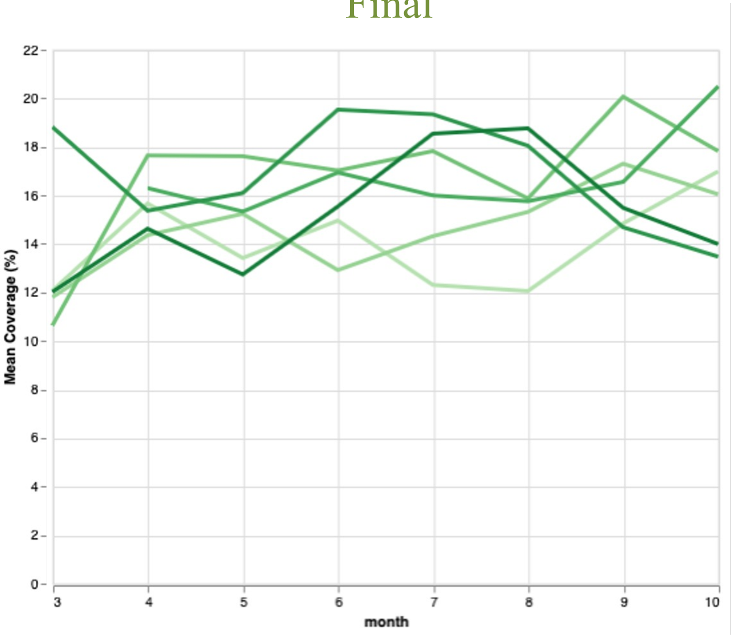
Coverage in Final polygons is about 50% greater than in Auto polygons, again.

Coverage in Medium polygons is 2-4 times greater than for Sparse polygons, but again, well below the official threshold (40%).

Auto

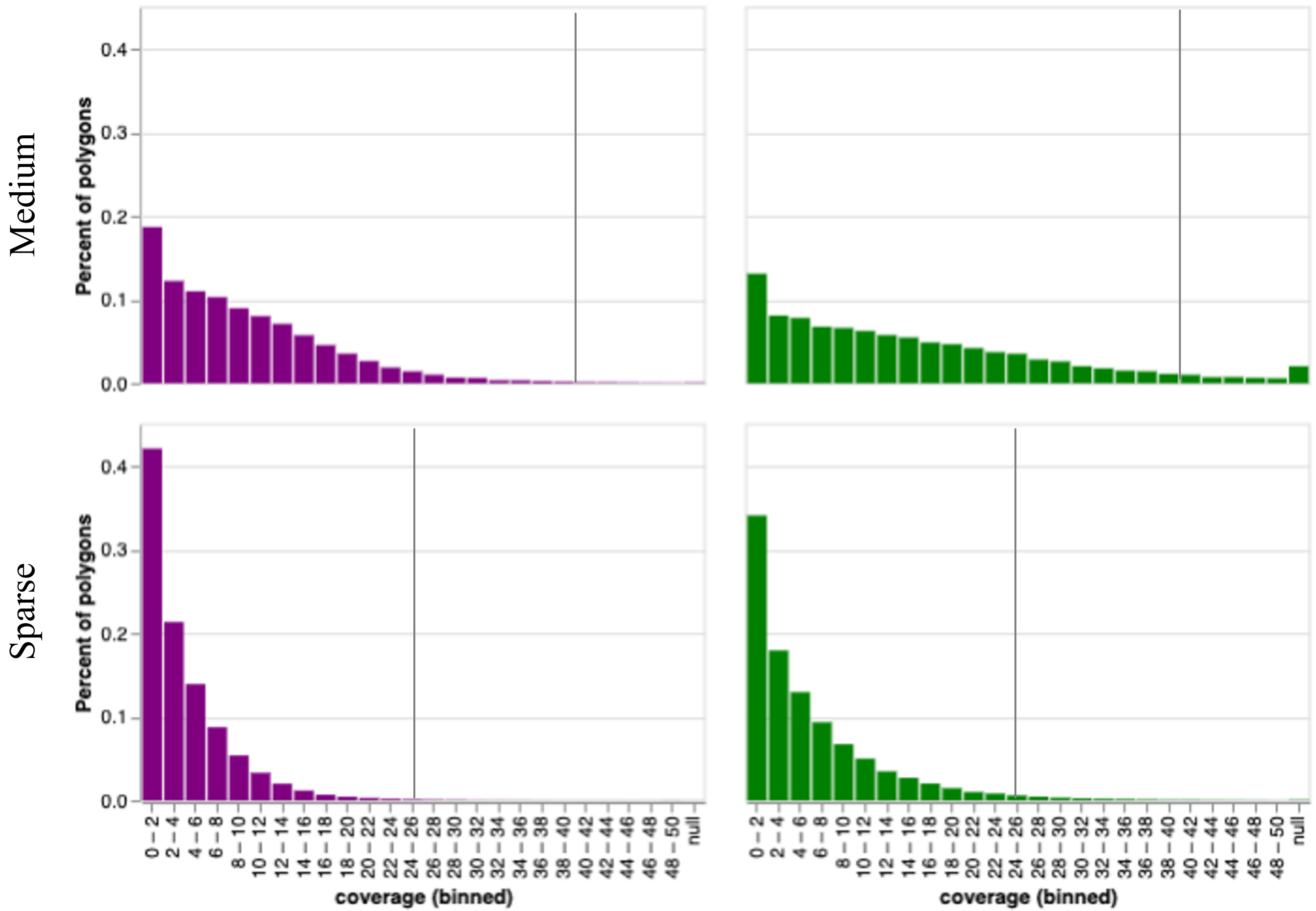


Final



TCF polygon coverage distributions

The larger polygons in the Auto are more likely to be completely empty—nearly 20% of Medium polygons have coverage < 2%—and less likely to meet threshold.



Summary

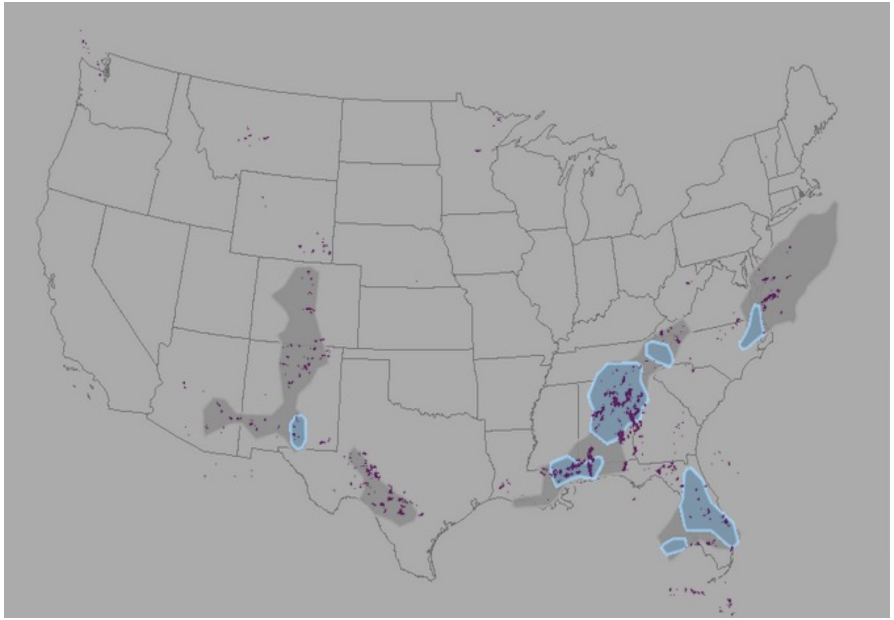
- Performance has changed little over the last six years.
  - Except the shift in the Final after 2019 toward smaller/fewer polygons, leading to a slight drop on skill
- Final forecasts reduce the airspace within a polygon by a factor of 2-3, relative to the Auto (post 2019).
- The smaller Final polygons mean, relative to Auto
  - Fewer hits, lower skill
  - Greater convective coverage within polygons (~50% greater)—less cautioned airspace that was perfectly good to fly through
- Nearly all polygon coverages are well below the official (25, 40%) thresholds.
  - Only 1 in 5 Final Medium polygons meet the Sparse coverage criterion
  - Only 1 in 20 Auto Medium polygons meet the Sparse coverage criterion

Extras

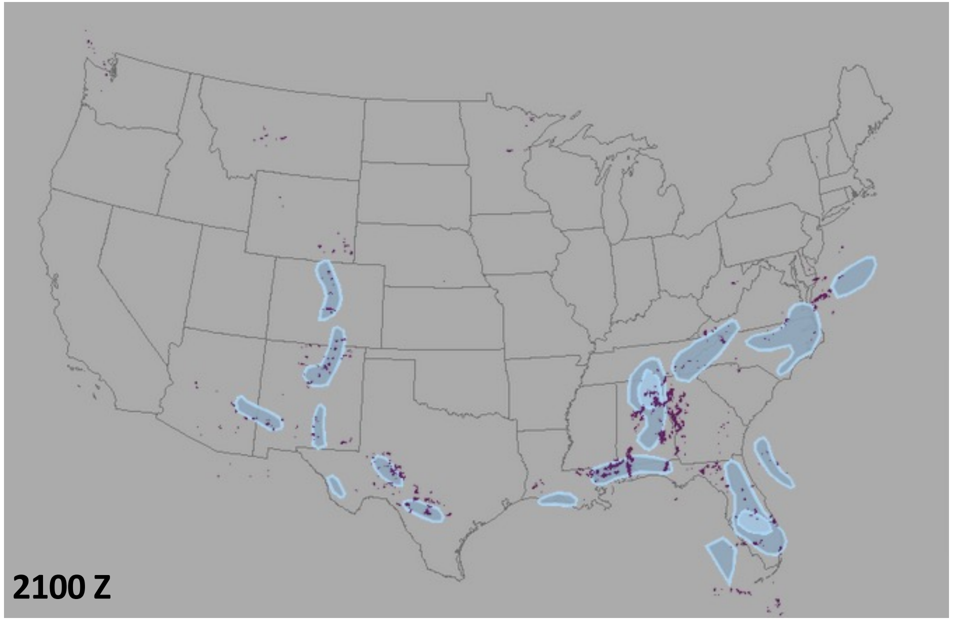
28 Aug—T-storms from mid-Atlantic to Gulf Coast to the Rockies

Twelve (12) polygons in the Auto; 18 in the Final.
The larger (and low confidence) polygons arguably capture the convective activity better, but is that desired more broadly?
Is there a good solution to a day like this? (The measured performance for this day was about average for the period.)

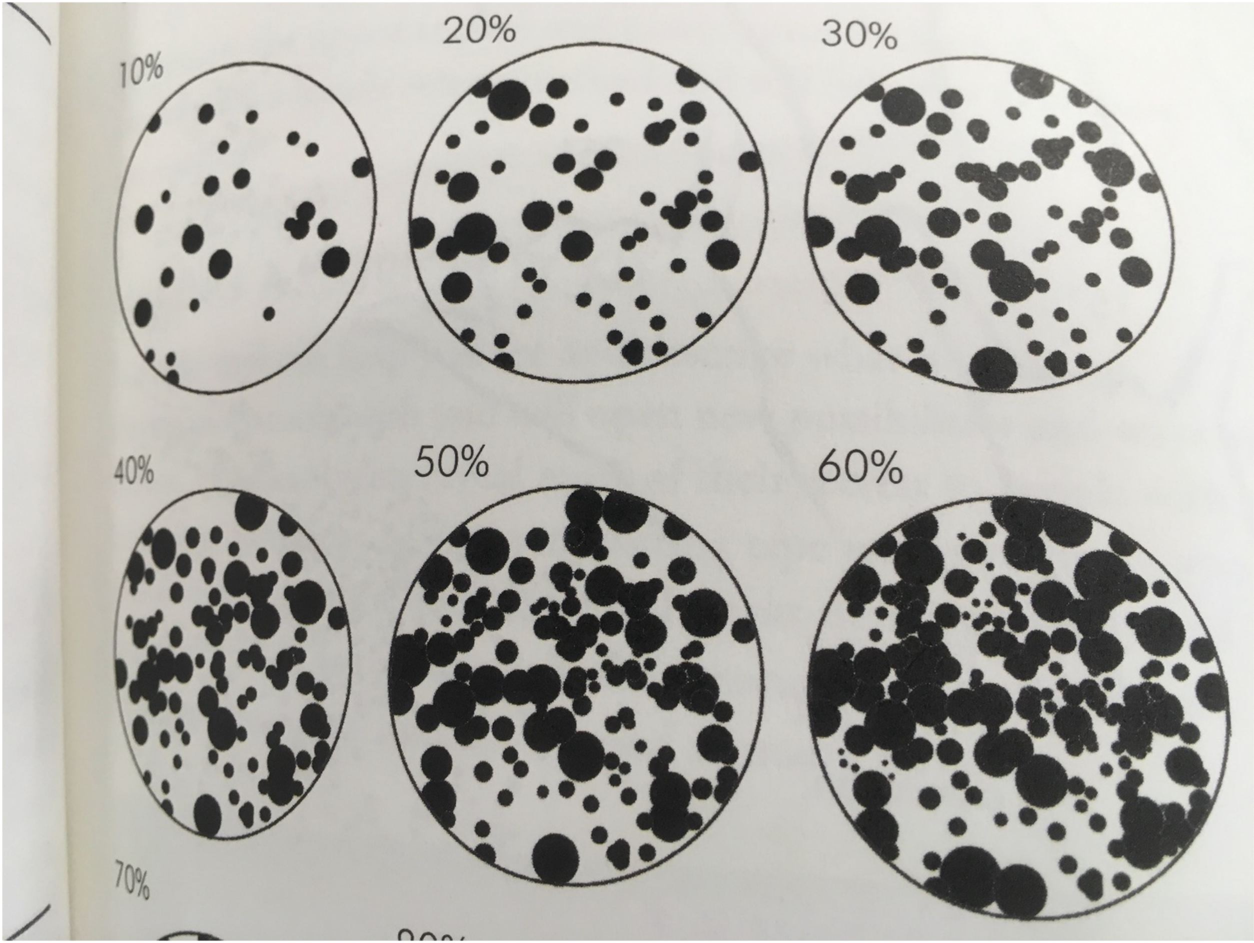
Auto



Final



Coverage examples



Polygons meeting coverage criteria

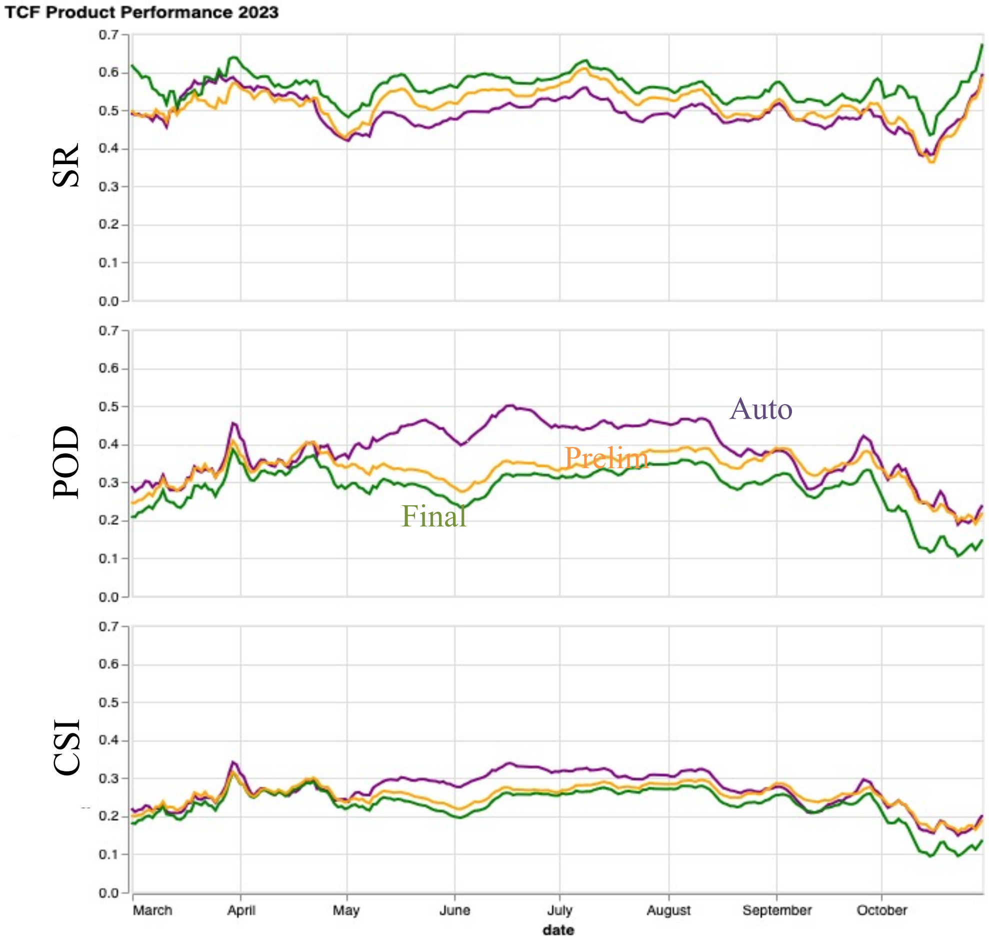
```
Final
sparse: yes=3155  tot=155804  %yes=2.025
medium: yes=1426  tot= 24816  %yes=5.746
med_sp: yes=5168  tot= 24816  %yes=20.825

Auto
sparse: yes= 503  tot=186738  %yes=0.269
medium: yes= 156  tot= 36152  %yes=0.432
med_sp: yes=1623  tot= 36152  %yes=4.489

Final  with at least 25% coverage = 4.608
Auto   with at least 25% coverage = 0.954
```

TCF Performance

2023



Verification Scores