NATIONAL WEATHER SERVICE

Towards Development of a Seasonal Verification System at the Environmental Modeling Center (EMC)

Aman Vashisht[2,1], Alicia Bentley[1], Shelley Melchior[2,1], Shannon Shields[3,1], Jason Levit[1]

[1]NOAA/NWS/NCEP/EMC [2]Lynker [3]SAIC

7 JUNE 2024

2024 S2S Community Workshop Session: Diagnostics, Validation, and Verification

Outline

- Introduction
- EMC Verification System (EVS) Overview
- EVS v2.0 - Seasonal component
- Seasonal Verification Variables
- Seasonal Verification Metrics and Validation Data
- Conclusion

Introduction

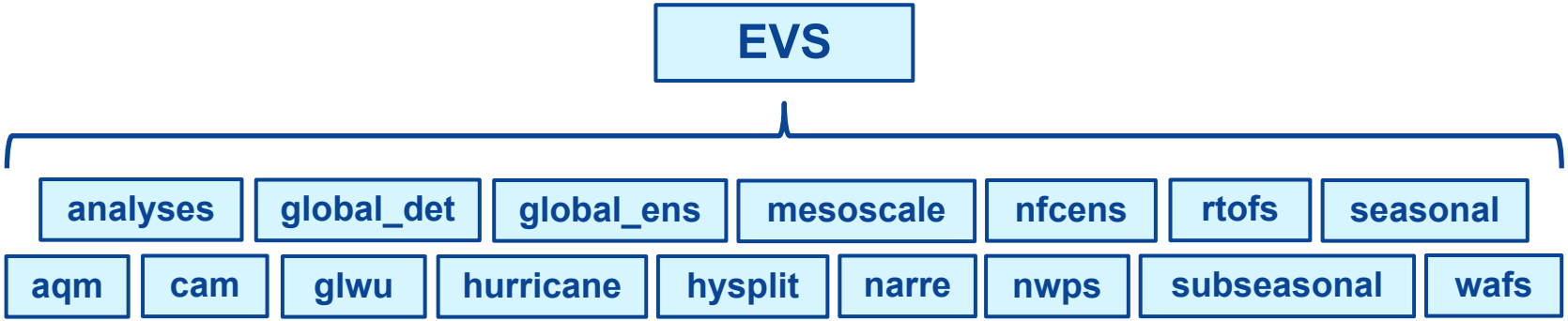
- Seasonal forecasts are increasingly being requested and utilized for water management, agriculture, energy markets, construction, insurance, tourism, financial markets etc.
    - Demoine & Kapnick (2024; Nature Communications): Investors wait for, or hedge against, the release of NOAA's Seasonal Winter and ENSO forecasts/outlooks, with each outlook affecting firms throughout the economy that have a total market capitalization of $6 trillion (Winter) and $13 trillion (ENSO)
- Seasonal verification is critical to assess models and their performance
- The Environmental Modeling Center (EMC) has recently created the EMC Verification System, which is now being updated to include a new Seasonal verification component

Introduction

- Seasonal verification timescale: Month to Year
- NCEP models that are forecasting or will be forecasting at this timescale:
  - Climate Forecast System version 2 (CFSv2) - current operational version of the CFS is version 2.3 as of 9 March 2022
  - Seasonal Forecast System version 1 (SFSv1) - currently in development

EMC Verification System (EVS)

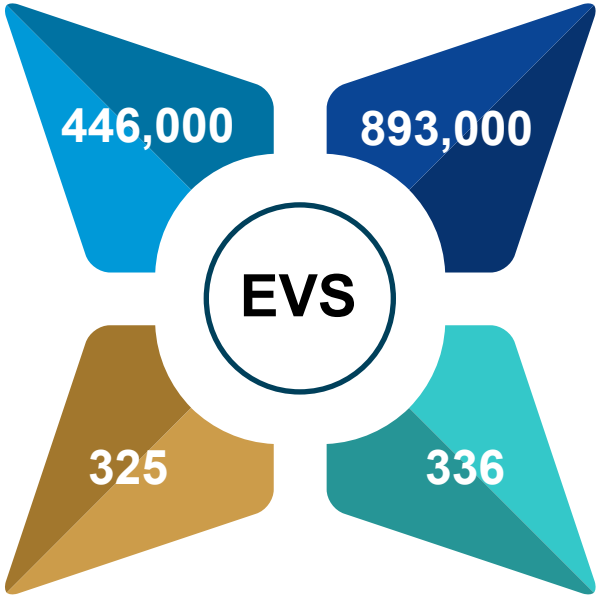
- EVS is a new software system used to assess **operational** NCEP model performance
- EVS routinely creates verification 1) statistics and 2) graphics in NCEP operations, allowing EMC to monitor **operational** NCEP model performance in near real time
  - Utilizes the Model Evaluation Tools (METplus) software package from DTC
- EVS will replace and expand upon the legacy verification software in NCEP operations



EMC Verification System (EVS) v1.0

~446,000 lines of code developed

~893,000 images generated per day



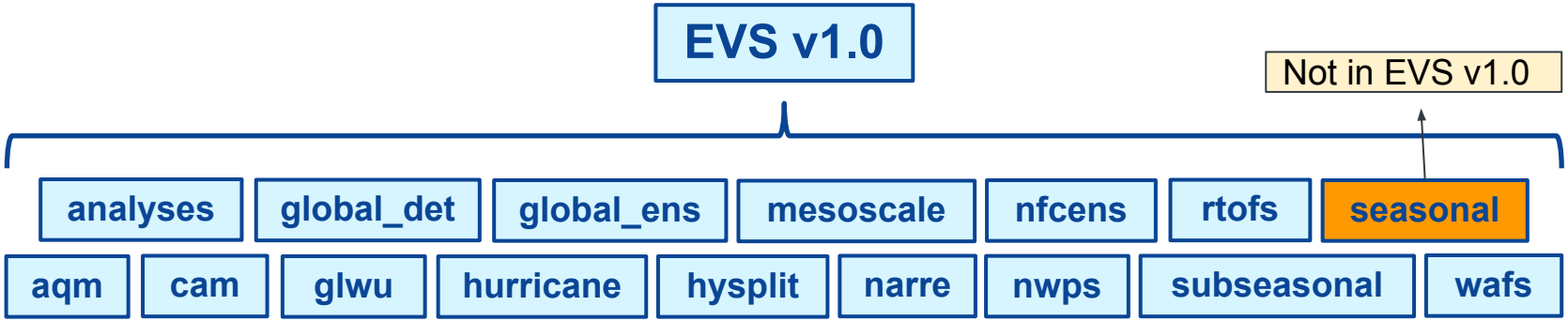
325 variables verified across the production suite

336 merged pull requests on Github

- Code Delivered to NCO: 8 December 2023
- 30-day IT Stability Test: 23 February 2024–24 March 2024
- Implement into Operations: 26 March 2024

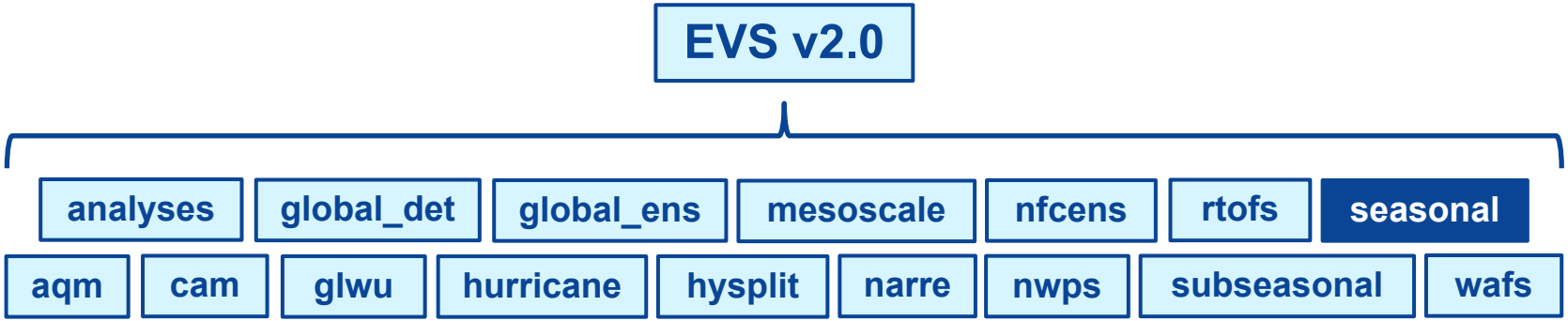
EMC Verification System (EVS) v1.0

- EVS is a new software system used to assess **operational** NCEP model performance
- EVS routinely create verification 1) statistics and 2) graphics in NCEP operations, allowing EMC to monitor **operational** NCEP model performance in near real time
  - Utilizes the Model Evaluation Tools (METplus) software package from DTC
- EVS will replace and expand upon the legacy verification software in NCEP operations



EMC Verification System (EVS) v2.0

- EVS v2.0 will expand upon the verification included in EVS v1.0
- EVS will include the **Seasonal verification** component to assess **operational** NCEP model performance
- Like other components, it will use **Model Evaluation Tools (METplus)** software package developed by DTC



EVS Seasonal Verification

- The seasonal component of EVS v2.0 will verify seasonal forecasts from operational models
    - Will verify CFS forecasts ranging from 1–9 months
    - CFS operational forecasts verification will serve as a baseline for comparison with future operational seasonal models (like the SFS which will produce forecasts up to 12 months)

- The verification metrics for the seasonal component of EVS will be derived from the 2021 DTC UFS Metrics Workshop

- Temporal resolution of seasonal verification: Monthly, 3–Monthly

- Spatial regions of seasonal verification: Tropics, CONUS, NH, SH, Arctic/Antarctic, Global etc.

EVS Seasonal Verification: Verification Variables

- A total of 30 verification variables have been planned to be included in the Seasonal component of EVS

- However, some of them are not included/coded in METplus yet, so the Seasonal component of EVS v2.0 will focus on **~18 verification fields/variables**

- Here are a few examples of variables that are important at the seasonal scale that will be included:
  - Oceanic Niño Index (ONI)
  - Sea Surface Temperatures (SST)
  - Sea Ice Concentration…etc.

- The next few slides will go into some depth for different verification fields, metrics and validation data

EVS Seasonal Verification: Verification Variables, Metrics and Validation Data

- **Oceanic Niño Index**

Model data: CFS, SFS (once in ops)	Validation data: GHRSST OSPO (Group for High Resolution Sea Surface Temperature Office of Satellite and Product Operations)
Model data resolution available: Monthly	Validation data resolution available: Daily

Metrics: RMSE, Bias
Regions: Niño 3.4

- **Sea Surface Temperatures**

Model data: CFS, SFS (once in ops)	Validation data: GHRSST OSPO
Model data resolution available: Monthly	Validation data resolution available: Daily

Metrics: RMSE, Bias
Regions: Tropics, NH, SH

EVS Seasonal Verification: Verification Variables, Metrics and Validation Data

- **Sea Ice Concentration**

Model data: CFS, SFS (once in ops)	Validation data: OSI-SAF*
Model data resolution available: Monthly	Validation data resolution available: Daily

Metrics: RMSE, Bias, Critical Success Index (CSI), Performance Diagram
Regions: Arctic, Antarctic

*Ocean and Sea Ice Satellite Application Facility

- **Sea Ice Edge**

Model data: CFS, SFS (once in ops)	Validation data: OSI-SAF
Model data resolution available: Monthly	Validation data resolution available: Daily

Metrics: Integrated Ice Edge Error (IIEE) (area where forecast and "truth" disagree on ice conc. being >/< 15%,i.e., sum of all areas where local sea ice extent is overestimated or underestimated; Goessling et al. 2016)
Regions: Arctic, Antarctic

EVS Seasonal Verification: Verification Variables, Metrics and Validation Data

- **OLR, OLR Anomalies**

Model data: CFS, SFS (once in ops)	Validation data: UMD OLR Analysis
Model data resolution available: Monthly	Validation data resolution available: Monthly

Metrics: RMSE,Bias,Anomaly Correlation Coefficient(ACC) [Climatology:CPC OLR Climo]
Regions: Global, 40 N - 40 S

- **Precipitation, Precipitation Anomalies**

Model data: CFS, SFS (once in ops)	Validation data: CCPA*, MRMS QPE*
Model data resolution available: Monthly	Validation data resolution available: Daily

Metrics: RMSE, Bias, Equitable Threat Score (ETS), Heidke Skill Score (HSS), Fractional Skill Score (FSS), Performance Diagram [Climatology needed]
Regions: CONUS, Alaska, Hawaii

*Climatology Calibrated Precipitation Analysis
*Multi-Radar Multi-Sensor Quantitative Precipitation Estimate

EVS Seasonal Verification: Verification Variables, Metrics and Validation Data

- **850-hPa Temperature and Temperature Anomalies**

Model data: CFS, SFS (once in ops)	Validation data: GFS Analysis
Model data resolution available: Monthly	Validation data resolution available: Daily

Metrics: ACC, Heidke Skill Score (HSS) [Climatology needed]
Regions: NH, SH, Tropics

- **2-meter Temperature and Temperature Anomalies**

Model data: CFS, SFS (once in ops)	Validation data: ECMWF Analysis, METARS*
Model data resolution available: Monthly	Validation data resolution available: 6-hourly

Metrics: RMSE, Bias, ACC, HSS [Climatology needed]
Regions: CONUS, NH, SH

*Meteorological Aerodrome Reports

EVS Seasonal Verification: Verification Variables, Metrics and Validation Data

- **500-hPa Geopotential Height and Geopotential Height Anomalies**

Model data: CFS, SFS (once in ops)	Validation data: GFS Analysis
Model data resolution available: Monthly	Validation data resolution available: Daily

Metrics: RMSE, Bias, ACC, Heidke Skill Score (HSS) [Climatology needed]
Regions: NH, SH, Tropics

- **850-hPa and 200-hPa U/V wind and 10-hPa Zonal Stratosphere wind**

Model data: CFS, SFS (once in ops)	Validation data: GFS Analysis
Model data resolution available: Monthly	Validation data resolution available: Daily

Metrics: RMSE, Bias
Regions: NH, SH, Tropics for U/V and 60 N for stratosphere zonal wind

EVS Seasonal Verification: Verification Variables, Metrics and Validation Data

- **Snow Accumulation**

Model data: CFS, SFS (once in ops)	Validation data: NOHRSC Snowfall Analysis
Model data resolution available: Monthly	Validation data resolution available: Daily

Metrics: Heidke Skill Score (HSS)
Regions: CONUS and CONUS regions

- **Blocking using Pelly-Hoskins Method**

Model data: CFS, SFS (once in ops)	Validation data: GFS Analysis
Model data resolution available: Monthly	Validation data resolution available: Daily

Metrics: RMSE, ACC, Ranked Probability Skill Score (RPSS)
Regions: NH, SH

*National Operational Hydrologic Remote Sensing Center

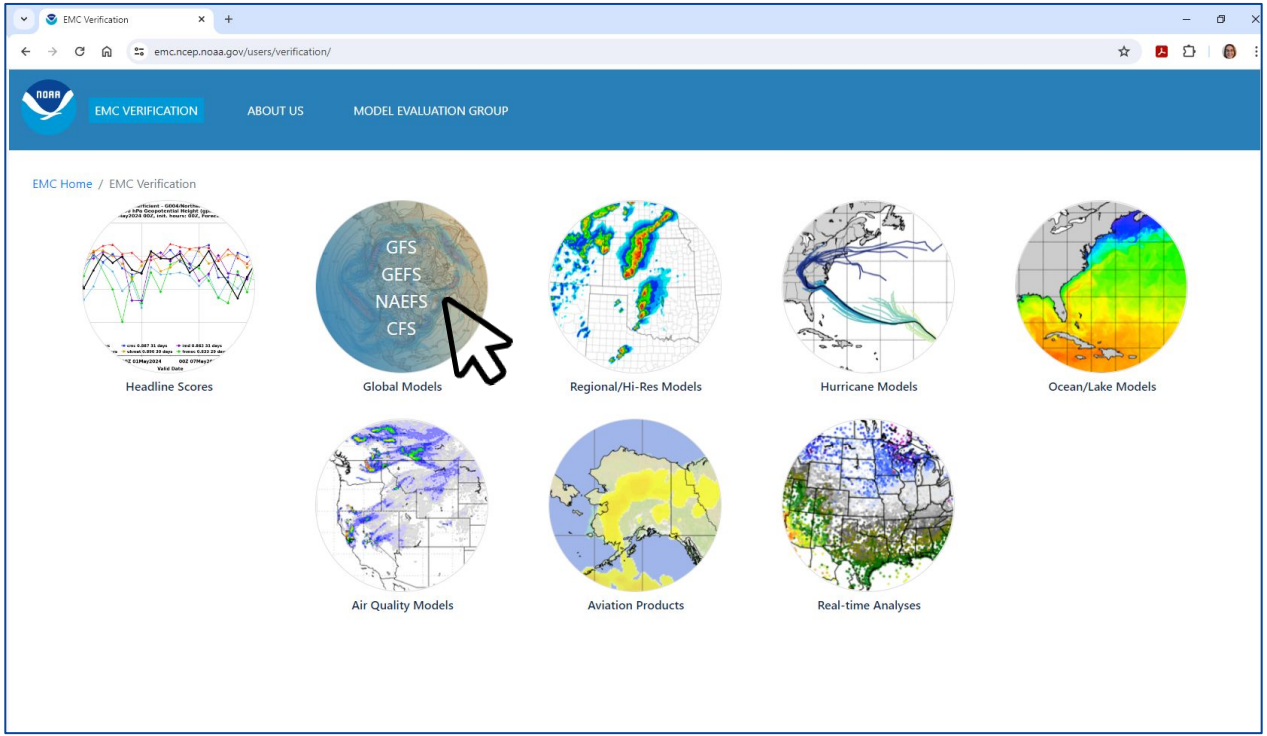
EVS Seasonal Verification: Verification Variables, Metrics and Validation Data

Additional fields/metrics to be developed in the Seasonal Verification after METplus capability developed and/or validation datasets for them become available:

- Soil Moisture (0-10 cm)
- QBO winds
- NAO index
- PNA index
- AO index
- AAO index
- East Asian Summer Monsoon Index
- Palmer Drought Severity Index (PDSI)
- Standard Precipitation Index (SPI)
- Soil Temperature (0-10 cm)
- Upper Ocean Heat Content

EMC Verification Website

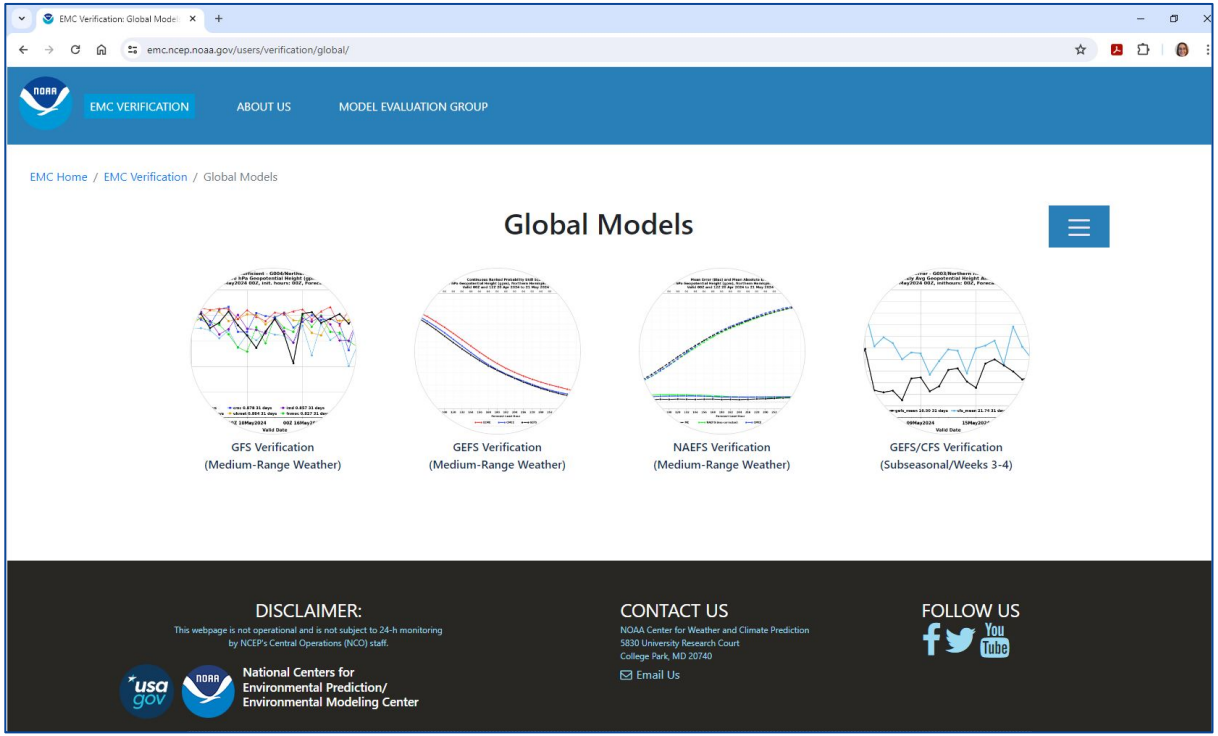
- EVS Seasonal Verification graphics will eventually be displayed under Global Models on EVS website



Link:
https://www.emc.ncep.noaa.gov/users/verification/

EMC Verification Website

- EVS Seasonal Verification graphics will eventually be displayed under Global Models on EVS website



Seasonal verification webpage would include:

- Option of selecting **G2G or G2O** verification
- Drop down menu options for **verification variables**
- Within each verification variable, drop down options **filter by**:
  - Region
  - Metric
  - Level
  - Time scale, etc.

Conclusions

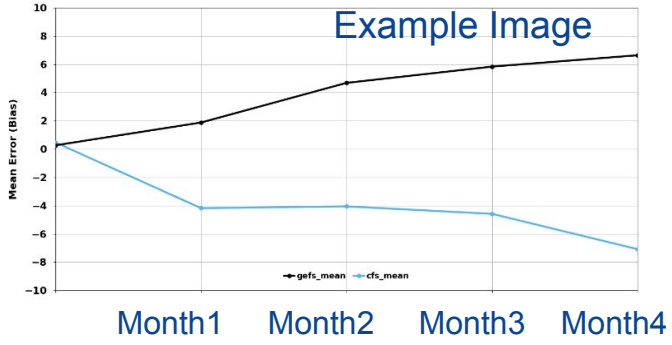
- EVS v2.0 Seasonal verification is being developed now!
  - Utilizes the Model Evaluation Tools (METplus) software package from DTC
- Operational models to be verified: CFSv2 and SFSv1 (once it's operational)
- Seasonal verification will range from 1 month to 9 months for CFSv2 and from 1 month to 12 months for SFSv1 (once it's operational)
- Temporal resolution of verification is monthly and 3–monthly
- Two types of verification: Grid-to-Grid and Grid-to-Obs (most Seasonal verification will use the Grid-to-Grid verification type)
- Several variables, metrics, and validation data sources were discussed

Extra Slides

Example for seasonal verification periods

Init: 0000 UTC 1 Jan 2024 F000 (same initialization date)

- Monthly verification
  - First 1-Month forecast verification period: Valid 0000 UTC 1 Feb 2023
  - Second 1-month forecast verification period: Valid 0000 UTC 1 March 2023
  - Third 1-month forecast verification period: Valid 0000 UTC 1 April 2023
  - …and so on
- 3-Monthly verification
  - First 3-month period: Valid 0000 UTC 1 April 2023
  - Second 3-month period: Valid 0000 UTC 1 July 2023
  - …and so on



Extra Slides

Credit: Collaboration from Australian Weather and Climate Research

CSI, ETS

Threat score (critical success index) - $TS = \frac{hits}{hits + misses + false\ alarms}$ (also denoted CSI)

Measures the fraction of observed and/or forecast events that were correctly predicted. It can be thought of as the *accuracy* when correct negatives have been removed from consideration, that is, *TS* is only concerned with forecasts that count. Sensitive to hits, penalizes both misses and false alarms. Does not distinguish source of forecast error. Depends on climatological frequency of events (poorer scores for rarer events) since some hits can occur purely due to random chance.

Equitable threat score (Gilbert skill score)- $ETS = \frac{hits - hits_{random}}{hits + misses + false\ alarms - hits_{random}}$ (also denoted GSS)

where $hits_{random} = \frac{(hits + misses)(hits + false\ alarms)}{total}$

Measures the fraction of observed and/or forecast events that were correctly predicted, adjusted for hits associated with random chance (for example, it is easier to correctly forecast rain occurrence in a wet climate than in a dry climate). The *ETS* is often used in the verification of rainfall in NWP models because its "equitability" allows scores to be compared more fairly across different regimes. Sensitive to hits. Because it penalises both misses and false alarms in the same way, it does not distinguish the source of forecast error.

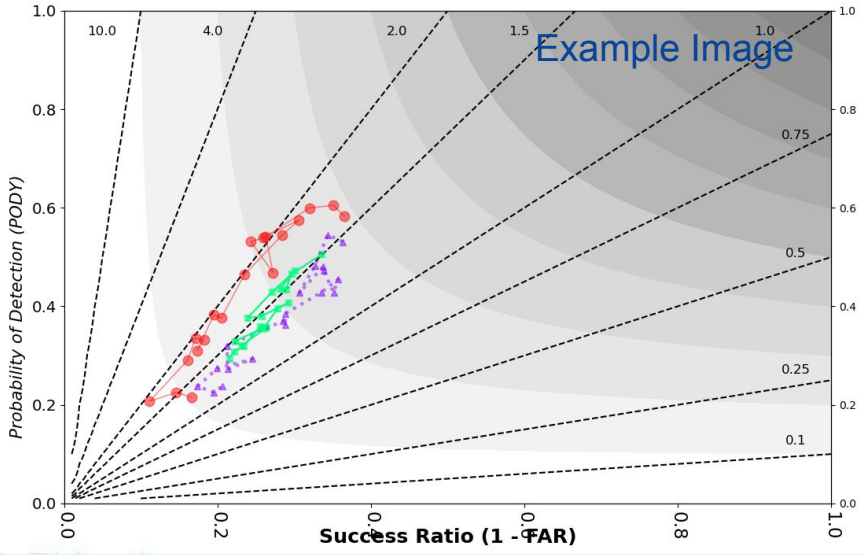
Extra Slides

Credit: METplotpy User's guide

Performance Diagram

Performance diagrams are used to show the relationship between categorical statistics, with axes representing detection and success (1 - false alarm) rates

Dashed lines: lines of equal frequency bias. Gray color curves:curves of equal Critical Success Index (CSI) connect the top of the plot to the right side.



Extra Slides

Credit: Collaboration from Australian Weather and Climate Research

HSS, FSS

Heidke skill score (Cohen's κ) - $$HSS = \frac{(hits + correct\ negatives) - (expected\ correct)_{random}}{N - (expected\ correct)_{random}}$$

Measures the fraction of correct forecasts after eliminating those forecasts which would be correct due purely to random chance.

$$FSS = 1 - \frac{\frac{1}{N}\sum_{N}(P_f - P_o)^2}{\frac{1}{N}\left[\sum_{N}P_f^2 + \sum_{N}P_o^2\right]}$$

where P_f is the forecast fraction, P_o is the observed fraction, and N is the number of spatial windows in the domain.

This approach directly compares the forecast and observed fractional coverage of grid-box events (rain exceeding a certain threshold, for example) in spatial windows of increasing size. FSS is sensitive to rare events (e.g., small rain areas).

Extra Slides

Credit: Collaboration from Australian Weather and Climate Research

RPSS

Ranked probability score - $$RPS = \frac{1}{M-1}\sum_{m=1}^{M}\left[\left(\sum_{k=1}^{m} p_k\right) - \left(\sum_{k=1}^{m} o_k\right)\right]^2$$

where M is the number of forecast categories, p_k is the predicted probability in forecast category k, and o_k is an indicator (0=no, 1=yes) for the observation in category k.

Measures the sum of squared differences in cumulative probability space for a multi-category probabilistic forecast. Penalizes forecasts more severely when their probabilities are further from the actual outcome.

Ranked probability skill score - $$RPSS = \frac{\overline{RPS} - \overline{RPS}_{reference}}{0 - \overline{RPS}_{reference}} = 1 - \frac{\overline{RPS}}{\overline{RPS}_{reference}}$$

Measures the improvement of the multi-category probabilistic forecast relative to a reference forecast (usually the long-term or sample climatology).