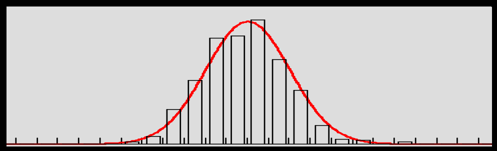
Part II:

Redefining the Ensemble Spread-Skill Relationship from a Probabilistic Perspective

Traditional Ensemble Spread-Skill Relationship

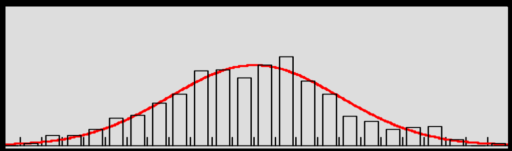
- Based on the premise that ensemble spread should provide a forecast of forecast error.

agreement



smaller forecast errors

disagreement



larger forecast errors

- Often characterized by the linear relationship between ensemble spread and forecast error -- the "spread-error correlation"
- Assumes:
  - A linear dependency between ensemble spread and forecast error
  - An end user that has a continuous sensitivity to forecast error

The Real Deal

- In theory, for a perfect ensemble of infinite size…
  - The strength of the correlation between ensemble spread (σ) and the ensemble mean forecast error ($|e_{EM}|$) is limited by the case-to-case spread variability (β).

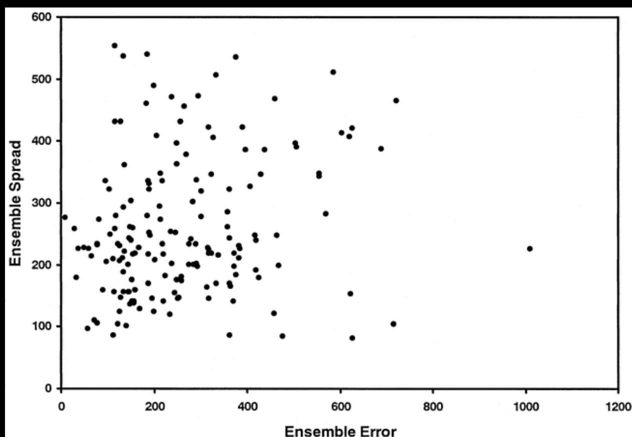
$$\rho^2(\sigma,|e_{EM}|) = \frac{2}{\pi} \frac{1-\exp(-\beta^2)}{1-\frac{2}{\pi}\exp(-\beta^2)} ; \beta=\mathrm{std}(\ln \sigma)$$

(Houtekamer, 1993)

  - Even with infinite spread variability, spread and error are not perfectly correlated ($\rho < 0.8$).

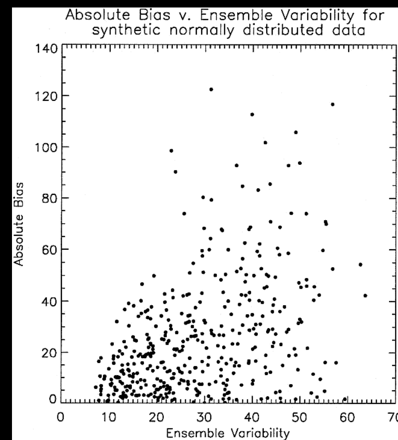
Disappointing Results

Tropical Cyclone Tracks



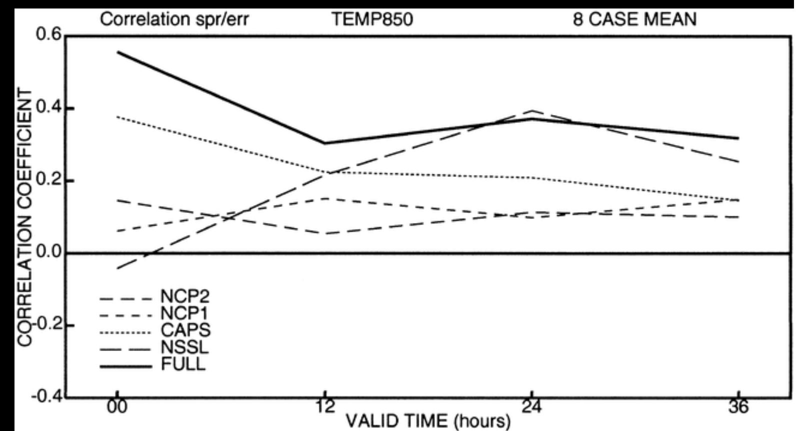
[c.f. Goerss 2000]

NCEP SREF Precipitation



[c.f. Hamill and Colucci 1998]

SAMEX '98 SREFs

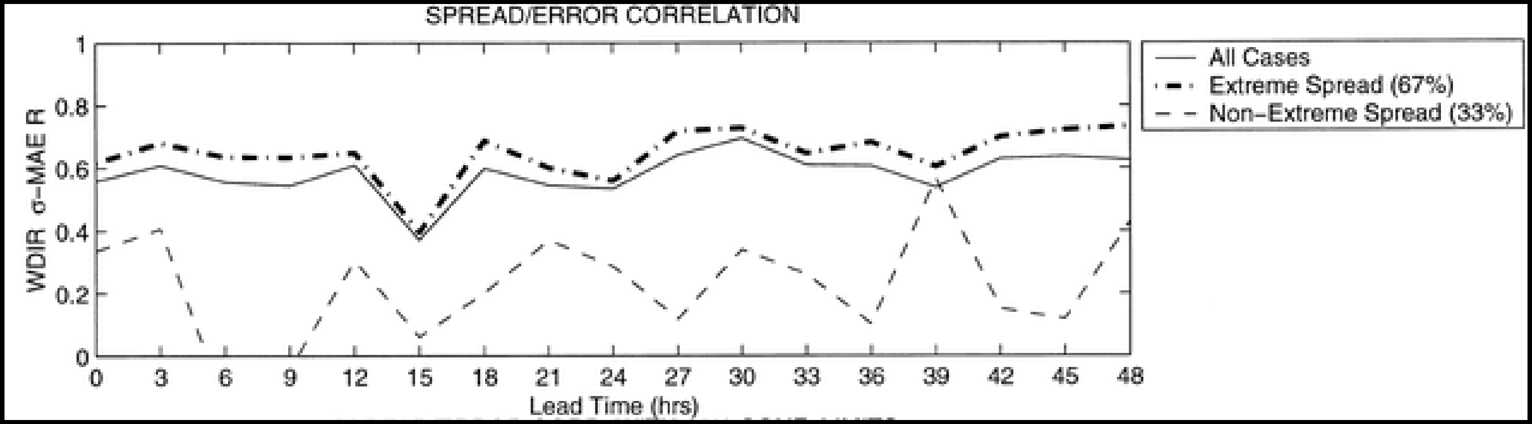


[c.f. Hou et al. 2001]

- **Highly scattered relationships, thus low correlations**
  - **Often less than 0.4**

Encouraging Results

UW MM5 SREF 10-m Wind Direction



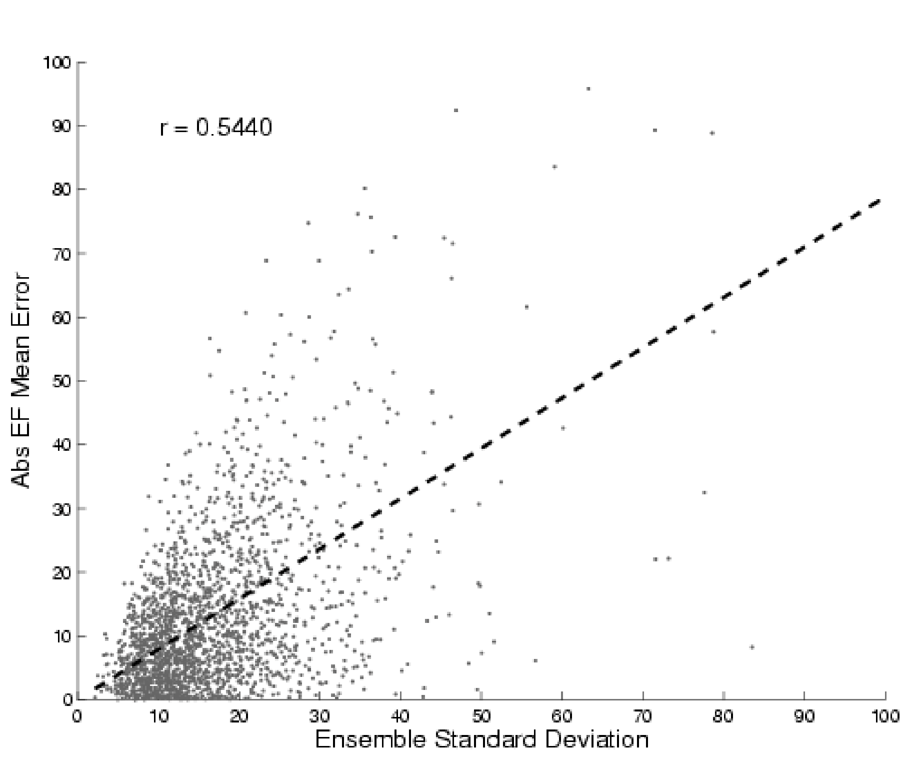
[c.f. Grimit and Mass 2002]

- More recent studies show that spatially-averaged spread-error correlations can be as high as 0.6-0.7
  (Grimit and Mass 2002, Stensrud and Yussouf 2003)

- Potentially higher correlations can be achieved by considering only cases with extreme spread

An Inherently Deterministic Approach

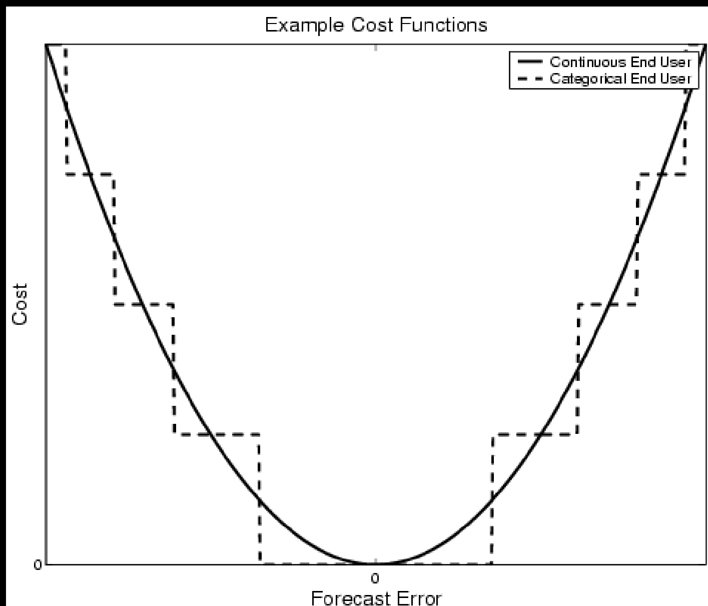
- The expected value of the absolute forecast error is estimated in the regression.
- Therefore, only an unsigned, deterministic error forecast is generated.
- The skill associated with such predictions is very limited.

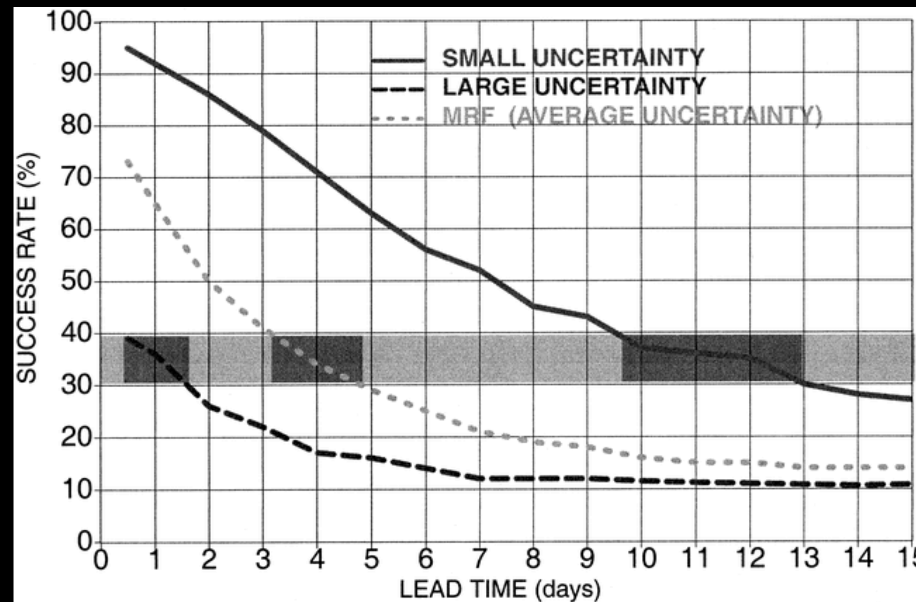


Idealized, statistical ensemble forecasts.

N = 2500
M = 50; β = 0.5

A Categorical Approach





c.f. Toth et al. 2001

- Some have concluded that categorical measures of forecast spread are more skillful predictors of forecast accuracy
  (Toth et al. 2001, Ziehmann 2001)
  - e.g. – statistical entropy (ENT), mode population (MOD)
  - Requires that forecasts/verification be divided into predetermined bins

- Need idealized Houtekamer-type investigation to verify

A Simple Stochastic Model of Spread-Skill

An extension of the Houtekamer (1993) model of spread-skill

PURPOSES:

1) To establish practical limits of forecast error predictability that could be expected given ideal ensemble forecasts of finite size.

2) To address the user-dependent nature of forecast error estimation by employing a variety of spread and error metrics.

3) To extend forecast error prediction to a probabilistic framework.

A Simple Stochastic Model of Spread-Skill

1. Draw today's "forecast uncertainty" from a log-normal distribution (Houtekamer 1993 model).

   $$\ln(\sigma) \sim N(\ln(\sigma_f), \beta^2)$$

2. Create synthetic ensemble forecasts by drawing M values from the "true" distribution.

   $$F_i \sim N(Z, \sigma^2) \;;\; i = 1,2,\ldots,M$$

3. Draw the verifying observation from the same "true" distribution (statistical consistency).

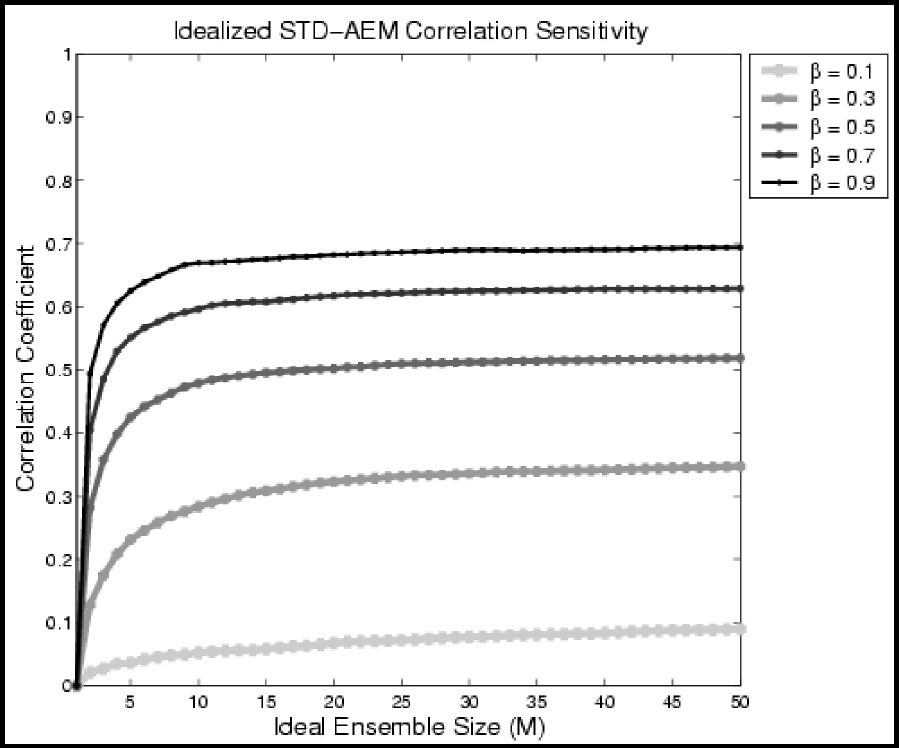
   $$V \sim N(Z, \sigma^2)$$

- Statistical ensemble forecasts at a single, arbitrary location
- 10^4 realizations (cases)
- Assumed:
  - Gaussian statistics
  - statistically consistent (perfectly reliable) ensemble forecasts
- Varied:
  - temporal spread variability (β)
  - finite ensemble size (M)
  - spread and skill metrics (continuous and categorical)

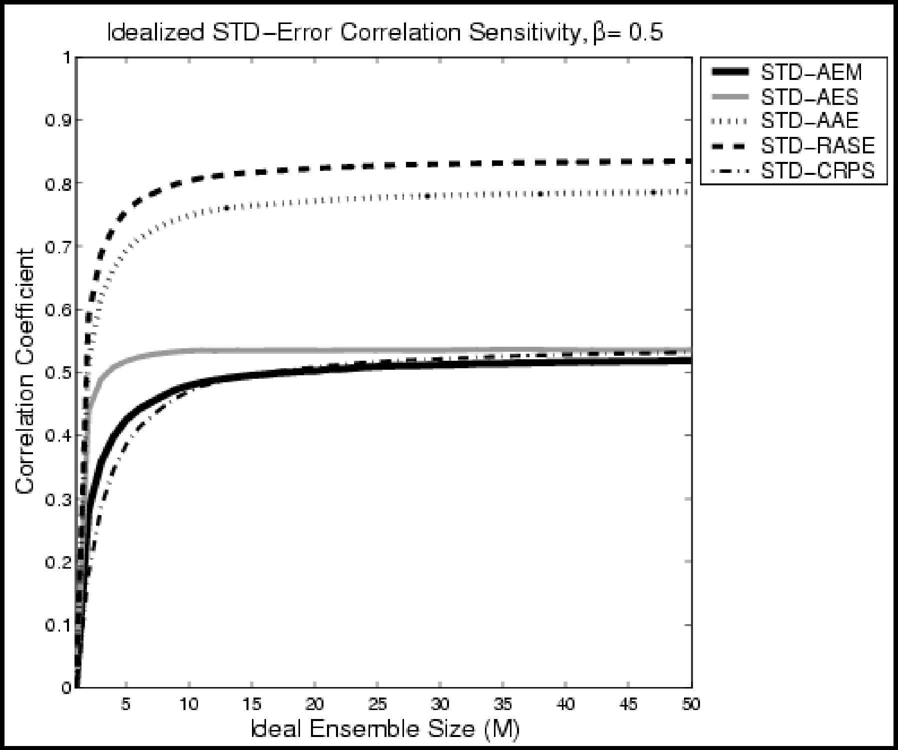
Idealized Spread-Error Correlations

N = 10000
β = 0.5

STD-AEM correlation



STD-error correlation



spread
STD = Standard Deviation

error
AEM = Absolute Error of the ensemble Mean

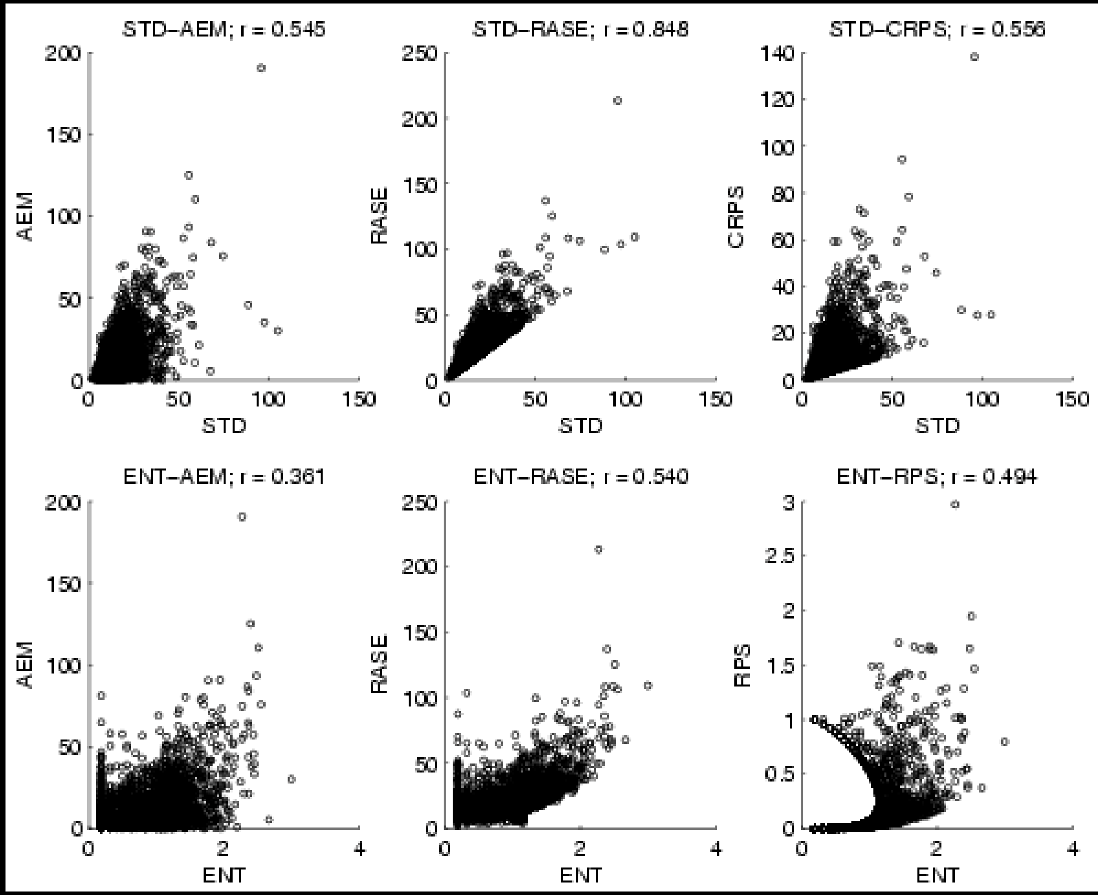
error
AES = Absolute Error of a Single ensemble member
AAE = ensemble-Average Absolute Error
RASE = square Root of ensemble-Average Squared Error
CRPS = Continuous Ranked Probability Score

Idealized Spread-Error Scatter Diagrams

N = 10000
M = 50; β = 0.5

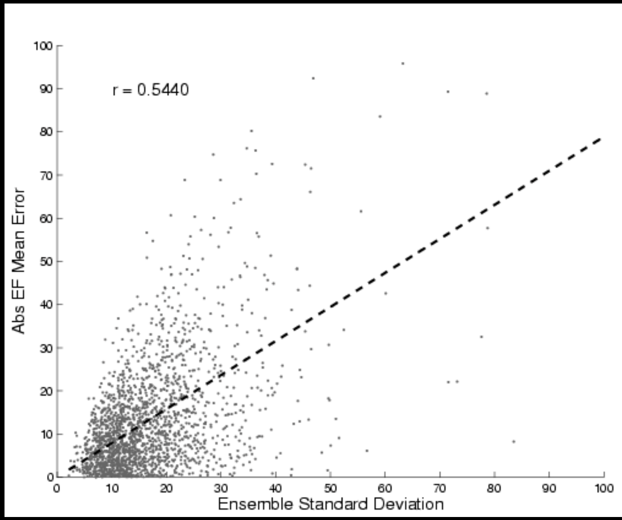
(continuous-continuous)

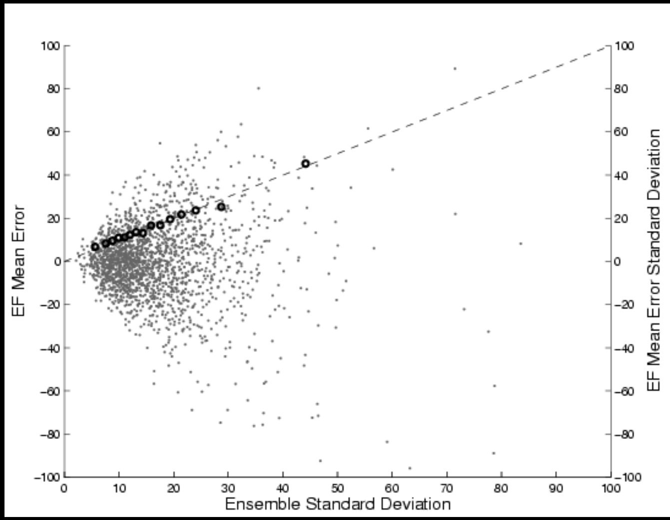
(categorical-continuous)



(categorical-categorical)

A Probabilistic Perspective

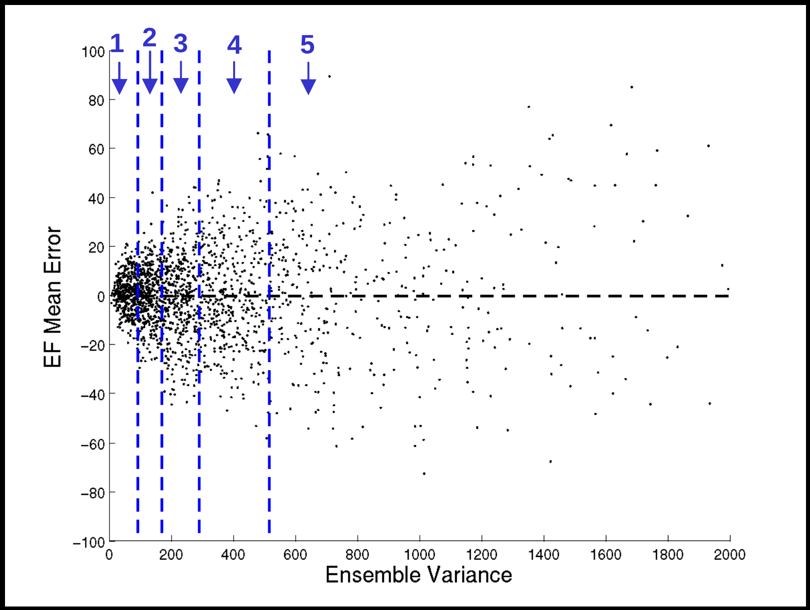




- Connection between statistical consistency and the spread-skill relationship:
  - Expect forecast variance and error variance to coincide
  - "Skill" part of spread-skill relationship needs to be understood as the error variance, not the error itself
  - Thus, statistical consistency and spread-skill association are <u>related</u> concepts!

Conditional Error Climatology (CEC)

- **Use historical errors, conditioned by spread category, as probabilistic forecast error predictions**
  - **Tradeoff between number of bins and number of samples**
  - **Variance-based conditional error climatology method: VAR-CEC**
  - **Evaluate skill by cross-validation, relative to the overall error climatology: ERR-CLI**



Idealized, statistical ensemble forecasts.

N = 2500
M = 50; β = 0.5

Idealized Probabilistic Error Forecast Skill

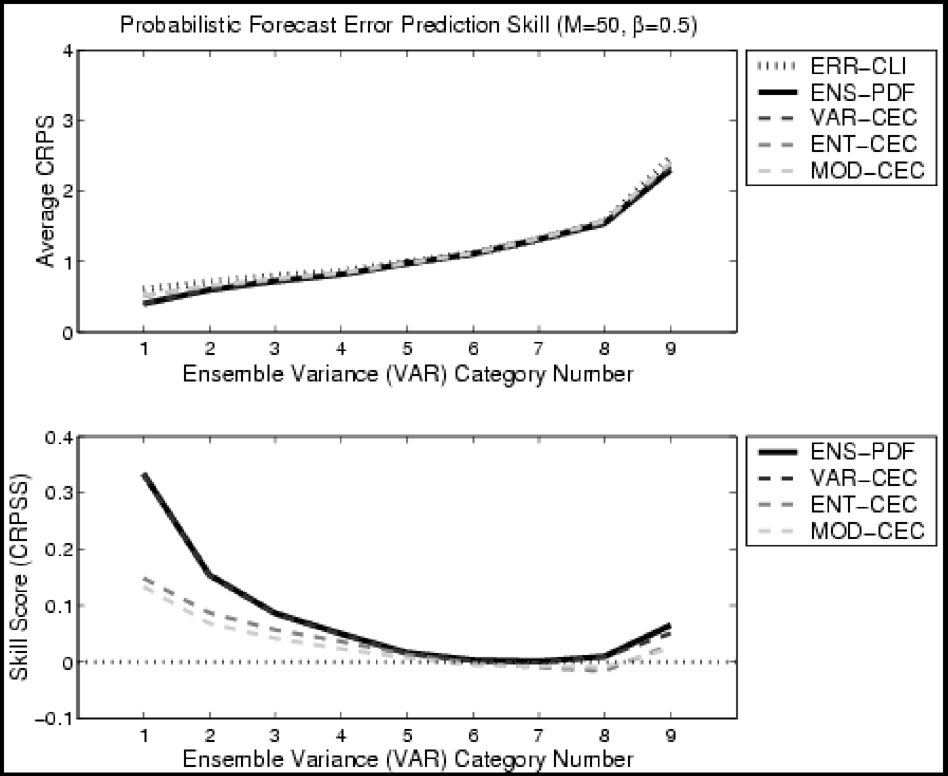
- May use the ensemble variance directly to get a probabilistic error forecast
  ENS-PDF
  - Most skillful approach if PDF is well-forecast

  ENS-PDF CRPSS = 0.060
  VAR-CEC CRPSS = 0.055
  ENT-CEC CRPSS = 0.027
  MOD-CEC CRPSS = 0.021

- VAR-CEC best among spread-based CEC methods when using a continuous verification

- Predictability highest for extreme spread cases
  - Reinforces earlier results

(continuous case)

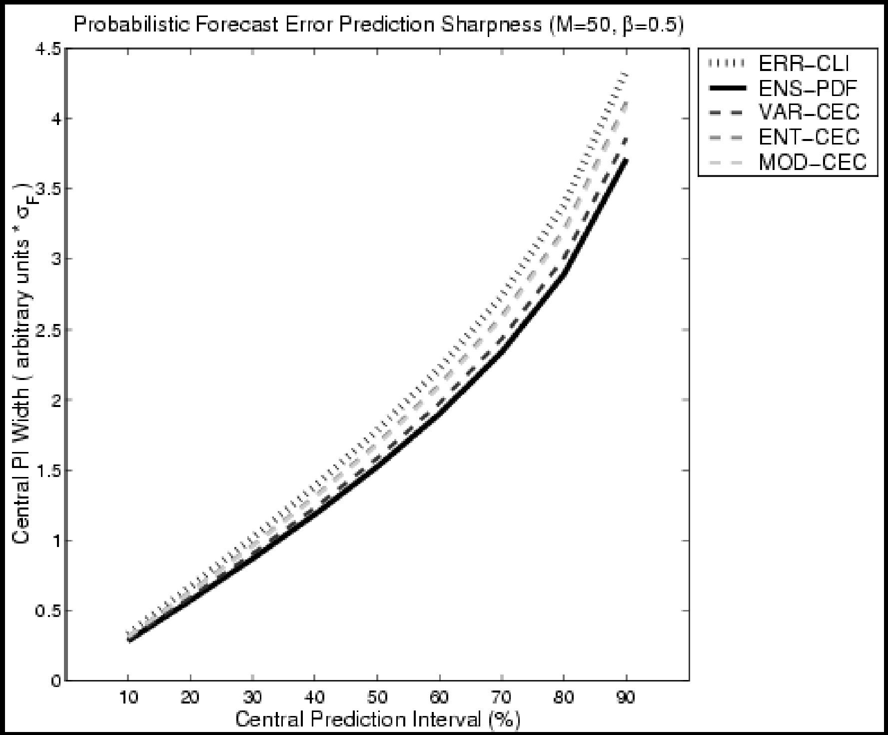


Idealized, statistical ensemble forecasts.

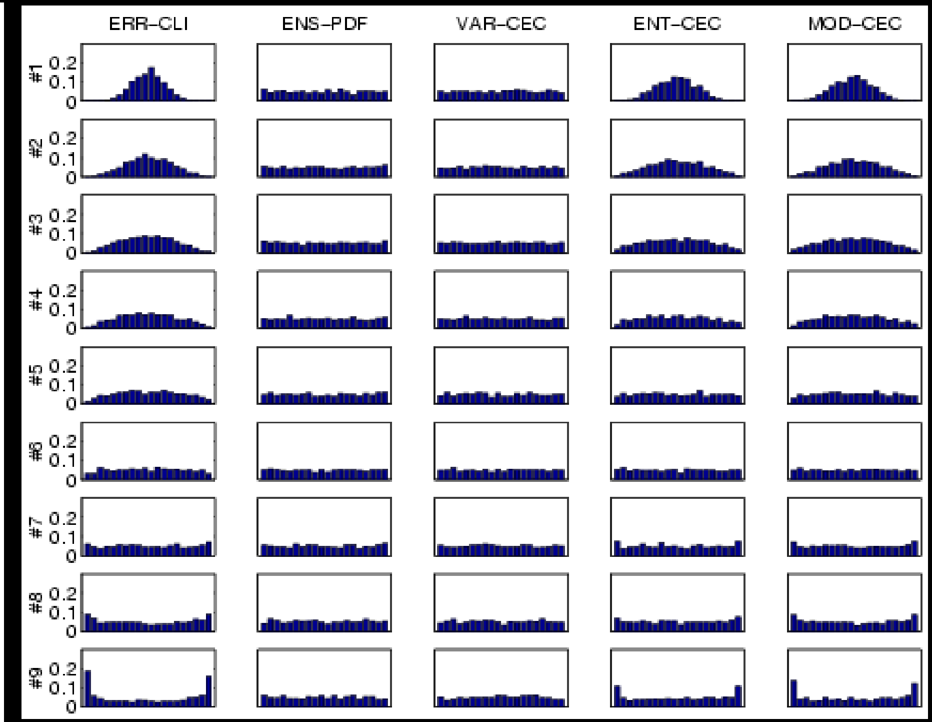
N = 10000
M = 50; β = 0.5

Idealized Probabilistic Error Forecast Skill

Sharpness

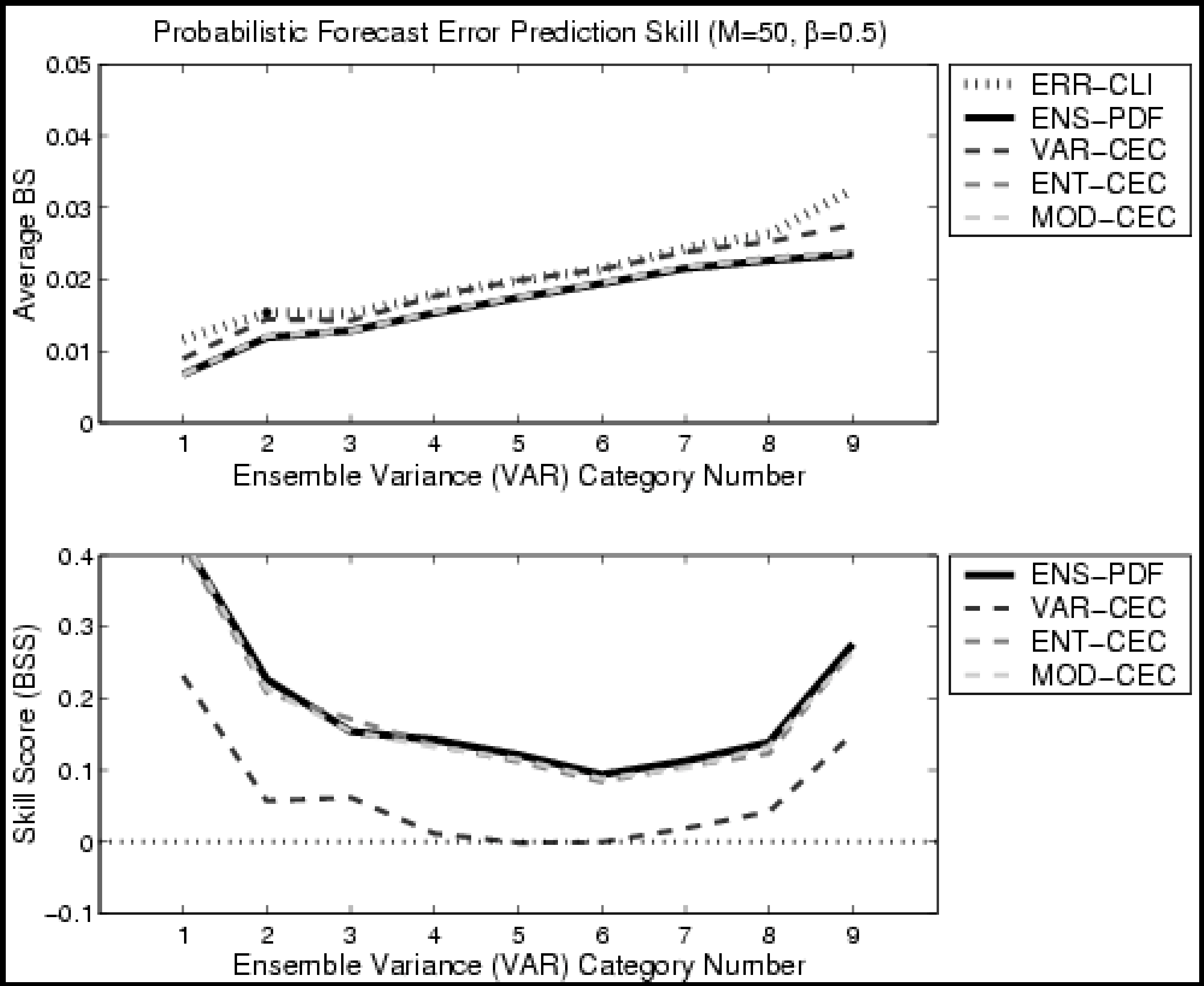


Calibration / Reliability



Idealized Probabilistic Error Forecast Skill

(categorical case)



Idealized, statistical ensemble forecasts.

N = 10000
M = 50; β = 0.5

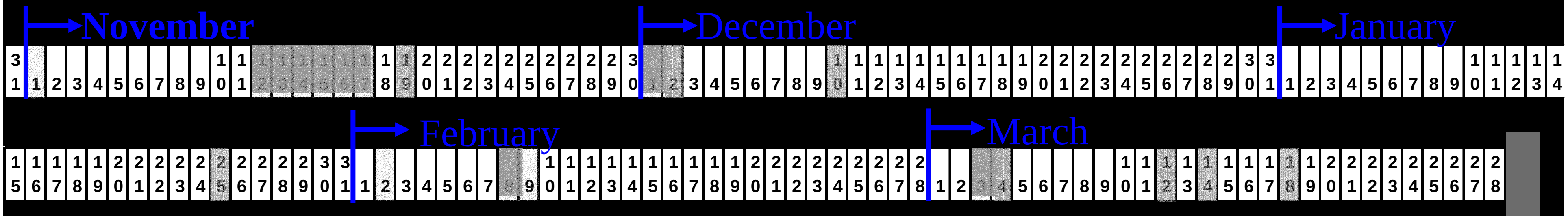
UW SREF System Summary

	Name	# of Members	EF Type	Initial Conditions	Forecast Model(s)	Forecast Cycle	Domain
Homegrown	ACME	17	SMMA	8 Ind. Analyses, 1 Centroid, 8 Mirrors	"Standard" MM5	00Z	36km, 12km
Homegrown	UWME	8	SMMA	Independent Analyses	"Standard" MM5	00Z	36km, 12km
Homegrown	UWME+	8	PMMA	" "	8 MM5 variations	00Z	36km, 12km
Imported	PME	8	MMMA	" "	8 "native" large-scale	00Z, 12Z	36km

ACME: Analysis-Centroid Mirroring Ensemble
PME: Poor-Man's Ensemble

SMMA: Single-Model Multi-Analysis
PMMA: Perturbed-Model Multi-Analysis
MMMA: Multi-model Multi-Analysis

Mesoscale SREF and Verification Data

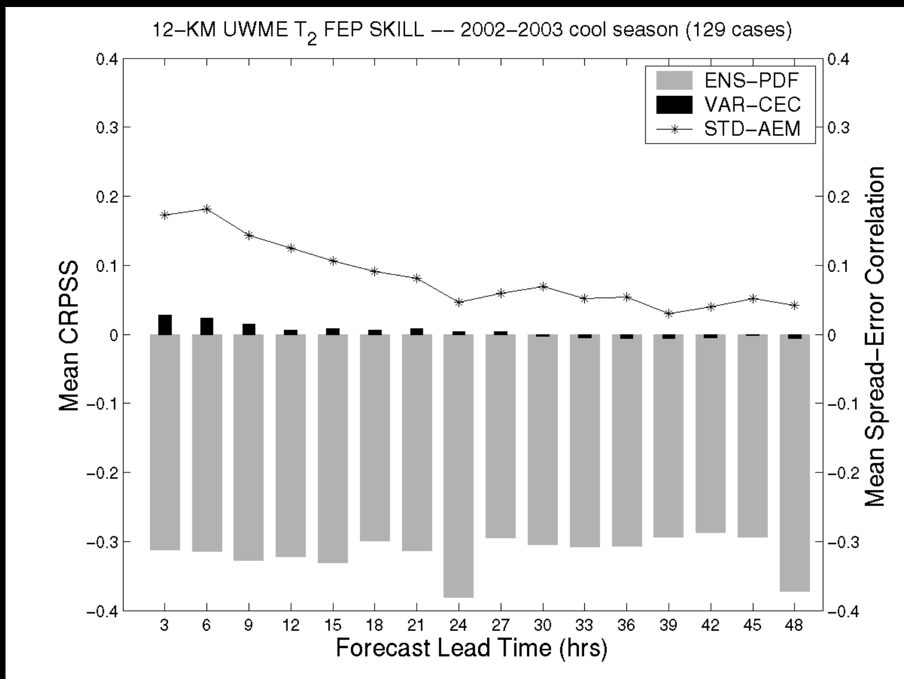


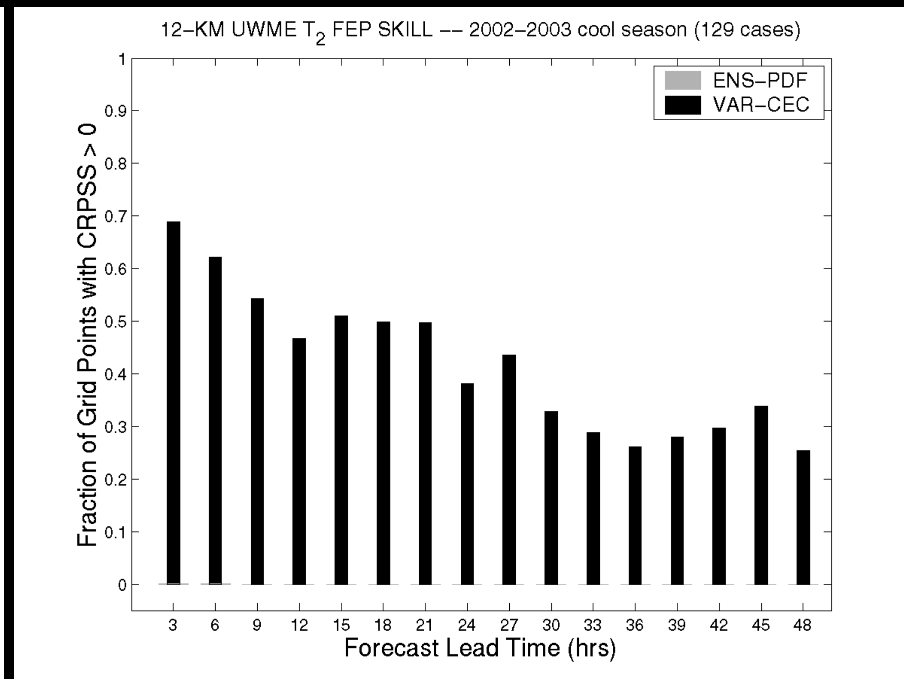
- Mesoscale SREF Data:
  - 129 cases (31 OCT 2002 – 28 MAR 2003)
    - 48h forecasts initialized at 0000 UTC
  - Parameters of Focus:
    - 12 km Domain: Temperature at 2m (**T_2**), Wind Speed and Direction at 10m (**$WSPD_{10}$**, **$WDIR_{10}$**)
  - Short-term mean bias correction
    - Separately applied to: each ensemble member, location, forecast lead time
    - Training window chosen to be 14 days

- Verification Data:
  - 12 km Domain: <u>RUC20 analysis</u>
    (NCEP 20 km mesoscale analysis)
    <u>observations</u>

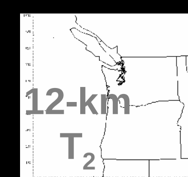
Real Probabilistic Error Forecast Skill

UWME (no bias correction)



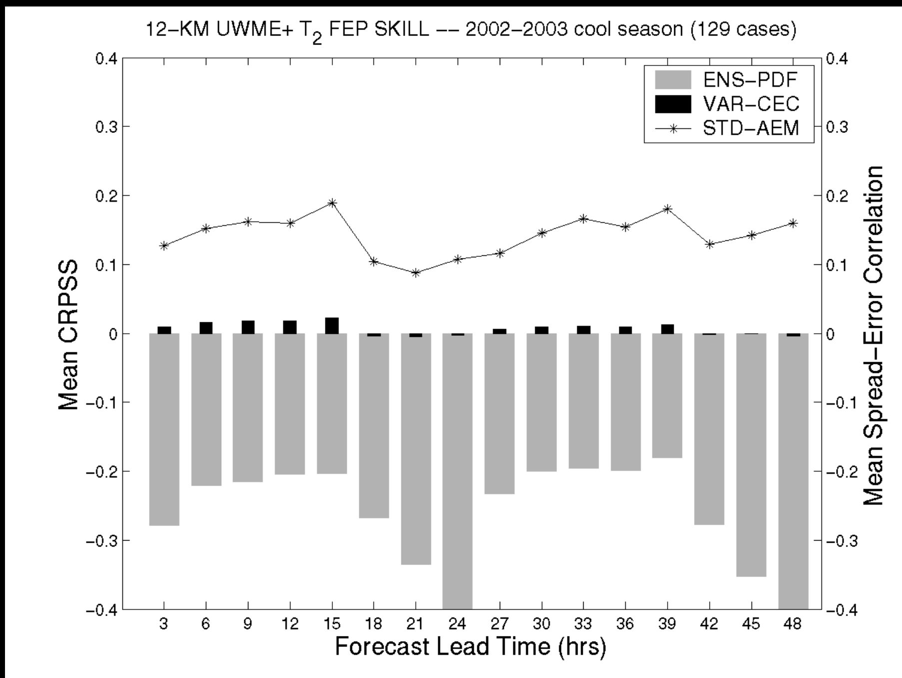


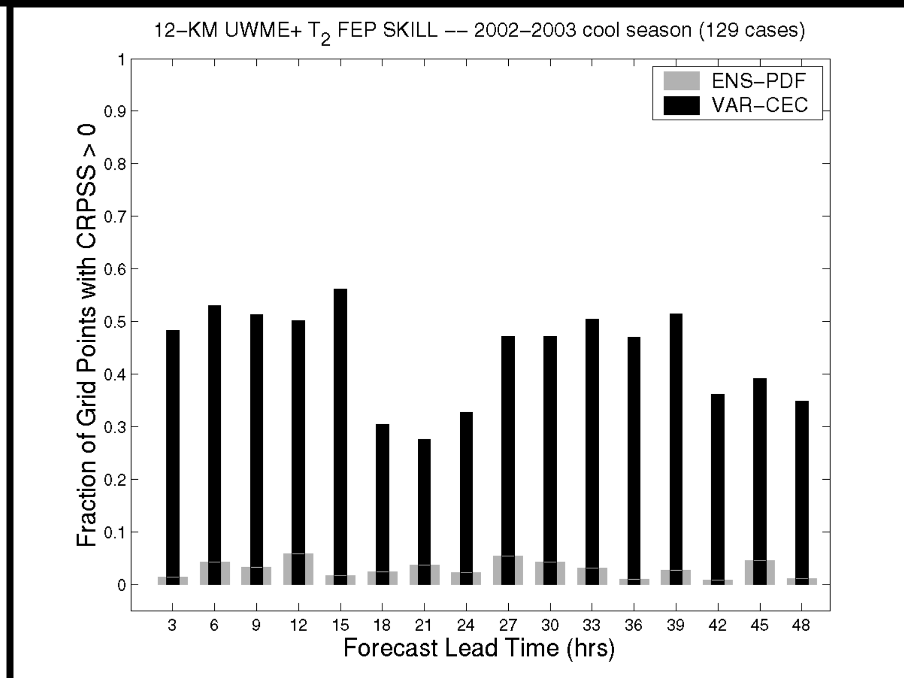
- VAR-CEC beats ENS-PDF handily
- VAR-CEC skill is generally small, but positive over 40-70% of the grid points through F24



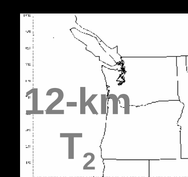
Real Probabilistic Error Forecast Skill

UWME+ (no bias correction)



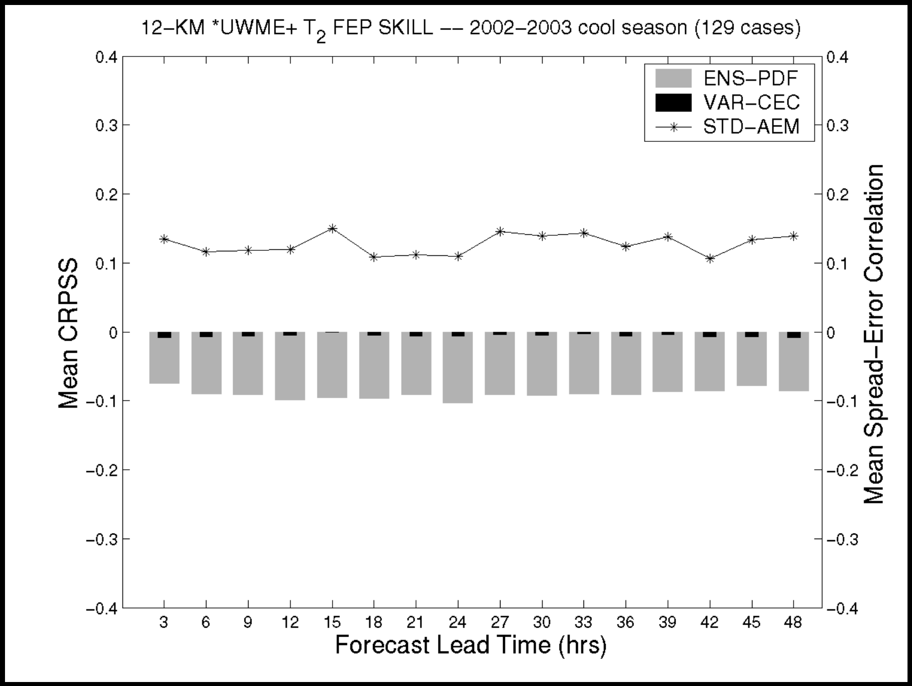


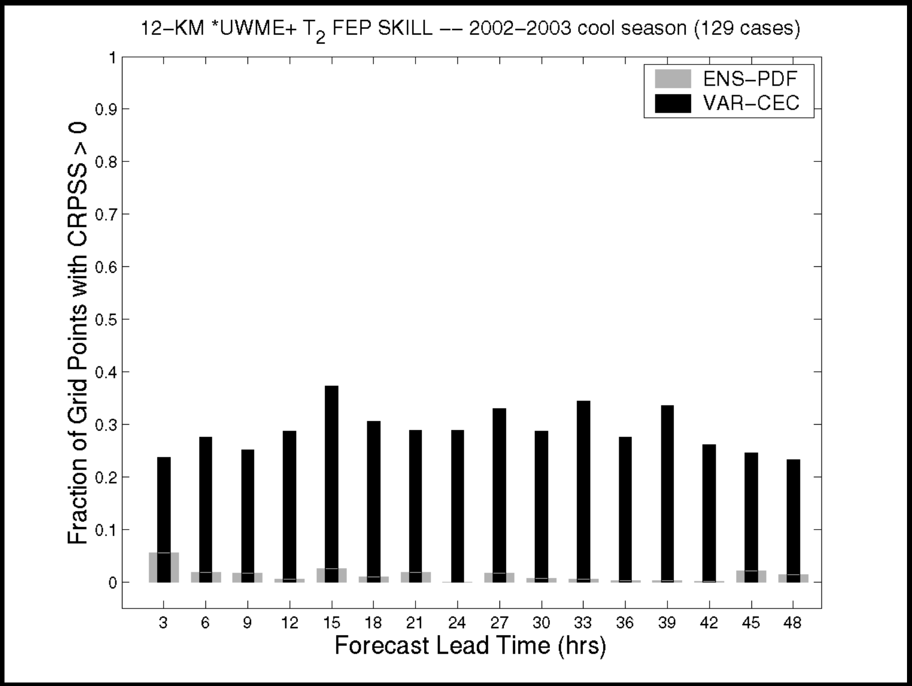
- UWME+ exhibits larger spread-error correlations
- Larger VAR-CEC skill (positive CRPSS into day-2 over 40-50% of the grid points)
- ENS-PDF improves (better raw PDF from UWME+)



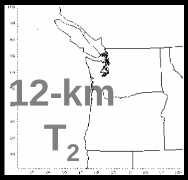
Effect of Post-Processing

UWME+ (14-day grid point bias correction)





- Bias correction reduces spread-error correlations and effectiveness of the VAR-CEC approach
  - Temporal spread variability decreases
- ENS-PDF closes the gap in performance, but is still below the baseline



Conclusions

- Traditional spread-error correlation is not the best way to describe the spread-skill relationship nor does it provide an adequate framework for making skillful forecast error predictions.
  - Probabilistic forecast error prediction is a good alternative.
  - If the true PDF is not well forecast, a spread-based CEC method provides a viable methodology.

- Continuous (categorical) measures of ensemble spread are most appropriate as forecast error predictors for end users with a continuous (categorical) cost function.

- Forecast error predictability is higher for cases with extreme spread, especially low spread cases.

- A simple bias correction improves ensemble forecast skill, but may also degrade forecast error predictability via the spread-based traditional and CEC methods.

QUESTIONS?

Contact Information

Eric P. Grimit, Ph.C.
University of Washington, Dept. of Atmospheric Sciences
Box 351640; Seattle, WA 98195
E-mail: epgrimit@atmos.washington.edu
Ph.: (206) 543-1456