Evaluation of Ensemble Icing Probability Forecasts in NCEP's SREF, VSREF and NARRE-TL Systems

Binbin Zhou[1,2], Jun Du[2], Geoff DeMigo[2]
and Robert Sallee[3]

1. I.M. System Group
2. Environmental Modeling Center of NCEP
3. Wyle Info Systems & Aviation Weather Center of NCEP

15th Conference on ARAM, Aug 1-4, 2011, Los Angles, CA

NCEP SREF/VSREF/NARRE-TL ICING products

Existed for a while but never been verified objectively

Different from AWC-FIP, SREF/VSREF/NARRE-TL's icing probability product is based on a simple diagnosis when:

T, RH and vertical velocity fall in certain ranges (suggested by AWC)

SREF Icing ensemble probability (6Eta, 5 NMM, 5ARW, 5 RSM):

Computed from 21 equal-weighted SREF members, 32 km res

VSREF/NARRE-TL Icing ensemble probability (6 previous RUC/RR, 4 previous NAM):

VSREF: Computed from 10 unequal-weighted VSREF members including 6 RUC and 4 NAM time-lagged runs, 13km resolution

NARRE-TL

Time Lagged N. America Rapid Refresh Ensemble System

  Similar to VSREF but with 6 RR and 4 NAM time-lagged runs

  RR: WRF ARW + GSI with larger domain + Alaska

  After Nov 0f 2011, RUC/VSREF will be stopped

ADDS (Aviation Digital Data Service, by AWC & NCAR)

 Analysis data, including CIP, FIP and others (GTG, V&C, etc …)

 CIP – Current Icing Potential (Obsv + RUC analysis)

 FIP – Forecast Icing Potential (by RUC analysis)

 Generated at AWC, GRIB data transferred to NCEP/EMC

ADDS/CIP

Combination of satellite data, radar, METAR, pilot report,
Lightning data and RUC to generate a 3-D diagnosis of icing
potential, super-cooed liquid water droplet (SLD) and icing
severity.

Choose CIP severity as icing truth

CIP Severity: 0, 1, 2, 3, and 4
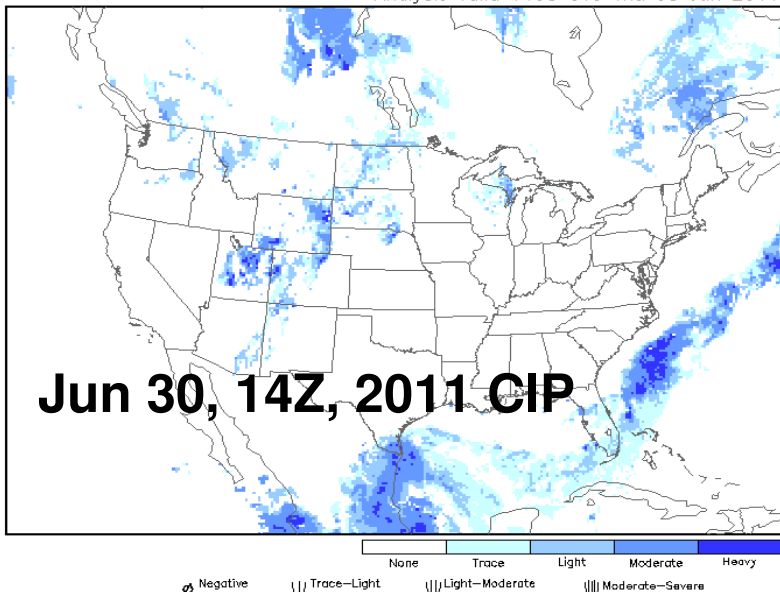   0 – no icing, 1 – trace, 2 – light, 3 – moderate , 4 – severe.
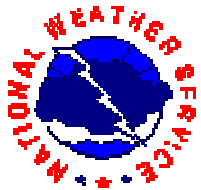
# Example: SREF, VSREF, NARRE-TL icing Prob: Compared to ADDS-CIP severity

By FAA policy CIP is a Supplementary Weather Product for enhanced situational awareness only and must be used with one or more primary products (safety decision) such as an AIRMET or SIGMET (see AIM 7-1-3).

Icing severity at FL190

Analysis valid 1400 UTC Thu 30 Jun 2011

Jun 30, 14Z, 2011 CIP

None   Trace   Light   Moderate   Heavy

Ø Negative   Trace-Light   Light-Moderate   Moderate-Severe

SREF: Probability of icing at FL180 06H FCST
from 09z Jun 30 2011. Verified Time: 15z 06/30/2011
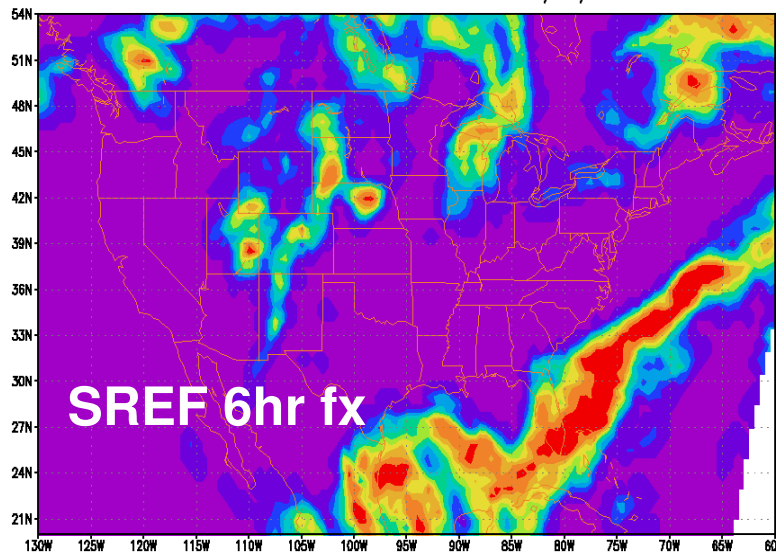
SREF 6hr fx

VSREF: Probability of Icing at FL180 06H FCST
from 08z Jun 30 2011. Verified Time: 14z 06/30/2011
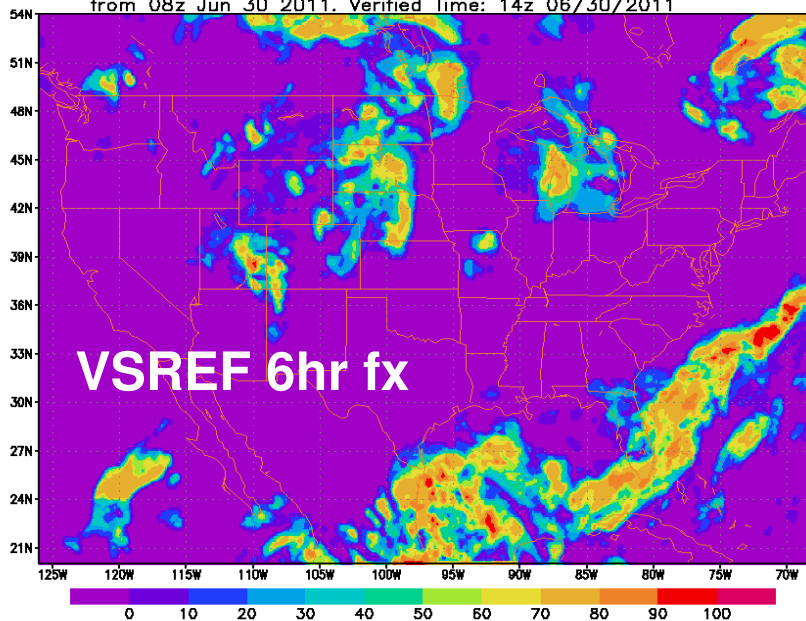
VSREF 6hr fx

0   10   20   30   40   50   60   70   80   90   100

NARRE: Probability of Icing at FL180 06H FCST
from 08z Jun 30 2011. Verified Time: 14z 06/30/2011

NARRE-TL 6hr fx

0   10   20   30   40   50   60   70   80   90   100
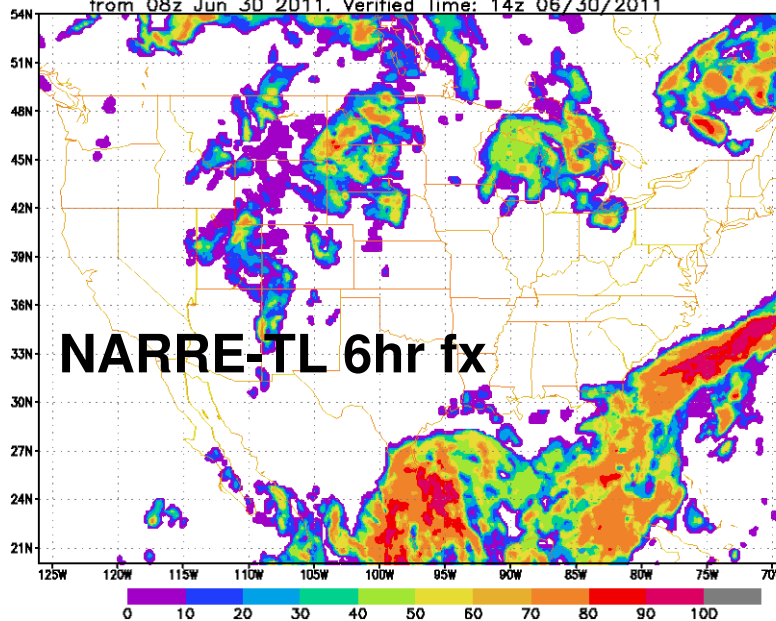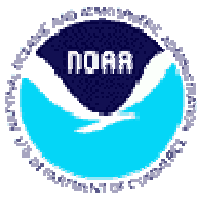
Verification Method

Prob-Event (P-E) grid-to-grid verification

Icing "Probability ≥ Pt" against Icing "Event (severity ≥ 1)"
where Pt is a series of ensemble prob thresholds: 10, 20, …90%

1. Deterministic: FHO (Fcst – Hit -- Obsv, under a prob thresholds Pt)

Fcst: Prob ≥ Pt ←→ Obsv: CIP ≥ 1

Hit:	**Prob ≥ Pt && CIP ≥ 1**
False alarm:	**Prob ≥ Pt && CIP < 1**
Miss:	**Prob < Pt && CIP ≥ 1**

Then compute POD, FAR, Bias, ETS, traditional scores

Repeat with different Pt (10%, … 50%, … 90 %)

2. Probabilistic: Use PBS (Probabilistic Brier Score)
to compute reliability within 10 probability bins

Verification time: 60 days from April 28 ~ June 28, 2011

Bias Comparison

SREF trends to over predict icing

VSREF and NARRE-TL under predict icing

Decrease with Prob thresholds

pod P900 ICETRC Error averaged by Threshold from 20110428 to 20110628

SREF

FL240
FL180
FL090
FL060
FL030

Pt: 1 10 20 30 40 50 60 70 80 90%

pod P900 ICETRC Error averaged by Threshold from 20110428 to 20110628

VSREF

FL240
FL180
FL090
FL060
FL030

Pt: 1 10 20 30 40 50 60 70 80 90%

NARRE-TL

FL240
FL180
FL090
FL060
FL030

Pt: 1 10 20 30 40 50 60 70 80 90%

POD comparison

SREF > VSREF/NARRE-TL

Very higher levels are worse

Decrease with prob thresholds

pofd P900 ICETRC Error averaged by Threshold from 20110428 to 20110628

SREF

FL240
FL180
FL090
FL060
FL030

Pt: 1   10   20   30   40   50   60   70   80   90%

pofd P900 ICETRC Error averaged by Threshold from 20110428 to 20110628

VSREF

FL240
FL180
FL090
FL060
FL030

Pt: 1   10   20   30   40   50   60   70   80   90%

NARRE-TL

FL240
FL180
FL090
FL060
FL030

Pt: 1   10   20   30   40   50   60   70   80   90%

FAR(rate) comparison

SREF > VSREF/NARRE-TL
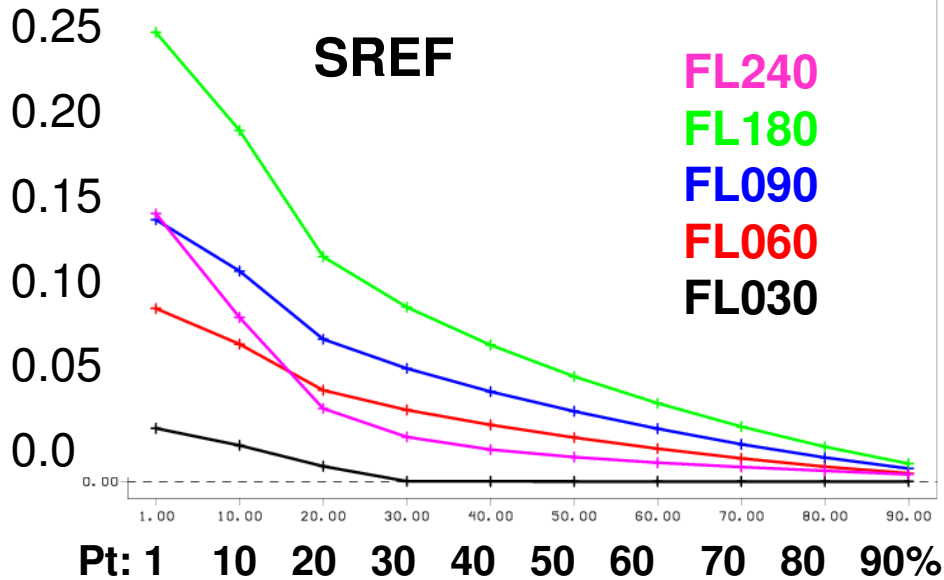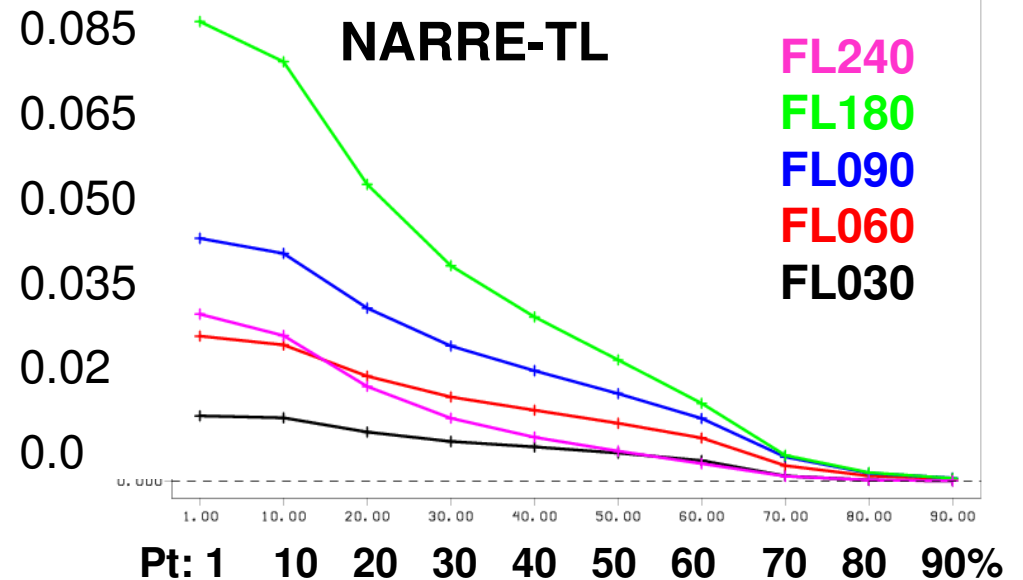
Decrease with prob thresholds

ets P900 ICETRC Error averaged by Threshold from 20110428 to 20110628

SREF	VARB: ICETRC	RGN: NORTH-AMERICA	LVL: P900	STAT: ETS
SREF	VARB: ICETRC	RGN: NORTH-AMERICA	LVL: P800	STAT: ETS
SREF	VARB: ICETRC	RGN: NORTH-AMERICA	LVL: P725	STAT: ETS
SREF	VARB: ICETRC	RGN: NORTH-AMERICA	LVL: P500	STAT: ETS
SREF	VARB: ICETRC	RGN: NORTH-AMERICA	LVL: P400	STAT: ETS

OBSERVATION COUNTS:
66493 66493 66493 66493 66493 66493 66493 66493 66493 66493

FL240
FL180
FL090
FL060
FL030

SREF

Pt: 1 10 20 30 40 50 60 70 80 90%
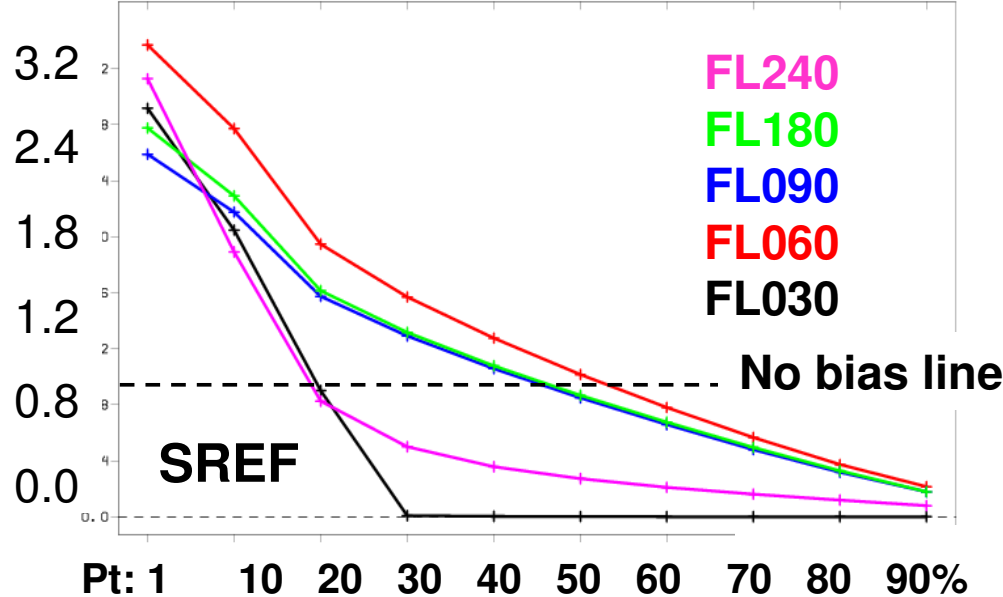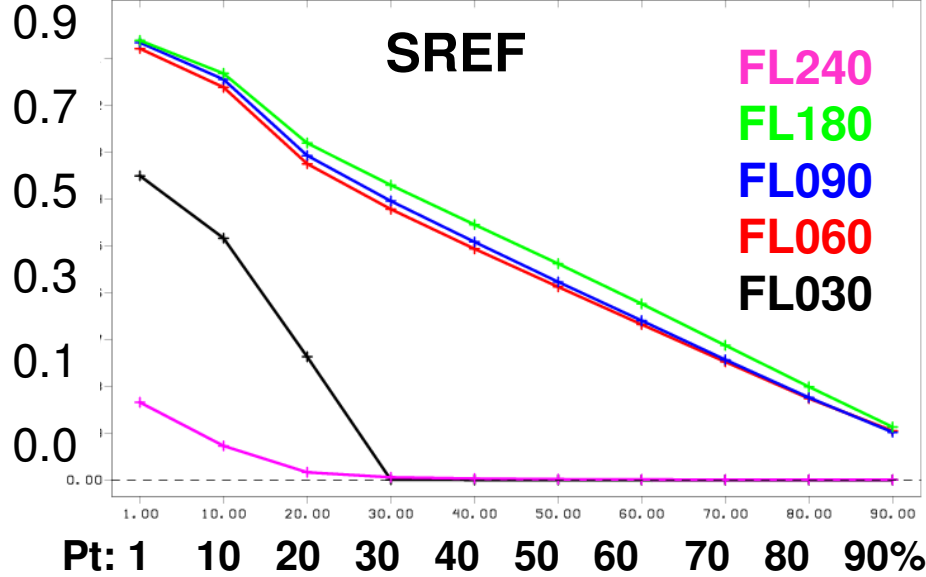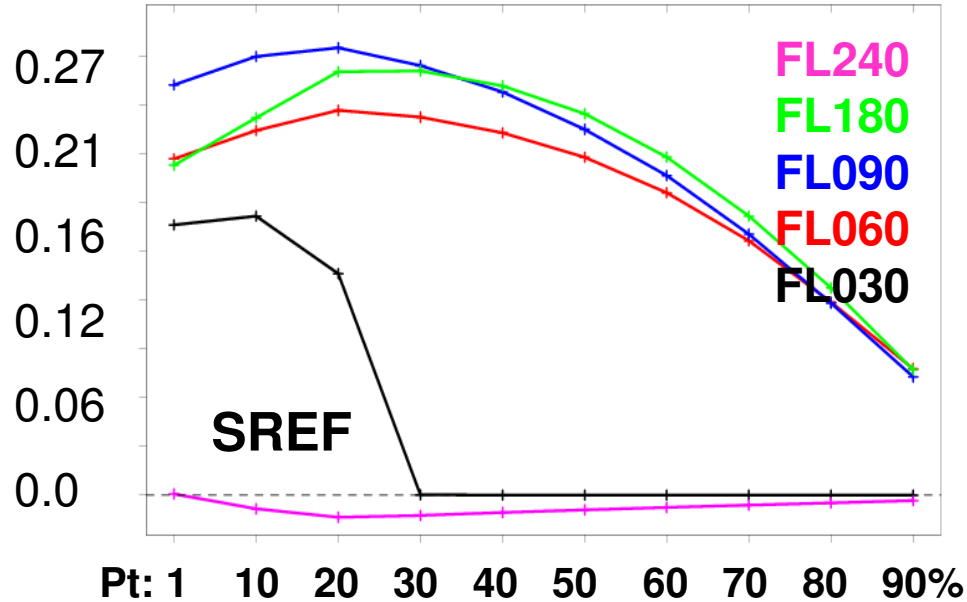
ets P900 ICETRC Error averaged by Threshold from 20110428 to 20110628

VSREF	VARB: ICETRC	RGN: CONUS	LVL: P900	STAT: ETS
VSREF	VARB: ICETRC	RGN: CONUS	LVL: P800	STAT: ETS
VSREF	VARB: ICETRC	RGN: CONUS	LVL: P725	STAT: ETS
VSREF	VARB: ICETRC	RGN: CONUS	LVL: P500	STAT: ETS
VSREF	VARB: ICETRC	RGN: CONUS	LVL: P400	STAT: ETS

OBSERVATION COUNTS:
17E05 17E05 17E05 17E05 17E05 17E05 17E05 17E05 17E05 17E05

FL240
FL180
FL090
FL060
FL030

VSREF

Pt: 1 10 20 30 40 50 60 70 80 90%

ETS (Equitable threat score)

NARRE-TL and VSREF better than SREF

Very high levels: worse

Very low levels: worse (SREF)

Decrease with prob thresholds

~ ETS of Precip but < FIP (0.3~0.5)

ets P900 ICETRC Error averaged by Threshold from 20110428 to 20110628

| NARRE | VARB: ICETRC | RGN: CONUS | LVL: P400 | STAT: ETS |

OBSERVATION COUNTS:
16E05 16E05 16E05 16E05 16E05 16E05 16E05 16E05 16E05 16E05

FL240
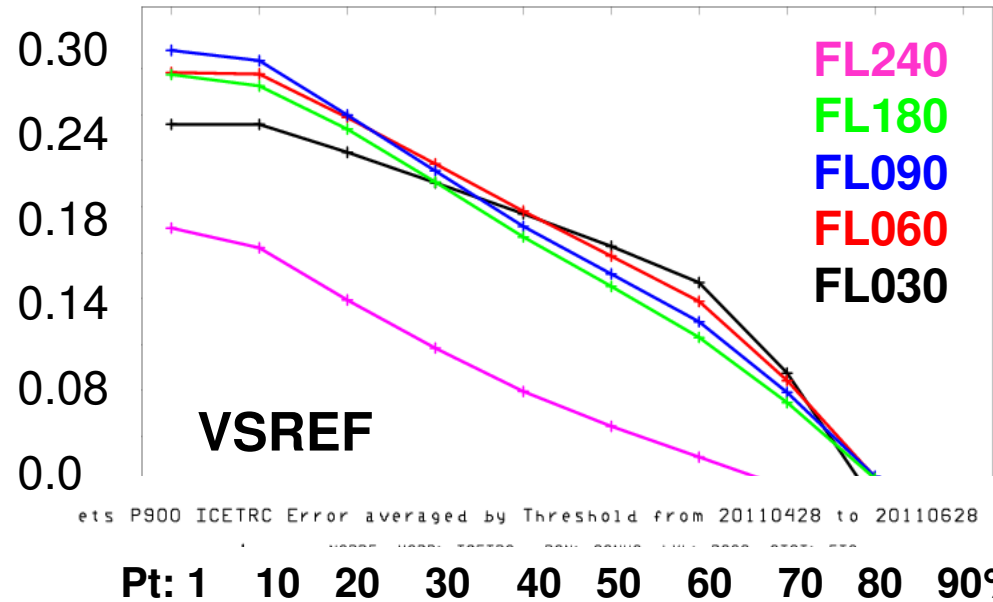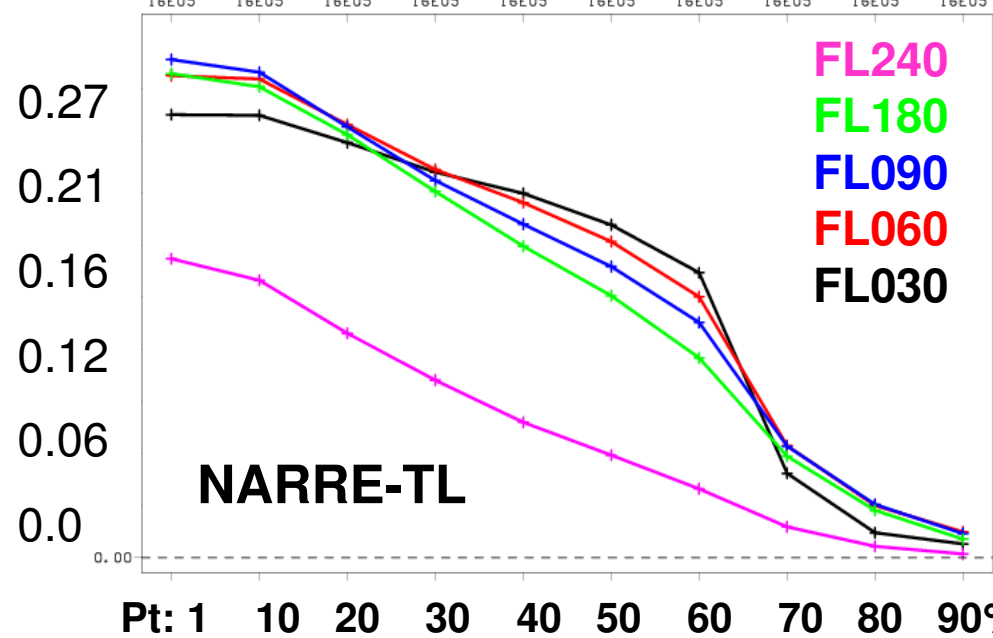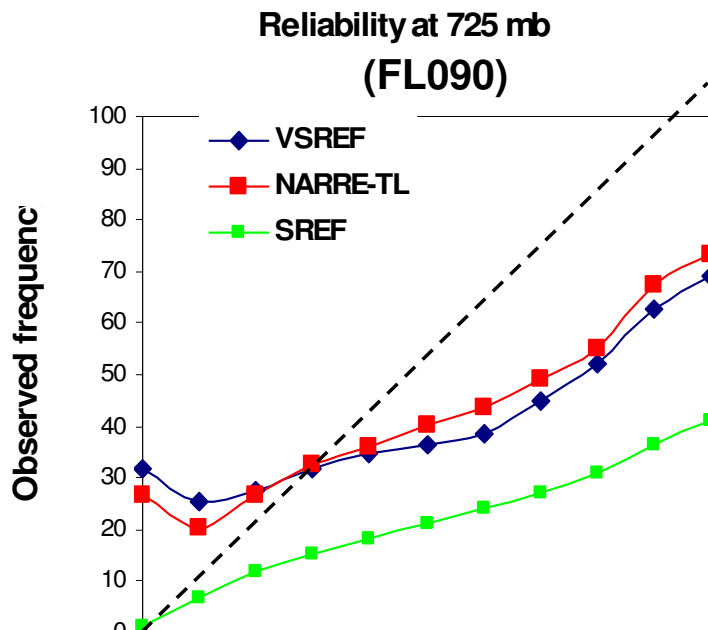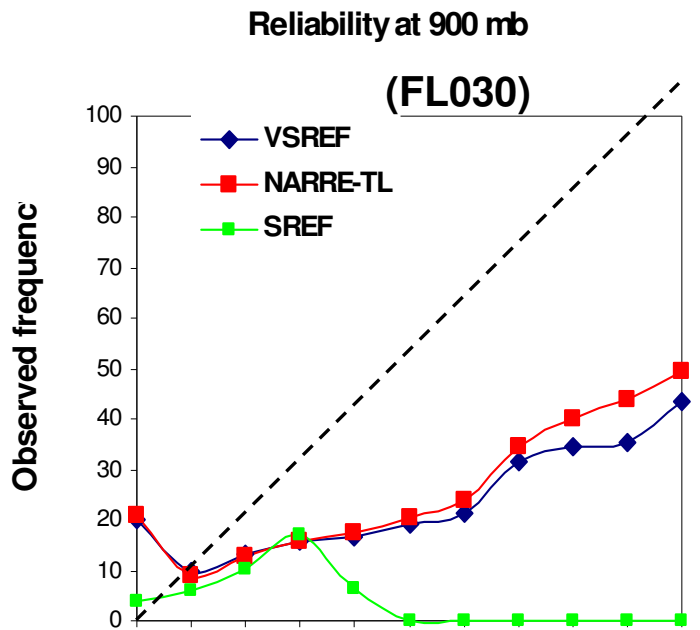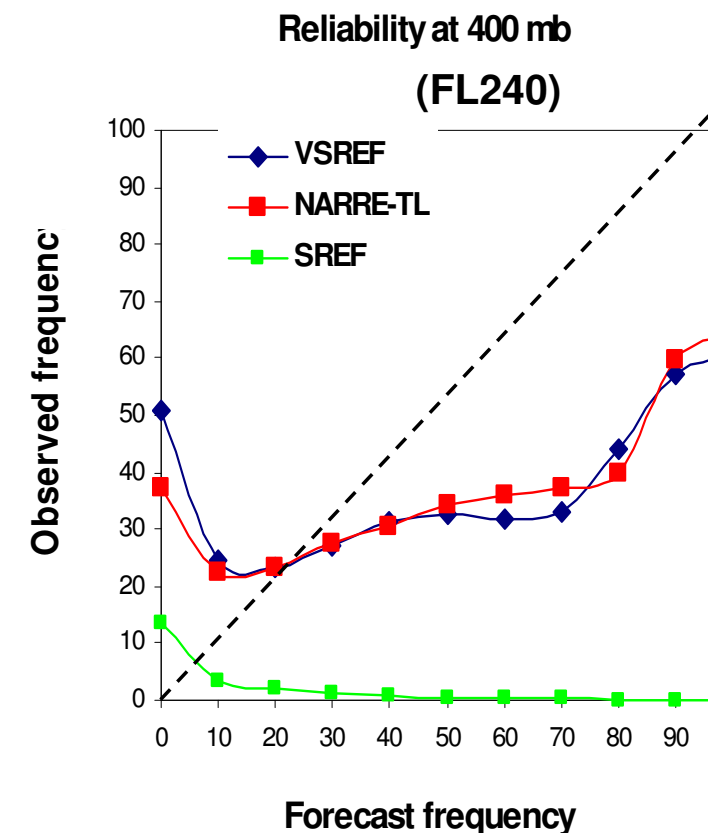FL180
FL090
FL060
FL030

NARRE-TL

Pt: 1 10 20 30 40 50 60 70 80 90%

Reliability at 900 mb (FL030)

Reliability at 725 mb (FL090)

Reliability at 500 mb (FL180)

Reliability at 400 mb (FL240)

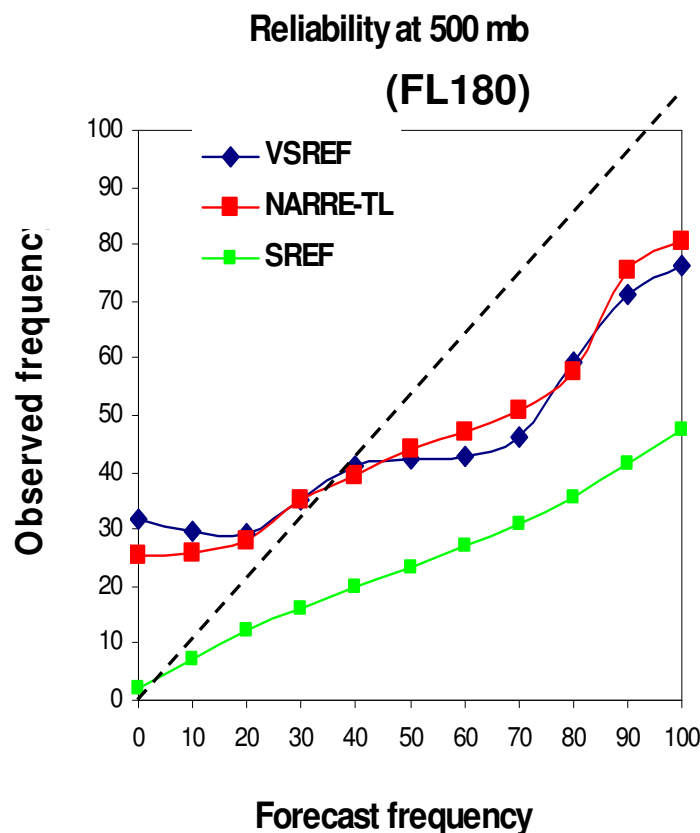Observed frequency vs Forecast frequency

Legend: VSREF, NARRE-TL, SREF

Reliability Plot

VSREF/NARRE-TL better than SREF

At mid-Flight levels

All have best reliability

NCEP
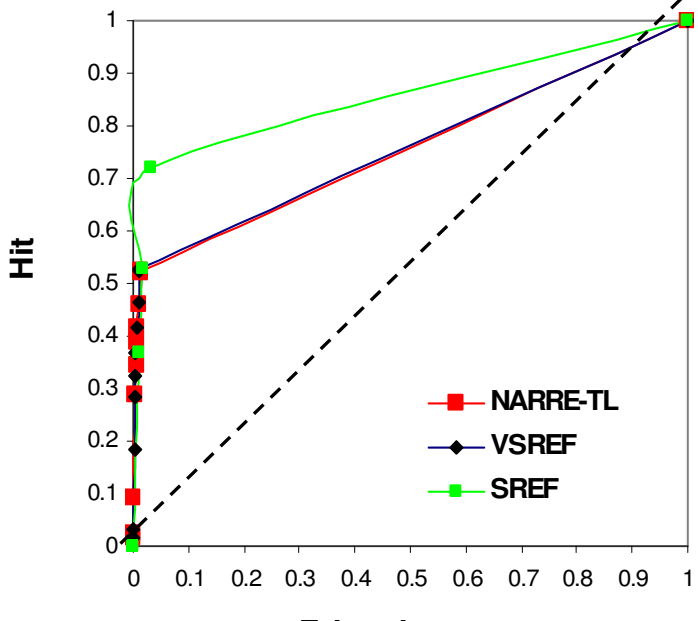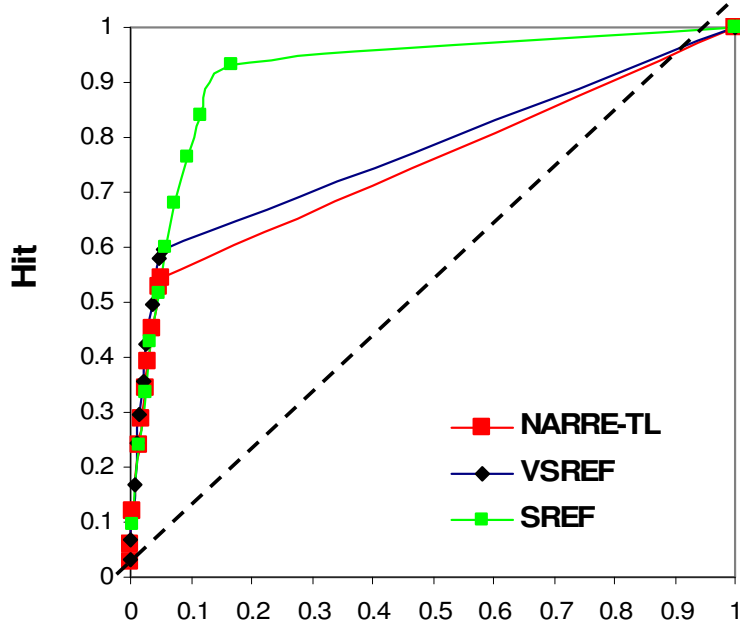
ROC diagram

All are skillful
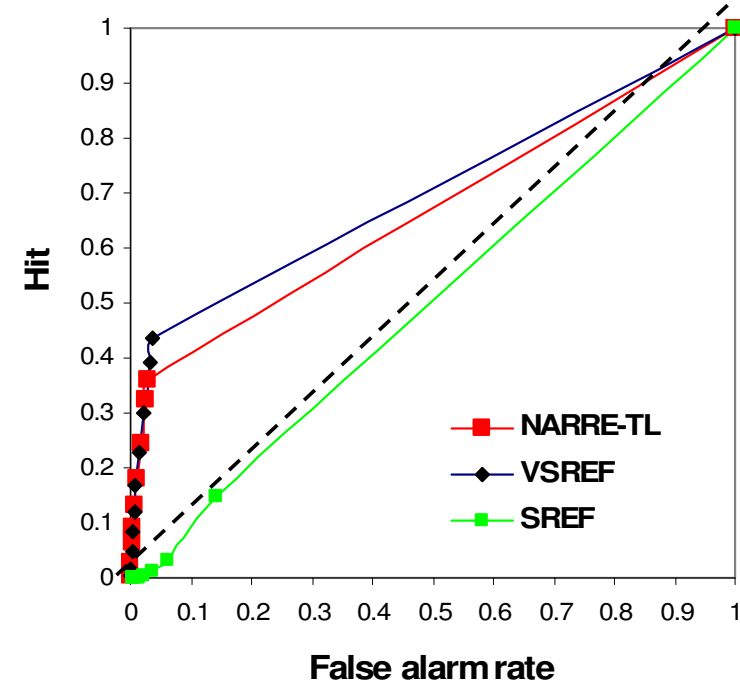
except for SREF
at lower level
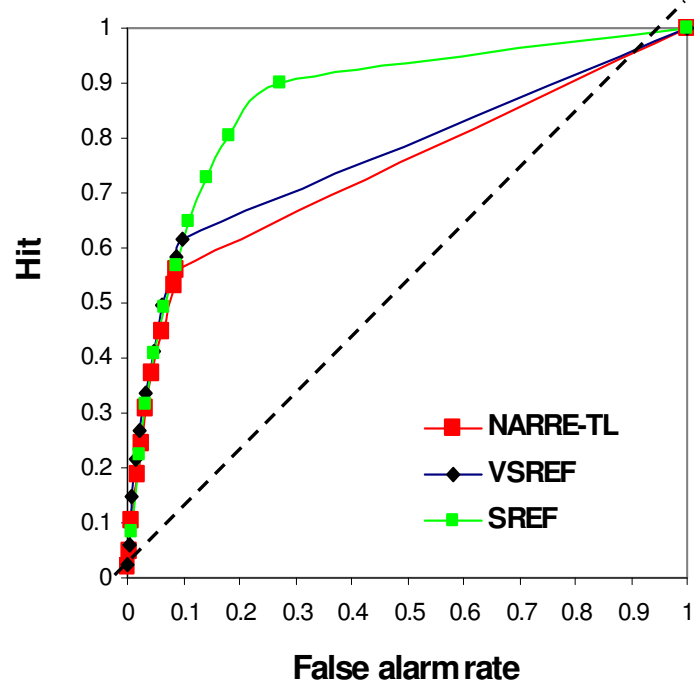
Conclusion

- Using "trace" CIP as truth, SREF/VSREF/NARRE-TL were verified over 60 days over CONUS

- Except for very high/low levels, SREF/VSREF/NARRE-TL icing probability forecasts have ETS score around 0.2~0.3 ~ to Precip but worse than FIP

- Very high/Low levels icing prob forecasts has worse scores

- Reliability plots: VSREF ~ NARRE-TL better than SREF

- VSREF/NARRE-TL/SREF have good ROC score except for SREF in very high levels