CPC Evaluations: Week-2 and Week-34

August 29, 2019

EMC 8th User's Ensemble Workshop

by Emerson LaJoie and Dan Collins

Outline:

- ✓ NAEFS (plus) Realtime Verification: Week-2
- ✓ SubX Hindcast Verification with a NAEFS focus: Week-34

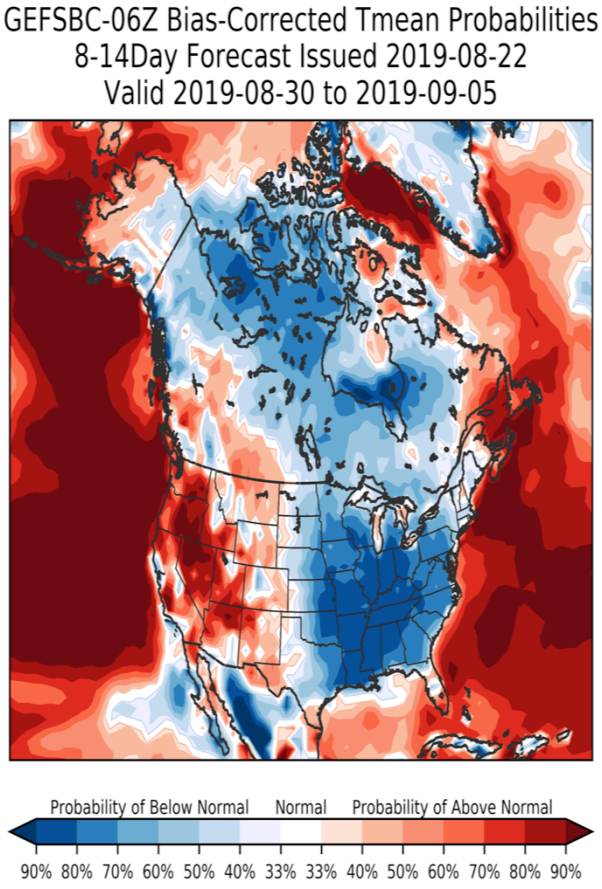
MAPP Modeling, Analysis, Predictions, and Projections

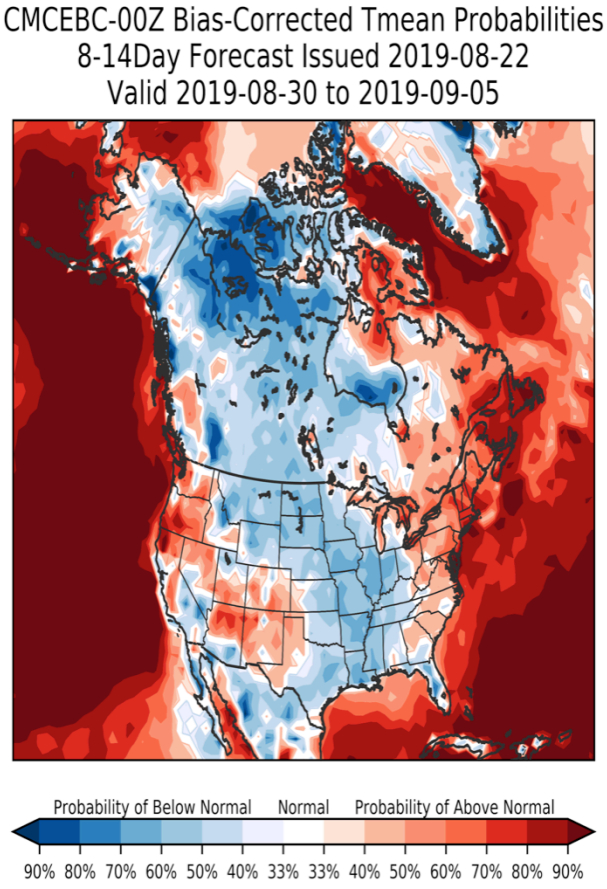
UNIVERSITY OF MIAMI ROSENSTIEL SCHOOL of MARINE & ATMOSPHERIC SCIENCE

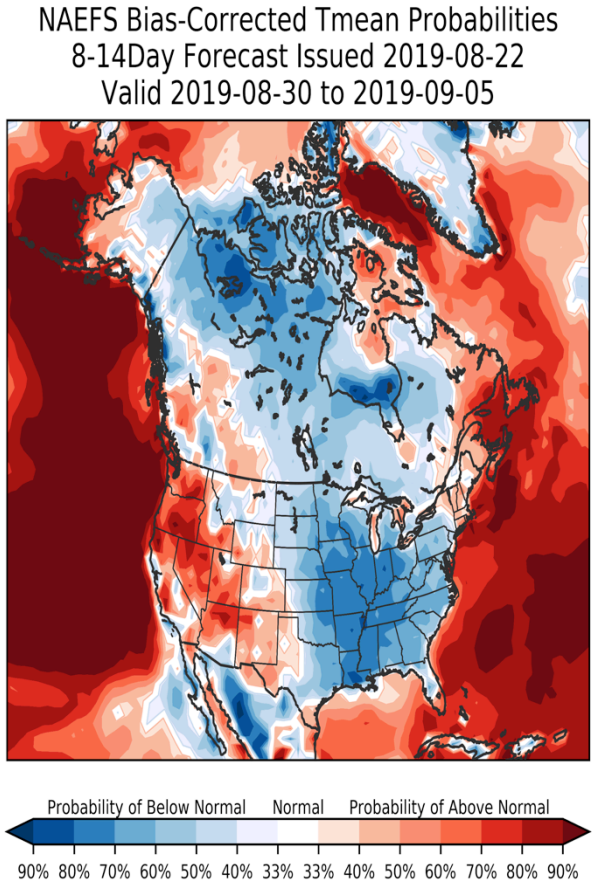
Week-2 NAEFS: Temperature Forecast Verification

- To include:
  - Raw GEFS, CMCE, ECENS
  - Autoblend
  - Consolidated
  - Bias corrected NAEFS

NAEFS Forecast Probabilities in support of CPC's Week-2 Outlook: Temperature

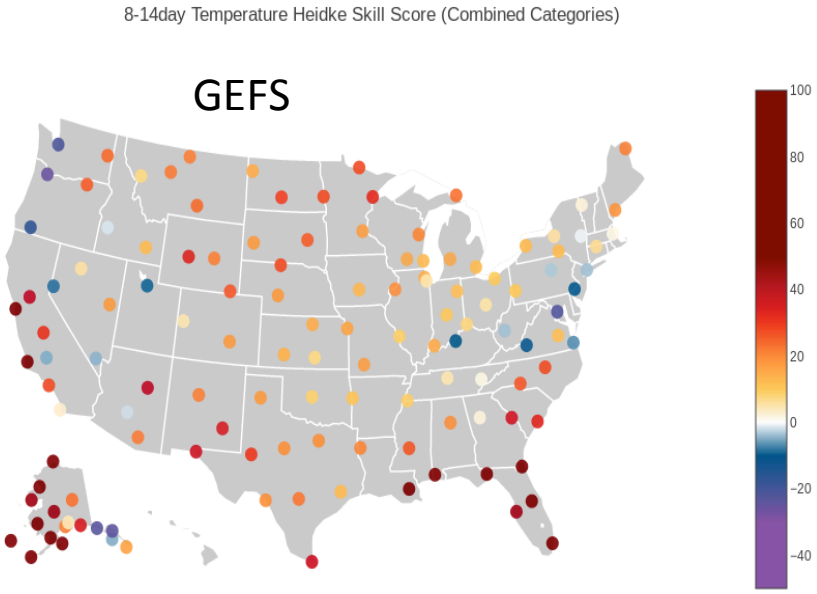


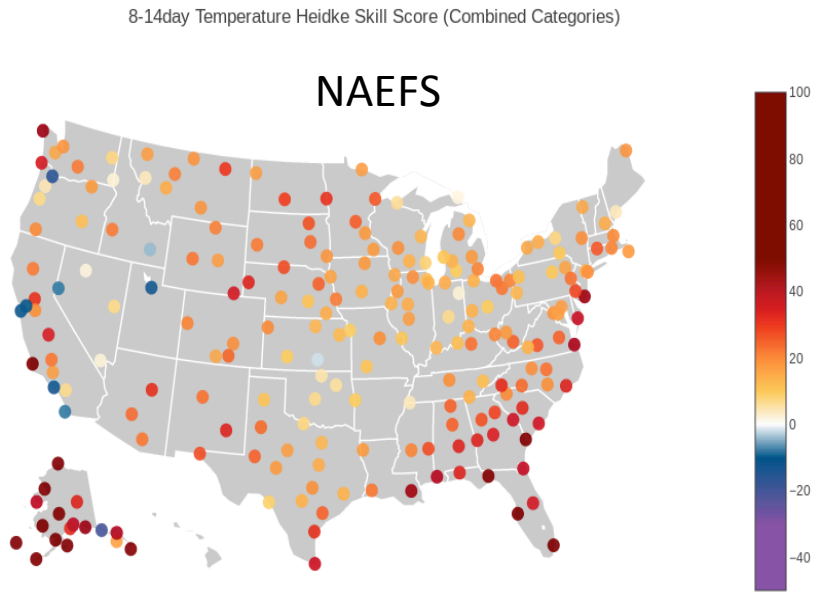


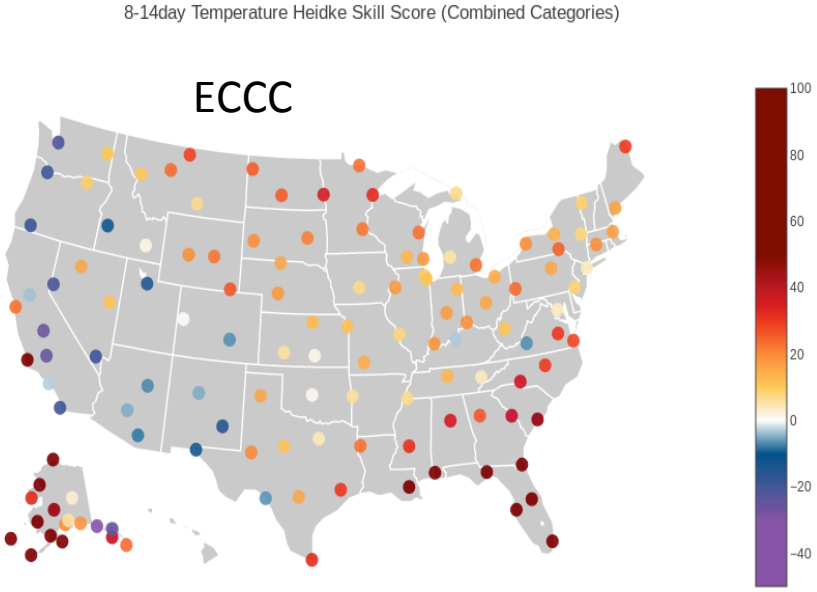


Temperature probabilities based on bias-corrected GEFS (left) and Environment Canada GEM ensemble forecasts (middle). Equal weighted combination for NAEFS (right).

Spatial Maps of 365-day Heidke Skill Scores for Temperature for GEFS, ECCC, and NAEFS

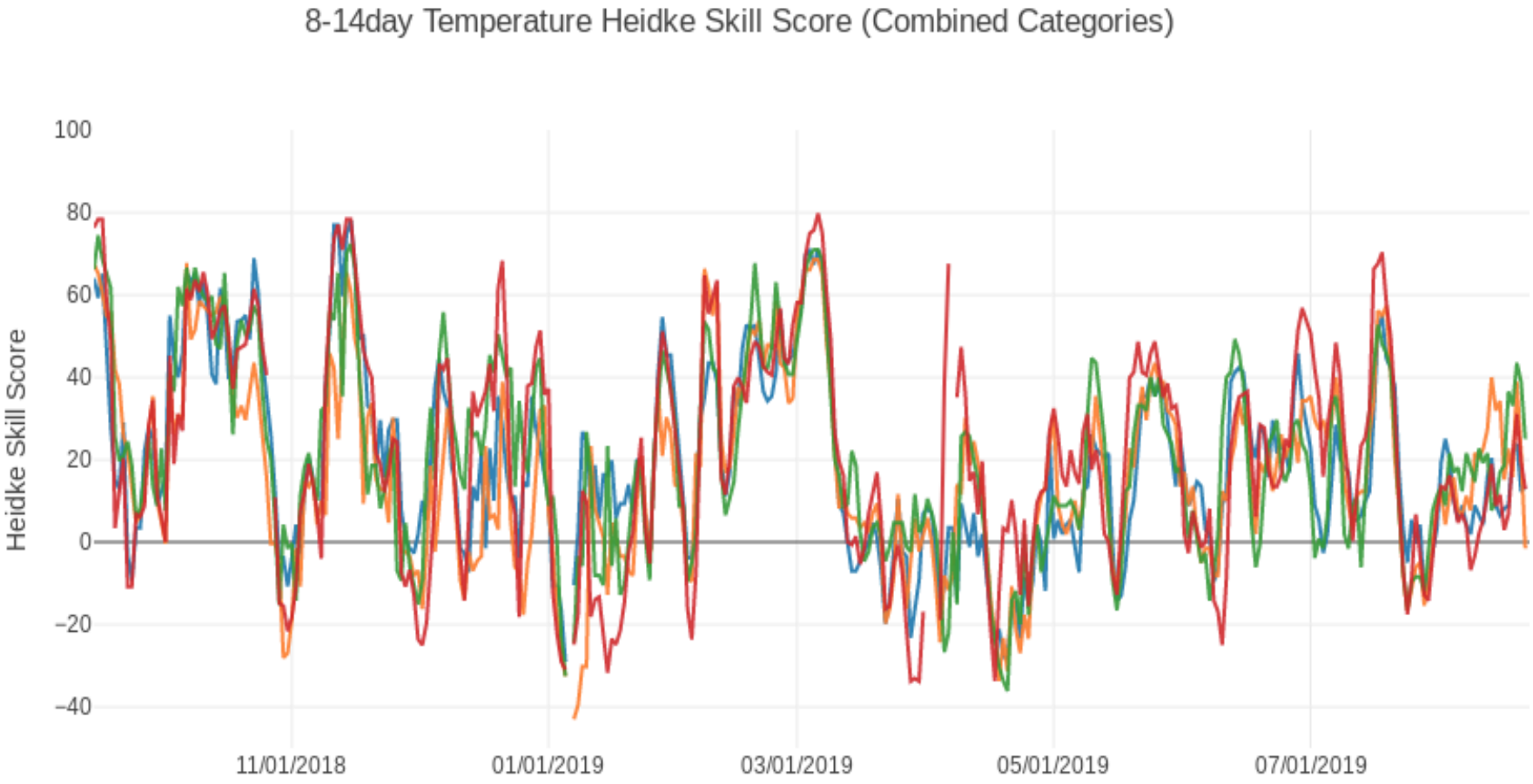






- Verification of NCEP GEFS, Canadian model and NAEFS for temperature on stations.
- Combined models have skill over essentially all regions.

Time Series of Heidke Skill Scores of Week-2 Temperature Outlook: GEFS, ECCC, ECMWF, and NAEFS

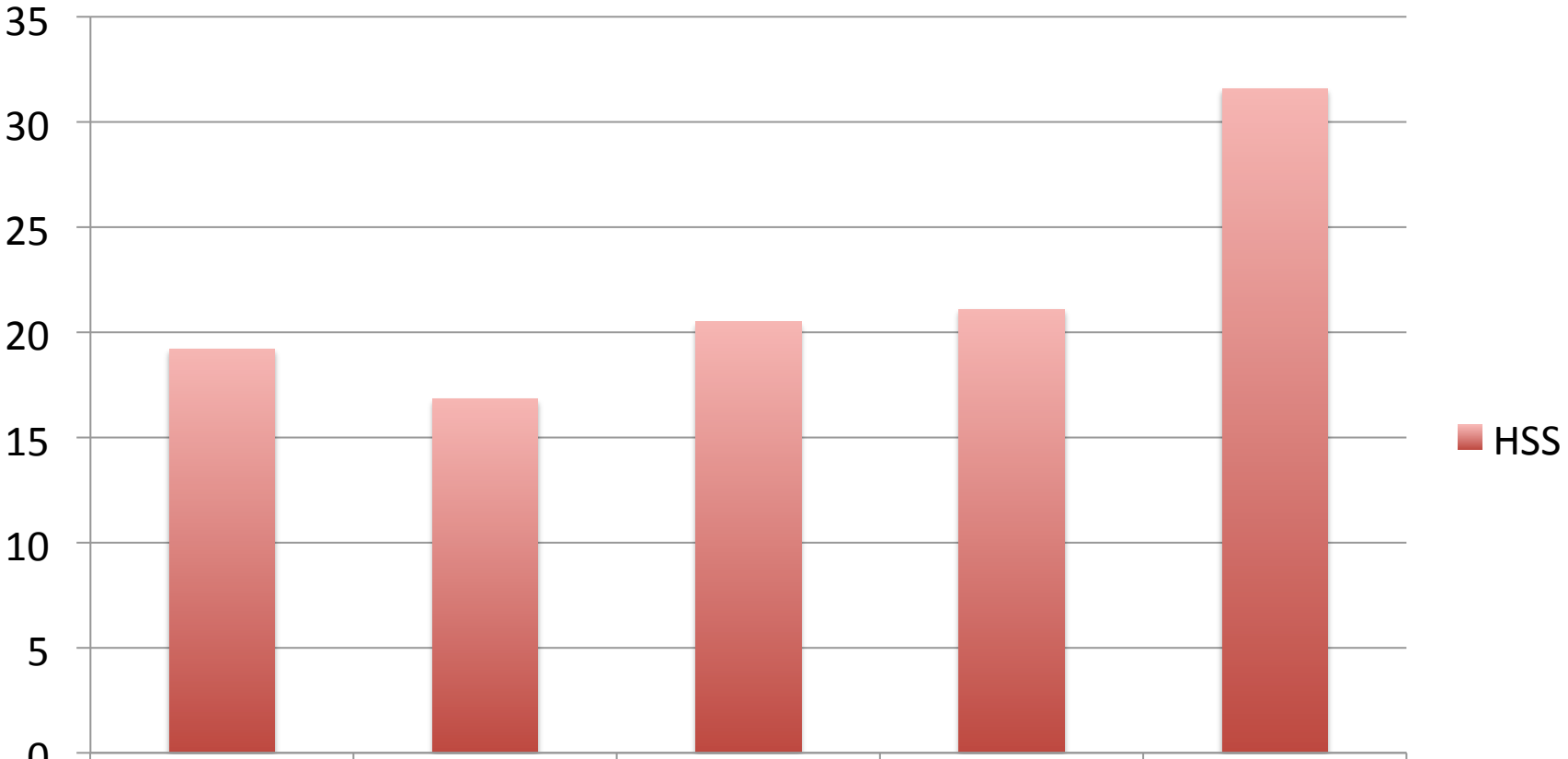


- 365-day time series of GEFS, GEM, ECMWF, and NAEFS.
- ECMWF has greater skill than GEFS or GEM, but less skill than NAEFS.

Average Scores

raw_gefs_al: *19.203*
raw_cmce_00z: *16.819*
raw_ecens_00z: *20.501*
BC_naefs: *21.087*

365-day Temperature Verification: Week-2 HSS Summary

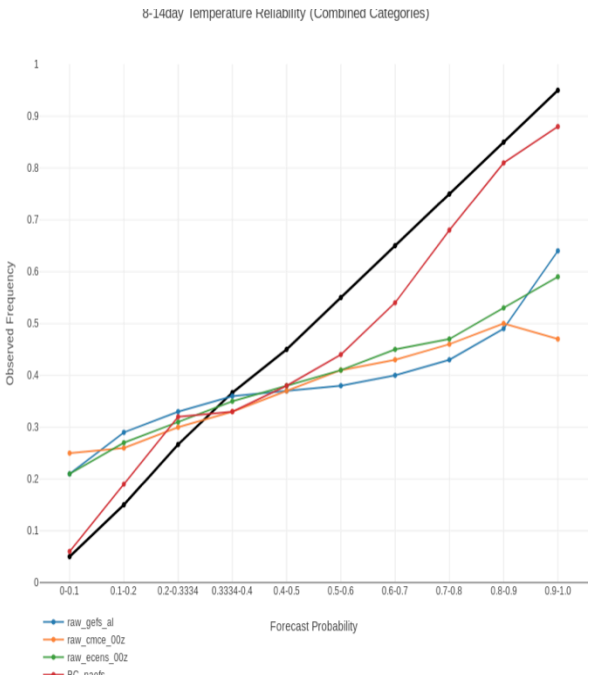


	Raw GEFS	Raw CMCE	Raw ECENS	BC NAEFS	Auto
HSS	19.203	16.819	20.501	21.087	31.581

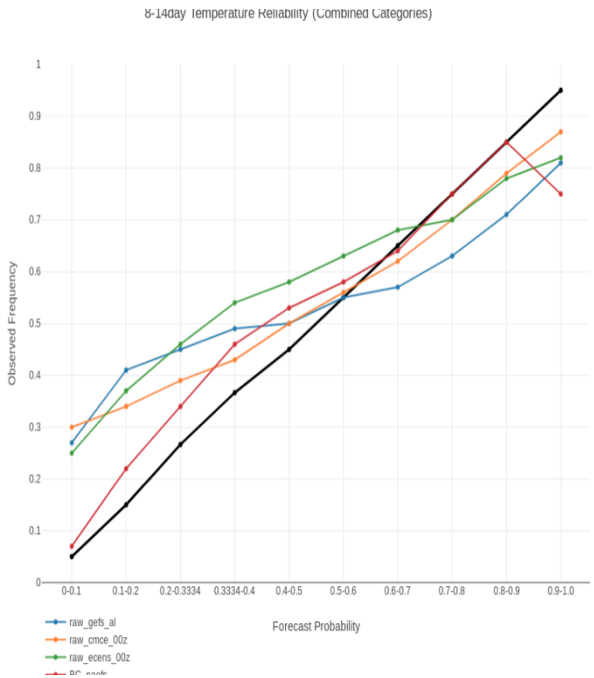
- ECMWF has greater skill than GEFS or GEM ensembles, but less skill than NAEFS.
- Autoblend is a combination of NAEFS, Calibrated GEFS and ECMWF (using reforecasts), and other model tools… including analogs (hybrid statistical-dynamical forecasts)
- Autoblend is CPC's primary forecast tool in week-2 and shows the best skill

365-day Reliability for GEFS, ECCC, ECMWF, and NAEFS:

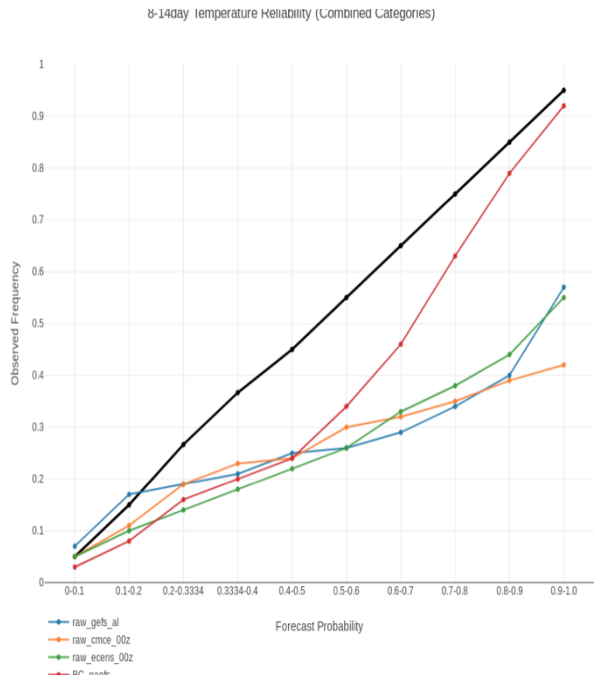
Three-Category



Above Normal



Below Normal



- **Reliability of GEFS, GEM, and ECMWF ensembles, and NAEFS.**
- **NAEFS MME has better reliability than individual models. Including for Above normal (center) and below normal (right)**
- **Below normal shows a bias and is forecast more than it occurs.**

Week 2 Summary:

- **MME's (NAEFS), blended tools, and calibrated consolidations of MME provide greater skill than individual ensembles**
- **Calibration using reforecasts leads to greater skill than simple bias corrections**
- **Additional thoughts….**
  - A calibrated and consolidated MME of the GEFS, GEM and ECMWF would benefit week-2 forecasts

Weeks 3-4 Subseasonal Experiment (SubX): Temperature and Precipitation Forecast Verification

SubX BY THE NUMBERS

7 Global Models

2 Years of *Real-time* Forecasts

15 Years of *Retrospective* Forecasts

3-4 week guidance for Climate Prediction Center Outlooks

SubX Protocol

- Prediction System Details up to Provider
- Real-time and Retrospective Systems Identical
- Reforecast Period: 1999-2014
- At Least 3 Ensemble Members
- Minimum Length: 32 Days
- Real-time Forecast Made Available to CPC *Every Thursday* by 10am of *Every week*
- Data on Uniform 1x1 Grid

Week-34: SubX Evaluation Details

- RMSE week-34 hindcast verification on temperature and precipitation over CONUS+Alaska (all months 1999-2014) from the SubX database
- Evaluation has four parts, designed with NAEFS models in mind:
  - Compare individual SubX models and the SubXMME to GEFS
  - Level 1: GEFS compared to GEFS+Model
  - Level 2: GEFS+GEM compared to GEFS+GEM+Model and the SubXMME
  - Level 3: GEFS+GEM+NESM compared to GEFS+GEM+NESM+Model and the SubXMME
    - About half of ECCC's forecasts are not included
    - ECCC is on the fly and upgrades often – presents some challenges
    - Two upgrades since this hindcast

Verification was performed with leads that match realtime – to capture realtime skill

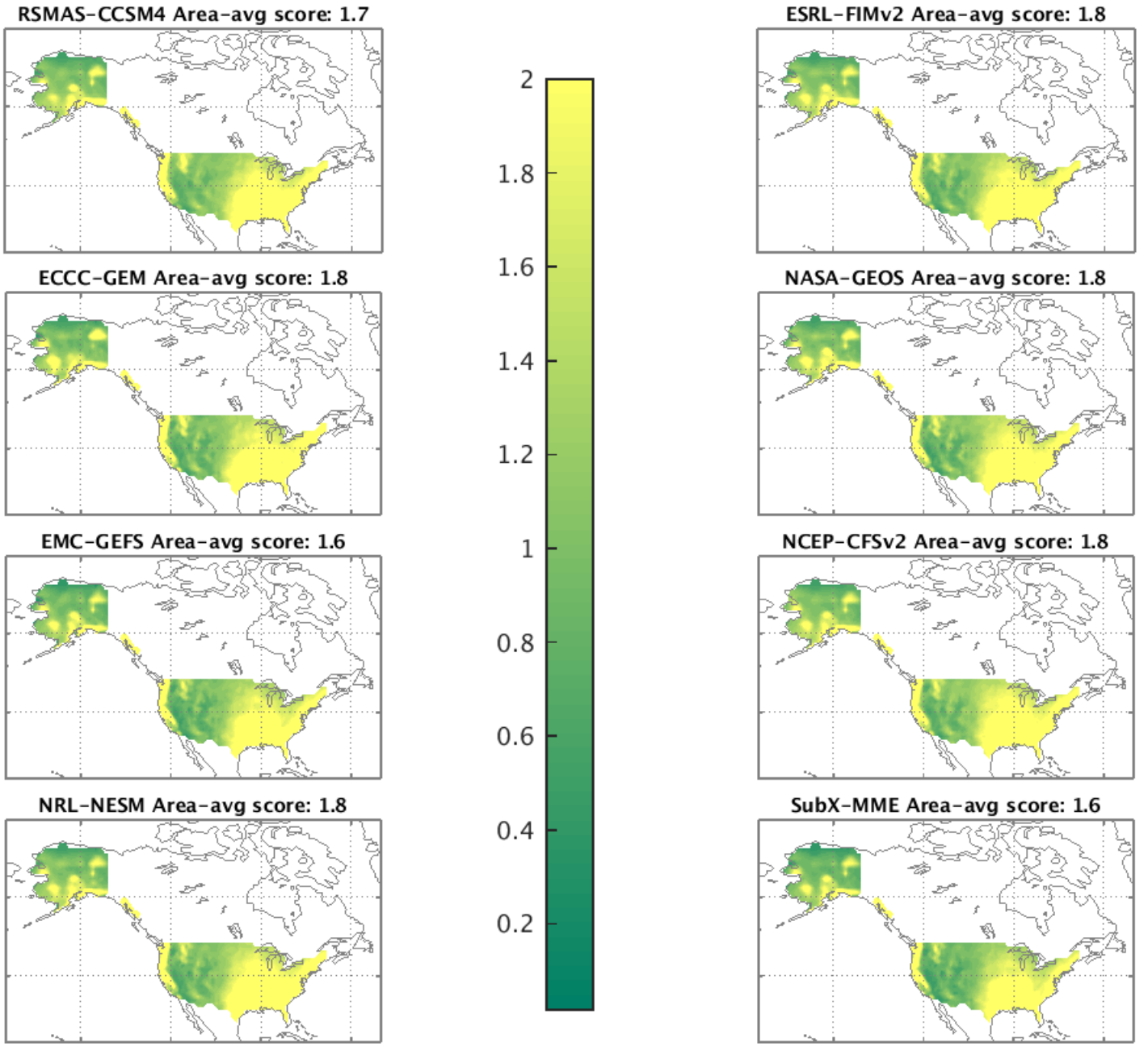
Week of Hindcast Dates and Target Dates	Jan 1	Jan 2	Jan 3	Jan 4	Jan 5	Jan 6	Jan 7	**Jan 8** **Forecast Day**	**Week 3-4 Outlook: Jan 22 – Feb 05**
Day of the week and Days to Target Dates	Fri 22:35	Sat 21:34	Sun 20:33	Mon 19:32	Tues 18:31	Wed 17:30	Thurs 16:29	**Fri 15:28**	**2 weeks from Sat + 13 days → WK34**
Center-Model	---------- **Reforecast Grab Period** ----------								
ECCC-GEM **4 members** 32 days							✹	**Forecast Day**	✹ **= Realtime**
EMC-GEFS **11 members** 35 days						✹		**Forecast Day**	
ESRL-FIMv2 **4 members** 32 days						✹		**Forecast Day**	
NASA-GEOS **4 members** 45 days								**Forecast Day**	*** GEOS5 roves in Realtime**
NCEP-CFSv2 **4 members** 44 days						✹		**Forecast Day**	
NRL-NESM **4 lagged members** 45 days		✹	✹	✹	✹			**Forecast Day**	
RSMAS-CCSM4 **3 members** 45 days			✹					**Forecast Day**	
Coming Soon: NCAR-CESM 10 members 45 days				✹				**Forecast Day**	

Spatial RMSE

Precipitation scores across the full hindcast

- **Individual Models and SubXMME**

Individual Models and SubXMME RMSE: PRECP for All Months (1999-2014)

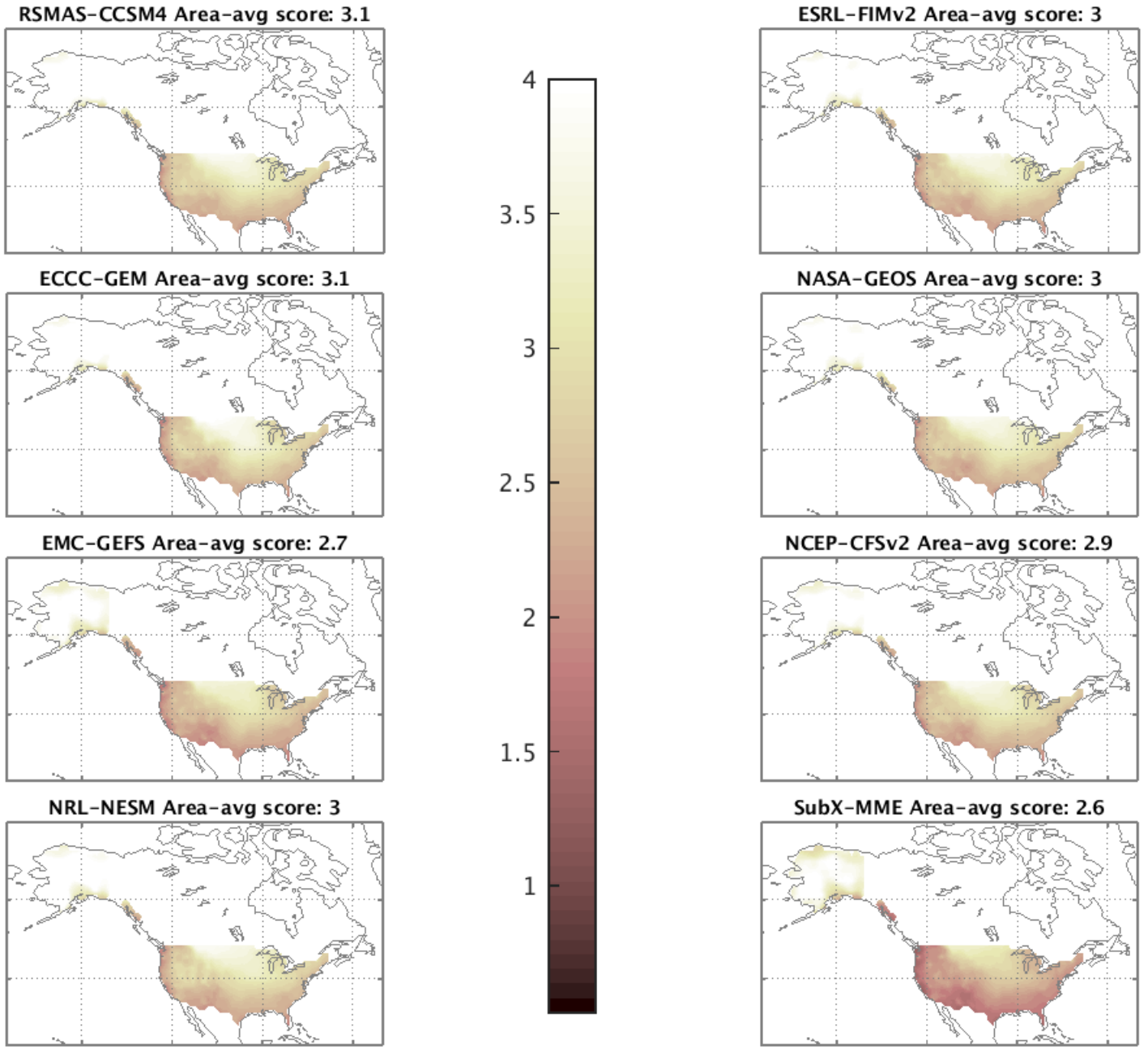


Spatial RMSE

Temperature scores across the full hindcast

- **Individual Models and SubXMME**

Individual Models and SubXMME RMSE: TAS2M for All Months (1999-2014)



SIGN TEST: RMSE

Precipitation scores across the full hindcast

DelSole and Tippett 2018: ***Forecast Comparison Based on Random Walks***

- ***Criterion for selecting most skillful model of a single forecast event***
- ***Based on a cumulative count of times a forecast was more skillful***
  - ***…think of a coin toss and the 50-50 chance of heads or tails…***
- ***Provides the probability of success of one model over another model***
- ***Test is not sensitive to comparing MMEs with models within the MME***
- ***Tempting to compare curves, but don't…***

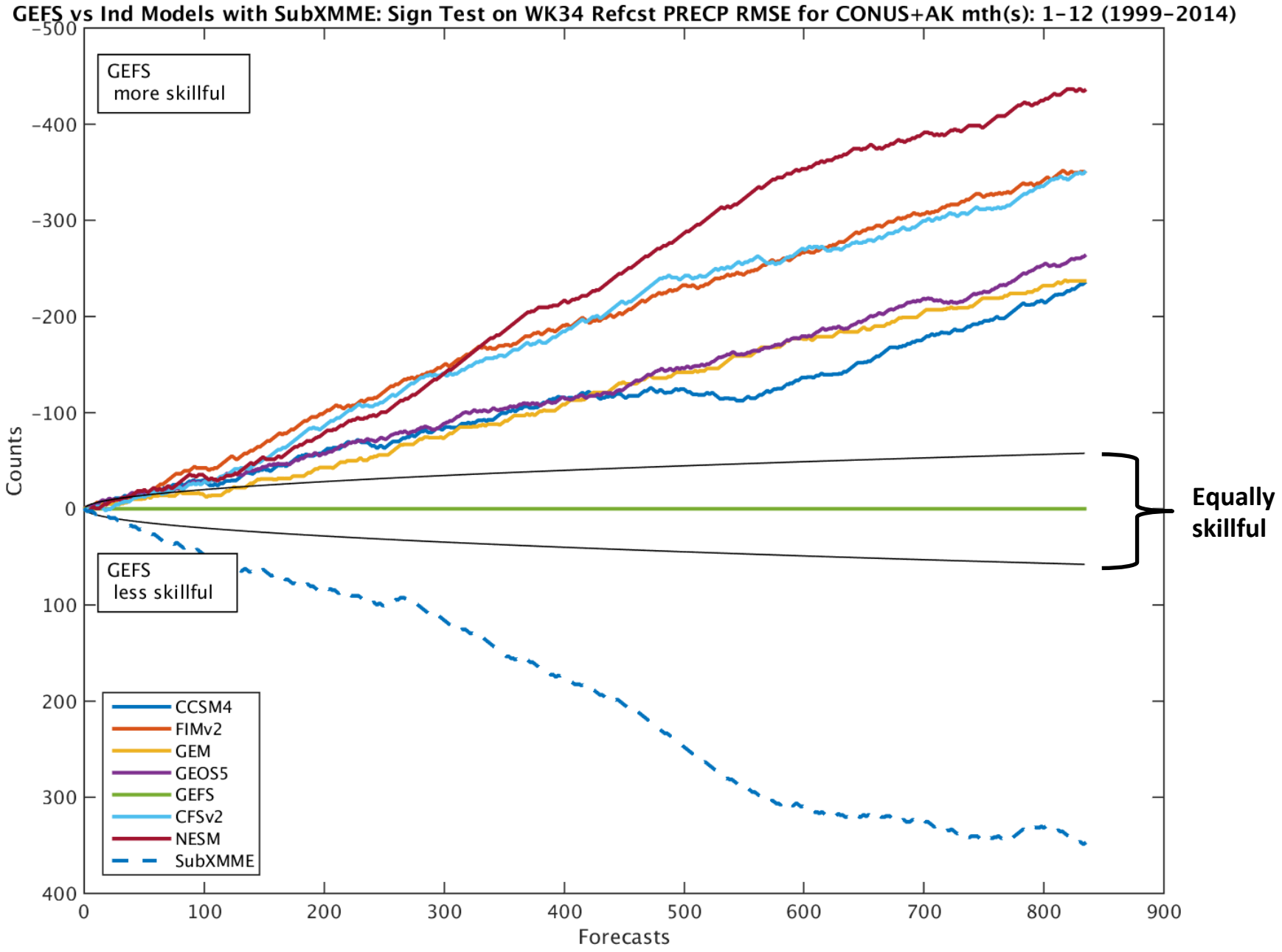
SIGN TEST: RMSE

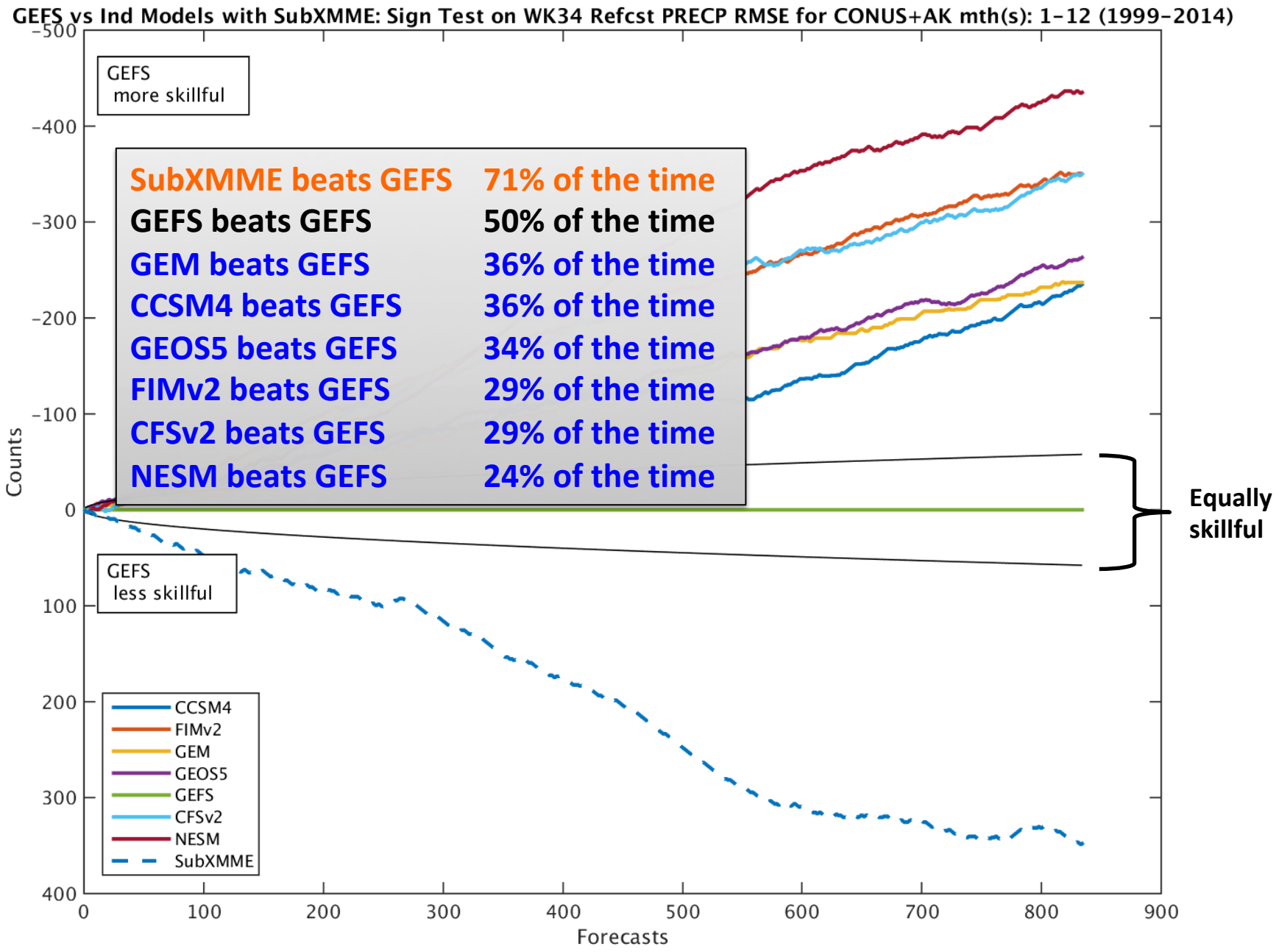
Precipitation scores across the full hindcast

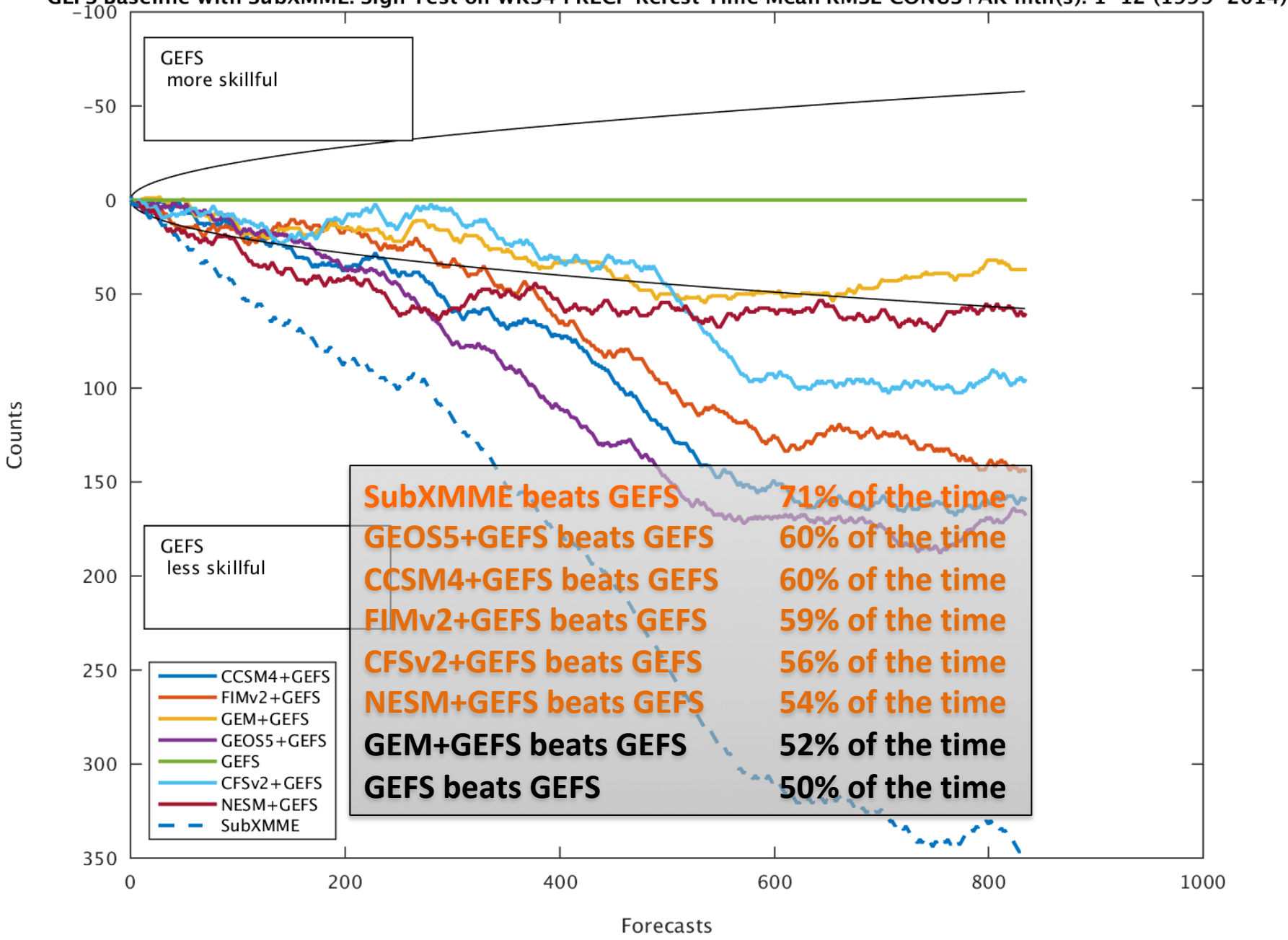
Method:

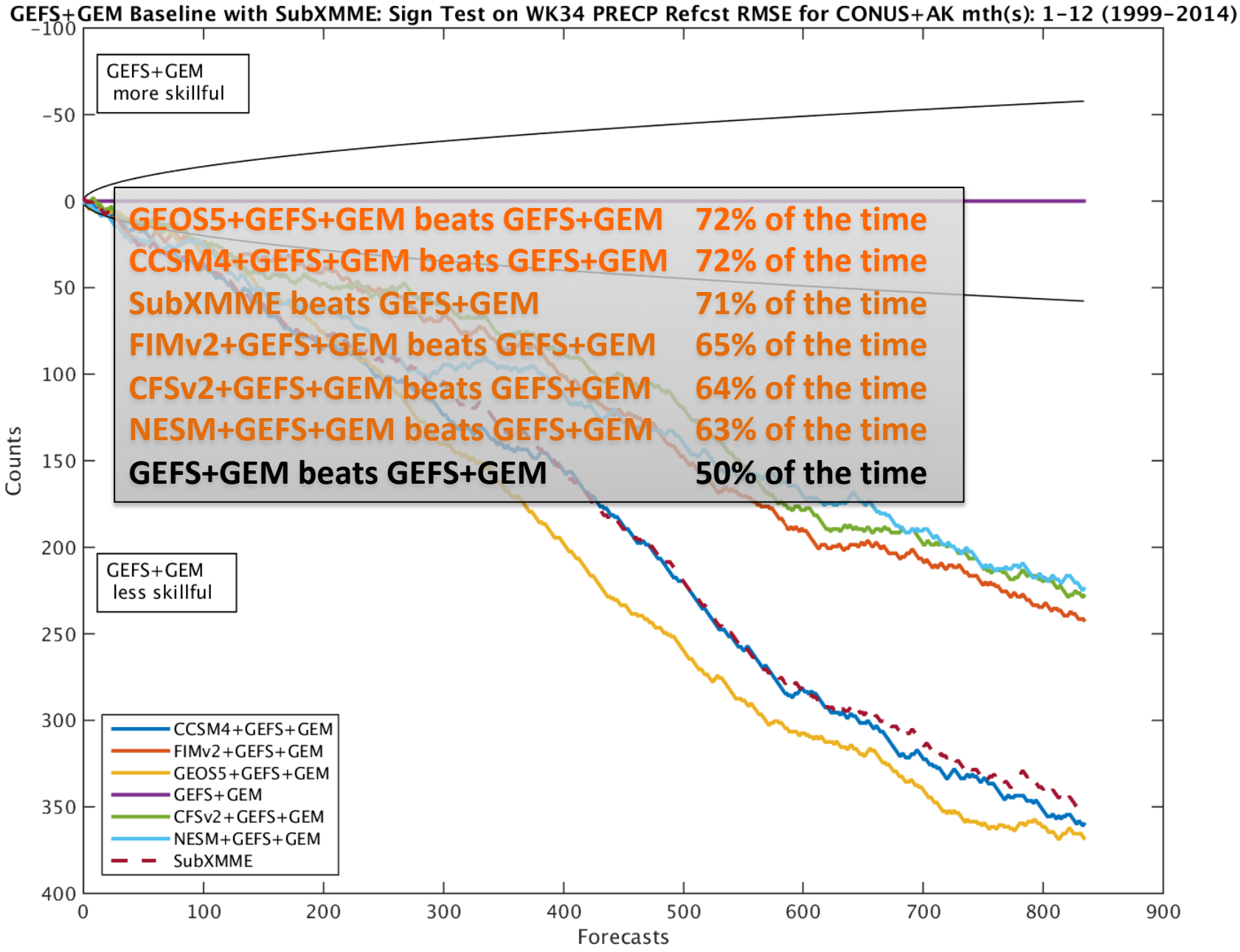
- **Model A minus Model B = sign of the difference (→ +1 or -1)**
- **Cumulative sum of those +/- 1s over all Forecasts → Counts**
- **Probability of success = (Total + Count / 2) * 100%**

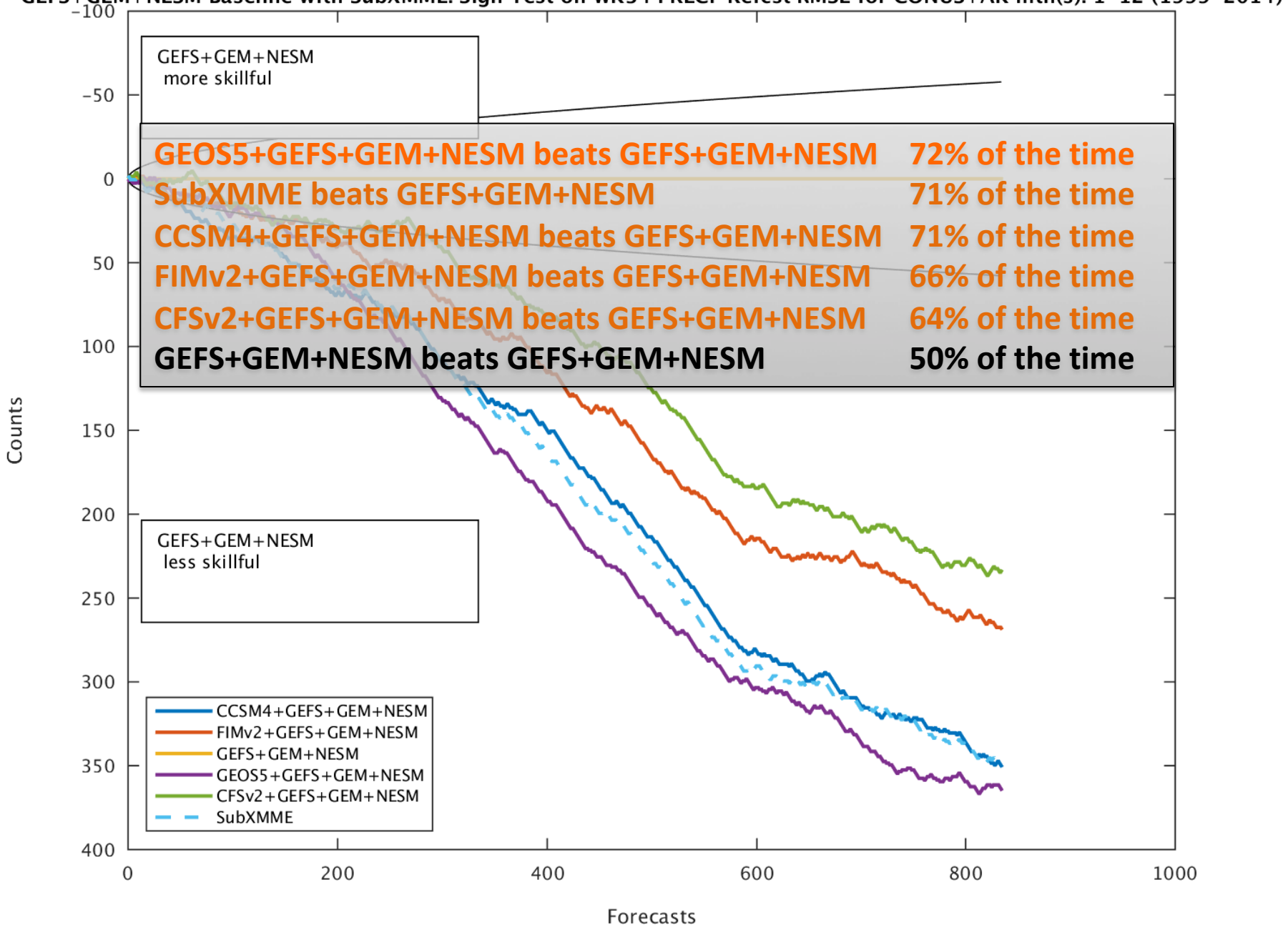
- **Individual Model Scores**
- **Three Levels**


GEFS vs Ind Models with SubXMME: Sign Test on WK34 Refcst PRECP RMSE for CONUS+AK mth(s): 1-12 (1999-2014)
GEFS
more skillful
GEFS
less skillful
Equally
skillful
CCSM4
FIMv2
GEM
GEOS5
GEFS
CFSv2
NESM
SubXMME
Counts
Forecasts
-500
-400
-300
-200
-100
0
100
200
300
400
0
100
200
300
400
500
600
700
800
900


GEFS vs Ind Models with SubXMME: Sign Test on WK34 Refcst PRECP RMSE for CONUS+AK mth(s): 1-12 (1999-2014)
GEFS more skillful
GEFS less skillful
SubXMME beats GEFS 71% of the time
GEFS beats GEFS 50% of the time
GEM beats GEFS 36% of the time
CCSM4 beats GEFS 36% of the time
GEOS5 beats GEFS 34% of the time
FIMv2 beats GEFS 29% of the time
CFSv2 beats GEFS 29% of the time
NESM beats GEFS 24% of the time
Equally skillful
CCSM4
FIMv2
GEM
GEOS5
GEFS
CFSv2
NESM
SubXMME
Counts
Forecasts
-500
-400
-300
-200
-100
0
100
200
300
400
0
100
200
300
400
500
600
700
800
900

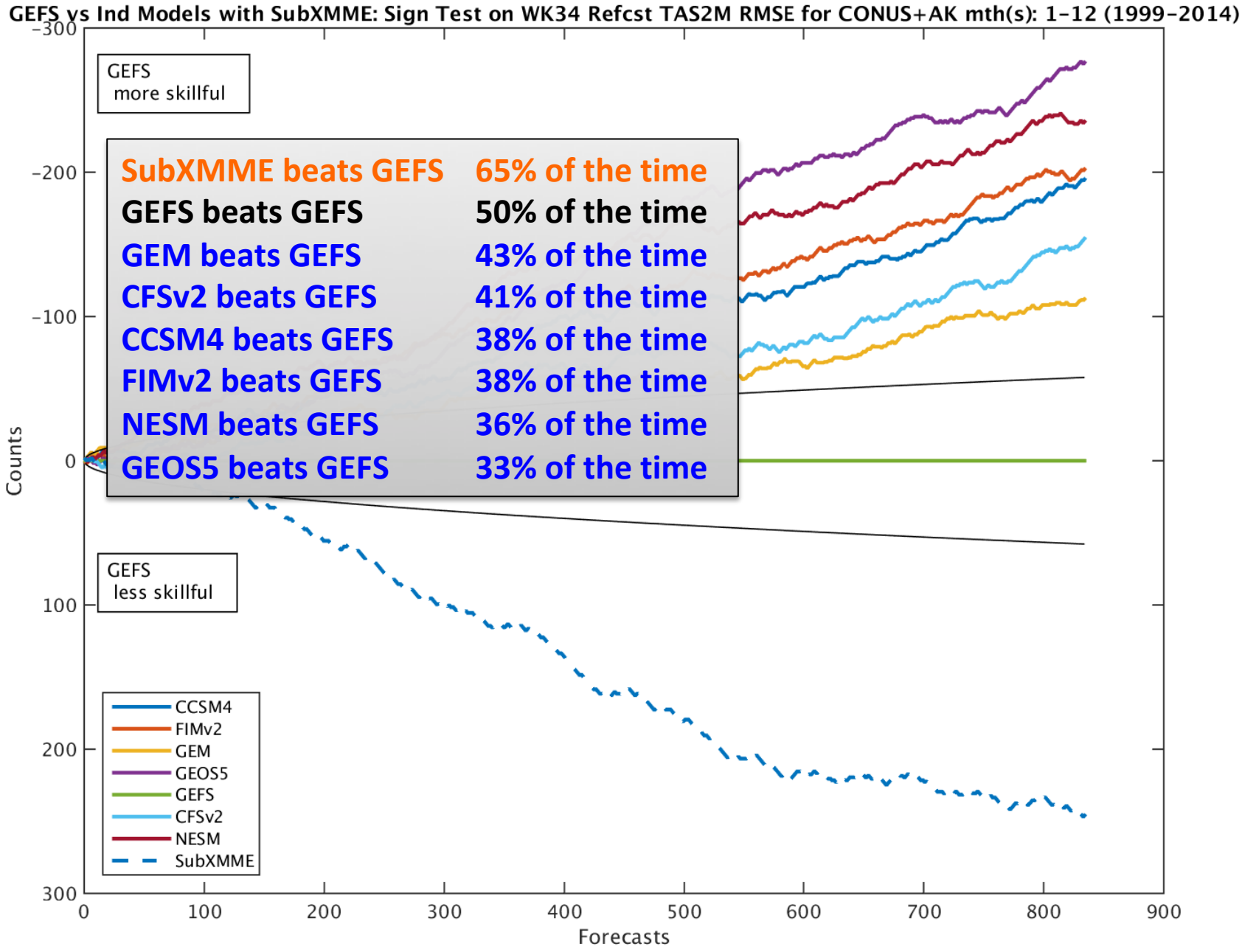

GEFS Baseline with SubXMME: Sign Test on WK34 PRECP Refcst Time Mean RMSE CONUS+AK mth(s): 1-12 (1999-2014)
GEFS
more skillful
GEFS
less skillful
CCSM4+GEFS
FIMv2+GEFS
GEM+GEFS
GEOS5+GEFS
GEFS
CFSv2+GEFS
NESM+GEFS
SubXMME
SubXMME beats GEFS 71% of the time
GEOS5+GEFS beats GEFS 60% of the time
CCSM4+GEFS beats GEFS 60% of the time
FIMv2+GEFS beats GEFS 59% of the time
CFSv2+GEFS beats GEFS 56% of the time
NESM+GEFS beats GEFS 54% of the time
GEM+GEFS beats GEFS 52% of the time
GEFS beats GEFS 50% of the time
Counts
Forecasts
-100
-50
0
50
100
150
200
250
300
350
0
200
400
600
800
1000

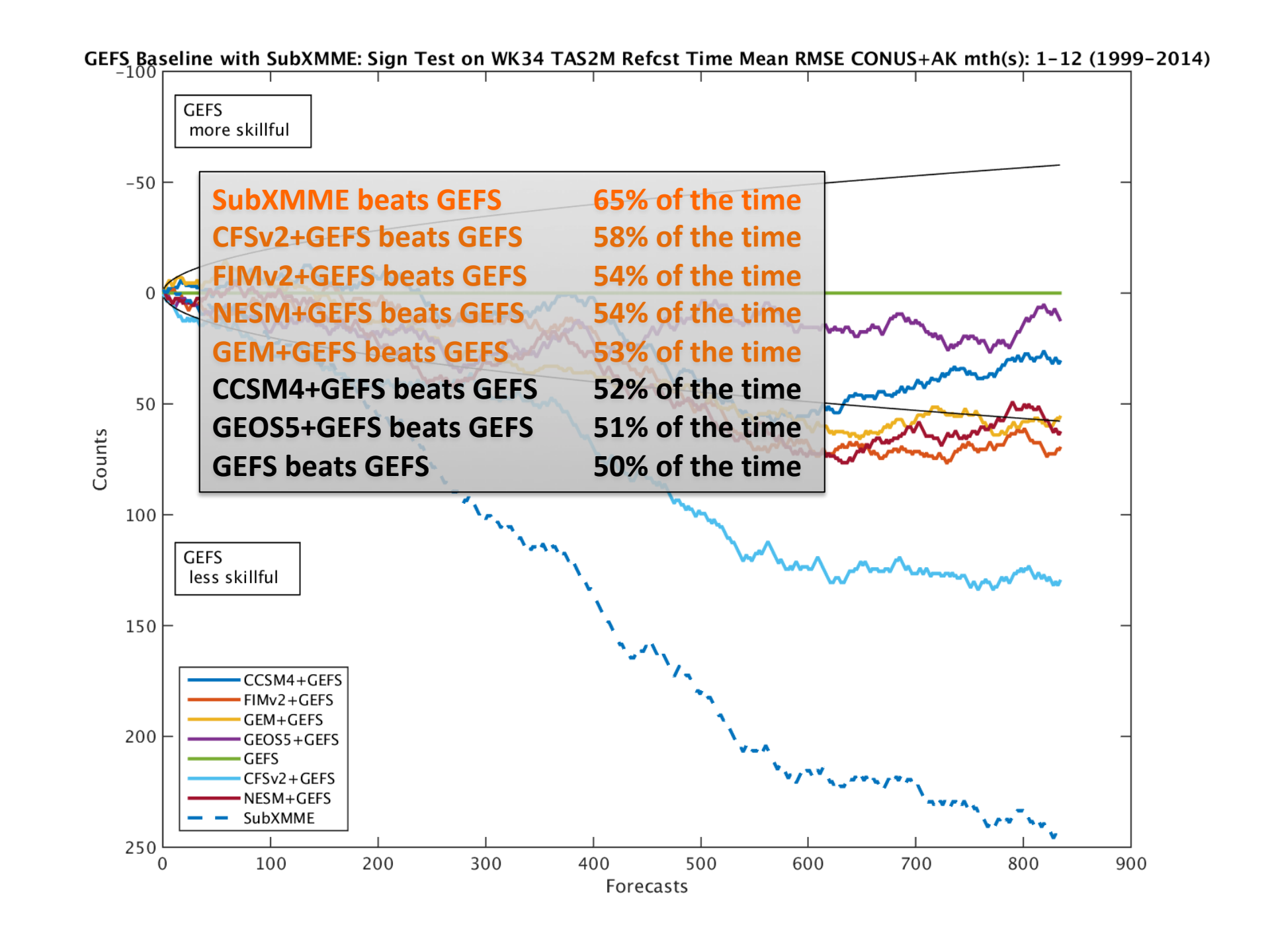

GEFS+GEM Baseline with SubXMME: Sign Test on WK34 PRECP Refcst RMSE for CONUS+AK mth(s): 1-12 (1999-2014)
GEFS+GEM more skillful
GEFS+GEM less skillful
GEOS5+GEFS+GEM beats GEFS+GEM 72% of the time
CCSM4+GEFS+GEM beats GEFS+GEM 72% of the time
SubXMME beats GEFS+GEM 71% of the time
FIMv2+GEFS+GEM beats GEFS+GEM 65% of the time
CFSv2+GEFS+GEM beats GEFS+GEM 64% of the time
NESM+GEFS+GEM beats GEFS+GEM 63% of the time
GEFS+GEM beats GEFS+GEM 50% of the time
CCSM4+GEFS+GEM
FIMv2+GEFS+GEM
GEOS5+GEFS+GEM
GEFS+GEM
CFSv2+GEFS+GEM
NESM+GEFS+GEM
SubXMME
Counts
Forecasts
-100
-50
0
50
100
150
200
250
300
350
400
0
100
200
300
400
500
600
700
800
900

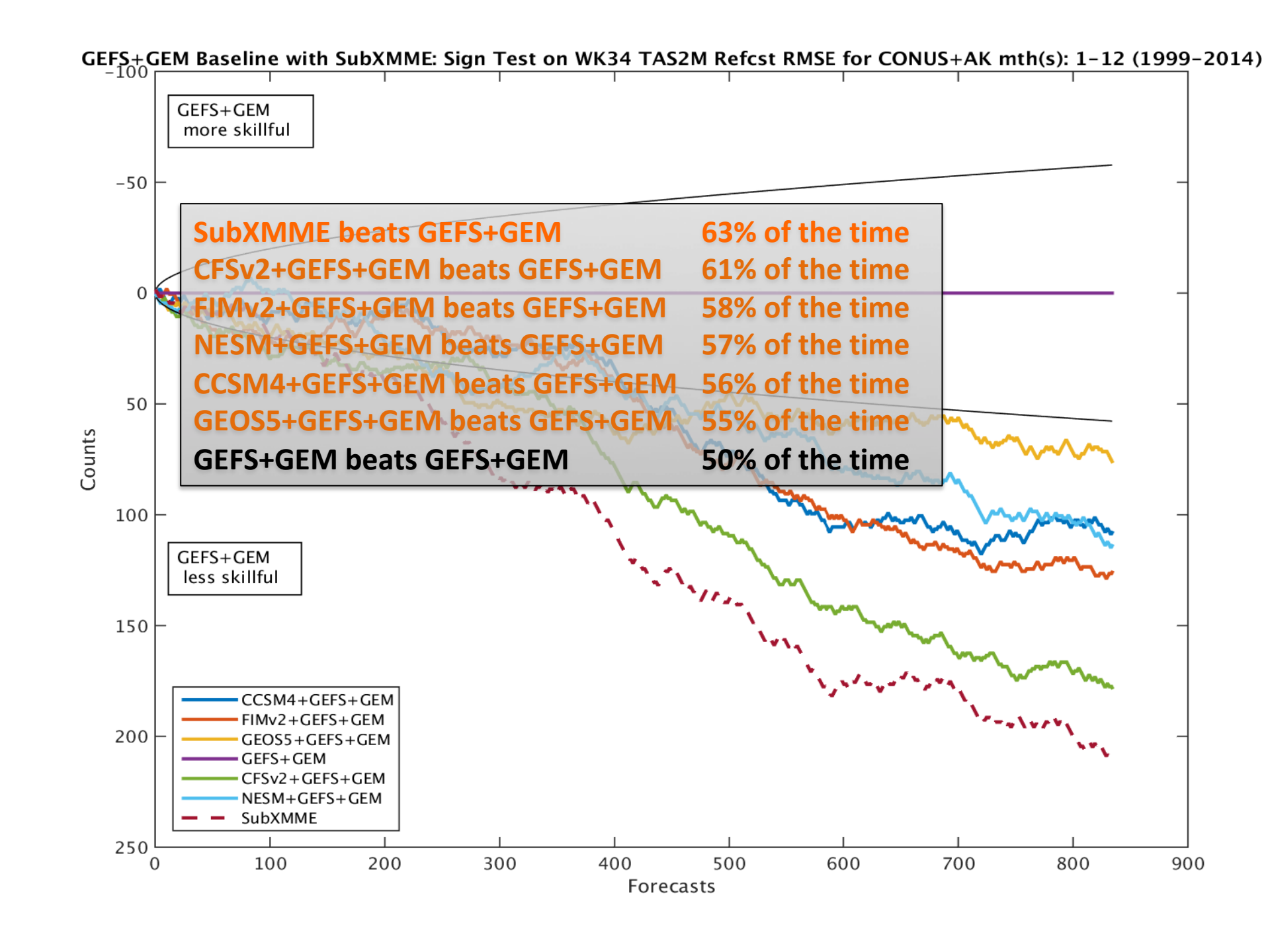

GEFS+GEM+NESM Baseline with SubXMME: Sign Test on WK34 PRECP Refcst RMSE for CONUS+AK mth(s): 1-12 (1999-2014)
GEFS+GEM+NESM
more skillful
GEOS5+GEFS+GEM+NESM beats GEFS+GEM+NESM 72% of the time
SubXMME beats GEFS+GEM+NESM 71% of the time
CCSM4+GEFS+GEM+NESM beats GEFS+GEM+NESM 71% of the time
FIMv2+GEFS+GEM+NESM beats GEFS+GEM+NESM 66% of the time
CFSv2+GEFS+GEM+NESM beats GEFS+GEM+NESM 64% of the time
GEFS+GEM+NESM beats GEFS+GEM+NESM 50% of the time
GEFS+GEM+NESM
less skillful
CCSM4+GEFS+GEM+NESM
FIMv2+GEFS+GEM+NESM
GEFS+GEM+NESM
GEOS5+GEFS+GEM+NESM
CFSv2+GEFS+GEM+NESM
SubXMME
Counts
Forecasts
-100
-50
0
50
100
150
200
250
300
350
400
0
200
400
600
800
1000

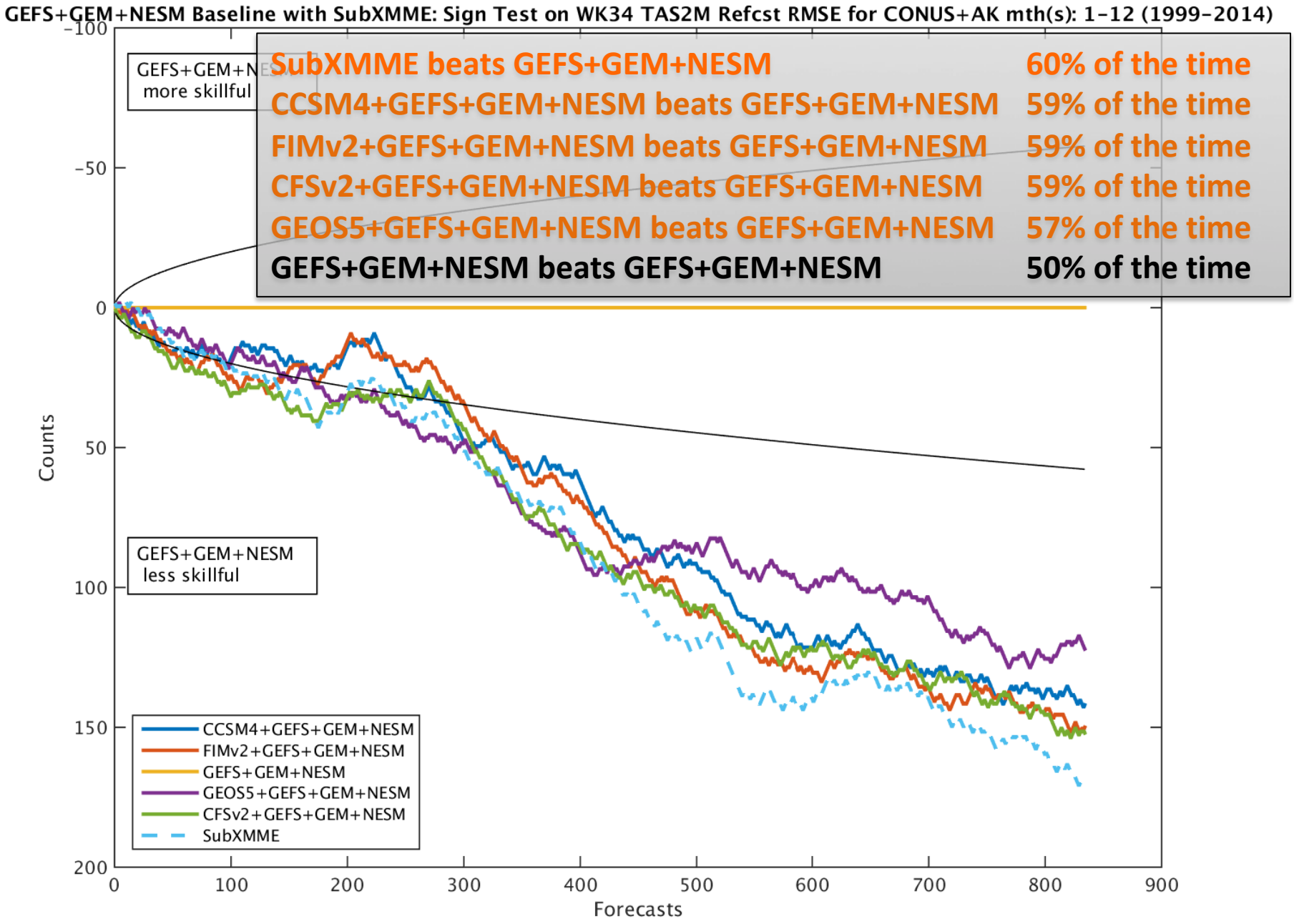
SIGN TEST: RMSE

Temperature scores across the full hindcast


GEFS vs Ind Models with SubXMME: Sign Test on WK34 Refcst TAS2M RMSE for CONUS+AK mth(s): 1-12 (1999-2014)
GEFS
more skillful
SubXMME beats GEFS 65% of the time
GEFS beats GEFS 50% of the time
GEM beats GEFS 43% of the time
CFSv2 beats GEFS 41% of the time
CCSM4 beats GEFS 38% of the time
FIMv2 beats GEFS 38% of the time
NESM beats GEFS 36% of the time
GEOS5 beats GEFS 33% of the time
GEFS
less skillful
CCSM4
FIMv2
GEM
GEOS5
GEFS
CFSv2
NESM
SubXMME
Counts
Forecasts
-300
-200
-100
0
100
200
300
0
100
200
300
400
500
600
700
800
900


GEFS Baseline with SubXMME: Sign Test on WK34 TAS2M Refcst Time Mean RMSE CONUS+AK mth(s): 1-12 (1999-2014)
GEFS more skillful
GEFS less skillful
SubXMME beats GEFS 65% of the time
CFSv2+GEFS beats GEFS 58% of the time
FIMv2+GEFS beats GEFS 54% of the time
NESM+GEFS beats GEFS 54% of the time
GEM+GEFS beats GEFS 53% of the time
CCSM4+GEFS beats GEFS 52% of the time
GEOS5+GEFS beats GEFS 51% of the time
GEFS beats GEFS 50% of the time
CCSM4+GEFS
FIMv2+GEFS
GEM+GEFS
GEOS5+GEFS
GEFS
CFSv2+GEFS
NESM+GEFS
SubXMME
Counts
Forecasts
-100
-50
0
50
100
150
200
250
0
100
200
300
400
500
600
700
800
900


GEFS+GEM Baseline with SubXMME: Sign Test on WK34 TAS2M Refcst RMSE for CONUS+AK mth(s): 1-12 (1999-2014)
GEFS+GEM more skillful
GEFS+GEM less skillful
SubXMME beats GEFS+GEM 63% of the time
CFSv2+GEFS+GEM beats GEFS+GEM 61% of the time
FIMv2+GEFS+GEM beats GEFS+GEM 58% of the time
NESM+GEFS+GEM beats GEFS+GEM 57% of the time
CCSM4+GEFS+GEM beats GEFS+GEM 56% of the time
GEOS5+GEFS+GEM beats GEFS+GEM 55% of the time
GEFS+GEM beats GEFS+GEM 50% of the time
CCSM4+GEFS+GEM
FIMv2+GEFS+GEM
GEOS5+GEFS+GEM
GEFS+GEM
CFSv2+GEFS+GEM
NESM+GEFS+GEM
SubXMME
Counts
Forecasts
-100
-50
0
50
100
150
200
250
0
100
200
300
400
500
600
700
800
900


GEFS+GEM+NESM Baseline with SubXMME: Sign Test on WK34 TAS2M Refcst RMSE for CONUS+AK mth(s): 1-12 (1999-2014)
SubXMME beats GEFS+GEM+NESM 60% of the time
CCSM4+GEFS+GEM+NESM beats GEFS+GEM+NESM 59% of the time
FIMv2+GEFS+GEM+NESM beats GEFS+GEM+NESM 59% of the time
CFSv2+GEFS+GEM+NESM beats GEFS+GEM+NESM 59% of the time
GEOS5+GEFS+GEM+NESM beats GEFS+GEM+NESM 57% of the time
GEFS+GEM+NESM beats GEFS+GEM+NESM 50% of the time
GEFS+GEM+NESM more skillful
GEFS+GEM+NESM less skillful
CCSM4+GEFS+GEM+NESM
FIMv2+GEFS+GEM+NESM
GEFS+GEM+NESM
GEOS5+GEFS+GEM+NESM
CFSv2+GEFS+GEM+NESM
SubXMME
Counts
Forecasts
-100
-50
0
50
100
150
200
0
100
200
300
400
500
600
700
800
900

SubX Weeks 3-4 Summary:

- **SubXMME is most frequently the most skillful forecast for both Regional Skill Scores and the Sign Test across multiple metrics**
- **As individual models, GEFS is most skillful, and also has the most members in the hindcast**
- **SubX models are adding skill to all three levels and for both precipitation and temperature. This is also generally true in the seasonal analyses for this metric and HSS, ACC, and BSS.**
- **Additional thoughts….**
  - It is likely that model diversity is adding value
  - Calibration
  - Weighting schemes
  - More realtime testing
    - value added to the operational suite?
    - SubX incorporated into a consolidation tool?